

IZA DP No. 7116

**The Impact of Scoring Weights on Price and
Quality Outcomes: An Application to the
Procurement of Welfare-to-Work Contracts**

Pierre Koning
Arthur van de Meerendonk

December 2012

The Impact of Scoring Weights on Price and Quality Outcomes: An Application to the Procurement of Welfare-to-Work Contracts

Pierre Koning

*VU University Amsterdam,
IZA and Ministry of Social Affairs and Employment*

Arthur van de Meerendonk

Maastricht University

Discussion Paper No. 7116
December 2012

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

The Impact of Scoring Weights on Price and Quality Outcomes: An Application to the Procurement of Welfare-to-Work Contracts*

This paper assesses the effects of changes in scoring weights in auctions, using a unique sample of biddings of private Welfare-to-Work (WTW) organizations to reintegrate groups of unemployed and disabled workers. WTW-organizations did not only bid on prices, but also received points for three proxies of the quality of their services: their reputation, a description of their intended methodology to reintegrate workers and the fraction of job placements they considered likely to achieve. We exploit the fact that the scoring weights of these items changed in the time period under consideration, while the broad classification of ('gross') worker types remained similar over time. Our estimation results show that increases in the scoring weights of the quality items result in higher price bids, particularly of WTW-firms that are bidding for the first time as entrants. For the WTW-firms that won the procurements, we find a higher weight for both the reputation and the methodology item to contribute to the job placement of workers.

JEL Classification: D44, I38

Keywords: scoring auctions, welfare-to-work

Corresponding author:

Pierre Koning
Ministry of Social Affairs and Employment
Directorate ASE, C6-02
P.O. Box 90801
2509 LV The Hague
The Netherlands
E-mail: PKoning@minszw.nl

* The Dutch social benefit administration (UWV) is kindly acknowledged for giving access to data on the procurement of Welfare-to-work services. Also, participants of the European Association of Labour Economics (EALE) in Bonn in September 2012, as well as participants of the Lustrum of the Tinbergen Institute in Amsterdam on September 28 in 2012, are thanked for their comments to earlier drafts of the paper.

1. INTRODUCTION

The procurement of public services accounts for approximately 17% of GDP of EU countries, making it a key economic activity of many governments (OECD 2011). Most literature in this field focuses on one dimensional pricing bids as the dominant type of procurement. Auctions in the late 1990s for telecom and utilities markets have spurred the applied literature on pricing auctions for these markets (see e.g. Klemperer (2000) for an overview). At the same time, however, the majority of real life procurements involve combined price and quality auctions instead, that is, multi-criteria auctions or scoring auctions (GHK 2010). There has been relatively little attention for scoring auctions for which the “economically most advantageous bid” can be either a bid with the highest quality for a given price or a combination of prices and quality bids.

Following Che’s (1993) seminal paper, some theoretical studies have looked into scoring auctions. Asker and Cantillon (2008, 2010) are among the most cited ones. They show that scoring auctions dominate beauty contests, bargaining, and price-only auctions in most instances in terms of efficiency and social welfare. Examples of applied studies on scoring auctions include the (public) procurement of public works investments (Blancas et al. 2011), IT services (Albano et al. 2008) and electric power generation (Wolak 2003). These studies typically address designs that apply to oligopolistic markets and relatively large projects. They do not address the procurement of social services, which is often characterized by many bidders and many recurrent auctions.

This paper assesses the effects of scoring weights on the level of price bids and the (ex post) quality of services that are procured. For this purpose, we use a unique sample

of biddings of (private) Welfare-to-Work (WTW) organizations to groups of unemployed and disabled workers in the Netherlands for the period 2003-2007. WTW-organizations did not only bid on the price per worker trajectory, but also received point scores for their reputation (i.e. job placements in the past), a description of their plans to reintegrate workers (i.e. their ‘methodology’) and the number of job placements they considered likely to achieve. The latter three items could be considered as proxies for the quality of their services. Teams of social benefit caseworkers awarded WTW-organizations with points on these four items, and the total number of points determined the assignment of one or more worker groups to these organizations.

We exploit the fact that the scoring weights that were attached to these four items changed over the time period under investigation, while the broad classification of worker groups that were procured remained similar. That is, the social benefit administration changed the scoring weights for some worker groups, and for some not. This enables us to perform ‘difference-in-difference’ analyses for contracted worker group types on a number of relevant outcome measures. These outcome measures do not only include the price bids that were made, but also the (ex post) job placement rates of the worker cohorts that were assigned to the winning WTW-organizations. This enables us to assess possible tradeoffs between the prices and ex post quality of WTW-services that were provided. In this respect, we would expect that a higher weight for the quality items increases the job placement rates, but at the cost of higher price levels.

Our analysis contributes to the literature in three aspects. First, to our knowledge this study is the first to provide an empirical analysis of the effects of the procurement mechanism in the area of social services. To date, the literature has not addressed the

procurement of social services – like in the care sector, or, like in this case, WTW-trajectories. Since the 1990s, governments in an increasing number of countries have taken the step of contracting out welfare-to-work services to private providers. Australia has been a pioneer in this respect, and the Netherlands, the UK and a number of other countries have followed its example (studies into these developments include Bruttel 2005, DWP 2005, Struyven 2006, Tergeist and Grubb 2006, and Finn 2008.) This literature, dealing with contracting WTW-services, however, is mostly descriptive and addresses institutional settings. Other studies analyze the effects of performance incentives on program outcomes of contracted WTW-firms, but not the design of the procurement mechanism (Koning and Heinrich 2010; Heinrich and Choi 2007; Heckman and Smith 2002; Finn 2008; Courty and Marschke 2004; Behaghel et al. 2012).

The second innovative aspect of this paper is that we empirically assess the link between (ex ante) scoring rules and bidding behavior. In this respect, our analysis is close to Albano et al. (2008), who investigate the relationship between price/quality weights on firm bids of (public) IT-services. They find that bid prices and quality are correlated in a counter-intuitive way. That is, higher quality in their findings is associated with lower prices. To our knowledge, we are the first to investigate how *changes* in scoring weights of price and quality items (i.e. reputation, methods used and job placements) affect not only the bidding behavior of WTW-organizations, but also the performance outcomes of (winning) WTW-organizations. Onderstal (2009) and Van de Meerendonk and Onderstal (2010) have shown (both theoretically and in a laboratory experiment with professional bidders) that bidding on prices for job placement services can induce a ‘race to the

bottom', impairing quality. Hence, it is instructive to find if bidding on prices in a multi-criteria auction has similar effects.

The final contribution of our paper relates to the role of selection effects in the bidding prices of WTW-contracts, as well the job placement rate rates of procurement winners. Price and job placement rate responses to changes in the scoring weights may both be driven by behavioral effects and changes in the composition of WTW-firms that decide to bid on the worker groups. With the data at hand, we test for the importance of selection effects by exploiting the fact that most WTW-organizations (particularly the incumbents firms) bid on worker groups in multiple worker programs and over multiple years, and with different scoring weights. We thus can control for changes in the *types* of bidding firms (with high or low price levels) – or the sample firms that won the procurements (with a high or low ability to return workers back to work).

According to our analysis, most of the estimation results are in line with economic intuition. Increases in all weights for quality proxies result in higher price bids and higher prices of winning WTW-firms. These effects are largest for procurements that were organized for entrants (where reputation weights were set to zero). An explanation could be that investments of new WTW-firms to increase their quality levels were more costly, resulting in higher response effects. This would also explain why entrant firms with higher price levels become less likely to bid when the job placement and methodology weights are higher. In addition, we only find the reputation item and the methodology item to contribute to the (ex post) job placement of workers, as compared to the price weight. As we find no strong evidence for selection effects – i.e., WTW-firms with higher return to work rates in general becoming more likely to win the procurement – we

argue that the reputation score weight probably encouraged WTW-firms to improve their performance ('behavioral' effects). For the methodology item, WTW firms probably invested more effort and resources in their actual methodology, which may have resulted in a better performance on worker groups if the procurement was won.

The remainder of this paper is organized as follows. Section 2 describes the institutional settings of the procurement system of WTW-trajectories, together with the structure of the bidding and program outcome data we use. In Section 3 we present some theoretical considerations on the effects of scoring weights. Section 4 discusses the empirical implementation of the data, whereas Section 5 presents the estimation results. Section 6 ends with some discussion.

2. INSTITUTIONS AND DATA

In the Netherlands the Disability Insurance (DI) and Unemployment Insurance (UI) programs are mandatory for all workers. Both DI and UI are implemented by the social benefit administration (UWV). Similar to the New Deal in the United Kingdom and Employment Services in Australia, providers are expected to offer unemployed and partially disabled clients mediation, job training or subsidized employment within twelve months after the start of their benefit spell. In the period under investigation (2003-2007), the Dutch social benefit administration contracted out the delivery of these WTW-services to private job placement service providers only.

The procurement of worker groups

UWV applied the 'Economically Most Advantageous Tender' (EMAT) method, with the simultaneous selection of (winning) WTW-organizations and the awarding of job placement contracts for worker groups. The Invitation for Tender Document (IFT) explained the selection procedure, the worker group characteristics, the services that were demanded, conditions for payment, the award criteria (or: bidding items) and the maximum number of points that could be obtained for these items. In all procurements that were used, these bidding items included the price that was offered, a description of the method that would be used to re-integrate workers, the experience with the relevant worker group type, and the number of job placements the WTW-organization considered likely to achieve. The latter three items could be regarded as proxies for the quality of WTW-firms.

It is important to notice that WTW-firms did not know how their bids were translated exactly into procurement points. In this perspective, the procurement could be characterized as a ‘black box’, with WTW-firms that were informed only on the importance of the bidding items (i.e. their weights). WTW-firms also knew that the scoring of prices was non- monotonous, i.e., it was only to a certain minimum value that price reductions lead to higher scoring points. For price bids that were too low, UWV considered re-integration unfeasible. For the remaining (quantitative) bidding items, the scoring formula was monotonically increasing in the exact biddings.

To start with, the points for the *price bid* were determined by an overall bid for the reward in case of job placement and (to a much lesser extent) a price bid for a commencement – i.e. when an individual reintegration plan of a client was approved by the social benefit administration. Note that we use the overall price bid as the variable that is explained in our empirical analysis. Second, the *methodology* item entailed a four to six page description of the activities that should lead to placement into jobs. This description was valued according to a number of criteria, e.g. on showing relevant knowledge on the worker groups, the intended intensity of contacts with clients, the provision of training facilities, the presence of employer networks etc. Third, the track record (or: *reputation*) of WTW-service providers was measured with information on the share of non-commencements of previous worker groups (i.e. ‘provider-induced selection’, see e.g. Koning and Heinrich 2010), past job placement rates, and the time it took to complete individual reintegration plans in the past; this information was derived from past performance with similar worker groups. In order to allow firms without a track record to bid on worker groups as well, UWV also organized separate procurements

without reputation weights; this constituted about 25% of the procured contracts (213 out of 843). The cohort sizes of these procurement were much smaller than those for incumbent firms.

Finally, the *job placement* bid should always exceed a minimum value that was listed in the IFT document. WTW-organizations were not committed to realize these job placement bids if they won the procurement, so the contract for awardees contained no penalty for underperforming.

Data structure

This paper uses registered data from the Dutch social benefit administration on procured WTW programs. Table 1 presents descriptive statistics of the characteristics of worker groups that were assigned to WTW-providers in 2003-2007, stratified by calendar years. The upper panel of the table provides an overview of the structure of the data (the classification of the worker groups, discussed in detail below), whereas the remainder shows sample statistics of the procurements and (ex post) program outcomes of worker groups. Table A in the appendix to this paper presents additional information on the structure, size and characterization of clients. In what follows, we explain the structure of the data, following its different classification levels.

**TABLE 1: Summary statistics of procurement, bidding and job placement data
(standard deviations in parentheses)**

	2003	2004	2005	2006	2007	All
# Gross worker types (I)	13	8	13	6	7	19
# Worker program types (II)	22	15	25	9	14	85
# Procurements (III)	235	80	222	120	156	843
- For incumbents	156	110	166	72	126	630
- For entrants	79	0	56	48	30	213
# Worker groups (IV)	333	263	382	194	261	1,433
# Biddings (V)	1,666	424	1,909	936	1,083	6,018
# Worker groups per procurement (IV/III)	1.417 (0.819)	2.391 (1.643)	1.721 (0.924)	1.617 (0.676)	1.673 (0.903)	1.700 (1.030)
# Biddings per procurement (V/III)	7.089 (5.388)	3.855 (2.695)	8.600 (7.595)	7.800 (5.191)	6.942 (4.514)	7.139 (5.810)
# Bidding WTW-firms	102	48	89	96	103	218
Price offers (per bid)	4,914 (2,018)	4,446 (1,497)	3,150 (1,913)	4,062 (1,947)	3,772 (1,755)	3,978 (2,021)
Number of clients (per worker group)	127 (163)	105 (113)	47 (77)	142 (558)	186 (936)	99 (360)
No Cure No Pay (per worker group)	0.140 (0.347)	0.135 (0.341)	0.335 (0.472)	0.214 (0.410)	0.176 (0.381)	0.222 (0.416)
Job placement rate (per worker group)	0.343 (0.110)	0.369 (0.105)	0.406 (0.213)	0.425 (0.142)	0.365 (0.118)	0.373 (0.135)
<i>Averages per procurement: incumbents</i>						
Price weight	0.186 (0.047)	0.208 (0.049)	0.285 (0.079)	0.230 (0.047)	0.237 (0.024)	0.226 (0.064)
Job placement weight	0.245 (0.067)	0.174 (0.047)	0.110 (0.007)	0.111 (.)	0.126 (0.024)	0.167 (0.071)
Reputation weight	0.313 (0.027)	0.364 (0.070)	0.338 (0.061)	0.365 (0.041)	0.279 (0.103)	0.326 (0.072)
Methodology weight	0.256 (0.108)	0.254 (0.052)	0.266 (0.054)	0.294 (0.039)	0.358 (0.135)	0.281 (0.100)

**TABLE 1, continued: Summary statistics
(standard deviations in parentheses)**

	2003	2004	2005	2006	2007	All
<i>Averages per procurement: entrants</i>						
Price weight	0.298 (0.086)		0.354 (0.034)	0.333 (.)	0.342 (0.019)	0.327 (0.060)
Job placement weight	0.362 (0.096)		0.164 (0.007)	0.167 (.)	0.247 (0.007)	0.248 (0.107)
Methodology weight	0.340 (0.152)		0.482 (0.034)	0.500 (.)	0.411 (0.012)	0.425 (0.116)

(i) Gross worker types

The first classification level of worker types entails a broad characterization of workers that holds for more years in the data and includes 19 categories. These “gross worker types” are observed in multiple years (see the first row of Table 1 and Table A for a description of them). Similar to Koning and Heinrich (2010), we also use this classification level as control dummies (which are discussed in the next section). Gross worker types are characterized by specific client groups in the UI and DI scheme, but some smaller groups are defined by other strata. The gross worker groups include, amongst others, immigrants, those with good or bad job prospects (after being profiled by social benefit administration caseworkers), disabled workers with special impairments, and older workers or young disabled workers without any work history.

(ii) Worker program types; scoring weights

The second classification level entails a more detailed description of the contract types and activities that have been contracted out in subsequent years – we therefore call them ‘worker program types’. Gross worker types are redundant with respect to the worker program types – that is, the gross worker types can be derived from the worker program types. The choice of the (85) program types for the gross worker types has varied over the years, depending on the specific needs pertinent for gross worker groups at a given time. Examples include programs aimed at guiding workers into self-employment, job training, job hunting and job mediation. Contract conditions, particularly those regarding the scoring weights and payment schemes, NCLP and NCNP, also varied according to these contract types.

Table 1 also shows the average scoring weights per procured contract, both for the incumbent and entrant WTW-firms. These weights are defined as the maximum number of points that could be obtained for the concerning item, divided by the total number of points for all four items together. Over the years, there were a number of measures that were taken by UWV, which is mirrored by changes in the (average) scoring weights over time. Generally, the timing of these changes differed between (gross) worker groups, enabling us to use a difference-in-difference design to estimate its effects on prices and job placement rates (see Section 4). First, both for incumbent and entrant firms the weight of the job placement bid was lowered in 2004, as they learned that WTW-organizations probably did not feel a strong commitment to their biddings on this item. Consequently, the other three items have gained in importance since then. Second, in 2005 UWV needed to cut costs on their WTW-trajectories, and therefore increased the price weight. And

finally, in 2007 UWV increased the scoring weight that was attached to the methodology bid, while decreasing that of the reputation weights. The argument was that this change would result in more competition between WTW-providers.

(iii) Procurements

The third classification level considers the actual procurements that were organized at a specific region and a specific moment in time (843 in total). We stated earlier that the conditions for these procurements – that is, the scoring weights and payment schemes – were set at the level of worker program types, and that about 25% of the procurements was organized for bidders without a track record – i.e. the entrant firms. As can be seen in Table 1, the number of procurements (and the number of program types) varied substantially over the years, with a maximum number of 235 in 2003 and a minimum number of 80 in 2004. This pattern is however less marked in terms of variation in the number of worker groups (which is the next classification level).

(iv) Worker groups

The fourth classification level is the actual worker group that was procured. With multiple worker groups that were assigned to (winning) WTW-organizations for many procurements, there were 1,433 worker groups that were procured in total.¹ For each worker group that was procured, we also observe the ex post outcomes of the WTW-programs. These include the number of clients and the job placement rates that were

¹ Note that the total number of procurement winners (1,389) in our sample that we could identify is somewhat lower than the number of worker groups that was procured (1,433). This can be

realized in the time period described by the contract (ranging from one to two years). On average this rate was 37.3%.

(v) Biddings

The fifth classification level entails the (6,018) biddings that were made by 218 WTW-organizations in total. Figure 1 displays the distribution of biddings that were made per procurement. Note that there were no procurements for which no biddings were observed, so we only observe procurements that succeeded in assigning one or more WTW-organizations. Table 1 shows that the average number of biddings per procurement was 7.1. Over the years, this number remained quite stable, except for 2004, when the average number of biddings per procurement was only 3.8. This can largely be explained by the fact that the majority of worker groups in that specific year consisted of young clients with impairments. For these clients, the number of specialized WTW-firms was limited. For all biddings, we observe the price bids and points awarded to the price bid, the methodology, job placements and for reputation. When aggregating these points, we can infer the winning bids of WTW-organizations. The probability of a bid to be awarded was 24% on average.

explained by the fact that for some procurements the number of bidders was insufficient to assign all worker groups.

FIGURE 1: Distribution of bidding numbers per worker group

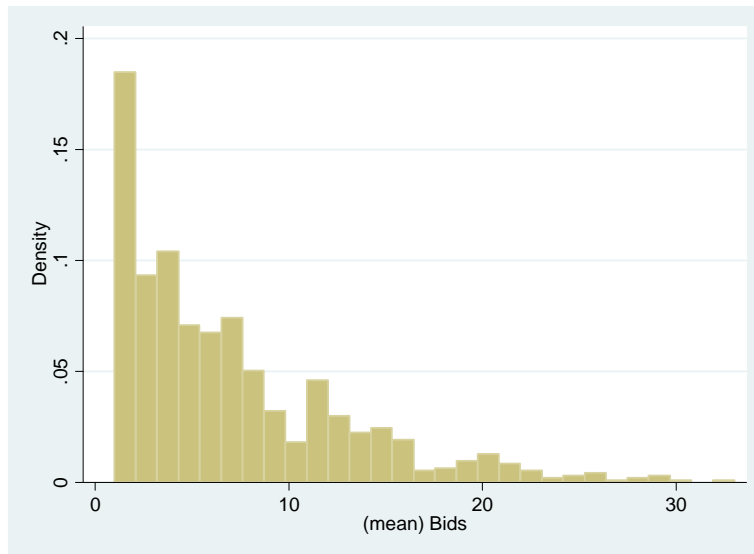


TABLE 2: Cross correlation of point scores on bidding items (normalized per procurement for incumbent WTW-firms); * indicates significance at 1%

	<i>Price points</i>	<i>Job placement points</i>	<i>Reputation points</i>	<i>Methodology points</i>
<i>Price points</i>	1.000			
<i>Job placement points</i>	0.103*	1.000		
<i>Reputation points</i>	0.074*	0.171*	1.000	
<i>Methodology points</i>	0.126*	0.253*	0.249*	1.000
<i>Price offer</i>	-0.371*	-0.068*	-0.028	-0.034

Table 2 shows the cross correlation of the bidding items that are adjusted for average points that were obtained per worker program type (i.e. the second classification level), together with the price bids. The table also shows the cross correlation of these point

scores with the adjusted price. As the table shows, all bidding scores are positively and significantly correlated. This suggests that more (less) efficient WTW-firms have higher (lower) scores on all items, rather than that firms either concentrate on prices or quality alone (i.e. substitution effects). As expected, we also find the price offer to be negatively correlated with the price points.

TABLE 3: Averages of WTW-firm types in markets, stratified according to number of yearly observations (standard deviations of means per WTW-firm)

	# Years in the sample				
	1	2	3	4	5
Number of WTW-firms	128	28	22	12	28
Regions	3.925 (0.231)	4.065 (0.487)	5.606 (0.453)	5.928 (0.451)	6.547 (0.167)
Gross worker groups	2.580 (0.353)	1.748 (0.295)	7.994 (0.848)	10.06 (1.033)	14.05 (0.594)
# Bids	4.898 (0.466)	7.643 (1.427)	33.32 (8.004)	48.50 (12.23)	133.1 (25.70)
Price offers	4,346.7 (262.0)	4,535.0 (397.3)	4,071.3 (302.0)	4,404 (341.6)	4,098 (139.8)
Fraction procurement winners	0.115 (0.022)	0.289 (0.067)	0.263 (0.060)	0.199 (0.032)	0.254 (0.026)

Finally, we analyzed the conduct and performance WTW-firms in more detail by tracking them over the years, regions and gross worker groups in our sample. Table 3 displays the number of years that WTW-firms are observed as bidders in our sample, resulting in five samples. Only 28 (13%) WTW-firms are observed in all years in our sample, whereas the majority of 108 firms (58%) is observed only in one year. Consequently, the ‘market’ for

UI and DI worker groups that were procured can well be characterized as highly dynamic, with many entrants and many firms that do not decide to bid for all procurements. This is particularly relevant for the first years of the procurement, when a large influx of new WTW-firms occurred.

Table 3 also presents the number of regions and gross worker groups the firms were bidding on; these statistics are averaged over WTW-firms. Clearly, the number of regions and gross worker groups that WTW-firms are bidding on increases in the number of years they are observed. In particular, the (small) sample of larger, incumbent WTW-firms is bidding on many gross worker groups and active in many regions. This contrasts to the large sample of smaller WTW-firms, that operate in less regions, appear to be more specialized and only bid in one of the procurement years in most cases. Particularly the latter sample of WTW-firms that is observed only in one year has markedly lower chances of winning a procurement.

3. THEORETICAL CONSIDERATIONS

Key to our analysis are the effect estimates of scoring weights on price bids and the job placement rates. From these estimates, we can infer possible tradeoffs between price and quality effects. In this section, we briefly set the stage by defining behavioral and selection effects as the drivers of these price and job placement effects.

Behavioral effects occur if WTW-firms adapt their price bids and/or exert less effort on job placement (if they are assigned a worker group). Less weight on the price score is associated with higher price bids, and, subsequently, more effort that is spent to return workers to work. Firms may also exert more effort if the reputation weight is increased, as higher job placement rates increase the prospect on winning future procurements. In light of the current institutional settings, such behavioral effects are especially relevant, as the procurement of WTW-trajectories was on incentive contracts. That is, winning bidders were awarded with the *opportunity* to receive payments for clients that were reintegrated. For NCNP contracts, payments were even made fully performance-contingent. Onderstal (2009) argues that WTW contracts are susceptible to moral hazard problems, as price bids in the procurement stage may become too low for WTW-providers to exert substantial effort on their clients. Putting weight on price bids may thus sharpen the tradeoff between prices and quality.

Selection effects occur if the scoring weights encourage or discourage certain types of WTW-firms to bid, or when certain types of WTW-organizations become more or less likely to win a procurement. In this respect, the information that comes from the bidding items can be considered as a screening device, particularly to select firms with sufficiently high return to work rates of workers. Regarding the effect of selection on

price bids, the impact of an increase in quality weights vis-à-vis the price weight is ambiguous. On the one hand, more efficient WTW-firms with higher prices may be attracted to the procurement, as their chances of winning may increase. On the other hand, if high quality WTW-firms are encouraged to bid, it depends on the correlation between price levels and quality levels what the impact on prices will be.

While this may seem trivial at first sight, the presupposed tradeoff between prices and quality critically hinges upon the accuracy of quality items. In this respect, all three quality items have their pros and cons. First, bids on *job placements* may in principle be informative on the quality of a WTW-firm if it reflects the capacity to return individuals to work – thus being a screening device to select the right firms – but they also can be characterized as “cheap talk” without any ex post commitment device. Moreover, for entrant firms the determination of this bid may be cumbersome, as there is no information on past performance to rely on. Second, the *reputation* item leaves no room for gaming and is objective by nature. The information on past experience with similar worker groups is however backward looking and not informative on the value added of the WTW-firm per se. And finally, points that are awarded for the *methodology* may be susceptible to subjectivity at the side of the social benefit administration. Still, if caseworkers are well informed on the key success factors of trajectories, the explanations on the methodology can be a way to uncover the expertise of WTW-organizations, thus being an appropriate screening device.

4. EMPIRICAL IMPLEMENTATION

Identification strategy

With the data at hand, we are able to assess the effects of scoring weights on prices and job placement rates. In doing this, it is essential to adequately control for unobserved heterogeneity and selection bias in our empirical analysis of scoring weights. For example, the assignment of weights to (worker program) types may be driven by (a priori) expected job prospects of the gross worker types. For both the price and outcome measures, we therefore use a difference-in-differences design that controls for calendar time (dummies) and gross worker types to obtain consistent estimates of the effects of scoring weights.

When following this difference-in-difference design, the effects from the scoring weights on prices and job placement rates are identified from two sources: (i) variation that is due to changes in *relative* scoring weights for gross worker groups in a given year, and (ii) remaining variation that is due to different worker program types for a given gross worker group and in a given year. Table B in the appendix to this paper shows that the first source of variation in the scoring weights is substantially larger than the second one, implying that the identification of effects essentially follows from interaction effects of calendar time and the gross worker groups. Stated differently, in our analysis changes in *relative* scoring weights of gross worker groups over time are the key exogenous variables. For this approach, we need the assumption to hold that business cycle or other trends affect all gross worker groups equally.

Price model²

To start with, we model the impact of the scoring weights on the log value of price offers as follows:

$$\text{Ln } P_{ijk}(r, t) = \alpha_p X_{ij} + \beta_p W_i + \sum_{rt} I_{rt} \zeta_{p,rt} + \sum_g I_g \eta_{p,g} + \varepsilon_{p,ijk} \quad (1)$$

where i ($i = 1, \dots, I$) is the worker program type, j ($j = 1, \dots, J$) indicates the procurement, k ($k = 1, \dots, K$) indicates the WTW-firm that is bidding, r ($r = 1, \dots, R$) is the region the procurement took place and t ($t = 1, \dots, T$) the year of the procurement. In our sample, we have 19 gross worker groups, $I = 85$ worker programs, $J = 843$ procurements, $R = 7$ regions, $T = 5$ years, and $K = 218$ WTW-organizations. \mathbf{X} is a matrix that includes the number of biddings for the procurement, (polynomials of) the number of worker groups that were procured and the minimum sample size of the worker group(s) that WTW-firms were informed on prior to the procurement. Vector α_b describes the impact of these variables. Matrix \mathbf{W} includes the weights of the job placement scores, the reputation score, the methodology scores and an indicator dummy for NCNP contracts. β_b describes the effect of these variables on the price bids. ζ_p is a vector that describes the impact of indicator dummies I_{rt} for each combination of a region and a year, so as to take account

² We argued earlier that selection and behavioral effects may affect the price biddings of WTW-firms. In contrast to the scoring points on the bidding items, the absolute price offer is the only source of information that can be used to assess the impact of scoring weights, as it is the only variable that can be compared *between* worker program types. In contrast, the assignment process of the score levels was set for each worker program separately, rendering any (direct) comparison between worker programs impossible.

of regional calendar time effects. Similarly, $\boldsymbol{\eta}_p$ is a vector for the gross worker dummies I_g that is included to control for gross)worker groups. Finally, we assume ε_p to be i.i.d. with $(0, \sigma_p^2)$. Equation [1] can be estimated by GLS, allowing for clustering effects for worker program types i .

To disentangle behavioral from selection effects, we exploit the fact that most WTW-organizations that participate in the procurement are observed at multiple (yearly) worker programs (and with different scoring weights). For this sample of WTW-firms, we can control for the fact that firms can vary in the price levels that they set. Obviously, this requirement holds for all incumbent firms in our sample. At the same time, most entrants that we observe in our sample are bidding at multiple gross worker groups at the same time, either as entrant or an incumbent firm. We thus re-specify the residual terms in equation [1] by allowing for fixed effects for WTW-firms k , indicated by vector $\boldsymbol{\theta}_q$:

$$\text{Ln } P_{ijk}(r, t) = \alpha_p X_{ij} + \beta_p W_i + \sum_{rt} I_{rt} \zeta_{p,rt} + \sum_g I_g \eta_{p,g} + \varepsilon_{p,ijk} \quad (2a)$$

with

$$\varepsilon_{p,ijk} = \theta_{p,k} + v_{p,ijk} \quad (2b)$$

where we assume v_p to be i.i.d. with $(0, \sigma_p^2)$. The comparison of the coefficient estimates of the scoring weights of equations [1] and [2] provides us with a natural test on selection effects. That is, more or less expensive WTW-firms may decide to continue or abstain from bidding in particular years, causing the (average) price level to change.

Job placement rate model

Using a similar setup with similar indices as for the price equation, we specify the job placement rates Q for the sample of WTW-organizations that were awarded with worker groups as follows:

$$Q_{ijk}(r, t) = \alpha_q X_{ij} + \beta_q W_i + \sum_{rt} I_{rt} \zeta_{q,rt} + \sum_g I_g \eta_{q,g} + \varepsilon_{q,ijk} \quad (3)$$

In equation [3], the coefficient values for the job placement rates model are now denoted by the subscript “ q ”. Similar to equation (1), we assume ε_q to be i.i.d. with $(0, \sigma_q^2)$. Equation [3] can be estimated with GLS, while allowing for clustering effects in worker program types i . To increase the efficiency of our estimates, we also use number of participants in the worker group as relative frequency weights. Take notice that matrix X with explanatory variables per procurement is reduced only to the log value of the number of clients in the worker group. As to the other variables (e.g. the information on the minimum worker group size), we argue that these were only relevant in the procurement stage.

As to the interpretation of the scoring weight coefficients, once more the question arises to what extent outcomes are driven by behavioral and selection effects. Behavioral effects stem from changes in the effort that is devoted to worker groups, whereas selection effects are due to changes in the efficiency of winning WTW-firms. Similar to equation [2], we can extend the model with fixed effects for WTW-firms, denoted by θ_q . This yields the following equations:

$$Q_{ijk}(r, t) = \alpha_q X_{ij} + \beta_q W_i + \sum_{rt} I_{rt} \zeta_{q,rt} + \sum_g I_g \eta_{q,g} + \varepsilon_{q,ijk} \quad (4a)$$

and

$$\varepsilon_{q,ijk} = \theta_{q,k} + v_{q,ijk} \quad (4b)$$

with v_w to be i.i.d. with $(0, \sigma_q^2)$. When supposing that changes in the job placement rates to scoring weights are fully driven by changes in the selection of types WTW-firms, we would expect to find no effect of the scoring weight for specification [4]. At the other extreme, when the coefficient values of the scoring weight remain unaffected, changes in job placement rates stem from behavioral effects, that is, WTW-firms changing their effort level as a response to changes in scoring weights.

5. ESTIMATION RESULTS

Price model

Table 4 displays the estimation results for the equations [1] and [2] for various samples of our data. In particular, we estimate coefficient values for (i) equation [1] for all price bids that are made by the WTW-organizations; (ii) equation [1] for the sample of procurement winners; (iii/iv) equations [1] and [2] for the sample of (procurements for) entrants; and (v/vi) equations [1] and [2] for the sample of (procurements for) incumbent firms.

Before discussing the results on the weights of the bidding items, we should bear in mind that contracts were either No Cure Less Pay (NCLP) or No Cure No Pay (NCNP). As NCNP was fully performance contingent and NCLP for 50% fixed, this means that, *ceteris paribus*, NCNP price bids had to exceed NCLP price bids to offset the lower expected payment. This is confirmed by the substantially higher price bids for the NCNP worker groups for all specifications that were estimated (with coefficient estimates ranging from 9 to 30%). Further, note that the regional calendar time dummies, (polynomials of) the number of worker groups per procurement and (polynomials of) the minimum number of workers were significant and substantial in all specifications.

**TABLE 4: Coefficient estimates for models of (log) price biddings
(2003-2007; standard errors in parentheses; *, ** and *** denote significance at 10,
5 and 1%, respectively)**

	(i)	(ii)	(iii)	(iv)	(v)	(vi)
Sample	All	Procurement winners	Incumbents	Incumbents	Entrants	Entrants
WTW firm fixed effects				X		X
Number of fixed effects				149		149
Reference group: price weight	—	—	—	—	—	—
Job placement weight	1.315*** (0.256)	0.759** (0.320)	0.957*** (0.293)	1.426*** (0.270)	3.664*** (0.766)	5.321*** (0.672)
Reputation weight	0.688*** (0.146)	0.817*** (0.191)	1.320*** (0.287)	1.698*** (0.269)	—	—
Methodology weight	1.560*** (0.145)	1.594*** (0.212)	2.046*** (0.211)	2.261*** (0.195)	2.139*** (0.447)	3.270*** (0.388)
No Cure No Pay	0.292*** (0.016)	0.256*** (0.033)	0.303*** (0.017)	0.299*** (0.017)	0.091*** (0.034)	0.033 (0.027)
Log number of biddings	0.026*** (0.010)	-0.015 (0.013)	0.029** (0.012)	0.010 (0.012)	-0.059** (0.029)	-0.070*** (0.023)
Gross worker group dummies (# = 19)	P = 0.000	P = 0.000	P = 0.000	P = 0.000	P = 0.000	P = 0.000
Regional calendar time dummies (# = 33)	P = 0.000	P = 0.000	P = 0.000	P = 0.000	P = 0.008	P = 0.000
# Worker groups procured (four polynomials)	P = 0.000	P = 0.000	P = 0.000	P = 0.000	P = 0.004	P = 0.000
Minimum # workers (four polynomials)	P = 0.000	P = 0.000	P = 0.000	P = 0.001	P = 0.000	P = 0.000
# Observations	5,906	1,389	4,562	4,562	1,344	1,344
Fraction var. due to FE				0.833		0.651
R-squared	0.669	0.724	0.704	0.868	0.347	0.687

In line with economic intuition, we find positive and significant coefficient values for all quality weights and for all specifications. For the total sample, an increase of 10%-point of the scoring weight for the job placement weight vis-à-vis the price scoring weight results in an increase in the average price bid of about 13%. Analogously, the estimated impacts of increases in the reputation and methodology weights are 7% and 16%, respectively. It thus seems that WTW-firms do respond to reduced price weights by increasing their price bids, probably because there is less price competition. WTW-firms may also exert more effort in terms of their plans to reintegrate workers, and therefore raise prices if the methodology weight increases. This may explain why the response effect is relatively high for the methodology weight.

When comparing these outcomes to the sample of procurement winners determining the actual prices that were set (i.e. column (ii)), we find no significant differences for any quality weight. This suggests that responses in the price bidding behavior of WTW-firms on these items are roughly equal among those who won the procurement and those who have not.

Columns (iii) and (iv) of Table 4 report the coefficient estimates for the sample of procurements that were organized for the incumbent firms. Similar to the full sample, bidding WTW-firms seem most responsive to increases of the methodology weight. Although the coefficient estimates of the quality scores change in similar directions, the inclusion of WTW-firm specific effects does not change these results significantly. We thus conclude that changes in price bids of incumbents are primarily driven by price setting behavior of similar WTW-firms, rather than by changes in the composition of bidders.

Interestingly, for entrants we find markedly higher effects on prices of the job placement weight bid than for incumbent bidders. In particular, we estimate an elasticity estimate of 3.664 (0.766) for entrants, compared to 0.957 (0.293) for the incumbent firms. The effect is even higher if we control for WTW-firm effects, amounting to an elasticity of 5.321 (0.672). When combining these results, it seems that entrants experienced and/or perceived procurements with high job placement weight as relatively costly, inducing entrant firms to bid higher prices, and at the same time discouraging particularly entrants with higher cost (and price) levels to bid. The former effect can be characterized as an incentive effect, whereas the latter as a selection effect. In a way, the high response to job placement weight increases is surprising, as there was no ex post commitment to the job placement bids. Higher job placement bids did not force WTW-firms to increase their costs if they would win the procurement. It may be that particularly the inexperienced entrant WTW-firms were not aware of this and perceived the job placement bid as binding.

Job placement rate model

Table 5 displays the coefficient estimates for the model for the job placement rates per worker group that was contracted. We present four model variants: (i) the model without WTW-firm fixed effects, as in equation [3]; (ii/iii) the same model for the sample of incumbents and entrants, respectively; and (iv) the model with fixed effect for WTW-firms, estimated for the sample of incumbent firms (i.e. equation [4]). Note that the number of entrants in our sample (215) is too limited to obtain reliable estimates with WTW-firm fixed effects.

**TABLE 5: Coefficient estimates for job placement rate model
(2003-2007; standard errors in parentheses; *, ** and *** denote significance at 10,
5 and 1%, respectively)**

	(i)	(ii)	(iii)	(iv)
Sample	All	Incumbents	Incumbents	Entrants
WTW-Fixed effects			X	
Number of fixed effects			75	
Reference group: price weight	—	—	—	—
Job placement weight	-0.097 (0.106)	0.134 (0.121)	0.260** (0.129)	-0.710 (0.525)
Reputation weight	0.119** (0.060)	0.331*** (0.099)	0.403*** (0.097)	—
Methodology weight	0.310*** (0.074)	0.464*** (0.083)	0.364*** (0.082)	-0.099 (0.298)
No Cure No Pay	0.062*** (0.010)	0.071** (0.010)	0.006 (0.015)	0.044*** (0.017)
Log worker group size	0.014*** (0.004)	0.014*** (0.004)	0.030** (0.014)	0.005 (0.007)
Gross worker type dummies (# = 19)	P = 0.000	P = 0.000	P = 0.000	P = 0.000
Regional calendar time dummies (# = 34)	P = 0.000	P = 0.000	P = 0.000	P = 0.087
# Observations	1,351	1,136	1,136	215
Fraction var. due to FE			0.550	
R-squared	0.788	0.804	0.860	0.739

To start with, we find the (set of) regional calendar time dummies to be significant for all four model variants. Moreover, the coefficient estimates of NCNP (vis-à-vis NCLP) contracts are positive in all cases. These results are in line with Koning and Heinrich (2010), who report somewhat lower effects for similar data but investigate the time period from 2002 to 2005 instead.

Regarding the scoring weights, it seems that increases in the methodology item's weight contributes mostly to the placement of workers into jobs. For the full sample, an increase in this weight with 10%-point – equaling about one standard deviation of this weight – results in an increase in the job placement rate with 3.1%-point. For procurements with incumbents this estimate is higher than for those with entrants. Comparing columns (ii) and (iii), we do not find strong evidence of the presence of selection effects that influence our results – that is, that an increased methodology weight helps in selecting firms that are more effective for other worker groups as well. Obviously, one should bear in mind here that this result holds for the (limited) changes in the weight that are observed in our data, and only for the sample of firms that won the procurement. It thus still may have been that the quality weight performed well as a first rudimentary screening device, to cancel out firms with low quality levels.

For the sample of incumbent firms we also find increases in the reputation weight to improve the job placement of workers, with coefficient estimates that are similar to those obtained for the methodology weight. This effect persists if we control for WTW-firm fixed effects, which again suggests that the reputation score is not used as a screening device. The alternative explanation is that higher job placement rates increase

the likelihood for WTW-firms to win procurements in the future, thus being an incentive for them to increase their effort levels.

Finally, according to our results there is no systematic evidence pointing at positive job placement effects of the job placement weight. It is only for the model with WTW-firm fixed effects that we find a weakly significant positive effect of this variable, suggesting that firms show higher effort levels of this weight increases. The effect is however offset by negative selection effects. That is, compared to the price bid item, this item does not seem informative on the quality of WTW-firms, measured in terms of ex post job placement rates.

Selection or incentive effects?

The picture that emerges from these results is that the reputation and methodology items are effective in increasing the placement in jobs of workers. More strikingly, however, it seems the scoring points on these items were not effective as a screening device to determine which firms are most efficient – at least not in the range of weights that we observe in our data. To get some more idea on the importance of selection and incentive effects, we tested the robustness of these results by running several additional regressions. Table 6 presents the results of these regressions for the scoring weight estimates for the sample of incumbent firms. In this table, we only present coefficient values for the scoring weight coefficients, together with some additional controls that are used in the robustness checks. As reference points to our earlier results, columns (i) and (ii) present the model outcomes for this group for equations [3] and [4], respectively.

TABLE 6: Robustness checks on weight coefficient estimates of job placement rate model, for the sample of incumbent WTW-firms^a
 (*, ** and *** denote significance at 10, 5 and 1%, respectively; standard deviations in parentheses)

	(i)	(ii)	(iii)	(iv)	(v)	(vi)
Fixed effects						
- WTW-firms			X			
- WTW x gross worker groups				X	X	
- WTW x calendar year						
# Fixed effects		75		227	163	
Reference group: price weight	—	—	—	—	—	—
Job placement weight	0.134 (0.121)	0.260** (0.123)	0.088 (0.117)	0.214 (0.154)	0.096 (0.145)	0.118 (0.125)
Reputation weight	0.331*** (0.099)	0.403*** (0.097)	0.352*** (0.102)	0.442*** (0.114)	0.370*** (0.124)	0.388*** (0.101)
Methodology weight	0.464*** (0.084)	0.364*** (0.082)	0.319*** (0.089)	0.516*** (0.102)	0.289*** (0.097)	0.415*** (0.086)
Price level, log value			0.014*** (0.004)			
Price level, log squared			1.414*** (0.158)			
Price bid points ^b						0.0026 (0.00029)
Job placement bid points ^b						-0.0068** (0.0030)
Reputation bid points ^b						0.0033 (0.0034)
Methodology bid points ^b						0.00063 (0.0030)

^a Note that the variable coefficients of NCNP, log worker group size and P-values of the regional calendar time dummies and gross worker type dummies are excluded.

^b The bid points on these items are normalized to Z-values, so as to obtain coefficient values describing the effect of one standard deviation.

A straightforward explanation for our results is that price bids were higher for procurements with high quality scoring weights. With rewards that were based on performance contracting (NCLP and NCNP), the incentive for WTW-firms to return clients to work thus was higher too. To test for the importance of price effects, column (iii) of Table 6 reports the model outcome of equation [3] with the log price level and the squared log price level as additional controls. We then find a small decrease in the coefficient value of the methodology weight. Furthermore, as expected, we the job placement rate increase in the price level – with an average elasticity coefficient of 0.06. When combining this outcome with the outcomes of equation [1] for the price model, the absolute impact of scoring weight changes on the job placement rates through price effects is only limited, amounting to 0.12%-point ($=2.06*0.06$). This small impact is in line with the (small) decrease in the methodology weight coefficient.

As a second potential explanation, one may argue the quality of WTW-firms which is screened upon is not well captured by the current specification. The assumption that WTW-firms are equally effective for different worker groups and during different time periods may be restrictive, so that we cannot adequately control for selection effects. We therefore tried more flexible specifications of fixed effects, so as to assess the importance of alternative forms of selection effects. First, we re-estimated equation [4] while allowing WTW-firm fixed effects that interact with gross worker groups (i.e. column (iv)). This specification recognizes the idea that WTW-firms may have comparative advantages or disadvantages for specific gross worker types. Second, we re-estimated the model with WTW-firm fixed effects that were allowed to vary over the years in our sample, while assuming that the effects were constant for gross worker types

(i.e. column (v)). This specification takes into account that the general performance of WTW-firms may improve or deteriorate over time.

Table 6 shows that both these model extensions do not substantially affect the coefficient values for both the reputation and the methodology weight that are estimated for the sample of incumbent firms. As a final test, we therefore re-estimated equation [3] once more for the sample of incumbent WTW-firms, but with the relative bid score points (measured as Z-values) as additional controls; the results of this are presented in column (vi). The intuition behind this approach is that the inclusion of the four point scores should take account of possible selection effects due to screening. In particular, the point scores were the actual information that was used to select the sample of winning firms. Interestingly, we find small and insignificant – and in one case even negative – coefficient effects for all point scores. This suggests that the point scores were not informative on the ex post quality of (winning) firms in general. More specifically, for the methodology score we find a shift of one standard deviation in the distribution of score points to result in an increase in the job placement rate of 0.06 %-point. Not surprisingly, compared to the outcomes in column (i), the estimates for the weight coefficients of the variables are hardly affected.

We conclude that there is no strong evidence pointing at selection effects for the job placement model, not even for the methodology bid. As another hypothesis, one may thus argue that there are spillovers from writing reintegration plans to the subsequent performance of firms. By investigating the specific needs of potential client groups, WTW-firms may be better prepared to make a success out of their trajectories.

6. CONCLUSION

In this paper we assessed the effects of scoring weights on the level of price bids and the (ex post) quality of services that are procured. For this purpose, we used a unique sample of biddings of (private) Welfare-to-Work (WTW) organizations to groups of unemployed and disabled workers in the Netherlands for the period 2003-2007. WTW-organizations did not only bid on the price per worker trajectory, but also received point scores for their reputation (i.e. job placements in the past), a description of their plans to reintegrate workers (i.e. their ‘methodology’) and the number of job placements they considered likely to achieve. We exploit the fact that the scoring weights that were attached to these four items changed over the time period under investigation, while the broad classification of (‘gross’) worker groups that were procured remained similar.

The overall picture of our analysis is that most of our results are in line with economic intuition. All three quality weights increase the price bids of WTW-firms, suggesting that lower price weights reduce price competition. The price responses are relatively substantial for the entrant firms, for which separate procurements were organized. It is likely that these firms are faced with relatively higher costs to draft accurate reintegration plans than incumbents do. Related to this, our results suggest that (potential) entrants with higher cost levels are discouraged to bid if the methodology weight increases. In addition, for the methodology and reputation item we find higher scoring weights to improve the job placement rates of firms that have won the procurement. The changes that result from these weights on both prices and job placement rates suggest the presence of price and quality tradeoffs.

Next to these findings, there are two issues that deserve special attention. First, it is questionable if firms – particularly entrants – were fully informed on the procurement process in all years in our sample. For the sample of entrants, we find substantial price response estimates for the job placement rate bid. Although these job placement rate bids were not binding ex post, entrants may not have been aware of this and therefore increased their price bids to cover costs that were expected to be higher. This effect may particularly apply to the first year of the procurements in our sample, when the job placement weight was markedly higher than those in the years thereafter.

The second issue relates to the role of selection effects in determining the job placement performance of firms. Particularly for the methodology weight, we would expect the point bids on this item to be informative on the (ex ante) quality of firms, thus being a screening device in the procurement stage. Using a variety of specifications capturing selection effects, we do however not find any supporting evidence. One should bear in mind here that this result holds for the (limited) changes in the weight that are observed in our data, and only for the sample of firms that won the procurement. It thus still may have been that the quality weight worked well as a first rudimentary screening device, to cancel out firms with low quality levels. Still, the improved performance of the sample of winning firms is more likely to result from incentive effects. Probably, there were spillover effects from writing accurate reintegration plans to the subsequent performance of firms. By investigating the specific needs of potential client groups, WTW-firms were better prepared to make a success out of their trajectories.

REFERENCES

- Albano, G., Dini, F., Zampino, R., 2008, Bidding for Complex Projects: Evidence from the Acquisitions of IT Services, Fondazione Eni Enrico Mattei, NOTA DI LAVORO 86.2008.
- Asker, J., Cantillon, E., 2008, Properties of scoring auctions. *RAND Journal of Economics* 39(1), 69-85.
- Behaghel, L., Crépon, B, Gurgand, M., 2012, Private and Public Provision of Counseling to Job-Seekers: Evidence from a Large Controlled Experiment, IZA Discussion paper 6518.
- Bergman, M., Lundberg, S., 2011, Tender Evaluation and Award Methodologies in Public Procurement. Umea University discussion paper, 821.
- Blancas, L., Chioda, L., Cordella, T., Oliveira, A., Várdu, F., 2011, Do Procurement Rules Impact Infrastructure Investment Efficiency? An Empirical Analysis of *Inversao das Fases* in Sao Paulo State. Policy Research Working Paper 5528, The World Bank.
- Bruttel, O., 2004. Contracting-out and Governance Mechanisms in the Public Employment Service. Wissenschaftszentrum Berlin für Sozialforschung (WZB), Discussion paper 2005-109.
- Courty, P., Marschke, G., 2004. An Empirical Investigation of Gaming Responses to Performance Incentives. *Journal of Labor Economics* 22(1), 23-56.
- Finn, D., 2008. The British 'welfare market'. Lessons from contracting out welfare to work programmes in Australia and the Netherlands, University of Portsmouth, <http://www.jrf.org.uk/sites/files/jrf/2306-welfare-unemployment-services.pdf> .
- GHK, 2010. Evaluation of SME's Access to Public Procurement Markets in the EU. Report to DG Enterprise and Industry. Contract No ENTR/E4/04/93/1/07/11.
- Heckman, J., Heinrich, C. J., Smith, J., 2002. The Performance of Performance Standards. *Journal of Human Resources* 38(4), 778-811.
- Heinrich, C.J., Choi Y., 2007. Performance-Based Contracting in Social Welfare Programs, *The American Review of Public Administration* 37(4), 409-435.
- Koning, P., Heinrich, C., 2010, Cream-skimming, Parking and Other Intended and Unintended Effects of Performance-Based Contracting in Social Welfare Services. La Follette School Working Paper No. 2010-021.
- McAfee, R., McMillan, J., 1987, Competition for agency contracts. *RAND Journal of Economics* 18(2), 296-307.
- OECD, 2007. Bribery in Public Procurement: Methods, Actors and Counter-Measures.

- OECD, 2011. Centralised Purchasing Systems in the European Union, Sigma Papers, No. 47, OECD Publishing.
- Onderstal, S., 2009. Bidding for the unemployed: An application of mechanism design to welfare-to-work programs. *European Economic Review* 53, 715-722.
- Struyven, L., Streurs, G., 2005. Design and redesign of a quasi-market for the reintegration of jobseekers: empirical evidence from Australia and the Netherlands. *Journal of European Social Policy* 15(3), 211-229.
- Van de Meerendonk, A., Onderstal, S., 2010, If You Pay Peanuts...: A Laboratory Experiment on Reward Schemes in Employment Service Contracting, *European Journal of Social Security*, 12(4), December 2010, 272-288.
- Wolak, F., 2003. Measuring Unilateral Market Power in Wholesale Electricity Markets: The California Market, 1998–2000. *American Economic Review* 93(2), 425–430.

TABLE A: Description and size of gross worker groups

	<i># Worker program types</i>	<i># Procured worker groups</i>	<i># Bids</i>
Disabled workers, migrants	7	74	271
Disabled workers, mild impairments	3	79	251
Disabled workers, mental impairments	5	109	550
Disabled workers, severe impairments	12	121	898
Disabled workers, health and safety services*	1	22	94
Mediation services, assessments, job hunting	8	121	912
Debt services*	1	15	20
Disabled workers, preventative activities for part time workers	4	38	256
Disabled workers, kidney patients	2	3	3
Telephone services for disabled workers	4	36	59
“Second chance” trajectories	5	101	603
Trajectories into self-employment	2	34	227
Disabled workers, visual and hearing impairments	2	13	18
Unemployed workers, migrants	3	58	196
Unemployed workers, highly educated	3	23	46
Unemployed workers, older than 50 years	4	106	306
Unemployed workers, bad job prospects	4	122	518
Disabled school leavers	14	186	624
Total	85	1,456	6,018

TABLE B: Variance decomposition of procurement, prices and job placement data

	<i>calendar time t</i>	<i>gross worker groups</i>	<i>calendar time * gross worker groups</i>	<i>program types</i>	<i>residual</i>
# Biddings per procurement	0.060	0.377	0.155	0.092	0.316
<i>Scoring weights: incumbent firms</i>					
Price weight	0.323	0.412	0.190	0.063	–
Job placement weight	0.620	0.203	0.165	0.013	–
Reputation weight	0.182	0.370	0.345	0.104	–
Methodology weight	0.159	0.462	0.354	0.025	–
<i>Scoring weights: entrants</i>					
Price weight	0.153	0.507	0.298	0.041	--
Job placement weight	0.711	0.195	0.091	0.003	
Methodology weight	0.357	0.422	0.211	0.010	
% NCNP contract	0.027	0.569	0.166	0.248	–
Price offers (log value, per bid)	0.142	0.471	0.041	0.107	0.248
Job placement (ex post)	0.037	0.513	0.032	0.051	0.370