Using two-part contracts to identify cross price elasticities: an application to family doctors[*]

Paul Rodríguez-Lesmes[†], Marcos Vera-Hernández[‡]

February 21, 2018

Abstract

We propose an empirical test for determining if rewarded tasks are cost complements or substitutes in a pay for performance scheme with kinks on linear task-specific reward functions. The test is based on the insensitivity of effort exerted on a particular task to variations in the price of competing tasks for agents who are bunched near the kinks. As a case study, we consider the case of the Quality and Outcomes Framework (QOF). This system accounts for nearly a quarter of family doctors' income and is the largest pay-for-performance programme for primary care services in the world. We found that changes introduced in the system in 2011 were tasks that were complements of many of the unmodified tasks. As a result, there is no evidence of effort-diversion as a result of the changes.

JEL: D8, I11, J2

1 Introduction

Principal-agent relationships are widespread in economics. Since Holmstrom and Milgrom [1991]'s seminar article, it is well understood that the agent's cost function plays a crucial role in a multitask environment, that is, when the agent must carry out more than one task. If tasks complement each other, rewarding one task will be enough to increase the production of an unrewarded task. If, however, tasks substitute for each other, rewarding one will reduce the effort exerted on the unrewarded task. Complementarities/substitutions across tasks not only play a role in the structure of incentive contracts, but also in job design. Whenever possible, tasks that are substitutes should be performed by different agents, each of them carrying out tasks that complement each other.

In this paper, we show how to recover from the data whether tasks are complements or substitutes when the agent faces a two-part linear contract, essentially a contract with two different piece-rate levels. Our approach exploits a change in the incentives faced by the agents, but in contrast to the literature, we can exploit nationwide incentive changes, and do not need that a "control" group, that is, a group of agents not eligible for the change of incentives.

[*] We would like to acknowledge the valuable comments and suggestions by David Bardey, Helmuth Cremer, Jonathan Kolstad, Albert Ma, Suphanit Piyapromdee, Luigi Siciliani, and participants at the 18th European Health Theory Workshop (Oslo, May 2017).

[†] Assistant Professor, Universidad del Rosario

[‡] Reader, University College London

Our main insight is that when agents face two-part linear contracts, a group of agents will naturally choose to produce at the level of the kink of the two-part contract, the level at which the piece rate changes. We show that these agents are insensitive to local changes in the incentives of other tasks, independently of how the tasks interact in the cost function. Hence, the individuals who self-select to produce at the kink will work as a "control" group. Because linear two-part contracts are quite prevalent (and we do not need an explicit control group), our method greatly expands the situations in which we can test for complementarities/substitutions in the agent's cost function. Empirically, the test consists on two steps. First, we test for the relevance of the kink by borrowing some concepts from taxation literature [Saez, 2010, Kleven, 2016]: the presence of kinks implies bunching in the distribution of the relevant output. Then, it is possible to detect if the kink on the payment's function is important for a task just by analysing its output density function. Second, a difference-in-differences (DiD) style of regression is suggested, where the control group corresponds to those agents at the kink. Here we are not assuming that the payments system is not affecting those agents, rather we rely on the insensitivity of those practices, which decided that their optimal effort level attained the output level of the kink, to small changes on the reward of other tasks.

We apply our method to identify whether different activities that family doctors perform are complements or substitutes in their cost function. Examples of the activities that we analyse include carrying out certain tests on diabetic patients, recording smoking history in *at risk* patients, or reviewing asthmatic patients with some minimum frequency, among others.

The types of activities that we analyse contrast with much of the existing empirical literature that has focused on much simpler activities.[1] This literature has focused on studying a specific case of multitasking: the trade-off between quantity and quality within a *single* activity. Because of the very nature of it, quantity and quality are either substitutes or independent at best, but complementarity is rightly dismissed.[2] Because we study genuinely different tasks (rather than the quantity and quality of a single task), the possibility of complementarities across them is real. It might well be, for instance, that the marginal cost of carrying out a test is smaller if another test is also being conducted during the same visit.

Monetary incentives are also used amongst professions with a large pro-social component, such as teachers and doctors. Although crowding out of intrinsic motivation is usually cited as a concern, multitasking is another one. Unsurprisingly, there is a reasonably large body of literature for "teaching to the test", and more generally whether teachers shift effort from unrewarded tasks to rewarded ones (see Neal [2011] for a review of US focused studies).[3] The evidence on health care probably lags behind that on education. Dumont et al. [2008] found that Canadian physicians who voluntary signed up to a contract that paid less for a specific quantity of consultations, increased the average time per consultation (an indicator of quality) as well as other activities unremunerated at the margin (i.e. teaching). Feng Lu [2012] exploited a mandatory quality disclosure policy and found that nursing homes improved scores on quality measures for the reported dimensions, but deteriorated in regard to unreported ones.

In this paper, we exploit the *Quality Outcomes Framework* (QOF), a programme established in 2004 that remunerated all family doctors in England according to their performance in a large

[1] See for instance, Lazear [2000], Shearer [2004], Kosfeld and Neckermann [2011], Bradler et al. [2013].

[2] Al-Ubaydli et al. [2012] finds that higher piece rates leads to higher quality when stuffing envelopes, but this is explained because the piece-rate mechanism signals to the agent that the principal has a good monitoring technology rather because there are complementarities in the cost function.

[3] Muralidharan and Sundararaman [2011], and Glewwe et al. [2010] are examples of developing country studies).

battery of indicators. There is a remuneration schedule for each rewarded indicator, which has a lower and an upper limit. The doctor's remuneration increases linearly as long as the indicator is between the lower and upper limit, and flattens out if the upper limit is passed. The programme was rolled out simultaneously across England, and any changes to the remuneration schedule also apply nationally. This makes it an ideal setting to apply the method that we develop in this paper.

The QOF is the largest primary care pay for performance programme worldwide [Roland and Olesen, 2016], and has already received some attention. Sutton et al. [2010] compared incentivised and unincentivised measures before and after the introduction of the program, a improvements in both measures which were higher for incentivised ones.[4] This approach relies on the assumption that incentivised and unincentivised measures would follow a common trend in the absence of the program. While important, this approach restrict the analysis to tasks that were initially regarded as less important, and which might be different enough to the originally incentivised ones making the parallel trends assumption a strong assumption. Moreover, despite being informative when the programme is introduced, this approach cannot be used to explore systems that have been in place for several years. Instead of relying on the implementation period of the QOF and analysing tasks which were not remunerated, our work analyses tasks that are currently remunerated in a period when QOF has already been in place for more than five years. This is possible because our test takes advantage the design of the payments system. Also, as there is a period in which there are changes to rewards (2010/11) where there was a net price drop in a set of modified indicators, preceded by a period without them (2009/10), we are able to distinguish effort response to the new rewards from variations linked to year-to-year variation, which are correlated with performance.

With our procedure, for those tasks which their remuneration was not modified between 2009 and 2011, we are able to assess whether each task is a complement or a substitute of the set of tasks that were modified or removed in this period. We found that there is no robust evidence of substitutability between tasks in the system, and if anything, several of them are complements. In particular, we found that the reduction on the price drop of certain tasks, principally on the clinical areas of diabetes and cardiovascular disease, resulted on a drop on the output of some indicators which were not modified but which reflect tasks in the same areas.

After this introduction, we present a basic model of multitasking with a two-part linear reward function for agents. Given that, we show the conditions under which we are able to identify complementarities/substitution in the cost function empirically. This is followed by a description of the QOF and the results of using our test on it. Finally, the conclusions are presented.

2 Model

Our test is based on the existence of a two-part linear tariff on a principal-agent relationship. In order to understand the intuition behind the test, we will start by presenting a simple version of the model without uncertainty. In this model, we will introduce the kink produced by a two-linear tariff and examine its implications. Later, we will consider how this main ideas would be affected by introducing uncertainty.

Consider a principal-agent relationship with two distinct tasks. The principal hires the agent to exert task-specific efforts (e_1, e_2). The principal benefits increasingly from the output of the two tasks $(x_{1,}x_2)$. The agent is paid according to $P(x_1, x_2; a_1, a_2) = T + a_1x_1 + a_2x_2$, where T

[4] Kaarboe and Siciliani [2011] motivate their multitasking model using the QOF. They argue, based on the results of Sutton et al. [2010], that quality dimensions in primary care might be complements.

represents a lump-sum payment, and a_i is the piece rate associated to x_i. The agent's cost function is given by $C(e_1,e_2;z)$. characterised by a parameter z. We assume that for $i \in 1,2$ we have that $\frac{\partial C}{\partial e_i} = C_i > 0$, $\frac{\partial^2 C}{\partial e_{ii}^2} = C_{ii} > 0$, $\frac{\partial C}{\partial z} > 0$, $\frac{\partial^2 C}{\partial z^2} > 0$, and that C is a convex function, but we do not restrict the sign of the cross-derivatives $C_{ij} = \frac{\partial^2 C}{\partial e_j \partial e_i}$, $i \neq j$. That is, while we know that it is increasingly costly to exert effort, we do not know if increasing effort in one task, increases or reduces the marginal cost of exerting effort on the other task. In the former case, the tasks are said to be substitutes, and in the latter they are complements. Our main goal is to estimate the sign C_{ij} to ascertain whether the tasks are complements or substitutes.

The agent takes the contract $P(x_1,x_2)$ as given, and decides optimal levels of effort in order to maximize his surplus, that is:

$$\max_{e_1,e_2} U = E\left[P(x_1,x_2;a_1,a_2) - C(e_1,e_2;z)\right] \tag{1}$$

2.1 Model without uncertainty

Let's assume that $x_{i=}e_i$ and that providers are heterogeneous only on an efficiency parameter z which is assigned in the population following a pdf $g(\cdot)$, or CDF $G(\cdot)$. Specifically, let's assume that $C(e_1,e_2;z) = \frac{1}{z}C(e_1,e_2)$. As a result, given a contract specified by $\{T,a_1,a_2\}$, the provider will solve:

$$\max_{e_1,e_2} U = (T + a_1 \cdot e_1 + a_2 \cdot e_2) - \frac{1}{z}C(e_1,e_2). \tag{2}$$

The first order conditions (FOC) of the problem are given by:[5]

$$a_i - \frac{1}{z}C_i = 0,\ i \in \{1,2\} \tag{3}$$

Essentially, the marginal benefit (a_i) of exerting effort has to be equal to the marginal cost $(\frac{1}{z}C_i)$. If we differentiate these FOC, we obtain:

$$da_i - \frac{1}{z}C_{ii}de_i - \frac{1}{z}C_{ij}de_j = 0,\ i \neq j,\ \ i,j \in \{1,2\} \tag{4}$$

This system of equations allows us to explore how optimal allocation of effort in each task would be adjusted as a response to variations in the piece-rates a_i and to the efficiency parameter z.

Proposition 1. *With a linear payment and without uncertainty, we have that* $\frac{de_1}{da_2} = \frac{-z \cdot C_{12}}{C_{11}C_{22} - C_{12}^2} > 0$*, and hence that the sign of* $\frac{de_1}{da_2}$ *is opposite to the sign of* C_{12}*. If the tasks are substitutes* $(C_{12} > 0)$*, we will have that* $\frac{de_1}{da_2} < 0$*. On the contrary, if the tasks are complements* $(C_{12} < 0)$ *then* $\frac{de_1}{da_2} > 0$*.*

Proof:

If we set $da_1 = 0$, that is, a_1 as the unchanged P4P incentive, we can obtain that

$$de_2 = -\frac{C_{11}}{C_{12}}de_1 \tag{5}$$

[5] The second order condition (SOC) is given by $C_{11}C_{22} - C_{12}^2 > 0$, which we assume to hold.

And hence, the impact of modifying the reward a_2 on e_2 is obtained by substituting (5) in the FOC of e_2 :

$$\frac{de_1}{da_2} = -\frac{z \cdot C_{12}}{C_{11}C_{22} - C_{12}^2} \tag{6}$$

Q.E.D.

If we consider that $da_2 = 0$ but $da_1 \neq 0$, we can derive the response on optimal effort for task 1, given variations in its own price. As expected, it is unambiguously positive:

$$\frac{de_1}{da_1} = \frac{z \cdot C_{22}}{C_{11}C_{22} - C_{12}^2} > 0 \tag{7}$$

Assumption 1. *We assume that* $\frac{de_1}{dz} = \frac{a_1 C_{22} - a_2 C_{12}}{C_{11}C_{22} - C_{12}^2} > 0$*, for any value of* z*.*

Note that this is a very natural assumption: if the agent becomes more efficient and its costs decreases, he will exert more effort. It is indeed guaranteed for the case of complements, because $C_{12} < 0$. For the case of substitutes, we need to assume that C_{22} is not too small compared to C_{12}. Otherwise, the agent might greatly increase e_2 and decrease e_1.

2.2 The role of kinks

Now, let's consider a two-part linear payment function, with a kink at $e_1 = UL$.[6] We consider a piece-rate for a given task varies at UL from $\underline{a}_1$to $\bar{a}_1$, as shown in Equation 8 below. As a notation convention, all objects denoted with a lower bar will be related to the contract when the output is below UL, and those with an upper bar for the contract when the output is above such a value. Following our specific application,[7] we will consider $\underline{a}_1 > \bar{a}_1$, so the marginal benefit of e_1 decreases discontinuously at $e_1 = UL$. Notice that this payment function also implies that the fix income jumps in order to maintain the total payment continuous at UL.

$$P(x_1, x_2; a_1, a_2) = \begin{cases} \underline{a}_1 x_1 + a_2 x_2 + T & if\, x_1 < UL \\ \bar{a}_1 x_1 + a_2 x_2 + T + (\underline{a}_1 - \bar{a}_1) \cdot UL & if\, x_1 \geq UL \end{cases} \tag{8}$$

Proposition 2. *Without uncertainty, the presence of a kink at* $e_1 = UL$ *implies that those providers with a* $z \in [\underline{z}, \bar{z}]$ *choose* $e_1^* = UL$*. Moreover,* $\frac{de_1}{da_2} = 0$ *for them.*

Proof:

Below the threshold UL, for a given z there is a optimal level of effort $\underline{e}_1(z) = e_1^*(z, \underline{a}_1, a_2)$. In particular, we assume that $\exists z = \underline{z}\ st\ \underline{e}_1(\underline{z}) = UL$. Above the threshold, $e_1 > UL$, there is also an optimal allocation $\bar{e}_1(\bar{z}) = e_1^*(z, \bar{a}_1, a_2)$, and we also assume that $\exists \underline{z}\ st\ \bar{e}_1(\bar{z}) = UL$.

[6] As will be described in the application section, the QoF is a three-part linear contract. It has a zero piece-rate below a first threshold, the *lower limit*, and above a second threshold, the *upper limit*. We will concentrate on what happens around the *upper limit* given that most of the agents are situated around or above it. Nevertheless, the model and empirical test detailed in this paper could potentially be formulated to the lower limit if there was enough information.

[7] The QoF presents an extreme scenario: $\underline{a}_1 > \bar{a}_1 = 0$. The results that we present here do not require a zero marginal benefit for unit of effort after the upper threshold. An alternative interpretation is that $\bar{a}_1$ represents the altruistic marginal benefit that the physicians obtain for improving their patients' health.

Given that $\underline{a}_1 > \bar{a}_1$, the optimal effort above UL, $\bar{e}_1(z) = e_1^*(z, \bar{a}_1, a_2)$, has to be smaller than the corresponding decision if there were no kink: $\underline{e}_1(z) > \bar{e}_1(z)\ \forall z$. In particular, $UL = \underline{e}_1(\underline{z}) > \bar{e}_1(\underline{z})$. This is due to Equation 7. Notice that it has to be the case that $e_1^*(z+\epsilon, a_1, a_2) > e_1^*(z, a_1, a_2)\ \forall \epsilon > 0$, which holds because of Assumption 1 ($a_1C_{22} - a_2C_{12} > 0$).[8] As a result, given that $UL > \bar{e}_1(\underline{z})$, it is required that $\bar{z} > \underline{z}$.

Those providers with a $z \in [\underline{z}, \bar{z}]$ have to choose $e_1^* = UL$, even though the FOC is not satisfied, because any deviation would be detriment of their utility. Let us consider the diagram on Figure 1 to illustrate the argument. Point A represents the decision of a provider with productivity $\underline{z}$, which is $e_1^* = UL$ as stated before. Point C does the same for the typical $\bar{z}$ provider, which also chooses $e_1^* = UL$.

Let us consider a provider with a productivity in between, $\tilde{z} \in (\underline{z}, \bar{z})$. Without the kink, the optimal decision under $\underline{e}_1(z)$ would have been point B'; however under the kinked payment function it is not optimal. At this point the marginal cost of exerting effort is larger than the marginal benefit of doing so, $\frac{C_1}{C_2}a_2 > a_1$ (from the FOC), so it is a better idea to reduce effort in order to enhance utility. An alternative scenario is to consider a world where $a_1 = \bar{a}_1, \forall e_1$; in such a scenario B'' would have been the choice. Once again, under the actual kinked function this is suboptimal. The provider is better off if effort is increased, as at that point $\frac{C_1}{C_2}a_2 < a_1$. As a result, due to the non-smoothness of the optimization problem, the provider is better off at point B, even though the FOCs do not hold. Notice that as $\frac{C_1}{C_2}a_2 \neq a_1$, the effect of a small variation in a_2 would have no impact on the allocation of a_1^*. As a result, $\frac{de_1}{da_2} = 0$ for those providers with a $z \in [\underline{z}, \bar{z}]$.

Proposition 3. *Without uncertainty, the presence of a kink at $e_1 = UL$ generates bunching on the distribution of effort on task one, $H(e_1)$.*

Proof:

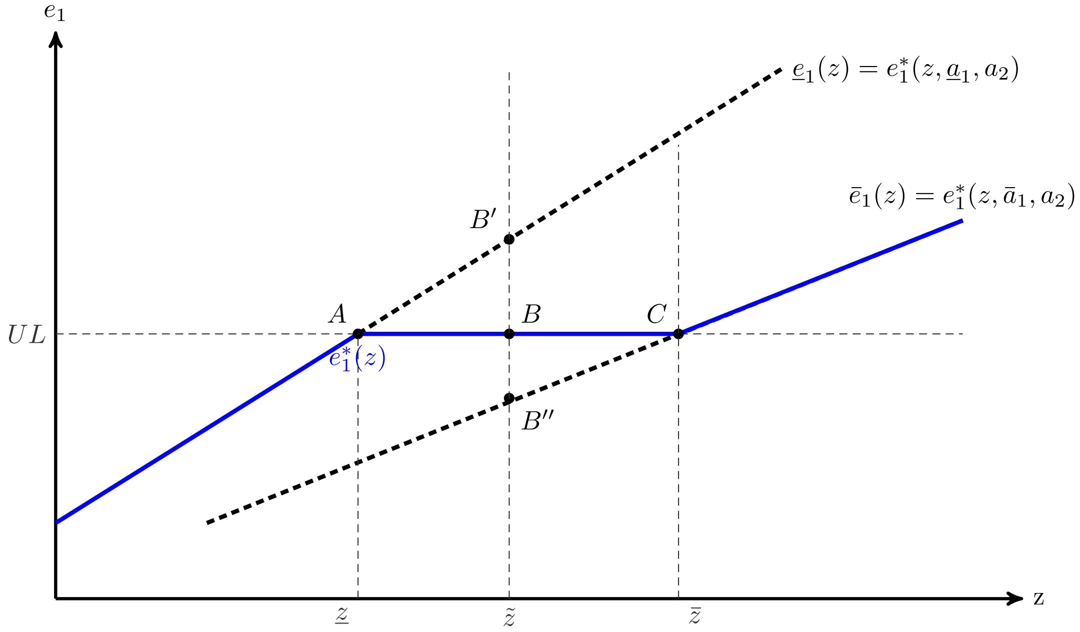
For this proof we provide an argument which follows Saez's [2010] reasoning for the taxation literature.[9] We define $H(\tilde{e}_1) = \Pr\left[e_1^*(z, a_1, a_2) \leq \tilde{e}_1\right] = \Pr\left[z \leq e_1^{*-1}(\tilde{e}_1; a_1, a_2)\right] = G\left[e_1^{*-1}(\tilde{e}_1; a_1, a_2)\right]$, where $e_1^{*-1}(\cdot)$ is the inverse function of $e_1^*(z)$. As explained above, $e_1^*(z)$ is piecewise defined, which is also the case for $H(\tilde{e}_1)$. Below UL we have $\underline{H}(\tilde{e}_1) = G\left[e_1^{*-1}(\tilde{e}_1; \underline{a}_1, a_2)\right]$, and above it the relevant function is $\bar{H}(\tilde{e}_1) = G\left[e_1^{*-1}(\tilde{e}_1; \bar{a}_1, a_2)\right]$. Given that all providers with a $\tilde{z} \in [\underline{z}, \bar{z}]$ have to choose $e_1^* = UL$, an entire mass that would have exerted an effort $\underline{e}_1(\tilde{z}) > UL$ if there were no kink is now collapsed at that single point and has a value of $b = h(UL) = \underline{H}(\underline{e}_1(\bar{z})) - \underline{H}(UL)$. Above $\bar{z}$, the distribution will follow $\bar{h}(e_1)$

Figure 2 extends the previous example and considers a uniform density $g(z)$ and how it transforms into $h(e_1)$. For $z < \underline{z}$, the kink makes no difference at all: $h(e_1) = \underline{h}(e_1)$. However, for those $z \in [\underline{z}, \bar{z}]$ there is a clear change. Without the kink, such provider would have exerted $e_1 \in [UL, UL + \Delta e]$, between points A and D in the figure, which would have followed the density $\underline{h}(e_1)$. Because of the kink, AD became AC and the entire area b is now collapsed into a unique spike at $e_1 = UL$. Finally, for $z > \bar{z}$ we have that optimal effort is given by $\bar{e}_1(z)$, which is reflected by density $\bar{h}(e_1)$. Notice that it is required that $1 - \bar{H}(UL) = 1 - \underline{H}(UL)$, so the final $H(e_1)$ is a valid CDF. This is reflected in the fact that all observations that would have covered $e_1 \in [UL + \Delta, \infty)$, are now spread into $e_1 \in [UL, \infty)$.[10]

[8] If the assumption does not hold, $e_1^* = 0$ as discussed before. A milder version would be when $\underline{a}_1 \cdot C_{22} - a_2C_{12} > 0$ but $\bar{a}_1 \cdot C_{22} - a_2C_{12} < 0$. In such a case $e_1^* = UL$ will always be preferred for all $z \geq \underline{z}$. This implies that there should not be no provider above UL, regardless of the value of z.

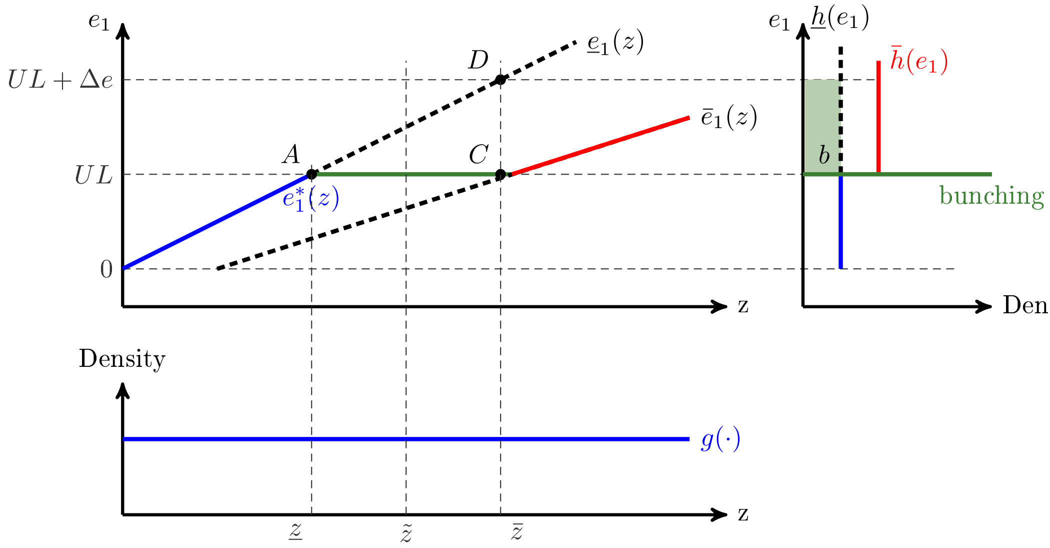
[9] See Kleven [2016] for a good and intuitive explanation on why bunching arises at income distribution as a results

Figure 1: The effect of a kink on rewards at $e_1 = UL$


e_1
$\underline{e}_1(z) = e_1^*(z, \underline{a}_1, a_2)$
$\bar{e}_1(z) = e_1^*(z, \bar{a}_1, a_2)$
B'
A
B
C
UL
$e_1^*(z)$
B''
z
$\underline{z}$
$\tilde{z}$
$\bar{z}$

Note: Providers' payment for task 1 effort below UL is $\underline{a}_1$, and above it is $\bar{a}_1 < \underline{a}_1$. It produces a piecewise optimal effort function $e_1^*(z) = \underline{e}_1(z) \times \mathbb{1}(e_1 \leq UL) + \bar{e}_1(z) \times \mathbb{1}(e_1 > UL)$, where $\underline{e}_1(z) = e_1^*(z, \underline{a}_1, a_2)$ and $\bar{e}_1(z) = e_1^*(z, \bar{a}_1, a_2)$. This diagram assumes constant second derivatives of function $C(e_1, e_2)$. It is also assumed that both tasks are substitutes, so the slope above UL is smaller than below it (see Assumption 1). Nevertheless, in the diagram $\tilde{a}_1 C_{22} - a_2 C_{12} > 0$, hence the values of e_1^* above UL are feasible.

Figure 2: The effect on e_1 density of a kink on rewards at $e_1 = UL$



Note: Providers' payment for task 1 effort below UL is $\underline{a}_1$, and above it is $\bar{a}_1 < \underline{a}_1$. It produces a piecewise optimal effort function $e_1^*(z) = \underline{e}_1(z) \times \mathbb{1}(e_1 \leq UL) + \bar{e}_1(z) \times \mathbb{1}(e_1 > UL)$, where $\underline{e}_1(z) = e_1^*(z, \underline{a}_1, a_2)$ and $\bar{e}_1(z) = e_1^*(z, \bar{a}_1, a_2)$. This diagram assumes constant second derivatives of function $C(e_1, e_2)$. It is also assumed that both tasks are substitutes, so the slope above UL is smaller than below it (see Assumption 1). Nevertheless, in the diagram $\tilde{a}_1 C_{22} - a_2 C_{12} > 0$, hence the values of e_1^* above UL are feasible.

2.3 Extensions

The model presented above provides the basic intuition for understanding the implications of the two-tariff payment system on the responsiveness of agents' effort in a task to marginal variations of the financial rewards of alternative ones. Nevertheless, the model does not consider realistic scenarios such as the presence of uncertainty, or the potential preference for certain cash-flow that agents might exhibit (risk aversion). Appendix A shows that under such conditions, in general terms the main results still hold. As expected, extreme scenarios such as very high levels of variance of the unobserved component of the output, will result in the irrelevance of the kink.

3 Empirical Test

In this section we present how we implement the test for determining whether a specific task is a complement or a substitute of a set of tasks for which there was an observed variation on the reward per unit of effect. The general concern is that if such a shock to the system occurs, normally it should affect all agents who are under the same contract. It involves two steps. First, two tests are presented in order to determine if there is bunching at the upper limit of a given indicator. It this is the case, for this specific indicator we can establish a set of agents that will not react to a variation in the reward per unit of effort in other tasks. These agents, who bunch themselves above UL, constitute a control group that motivates a difference-in-differences (DiD) approach (Equation 15 below). As a treatment group, agents that originally reported a level of output below the kink point UL are selected.

In the subsections below, the motivation and identification arguments for the DiD are discussed first, and the tests for bunching afterwards.

3.1 Test Specification

Our object of interest is the sign of C_{12}. According to proposition 1, the sign of C_{12} is the same as the sign of $\frac{de_1}{da_2}$. In this section, we explain how we can use data from a random sample of agents to estimate the sign of $\frac{de_1}{da_2}$ (and hence the sign of C_{12}).

Assume that we have available a random sample of N agents, observed consecutively for three time periods (t = 1,2,3). For each agent and time period we observe their task 1 output, that is, $\{x_{1it}\}_{i=1,t=1}^{N,3}$. Assume that the payment function for output x_1 is exactly as (8) and is the same in the three time periods. On the contract, assume that the piece rate for task 2 output is the same in the first two time periods, but changes in the third time period: $a_{2t}{=}a_2^{'}$ if $t = 1, 2$; and $a_{2t}{=}a_2^{''}$ if $t = 3$. Without loss of generality, we assume that $a_2^{''}{<}a_2^{'}$.

We will represent agent i's observed level of task 1 output at time t by:

$$x_{1it} = e_1^* \left(a_{2t}, z(i)\right) + \theta_{1i} + \lambda_{1t} + \epsilon_{1it}, \tag{9}$$

of the presense of kinks on income taxes.

[10] For the uniform example in Figure 2, this means that the maximum value of e_1 will fall, but the density at any point will be larger ($\bar{h}(e_1) > \underline{h}(e_1)$ for $e_1 \in [UL, \bar{e}_1(z^{max})]$). See the example in the Appendix for more details.

where $e_1^*(a_{2t}, z(i))$ represents agent i's effort choice on task 1 when he faces a_{2t} as task 2 piece rate, and $z(i)$ is agent i's efficiency parameter.[11]We allow for the measured level of x_{1it} to differ from the agent's optimal choice due to a agent fixed component, θ_{1i}, a time component common across agents, λ_{1t}, and an independent and identically distributed random error term ϵ_{1it}, which exhibits zero mean and finite variance.

Using the above, the change in agent i's observed task 1 output between the third and second time period is given by:

$$x_{1i3} - x_{1i2} = e_1^*(a_2'', z(i)) - e_1^*(a_2', z(i)) + \lambda_{13} - \lambda_{12} + \epsilon_{1i3} - \epsilon_{1i2},$$

where we are using that the task 2 piece rate, a_{2t}, changed from $a_2^{'}$ to $a_2^{''}$ between these two time periods. We will specialise the above expression according to whether the agent's efficiency parameter, $z(i)$, is such that $z(i) \in [\underline{z}, \bar{z}]$, and hence agent i's optimal effort corresponds to the kink ($e_1^* = UL$), or when $z(i) < \underline{z}$, and hence the exerted effort is higher. Moreover, we assume that $a_2^{'}$, and $a_2^{''}$ are sufficiently close, so that $e_1^*(a_2', z(i)) = e_1^*(a_2'', z(i)) = UL$ if $z(i) \in [\underline{z}, \bar{z}]$ (see proposition 2). This means that $e_1^*(a_2'', z(i)) - e_1^*(a_2', z(i)) = 0$ for the group of agents for which $z(i) \in [\underline{z}, \bar{z}]$. Hence, we have that:

$$x_{1i3} - x_{1i2} = \lambda_{13} - \lambda_{12} + \epsilon_{1i3} - \epsilon_{1i2} \quad \text{if} \quad z(i) \in [\underline{z}, \bar{z}] \tag{10}$$

$$x_{1i3} - x_{1i2} = e_1^*(a_2'', z(i)) - e_1^*(a_2', z(i)) + \lambda_{13} - \lambda_{12} + \epsilon_{1i3} - \epsilon_{1i2} \quad \text{if} \quad z(i) < \underline{z} \tag{11}$$

Taking expectations of (10) and (11) over the relevant group of agents, and subtracting one from the other, we have that:

$$\Delta{=}E_{i\epsilon\{i:z(i)<\underline{z}\}}[x_{1i3} - x_{1i2}] - E_{i\epsilon\{i:z(i)\in[\underline{z},\bar{z}]\}}(x_{1i3} - x_{1i2}) = E_{i\epsilon\{i:z(i)<\underline{z}\}}[e_1^*(a_2'', z(j)) - e_1^*(a_2', z(j))] \tag{12}$$

.

Note that the left hand side of (12), $E_{i\epsilon\{i:z(i)<\underline{z}\}}[e_1^*(a_2'', z(j)) - e_1^*(a_2', z(j))]$, is the discrete approximation to (-$\frac{de_1}{da_2}$) (averaged over the set of agents i for which $z(i) < \underline{z}$), whose sign is the same as the sign of C_{12}, our object of interest, and hence the sign of C_{12}. [12] We can estimate the sign of $E_{i\epsilon\{i:z(i)<\underline{z}\}}[e_1^*(a_2'', z(j)) - e_1^*(a_2', z(j))]$, by estimating the sign of the coefficient γ_1 in the following difference-in-difference regression:

$$x_{1i3} - x_{1i2} = \gamma_1 \mathbb{1}(z(i) < \underline{z}) + v_{ijt} \tag{13}$$

which implicitly uses the idea that those agents whose $z(i)$ is between $[\underline{z}, \bar{z}]$ can be used as a control group, because they choose to be at the kink of the payment function of $x_1$and hence are insensitive to small changes in a_2, the piece rate of the other task: x_2.

A problem with implementing (13) is that neither $z(j)$ nor $\underline{z}$ will generally be observable to the econometrician. To address this problem, one could estimate the following regression diff-in-diff regression:

[11] For ease of notation, we do not make explicit that the agent's optimal choice of task 1 effort, $e_1^*(\cdot)$, also depends on the payment function of x_1 as well as agent $i's$ cost function. These elements are assumed to be constant along the sample period.

[12] Note that we place a minus in front of $\frac{de_1}{da_2}$ because we assumed that $a_2''{<}a_2'$.

$$x_{1i3} - x_{1i2} = \beta_1 \mathbb{1}(x_{1i2} < UL) + v'_{ijt} \tag{14}$$

where we are using the idea that those agents whose $z(i)< \underline{z}$ are those that have a output level below the kink ($x_{1i2} < UL$), because the individuals that choose to produce at the kink (UL) are those with $z(i) \in [\underline{z}, \bar{z}]$. While it is feasible to estimate (14), a problem is that x_{1i2} depends on the random component $\varepsilon_{1,i2}$, which introduces a bias due to mean reversion. That is, there might be agents for which $e^*_{1i2} > UL$ but due to a large negative transitory shock, $\varepsilon_{1,i2} < 0$, they end up with $x_{1i2} < UL$. In the following time period, $t = 3$, we expect x_{1i3} to be larger or equal to UL, even if $a_{2t}$was the same in both $t = 2$ and $t = 3$. To net out this mean reversion bias, we need to estimate the following regression:

$$x_{1it} - x_{1it-1} = \alpha_1 \mathbb{1}(x_{1it-1} < UL) + \alpha_2 \mathbb{1}(t = 3) + \alpha_3 \mathbb{1}(x_{1it-1} < UL) \cdot \mathbb{1}(t = 3) + v''_{ijt},\ t = 2, 3 \tag{15}$$

where the estimate of α_1 absorbs the mean reversion effect, and the sign of the estimate of α_3 will have the same sign as $E_{i\epsilon\{i:z(i)<\underline{z}\}}[e^*_1\left(a''_2, z(j)\right) - e^*_1\left(a'_2, z(j)\right)]$, and hence the same sign as C_{12}.

An additional characteristic of payment systems is that groups of tasks might share important characteristics which make their output to be correlated beyond the cost function. In the case of QOF, clinical indicators are grouped in areas given by specific pathologies and domains of care (e.g. diabetes, chronic heart disease). For all indicators j and h which are part of a set J, we can allow $cov(v_{jit}, v_{hit}) \neq 0$. This is implemented using seemingly unrelated regressions (SU)R and augmenting the econometric model in order to cover the entire domain of x_j. In order to implement the SUR system of equations, each of them have to rewritten as follows:

$$\begin{aligned} x_{jit} - x_{jit-1} \quad = \quad & \alpha_1 \mathbb{1}\left\{x_{jit-1} \in [UL - 10, UL)\right\} + \alpha_2 \mathbb{1}\left\{t = 3\right\} + \alpha_3 \mathbb{1}\left\{x_{jit-1} \in [UL - 10, UL)\right\} \cdot \mathbb{1}\left\{t = 3\right\} \\ & + \alpha_4 \mathbb{1}\left\{x_{jit-1} \in [0, UL - 10)\right\} + \alpha_5 \mathbb{1}\left\{x_{jit-1} \in [0, UL - 10)\right\} \cdot \mathbb{1}\left\{t = 3\right\} \\ & + \alpha_6 \mathbb{1}\left\{x_{jit-1} \in (UL + 5, 100]\right\} + \alpha_7 \mathbb{1}\left\{x_{jit-1} \in (UL + 5, 100]\right\} \cdot \mathbb{1}\left\{t = 3\right\} + v_{jit} \end{aligned}$$

where the additional terms correspond to the excluded domains for the given indicator. The censoring is done in this format as there are practices which are located close to the kink for some indicators but not for others. Thus, restricting the sample to only those practices which are always near the kink might induce particular selection of the sample in terms of relevant unobserved characteristics.

3.2 Detection of Bunching

It is necessary to construct a counterfactual distribution of achievement in order to detect the existence of bunching. First, we consider the basic strategy for bunching developed by Kleven [2016]: fit a parametric model on the observed distribution excluding an interval around UL, and compare it with the observed distribution. Moreover, if financial rewards play a big role in effort allocation, they will affect the entire shape of the distribution above UL, not only an interval around the threshold. For this reason, we borrow a concept from regression discontinuity design. Essentially, if agents' effort is the main driver of achievement, this will produce not only bunching at UL but a discontinuity on the density at that point. By running a standard McCrary [2008] test, we can determine if this is the case for a given estimator without imposing an assumption on the endogenous shape of the density.

In both exercises, our output variables are the histograms of the indicators. For this purpose we define bins on achievement following McCrary's procedure ($\tilde{x}_h$) and count the number of agents in each bin (n_{hj}).[13]

Bunching strategy We fit restricted cubic splines on the histogram excluding the interval $[UL_j, UL_j + L]$.[14] This strategy essentially splits the domain into segments defined by K knots (joint points) in order to fit the histogram (n_{hj}) of indicator j with a piece-wise cubic polynomial in the middle segments, and a linear function in the first and last ones. It requires the transformation of the domain variable (the midpoint of the bins, $\tilde{x}_h$) into $K-1$ constructed variables $\left(X_{jh}^{(k)}\right)$ that ensure that the resulting function's first and second derivatives are the same.[15] Such variables are included in the linear expression presented in Equation 16 which also considers dummy variables that indicate the presence of an excluded bin ($\mathbb{1}\left\{\tilde{x}_h = l\right\}, \forall l \in [UL_j, UL_j + L]$). The error term, u_{jh}, is assumed to be i.i.d. and normally distributed.

$$n_{jh} = \sum_{k=1}^{K} \omega_k X_{jh}^{(k)} + \sum_{l=UL}^{UL+L} \gamma_l \mathbb{1}\left\{\tilde{x}_h = l\right\} + u_{jh} \tag{16}$$

After the vector of parameters $\{\omega, \gamma\}$ is estimated, the counterfactual density is the predicted value of this equation without the dummies for the excluded range's contribution: $\hat{n}_{jh} = \sum_{k=1}^{K} \hat{\omega}_k X_{jh}^{(k)}$. Then, the excess number of observations that bunch above UL relative to the calculated counterfactual is the difference between the observed and counterfactual histograms in the excluded range. This is equivalent to the sum of the omitted dummies γ:

$$\tilde{b}_j = \sum_{l=UL}^{UL+L} \hat{\gamma}_l = \sum_{l=UL}^{UL+L} \left(n_{jh} - \hat{n}_{jh}\right)$$

Following Chetty et al. [2009], we compare the amount of excess bunching with the average density per 1 pp. in the excluded range

$$b_j = \frac{\tilde{b}_j}{\frac{1}{L+1}\sum_{l=UL}^{UL+L} \hat{n}_{jh}}$$

In case there is bunching, the estimated b_j overestimates the amount of it. The reason is that it does not consider that some of the bunched observations in the interval $[UL_j, UL_j + L]$ should be above $UL_j + L$ in the counterfactual distribution, as predicted by the model.[16] As our goal is

[13] More precisely,

$$n_{jh} = \sum_{i=1}^{N} \mathbb{1}\left\{\frac{\tilde{x}_h - \tilde{x}_{h-1}}{2} \leq x_{ij} < \frac{\tilde{x}_{h+1} - \tilde{x}_h}{2}\right\} \;,\; \tilde{x}_h \in \{0.5, 1, 1.5, ..., 99.5\}$$

[14] While Kleven [2016] recommends polynomials, such functions might produce poor approximations in certain cases [Harrell, 2015, Chap 2.4.2]. Spline interpolation is a parametric approach that is as easy to implement as a polynomial, without several of its limitations.

[15] The procedure was implemented in STATA 13 using mkspline command, using 5 to 7 knots determined by percentiles recommended in Harrell [2015, Chap 2.4.6].

[16] Chetty et al. [2009] correct for this using an iterative procedure in which the area above $UL_j + L$ is artificially increased in such a way that the area under both the observed and counterfactual densities is the same.

to determine whether or not there is bunching, we perform a joint significance test of the omitted dummies from Equation 16:

$$H_0 : \sum_{l=UL}^{UL+L} \hat{\gamma}_l = 0 \tag{17}$$

RDD strategy In the context of the regression discontinuity design (RDD), McCrary [2008] introduced a test for the continuity of the log-density $g(x)$ at a given point:

$$\iota = \ln \lim_{\tilde{x} \downarrow UL} g(\tilde{x}) - \ln \lim_{\tilde{x} \uparrow UL} g(\tilde{x})$$

The basic idea behind it is that if a treatment were assigned according to being above or below such a point, individuals would try to 'choose' their position in the domain in order to obtain or avoid the treatment. Such self-selection would induce a discontinuity on the density. In the bunching literature a discontinuity is not necessary as it allows for a noisy relationship between individual choices and observed outcomes. However, if such a noise is not present, the excess of density at one point will induce a drastic change in the density at such a point.

The estimation of the jump on the log-density, $\hat{\imath}$, is undertaken following McCrary's procedure. First, the bin size is determined according to the standard deviation of the indicator and the total number of indicators. Second, a bandwidth is selected based on the non-parametric estimator literature.[17] Given the bandwidth, local linear regressions are fitted to both sides of UL. Finally, the estimator tests whether the fitted function is continuous at UL.

3.3 The importance of bunching

The presence of the kink at UL is essential for the test. If there were no corner solution near this point, the expression in Equation 12 would deliver misleading results. From Equation 6, and assuming as before that the sole source of heterogeneity is the efficiency parameter z, we can derive the predicted sign of Equation 12 if the UL does not produce bunching. As shown in Figure 2, a higher level of z implies a higher level of e_1. As a result, when we compare the e_1^* response to variation in a_2 for an agent with high x_1 with one with low x_1, we are comparing an agent with a high vs. low value of z. Then, the essential question here is how $\frac{de_1}{da_2}$ changes along z. Equation 18 answers that question, and shows that its sign is determined by the sign of C_{12}, just like the derivative itself. For substitutes ($C_{12} > 0$), the derivative is negative ($\frac{de_1}{da_2} < 0$) and becomes even more negative with higher values of the productivity parameter $\left(\frac{d^2 e_1}{dz da_2} < 0\right)$. For complements the opposite is true.

$$\frac{d^2 e_1}{dz da_2} = -\frac{C_{12}}{C_{11} C_{22} - C_{12}^2} \tag{18}$$

Equation 18 has a strong implication for the test described above. Essentially, if the sorting is based on overall productivity, z, and there is no bunching, the term Δ presented in Equation 12 will produce a result that is opposite to the test result. In order to illustrate this, let us compare the

[17] In a few cases, the suggested optimal bandwidth is beyond the domain of the indicator (i.e. upper limit above 100%). In such case, we set the bandwidth to be equal to $100 - UL$.

response of two practices, one below UL with a productivity $\underline{z}$ and the other above such a cut-off with $\bar{z}$. The sorting of e_1^* implies that $\bar{z} = \underline{z} + \iota$, where $\iota > 0$. As shown below, if we approximate Equation 12 with derivative, it is clear that the sign of Δ is the same as the sign of C_{12}, exactly the opposite result from the one stated in the test description.

$$
\begin{aligned}
\Delta &= E_{i\epsilon\{i:z(i)<\underline{z}\}}[x_{1i3} - x_{1i2}] - E_{i\epsilon\{i:z(i)\in[\underline{z},\bar{z}]\}}(x_{1i3} - x_{1i2}) \\
&\approx \frac{de_1(\underline{z})}{da_2} - \frac{de_1(\bar{z})}{da_2} \\
&= -\frac{\underline{z} \cdot C_{12}}{C_{11}C_{22} - C_{12}^2} + \frac{\bar{z} \cdot C_{12}}{C_{11}C_{22} - C_{12}^2} \\
&= (\bar{z} - \underline{z}) \frac{C_{12}}{C_{11}C_{22} - C_{12}^2} \\
&= \frac{\iota \cdot C_{12}}{C_{11}C_{22} - C_{12}^2}
\end{aligned}
$$

The previous derivation was based on particular sorting with respect to overall efficiency z. However, sorting might be along other dimensions so no reliable test can be derived based on such a difference. For instance, if heterogeneity is only based on the efficiency of task 2, Δ might always be negative regardless of the sign of C_{12}. See the example in Appendix B.1 for more details.

4 An application: The Quality and Outcomes Framework

4.1 Background

The program that we analyse, the *Quality and Outcomes Framework* (QOF), was introduced in 2004 as part of major reform with the aim of improving service and reducing inequality in the quality of care received. It is a financial reward system for achieving a set of administrative and clinical goals. The level of achievement of these goals is monitored by a regional commissioner. Every year, the NHS and the physicians trade union, the *British Medical Association*, negotiate which indicators should be included and how much money should be paid for each one. Rewards are defined according to a point system, which is based on indicators. Administrative indicators are usually binary questions, where the practice obtains all of the points assigned to an indicator if a certain requirement is fulfilled. On the other hand, most clinical indicators are a non-linear function of the proportion of patients that received a certain standard of care. This will be explained in detail in the next section. Changes to the system have been proposed by the *National Institute for Health and Care Excellence* (NICE), but still have to be negotiated by the interested parties. These indicators are one of the most significant contributions of the program, as they provide an image of the quality of primary care services that was not available before. All the information is published yearly by the NHS at GP practice level and is the main source of data for the present study.[18]

Clinical indicators are related to management of chronic diseases and public health concerns. They cover chronic patients that require specific treatments such as those with coronary heart disease, heart failure or diabetes. Moreover, it involves lifestyle advice for smoking, obesity and

[18] Currently date is archived by *NHS Digital* at `http://digital.nhs.uk/qof`.

primary prevention of cardiovascular diseases in general. Since their introduction, several areas have been removed or introduced or indicators replaced.

Analysis of multitasking on the QOF starts with the introduction of the system. The first order concern was to determine whether the programme had a negative impact on unmeasured (thus, unrewarded) indicators of care, one of the possible outcomes predicted by Holmstrom and Milgrom [1991] and Baker [1992]. Sutton et al. [2010] studied a panel of medical records collected before and after the introduction of the programme in Scotland, which included both rewarded and unrewarded outcomes. They claim that after the introduction of the programme there was an improvement in record-keeping for both type of outcomes with respect to the pre-programme trend, but this was larger for those rewarded measures. This was the case for recordings on blood pressure, cholesterol and smoking, which were rewarded, against BMI and alcohol consumption, which were not. Doran et al. [2011] did a similar exercise for a sample of practices in England, but in this case they had access to prescription and biomarkers data, and they obtained similar results. In both studies, as unrewarded measures are affected by the reallocation of effort generated by the introduction of rewards, the identification of the effects of multitasking relies on the validity of using extrapolated pre-treatment trends as a counter-factual. This has also motivated theoretical work on the optimal design of the system. such as Eggleston [2005] and Kaarboe and Siciliani [2011].

As the QOF is adjusted almost every year, a second generation of the analysis followed these innovations. A first set of changes was introduced in 2005/06, where the payment thresholds were revised for some indicators making it more difficult to achieve the maximum number of points. Feng et al. [2015] compared the evolution of the modified and unmodified indicators in Scotland, and showed that performance increased for the affected measures.

A final element to consider is gaming of the system. The main concern is called exception reporting for clinical indicators, which consists of declaring that a patient should not be treated according to the QOF guidelines due to specific health conditions. By increasing the number of excepted patients, the relevant indicator will increase without providing extra services. Gravelle et al. [2010] showed that GP practices exempt relatively more patients from being considered for some of the clinical indicators if the overall achievement in the previous year was below UL, than if it was above this threshold. For our purposes, cheating implies that some practices with productivity $z_0 - \eta$ would report having productivity z_0. This would be a problem for our estimates if those cheating above UL adjusted their reported effort in response to changes in the price of alternative tasks.

Panel A of Table 1 presents the number of practices in the financial years 2009, 2010 and 2011 and their average number of patients (list size). There are around 8000 GP practices covering on average 7000 patients. Panel B shows the mean achievement per domain in each year, which is very close to 100% in all years. The big increase from 2010 to 2011 is due to the removal of some of the indicators, which will be discussed in the next section. Panel C presents the total clinical points (2009) assigned to those conditions with the highest prevalence in the population, according to the QOF data reports. Such points assignments provide an idea on the areas where the NHS considered it a priority to improve and standardize health care. In 2009, diabetes was the most rewarded clinical area with 100 points out of 697 available for the clinical indicator, followed by hypertension and CHD. While these are also some of the most common chronic conditions, relevance is not the sole criteria. For instance management of new cases of depression in the previous years received more points than asthma, even though the latter was the second most common chronic disease after hypertension.

Table 1: GP Practices and QOF Descriptives

Panel A: Main Characteristics	Average by practice and year		
	2009	2010	2011
Number of patients (list size)	6602.84	6691.28	6835.62
Number of practices	8305	8359	8124
Panel B: QOF achievement	Average by practice and year		
	2009	2010	2011
Clinical	95.86	96.75	97.01
Organisational	96.34	97.36	96.37
Patient Experience	71.47	72.60	98.95
Additional Services	95.35	97.13	97.02
Total	93.69	94.66	96.91
Panel C: Selected Raw Prevalences and QOF points for 2009			
	Points	Mean	Std Dev
Diabetes †	100	4.28	1.85
Hypertension	81	13.53	4.79
Asthma	45	5.95	2.29
Coronary Heart Disease	87	3.45	1.49
Depression new cases †	53	0.76	0.80

Notes: Own calculations based on QOF data published in NHS Digital. † Diabetes raw prevalence is underestimated as it is calculated as the number of individuals aged 17 and over with diagnosed types I or II, over the total list size (without age distinction). New cases of depression are those patients diagnosed with the disease during the last financial year (April 1 to March 31).

4.2 Payment system

In our analysis we will consider that for a GP practice, the marginal benefit of exerting effort on a task is a linear function that involves both altruism and monetary payments. Hence, the marginal reward above UL for task j, which we called $\bar{a}_1$ in subsection (2.2), refers to the altruistic motive.[19] Our analysis is based on data from the years 2009 to 2011. In 2009 and 2010, GPs could obtain up to 1000 points: 697 for the clinical domain, 167.5 for the organizational domain, 91.5 for patient experience, and 44 for additional services. In 2011, the clinical domain was reduced to 661 and patient experience to 33, and 262 points were rellocated to organizational indicators. Points are translated into income depending on the size of the practice and how common the underlying health condition is in the practice's population.[20]

Monetary payments in the QOF are determined by achievement according to a set of indicators, of which there are two main types: binary and ratios. The former gives a fixed amount of points if a condition is attained.[21] For instance, indicator BP1 gives 6 points if there is a register of people with established hypertension, or 0 points if there is not. On the other hand, the awarded points for ratio based indicators depend on the number of patients that should potentially receive a given treatment (denominator), and the number of those who effectively receive it (numerator) during a

[19] The assumption of a linear benefit to patients' welfare is relaxed by Kaarboe and Siciliani [2011]. In such a scenario, the relevant function is not $C(\cdot)$ but $B(\cdot) - C(\cdot)$, hence our results will signal complementarity or substituiability of this function.

[20] See Appendix C for further details.

[21] Some administrative indicators also involve ratios. For instance, if there are less than 5 years of records of the blood pressure of patients for 80% of the patients aged 45 and over (indicator RECORD17). In those cases, the number of points allocated follow a binary allocation instead of a piece-rate reward system.

specific period of time.[22] For instance, the definition below for indicators DM17 and ASTHMA6.

> **Indicator ASTHMA6:** The percentage of patients with asthma who have had an asthma review in the previous 15 months
>
> **Indicator DM17:** The percentage of patients with diabetes whose last measured total cholesterol within the previous 15 months was 5 mmol/l or less

If achievement is below a lower limit (LL_j) zero points are awarded, and if it above the upper limit (UL_j) the maximum amount of available points for indicator j are awarded.

Returning to the DM17 indicator example, the lower limit is $LL = 40\%$ and the upper limit is $UL = 70\%$. Then, if at least 70 out of every 100 patients with diabetes have total cholesterol of 5mmol/l or less in the last 15 months, the practice will receive 6 points, the total number of points allocated to this indicator. For ASTHMA6 there are 20 points available and it has the same thresholds $LL = 40\%$ and $UL = 70\%$. A graphic representation of such an assignment rule is presented in the top diagrams in Figure 3, where the horizontal axis presents the possible levels of achievement and the vertical axis represents the number of points that would be awarded according to the QOF rules. Figure 3 also presents histograms for the actual achievement attained by GP practices in each indicator for the 8301 practices in the 2009/10 financial year.[23] From these densities, there are two main points to remark on. First, there are few practices at or close to the lower limit LL; and in fact, most of the distribution is above the UL. The mean achievement for ASHTMA6 was 80% and 83% for DM17 (see Table 2). Less than 6% of the practices attained a level below UL for ASTHMA6, while for DM17 this figure was 2.5%. This is a common element in all indicators that initially exceeded the expectations of the policymakers [Gregory, 2009]. As a result, the main focus of this project is the role of the UL, hence the LL will not be discussed.

Second, as seen for the case of ASHTMA6, there is a sudden increase in the density at UL; in other words, there is *bunching* above the threshold, which is an usual feature of the data produced by discontinuities in budget constraints [Saez, 2010]. However, this is not the case for all of the indicators. This seems to be the case of indicator DM17. According to the model discussed before, this might be either because the financial reward has a minimum impact on the motivation of physicians for accomplishing the goal or due to substantial noise between effort and the measured achievement indicator.[24] Another typical reason for not detecting bunching, the measurement error [Kleven, 2016], is a problem for the present study as the QOF data are based on administrative records for a large number of GP practices.

The other main source of variation in the data is time. Given that between 2009/10 and 2010/11 there were not changes to the QOF indicators, we can understand how achievement changes from period to period. First, while achievement is persistent, there is substantial year-to-year variation. The autocorrelation coefficients are 0.54 for ASTHMA6 and 0.6 for DM17. Second, practices below the UL in one year tend to increase their achievement in the next one. The mean variation for ASTHMA6 is 11 pp. (SD = 14.8 pp.) for those practices below the UL in 2009, but it is -0.2 pp.

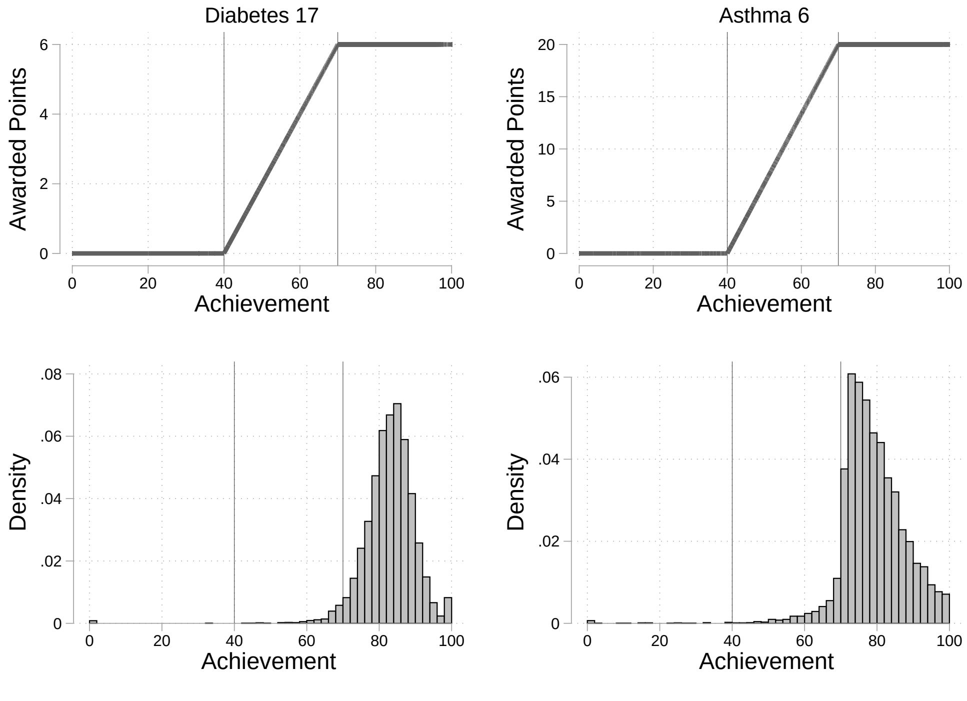
[22] In principle, payment is retrospective, but it is possible to obtain advance payments based on previous year's performance, which are known as *aspiration payments*. More details are available from the BMA [2013].

[23] This includes practices without any cases of hypertension (5 practices) or asthma (8 cases). In those scenarios, zero points are given.

[24] For instance, the staff of the GP practice might have complete control in keeping records of tests or ensuring that patients with a given condition are prescribed a given drug. However, ensuring that the levels of cholesterol of their patients are within certain range, as required by indicator DM17 discussed before, might depend on many actions not controlled by providers. Indeed, Fichera et al. [2014] present a game in which physicians and doctors interact using their available tools, prescriptions and lifestyle, in response to QOF incentives.

(SD = 6.7 pp.) for those above it. Such a mean difference is different from 0 at the 99% level. The same happens for DM17, but with a difference of means of 9 pp. Descriptive statistics for the other indicators are presented in Table 7 in the appendix as the pattern is the same.

Figure 3: Points reward function and achievement density for Diabetes 17 (DM17) and ASHTMA6 (2009/10)



Notes: Own calculations based on QOF data archived at NHS Digital.

Table 2: DM17 and ASTHMA3 QOF indicators descriptives for 2010/11

		(1)	(2)	(3)	(4)	(5)	(6)
Indicator	UL	Number	$E[x_t]$	P$[x_t < UL]$	$\rho(x_t)$	$E[x_t - x_{t-1}$ $\vert x_{t-1} < UL]$	$E[x_t - x_{t-1}$ $\vert x_{t-1} > UL]$
ASTHMA06	70%	8245	79.58	5.29	0.54	11.03	-0.19
DM17	70%	8245	82.73	2.43	0.60	8.70	-0.55

Notes: Own calculations based on QOF data. **Number:** Number of GP practices, including those with 0 elegible patients for the given indicator. $E[x_t]$ **:** Average achievement per indicator. $\mathbf{P}[x_t < UL]$ **:** Proportion of practices with an achivement below UL. $\rho(x_t)$ **:** Correlation between 2010 and 2009 achivement.

4.3 The 2011 changes

While QOF is normally revised every year, there was no change between 2009 and 2010 after an agreement between BMA and NHS during the H1N1 vaccination program [NHS Employers, 2010].

However, between 2010 and 2011 there were major changes that we will interpret as a net reduction in the financial reward per unit of effort for part of the clinical indicators. This time-frame between 2009 and 2011 will be the main source of data for our analysis.

There are in total 1000 QOF points in all three years, but several indicators were either removed, modified or replaced by new ones. We have summarized them in three broad categories presented in Table 3. First, those that imply a reduction in the financial reward per unit of effort; second, those that we interpret as an increase in the marginal benefit; and third, those whose nature is ambiguous. A more detailed explanation of these changes is presented in Table 8 in the appendix.

In the first category (reduction in the financial reward per unit of effort), we include indicators that are withdrawn,[25] increases in UL (which will obviously flatten the slope of the reward function)[26] or changes that consisted of a reduction in the number of points allocated to the indicator. In total 143 of the original clinical points are affected. A different type of change also implied a reduction in the financial reward per unit of effort: these were wording amendments in which the goal definition changed to require either additional tasks or reduce the reference time of the indicator.

The second category (ambiguous change) covers several word amendments that are not straightforward to classify. In these cases typically a more precise definition of the goal to be accomplished is accompanied by additional points in compensation. In total 51 of the original points are in this category.

The third category (increase in the financial reward per unit of effort) includes new indicators as well as old ones with goals that are easier to achieve. The new indicators, covering 12 points, refer to tasks that were not financially rewarded before. Also, for one indicator (17 points) the new wording relaxed the goal defined in the original version.

As we can see, in terms of clinical indicators, the total amount of points related to a reward drop are larger than those associated with an increase, even if we consider all ambiguous changes as increases. Hence, we interpret the overall changes in 2011 as an overall reduction in the marginal payment per unit of effort.

Administrative indicators suffered a major modification in 2011. Two thirds of the *patient experience* domain were removed in favour of the new *quality and productivity* indicators. Practices had to agree a plan with the primary care organisations consisting of three main goals for prescribing (28 points), outpatient referrals (21 points) and emergency admissions (47.5 points). The exact indicator definition and its upper threshold was defined at local level. The objective of the indicators was to reduce costs for the PCT by improving the cost-efficiency of prescribing and by treating more patients at primary care level, reducing both referrals and emergency admission rates.

For the reasons given above, we consider that the main objective of the changes was to *tighten-up* the requirements for obtaining rewards, at least on the clinical side. We will not discuss the administrative indicators, given that almost an entire domain was replaced with an other: the perceived time for getting an appointment was replaced with meetings related to prescribing and other supervised improvement plans designed by the PCT. Because these are administrative tasks, we assume that they were not carried out by doctors themselves and hence that they do not alter

[25] Clinical retired indicators were almost a requirement for measuring other QOF indicators. For instance, indicator CH5 was about having a recent blood pressure record for patients who suffered from coronary heart disease but CHD6 rewards practices for keeping the blood pressure of these patients controlled.

[26] See Equation 25 in the appendix. While the initial proposal was to redefine the UL and make them a function of the underlying indicator distribution in 2011 (match the 75th percentile), the negotiations delivered a slow-paced plan. By 2011 two ULs had increased by one pp. However in 2012 both the lower and upper limits were increased by between 4 to 10 pp. for 13 indicators [Doran et al., 2014].

Table 3: Changes in QOF 2011 with respect to 2009-2010

Panel A. Clinical Indicators

Price Interpretation (Total Points)	**Status**	**Description**	**Points**
Reduction (143 to 87)	Withdrawn	No longer rewarded tasks	32
	Points reduced	Number of assigned points per indicator was reduced.	26 to 22
	Upper Limit Increased	Increase on UL	22
	Replacement I	New wording with more strict definition of a goal or a reduced time-frame for accomplishing it	18
	Replacement II	Decrease in points and new wording is more detailed	45 to 25
Ambiguous (51 to 59)	Replacement III	Harder to accomplish or more detailed goals but compensated with extra points	51 to 59
Increase (29)	Replacement IV	Reference cutoff relaxed	17
	New	New tasks to be rewarded	12
NA (486)	Replacement V	Similar or same wording, but expressed in new units or highlight recent changes on diagnostic procedures.	32
	Unchanged	No change on points, thresholds or wording	454

Panel B. Non-Clinical Indicators

Price Interpretation	**Status**	**Description**	**Points**
Reduction	Retirements	No longer rewarded tasks	60.5
Increase	New	New tasks to be rewarded	96.5
NA	Unchanged	No change on either points or wording	242.5

Note: Authors' interpretation based on NHS Employers public documents.

the marginal cost of clinical effort.

5 Results

The results are presented in two steps. First, we assess the validity of the test by checking for a discontinuity and/or for bunching at the upper limit (UL). Second, we test the sign of the response on effort to a price drop in alternative tasks, on those indicators that were not affected by the QOF 2011 changes.

For the bunching analysis, we pool data from both years 2009 and 2010 and set 10 pp. an estimation window below and above UL. We also discard the bins corresponding to 100%, which is hard to fit with a continuous density function. Figure 4a presents a graphical representation of the McCrary test for continuity on the density at UL for indicators DM17 and ASTHMA6, our examples discussed in the previous section. Both graphs present the histogram (n_{hj}), and the fitted models to both sides of UL . For ASTHMA6 there is clear evidence of the existence of a discontinuity as the null hypothesis that both approximated log-densities are the same at UL is rejected. In both cases the test suggest the presence of a discontinuity on the density at UL. For DM17 such a null cannot be rejected at the 95% level, but it is at the 90% level. Table 4 presents this exercise (Column 4) for each indicator (rows) given a McCrary's default calculations for bin size (Column 2) and bandwidth (Column 3).

The calculation of the amount of excess bunching for both indicators is presented in Figure 4b. Apart from the histogram (n_{hj}), these figures present the fitted model including dummies γ covering $[UL, UL + 5\,\text{pp.}]$ (orange line) and excluding them from the prediction (black line). For DM17, the difference between the histogram and the counterfactual difference is of 42% of the average density in the interval; and for ASTHMA6 it is 107%. Both estimates are significant at the 95% level. However, such estimates are sensible to the number of knots in the spline, the excluded range size L, and the estimation window. Varying the configuration of such parameters we obtain very different point estimates. Columns 5 to 9 in Table 4 present several configurations of an excluded range from $L = 2$ to $L = 7$, 5 and 7 knots, and estimation windows of 10 and 20. For DM17 an estimate of b between -90% and 43%; and for ASTHMA6 it is around 60% to 417%. Despite such large differences, the null in Equation 17 is not rejected for ASTHMA6. On the other hand, for DM17 the null is rejected in 3 out of 5 of the explored specifications.

Given the results stated above, there is clear evidence that the upper limit has an effect on practices, effort allocation for ASTHMA6, but this is not as clear for DM17. Therefore the test is likely to be informative for the first but not the second indicator. Table 4 also suggests that for indicators DM22,[27] SMOKE3[28] and THYROI02[29] there is no evidence of bunching. Table 9 in the appendix presents definitions and graphs equivalent to Figures 4a and 4b for these indicators.

Table 4: QOF indicators corner test

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Indicator	UL	BS	BW	DC Test	w=10, h=2, k=5	w=10, h=3, k=5	w=10, h=3, k=7	w=20, h=3, k=5	w=20, h=5, k=5
AF03	90	0.06	6.38	1.82 ***	123.1 ***	242.4 ***	183.3 ***	176.9 ***	345.4 ***
				[30.72]	[6.53]	[10.96]	[6.11]	[9.34]	[9.73]
AF04	90	0.19	10.00	2.50 ***	171.4 ***	184.7 ***	237.4 ***	137.1 ***	183.7 ***
				[26.28]	[4.10]	[3.88]	[2.67]	[3.82]	[3.17]
ASTHMA03	80	0.11	10.00	1.76 ***	200.5 ***	154.5	89.1	150.4 **	295.5 ***
				[20.65]	[3.31]	[1.41]	[0.51]	[2.13]	[3.32]
ASTHMA06	70	0.13	6.05	0.92 ***	49.3 ***	107.4 ***	57.1 ***	125.4 ***	416.9 ***
				[12.28]	[4.68]	[9.33]	[5.19]	[6.36]	[19.66]
ASTHMA08	80	0.14	10.00	1.93 ***	205.8 ***	254.8 ***	193.4	256.6 ***	492.7 ***
				[27.49]	[4.98]	[3.42]	[1.59]	[5.11]	[8.02]
BP5	70	0.10	5.26	0.31 ***	13.9 *	37.5 ***	42.5 ***	24.1 ***	18.8 **
				[3.60]	[1.92]	[3.99]	[2.83]	[3.94]	[2.32]
CANCER03	90	0.28	10.00	1.41 ***	160.2 ***	235.6 ***	195.5 **	229.2 ***	352.1 ***
				[26.48]	[3.36]	[4.05]	[2.44]	[6.30]	[5.83]
CHD08	70	0.10	6.59	0.30 ***	37.8 ***	39.1 **	68.8 **	14.7	-56.2 ***
				[2.59]	[4.27]	[2.10]	[2.54]	[0.85]	[-2.91]
CHD09	90	0.05	4.57	1.20 ***	77.1 ***	136.2 ***	85.4 ***	57.7 ***	208.4 ***
				[18.55]	[8.68]	[11.54]	[5.81]	[5.48]	[9.06]
CHD10	60	0.15	7.69	1.30 ***	91.9 ***	128.3 ***	112.7 ***	112.4 ***	237.5 ***
				[12.46]	[7.18]	[8.89]	[4.60]	[6.19]	[9.17]
CHD12	90	0.08	4.82	1.21 ***	118.2 ***	222.2 ***	153.4 ***	161.5 ***	323.7 ***
				[22.56]	[10.79]	[20.70]	[17.18]	[16.06]	[14.33]
CKD02	90	0.04	3.42	0.83 ***	289.9 ***	141.9 ***	-1279.0 ***	398.2 **	-166.1 **
				[3.52]	[6.71]	[2.83]	[5.35]	[2.30]	[-2.02]
CKD03	70	0.13	7.73	0.69 ***	82.9 ***	136.9 ***	100.9 ***	164.0 ***	343.5 ***
				[16.24]	[10.73]	[11.64]	[6.81]	[11.02]	[18.86]
CKD05	80	0.16	10.00	2.40 ***	466.9 ***	249.9	83.9	243.7 *	413.4 **

Continued on next page

[27] Based on having a record of glomerular filtration rate (GFR), which measures kidney function.

[28] Proportion of individuals affected by several chronic conditions who are referred to smoking cessation advice.

[29] Record on thyroid function tests.

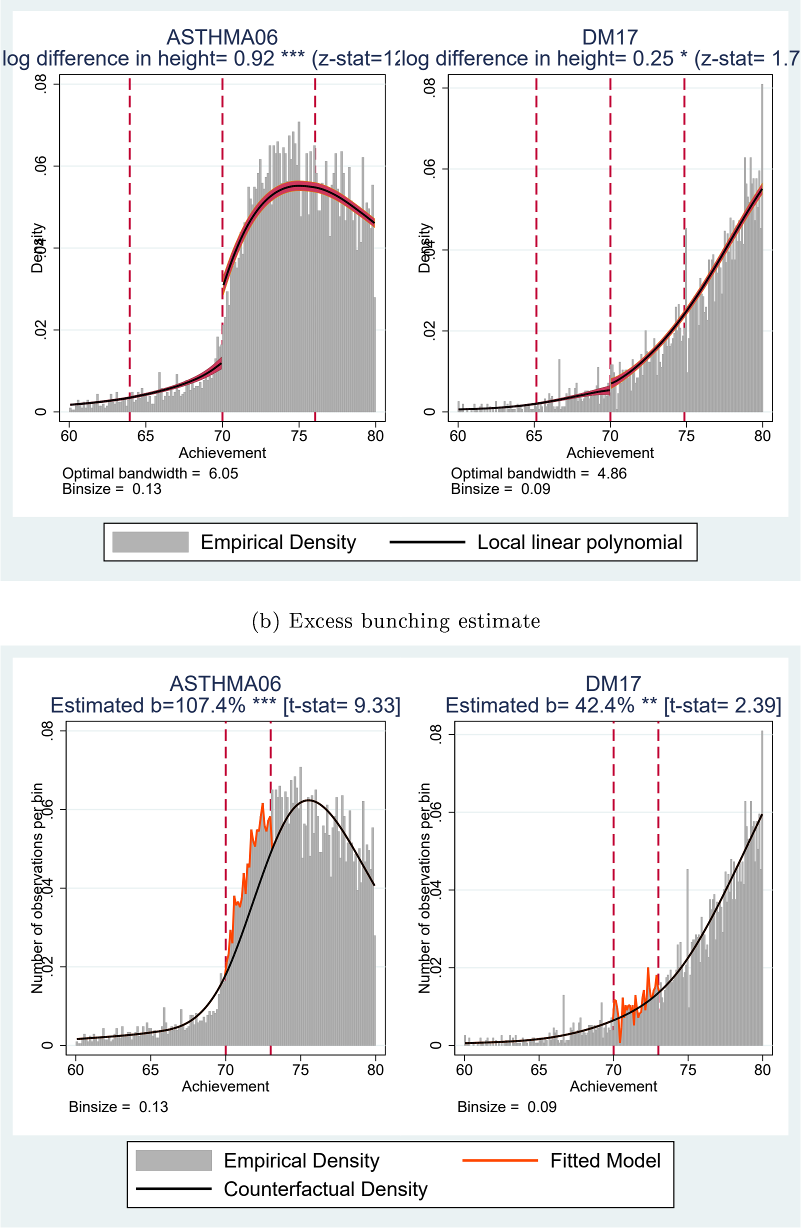
Table 4: (Continued)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Indicator	UL	BS	BW	DC Test	w=10, h=2, k=5	w=10, h=3, k=5	w=10, h=3, k=7	w=20, h=3, k=5	w=20, h=5, k=5
				[18.73]	[3.21]	[1.11]	[0.29]	[1.72]	[2.48]
CKD06	80	0.22	8.89	0.85 ***	77.3 ***	113.6 ***	109.9 ***	109.8 ***	259.6 ***
				[16.84]	[5.82]	[5.70]	[3.40]	[5.11]	[8.42]
CVD01	70	0.28	10.00	1.11 ***	164.8 ***	254.3 **	200.1	179.7 **	499.0 ***
				[13.83]	[4.00]	[2.32]	[1.34]	[2.01]	[5.17]
CVD02	70	0.22	10.00	1.02 ***	97.4 ***	118.1 **	97.6	79.6 *	71.4
				[10.77]	[3.84]	[2.50]	[1.40]	[1.78]	[0.89]
DEM02	60	0.19	10.00	2.31 ***	366.9 ***	359.9 ***	717.4 *	282.1 **	338.2
				[13.60]	[4.56]	[4.71]	[1.92]	[2.40]	[1.30]
DM2	90	0.06	4.22	0.20 **	4.9	-34.2 ***	39.3 ***	-106.8 ***	-141.5 ***
				[2.33]	[0.53]	[-3.78]	[3.16]	[-10.34]	[-6.49]
DM10	90	0.12	4.98	0.74 ***	48.9 ***	108.5 ***	57.6 ***	58.9 ***	214.8 ***
				[13.74]	[4.80]	[9.03]	[6.26]	[5.46]	[9.35]
DM13	90	0.14	4.95	0.75 ***	74.0 ***	175.6 ***	113.8 ***	159.1 ***	400.7 ***
				[16.51]	[5.69]	[11.93]	[14.40]	[12.83]	[19.51]
DM15	80	0.13	10.00	1.45 ***	306.7 ***	223.7 *	172.0	194.6 **	262.3 ***
				[19.09]	[4.69]	[1.83]	[0.84]	[2.57]	[2.70]
DM17	70	0.09	4.86	0.25 *	0.0	42.4 **	42.6	42.5	-93.2 ***
				[1.76]	[0.49]	[2.39]	[1.60]	[1.54]	[-3.44]
DM18	85	0.09	5.60	0.46 ***	30.0 ***	27.7 ***	42.1 ***	14.5 *	65.9 ***
				[6.41]	[4.70]	[3.08]	[3.19]	[1.90]	[6.50]
DM21	90	0.12	5.30	1.07 ***	119.7 ***	222.1 ***	156.1 ***	186.2 ***	376.1 ***
				[22.39]	[9.26]	[17.40]	[14.53]	[15.90]	[21.36]
DM22	90	0.05	4.21	0.27 *	76.2 ***	34.1	601.1 ***	-9.8	-143.7 ***
				[1.87]	[5.44]	[1.63]	[5.28]	[-0.32]	[-2.82]
EPILEP06	90	0.11	8.54	1.38 ***	84.6 ***	66.9 **	181.1 ***	-3.9	-72.3 *
				[19.48]	[2.70]	[1.99]	[2.69]	[-0.18]	[-1.66]
EPILEP08	70	0.21	10.00	1.18 ***	145.7 ***	245.2 ***	208.6 **	250.8 ***	288.7 **
				[23.80]	[4.44]	[3.43]	[2.15]	[3.75]	[2.33]
HF02	90	0.17	10.00	2.41 ***	222.9 ***	259.5 ***	257.5 **	232.6 ***	336.4 ***
				[29.80]	[3.90]	[4.31]	[2.65]	[5.14]	[6.12]
HF03	80	0.11	10.00	2.19 ***	399.7 ***	240.8 *	246.3	221.1 **	293.6 **
				[21.12]	[4.66]	[1.72]	[0.95]	[2.40]	[2.50]
HF04	60	0.19	10.00	2.51 ***	662.3 ***	675.1 ***	-4475.9 *	580.8 **	274.0
				[11.53]	[3.67]	[4.94]	[1.92]	[2.39]	[0.68]
SMOKE03	90	0.04	3.15	0.14	-9.7	-91.0 ***	104.7 ***	-165.7 ***	-238.9 ***
				[1.17]	[-0.58]	[-6.60]	[5.32]	[-12.89]	[-10.12]
SMOKE04	90	0.08	4.86	1.32 ***	132.9 ***	299.2 ***	165.9 ***	272.4 ***	662.6 ***
				[24.79]	[9.08]	[18.04]	[14.88]	[18.08]	[26.30]
STROKE07	90	0.10	6.58	1.12 ***	98.1 ***	179.4 ***	125.2 ***	142.5 ***	308.2 ***
				[25.10]	[6.41]	[8.71]	[4.76]	[8.74]	[11.01]
STROKE08	60	0.13	9.60	0.47 ***	89.9 ***	116.4 ***	114.5	114.5 **	90.4
				[4.25]	[2.99]	[2.85]	[1.30]	[2.49]	[1.12]
STROKE10	85	0.11	7.79	0.98 ***	64.6 ***	103.9 ***	95.3 ***	113.3 ***	262.9 ***
				[17.87]	[4.18]	[5.84]	[3.64]	[7.16]	[11.52]
STROKE12	90	0.07	5.96	1.51 ***	88.3 ***	155.6 ***	128.2 ***	84.2 ***	217.0 ***
				[24.03]	[4.39]	[5.91]	[3.21]	[3.75]	[5.30]
STROKE13	80	0.19	10.00	2.32 ***	450.8 ***	267.3 *	165.3	259.9 **	455.9 ***
				[21.42]	[5.27]	[1.72]	[0.68]	[2.55]	[3.52]
THYROI02	90	0.05	4.48	0.04	-20.3 *	-50.7 ***	33.8	-122.2 ***	-144.6 ***
				[0.31]	[-1.97]	[-4.09]	[1.35]	[-10.19]	[-5.84]

Notes: Own calculations based on QOF data. McCrary test on the continuity of the density at the threshold. Optimal bin sizes (BS) and bandwidths (BW) for each indicator are chosen following McCrary implementation of the test. Significance: * 10%, ** 5%, *** 1%.

Table 5 presents the second part in which we estimate regression (15), where x_1 refers to indicators whose rewards remained unchanged throughout the three years that we consider (2009-2011). We exclude from the analysis those indicators in which the test would not be not valid because bunching was not detected. For each indicator (rows), the table reports the number of observations

Figure 4: Testing for Bunching



(a) McCrary Test

(b) Excess bunching estimate

Note: In sub-figure (a), the empirical densities to both sides of the threshold UL are smoothed using a local linear regression within the given bandwidth of UL (vertical lines). These smoothed functions are presented with a 95% CI. In sub-figure (b), the empirical density is fitted with a restricted cubic spline based on 5 knots. Domain was restricted to a 10 pp. window around UL, and the excluded range is $[UL_j, UL_j + 5pp.]$

above and below the threshold within a 5 pp. window to both sides of the UL (Columns 1 and 2). Such is the selected sample for estimating the parameters of regression (15): columns 3 to 5 of the table presents estimates for α_1, α_2, and α_3. In order to estimate the model, first differences with respect to time are obtained for each GP practice between 2010 and 2011, and between 2009 and 2010. Such a variable is the outcome of the equation. We also construct a binary variable that indicates whether the practice was below the UL in years 2009 and 2010 ($\mathbb{1}(x_{1j,t-1} < UL_j)$), and another that indicates whether we are observing data from the variation 2010 to 2011 ($\mathbb{1}(t = 2011)$). The sample is restricted to a window of $[UL - 10, UL + 5]$. In particular, we are interested in the sign of α_3. Given that we observed a net reduction in the marginal benefit of alternative tasks, a negative sign of $\hat{\alpha}_3$ indicates a positive cross-derivative $\left(\frac{de_1}{da_2} > 0\right)$ which indicates that the analysed task are complementary to the tasks affected by the 2011 changes. This does not mean that the task is a complement of all modified indicators, but that overall, the net response is equivalent to complements. Another possibility is that the task is a substitute only of those tasks for which the marginal reward was increased instead of reduced. This is less likely as the majority of changes correspond to a decrease, rather than an increase, but we cannot rule out such a possibility.

We also note that, for some indicators, we might not be able to reject the hypothesis that$\alpha_3 = 0$ because of lack of power. In particular, there are some indicators that have a very small number of practices below the threshold. For instance, for HF04 there are only 45 practices below UL in comparison with 669 above it.

On one hand, we find that AF03, AF04, ASTHMA06, CKD06, CVD02, DM10 and DM13 are complements of the overall modified indicators: effort was reduced in response to the net reduction in incentives in other indicators. The first (AF03), is the percentage of patients with atrial fibrillation (a rapid and irregular heartbeat) who are being treated with a anticoagulant drug therapy, while the second (AF04) is the percentage of those patients who had their diagnosis confirmed by an specialist or with a specialised test. The third (ASTHMA06) is the percentage of asthmatic patients who had a review of their disease progression in the last 15 months. The fourth (CKD06) is the percentage of patients with chronic kidney disease who have a record for a test that checks their kidney status. The fifth (CVD02) corresponds to those patients diagnosed with hypertension who received lifestyle advice. The last two refer to diabetic patients: DM10 is on having records of neuropathy testing (nerve disorders) and DM13 records of micro-albumuria testing (kidney's status) for diabetic patients. On the other hand, DM18 is the only substitute task identified. This indicator is based on the proportion of diabetic patients who were immunised against influenza.

Alternative estimation windows are considered in Table 6. In this table, each cell presented is an estimate of α_3 considering a sample of $[UL - l, UL + k]$. This table is restricted to those cases in which the hypothesis $\alpha_3 = 0$ is rejected at least once. This means that Column 3 of Table 5 corresponds to the fourth column ($l = 10, k = 5$) of Table 6. Estimates for AF04, CKD06, CVD02 and DM13 are stable across the different specifications. Table 10 in the appendix presents definitions and graphs with the bunching test for these indicators.

The diabetes mellitus (DM) area suffered several changes. There were changes in payments for keeping blood pressure of patients controlled and on records of foot examination. Also, financial rewards for keeping records of plasma glucose concentration, blood pressure and cholesterol were removed. CVD02 is directly related to handling hypertense population, which is directly linked to keeping track of such population. Given that both DM10 and DM13 are also records of recent tests, it seems plausible that such tasks are complements. However, while the result is not robust to the specification, DM18 indicator seems to be working in the opposite direction. The reason for this might be that while we are talking about diabetic patients, here we are considering immunisation, which is tasks relatively different to all other cardiovascular care activities included in the QOF.

Neither the chronic kidney disease nor the atrial fibrillation indicators were modified in 2011. Nevertheless, AF04 and CKD06 are affected by other indicators' changes. AF04 measures the proportion of individuals diagnosed with ECG or by a specialist. CKD06 rewards keeping a record of albumin creatinine ratio, which is a specific measure related to kidney disease.

Table 5: QOF indicators results SUR: Control: UL + 5 pp., Responsive practices: UL - 10 pp.

		(1)	(2)	(3)	(4)	(5)	
		Descriptives †		Estim. Regression Coefficients			Classif.
Indicator	UL	N Below	N Above	BELOW α_1	AFTER α_2	INTER α_3	
AF03	90%	544	4287	0.039^{***}	0.001	-0.009^{**}	Comp
				(0.003)	(0.001)	(0.003)	
AF04	90%	284	1630	0.047^{***}	-0.003^{**}	-0.013^{**}	Comp
				(0.004)	(0.001)	(0.006)	
ASTHMA03	80%	248	1449	0.044^{***}	-0.004	-0.008	
				(0.006)	(0.003)	(0.009)	
ASTHMA06	90%	421	2143	0.050^{***}	-0.008^{***}	-0.016^{***}	Comp
				(0.004)	(0.002)	(0.006)	
ASTHMA08	80%	325	1984	0.042^{***}	-0.006^{***}	-0.013^{*}	
				(0.005)	(0.002)	(0.007)	
BP5	70%	461	1406	0.029^{***}	-0.002	-0.004	
				(0.003)	(0.002)	(0.005)	
CANCER03	90%	897	1768	0.027^{***}	-0.002	-0.007	
				(0.004)	(0.002)	(0.006)	
CHD08	70%	265	761	0.039^{***}	-0.020^{***}	-0.009	
				(0.005)	(0.003)	(0.007)	
CHD09	90%	528	4382	0.025^{***}	0.001	-0.003	
				(0.002)	(0.001)	(0.003)	
CHD10	60%	175	1066	0.048^{***}	-0.001	-0.005	
				(0.008)	(0.003)	(0.010)	
CHD12	90%	1240	4434	0.024^{***}	-0.004^{***}	0.003	
				(0.002)	(0.001)	(0.002)	
CKD02	90%	58	717	0.048^{***}	0.000	-0.005	
				(0.006)	(0.003)	(0.008)	
CKD03	70%	1498	2384	0.021^{***}	0.009^{***}	0.003	
				(0.002)	(0.002)	(0.003)	
CKD05	80%	152	695	0.047^{***}	-0.013^{***}	-0.007	
				(0.009)	(0.004)	(0.014)	
CKD06	80%	1167	1938	0.039^{***}	-0.013^{***}	-0.020^{***}	Comp
				(0.003)	(0.002)	(0.004)	
CVD01	70%	400	744	0.037^{***}	-0.009	-0.003	
				(0.010)	(0.007)	(0.013)	
CVD02	70%	245	593	0.073^{***}	-0.001	-0.041^{***}	Comp
				(0.011)	(0.006)	(0.014)	
DEM02	60%	102	542	0.085^{***}	0.014^{**}	-0.045	
				(0.024)	(0.007)	(0.034)	
DM2	90%	555	2852	0.031^{***}	-0.000	-0.003	
				(0.002)	(0.001)	(0.003)	
DM10	90%	1448	3920	0.027^{***}	0.001	-0.004^{**}	Comp
				(0.002)	(0.001)	(0.002)	
DM13	90%	2151	3755	0.024^{***}	-0.002^{*}	-0.006^{***}	Comp
				(0.001)	(0.001)	(0.002)	
DM15	80%	326	1093	0.034^{***}	0.002	-0.009	
				(0.006)	(0.003)	(0.008)	
DM18	85%	711	2295	0.023^{***}	-0.003^{**}	0.008^{**}	Subs
				(0.003)	(0.001)	(0.004)	
DM21	90%	1585	3837	0.023^{***}	-0.002^{*}	0.000	
				(0.002)	(0.001)	(0.002)	

Continued on next page

Table 5: (Continued)

		(1)	(2)	(3)	(4)	(5)	
		Descriptives		Estim. Regression Coefficients			Classif.
Indicator	UL	N Below	N Above	BELOW α_1	AFTER α_2	INTER α_3	
EPILEP06	90%	489	2376	0.036^{***}	-0.004^{***}	-0.006	
				(0.004)	(0.001)	(0.005)	
EPILEP08	70%	1139	1896	0.024^{***}	0.013^{***}	-0.009^{*}	
				(0.004)	(0.003)	(0.005)	
HF02	90%	463	1860	0.033^{***}	-0.000	0.001	
				(0.004)	(0.001)	(0.005)	
HF03	80%	186	1223	0.043^{***}	-0.003	0.007	
				(0.008)	(0.003)	(0.011)	
HF04	60%	75	367	0.118^{***}	0.016^{*}	-0.026	
				(0.028)	(0.009)	(0.045)	
SMOKE04	90%	950	4411	0.027^{***}	0.001	-0.003	
				(0.002)	(0.001)	(0.003)	
STROKE07	90%	1550	3947	0.022^{***}	-0.003^{***}	-0.003	
				(0.001)	(0.001)	(0.002)	
STROKE08	60%	210	397	0.036^{***}	-0.011^{**}	0.001	
				(0.008)	(0.005)	(0.011)	
STROKE10	85%	904	2724	0.028^{***}	-0.002	0.003	
				(0.003)	(0.002)	(0.004)	
STROKE12	90%	580	3735	0.034^{***}	-0.000	-0.003	
				(0.003)	(0.001)	(0.004)	
STROKE13	80%	191	1116	0.035^{***}	-0.006^{**}	-0.009	
				(0.009)	(0.003)	(0.012)	

Notes: Own calculations based on QOF data. **BELOW:** To have attained below the respective upper thershold in the first year of the variation (2009 for 2009-2010 and 2010 for 2010-2011). **AFTER:** 2010 to 2011 variation. **AFTER:** Interaction between INTER and AFTER. Clustered at partice-level standard errors in parenthesis, which are allowed to be correlated between indicators of the same diagnostic area. † practices' descriptives are presented according to 2010 achievement, within the 5 points window around UL . Significance: ** 5%, *** 1%.

Table 6: QOF indicators results: Multiple windows

Estimate of α_3 under the sample in $[UL-l, UL+k]$
Presents only indicators for which $\alpha_3 = 0$ is rejected in at least one specification.

	Entire interval				Removing $[UL-1, UL+1]$			
	k=5 pp. above UL		k=3 pp. above UL		k=5 pp. above UL		k=3 pp. above UL	
Indicator	l=10	l=5	l=10	l=5	l=10	l=5	l=10	l=5
AF03 (UL=90)	-0.007^{**}	-0.001	-0.007^{**}	-0.001	-0.009^{*}	0.000	-0.009^{*}	0.000
	(0.003)	(0.003)	(0.003)	(0.003)	(0.005)	(0.005)	(0.005)	(0.005)
AF04 (UL=90)	-0.016^{***}	-0.015^{***}	-0.016^{***}	-0.015^{***}	-0.023^{**}	-0.027^{***}	-0.023^{**}	-0.027^{***}
	(0.006)	(0.005)	(0.006)	(0.005)	(0.009)	(0.009)	(0.009)	(0.009)
ASTHMA06 (UL=70)	-0.017^{***}	-0.009	-0.012^{*}	-0.005	-0.020^{**}	-0.010	-0.016^{**}	-0.006
	(0.006)	(0.007)	(0.007)	(0.007)	(0.008)	(0.009)	(0.008)	(0.009)
ASTHMA08 (UL=80)	-0.015^{**}	-0.007	-0.013	-0.004	-0.019^{**}	-0.010	-0.017^{*}	-0.008
	(0.008)	(0.008)	(0.008)	(0.008)	(0.009)	(0.010)	(0.009)	(0.010)
CHD08 (UL=70)	-0.012	-0.012	-0.016^{*}	-0.017^{*}	-0.017^{*}	-0.022^{**}	-0.022^{**}	-0.028^{**}
	(0.008)	(0.009)	(0.008)	(0.009)	(0.009)	(0.011)	(0.010)	(0.011)
CKD06 (UL=80)	-0.021^{***}	-0.018^{***}	-0.018^{***}	-0.015^{**}	-0.021^{***}	-0.018^{***}	-0.017^{***}	-0.014^{**}
	(0.005)	(0.006)	(0.005)	(0.006)	(0.005)	(0.006)	(0.006)	(0.007)
CVD02 (UL=70)	-0.040^{***}	-0.023	-0.037^{**}	-0.020	-0.061^{***}	-0.052^{**}	-0.055^{***}	-0.046^{**}
	(0.014)	(0.018)	(0.017)	(0.020)	(0.016)	(0.021)	(0.018)	(0.023)
DM10 (UL=90)	-0.003	-0.006^{*}	-0.001	-0.004	-0.003	-0.007^{*}	-0.001	-0.005
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)	(0.004)
DM13 (UL=90)	-0.006^{***}	-0.005^{**}	-0.006^{**}	-0.005^{*}	-0.007^{***}	-0.005^{*}	-0.007^{***}	-0.005^{*}
	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)	(0.003)	(0.003)
DM18 (UL=85)	0.009^{**}	0.007	0.008^{*}	0.006	0.011^{**}	0.008	0.010^{**}	0.007
	(0.004)	(0.004)	(0.004)	(0.005)	(0.004)	(0.005)	(0.005)	(0.005)
EPILEP08 (UL=70)	-0.009^{*}	-0.013^{*}	-0.002	-0.005	-0.012^{*}	-0.020^{**}	-0.003	-0.010
	(0.006)	(0.007)	(0.006)	(0.007)	(0.006)	(0.008)	(0.007)	(0.009)
STROKE10 (UL=85)	0.009^{**}	0.006	0.007	0.003	0.010^{**}	0.005	0.008	0.003
	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)

Notes: Own calculations based on QOF data. Clustered at practice-level standard errors in parenthesis.
Significance: ** 5%, *** 1%.

6 Conclusion

This paper introduces a test for complementarities/substitutions in the agent's cost function in a multitasking setting when there is a two-part linear contract. It works by considering as a "control" group those agents who self-select into levels of effort that corresponds to the "kink" in the reward function, that is, at the threshold where there is a sudden change in the marginal benefit for exerting effort in a given task. For these agents, there is a wedge between the marginal benefit and marginal cost of effort, and hence, small changes in incentives will not alter their effort allocation (and hence can be used as a control group). The test consists of two steps: first, determining whether the kink produces "bunching" in the distribution of achievement at the threshold, and if that is the case, a difference in differences estimator identifies the desired characteristic of the cost function.

As a case of study we have analysed a pay for performance scheme for family doctors in the UK, the Quality and Outcomes Framework (QOF). We have shown that changes introduced in 2010/11, which we understand as a net price drop in a set of modified indicators, revealed that several indicators are in fact complements. This might be because most clinical indicators refer to chronic patients, who not unusually have several co-morbidities.

References

Omar Al-Ubaydli, Steffen Andersen, Uri Gneezy, and John A. List. Carrots that look like sticks: toward an understanding of multitasking incentive schemes. *NBER Working Paper*, 18453, 2012.

George P Baker. Incentive contracts and performance measurement. *Journal of political Economy*, pages 598–614, 1992.

BMA. Focus on qof payments. `https://www.bma.org.uk/-/media/files/pdfs/practical%20advice%20at%20work/contracts/independent%20contractors/qof%20guidance/focusonqofpaymentsnov2013.pdf`, 2013. Accesed: 2016-08-09.

Patrick Bolton and Mathias Dewatripont. *Contract theory*. MIT press, 2005.

C. Bradler, R. Dur, S. Neckermann, and A. Non. Employee recognition and performance: A field experiment. *CESifo Working Paper*, 4164, 2013.

Raj Chetty, John N Friedman, Tore Olsen, and Luigi Pistaferri. Adjustment costs, firm responses, and micro vs. macro labor supply elasticities: Evidence from danish tax records. Technical report, National Bureau of Economic Research, 2009.

Tim Doran, Evangelos Kontopantelis, Jose M Valderas, Stephen Campbell, Martin Roland, Chris Salisbury, and David Reeves. Effect of financial incentives on incentivised and non-incentivised clinical activities: longitudinal analysis of data from the uk quality and outcomes framework. *Bmj*, 342:d3590, 2011.

Tim Doran, Evangelos Kontopantelis, David Reeves, Matthew Sutton, and Andrew M Ryan. Setting performance targets in pay for performance programmes: what can we learn from qof? *BMJ*, 348, 2014. doi: 10.1136/bmj.g1595. URL `http://www.bmj.com/content/348/bmj.g1595`.

Etienne Dumont, Bernard Fortin, Nicolas Jacquemet, and Bruce Shearer. Physicians' multitasking and incentives: Empirical evidence from a natural experiment. *Journal of Health Economics*, 27(6):1436 – 1450, 2008. ISSN 0167-6296. doi: http://dx.doi.org/10.1016/j.jhealeco.2008.07.010. URL `http://www.sciencedirect.com/science/article/pii/S016762960800091X`.

Karen Eggleston. Multitasking and mixed systems for provider payment. *Journal of Health Economics*, 24(1):211–223, 2005.

Yan Feng, Ada Ma, Shelley Farrar, and Matt Sutton. The tougher the better: an economic analysis of increased payment thresholds on the performance of general practices. *Health economics*, 24(3):353–371, 2015.

Susan Feng Lu. Multitasking, information disclosure, and product quality: Evidence from nursing homes. *Journal of Economics & Management Strategy*, 21(3):673–705, 2012.

Eleonora Fichera, James Banks, and Matt Sutton. Health behaviours and the patient-doctor interaction: The double moral hazard problem. Technical report, Economics, The University of Manchester, 2014.

Paul Glewwe, Nauman Illias, and Michael Kremer. Teacher incentives. *American Economic Journal: Applied Economics*, 2(3):205–227, 2010.

Hugh Gravelle, Matt Sutton, and Ada Ma. Doctor behaviour under a pay for performance contract: Treating, cheating and case finding?*. *The Economic Journal*, 120(542):F129–F156, 2010.

Sarah Gregory. *General practice in England: An overview*. King's Fund, 2009.

Frank Harrell. *Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis*. Springer, 2015.

Bengt Holmstrom and Paul Milgrom. Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, & Organization*, 7:24–52, 1991.

Oddvar Kaarboe and Luigi Siciliani. Multi-tasking, quality and pay for performance. *Health Economics*, 20(2):225–238, 2011.

Henrik J Kleven. Bunching. *Annual Review of Economics*, 8(1), 2016.

M. Kosfeld and S. Neckermann. Getting more work for nothing? symbolic awards and worker performance. *American Economic Journal: Micoeconomics*, 3:86–99, 2011.

Lazear. Performance pay and productivity. *American Economic Review*, 90(5):1346–61, 2000.

Justin McCrary. Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, 142(2):698 – 714, 2008. ISSN 0304-4076. doi: http://dx.doi.org/10.1016/j.jeconom.2007.05.005. URL `http://www.sciencedirect.com/science/article/pii/S0304407607001133`. The regression discontinuity design: Theory and applications.

Karthik Muralidharan and Venkatesh Sundararaman. Teacher performance pay: Experimental evidence from india. *Journal of Political Economy*, 119:39–77, 2011.

Derek Neal. The design of performance pay in education. In Eric Hanushek, Steve Machin, and Ludger Woessmann, editors, *Handbook of Economics of Education*, volume 4, pages 495–550. Elsevier, Oxford, 2011.

NHS Employers. Changes to qof 2010/11. URL: http://www.nhsemployers.org/your-workforce/primary-care-contacts/general-medical-services/quality-and-outcomes-framework/changes-to-qof-201011, May 2010. URL `http://www.nhsemployers.org/your-workforce/primary-care-contacts/general-medical-services/quality-and-outcomes-framework/changes-to-qof-201011`. Accessed: 2016-08-09.

Martin Roland and Frede Olesen. Can pay for performance improve the quality of primary care? *BMJ*, 354:i4058, 2016.

Emmanuel Saez. Do taxpayers bunch at kink points? *American Economic Journal: Economic Policy*, 2(3):180–212, 2010.

B. Shearer. Piece rates, fixed wages and incentives: Evidence from a field experiment. *Review of Economic Studies*, 71(2): 513–34, 2004.

Matt Sutton, Ross Elder, Bruce Guthrie, and Graham Watt. Record rewards: the effects of targeted quality incentives on the recording of risk factors by primary care providers. *Health economics*, 19(1):1–13, 2010.

A Uncertainty

A common characteristic of multitasking models is the role of uncertainty.[30] In particular, Holmstrom and Milgrom [1991] discuss the role of using noisy signals for rewarding agents. Let us consider $x_1 = e_1 + \varepsilon_1$, where ε_1 is distributed according to $F(\cdot)$, which is a twice differentiable CDF, with PDF $f(\cdot)$which correponds to a symmetric unimodal distribution with mean 0. Let us assume that uncertainty has an impact of $-\Omega < 0$ on utility, which is considered here only because the two-part contract of the QOF has this feature which might not be present in other pay-for-performance schemes, and which will be discussed below in depth.[31] Hence, we can write their problem as follows.

[30] For our particular application, the model without uncertainty is not necessarily too simplistic. This is because the payment is based on the aggregate outcome of the doctor's patients, and hence the noise might be averaged out.

[31] This would be the case with preferences that exhibit absolute risk aversion η. For example:

$$\begin{aligned}\max_{e_1,e_2\in[0,1]} U &= E\left[u(\alpha B(e_1,e_2) + (T + \phi(p_1, e_1 + \varepsilon_1) + a_2 e_2 - \frac{1}{z}C(e_1,e_2))\right] \\ &= E[-e^{-\eta(\alpha B(e_1,e_2)+(T+\phi(p_1,e_1+\varepsilon_1)+\tilde{a}_1 e_1+a_2 e_2)-\frac{1}{z}C(e_1,e_2))}] \qquad (19)\end{aligned}$$

With a linear tariff $\phi_1(x_1) = p_1 x_1 = p_1 e_1 + p_1 \varepsilon_1$ the problem can be expressed in terms of the certainty equivalent $\hat{U}$. Where, despite risk aversion, the noise plays no role in the allocation of effort. This is because a provider's choices do not affect the expected value of the reward for attaining a certain level of performance.

$$\max_{e_1,e_2\in[0,1]} \hat{U} = \alpha B(e_1,e_2) + (T + a_1 e_1 + a_2 e_2) - C(e_1,e_2) - \frac{1}{2}\eta(p_1^2\sigma_1^2) \qquad (20)$$

$$\begin{aligned}\max_{e_1,e_2\in[0,1]} U =& \Pr\left[e_1 < UL - \varepsilon_1\right]\cdot\left\{E\left[(T+\underline{a}_1\cdot e_1+\underline{a}_1\cdot\varepsilon_1+a_2e_2)-\frac{1}{z}C(e_1,e_2)\right]-\Omega\right\}\\ &+\Pr\left[e_1 \geq UL-\varepsilon_1\right]\cdot\left\{(T+p_1UL+\bar{a}_1e_1+a_2e_2)-\frac{1}{z}C(e_1,e_2)\right\}\end{aligned}$$

Where replacing the probabilities with the densities results in the following optimisation problem.

$$\begin{aligned}\max_{e_1,e_2\in[0,1]} U =& F\left(UL-e_1\right)\cdot\left[(e_1-UL)\cdot p_1-\Omega\right]\\ &+p_1\cdot UL+\bar{a}_1e_1+a_2e_2-\frac{1}{z}C(e_1,e_2)\end{aligned}$$

The FOC for e_2 is still the one presented in equation 3, but for e_1 it is now required to consider the probability of attaining an output above UL, as shown below.

$$FOC_1 := \bar{a}_1-\frac{1}{z}C_1+\left\{F\left(UL-e_1\right)\cdot p_1-f(UL-e_1)\cdot\left[(e_1-UL)\cdot p_1-\Omega\right]\right\}=0 \tag{21}$$

Whether the output is above or below UL is important because as long as $e_1+\varepsilon_1 < UL$, part of the marginal financial return $p_1=\underline{a}_1-\bar{a}_1$ is subject to uncertainty (intrinsic motivation is not subject to it). However, by exerting more effort, the probability of loosing such financial reward decreases. Here, for the sake of simplicity, the preference for certainty is modelled by introducing the penalty Ω, which is avoided if the output overcomes the UL threshold. In other words, Ω captures the value that agents give to obtain a certain reward as opposite to depend on the volatile result that is obtained below UL. As mentioned above, notice that certainty above UL for the financial reward is a specific consideration of the QOF: practices know that if they attain certain performance level, they will know for sure how much income they will have. Thus, this Ω responds more to the specific programme characteristics than to a general consideration of risk aversion: in the QOF, practices have an extra incentive for trying to perform above UL and as a result it is important to understand how it would affect the test.

The next step is to obtain the marginal variation in optimal effort on task 1 with respect to the reward on task 2 following the same procedure as in the case without uncertainty.

$$\frac{de_1}{da_2}=-\frac{z\cdot C_{12}}{C_{11}C_{22}-C_{12}^2+p_1\cdot z\cdot C_{22}\cdot f(UL-e_1)\cdot\left\{2+\frac{f'(UL-e_1)}{f(UL-e_1)}\cdot\left[\frac{1}{p_1}\Omega+UL-e_1\right]\right\}} \tag{22}$$

This expression is the equivalent to the certainty-case equation 6. Here, there is an extra term in the denominator, which in general should still be positive as it is equivalent to the SOC. Thus, as before the sign is determined by C_{12}, but the magnitude is a function of current effort e_1. Hence, Proposition 1 is not affected by the presence of either risk or uncertainty.

Proposition 2 requires further analysis.

As with the no-uncertainty scenario, we can derive how general efficiency z is related to e_1^*. The expression below is the equivalent to the derivate present is assumption 1.

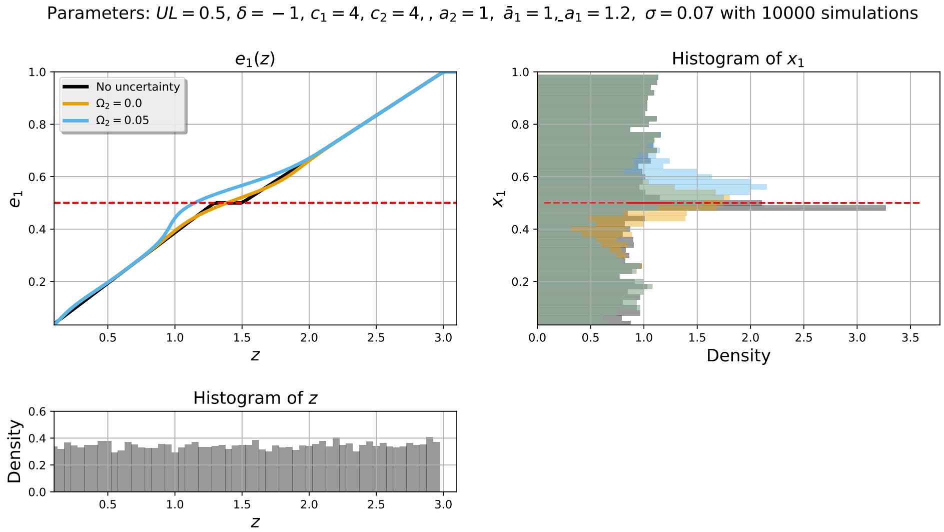
$$\frac{de_1}{dz}=\frac{\left\{\bar{a}_1+p_1\cdot\left[F\left(UL-e_1\right)+f(UL-e_1)\cdot\left(\frac{1}{p_1}\Omega+UL-e_1\right)\right]\right\}C_{22}-a_2C_{12}}{C_{11}C_{22}-C_{12}^2+p_1\cdot z\cdot C_{22}\cdot f(UL-e_1)\cdot\left\{2+\frac{f'(UL-e_1)}{f(UL-e_1)}\cdot\left[\frac{1}{p_1}\Omega+UL-e_1\right]\right\}} \tag{23}$$

Prior to discuss this equation in depth, the simulation exercise in Figure 5 will be useful to illustrate how Equations 22 and 23 compare with the expresions in the no-uncertainty case. This Figure follows the same configuration as the diagram presented in Figure 2. In this simulation, a cost function with constant second order derivatives is assumed. The noise on the task's result is assumed to follow a normal distribution. The provided parameters imply that both tasks are substitutes, and parameter z is drawn from a uniform distribution. The figure considers three cases: first, in black, the policy rules for e_1^* derived with no-uncertainty (black); second, with uncertainty but without risk aversion (orange), and finally including risk aversion (light blue).

Let us consider the case without risk aversion, $\Omega=0$. As shown in the graph, uncertainty essentially smooths out the corners of optimal effort $e_1^*(z)$. Moreover, the slope $\frac{\partial e_1}{\partial z}$ is always positive, as predicted by Equation 23. While introducing noise removes the idea of corner solution, it still generates bunching at UL as the slope becomes smaller rapidly near this threshold.

Let us consider first the denominator of Equation 23, and in particular, its last term: $\left(\frac{f'(UL-e_1)}{f(UL-e_1)}\cdot\left[UL-e_1\right]<0\right)$. When $e_1^*<UL$, it is implied that $(UL-e_1)<0$ which also means that $f'(UL-e_1)>0$.

Figure 5: Simulation exercise



Note: Parameters z drawn from a beta distribution with parameters $(5, 2)$ multiplied by 3. The cost function is defined as $C(e_1, e_2; z, c_1, c_2, \delta) = \frac{1}{z} \cdot (\frac{1}{2}(c_1 e_1^2 + c_2 e_2^2) + \delta e_1 e_2)$. For the cases with uncertainty, $x_1 = e_1 + v_1$ where $v_1 \sim N(0, \sigma)$

Hence, the entire term is negative $\left(\frac{f'(UL-e_1)}{f(UL-e_1)} \cdot [UL - e_1] < 0\right)$, so the denominator will become smaller as e_1 moves away from UL. When $e_1^* > UL$, exactly the same happens as when $f'(\cdot) < 0$ and $(UL - e_1) > 0$. Hence, the further e_1 is from UL, the larger the derivative, at least until it becomes equal to the no-uncertainty case when $f(UL - e_1) \rightarrow 0$.

Risk aversion plays an important role as observed in the example in Figure 5 ($\Omega = 0.05$). In the denominator, the term $\left(\frac{f'(UL-e_1)}{f(UL-e_1)} \cdot \left[\frac{1}{p_1}\Omega + UL - e_1\right]\right)$ changes the sign near UL three times. First, below UL, it makes the slope even larger, as it goes in the same direction as $UL-e_1$and the denominator becomes smaller. Second, in the interval $e_1^* \in [UL, \frac{1}{p_1}\Omega+UL]$, the term $f'(\cdot)$ becomes positive so the denominator is larger and then the derivative $\frac{de_1}{da_2}$ is smaller. Finally, when $e_1^* \geq \frac{1}{p_1}\Omega+UL$, the derivative starts to grow again. The implication for the distribution of x_1 is that the bunching will be centred above UL.

While the numerator of Equation 23 is also a function of $f(\cdot)$ and risk aversion, it plays a less important role in the graph of $e_1^*(z)$. The term $\left[F\left(UL - e_1\right) + f(UL - e_1) \cdot \left(\frac{1}{p_1}\Omega + UL - e_1\right)\right]$ decreases as e_1 departs from 0. This is because $F(UL - e_1)$ decreases with e_1, and so does $\left(\frac{1}{p_1}\Omega + UL - e_1\right)$. This effect is present both above and below UL.

Proposition 4. *In the presence of uncertainty on the task result, and if $f(\cdot)$ corresponds to a symmetric unimodal distribution with mean 0, $\frac{de_1}{da_2}$ becomes larger in absolute value as e_1 moves away from $\frac{1}{p_1}\Omega + UL$.*

This proposition replaces Proposition 2, as $\frac{de_1}{da_2}$ is not required to be 0 at UL anymore. The denominator in Equation 6 is the same as in Equation 23, so the same attenuation pattern when e_1 is just above UL can be expected. The main difference is that the sign is given by parameter C_{12} and that $f(\cdot)$ and risk aversion are present only in the denominator. Figure 6 presents two additional examples. The graphs on the left correspond to a cost function that exhibits substitution between tasks, while the ones on the right come from complementary tasks. The top graphs show optimal effort exerted on task 1 as a function of the price of task 2, for each of the cost functions and considering no-uncertainty (black), uncertainty (orange) and risk aversion (light blue). In the second row, the figure presents the first derivative of the graphs above, $\frac{de_1}{da_2}$. In both types of cost function, the derivatives are closest to zero when $e_1 = UL$ or is above it. For the case of substitutes, there are two additional cases in which the derivative is zero; those are corner solutions in which either $e_2^* = 0$ or $e_2^* = 1$.

B Model Examples

A simple cost function that captures both substitutability and complementarity is presented in Bolton and Dewatripont [2005]: $C(e_1, e_2; \theta = \{z, c_1, c_2, \delta\}) = \frac{1}{z} \cdot (\frac{1}{2}(c_1 e_1^2 + c_2 e_2^2) + \delta e_1 e_2)$ under the assumption that $\delta < \sqrt{c_1 c_2}$, $c_i > 0 \, \forall i$. As a result we can characterize the second derivatives with each parameter $C_{ii} = \frac{1}{z} \cdot c_i$ and $C_{ij} = \frac{1}{z} \cdot \delta$, $\forall i \neq j$.

B.1 No uncertainty

Given our function $\phi_i(x_i)$, for an optimal level of effort below UL_1 , the optimal levels of effort are given by

$$e_1^* = z \cdot \frac{a_1 c_2 - \delta a_2}{c_1 c_2 - \delta^2} \text{ , } e_2^* = z \cdot \frac{a_2 c_1 - \delta a_1}{c_1 c_2 - \delta^2}$$

Hence, Equation 6 becomes:

$$\frac{de_1}{da_2} = z \cdot \frac{-\delta}{c_1 c_2 - \delta^2}$$

Where it is clear that the sign of δ dominates the response to the incentives: if it is negative, then the tasks are complements as the marginal cost of one of the tasks is reduced when the effort of the other is increased (similar to the concept of economies of scope). However, notice that if we are above the threshold UL_1, two options should be considered

$$e_1^* = z \cdot \frac{\tilde{a}_1 c_2 - \delta a_2}{c_1 c_2 - \delta^2} \text{ , } e_2^* = z \cdot \frac{a_2 c_1}{c_1 c_2 - \delta^2} \qquad \text{and} \qquad e_1^* = UL_1 \text{ , } e_2^* = \frac{z \cdot a_2 - \delta UL_1}{c_2}$$

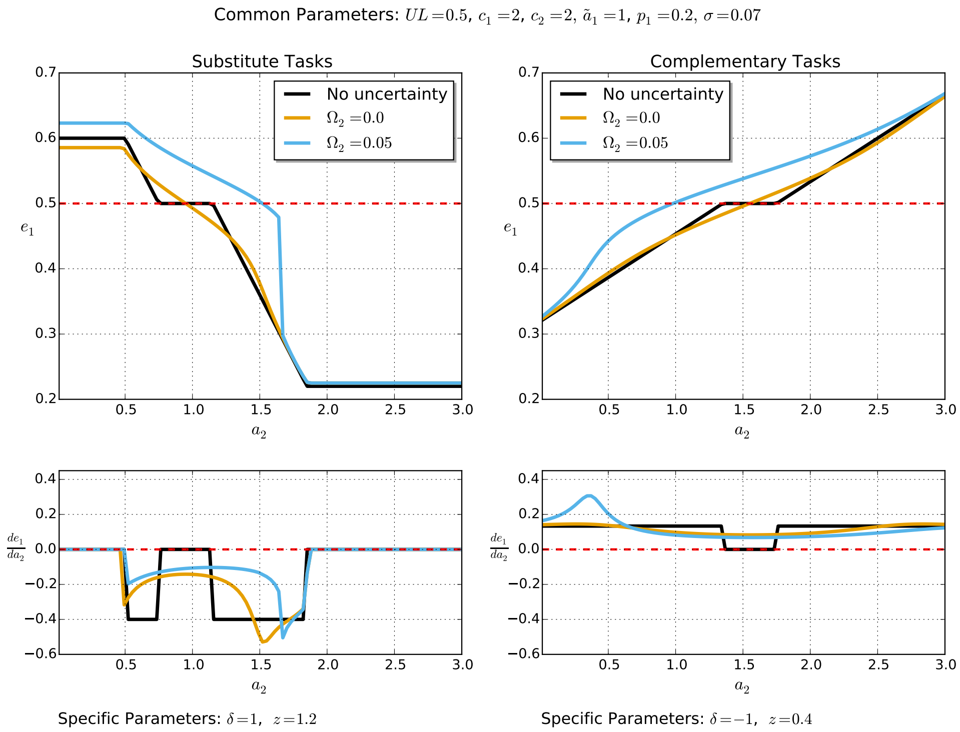
As a result:

1) If $\delta > 0$ (substitutes), at most, it is optimal to exert an effort level $e_1 = UL_1$, so it is expected that $\frac{de_1}{da_2}|_{e_1^* \geq UL_1} = 0$. Below that level, effort in task 1 it is decreasing with respect to the other task price: $\frac{de_1}{da_2}|_{e_1^* \geq UL_1} \leq 0$

2) If $\delta < 0$ (complements), below a cutoff $\bar{a}_2$ it is optimal to exert an effort level $e_1 = UL_1$, but above such a price cutoff, $\frac{de_1}{da_2} > 0$.

The result is a three section supply of effort 1. For substitutes it is flat, and then it decreases until it is optimal not to do any effort; and for complements it is increasing, flat and then increasing.

Figure 6: Simulation exercise: $x_1(a_2)$



Note: The cost function is defined as $C(e_1, e_2; z, c_1, c_2, \delta) = \frac{1}{z} \cdot (\frac{1}{2}(c_1 e_1^2 + c_2 e_2^2) + \delta e_1 e_2)$. For the cases with uncertainty, $x_1 = e_1 + v_1$ where $v_1 \sim N(0, \sigma)$

Kink With our current restrictions, it is straightforward to obtain the density of e_1^*. Here, $\bar{H}(\tilde{e}_1) = G\left[e_1^{*-1}\left(\tilde{e}_1; \bar{a}_1, a_2\right)\right] = G\left[\tilde{e}_1 \frac{c_1 c_2 - \delta^2}{\bar{a}_1 c_2 - \delta a_2}\right]$. Then, $\bar{h}(\tilde{e}_1) = g\left[\tilde{e}_1 \frac{c_1 c_2 - \delta^2}{\bar{a}_1 c_2 - \delta a_2}\right] \cdot \frac{c_1 c_2 - \delta^2}{\bar{a}_1 c_2 - \delta a_2}$ and similarly $\underline{h}(\tilde{e}_1) = g\left[\tilde{e}_1 \frac{c_1 c_2 - \delta^2}{(\bar{a}_1 + p_1)\cdot c_2 - \delta a_2}\right] \cdot \frac{c_1 c_2 - \delta^2}{(\bar{a}_1 + p_1)\cdot c_2 - \delta a_2}$.

Let us consider the point $\hat{e} = \tilde{e}_1 \frac{(\bar{a}_1 + p_1)\cdot c_2 - \delta a_2}{\bar{a}_1 c_2 - \delta a_2}$. If we consider the density without kink $\underline{h}(\hat{e}) = g\left[\tilde{e}_1 \frac{c_1 c_2 - \delta^2}{\bar{a}_1 c_2 - \delta a_2}\right] \cdot \frac{c_1 c_2 - \delta^2}{(\bar{a}_1 + p_1)\cdot c_2 - \delta a_2}$. We can re-express it as $g\left[\tilde{e}_1 \frac{c_1 c_2 - \delta^2}{\bar{a}_1 c_2 - \delta a_2}\right] = \underline{h}(\hat{e}) \cdot \frac{(\bar{a}_1 + p_1)\cdot c_2 - \delta a_2}{c_1 c_2 - \delta^2}$. Replacing this term in the density above UL, we can express the density of e_1^* in terms of $\underline{h}(\cdot)$, as shown below:

$$h(\tilde{e}_1) = \begin{cases} \underline{h}\left(\tilde{e}_1\right) & \text{if } \tilde{e}_1 < UL \\ b & \text{if } \tilde{e}_1 = UL \\ \underline{h}\left(\tilde{e}_1 \frac{(\bar{a}_1 + p_1)\cdot c_2 - \delta a_2}{\bar{a}_1 c_2 - \delta a_2}\right) \cdot \frac{(\bar{a}_1 + p_1)\cdot c_2 - \delta a_2}{\bar{a}_1 c_2 - \delta a_2} & \text{if } \tilde{e}_1 > UL \end{cases} \tag{24}$$

Notice that near UL, there is a discontinuity on the density even if we do not consider the bunching mass at UL. Below UL the density is $\underline{h}\left(\tilde{e}_1\right)$, but above it, the density is larger for a constant $\underline{h}\left(\tilde{e}_1\right)$. This is evident in the example of figure 5, where $\underline{h}\left(\cdot\right)$ is a constant as $g(\cdot)$ is uniformly distributed.

Comparative Statics What can generate the distribution over e_1? Let us consider only interior solutions ($a_1 c_2 - \delta a_2 > 0$ and $a_2 c_1 - \delta a_1 > 0$) and let Δ_T be the difference between the slopes of e_1^* with respect to a_2 above and below a given point T

$$\Delta_T = \frac{de_1}{da_2}|_{e_1^* < T} - \frac{de_1}{da_2}|_{e_1^* \geq T}$$

Heterogeneity on c_1 If the distribution on e_1 is due to efficiency on task 1, the sign of Δ is informative about the sign of δ. The resulting sorting on e_1 due to variation in e_1 is the same regardless of the nature of the cost function, while the size of the e_1^* slope with respect to a_2 depends on it.

	$\frac{\partial e_1}{\partial c_1}$	$\frac{\partial^2 e_1}{\partial a_2 \partial c_1}$	Below - Above (Δ_T)
	$-\left(a_1 c_2 - \delta a_2\right) \cdot z^{-1} \cdot \left(c_1 c_2 - \delta^2\right)^{-2} c_2$	$\delta \cdot z^{-1} \cdot \left(c_1 c_2 - \delta^2\right)^{-2} c_2$	$\frac{-\delta}{z \cdot \left(\bar{c}_1 c_2 - \delta^2\right)} - \frac{-\delta}{z \cdot \left(\underline{c}_1 c_2 - \delta^2\right)}$ First term is smaller in abs. val as its denominator is larger
$\delta < 0$ (Complements)	< 0	< 0	< 0
$\delta > 0$ (Substitutes)	< 0	> 0	> 0

Heterogeneity on c_2 If the distribution on e_1 is due to the efficiency on task 2, the sign of Δ is not informative about the sign of δ. In this case, both the sorting and the size of the e_1^* slope with respect to a_2 depend on the nature of the costs function.

	$\frac{\partial e_1}{\partial c_2}$	$\frac{\partial^2 e_1}{\partial a_2 \partial c_2}$	Below - Above (Δ_T)
	$\delta\left(a_2 c_1 - a_1 \delta\right) \cdot z^{-1} \cdot \left(c_1 c_2 - \delta^2\right)^{-2}$	$\delta \cdot z^{-1} \cdot \left(c_1 c_2 - \delta^2\right)^{-2} c_1$	Depends on δ
$\delta < 0$ (Complements)	< 0	< 0	< 0
$\delta > 0$ (Substitutes)	> 0	> 0	< 0 !!!!

Heterogeneity on δ If the distribution on e_1 is due to the degree of complementarity/sustituibility, the sign of Δ can only detect substitutes. Here, the size of the e_1^* slope with respect to a_2 depend on the nature of the costs function but the sorting depends on the value of other parameters. If tasks are substitutes and $e_1^* > \frac{a_2}{2\delta z^2}$, the sorting will be positive. In that case it is possible to say that the tasks are substitutes by observing a positive Δ, but if this term is positive it is not possible to deduce the sign of δ.

	$\frac{\partial e_1}{\partial \delta}$	$\frac{\partial^2 e_1}{\partial a_2 \partial \delta}$	Below - Above (Δ_T)
	$\left(-a_2\left(c_1 c_2 - \delta^2\right) + 2\delta\left(a_1 c_2 - \delta a_2\right)\right) \cdot z^{-1} \cdot \left(c_1 c_2 - \delta^2\right)^{-2}$ $\left(-a_2 + 2\delta z^2 \frac{a_1 c_2 - \delta a_2}{z\left(c_1 c_2 - \delta^2\right)}\right) \cdot z^{-1} \cdot \left(c_1 c_2 - \delta^2\right)^{-1}$ $\left(-a_2 + 2\delta z^2 e_1^*\right) \cdot z^{-1} \cdot \left(c_1 c_2 - \delta^2\right)^{-1}$	$-\left(\delta^2 + c_1 c_2\right) \cdot z^{-1} \cdot \left(c_1 c_2 - \delta^2\right)^{-2} c_1$	Depends on δ
$\delta < 0$ (Complements)	< 0	< 0	< 0
$\delta > 0$ (Substitutes)	If $-a_2\left(c_1 c_2 - \delta^2\right) + 2\delta\left(a_1 c_2 - \delta a_2\right) > 0$, then > 0	< 0	> 0
	If $-a_2\left(c_1 c_2 - \delta^2\right) + 2\delta\left(a_1 c_2 - \delta a_2\right) < 0$, then < 0	<0	< 0 !!!!

B.2 With uncertainty

Adding the functional form $C = \frac{1}{z} \cdot \left(\frac{1}{2}(c_1 e_1^2 + c_2 e_2^2) + \delta e_1 e_2\right)$. Also, assume $\varepsilon_1 \sim N(0, \sigma_1)$, which will allow us to work with the standard normal distribution. An additional element is the inclusion of the penalty Ω for uncertainty. For instance, this term will be equal to $\frac{1}{2}\eta(a_1^2 \sigma_1^2)$ if we consider an exponential utility function $u(p) = -exp(-\eta \cdot p)$, where $\eta$is the absolute risk aversion coefficient.

$$FOC_1 : = z \cdot \left\{ a_1 \Phi \left(\frac{UL - e_1}{\sigma_1} \right) + \frac{1}{\sigma_1} \phi \left(\frac{UL - e_1}{\sigma_1} \right) \cdot [(UL - e_1) \cdot a_1 + \Omega] \right\} - c_1 e_1 - \delta e_2 = 0$$

$$FOC_2 := z \cdot a_2 - c_2 e_2 - \delta e_1 = 0$$

And the equivalent of Equation 22

$$\frac{de_1}{da_2} = z \cdot \frac{-\delta}{c_1 c_2 - \delta^2 + z \cdot c_2 \left\{ \frac{1}{\sigma^2} \phi' \left(\frac{UL - e_1}{\sigma_1} \right) \cdot [(UL - e_1) \cdot a_1 + \Omega] + 2a_1 \frac{1}{\sigma} \phi \left(\frac{UL - e_1}{\sigma_1} \right) \right\}}$$

Given that for the standard normal pdf it holds that $\phi'(x) = -x\phi(x)$

$$\begin{aligned} \frac{de_1}{da_2} &= z \cdot \frac{-\delta}{c_1 c_2 - \delta^2 + z \cdot c_2 \left\{ -\frac{1}{\sigma^3} \phi \left(\frac{UL - e_1}{\sigma_1} \right) \cdot [UL - e_1] \cdot [(UL - e_1) \cdot a_1 + \Omega] + 2a_1 \frac{1}{\sigma} \phi \left(\frac{UL - e_1}{\sigma_1} \right) \right\}} \\ &= z \cdot \frac{-\delta}{c_1 c_2 - \delta^2 + a_1 \cdot z \cdot c_2 \cdot \frac{1}{\sigma} \cdot \phi \left(\frac{UL - e_1}{\sigma_1} \right) \left\{ 2 - \frac{1}{\sigma^2} \cdot [UL - e_1] \cdot \left[\frac{1}{a_1} \Omega + UL - e_1 \right] \right\}} \end{aligned}$$

As in the general case, being far from UL implies a larger slope (in absolute value). This is an effect that is attenuated by risk aversion below UL. Above such cut-off, risk aversion makes the derivative larger in absolute value. In this particular case, being very far from UL implies that the derivative will be equivalent to the non-uncertainty case.

C QOF Payment

Equation 26 shows how ratio indicators are translated into income for a practice i. Essentially, achievement x_i of indicator j is translated into points, and such points into yearly income. First, points are allocated according to a non-linear tariff that depends on two indicator specific thresholds. Below the lower limit (LL_j) zero points are awarded, and above the upper limit (UL_j) the maximum amount of available points for indicator j is awarded (Equation 25). The resulting figure is adjusted with respect to the relative size of the practice (*contractor population index*, CPI_i), and to the relative prevalence of the specific condition rewarded for clinical indicators (PF_{ij}). The achievement factor is multiplied by the CPI index and the prevalence factors, and by the price per point (Equation 26). The CPI captures the size of the practice, and is calculated as the number of patients in the practice relative to the figure 5891, which was the 2003 average list size.[32] The prevalence factor measures how commonly the condition is treated in indicator j, relative to the national average.

$$\begin{aligned} x_{ij} &= \frac{\text{Numerator}_{ij}}{\text{Denominator}_{ij}} \\ AF_{ij} &= \begin{cases} 0 & \text{if } x_i \leq LL_j \\ (x_i - LL_j) \cdot \frac{\text{Avail. Points}_j}{UL_j - LL_j} & \text{if } x_i > LL_j \\ \text{Avail. Points}_j & \text{if } x_i \geq UL_j \end{cases} \qquad (25) \\ CPI_i &= \frac{list_i}{5891} \\ PF_{ij} &= \frac{denom_{ij}/list_i|X_{ij}}{E[denom/list|X]} \text{ , where } X \text{ are specific conditions} \\ P_{ij} &= (\text{Value per point in č}) \cdot AF_{ij} \cdot CPI_i \cdot PF_{ij} \qquad (26) \end{aligned}$$

D Additional Tables

[32] Since 2013 this figure has been updated annually. More details are available from BMA [2013].

Table 7: QOF indicators descriptives for 2010/11

		(1)	(2)	(3)	(4)	(5)	(6)
Indicator	UL	Number	$E[x_t]$	P$[x_t < UL]$	$\rho(x_t)$	$E[x_t - x_{t-1}$	$E[x_t - x_{t-1}$
						$\|x_{t-1} < UL]$	$\|x_{t-1} > UL]$
AF03	90%	8245	93.82	7.14	0.50	9.88	-0.55
AF04	90%	8245	95.28	6.43	0.51	26.20	-1.39
ASTHMA03	80%	8245	90.00	4.69	0.41	18.64	-0.76
ASTHMA06	70%	8245	79.58	5.29	0.54	11.03	-0.19
ASTHMA08	80%	8245	87.89	6.37	0.46	15.63	-0.98
BP5	70%	8245	79.68	5.17	0.64	5.87	-0.09
CANCER03	90%	8245	92.75	17.84	0.34	18.18	-2.69
CHD08	70%	8245	81.90	3.51	0.56	11.30	-0.41
CHD09	90%	8245	93.58	7.56	0.46	4.30	-0.66
CHD10	60%	8245	74.91	2.60	0.67	13.30	-0.70
CHD12	90%	8245	92.73	16.53	0.48	5.17	-0.31
CKD02	90%	8245	97.26	1.29	0.41	26.83	-0.37
CKD03	70%	8245	74.86	21.73	0.52	5.58	-1.72
CKD05	80%	8245	90.78	6.03	0.46	40.20	-2.70
CKD06	80%	8245	82.35	24.29	0.53	14.80	-1.33
CVD01	70%	8245	80.12	14.71	0.44	26.06	-5.50
CVD02	70%	8245	82.61	7.94	0.37	34.13	-5.68
DEM02	60%	8245	80.54	3.04	0.42	36.15	-0.92
DM2	90%	8245	94.87	7.00	0.54	4.47	-0.36
DM10	90%	8245	91.39	22.84	0.58	4.93	-0.69
DM13	90%	8245	88.80	37.48	0.65	3.38	-1.38
DM15	80%	8245	89.28	8.07	0.53	20.31	-1.66
DM17	70%	8245	82.73	2.43	0.60	8.70	-0.55
DM18	85%	8245	91.19	9.76	0.47	5.99	-0.17
DM21	90%	8245	91.08	24.33	0.52	5.46	-0.97
DM22	90%	8245	96.95	2.44	0.44	7.78	-0.02
EPILEP06	90%	8245	95.62	6.95	0.27	13.07	-0.64
EPILEP08	70%	8245	73.96	26.14	0.56	7.72	-3.09
HF02	90%	8245	95.46	8.02	0.51	17.25	-1.03
HF03	80%	8245	90.26	4.24	0.46	27.42	-1.10
HF04	60%	8245	83.15	3.26	0.47	41.65	-1.39
SMOKE03	90%	8245	95.61	2.66	0.52	5.69	0.00
SMOKE04	90%	8245	93.07	12.48	0.44	4.98	-0.72
STROKE07	90%	8245	91.49	23.91	0.43	5.01	-1.08
STROKE08	60%	8245	77.18	3.07	0.50	20.72	-0.57
STROKE10	85%	8245	90.09	13.45	0.40	8.97	-0.49
STROKE12	90%	8245	93.79	8.98	0.45	9.97	-0.93
STROKE13	80%	8245	88.90	7.51	0.58	25.92	-1.87
THYROI02	90%	8245	95.81	3.24	0.41	10.46	-0.11

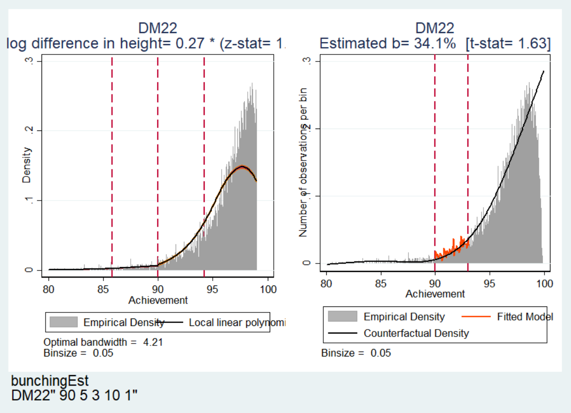
Notes: Own calculations based on QOF data. **Number:** Number of GP practices, including those with 0 elegible patients for the given indicator. $E[x_t]$ **:** Average achievement per indicator. **P**$[x_t < UL]$ **:** Proportion of practices with an achivement below UL. $\rho(x_t)$ **:** Correlation between 2010 and 2009 achivement.

Table 8: Detailed Changes in QOF 2011 clinical indicators with respect to 2009-2010

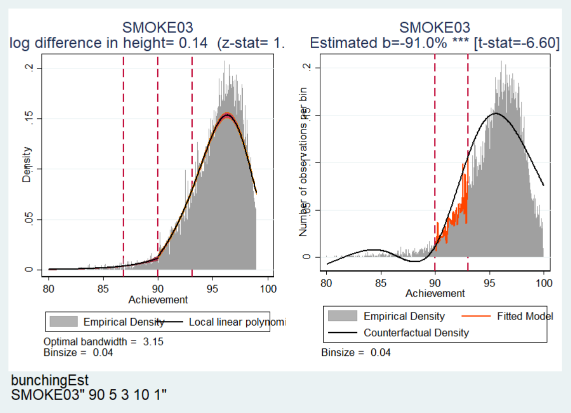
Status	**Description**	**Affected Indicators**	**Price Interpretation**	**Points**
Retirements	These tasks are not rewarded anymore. Clinical indicators are about having a recent record of certain physical measures, or reviews.	CHD5, CHD7, DM5, DM11, DM16, EPILEPSY7, MH7, STROKE5	Reduction	32
Points reduced	Number of assigned points per indicator was reduced.†	BP4, DEP1	Reduction	26 to 22
Upper Limit Increased	Small increase from 70% to 71%. ♠	CHD6, STROKE6	Reduction	22
Replacement I	For indicators PP01, MH04, MH05, the time for accomplishing a given goal was reduced. For CHD2, the optional specialist referral was made compulsory.	PP01, MH04, MH05, CHD2	Reduction	18
Replacement II	Decrease in points and new wording is more precise and requires actions at the moment of diagnosis instead of treatment starting point.	DEP2, DEP3	Reduction	45 to 25
Replacement III	Most of these indicators were replaced by versions which are harder to accomplish. In a few of them this was compensated with extra points, but in some others there was a reduction as well: • For CHD11/CHD14 there is an increase from 7 to 10 points in exchange for prescribing aspirin and statins on top of an ACE inhibitor or alternative blood pressure treatments. • Requirements for DM9 were increased from checking peripheral pulses to a more comprehensive foot examination. It was also increased from 3 to 4 points. • Indicator DM12 was split into DM30 and DM31, keeping the same number of points. It asked for a percentage of patients below a given blood pressure target (145/85). It was replaced by two targets, one slightly below the original (140/80), and one notoriously above (150/90). • Indicator MH09 was split into MH11, MH12, MH13, MH14, MH15 and MH16. It moved from 23 to 27 points. The original indicator was general and imprecise ("routine health promotion and prevention advice appropriate to their age and health status"), while the replacements ask for specific measurements depending on age and gender.	CHD11/CHD14, DM9, DM12 (DM30,DM31), MH09 (MH11, MH12, MH13, MH14, MH15 and MH16)	Ambiguous	51 to 59
Replacement IV	The cutoff was relaxed from last HbA1C to be 7% or less, to HbA1C to be 7.5% or less	DM23/DM26	Increase	17
Replacement V	Similar or the same wording, but the recoding was done in order to highlight recent changes in diagnostic procedures. For diabetes indicators the wording is explicit about new measurement standards.	COPD1/COPD14, COPD12/COPD15, MH6/MH10, DM24/DM27, DM25/DM28	-	32
New	These are tasks that were not considered before. Three new clinical indicators, on dementia, epilepsy and learning disabilities.	DEM3, EPILEPSY 9, LD2	Increase	12
Unchanged	No change on points, thresholds or wording		-	454

Note: This corresponds to our interpretation based on NHS Employers public documents. † Does not include indicators which wording was amended as DEP2 and DEP3. ♠ Does not include DM12/DM30, which is an indicator that its wording was also amended.

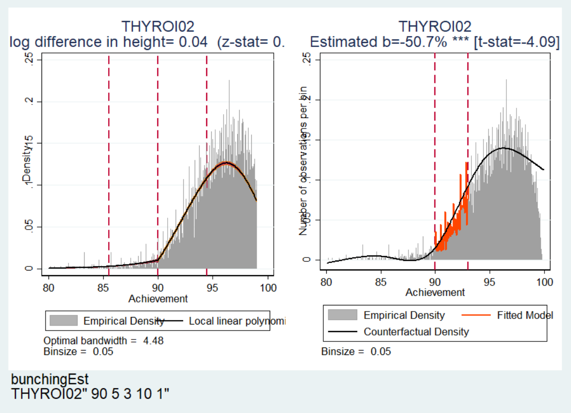
Table 9: Bunching tests for selected indicators I



DM22: The percentage of patients with diabetes who have a record of estimated glomerular filtration rate (eGFR) or serum creatinine testing in the previous 15 months.
3 points. LL=40, UL=90.

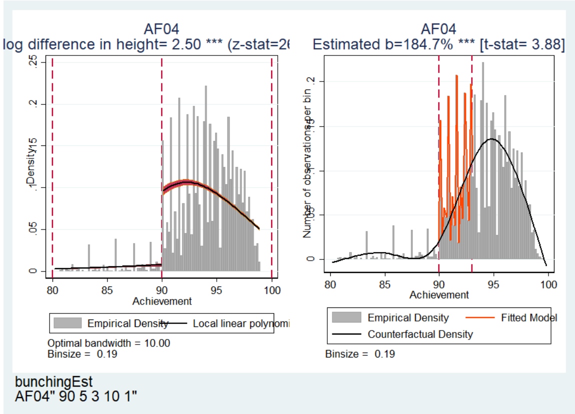


SMOKE3: The percentage of patients with any or any combination of the following conditions: coronary heart disease, stroke or TIA, hypertension, diabetes, COPD, CKD, asthma, schizophrenia, bipolar affective disorder or other psychoses whose notes record smoking status in the previous 15 months (except those who have never smoked where smoking status need only be recorded once since diagnosis)
30 points. LL=40, UL=90.

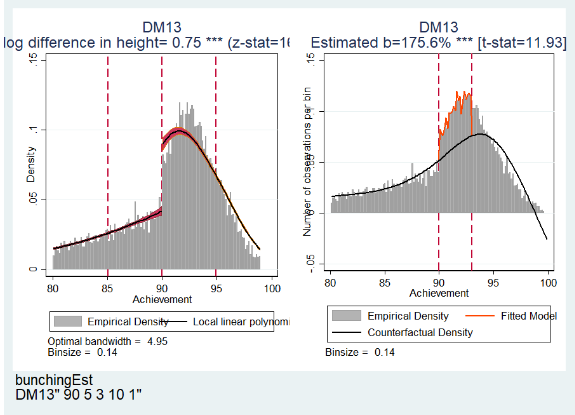


THYROID2: The percentage of patients with hypothyroidism with thyroid function tests recorded in the previous 15 months
6 points. LL=40, UL=90.

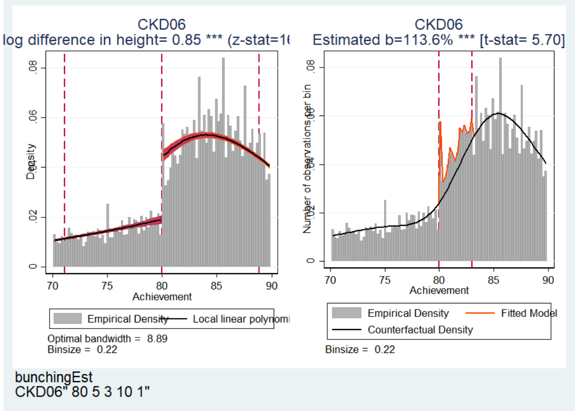
Table 10: Bunching tests for selected indicators II



AF04: The percentage of patients with atrial fibrillation diagnosed after 1st April 2008 with ECG or specialist confirmed diagnosis.
10 points. LL=40, UL=90.



DM13: The percentage of patients with diabetes who have a record of micro-albuminuria testing in the previous 15 months (exception reporting for patients with proteinuria)
3 points. LL=40, UL=90.



CKD06: The percentage of patients on the CKD register whose notes have a record of an albumin:creatinine ratio (or protein:creatinine ratio) value in the previous 15 months
6 points. LL=40, UL=80.