CAMBRIDGE
UNIVERSITY PRESS

ARTICLE

Outcome-based contracting and gaming practices in marketised public employment services. Dilemmas from the Italian case

Gianluca Scarano[iD]

University of Turin, Turin, Italy
Email: gianluca.scarano@unito.it

Abstract

Concerns about the unintended effects of marketised public employment services are increasingly expressed because there is mounting evidence that such services are frequently characterised by various gaming practices on the part of their providers. To prevent these unintended consequences, payment-by-result approaches have been progressively strengthened.

The aim of the research reported in this article was to investigate the extent to which such approaches are able to make service providers accountable for client outcomes. The study used two Italian regional cases in order to compare different and alternative contracting arrangements: Lombardy (outcome-based payments), and Emilia-Romagna (fixed payments). Drawing on rich administrative databases, the analysis relied on a quantitative methodology based on propensity score matching and logistic regressions. Even if outcome-based contracting can make service providers financially accountable for the service outcome, the results of the analysis show that it does not sufficiently prevent gaming practices, resulting in inequity among the services provided.

1. Introduction

Over the past three decades, both micro and macro studies have greatly improved the understanding of public employment services (PES) and activation policies. In response to the high levels of long-term unemployment and constraints resulting from the fiscal-austerity measures implemented, national governments have increasingly encouraged the imposition of market forces on PES in order to drive innovation, efficiency and specific know-how (Fay, 1997; Greer *et al.*, 2017; Sol and Westerveld, 2005),

---

Institutional affiliation where the work was conducted: University of Milan.

Contracting arrangements have increasingly resulted in the creation of quasi-markets in the delivery of employment services. Over the years, national reforms have been pursued on the basis of high expectations regarding the ability of quasi-markets to deliver effective and efficient services. However, such high expectations of quasi-markets have not always been fulfilled. Concerns about the unintended effects of quasi-marketised employment services have also been expressed because there is evidence that the providers of such services frequently engage in various gaming practices, especially when dealing with clients faced by complex employment barriers. The most recurrent of these practices are creaming and parking (Carter and Whitworth, 2015; Greer *et al.*, 2018). Such considerations have extended the debate to include the kind of market governance instruments that should be applied to prevent gaming practices (Bredgaard and Larsen, 2008; Bruttel, 2005). Accordingly, national reforms have progressively shifted towards the introduction and strengthening of performance signals in payment structures, in exchange for giving private service providers greater flexibility in choosing the service delivery method to be used (Finn, 2010, 2012; Tomkinson, 2016). Nevertheless, outcome-based contracting may generate concerns about the inequitable treatment of 'harder-to-help' clients as well (Carter and Whitworth, 2015).

The aim of the research reported in this article was to investigate the extent to which the outcome-based contracting model could make service providers accountable for client outcomes by comparing that model with different and alternative contracting arrangements.

In fact, even if several studies have examined outcome-based contracting, few studies have empirically compared it with other contracting models (Jantz *et al.*, 2018). In particular, the present study focuses on the Italian context. One of the most significant aspects of Italy is that it is characterised by regionalised PES. This allowed the comparison of regions that are similar to each other from the economic viewpoint but very different from each other from the perspective of contracting models in PES. In what follows, the two Italian regions Lombardy and Emilia-Romagna are compared. They can be considered highly dissimilar cases in that Lombardy has an outcome-based quasi-marketised contracting model while Emilia-Romagna has a fixed-payment-based public-oriented contracting model.

The availability and quality of rich administrative databases enabled the use of a statistical approach for a broader assessment of the significant differences that may emerge between the two aforementioned contracting models.

The rest of this article is organised as follows. The next section discusses the theoretical debate concerning outcome-based contracting, and it presents the two cases that were analysed in the research reported. The third section introduces the data and methods that were applied. The final section presents the results of the study, discusses those results, and makes some concluding remarks.

2. Background

2.1. Outcome-Based Contracting: Purposes and Potential Criticisms

The introduction of quasi-markets into PES stemmed from the pressure to improve their efficiency, effectiveness, and responsiveness to individual needs, and to provide

more choices for the users (Webster and Harding, 2000). A typical quasi-market involves a clear-cut split between service purchasers and service providers which reproduces the market's characteristic of independent agents competing with each other. Public agencies can continue to act as one of the competing actors, while service providers are not necessarily private for-profit organisations (Struyven and Steurs, 2004).

This context is frequently characterized by the adoption of a 'voucher system' which enables clients to 'purchase' services. According to this system, funds go directly to the clients, who can 'shop around' different service providers in the market (Struyven and Steurs, 2004; Le Grand, 2006),

The central concern is still the difficulty of resolving the tension between efficacy and equity and the risk of moral hazard, particularly because providers will use contracts to their advantage by focusing on those clients who are easiest to help in order to maximise profits (Koning and Heinrich, 2013; Carter and Whitworth, 2015; Greer et al., 2018; Considine et al., 2020). In particular, when a large degree of information asymmetry exists among the government, the service provider and the clients, the contracted provider knows that the quality of the services that it will deliver is difficult to determine. Hence, it may shift the public resources to initiatives where the assessment of its net employment contribution as a provider may be confounded by external factors related to the business cycle or to labour-demand trends (Hill, 2013). The economic interest, considered in its original purpose as a vehicle for effectiveness, may generate efficiency savings and cost cutting, which will make harder-to-help clients more vulnerable. This is because there is an inevitable variation in the employment likelihoods of different clients. In the experience of marketised employment services, unintended consequences often take the form of two phenomena: 'creaming' and 'parking'. In employment services, creaming and parking are forms of 'adverse selection' whereby clients are selected for assistance in inverse proportion to their need. Creaming practice may be evident in how job-ready clients are 'creamed off' for services so that providers can easily claim payments by neglecting clients more distant from the labour market. Parking practice, on the other hand, concerns provider behaviour that avoids devoting time or resources to clients with more significant barriers to employment (Carter and Whitworth, 2015). The practices of creaming and parking are clearly contrary to the explicit political intention of reducing the performance gap between easier- and harder-to-help clients.

One of the most recurrent strategies used to mitigate service providers' gaming practices has been to strengthen performance signals in payment structures in exchange for giving service providers greater flexibility in choosing the service delivery methodology to be used. This 'black box' approach refers to contracting based on a payment-by-result system where service providers receive greater leeway in the services that they provide the clients. The success of this approach largely depends on the service providers' financial accountability for their service outcomes, which is the main market governance instrument used to control them (Finn, 2012; Considine et al., 2020),

There is a heated debate on the various aspects of the payment-by-result system and the responses that they induce. Several studies have been conducted to determine if the model is more effective in achieving results favourable for the clients and

to identify the unintended consequences of its use. However, the effects of the use of a particular model are often uncertain and contingent on the context and implementation (Van Berkel and Van der Aa, 2005; Bredgaard and Larsen, 2008). For Finn (2010), outcome-based contracting is a potential driver of innovation, flexibility and efficiency savings. Similarly, Pattison (2012) considers flexibility in contracting through the avoidance of public procurement rules to be a key advantage of outcome-based contracting in the delivery of employment services. Courty *et al.* (2005), focusing on two US federal job training programmes, showed that strengthening performance signals in payment structures does not fully eliminate cream-skimming. Indeed, the flexibility given to providers in the services made available may generate incentives to prioritise short-run, 'quick-fix'-type job placement activities in lieu of longer-term activities with more training content. Koning and Heinrich (2013) highlight how unintended practices, such as creaming, are expected to increase if payments are performance-based, because workers with bad *a priori* job prospects will increase the risk of no (or lower) payments. At the same time, this response is conditional on the extent to which providers are able to select clients, and to select them in ways that influence the outcomes.

In some settings, these considerations have even led to a retreat in the use of outcome-based payments, with some governments reverting to fixed-payment schemes for private providers (Hefetz and Warner, 2004). In general, a balance between outcome-based payment and other payment structures for providers, such as fixed-payment schemes, is hard to establish, and contracts are frequently adjusted to prevent unintended consequences (Considine *et al.*, 2011).

Contracting on a payment-by-result basis should at least demonstrate its ability to prevent the parking of clients by proving that a client has obtained either continuous or cumulative employment. This, however, does not eliminate the risk that the contracted provider may rely on less expensive programmes that concentrate solely on matching candidates with vacancies related to unsustainable employment, promoting a work-first ethos (Hill, 2013). The main challenge is to design payment structures that can increase the proportion of payments contingent on the employment outcomes of harder-to-help clients. It should thus become the priority of the government to provide equitable employment support for all clients regardless of the type of employment barrier faced by them. To improve the performance signals related to harder-to-help groups, the outcome-based contracting model often includes payment groups to which different clients are assigned according to some notion of the average difficulty of securing transitions to employment for them. Such payment groups are designed in such a way that the remuneration increases in proportion to how hard it is to help a client find employment. In this regard, profiling tools are used to sort jobseekers into groups with a similar risk level of work resumption, and in turn to determine their level of access to different treatment levels. In general, profiling systems have been developed in many countries to make labour market integration more effective by targeting services and resources (Hasluck, 2004; Loxha and Morgandi, 2014). In a number of cases, they are followed by case management using counsellors' experience to facilitate individualised support. Problems arise when profiling systems do not adequately discriminate 'within' the target group among its members. That is because members of any target group may not necessarily be homogeneous in terms of risk level of work

resumption. Therefore, profiling systems do not fully prevent the risk that service providers will still select those clients who are easiest to serve within these sub-populations of clients (Courty *et al.*, 2005).

2.2. A Fragmented Picture: The Case of the Italian PES

In spite of the differences among countries in terms of their PES-related reforms, many of the policy changes that have occurred in some countries since the 1990s may have similar trends concerning marketisation and the introduction of performance signals into their payment models.

Among these countries is Italy, which is usually considered a 'slow moderniser' or a 'latecomer country' (Van Berkel *et al.*, 2011). The peculiarity of Italy resides in the fact that the quasi-marketisation process has proceeded in a very fragmented way in the country. An intensive period of reforms from 1997 to 2003 promoted a new multi-level architecture that allocated labour market policies competence to regional governments (Zimmermann *et al.*, 2014). It then became possible for regional governments to determine the proportion of services to be delivered by for-profit service providers, and the contracting model to be adopted. Consequently, decentralised decisions on outsourcing contributed to the increase in regional-level discretion in the implementation of marketised services, thus jeopardising the implementation of market-based strategies in the country (OECD, 2019).

For research purposes, the research sought to take advantage of the differential implementation of the marketisation and contracting models from region to region in Italy. In particular, there are two neighbouring regions, with similar economic backgrounds, that can approximate two opposite models in this field: Lombardy and Emilia-Romagna. Figure 1 compares the two regions from the perspective of labour market characteristics. It does so by using three indicators – activity rate, employment rate and unemployment rate – between 2007 and 2019. It is evident that the three indicators present almost identical levels over time.

In the two cases, the contracting models were implemented at almost the same time (2005 for Emilia-Romagna; 2006 for Lombardy).

Lombardy can be considered the region at the forefront of the Italian marketisation of employment services. It was the first to develop a voucher model of a quasi-market in 2006 (Regional Law 22/2006) known as the '*dote* system' (*dote* literally means 'endowment'). In this system, public employment offices are competing actors, because public and private providers are given full parity of rights and duties by means of an accreditation process. Accreditation is granted to any organisation that fulfils pre-defined formal requirements. This gives rise to a vast array of non-public providers participating in the quasi-market (Sabatinelli and Villa, 2015). In this context, the voucher model is supposed to enable the client to 'shop around' among providers so that quality is enhanced via competition (Thuy *et al.*, 2001; Giubileo and Parma, 2013).

Accredited service providers compete to obtain vouchers from clients and then apply to the regional administration for reimbursement: 75% of their total revenue is contingent on their success in obtaining regular employment contracts for their clients. In particular, payments are made only for clients for which an employment contract with a minimum duration of 6 months has been obtained.
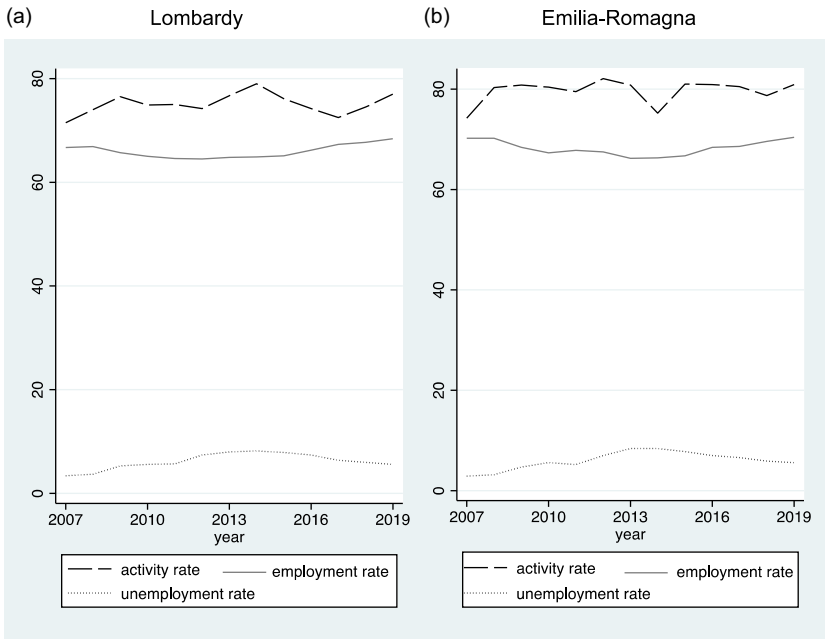
Source: Eurostat, own calculations.

Figure 1. Activity rate, employment rate and unemployment rate in (a) Lombardy and (b) Emilia-Romagna, 2007–2019.

At the same time, to improve the performance of service providers in relation to harder-to-help groups, outcome payments are given to three different payment groups. In the profiling phase after clients are taken on by a service provider, they are given a score for their employment likelihood (the lower the employment likelihood, the higher the score). On the basis of their scores, clients are assigned to a payment group. The higher their score, the higher the payment attached to the voucher (see Supplementary Information 1 for more details).

Emilia-Romagna is a region that chose to adopt an approach characterised by a public-hierarchical logic, which is the opposite of Lombardy's quasi-marketised system. Emilia-Romagna has implemented the traditional contracting approach, in which publicly-funded services are mainly delivered through in-house provision, with some parts usually being outsourced to other non-public service providers via periodic tenders (Struyven and Steurs, 2004). In 2005, the regional government reformed local labour market policies (Regional Law 17/2005), and it was established that private providers must simply support, but never replace, the role of public ones, which remain at the core of activation policy management (Scarano, 2020).

On the basis of a complementary dynamic rather than a competitive one (as in quasi-marketised services), outsourced services are only those that the public service providers are not able to deliver themselves. This dynamic is based

on a fee-for-service scheme that administers the fulfilment of the public and private service operators' mutual obligations, which are strictly defined (Struyven, 2014; Tomkinson, 2016). Clients have no freedom to choose among the available service providers, which are selected for them by the public employment offices. The logic behind this model is that the limited proportion of services outsourced should reduce the cost of monitoring external for-profit service providers, thereby avoiding the necessity of outcome-based payments.

At the same time, the contracting model has become a further important factor in influencing the choice of profiling method. In Lombardy, where more extensive marketisation is in place and the regional administration mainly performs a monitoring role, objective diagnostics relying on quantitative approaches are of particular importance. In contrast, the preservation of tools and resources for less extensive outsourcing and more in-house service delivery are associated with a greater centrality of the counsellors in the diagnosis of jobseekers, as in Emilia-Romagna, whereas diagnosis of the employment likelihood of clients depends on counsellors' judgement (Loxha and Morgandi, 2014) (see Supplementary Information 1).

3. Methods

To assess the degree to which each of the aforementioned contracting models effectively eliminates service providers' incentives to cream and park, a rigorous method of measuring the net contributions of service providers is necessary. To determine this degree, the actual situation of the clients (having obtained a job after services provided to them by the service provider) was compared to their relative 'counterfactual situation' (the situation that would have been observed for the same clients in the absence of the service provider's services).

Adopted for this purpose was propensity score matching, i.e. a statistical matching technique that forms the control group among untreated subjects with observable characteristics that are the most similar to those of the treated subjects (Rosenbaum and Rubin, 1983; Caliendo and Kopening, 2008). In the analysis reported by this article, the treatment and control groups were distinguished from each other by using the procedures required for PES registration. All unemployed people have to register, after which they are given a certification of their unemployment condition. However, not all registered unemployed people become PES clients because their participation in the PES activation initiatives is still voluntary. Those who were not participating in such initiatives constituted the control group, and those who were participating in the initiatives represented the treatment group (see Supplementary Information 2 for more details). The 'treatment' corresponded to any measure that a first-time unemployed PES client could receive during the period considered (Sianesi, 2004). The measures considered in the analysis referred mainly to employment assistance and training. The former corresponds to counselling, orientation and placement measures. The latter concerns interventions aimed at skill contents. However, employment assistance is the most widely used measure for the majority of clients (see Supplementary Information 3).

Because no national data on the aforementioned types of information have been collected to date, to explore the pertinent issues, the analysis reported in this study

drew on the rich local administrative databases made available by the relevant Italian regional authorities.[1] The focus was on the two most important and representative districts of each region: Milan Metropolitan City for Lombardy, and Bologna Metropolitan City for Emilia-Romagna. The said data allowed the inclusion of almost the entire population of jobseekers in the areas observed. The Lombardy data pertained to the period from January 1 to September 30, 2014 while the Emilia-Romagna data pertained to the period from January 1 to September 30, 2015.[2]

The final Lombardy dataset comprised 33,389 observations (6,862 participants; 26,527 non-participants). The final Emilia-Romagna dataset, on the other hand, comprised 20,014 observations (7,036 participants; 12,978 non-participants).

The administrative data obtained enabled differentiation of the subjects on the basis of age, gender, level of educational attainment, citizenship, month of registration and length of unemployment spell (descriptive statistics are presented in Supplementary Information 4) for analysis purposes.[3] These characteristics were used as independent variables to estimate (by means of probit regression) the propensity score $p(X_i)$, which is the conditional probability of receiving a treatment given pre-treatment characteristics (Rosenbaum and Rubin, 1983) (1).

$$p(X_i) = \Pr\left(T_i = 1 \mid X_i\right) \tag{1}$$

where T= {1, 0} is the indicator of exposure to the treatment and X represents pre-treatment covariates.

Probit regression allowed identification of a subject's characteristics that influenced him/her to decide to participate in the PES activation initiatives (Supplementary Information 5). Radius matching based on the propensity score was performed, and a 1% caliper was imposed. Furthermore, the matching was limited to observations concerning the common support even if this had never been an issue (Supplementary Information 6). The matching of the estimated propensity scores sufficiently balanced the distribution of all the covariates between the two groups (Supplementary Information 6), meaning that the mean standardised bias was considerably reduced after matching (Caliendo and Kopening, 2008). The net impact of the services delivered was then estimated as the average treatment effect on the treated[4] (ATT)(2).

$$ATT = E[\, Y_i(1 \mid T_i = 1) - Y_i(0 \mid T_i = 1)] \tag{2}$$

where $Y_i(1)$ is outcome under treatment and $Y_i(0)$ is outcome under no treatment.

In particular, ATT was estimated on the basis of two different outcome variables: the employment condition, and the first employment obtained. The first outcome variable was the probability of being employed at a specific point in time after the beginning of the treatment. In this case, all the employment contracts obtained by the clients after being treated were considered. The second outcome variable, on the other hand, was the probability of being employed with account taken of only the first employment obtained after the services were provided. The purpose of this was to investigate if contracting affects the duration of the employment provided. Thus, the stability of the first employment was considered a proxy for job quality (Koning and Heinrich, 2013; Dengler, 2019). In both cases included in the analysis, the observation period for the outcome was up to 2 years (up to October 2016 for Lombardy and up to October 2017 for Emilia-Romagna).

Table 1. Impact of the services provided identified by observing the employment condition after the treatment

a) Lombardy					
Matching	after 3 months	after 6 months	after 12 months	after 18 months	after 24 months
Before matching	0.170 (0.006)	0.169 (0.006)	0.146 (0.006)	0.119 (0.006)	0.117 (0.006)
ATT After matching	0.232*** (0.007)	0.239*** (0.007)	0.240*** (0.006)	0.228*** (0.006)	0.231*** (0.006)
b) Emilia-Romagna					
Matching	after 3 months	after 6 months	after 12 months	after 18 months	after 24 months
Before matching	0.056 (0.007)	0.038 (0.007)	−0.009 (0.007)	0.028 (0.006)	0.030 (0.006)
ATT After matching	0.042*** (0.004)	0.021*** (0.004)	−0.024*** (0.006)	0.014* (0.007)	0.015** (0.006)

Bootstrap standard errors in parentheses.
***$p < 0.01$. **$p < 0.05$.* $p < 0.1$.

It is important to highlight that the conditional-independence assumption (CIA) had to hold to prevent the analysis results from becoming biased when the outcome was partly dependent on the selection of the individuals to be treated. In the analysis reported in this study, it was assumed that information about the previous period of unemployment met this condition considering that information on labour market histories strongly influenced inclusion in the treatment group (Caliendo et al., 2017).

Once the net contributions of service providers had been detected, binary logistic regressions were used to estimate the clients' predicted probability of transition to sustained employment given their characteristics. This focus reflects the present study's aim of presenting explanatory analysis results and ideas about the equity of the treatment, which will be further discussed below.

4. Findings

4.1 To what extent does outcome-based contracting contribute to employment?

This section presents the estimates of the net impact of the services on the probability of being employed after the treatment. Table 1 shows the overall effect of the services provided 3, 6, 12, 18 and 24 months after the beginning of the treatment.

Table 1a shows that in the case of Lombardy, the impact of the services provided on the employment condition was generally high. In particular, the net impact was 23 percentage points after only 3 months, and it remained stable at around that level even in the long run, until the observation threshold of 24 months. In the case of Emilia-Romagna (Table 1b), the net impact of the services provided was generally very low. It tended to be high in the short term and decreased until it reached almost the null level in the long term.

Table 2. Impact of the services provided identified by observing the first employment obtained after the treatment

a) Lombardy				
Matching	>= 6 months	>= 12 months	>= 18 months	>= 24 months
Before matching	0.232 (0.004)	0.095 (0.002)	0.080 (0.003)	0.076 (0.002)
ATT After matching	0.227*** (0.006)	0.120*** (0.004)	0.101*** (0.003)	0.098*** (0.003)
b) Emilia-Romagna				
Matching	>= 6 months	>= 12 months	>= 18 months	>= 24 months
Before matching	−0.009 (0.006)	−0.004 (0.005)	−0.003 (0.005)	−0.005 (0.005)
ATT After matching	−0.006 (0.006)	−0.003 (0.005)	−0.002 (0.005)	−0.004 (0.004)

Bootstrap standard errors in parentheses.
*** $p < 0.01$. ** $p < 0.05$. * $p < 0.1$.

The matchings were estimated again, but after changing the outcome variable, considering only the first employment contract obtained by each client after the treatment. Table 2 then compares the two cases investigated in the present study with regard to the probability of securing an employment contract with minimum duration of 6, 12, 18 and 24 months.

For Lombardy (Table 2a), in the case of the 6-month-duration employment, the net impact of the services provided was still high: 22 percentage points, a figure comparable to that found when all the employment contracts obtained after the treatment were considered, as in the previous estimations (Table 1a). The impact was massively reduced, however, when the outcome variable was the probability of obtaining an employment contract with a minimum duration of 1 year. To be noted is that the net impact of the services provided in this case was half of the probability of securing an employment contract with a minimum duration of 6 months.

Again in the Lombardy case, when the longer contract durations were examined, it was found that the net impact stabilised at around 10 percentage points. By contrast, in the Emilia-Romagna case, the effect was always completely null regardless of the stability of the first employment contract obtained (Table 2b). In all the estimations, changing the outcome variable from this perspective did not produce any difference.

4.2 What happens when dealing with different employment barriers?

As mentioned earlier, the Lombardy model was designed to take into account the complex employment barriers of different clients by making payments contingent on the employment outcomes of harder-to-help groups. As a result, the outcome payment increases when the client belongs to a higher-payment group as s/he is farther from employment.

Nevertheless, it may not be sufficient to check only whether the service providers are able to provide jobs for the hardest-to-help clients. The prevalence of short-term contracts as the first employment obtained after the treatment in the Lombardy

case, as shown in particular with respect to the results presented in Table 2a, might be the result of the service providers' behaviour of striving to obtain only the minimum-length contracts required to get payments. Therefore, the crucial question is not only 'if' the service providers provide jobs but also 'what' kinds of jobs they provide and 'to whom' they provide them.

The analysis therefore had to consider the difference in the effect when a different-duration contract was taken into account. A short-term first contract was distinguished from a long-term first contract on the basis of the probability of obtaining a contract of 1 year or less at most and the probability of obtaining a contract with a duration of more than 1 year. The analysis used logistic regressions in which the two aforementioned probabilities were the dependent variables. The set of independent variables was almost the same as those that were used to estimate the propensity scores, including the personal observable characteristics available from the administrative data. The set of covariates was augmented, however, with the interaction between each of the aforementioned variables and the dichotomic treatment variable.[5] The aim was to test whether the service intervention changed the relationship between each personal observable characteristic and the probability of obtaining a contract with a duration of less/more than 1 year. At the same time, to account for the selection assignment differences between the treatment and control groups, weights were created on the basis of the propensity scores[6] (McCaffrey et al., 2004). The observations off the common support were excluded.

The analysis results are presented in Tables 3 and 4, where only the relevant coefficients of the interactions are shown (the complete results are available in Supplementary Information 7–16).

As can be seen in Table 3, in relation to the probability of obtaining a contract with a maximum duration of 1 year, some estimates of the effects of the interactions are significant, particularly with respect to the Lombardy case and concerning age, level of educational attainment, and length of unemployment spell.

It was found that in Lombardy, the older participants who received the treatment were more likely to obtain a contract lasting for a maximum of 1 year than were the younger participants (ages 25–34: $\beta = 0.578$, 95% CI [0.401; 0.755]; 35–44: $\beta = 1.088$, 95% CI [0.886; 1.288]; 45–54: $\beta = 0.868$, 95% CI [0.651; 1.083]; 55–64: $\beta = 0.793$, 95% CI [0.445; 1.140]).

With regard to level of educational attainment, the participants with a higher level of educational attainment who received the treatment were less likely to obtain a contract lasting for a maximum of 1 year than those with compulsory schooling (diploma: $\beta = -0.902$, 95% CI [−1.065; −0.737]; university degree: $\beta = -1.377$, 95% CI [−1.572; −1.182]). In relation to the length of unemployment spell, the participants who received the treatment with an unemployment spell between from 6 to 12 months were positively associated with the probability of obtaining a contract with a maximum duration of 1 year ($\beta = 0.301$, 95% CI [0.008; 0.593]) while those who received the treatment with an unemployment spell of more than 2 years were negatively associated with the same probability ($\beta = -0.261$, 95% CI [−0.404; −0.117]).

With respect to the Emilia-Romagna case, there was no significant estimate in relation to the aforementioned probability.

As can be seen in Table 4, both the Lombardy and Emilia-Romagna cases present several significant estimates in relation to the probability of obtaining a contract

Table 3. Coefficients of the interactions between the personal observable characteristics and the treatment with regard to the probability of obtaining a contract lasting for a maximum of 1 year as the first employment after the treatment

Variables	Lombardy	Emilia-Romagna
Age		
15-24 [ref.]		
T*25-34	0.578***	0.091
	[0.401 0.754]	[−0.121 0.304]
T*35-44	1.088***	−0.201*
	[0.886 1.288]	[−0.421 0.019]
T*45-54	0.868***	−0.120
	[0.651 1.083]	[−0.346 0.106]
T*55-64	0.793***	0.064
	[0.445 1.140]	[−0.237 0.366]
Gender		
T*female	−0.003	−0.037
	[−0.131 0.124]	[−0.173 0.098]
Education		
compulsory[ref.]		
T*diploma	−0.902***	−0.126
	[−1.065 −0.737]	[−0.283 0.031]
T*university degree	−1.377***	0.028
	[−1.572 −1.182]	[−0.155 0.211]
Citizenship		
T*foreigner	−0.063	−0.070
	[−0.274 0.147]	[−0.302 0.161]
Unemployment Spell		
≤ 3 months[ref.]		
T* 3-6 months	−0.041	−0.034
	[−0.355 0.272]	[−0.339 0.270]
T* 6-12 months	0.301**	0.153
	[0.008 0.593]	[−0.200 0.506]
T* 1-2 years	0.116	0.068
	[−0.168 0.401]	[−0.222 0.359]
T* > 2 years	−0.261***	0.162*
	[−0.404 −0.117]	[−0.005 0.329]

Notes. p values: *p < 0.05; **p < 0.01; ***p < 0.001. Confidence intervals in parentheses.
See Supplementary Information 7-11 for full models.

Table 4. Coefficients of the interactions between the personal observable characteristics and the treatment with regard to the probability of obtaining a contract lasting for more than 1 year as the first employment after the treatment

Variables	Lombardy	Emilia-Romagna
Age		
15-24 [ref.]		
T*25-34	0.004	0.019
	[−0.361 0.370]	[−0.235 0.275]
T*35-44	−0.832***	0.327**
	[−1.228 −0.436]	[0.067 0.587]
T*45-54	−0.387*	0.327**
	[−0.777 0.002]	[0.065 0.589]
T*55-64	0.317	0.681***
	[−0.166 0.801]	[0.356 1.005]
Gender		
T*female	−0.995***	0.153*
	[−1.218 −0.771]	[−0.003 0.310]
Education		
compulsory[ref.]		
T*diploma	1.035***	−0.125
	[0.748 1.322]	[−0.306 0.055]
T*university degree	1.521***	−0.226**
	[1.091 1.949]	[−0.435 −0.015]
Citizenship		
T*foreigner	−1.658***	0.395***
	[−2.036 −1.280]	[0.137 0.651]
Unemployment Spell		
≤ 3 months[ref.]		
T* 3-6 months	−0.350	0.011
	[−1.077 0.378]	[−0.321 0.344]
T* 6-12 months	0.0445	0.429**
	[−0.739 0.828]	[0.014 0.843]
T* 1-2 years	0.176	0.331**
	[−0.465 0.818]	[0.015 0.646]
T* > 2 years	−0.095	−0.122
	[−0.314 0.123]	[−0.323 0.079]

Notes. p values: *p < 0.05; **p < 0.01; ***p < 0.001. Confidence intervals in parentheses.
See Supplementary Information 12-16 for full models.

with a duration of more than 1 year. In Lombardy, differently from before, the participants aged 35–44 were negatively associated with the probability of obtaining a contract lasting for more than 1 year (β = −0.832, 95% CI [−1.228; −0.436]). The female participants showed a lower probability of obtaining a contract lasting for more than 1 year compared to the male participants (β = −0.995, 95% CI [−1.218; −0.771]). With regard to the level of educational attainment, those receiving the treatment with a higher level of educational attainment were more likely to obtain a contract lasting for more than 1 year than the participants with only compulsory schooling (diploma: β = 1.035, 95% CI [0.748; 1.322]; university degree: β = 1.521, 95% CI [1.091; 1.949]). The foreigner participants showed a lower probability of obtaining a contract lasting for more than 1 year compared to the Italians who received the treatment (β = −1.658, 95% CI [−2.036; −1.280]). The length of unemployment spell showed no significant effect on the probability of obtaining a contract lasting for more than 1 year.

The picture is different for Emilia Romagna, particularly with regard to the older participants and those with the lowest level of educational attainment, foreign citizenship and a medium-long to long unemployment spell who received the treatment. In this case, to a large extent, the older participants were more likely to obtain a contract lasting for more than 1 year than the younger participants (ages 35–44: β = 0.327, 95% CI [0.067; 0.587]; 45–54: β = 0.327, 95% CI [0.065; 0.589]; 55–64: β = 0.681, 95% CI [0.356; 1.005]). Only with respect to those aged 25–34 was the estimate not significant. With regard to the level of educational attainment, there was no significant estimate with respect to diploma, but the participants with a university degree were less likely to obtain a contract lasting for more than 1 year than were those with only compulsory schooling. Those foreigners who received the treatment showed a higher probability of obtaining a contract lasting for more than 1 year compared to the Italians who received the treatment (β = 0.395, 95% CI [0.137; 0.651]). Finally, the participants who had been unemployed from 6 months to 2 years were more likely to obtain a contract lasting for more than 1 year than were those with a shorter unemployment spell (6- to 12-month unemployment spell: β = 0.429, 95% CI [0.014; 0.843]; 1- to 2-year unemployment spell: β = 0.331, 95% CI [0.015; 0.646]).

In general, due to the problem of unobserved heterogeneity, logistic regression should be treated with caution. The analysis reported in the present study tried to cope with this limitation by re-estimating the results with linear probability models as robustness checks (Mood, 2010). The coefficients of the interactions between treatment and the personal observable characteristics were in line with the results obtained via logistic regressions (Supplementary Information 17-26).

5. Discussion

As already highlighted in the previous sections, tensions and risks in relation to conflicting policy aims seem to remain a constant problem within the framework of marketised employment services. Policymakers oscillate between encouraging competition and freedom in service delivery, on the one hand, and strengthening

control on the other, in order to avoid the unintended consequences of market governance (Jantz *et al.*, 2018).

The wide gap between Lombardy and Emilia-Romagna in the net contribution of the service providers to getting jobs for their clients may be considered prime evidence that the outcome-based payment scheme is very effective in preventing the risk of the parking of clients. Lombard clients have greater chances of being employed while Emilian clients have only a slightly higher probability of being employed compared to other jobseekers.

Service providers could then prove to be financially accountable for achieving the contracted outcome. However, many of these additional jobs lasted for only a maximum of 1 year, which probably indicates that the clients were poorly matched to the employment that they had received (Hill, 2013). The experience of Lombardy then seems to induce service providers to concentrate solely on matching candidates with vacancies and promoting a quick-job-entry approach. The threshold of 6 months required as the period of employment proved to be weakly related to the long-run impacts of the programme, emphasizing short-run job-placement activities (Courty and Marschke, 2003). This standard is likely to send the wrong signals and to stimulate behavioral responses with negative implications (Courty *et al.*, 2005). Because such a contract length suffices for the service providers to receive their expected payments, obtaining a contract for their clients with just that length is the fastest way for the providers to maximise their profit.

Lombard policymakers have attempted to take performance signals into account in relation to disadvantaged groups by designing three payment groups into which clients are grouped according to a particular metric of their employment likelihood. Nevertheless, some harder-to-help groups seem more likely to be given only short-term employment. In this respect, the creaming practice may still be evident in the way service providers maximise their profits by providing harder-to-help groups (who are included in the payment groups ensuring higher remuneration) with only short-term employment. It is likely that providers have been induced to focus on needier target populations, but the payment structure incentivises them to emphasise short-run job-placement activities within these sub-populations of clients in lieu of longer-term initiatives with more human capital content (Courty and Marschke, 2003).

Conversely, considering the evidence from Emilia-Romagna, the payment for the services rendered to the less employable may be more equitable through the fixed-payment scheme. Some harder-to-help people assisted in finding a job are able to obtain longer employment contracts. This may be the result of initiatives that specifically target expanding employment opportunities for disadvantaged groups (Scarano, 2020). In this regard, the more intensive efforts made to assist particular disadvantaged groups are probably related to a decrease in the overall effectiveness of the system (Koning and Heinrich, 2013).

Some limitations of the analysis reported in this article should be highlighted, starting with the methods employed. While the reliance on register data made it possible to build large datasets, it should be noted that this information is not collected for research purposes. It is not possible to identify personality or behavioural traits (which can influence an individual's labour market performance) from the administrative archives. Thus, more emphasis should have been put on the

labour market history, which could enable a large part of the information on usually unobserved variables to be implicitly captured (Caliendo *et al.*, 2017). The analysis reported in this study attempted to make the most of the available information, taking into account the different lengths of unemployment spells. If further possibilities arise from the available data, it is reasonable to consider that the estimation of the impact of the services may also be affected by how the labour market history is accounted for and by whether a larger set of covariates is used. Richer administrative databases adopting a longitudinal perspective may enable the construction of discrete-time models that control for unobserved heterogeneity (Hohmeyer and Lietzmann, 2020). At the same time, the role of frontline staff, who work directly with jobseekers, should not be overlooked. In quasi-marketised employment services, the freedom of choice for clients should consider that the choice may be imperfectly informed and experienced in different ways, so that it would affect some groups with lower capabilities (Sabatinelli and Villa, 2015). In these respects, information may be drawn from surveys of frontline employment services staff operating in contracted providers and public offices. Furthermore, more qualitative approaches, using exploratory interviews, may help to elicit information on how PES staff carry out the service delivery tasks required for assisting jobseekers to find work (Considine *et al.*, 2011; Freier and Senghaas, 2022).

6. Conclusion

The results of the foregoing analysis show that an outcome-based contracting model can make service providers financially accountable. However, the empirical evidence suggests that performance measures need to be carefully identified to promote long-term program objectives while simultaneously generating more readily available performance information for program management (Heinrich, 2011).

The analysis reported in this study was meant to identify some factors that need to be considered in particular when service providers' risk of failure is greater due to their clients' complex employment barriers. In this respect, the effectiveness of performance measures depends on the ability of their designers to identify dimensions on the basis of which gaming practices take place, doing so on the basis of the heterogeneity among participants (Courty *et al.*, 2005, 2011). In order to deal with the limitations described there is still no consensus in the academic and institutional debate on how outcomes should be measured and evaluated.

It is important to bear in mind that policy effects are often contingent on the context and implementation (Bredgaard and Larsen, 2008; Carter and Whitworth, 2015). Therefore, the meaning of outcomes may vary according to socioeconomic contexts. In regions characterised by highly dynamic labour markets, the achievement of a short-term employment contract may not be particularly meaningful without a corresponding increase in employability (Sabatinelli and Villa, 2015). Different approaches to setting performance measures may consider the opportuneness of updating payment structures and human-capital content from time to time. This could be useful for considering variations in economic features of local areas, participant characteristics, and services delivered (Courty *et al.*, 2005). In this view,

performance tracking systems gain in importance for constructing estimates based on data from past performance outcomes.

Failure to meet the support needs of harder-to-help clients may also reflect the fact that service providers (both public and private) may simply not have the economic means and sources of information to consider certain characteristics of jobseekers or their local areas that are relevant to their (un)employment (Carter and Whitworth, 2015). As already highlighted, in both the cases examined, training initiatives are very limited (Supplementary Information 3). Even if the analysis in this study has focused on two well-developed areas of Italy, it is important to note that the country is still marked by a lack of public resources for PES, even during the years of the Great Recession (OECD, 2019). In the framework of the Next Generation EU (NGEU) economic recovery package to support member states adversely impacted by the COVID-19 pandemic, the latest Italian reforms have been accompanied by a renewed emphasis on the activation conditionality of the new benefit recipients. This perspective makes the improvement of PES a more urgent issue, including contracting design – aspects that are of crucial importance especially when dealing with latecomer countries from the viewpoint of PES marketisation.

Supplementary material. To view supplementary material for this article, please visit https://doi.org/10.1017/S0047279422001003

Competing interests. The author declares none.

Notes

1 These sources were released by the labour market agencies of the two subject regions in this study: *Polis* for Lombardy and *Agenzia Regionale per il Lavoro* for Emilia-Romagna.
2 The difference in the observed time interval is only due to data availability and does not imply any substantial difference in the observation.
3 Each of the variables distinguishes the following categories: ages 15–24, 25–34, 35–44, 45–54, 55-64; gender is 1 when female and 0 when male; level of educational attainment is 1 for compulsory schooling, 2 for high school diploma and 3 for university degree; citizenship is 1 when a foreigner and 0 when Italian; length of unemployment spell distinguishes between up to 3 months, more than 3 and up to 6 months, more than 6 and up to 12 months, more than 1 year and up to 2 years, more than 2 years; month of registration is from January to September.
4 As the standard errors of the ATT did not take into account the estimated propensity scores, they were replaced by means of the bootstrap technique performed with 100 re-samplings.
5 The interactions do not include the month of registration as unemployed.
6 The weights were created automatically by using *Psmatch2* Stata command, by assigning 1 to each treated units and a value <1 to non-treated units, describing the relationship between pretreatment characteristics and treatment assignment with minimal approximation error.

References

Bredgaard, T. and Larsen, F. (2008), 'Quasi-Markets in Employment Policy: Do They Deliver on Promises?', *Social Policy and Society*, **7**, 3, 341–352.
Bruttel, O. (2005), 'Contracting-out and the Governance Mechanisms in the Public Employment Service', WZB Discussion Paper 109, Berlin.
Caliendo, M. and Kopening, S. (2008), 'Some practical guidance for the implementation of propensity score matching', *Journal of Economic Surveys*, **22**, 1, 31–72.

Caliendo, M., Mahlstedt, R. and Mitnik, O.A. (2017), 'Unobservable, but unimportant? The relevance of usually unobserved variables for the evaluation of labor market policies', *Labour Economics*, **46**, 14–25.

Carter, E., and Whitworth, A. (2015), 'Creaming and parking in quasi-marketized welfare-to-work schemes: Designed out of or designed into the UK Work Programme?', *Journal of Social Policy*, **44**, 2, 277–296.

Considine, M., Lewis, J.M. and O'Sullivan, S. (2011), 'Quasi-Markets and Service Delivery Flexibility Following a Decade of Employment Assistance Reform in Australia', *Journal of Social Policy*, **40**, 4, 811–833.

Considine, M., O'Sullivan, S., McGann, M. and Nguyen, P. (2020), 'Contracting personalization by results: Comparing marketization reforms in the UK and Australia', *Public Administration*, **98**, 4, 873–890.

Courty, P. and Marschke, G. (2003), 'Dynamics of Performance-Measurement Systems', *Oxford Review of Economic Policy*, **19**, 2, 268–284.

Courty, P., Heinrich, C. and Marschke, G. (2005), 'Setting the Standard in Performance Measurement Systems', *International Public Management Journal*, **8**, 3, 321–347.

Courty, P., Kim, D. H. and Marschke, G. (2011), 'Curbing cream-skimming: Evidence on enrolment incentives', *Labour Economics*, **18**, 5, 643–655.

Dengler, K. (2019), 'Effectiveness of Active Labour Market Programmes on the Job Quality of Welfare Recipients in Germany', *Journal of Social Policy*, **48**, 4, 807–838.

Fay, G. (1997), 'Making the Public Employment Service more effective through the introduction of market signals', OECD Labour Market Occasional Paper 25, Paris.

Finn, D. (2010), 'Outsourcing Employment Programmes: Contract Design and Differential Prices', *European Journal of Social Security*, **12**, 4, 289–302.

Finn, D. (2012), 'Subcontracting in Public Employment Services: the design and delivery of "outcome based" and "black box" contracts', European Commission DG Employment, Social Affairs and Inclusion - Analytical Paper, Bruxelles.

Freier, C. and Senghaas, M. (2022), 'Placement Advisors as Innovators. How Professionals Use Enhanced Discretion in Germany's Public Employment Services', *Journal of Social Policy*, **51**,1, 155–172.

Giubileo, F. and Parma, A. (2013), 'Esperienze di welfare occupazionale in Lombardia. Una prima valutazione sui rendimenti della Dote Lavoro', *Autonomie locali e servizi sociali*, **1**, 37–53.

Greer, I., Breidhal, K.N., Knuth, M. and Larsen, F. (2017), *The Marketization of Employment Services: The Dilemmas of Europe's Work-First Welfare States*, Oxford: Oxford University Press.

Greer, I., Schulte, L., and Symon, G. (2018), 'Creaming and parking in marketized employment services: An Anglo-German comparison', *Human Relations*, **71**, 11, 1427–1453.

Hasluck, C. (2004), 'Targeting Services in the Individual Customer Strategy: The Role of Profiling. A Review of the Research Evidence', Department for Work and Pensions Research Report 192, London.

Hefetz, A. and Warner, M.E. (2004), 'Privatization and Its Reverse: Explaining the Dynamics of the Government Contracting Process', *Journal of Public Administration Research and Theory*, **14**, 2, 171–190.

Heinrich, C. J. (2011), 'The Performance of Performance Standards', *Employment Research*, **18**, 2, 4–5.

Hill, J. G. (2013), 'The marketization of employment services and the British Work Programme', *Competition and Change*, **17**, 2, 197–207.

Hohmeyer, K. and Lietzmann, T. (2020), 'Persistence of Welfare Receipt and Unemployment in Germany: Determinants and Duration Dependence', *Journal of Social Policy*, **49**, 2, 299–322.

Jantz, B., Klenk, T., Larsen, F., and Wiggan, J. (2018), 'Marketization and varieties of accountability relationships in employment services: Comparing Denmark, Germany and Great Britain', *Administration and Society*, **50**, 3, 321–345.

Koning, P., and Heinrich, C.J. (2013), 'Cream-Skimming, Parking and Other Intended and Unintended Effects of Performance-Based Contracting in Social Welfare Services', *Journal of Policy Analysis and Management*, **32**, 3, 461–483.

Le Grand, J. (2006), 'Equality and Choice in Public Services', *Social Research*, **73**, 2, 695–710.

Loxha, A., and Morgandi, M. (2014), 'Profiling the Unemployed: a review of OECD experiences and implications for emerging economies', Social Protection & Labour Discussion Paper 1424, Washington DC.

McCaffrey, D.F., Ridgeway, G. and Morral, A.R. (2004), 'Propensity score estimation with boosted regression for evaluating causal effects in observational studies', *Psychol Methods*, **9**, 4, 403–425.

Mood, C. (2010), 'Logistic Regression: Why We Cannot Do What We Think We Can Do, and What We Can Do, and What We Can Do About It', *European Sociological Review*, **26**, 1, 67–82.

OECD (2019), *Connecting People with Jobs: Strenghtening Active Labour Market Policies in Italy*, Paris: OECD Publishing.

Pattison, V. (2012), 'Payment by results: Delivering high quality public services in straitened fiscal times', *Local Economy*, **27**, 5/6, 465–474.

Rosenbaum, P.R. and Rubin, D.B. (1983), 'The Central Role of Propensity Score in Observational Studies for Causal Effects', *Biometrika*, **70**, 1, 41–55.

Sabatinelli, S. and Villa, M. (2015), 'Happy ever after in the quasi-market place? The dowry logic of active labour policy in the Lombardy Region', *International Journal of Sociology and Social Policy*, **35**, 11/12, 812–827.

Scarano, G. (2020), 'Alternative models of activation policies: the experience of public oriented services', *International Journal of Sociology & Social Policy*, **40**, 3/4, 382–408.

Sianesi, B. (2004), 'An evaluation of the Swedish system of active labor market programs in the 1990s', *The Review of Economics and Statistics*, **86**, 1, 133–155.

Sol, E. and Westerveld, M. (2005), *Contractualism in Employment Services. A New Form of Welfare Governance*, The Hague: Kluwer Law International.

Struyven, L. (2014), 'Varieties of market competition in public employment services – a comparison of the emergence and evolution of the new system in Australia, The Netherlands and Belgium', *Social Policy & Administration*, **48**, 2, 149–168.

Struyven, L. and Steurs, G. (2004), 'Quasi-market reforms in employment and training services: first experiences and evaluation results', in P. Descy, and M. Tessaring, (eds.), *Evaluation and impact of education and training: the value of learning, Third report on vocational training research in Europe*, Luxembourg: Office for Official Publications of the European Communities.

Thuy, P., Hansen, E. and Price, D. (2001), *The employment service in a changing labour market*, Geneva: International Labour Office.

Tomkinson, E. (2016), 'Outcome-based contracting for human services', *Evidence Base*, **1**, 1, 1–20.

Van Berkel, R. and Van der Aa, P. (2005), 'The marketization of activation services: A modern panacea? Some lessons from the Dutch experience', *Journal of European Social Policy*, **15**, 4, 329–345.

Van Berkel, R., De Graaf, W. and Sirovatka, T. (eds.) (2011), *The Governance of Active Welfare States in Europe*, Basingstoke: Palgrave and MacMillan.

Webster, E. and Harding, G. (2000), 'Outsourcing Public Employment Services: the Australian experience', Melbourne Institute Working Paper 4/00, Melbourne.

Zimmermann, K., Aurich, P., Graziano, P.R., and Fuertes, V. (2014), 'Local Worlds of Marketization – Employment Policies in Garmany, Italy and the UK Compared', *Social Policy and Administration*, **48**, 2, 127–48.