The London School of Economics and Political Science

Essays in Labour Economics

Nitika Bagaria

A thesis submitted to the Department of Economics of the London School of Economics for the degree of Doctor of Philosophy, London, July 2014

To my family

Declaration

I certify that the thesis I have presented for examination for the MPhil/PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

I declare that my thesis consists of **57,382** words.

Statement of conjoint work

I confirm that Chapter 2 was jointly co-authored with Professor John Van Reenen and Professor Barbara Petrongolo. I contributed 35% of this work.

I confirm that Chapter 4 was jointly co-authored with Professor John Van Reenen and Dawn Holland. I contributed 35% of this work.

I confirm that Chapter 5 was jointly co-authored with Dr. Novella Bottini and Dr. Miguel Coelho. I contributed 35% of this work.

Acknowledgements

I am indebted to John Van Reenen, Steve Pischke and Robin Burgess for their patient advice. I also thank Barbara Petrongolo, Alan Manning, Jonathan Portes, Simon Kirby, Ghazala Azmat, Jonathan Wadsworth, Esteban Aucejo, Guy Michaels, Anna Valero, Andy Feng, Marta de-Phillippis, Katalin Szemeredi, Isabelle Roland, Claudia Steinwender, Joao Paolo Pessoa, Pedro Alves and seminar participants at the London School of Economics and IFS for helpful comments. I would also like to thank Bill Wells of DWP for his insights. I would also like to thank Angus Armstrong, Richard Layard, Katerina Lisenkova and Jonathan Portes for useful discussion and comments on the modelling work for Chapter 4. I would like to thank the ESRC for funding my doctoral research through the Centre for Economic Performance.

Abstract

My thesis investigates the role of incentives for employees and the phenomenon of labour market hysteresis in driving productivity and employment in the labor market in the UK. Chapter 1 summarizes the thesis.

The second chapter estimates the impact of introducing an explicit points-based system in favour of finding jobs for the disabled in a UK job placement agency. Using dynamic analysis, we find that in the long-run the policy improved disabled outflows by 6% and had an insignificant effect on JSA outflows. In the short-run, the policy had a negative impact on JSA outflows that declined by 2%. This is consistent with a model where information helps both groups, but incentives offset this for the able and reinforce it for the disabled. The third chapter studies how incentives are weakened in a public sector organization when rewards are based on team output rather than individual output. With the introduction of team rewards, employees are likely to free-ride on each other's efforts. I find compelling evidence that this indeed occurs. Peer monitoring, may however, limit free-riding in teams. I formalize the impact under two benchmark models to ascertain the relative impact of peer monitoring and free-riding. Using difference-in-differences estimators, I find that consistent with a degree of peer monitoring, the dilution effect is smaller when peer monitoring is easier. The fourth chapter models the phenomenon of labor market hysteresis in a macroeconomic model to determine its impact on macroeconomic outcomes. In particular, we study its role in determining the impact of the scale and timing of UK's fiscal consolidation programme on output and unemployment in the UK. Finally, the last chapter studies employee incentives in the context of education. Motivated by a diagnosis of increasing inequality in UK's educational attainment in secondary education, we recommend a flexible school system and improved school and teacher governance.

Contents

Chapter 1: Introduction

In the introduction, I summarize the chapters. In this thesis, the second chapter estimates the impact of introducing an explicit points-based system in favour of finding jobs for the disabled in a UK job placement agency. Disability rolls have escalated in developed nations over the last 40 years. But they stopped rising in the UK when a welfare reform was introduced that physically integrated the disability benefit (IB) with unemployment insurance. This policy reform improved job information and relatively sharpened bureaucratic incentives to find jobs for the disabled (relative to those on UI) by introducing a points-based system. The policy roll-out was staggered both geographically and across time in a quasi-random way that we exploit to evaluate the change falls in welfare rolls. In the long-run the policy improved disabled outflows by 6% and had an (insignificant) 1% increase in UI outflows. In the short-run, the policy had a negative impact, particularly on the UI outflows that declined by 2%. This is consistent with a model where information helps both groups, but incentives offset this for the able and reinforce it for the disabled. A cost-benefit calculation is very supportive of the policy, but the costs of the organizational disruption implies that benefits take 5 years to exceed costs making it unattractive for (myopic) politicians. This illustrates the difficult political economy of welfare reform.

The third chapter explores the impact of introducing team based rewards in a public sector organization. In my context, the new rewards switch from being individually based to being based in a team at the district level. Districts are organized into spatially dispersed sub-teams (i.e. offices within districts). In this setting, I explore whether it is really the case that incentives are weakened when rewards are based on team output rather than individual output. I find compelling evidence that this indeed occurs. With the introduction of team rewards, employees are more likely to free-ride on each other's efforts in large teams. I find support for team size determining the degree to which incentives are weakened. Peer monitoring, may, however, limit free-riding in teams. I formalize the impact on output under two benchmark models to ascertain the relative impact of peer monitoring and free-riding. To investigate the model's predictions empirically, I exploit the national rollout of a incentive structure in the UK public employment service (Jobcentre Plus) in the mid-2000s. Using difference-in-differences estimators, I find that the team based measures dilute incentives. Consistent with a degree of peer monitoring, however, the dilution effect is smaller when peer monitoring is easier (as captured by the concentration of teams across offices within a district).

The fourth chapter models the phenomenon of labor market hysteresis in a macroeconomic model. In 2009-10, the UK's budget deficit was about 11 per cent of GDP. A credible plan for fiscal consolidation was introduced in the UK over the fiscal years 2011-12 to 2016-17. In this chapter, we assess the impact of the scale and timing of this fiscal consolidation programme on output and unemployment in the UK. During a prolonged period of depression when unemployment is well above most estimates of the NAIRU, the impact of fiscal tightening may be different from that in normal times. We contrast three scenarios: the consolidation plan implemented during a depression; the same plan, but with implementation delayed for three years when the economy has recovered; and no consolidation at all. The modelling confirms that doing nothing was not an option and would have led to unsustainable debt ratios. Under both our "immediate consolidation" scenario and the "delayed consolidation", the necessary increases in taxes and reductions in spending reduce growth and increase unemployment, as expected. But our estimates indicate that the impact would have been substantially less, and less long-lasting, if consolidation had been delayed until more normal times. The impact is partly driven by the heightened magnitude of fiscal multipliers, and exacerbated by the prolongation of their impact due to hysteresis effects. The cumulative loss of output over the period 2011-21 amounts to about £239 billion in 2010 prices, or about 16 per cent of 2010 GDP. And unemployment is considerably higher for longer - still 1 percentage point higher even in 2019.

Finally, the last chapter studies employee incentives in the context of education. Economic theory and evidence shows that in the long-run human capital is the critical input for growth. While indicators of average educational outcomes at the secondary level, in the UK, tend to show significant improvements over time, they mask the fact that the UK has a long tail of poor (secondary) education performance compared to other countries. This holds back growth and social mobility. The incentives for schools to focus on the performance of children from disadvantaged backgrounds are weak. Dissemination of high quality teaching through the school system depends fundamentally on school incentives - performance measures, autonomy and competition. We propose a flexible system for education, which gives schools greater autonomy and the ability to grow within a national accountability framework that places a premium on radically raising the standards of the bottom ability group. Together with improved choice for parents, better quality information (across the entire distribution of achievement) and more effective incentives for teachers and schools, this will improve the quality of teaching.

Chapter 2: Can helping the sick hurt the able? - Incentives, Information and Disruption in a Welfare to Work reform

Abstract: Disability rolls have escalated in developed nations over the last 40 years. But they stopped rising in the UK when a welfare reform was introduced that physically integrated the disability benefit (IB) with unemployment insurance. This policy reform improved job information and relatively sharpened bureaucratic incentives to find jobs for the disabled (relative to those on UI) by introducing a points-based system. The policy roll-out was staggered both geographically and across time in a quasi-random way that we exploit to evaluate the change falls in welfare rolls. In the long-run the policy improved disabled outflows by 6% and had an (insignificant) 1% increase in UI outflows. In the short-run, the policy had a negative impact, particularly on the UI outflows that declined by 2%. This is consistent with a model where information helps both groups, but incentives offset this for the able and reinforce it for the disabled. A cost-benefit calculation is very supportive of the policy, but the costs of the organizational disruption implies that benefits take 5 years to exceed costs making it unattractive for (myopic) politicians. This illustrates the difficult political economy of welfare reform.

Acknowledgements: We would like to thank the ESRC for funding this research through the Centre for Economic Performance. IFS generously provided some of the estimates of benefit receipt.

1 Introduction

Disability rolls have risen almost inexorably in the US and other advanced nations in the last forty years (Autor and Duggan, 2003). In the US, Social Security Disability Insurance (SSDI) cash transfers have tripled from $40bn in 1979 to $124bn in 2010. Its share of total social security payments rose from 10% in 1988 to 20% in 2009 (Autor, 2012). Figure 1 shows that the numbers on the equivalent UK scheme, Incapacity Benefit (IB) also rose spectacularly: from 400,000 in 1977 to 2m in 2009. Unlike the US, however, the increase seemed to stop from the early 2000s. In 2001, a major policy reform was introduced which was designed to address the large numbers on IB. The agencies responsible for welfare benefits for working age people were merged into one, "Job Centre Plus". The aim was to give a much greater focus on getting those on IB and other benefits back into jobs. The main other group, by numbers, were those on unemployment insurance, called Job Seekers' Allowance (JSA).

The reform physically integrated the offices where welfare claimants go to have work-focused interviews, collect checks and look for jobs. It simultaneously increased *information* and changed bureaucratic *incentives* (the formal system for receipt of welfare was unchanged). Information was improved for all groups by a major new IT system and new buildings. The people in charge of helping welfare claimants find jobs had changed incentives with an explicit points system introduced with three times as many points given for placing a disabled person into a job than for the unemployed. Points fed into career progression.

In our empirical analysis we exploit the policy roll-out that was staggered over time and across geographical areas (we use quarterly administrative data based on districts, similar to US counties[1], over a 9 year period). We identify the policy impact by comparing the change in exit rates for disability and unemployment benefit claimants in districts treated at a point in time to that in districts treated at an earlier or later date. Information on benefit claimants at the district level is provided by the Department for Work and Pensions, and we use quarterly series for stocks, inflows and outflows for

[1] The average population of a district was 120,000 in the 2001 census. In comparison, the average population of a US county is 100,000 people.

various categories of welfare benefits, disaggregated by age and district, and, for the case of JSA claims, disaggregated by destination (e.g. to job vs. non-participation).

We find two main results. First, there were significant organizational disruption costs from the policy, with outflows from disability and unemployment benefits initially declining after the introduction of the policy, and more markedly so for JSA than IB recipients. Second, in the long-run there were significant positive effects of disabled outflows, whereas unemployment outflows, although positive, are small and insignificant. These patterns are consistent with a simple model whereby bureaucratic efforts to reduce disabled increased, but efforts to reduce unemployment rolls decreased. However, lain over this there was a long-term positive effect on both groups from better information but a short-run negative effect from adjustment costs due to organizational disruption.

Using our results we present a cost-benefit analysis showing that the policy has a large welfare gain. However, the presence of significant short-run costs from disruption and sunk set-up costs highlights why such welfare changes are hard to implement. We estimate it takes five years for the reform to break even, which is beyond the time horizon of even benign policy-makers.

This chapter links to several literatures. First, the issue of welfare reform has come to the fore with the Great Recession of 2008-2009. The rise in unemployment has been much less than expected in the UK[2] and Germany, and both countries experienced significant welfare reforms prior to the crisis. Second, a growing literature has highlighted how incentive systems can be used to improve efficiency, particularly in the public sector (see Besley and Ghatak, 2005 on theory; Meyer, 1995, for a survey of findings from social experiments or Bloom and Van Reenen, 2011, for a general survey). For example, Heckman et al. (2011) found bureaucrats rewards affected cream skimming and showed significant differences between short-run and long-run programme impacts. We contribute to this literature by emphasizing the multitasking aspects in the provision of effort in government organizations. A body of work in this literature has studied the effects of financial incentives to benefit recipients on the

[2] On the UK case see Blundell, Crawford and Jin (2013); Gregg, Machin and Salgade (2013) and Pessoa and Van Reenen (2013).

duration of unemployment (see Lalive et al. 2007 for recent evidence), while the role of explicit incentives in the provision of job placement services is to date largely unexplored.

The chapter is organized as follows. In section 2 we describe the institutional framework in the UK and the hypotheses we test, in section 3 we outline the data used in the empirical analysis, while in section 4 we report the analysis and results of how the treatment impacts inflow into and outflow from different benefit categories. In section 5 we examine the robustness of our results to different specifications and in section 6 we perform a simple cost-benefit evaluation of Jobcentre Plus. Section 7 concludes.

2 The Institutional Framework

2.1 The Jobcentre Plus system

There were major infrastructure changes in the delivery of public employment and benefits services in the UK between 2001 and 2008. The change was part of a wider policy emphasis on Welfare-to-Work initiatives[3] that sought to increase labour market activity. In March 2000, the Prime Minister announced the establishment of Jobcentre Plus that would deal with people of working age, to deliver a single, work-focused, integrated service to both employers and benefit claimants of working age in UK. Since October 2001, the Employment Services (ES) and Benefits Agency (BA) were integrated into one organization. The new organization which combined benefit advice with job placement services was called Jobcentre Plus. The integration took place in six waves between 2001 and 2008.

There were two broad changes as a result of the Jobcentre Plus policy relating broadly to information and incentives. On information, the physical organization of offices as the two types of services (help with finding jobs and giving out welfare checks) were now delivered under one roof. There was a massive investment in improved information technology (IT) systems and organizational restructuring. The average size of an office was increased as buildings were combined, re-built and offices refurbished. Aggregate floor space decreased by 20% as did the total number of staff

[3] The guiding principle of welfare reform was "work for those who can and security for those who cannot" (Hyde et al., 2002).

even though operating costs per square meter by 12% because of high quality infrastructure and locations. Overall annual administrative costs per year were reduced by £240m although (separate) the sunk costs of re-organization for around £1.8bn (National Audit Office, 2008).

The second major change was the introduction of explicit performance targets called Job Entry Targets (JET). As opposed to the previous system of national-level targets for the number of beneficiaries to place into jobs, under the new regime every benefit officer who helped a benefit claimant into a job was awarded a certain number of *explicit* Job Entry Target points varying by the category of the benefit claimant. In addition, there was a district-level target in terms of the number of points to achieve each quarter. These performance standards acted like a performance benchmark for the managers and mattered for the career prospects of the benefit officers.[4]

2.2 Framework

Theoretically, we distinguish between three different effects on benefit officers' job placement activity as a result of the introduction of the Jobcentre Plus. Firstly, the physical reorganization, installation of new IT systems and estate rationalization would have an immediate disruptive effect, as the provision of services would have been disrupted and unavailable for some periods of time during the initial phase of the introduction of Jobcentre Plus. This would lead to a reduction in the productivity of the officers in the short-run. We expect the disruption effect to be broadly similar across all benefit groups. This effect should decay over time as officers settle into the new system.

We distinguish this from the second hypothesis about the long-run effect of system restructuring and modern IT systems on efficiency. Benefit officers are now able to use the new IT systems which facilitate various manual tasks such as recording job entries and keeping records of beneficiaries. Increased automation of services would improve the speed and accuracy with which benefits applications were processed. This reduces

[4] The UK welfare system had introduced performance benchmarking since the early 1980s (Propper and Wilson 2003; Bagaria et al, 2013). They have been designed according to targets embodied in the Public Service Arrangements (PSAs) of different government agencies. Makinson (2000) describes the performance standards in the Employment Service, The Benefits Agency, HM Customs and Excise and Inland Revenue. These mostly consisted of national-level targets for the number of beneficiaries to place into jobs, without explicit rewards at the individual or local level.

operating costs as well as the time officers spend on these back office functions and enables them to focus on conducting more client-facing job finding interviews. Thus, in the long run, we expect an increase in job placements for all benefit groups as the provision of welfare services becomes more efficient. Again, this effect should be broadly similar across all benefit recipients.

Thirdly, the effect of the JET implies a shift in bureaucratic incentives in favour of Incapacity Benefit claimants. Before the introduction of Jobcentre Plus, there were broad national level targets for job placements and sub-targets for different benefit categories. For example, in 2001 there was a national target to place 1.36m jobless people into work accompanied by a sub-target to place 275,000 disadvantaged[5] into work. The explicit award points under the new JET system were designed to reflect organizational priorities towards the Incapacity Benefit claimants. As shown in Table A1, a benefit officer was awarded fifty per cent more points if he/she placed a person on Incapacity Benefit[6] into a job than a long-term JSA beneficiary, and three times more points relative to a short-term JSA beneficiary. Given that the benefit officers had to achieve a quarterly target number of points, this should incentivize them to focus on placing the IB claimants.

Consider a multi-tasking model with fixed inputs along the lines of Holmstrom and Milgrom (1991). Assume that the Jobcentre officers have a given stock of “inputs”. They can apply different amounts of this “input” to different individual clients (benefit claimants). These inputs affect the outcomes experienced by claimants. In our context, the input variable represents staff time for interviews and the direct costs of the services provided. After the introduction of the explicit weighting system, we expect them to re-organize the manner in which they allocate their efforts[7]. They would now focus more on and increase effort in placing Incapacity Benefit claimants. Given a fixed stock of inputs, this should adversely affect the job placement efforts and outcomes for those on Job Seeker’s Allowance claimants. Thus, Jobcentre Plus is likely to have a larger impact on IB recipients than JSA recipients.

---

[5] This included those with disabilities, participants in New Deal for Lone Parents, partners of continuously unemployed for 26 weeks, homeless people and qualifying ex-offenders.

[6] We discuss the other group, Lone Parents in Section 6.

[7] Unfortunately, we are not able to measure staff inputs, but we can observe participant outcomes. So, in a sense this is a reduced form estimation.

2.3 Identification

We exploit the staggered roll-out of the Jobcentre Plus offices across Local Authority Districts in the UK to identify the causal impact of the policy. The switch to Jobcentre Plus was phased in over six waves, as illustrated in Figure 2. The figure shows the additional districts covered under each wave. The first wave begun on 1st October 2001 in 32 districts, the second wave began in October 2002 with 27 more districts and by the first quarter of 2008, almost 100% of the country was covered. Figure 2 presents a map of the policy roll out, showing no obvious patterns of geographic clusters that adopted the policy at the same time.

We consider treatment as the "go live" for a district. One concern is that districts selected into treatment were not randomly assigned into treatment. To examine this, we check for differences in pre-trends between the treated and non-treated for various benefit categories (see section 4).

There are multiple offices in a district (between 32 and 171) and we also considered using the penetration across offices within a district. However, we found that although observables could not predict which districts were treated, there did appear to be a systematic component of which offices within a district were treated. It is likely that a district treated the offices with a higher benefit outflow rate to begin first (Table A2). [8] Further, the points system was formally at the district level.

We use a difference-in-differences framework to identify the causal impact of Jobcentre Plus. Since all districts are treated eventually, effectively we are comparing districts which are treated in a particular year and quarter to those who are treated at a later stage. Our main outcomes are the number of exits from disability and unemployment benefits in each quarter in each district. We further disaggregate outflows by age groups - young and old. The young are defined as being aged between 18 to 24 years and the older group consists of people aged 25-59 years.[9]

[8] National Audit Office (2008) states that "Whilst an overall vision of the service improvements was successfully communicated from the centre, the detailed planning of the roll-out was delegated to the districts…. Implementation of Jobcentre Plus was a locally driven process" and that "Localised planning allowed Jobcentre Plus to make early progress with the roll-out, as the districts which were ready first could be scheduled for early roll-out".

[9] We also examine other age splits as robustness checks (Table A6).

One potential concern is that jobseekers may be manipulating the benefit category that they apply to, thus affecting the composition of the claimant stock in each clients' group and the corresponding outflow rate from benefits. For instance, benefit applicants may choose to enter the caseload as a disabled person rather than under Job Seeker's Allowance. In the UK context, it is rather difficult to be classified as disabled to claim Incapacity Benefit since it requires a medical certificate and the conditions required to qualify for receiving Incapacity Benefit have been made stricter over time.[10] A related concern is that the introduction of Jobcentre Plus may affect jobseekers' decisions whether to sign-on at all for benefits. To examine this further, we also analyse the impact of the Jobcentre Plus on the *inflows* into the different categories in section 4.

3 Data

We use administrative panel provided by the UK Department of Work and Pensions that cover the welfare population. The Job Seeker's Allowance database contains monthly information from 1983 on the stocks, inflows and outflows of recipients' unemployment benefits. The data is available at the Local Authority District across Great Britain and there are 406 districts defined on a consistent basis.[11] The data can be disaggregated in various ways, and we focus on cuts by age.

The second dataset contains *quarterly* data from 1999Q3 at the district level on other welfare benefits including the key disability benefit, Incapacity Benefit (IB). To be consistent across the two datasets, we aggregate the monthly JSA data to the quarterly level, but use the monthly information as a robustness check. We estimate all our

[10] For instance, in 1999, the Welfare Reform and Pensions Bill introduced '*continuing assessment of possibility of returning to work*' (Burchardt, 1999). While the criteria for benefit receipt remained unaltered, the significant change was the collection of additional information focussing on the abilities of the claimant at intervals and the allocation of a personal adviser to oversee each claim. New claimants were also required to attend an interview at the beginning of the claim, and any time thereafter, to discuss possibilities for returning to work. More recently, applicants to IB will have to go to a Work Capability Assessment during the first 13 weeks of IB. This was aimed to see if the illness or disability affected the claimant's ability to work.

[11] Local government in England operates under either a single-tier system of unitary authorities and London boroughs, or a two-tier system of counties and district councils. The spatial units in our analysis include the unitary authorities, London boroughs and districts within counties. There are 352 such units in England. Local government in Scotland is organized through 32 unitary authorities. Since 1 April 1996, local government in Wales is organized through 22 single-tier principal areas. The Scottish and Welsh unit areas are also included in our sample.

specifications on a consistent time period of nine years (36 quarters) 1999Q3 to 2008Q2, the quarter before the collapse of Lehman's, which triggered the Great Recession and a huge upsurge of unemployment.

Descriptive statistics are presented in Table 1. Columns (1) and (2) refer to national aggregates per quarter, and columns (3) and (4) refer to (unweighted) averages across districts, age groups and quarters. The aggregate outflow rate from JSA is 70% consisting of an average outflow of 653,819 per quarter and a stock of 939,267 per quarter in the country as a whole. Inflows were a bit lower than this (648,957 per quarter) as unemployment was falling over the sample period. Outflow rates for IB were much lower at 3% per quarter: as is well known far fewer people leave the stock of disability rolls than unemployment.

3.1 Jobcentre Plus and Benefit Flows

We estimate benefit outflow equations in a difference-in-differences framework. We start by estimating a static model to estimate the *average* effect over time arising from the change in bureaucratic incentives. The specification is:

$$\ln Y_{ait}^{B} = \beta^{B} D_{it} + \gamma_{1}^{B} \ln U_{ait-1}^{B} + \gamma_{2}^{B} \ln U_{a'it-1}^{B} + \delta_{ai}^{B} + \delta_{at}^{B} + \varepsilon_{ait}^{B} \qquad (1)$$

where Y_{ait}^{B} is the number of people in age group *a* leaving the benefit register *B* (JSA or IB) in district *i* at time *t* (quarter). D_{it} denotes a treatment dummy which turns on in the quarter when the first office in district *i* is treated. The coefficient is identified by the fact that the policy was rolled out in six waves with different districts being treated in each wave. One robustness test we consider is to allow β^{B} being different in each wave and showing that the effect looks remarkably similar across waves when the post-wave window is kept fixed. As noted above, we found that the timing of when a district was treated appeared to be unrelated to observables.

We include as controls the stock of claimants of benefit *B* at the end of the previous quarter for the own age group, U_{ait-1}^{B} as well as other age groups, *a'* (old/young respectively in the baseline specification). Our preferred specifications include a full set of fixed effects (district by age group) and age by time dummies, but we also show more restrictive specifications just including separate district, age and time effects. We cluster the standard errors at the district level, which is the level at which the policy is defined.

In equation (1) the treatment effect is summarized by the coefficient β^B which is an average over all the post-treatment quarters. Our adjustment cost theory, however, suggests that there should be a distinct dynamic pattern of change with the positive policy effects being dampened at first by organizational disruption. Hence, we allow the policy effect to be different depending on how many quarters have elapsed since the policy.

$$\ln Y_{ait}^B = \sum_{\tau=1}^{7} \beta_\tau^B D_{it+\tau} + \beta_{LR}^B D_{LR} + \gamma_1^B \ln U_{ait-1}^B + \gamma_2^B \ln U_{a'it-1}^B + \delta_{ai}^B + \delta_{at}^B + \varepsilon_{ait}^B \quad (2)$$

The $D_{it+\tau}$ term is broken down by period after the policy begins so D_{i1} is the quarter when the policy is turned on, D_{i2} is the first quarter after the policy is in effect and so on. D_{LR} is the "long-run", defined as 8 quarters or more after the policy is in place. Since the last district to have the policy was Wave 6 in 2006Q3 we have at least two years of post-policy experience for all districts. Ending the dynamics after two years is somewhat arbitrary, but the coefficients seem stable afterwards and we show that the qualitative results are robust to alternative dynamic specifications (e.g. Table A4)

The disruption hypothesis suggests that the $\beta_3^B > \beta_2^B > \beta_1^B$ and so on. In other words, the initial negative disruption effects unwind as the new organizational structure settles down. The incentives hypothesis suggests that the positive effects on IB should be greater than JSA i.e. $\beta^{IB} > \beta^{JSA}$.

4 Results

In this section we present both average impact of Jobcentre Plus on the outflow from benefits, as well as its dynamic evolution, and then turn to examining its effect on inflows.

4.1 Basic Results on Outflows from benefits

In column (1) of Table 2 we estimate equation (1) where the dependent variable is the ln (total outflow) from unemployment (JSA claimants). Our controls include the stock of unemployed claimants at the end of the previous quarter by age group, time (quarter by year), age and district dummies. The coefficient on the post-policy dummy is negative and significant with a value of -0.0153 suggesting that a treated district on average, a 1.5% *decrease* in unemployment outflows. Given the average unemployment outflow of about 650,000 per quarter this implies just under 10,000 more people staying on unemployment benefits. This overall impact is consistent with the idea of disruption effects and/or that the job entry point system gave incentives for benefit officers to substitute effort away from the unemployed and towards the disabled. The lagged stock of own age unemployed claimants enters with a significant positive coefficient as expected (the larger stock the larger the flow, all else equal). The coefficient on the stock of the other age group is negative, suggesting job competition effects across age groups. In column (2), we include a full set of fixed effects (district by age dummies) and again find a significant negative coefficient on the post-policy variable and this result remains basically unchanged when we also include age interacted with time effects in column (3).

We repeat the same specifications in columns (4)-(6) of Table 2 for the (log) outflow from IB, and condition on the lagged IB stocks on the right hand side. In column (4) we estimate the specification analogous to column (1) and find a positive and weakly significant coefficient of 0.0166 on the post-policy dummy. This suggests about a 1.7% increase in total outflows, which given a sample average outflow of 56,000 people means an additional 1,000 fewer people off the IB register. In column (5), we include a full set of fixed effects and in column (6) we include age*time dummies. The treatment effect falls slightly to 0.0151, broadly equal and opposite to the policy coefficient in the unemployment outflow equation.[12]

The specifications in Table 2 just look at the post-policy period as a whole without examining the dynamics of the policy effect. To assess short-run effects arising

[12] The results are robust to conditioning on stocks (by age group) of other benefit recipients (i.e. IB and lone parent stocks in JSA outflow equations, JSA and lone parent stocks in IB outflow equations).

from organizational disruption, Table 3 probes the dynamics more carefully, allowing a differential effect in each of the quarters after the policy switches on (up to the eighth quarter after introduction) as in equation (2).

Interestingly, the coefficients show a consistent dynamic pattern, being negative in the quarter immediately after the policy's introduction, but then becoming more positive over time. We detect negative impacts on JSA outflows for the first 5 quarters, but these cease to be significant by quarters 6 and 7 and actually turn positive for quarter 8 and beyond. This suggests that after two years there is a positive effect of 1.2% on outflows due to the policy, although this effect is not significantly different from zero. In contrast, for IB outflows, although we find a negative effect in the first quarter after the policy is introduced, this turns positive by the second quarter. This positive effect gradually becomes larger and more significant until in the long-run it suggests 6.1% more disabled people left the register in our most general specification of column (6).

These dynamic responses are presented graphically in Figure 4 and highlight our two main findings. First, the long-run effect is positive for both forms of welfare, but it is clearly much stronger for disability benefits (Panel B) than unemployment benefits (Panel A). Second, there is initially a negative effect for both benefits of the policy change, but this is much stronger for unemployment than disability benefits.

The interpretation of our results is that the more positive effect of the policy on disability compared to unemployment is driven by the new incentive system, such that officers devote more effort to helping the disabled into new jobs than the unemployed after the policy change. Overlaid on this, however, is an initial disruption effect as buildings and new systems bed down and a generally positive effect on both groups from improved information. This depresses all outflows and is an adjustment cost of the policy.

An alternative explanation would be that incentives do not matter but somehow the information treatment had a disproportionately larger effect on IB claimants than the unemployed. It is not obvious why this should be, but we will look at a more refined test of the incentives hypothesis involving a third group of welfare recipients (lone parents) where bureaucratic incentives are somewhere in-between those for the other two groups.

We find results again consistent with the incentive hypothesis when looking at outflows from benefits for this third group (see Section 6).

4.2 Pre-policy trends?

An identification threat to any difference-in-differences estimation is the existence of differential pre-policy trends. For example, if districts initially selected for treatment were those in which IB outflows were already increasing (and/or JSA outflows decreasing), we would estimate a positive (respectively, negative) impact of treatment even in the case in which the policy had no real effect. To investigate this we look at pre-treatment trends by estimating the following augmented specification of equation (1):

$$\ln Y_{ait}^{B} = \sum_{k=1}^{K} \beta_k^B D_{it-k} + \beta^B D_{it} + \gamma_1^B \ln U_{ait-1}^{B} + \gamma_2^B \ln U_{a'it-1}^{B} + \delta_{ai}^{B} + \delta_{at}^{B} + \varepsilon_{ait}^{B} \qquad (3)$$

The first term on the right hand side of equation (3), $\sum_{k=1}^{K} \beta_k^B D_{it-k}$ allows for pre-policy trends. The results are reported in Table 4. Column (1) replicates our baseline specification for JSA outflows (column (3) of Table 2) for comparison. In column (2) we include four pre-treatment lags and note that the coefficients on the pre-treatment dummies are jointly insignificant (F-test =1.88). We perform the same specifications for IB in columns (3) and (4) and again find no evidence of pre-treatment effects.

Although this is reassuring, one caveat is that the individual dummy for the quarter immediately prior to treatment is significant at the 10% level for JSA in column (2). This could be due to the fact that our treatment indicator is based on the true "go live" date of Jobcentre Plus and there is likely to be some organizational disruption in advance of that date, which could spill into the previous quarter.[13] This would reduce the benefits of the policy.

[13] National Audit Office (2008) states that "It introduces a radical shift from the former impersonal surroundings of the Jobcentre and Social Security offices to a modern retail-style environment and has a major impact on the way in which staff interact with customers and hence the quality of service provided."

4.3 Inflow Rates

Our analysis focuses on the intended Jobcentre Plus outcome to increase the outflow rates from benefits, but a possible side effect is that the inflow rate into benefits also changes as a consequence of the policy change. For example, individuals may have an incentive to apply for IB rather than JSA if they perceive the new regime as "tougher" for JSA claimants, thereby inducing a change in size and composition of IB and JSA stocks and in the corresponding outflow rates. The resulting bias in the estimated policy effect is hard to sign. One would expect an upward bias in the estimated effect of the policy on the JSA outflow (and a downward bias for the IB outflow) if the dissuaded individuals were less motivated in their job search and more weakly attached to the labour force – and vice versa.

To examine this issue directly we analyse the impact of the Jobcentre Plus on *inflows* into JSA and IB. We estimate a specification similar to equation (1) but use the inflow into each benefit category as the dependent variable instead of the outflow:

$$\ln(INFLOW_{ait}^{B}) = \sum_{\tau=0}^{7} \beta_{\tau}^{B} D_{it+\tau} + \beta_{LR}^{B} D_{LR} + \rho \ln Pop_{ait} + \delta_{ai}^{B} + \delta_{at}^{B} + \varepsilon_{ait}^{B} \qquad (4)$$

In the outflow equation we controlled for the stock of existing benefit claimants, and the corresponding stock in the inflow equations is the age-specific population (Pop_{ait}). Ideally, as inflows (mostly) consist of people flowing from employment to unemployment, one should control for local employment on the right-hand side. But in the absence of high-frequency employment data at the district level we use the population figures as a proxy.[14]

In column (1) of Table 5 we show that, on average, Jobcentre Plus had no significant effect on the inflows into JSA. Along similar lines, column (3) shows no evidence of significant average effects on inflows into IB benefits. However, when we look at the dynamic impact on inflows in columns (2) and (4), we find that the policy

[14] We assign the mid-year population estimate from www.nomisweb.com (taken on the 30th of June each year) to all the quarters in the year. Using interpolated quarterly population estimates (from the mid-year estimates) does not change our results.

had initially a negative and significant impact on inflows into both benefits, although these become positive and insignificant in the long-run.

To address whether this could be a concern for our results because of changing selection, we repeat outflow regressions controlling for various lags of the corresponding inflows. The results are reported in Table 6, where all specifications include fixed effects for both district*age and age*time interactions.[15] Columns (1)-(3) refer to JSA outflows. Although the coefficients on the inflow variables, whether one or four lags, are positive and significant as one would expect (since more recent welfare recipients are more likely to leave), our main results are robust to their inclusion. To see this, in column (3) we report our baseline equation (2) on the same sample as column (2), as some observations are lost when we condition on lagged inflows. The long-run effects in columns (1) and (2) are almost identical to those in our baseline specification of column (3), and the dynamic effects only slightly muted by the inclusion of inflows. Columns (3)-(6) refer to IB outflows, and all coefficients measuring the impact of policy are both qualitatively and quantitatively similar across specifications. In particular, the long-run positive effect of the policy on IB is still significant and only falls slightly from 0.0547 (column (6)) to 0.0503 in column (5). Hence, despite some effects in inflows, any change in composition arising from this does not appear to substantially affect our results. Figures 5 and 6 illustrate this graphically.

4.4 Outflows to employment vs. other destinations

The JSA (but not the IB) database allows us to disaggregate the outflows into alternative destinations, and in particular to look at outflows into work separately from outflows into other states (such as different benefits, training, inactivity, etc.) The results of this analysis are reported in Table 7, where columns (1) and (2) refer to outflows into work, while columns (3) and (4) refer to other destinations. The broad pattern for either destination looks similar to the overall outflow results, although the estimated effects appear stronger especially in the short run when looking at outflows into work rather than other destinations.

[15] In alternative specifications, we explicitly control for the duration composition of the stock of benefit claimants at the end of the previous quarter, and find that the baseline outflow results are robust.

Negative effects on JSA outflows into both work and non-work can be rationalized if one takes into account the “stick” (job search monitoring) and “carrot” (search effort assistance) components of the interactions between JSA claimants and dedicated staff at Job Centres. The change in the incentive structure implied that JSA claimants would receive less assistance with the job search process than before, thus lowering their job finding rates, at least in the short run. But insofar as poorer job search assistance also implied less frequent contact with JSA claimants, one may expect looser monitoring and fewer transitions off benefits due to sanctions or discouragement (see also Manning, 2009, and Petrongolo, 2009, for the effects of monitoring on the time spent on JSA benefits).

Another interesting point to be noted about columns (1) and (3) is that the congestion effect stemming from job competition by jobseekers from other age groups is clearly not present in the JSA outflow into other destinations, as the other age group could be competing for jobs in the labour market, but not for other destinations.

Overall, the results in both specifications in Table 7 are comparable to the earlier results on total outflow in Table 2. This reinforces the validity of using total outflow as our dependent variable to proxy for outflow to work.

5 Cost-benefit evaluation

We perform a simple cost-benefit analysis. First, we consider the immediate introduction of the policy as the staggered roll-out would not offer much general insight into costs and benefits of similar hypothetical policies. We conduct two thought experiments. First, we assume away the transitional disruption costs and consider reaching the steady state immediately. This is a best case benchmark. Second, we explicitly incorporate the dynamic effects estimated in Figure 3 and calculate social welfare. This produces lower benefits because (i) there are the disruption effects causing an initial increase in the welfare rolls and (ii) these costs and the sunk set up costs are borne in the short-run whereas the benefits are in the longer-run, so discounting will further reduce the social present value of the policy.

In performing the cost-benefit we take into account (i) the savings in administration costs implied by the reorganization of the welfare system; (ii) the increase in output implied by the impact of the policy on job finding; (iii) the net exchequer savings which enter into welfare through a lower deadweight loss taxation (the rest simply being transfers); (iv) the sunk set-up costs. We abstract away from the leisure gains of those on welfare.

5.1 Long-run Cost Benefit Analysis

The results of the steady-state analysis are represented in Table 8. According to audit reports the annual running costs post-policy were £3.3bn, about £240m lower than pre-policy as shown in rows 1-3. However, the sunk set up cost was estimated at £1.8bn (row 9). The steady-state impact of Jobcentre Plus on job creation is obtained from the long-run estimates reported in columns (3) and (6) of Table 3.

Conservatively, we assume that the long-run policy impact on unemployment outflows is zero, as although the point estimate is positive (0.012), it is insignificantly different from zero. We set the long-term impact on IB exits at 0.061. Using this estimate, we obtain the implied steady-state fall in the IB rate (IB stock over population), according to a flow model of IB entry and exit, as shown in Appendix B. Not all of these exits will be to jobs. Using the Labour Force Survey (LFS) quarterly panel data 1998Q2-2002Q2 (pre-policy) we observe that 30% of IB exists are to jobs, while 70% of terminations transit into other benefits or out of the labour force. We assume these non-employment exits would be to other benefits with cost on average equivalent to IB. This implies that IB spells that do not terminate into employment do not contribute to either job creation or to benefit savings and is again a conservative estimate of policy benefits, as many IB exits will be to zero welfare.

We consider three possible wage outcomes (our proxy for additional output) for individuals finding employment after a spell on IB: the minimum wage, the empirical mean wage for individuals ending a benefit spell (again from LFS), and the median wage in the overall wage distribution (from the ASHE 1% sample of taxpayers). The middle case seems the most realistic but the minimum wage and median wage scenarios provide useful lower and upper bounds, respectively. Columns (1) to (3) of Table 8 correspond to the three alternative wage outcomes considered. Row 4 reports weekly earnings for each wage outcome, and row 5 reports the increase in GDP obtained by combining wage levels with job creation resulting from IB exits. The overall GDP gains range between £0.5bn and £1.6bn p.a.

Row 6 reports the net gain resulting from a reduced deadweight cost of taxation. This is set to 40%[16] of the lower net exchequer cost arising from increased tax revenues and lower benefit payments. The mean IB payment in 2000 was £74.71 per week. When an IB recipient finds a job, this benefit saving is accompanied by a change in the tax revenue that depends on the earnings and the household composition of the recipient. We used the IFS TAXBEN[17] simulation model to approximate these additional taxes and benefits for the 30% of IB exits who found jobs.[18] Combining these elements produces a benefit from a lower deadweight loss between £85m and £90m.

The sum of the three components reported in rows 3, 5 and 6 of Table 8 represents the total annual benefit of the policy in steady state. This implies an annual net benefit of £900m-£2bn. This benefit needs to be compared to the sunk set-up cost of £1.9bn. It is clear that the policy covers the sunk costs of programme introduction in just over 2 years even on conservative assumptions. If we use the 3.5% social discount rate used by the UK government (HM Treasury, 2003) our cost-benefit analysis implies a net benefit of Jobcentre Plus in excess of £20bn, even on the most conservative assumptions (row 10).

5.2 Cost Benefit Analysis with transitional dynamics

The previous calculations are unrealistic as we are ignoring the transitional dynamics. We use all estimates of the dynamic effects of policy from columns (3) and (6) of Table 3, for each quarter since the policy change. However, we cannot impose the steady-

[16] This follows Gruber (2011).

[17] Estimates were provided by Barra Roantree of the Institute for Fiscal Studies using the IFS tax and benefit micro-simulation model, TAXBEN.

[18] We consider two household types, a single adult and a couple with two dependent children, and obtain the associated tax payments. We assume that two thirds of IB exits are represented by single adults, while the remaining third is represented by members of couples with two children consistent with our estimates from the LFS 1998Q2-2002Q2.

assumptions used to computer the steady-state rise in the number of jobs, as this would be equivalent to assuming that the JSA and the IB rates reach their steady state levels within a quarter. We thus simply obtain the out-of-steady-state number of jobs created as the predicted change in the benefit outflow in the relevant quarter, net of job separations during that quarter. With labour market churning, some of the workers who find jobs separate in subsequent quarters. We estimate these flows from the (pre-policy) LFS panel.[19] For individuals who were on JSA and found jobs 2.3% lost them in the next quarter and for IB the number was 0.5%.[20]

The three earnings scenarios, as well as the running costs, are the same as those considered for the steady state analysis of Table 8. However, we now need to track GDP and net exchequer gains or losses for each quarter since the policy change, for both JSA and IB recipients. With discounting, this will reduce the present value of the policy change because the losses are front-loaded. We maintain all other assumptions on job finding rates for IB and again use the empirical data in the LFS that shows that for JSA recipients, 70% of exits where to jobs.[21]

The evolution of costs and benefits over time is represented in Figure 7. The flat, solid line represents the set-up costs, while the three dashed lines represent cumulative benefits since the quarter in which the policy turns on, for the three different levels of earnings. Regardless of the earnings assumptions, flow social benefits eventually exceed the costs so although incorporating dynamics substantially dampens down the net benefits (by almost an order of magnitude), it does not reverse the earlier positive assessment of the program. The present value of the net benefit of the reform is about £3-£5bn which outweighs the £2bn sunk cost.

It is important that in the baseline case (middle dashed line in Figure 7), it takes a full *five years* after policy introduction for net benefits to exceed set-up costs. This is mainly because of declines in unemployment outflows during the first two years of the

[19] The job separation rate is obtained as the ratio of inflows into IB (JSA) to the employed population of working age.

[20] These quantitative results are very similar to an analytical approximation of the change in employment rates during the transition to a new steady state (see Appendix B2).

[21] For the benefit and tax simulation we assume that 70% of JSA exits who find jobs live alone, while 30% live in a couple with two children. For IB, about 67% of those who find jobs live alone, while 33% live in a couple with two children.

new regime. Only after five years are job entry gains sufficient to compensate both the initial job entry losses and the set-up cost.

So although this is a policy which clearly passes the cost-benefit test, even a benign policy maker will not cover the costs of introduction for five years. Constitutionally, UK general elections have to be held at least once every five years (it is typically four years) and the tenure of a minister is usually only two years. These form of welfare policies are like classic investment decisions. Since a politician's discount rate will be much higher than the social discount rate we should expect systematic under-investment, which is what we indeed generally see (e.g. Aghion et al, 2013).

Given this, how is welfare reform ever possible? Sometimes a crisis hits, forcing radical reforms. But perhaps the large majorities enjoyed by Prime Minister Tony Blair in 1997 and 2001 enabled the government to pursue longer run policies, at least in welfare reform where there is substantial cross-party consensus.

6 Robustness and Extensions

6.1 Disaggregating the treatment effect by wave

As discussed above, the roll out of the policy was introduced in six staggered waves across the country. Our baseline estimates exploit all waves for identification, but an important issue is whether there is any heterogeneity in the treatment effect across different waves. For example, a legitimate worry could be that the dynamic effects that we identify in Figure 3 may be instead due to averaging over different effects in earlier and later waves.

To investigate this we estimated equation (1) separately for each wave of the policy roll-out. In order to avoid conflating the dynamics with wave effects we kept to a fixed post treatment window of one year. The results are reported in Table 9. Although the standard errors are larger as the number of observations is substantially reduced, the estimated treatment effect is remarkably stable across the different waves. Panel A refers to JSA outflows. Compared to the pooled effect we estimated in Table 2 of -0.015, wave-specific estimates range from -0.10 (wave 5) to -0.20 (wave 1) which is a reasonably tight bound. IB estimates in Panel B are generally higher (a range of 0.011 to

0.32) than the pooled estimate of 0.015, suggesting, if anything, that we might be underestimating the positive effect of the programme using the parsimonious specification of equation (1).

We also considered alternative specifications such as restricting the comparison areas to those that had not been treated (Table A3), which lead to similar results.

6.2 Lone Parents' Benefits

Besides JSA and IB, Lone Parents (single moms) on income support (LP) are the third big category on welfare rolls. From Table A1 we see that the introduction of job entry points increased bureaucratic incentives to place IB recipients into work and reduced the incentives to place JSA recipients into work. However, we also see that the points awarded to placing a lone parent in a job is the same as placing some on IB. Hence, might expect to see this group responding in similar ways to the policy as the IB group.

The lone parent groups were not subject to the same pressures to look for jobs as the other two groups, however, so this means the policy treatment may be weaker. There were also a raft of other policies aimed at lone parents during the same time period including a large increase in the generosity of in-work benefits for lone parents (similar to EITC) and a voluntary job assistance programme ("New Deal for Lone Parents"). These changes may contaminate our tests.

In column (1) of Table 10 we estimate equation (1) for welfare outflows for lone parents and obtain an average decrease of about 1.3% after the policy, only slightly below the 1.5% for IB. However, when testing for the presence of pre-trends in column (2) based on the analogue of equation (3), we find that the joint F-test rejects the hypothesis of no pre-policy trends (F=3.894). Whereas we did not find evidence of differential pre-treatment trends for the unemployed or disabled (Table 4). We attempt to control for these pre-trends by including district-specific trends in column (3) and the joint F-test for pre-trends is now insignificant. Similar to the unemployment and disability results, however, we do find a negative effect one quarter before treatment, consistent with the impact of organizational restructuring which impedes service delivery even before the true "go live" date. When pre-treatment dummies are dropped in column (4), the coefficient on the policy variable is -0.001 and insignificant. This value falls about half way between the IB and JSA effects. In column (5) we estimate a dynamic specification, and find an initial negative impact, which becomes positive by quarter 6 and is significantly positive in the long run. The long-run effect of 2.5% is smaller than the long-run IB effect of 6.1% but larger than the JSA effect of 1.2% (Figure 8).

Overall, the treatment effects on LP appear to lie between the effects of JSA and IB. Just like the other benefits there appears to be an initial negative effect which we interpret as an organizational disruption effect. However, in the long-run there is a positive improvement consistent with an improvement in incentives. Given the evidence of pre-policy effects, we place less weight on these results than for the other two groups, but the results do seem broadly supportive of our general story.

6.3 Spillover Effects

One concern with these estimates is that, in common with standard difference in difference approaches we do not look at general equilibrium effects of the policy. For example, Crepon et al. (2012) find that there can be unintended negative externalities of active labour market policies as the higher outflows from one benefit group take jobs at the expense of others, especially in depressed labour markets. We examine this idea by looking at outflows in neighbours to treated districts. For example, we estimated:

$$\ln Y_{ait}^{B} = \beta^{B} D_{it} + \mu^{B} NBR_{it} + \gamma_{1}^{B} \ln U_{ait-1}^{B} + \gamma_{2}^{B} \ln U_{a'it-1}^{B} + \delta_{ai}^{B} + \delta_{at}^{B} + \varepsilon_{ait}^{B} \qquad (5)$$

We capture spill overs using a dummy (NBR) that is unity in the quarter when a district's neighbours is treated and zero otherwise. The associated effect is captured by the parameter μ^{B} and is identified by the fact that different districts' had their first neighbour treated in different quarters. We define neighbouring districts as those with centroids within 10 km of the centroid of the main district.

The results are shown in Table 11. The sample size decreases due to the fact that the estimation is based on the first five waves since all neighbours are anyway treated by the sixth wave. Some of them have neighbours further away. The baseline impacts on

JSA and IB hold true even in this sample as seen in columns (1) and (3). In column (2), the coefficient on NBR_{it} is positive, consistent with spillover as the unemployed find it easier to get jobs due to lower JSA outflows in the treated areas. The coefficient is statistically insignificant, however. Similarly, in column (4), the coefficient on NBR_{it} is positive but insignificant.

We investigated a large number of other specifications to examine such congestion effects such as looking at other bandwidths than 10km, examining the proportion of treated neighbours rather than a discrete dummy, weighting by distance, interacting the policy and spillover effects with measures of labor market rightness (using vacancy rates), interacting the policy effects with lagged stocks of benefit claimants, etc. In no case could we find evidence that the policy had significant effects on other groups.

6.4 Other Robustness Tests

We have subjected our results to many other robustness tests, some of which we note here.

Alternative dynamic specifications. Other dynamic patterns reported in Table A4 again confirm the robustness of our main specifications.

Weighting. A worry is that our results could be driven by a few small districts. To address this concern, we weight observations by the district level age-specific benefit caseload in a pre-policy period (1999Q3). Table A5 reports the results for equation (1) using this weighting system. Column (1) has a coefficient for JSA of 2.5% Dynamic effects reported in column (2) are instead very similar to those of Table 3. For IB, the average effect reported in column (3) for IB is now lower than in the unweighted regression. We interpret this result as showing that smaller districts seemed to be more affected by the treatment for IB. The short and long run effects of the treatment are however very similar to those from the unweighted regression.

Estimates at Monthly Frequency. We are able to estimate JSA (but not IB) outflow equations at the monthly, rather than quarterly, frequency. The dependent variable is now the monthly outflow from JSA, having included the stock at the end of the previous month as a control. Column (1) in Table A7 shows a policy coefficient unemployment outflows of -1.6%, which is very close to the baseline 1.5%. The dynamic results in column (2) are also similar to the baseline.

Heterogeneity of the policy Effect. We investigated whether the treatment effects were heterogeneous in interesting ways across different groups. For example we looked at whether the coefficients in Table 3 columns (3) and (6) were different for welfare recipients of different ages, benefit durations, regions (e.g. London vs. others), and so on. What was surprising was that we did not find evidence for much systematic heterogeneity across these groups.

7 Conclusions

The UK embarked on a major change in the administration of welfare benefits for the unemployed and the disabled in 2001 with the introduction of Jobcentre Plus. Bureaucratic incentives to place the disabled into jobs were sharpened and offices were re-organised to be more efficient (e.g. in their use of IT). At the same time, the growth of the stock of those on disability benefits (IB) which had been rising for 30 years (like the US and other advanced countries) stopped increasing. We evaluate this policy in the light of a model with incentives, information and adjustment costs. We show that there are potentially two unintended consequences of the change. First, the relative incentives to place the unemployed (JSA claimants) into jobs fell. Second, the re-organization of the job centres can be expected to have disruption costs which reduce outflow rates temporarily and are an additional (and generally ignored) cost of such changes.

We found several striking results that are consistent with the existence of incentive and organization effects. First, we find an increase in the outflow rates of the disabled and unemployed in the long-run (after two years), but the effects are much larger and only significant for the disabled. In the long-run outflows from disability benefits were 6.1% higher (and this was statistically significant) and JSA outflows 1.2% higher (and this was statistically insignificant). Second, there is evidence of important disruption effects with outflow rates falling after policy introductions, especially for the unemployed.

A quantitative simulation of the policy suggests that the short-run costs are easily outweighed by long-run benefits. However, the benefits of the program do take some time to be visible and this poses a problem for policy-makers whose time horizons may be much shorter than a social planner. This reveals the political economy problem at the heart of welfare reform: changes to the administration of the benefit system that have long-run benefits have significant short-run costs and this makes it hard to build up a coalition for change.

There are many directions we want to take this work. To what extent does the increased labour supply lead to lower equilibrium wages (not just due to compositional changes)? Can we unbundle further some of the elements of the policy to distinguish incentives effects from information (which conceivably could be more important for the disabled)? These are areas we are actively engaged in exploring.

References

Aghion, Philippe, Tim Besley, John Browne, Francesco Caselli, Richard Lambert, Rachel Lomax, Nick Stern and John Van Reenen (2013) *Investing for Prosperity: Report of the LSE Growth Commission,* LSE: London.

Autor, David (2012) "The Unsustainable Rise of the Disability Rolls in the United States: Causes, Consequences and Policy Options", Brookings Papers.

Autor, David and Mark Duggan (2003) "The rise of disability rolls and the decline in unemployment" *Quarterly Journal of Economics* 188(1), 157-206.

Bagaria, Nitika, Novella Bottini and Miguel Coelho (2013), "Human Capital and Economic Growth: A Focus on Primary and Secondary Education in the UK", *LSE Growth Commission Report*.

Barnes Martin and Richard Poynter (1997), *Jobseeker's Allowance Handbook*, Child Poverty Action Group, London.

Beatty, Christina and Steve Fothergill (2009), "Incapacity Benefits in the UK: An Issue of Health or Jobs?", Sheffiled Hallam University mimeo.

Besley, Tim and Maitreesh Ghatak (2005), "Competition and Incentives with Motivated Agents", *American Economic Review* 95(3): 616-636.

Bloom, Nick and John Van Reenen (2011), "Human Resource Management and Productivity", *Handbook of Labor Economics Volume 4B* in Orley Ashenfelter and David Card (eds), Chapter 19 1697-1769.

Blundell, Richard, Monica Costa Dias, Costas Meghir and John Van Reenen (2004), "Evaluating the Employment Impact of a mandatory job search assistance programme", *Journal of the European Economics Association,* 2(4): 569-606.

Blundell, Richard, Claire Crawford and Wenchao Jin (2013), "Employment Composition, Hourly Wages and the Productivity Puzzle", *Economic Journal*, forthcoming.

Burchardt, Tania (1999), "The Evolution of Disability Benefits in the UK: Re-weighting the basket", mimeo.

Crepon, Bruno, Esther Duflo, Marc Gurgand, Roland Rathelot and Philippe Zamora (2012), “Do Labour Market Policies have Displacement effects? Evidence from a clustered randomized experiment”, *Quarterly Journal of* Economics, 128 (2): 531-380.

Delong, Brad and Larry Summers (2012), “Fiscal Policy in a Depressed Economy” *Brookings Papers on Economic Activity*, Spring, 233-274.

Dolton, P. and D. O'Neill (1996), "Unemployment Duration and the Restart Effect: Some Experimental Evidence", *Economic Journal*, 106: 387-400.

Dolton, P. and D. O'Neill (2002), "The Long-Run Effects of Unemployment Monitoring and Work-Search Programs: Experimental Evidence from the United Kingdom", *Journal of Labor Economics,* 20, 381-403.

Gregg, Paul, Stephen Machin and Maina Salgade (2013), “Real Wages and Unemployment in the Big Squeeze”, *Economic Journal*, forthcoming.

Gregg, Paul (2009), “Welfare Reform and Lone Parents in the UK”, *The Economic Journal*, 119:F38-F65.

Gruber, Jonathan (2011), “Public Finance and Public Policy”, Worth Publishers.

Heckman, James, Carolyn J. Heinrich, Pascal Courty, Gerald Marschke, and Jeffrey Smith (2011), “The Performance of Performance Standards”, Upjohn Institute.

HM Treasury (2004) “Opportunity for all: The strength to take the long-term decisions for Britain”, Pre-Budget Report, London: HMSO.

HM Treasury (2003) “The Green Book: Appraisal and evaluation in Central Government”, London: HMSO.

Holmström, Bengt and Paul Milgrom (1991), “Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design”, *Journal of Law, Economics, and Organization* 7(0): 24–52.

Hyde, Mark, John Dixon and Melanie Joyner (2002), “Work for Those Who Can and Security for Those Who Cannot: The United Kingdom’s New Social Security Reform Agenda”, *International Social Security Review,* 52(4): 69-86.

Johnson, Terry R. and Daniel H. Klepinger (1994), "Experimental Evidence on Unemployment Insurance Work-Search Policies", *Journal of Human Resources,* 29: 695-717.

Lalive, Rafael, Jan van Ours and Josef Zweimüller (2007), "How Changes in Financial Incentives Affect the Duration of Unemployment", *Review of Economic Studies,* 73: 1008-1038.

Layard, Richard (2000), "Welfare to Work and the New Deal", *The Business Economist* 31, 3, 28-40.

Manning, Alan (2009), "You Can't Always get What You Want: The Impact of the UK Jobseeker's Allowance", *Labour Economics*, 16: 239-250.

Makinson, John (2000), "Incentives for Change: Rewarding Performance in National Government Networks", Public Services Productivity Panel, HMT.

Meyer, Bruce (1995), "Lessons from the U.S. Unemployment Insurance Experiments", *Journal of Economic Literature,* 33: 91-131.

Mueller, Andreas I.**,** Jesse Rothstein and Till M. von Wachter (2013), "Unemployment Insurance and Disability Insurance in the Great Recession", NBER Working Paper no. 19672.

National Audit Office (2008), "The roll-out of the Jobcentre Plus Office network", London: HMSO.

Office for National Statistics, 2011 Census: Digitised Boundary Data (England and Wales), UK Data Service Census Support. Downloaded from: http://edina.ac.uk/census.

Pessoa, Joao and John Van Reenen (2013), "The UK Productivity and Jobs Puzzle: Does the Answer Lie in Labour Market Flexibility?", *Economic Journal*, forthcoming.

Petrongolo, Barbara (2009), "What Are the Long-Term Effects of UI? Evidence from the UK JSA Reform", *Journal of Public Economics,* 93: 1234-1253.

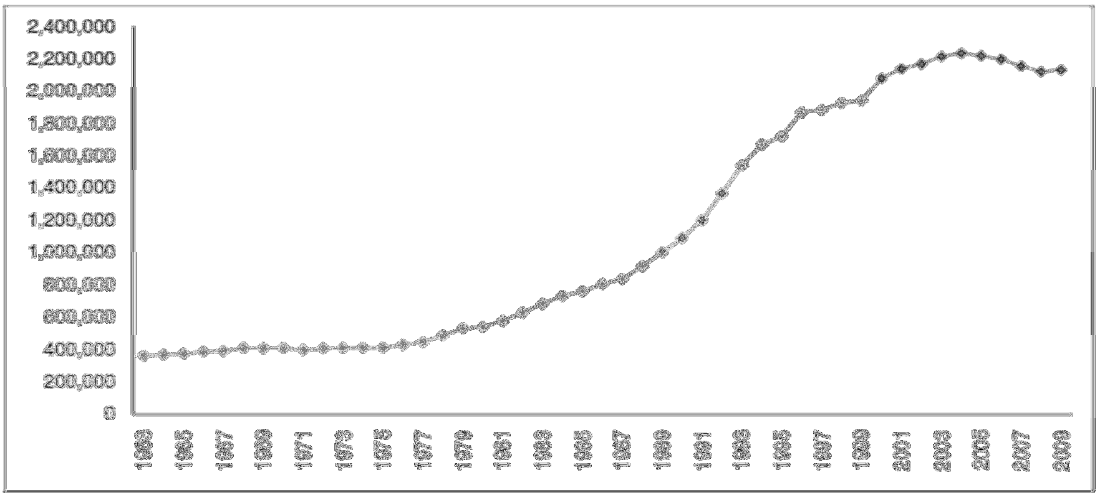
Propper, Carol and Deborah Wilson, (2003), "The Use and Usefulness of Performance Measures in the Public Sector", CMPO Working Paper 03/073.

Riley, Rebecca, Helen Bewley, Simon Kirby, Ana Rincon-Aznar and Anitha George (2011), *The introduction of Jobcentre Plus: An evaluation of labour market impacts*, DWP Research Report No 9781.

Van Reenen, John (2004) "Active Labour Market Policies and the British New Deal for Youth in Context" in Blundell, R., Card, D. and Freeman, R. *Seeking a Premier Economy*.

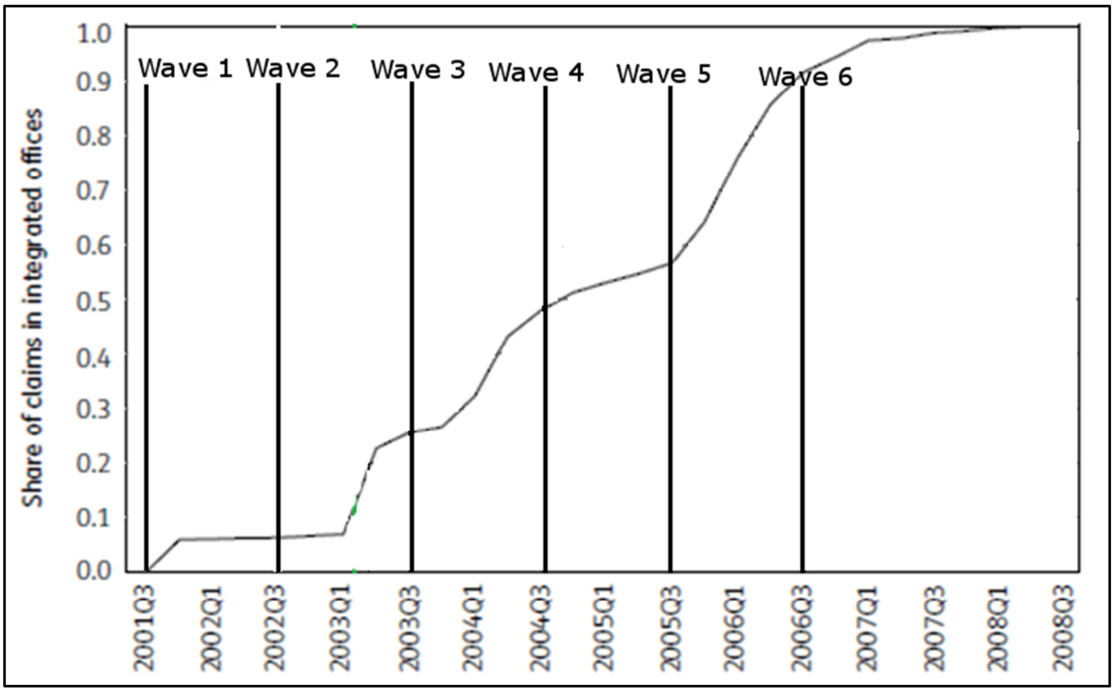
Tables and Figures

Figure 1: Incapacity Claimants of working age – 1963 -2009



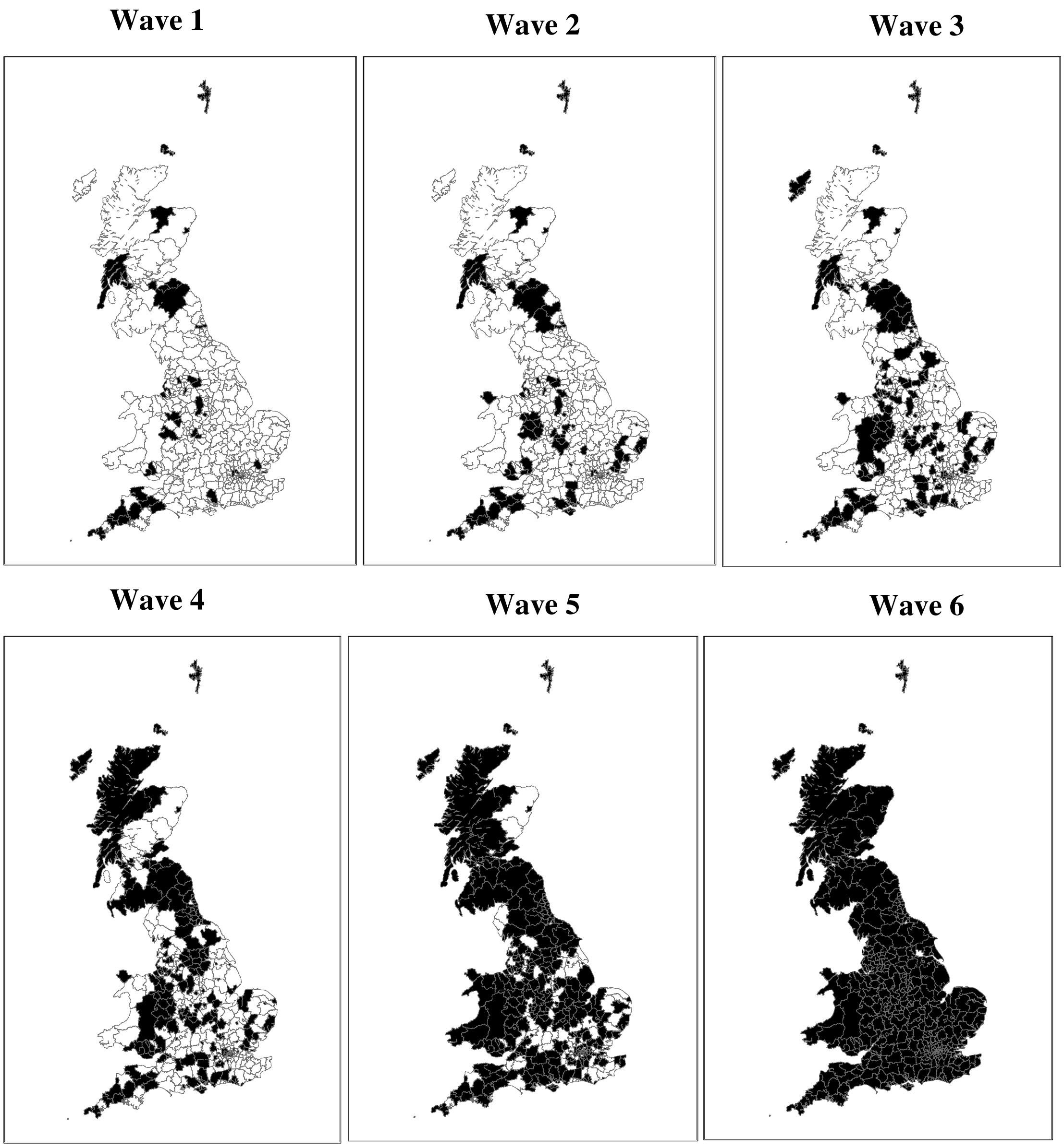
Source: Beatty and Fothergill (2009)

Figure 2: The Staggered Roll-out of the "Jobcentre Plus" Policy



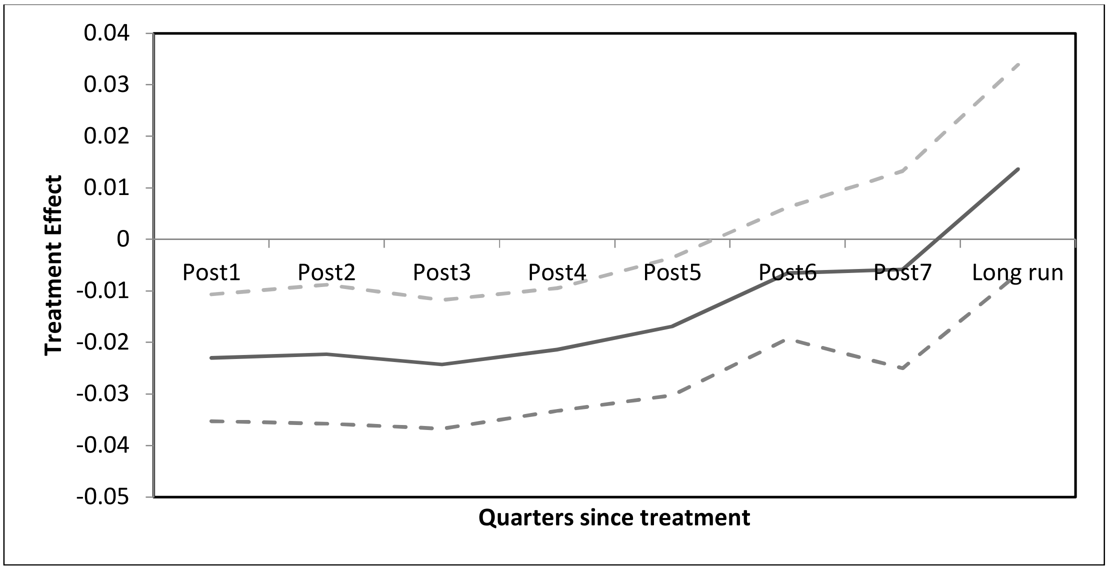
Notes: The vertical lines indicate the six waves of the roll-out of the policy (at the start of each wave at least one office switched to the new regime in a district). In Wave 1 there were 32 districts, in Wave 2 there were 27 districts, in Wave 3 there were 36 districts, in Wave 4 there were 28 districts, in Wave 5 there were 135 districts and in Wave 6 there were 148 districts. The line shows the proportion of JSA claimants who were affected by the policy (i.e. each office is weighted by the stock of JSA claimants in the quarter that the policy was turned on. **Source**: Riley et al (2011)

Figure 3: Spatial Map of Diffusion, by Wave



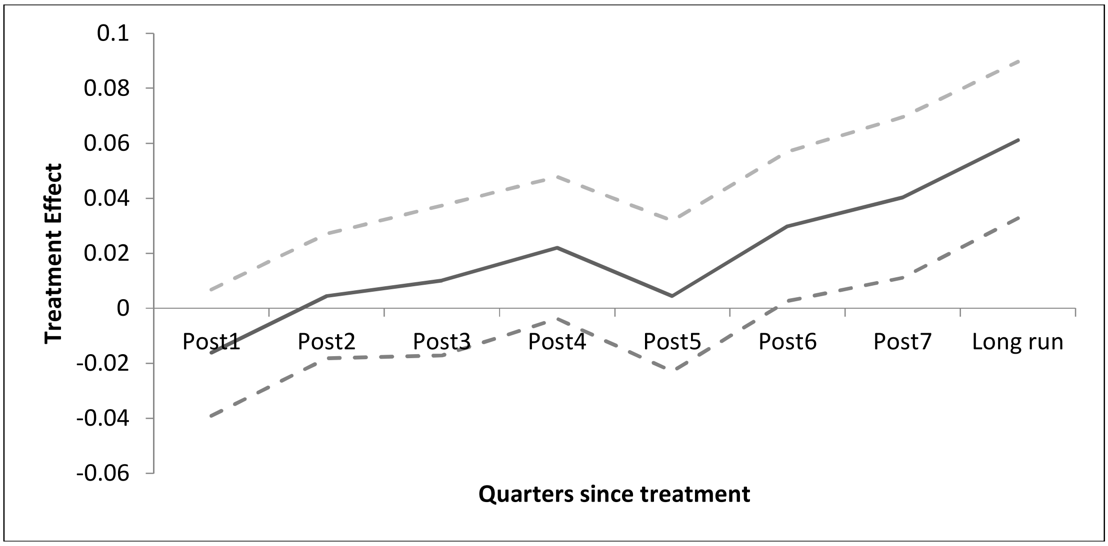
Notes: The maps show the additional districts covered under each wave. The treated districts are shaded in black. The first wave begun on 1st October 2001. In Wave 1 there were 32 districts, in Wave 2 there were 27 districts, in Wave 3 there were 36 districts, in Wave 4 there were 28 districts, in Wave 5 there were 135 districts and in Wave 6 there were 148 districts. By the first quarter of 2008, almost 100% of the country was covered.

Figure 4 Panel A: Dynamic Effects on JSA Outflow



Notes: The sample is a panel of 406 districts from 1999Q3 to 2008Q2. The dependent variable is ln (outflow) for JSA. Standard errors are clustered at the district level. The middle line denotes the estimated coefficients for the dynamic specification in equation (2). The top and bottom lines denote the 95% confidence intervals.

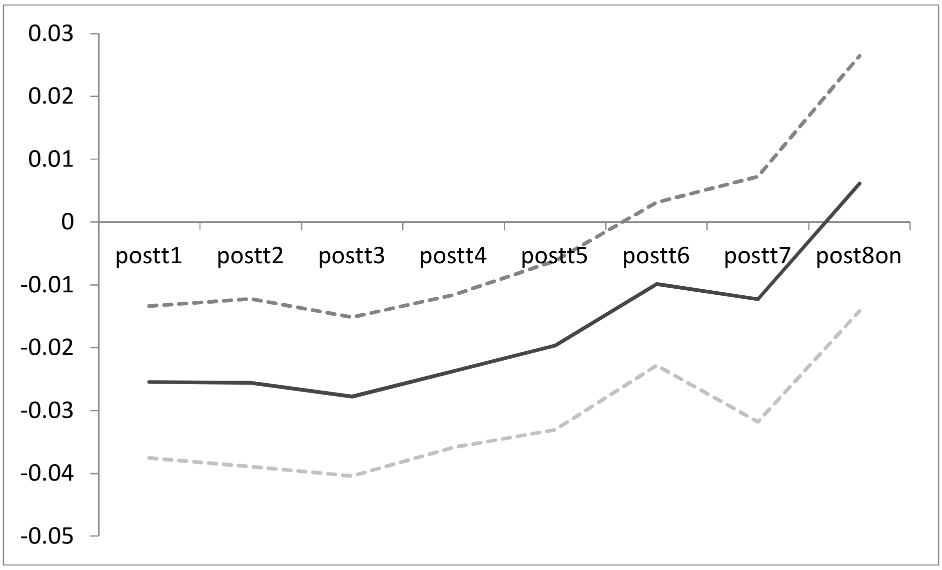
Figure 4 Panel B: Dynamic Effects on IB Outflow



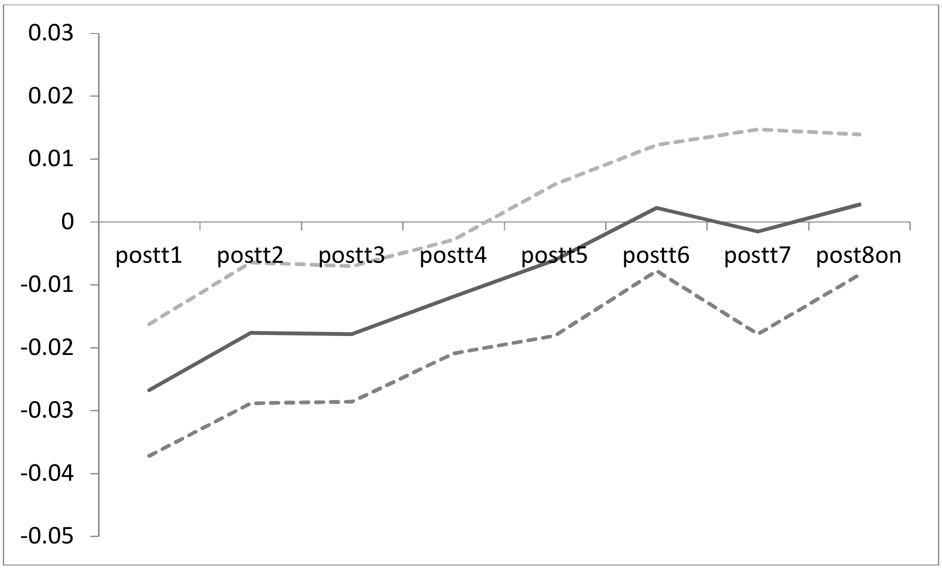
Notes: The sample is a panel of 406 districts from 1999Q3 to 2008Q2. The dependent variable is ln (outflow) for IB. Standard errors are clustered at the district level. The middle line denotes the estimated coefficients for the dynamic specification in equation (2). The top and bottom lines denote the 95% confidence intervals.

Figure 5: Dynamic Effects on JSA Outflows, with and without JSA inflow controls

Panel A: No controls for JSA Inflows



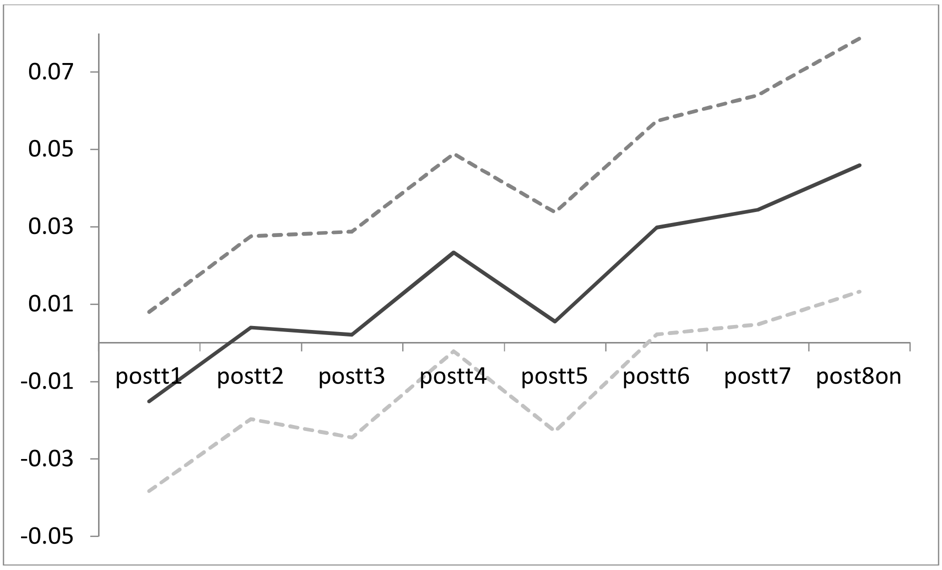
B: Controlling for a fourth order distributed lag of JSA Inflows



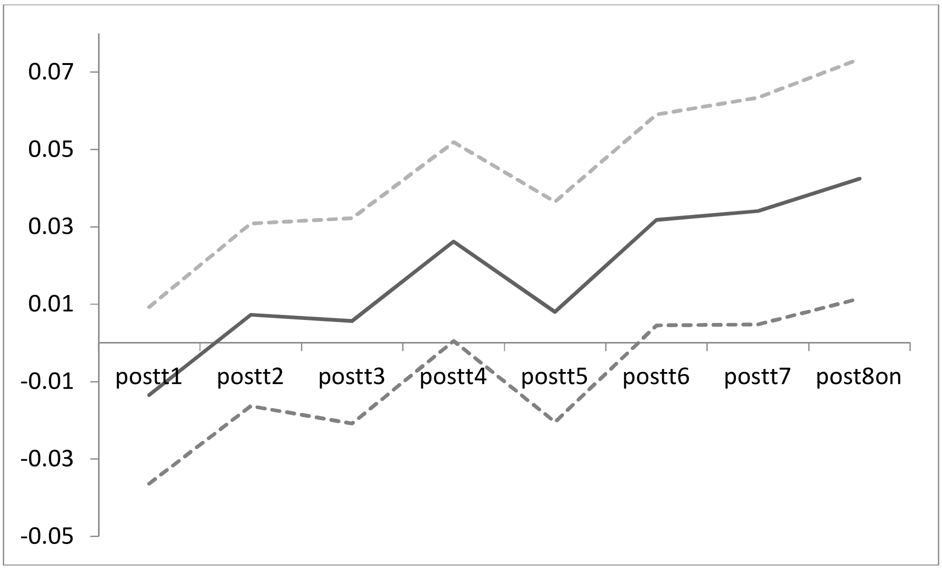
Notes: These are the coefficients in Table 6 column (2) and column (3). The dependent variable is ln (outflow) for JSA. Standard errors are clustered at the district level. The middle line denotes the estimated coefficients and the top and bottom dashed lines denote the 95% confidence intervals.

Figure 6: Dynamic Effects on IB Outflows, with and without IB inflow controls

Panel A: No controls for IB Inflows

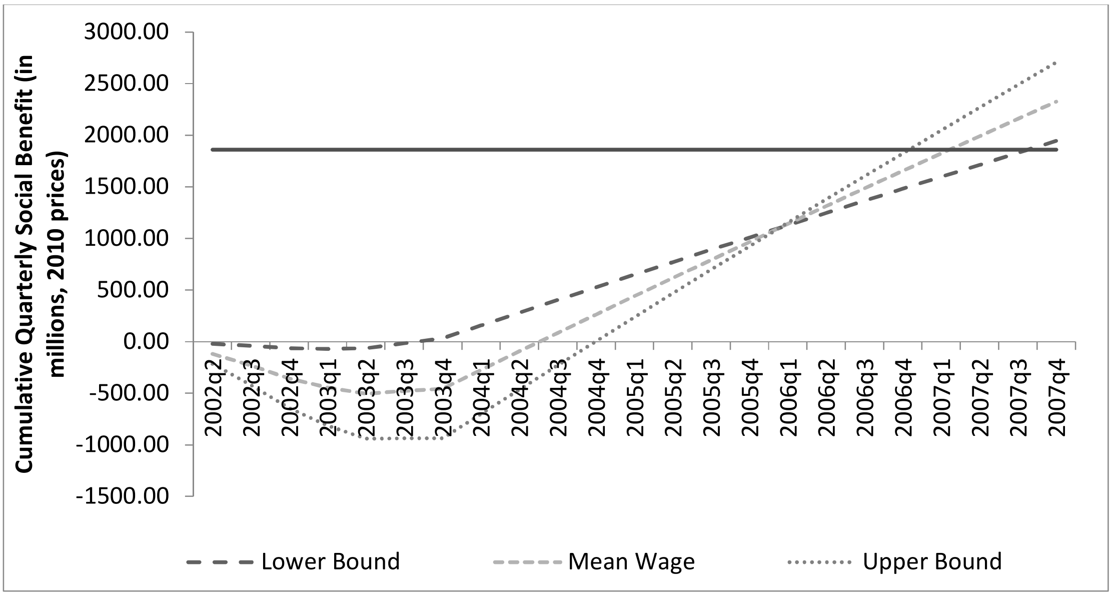


Panel B: Controlling for a fourth order distributed lag of IB Inflows



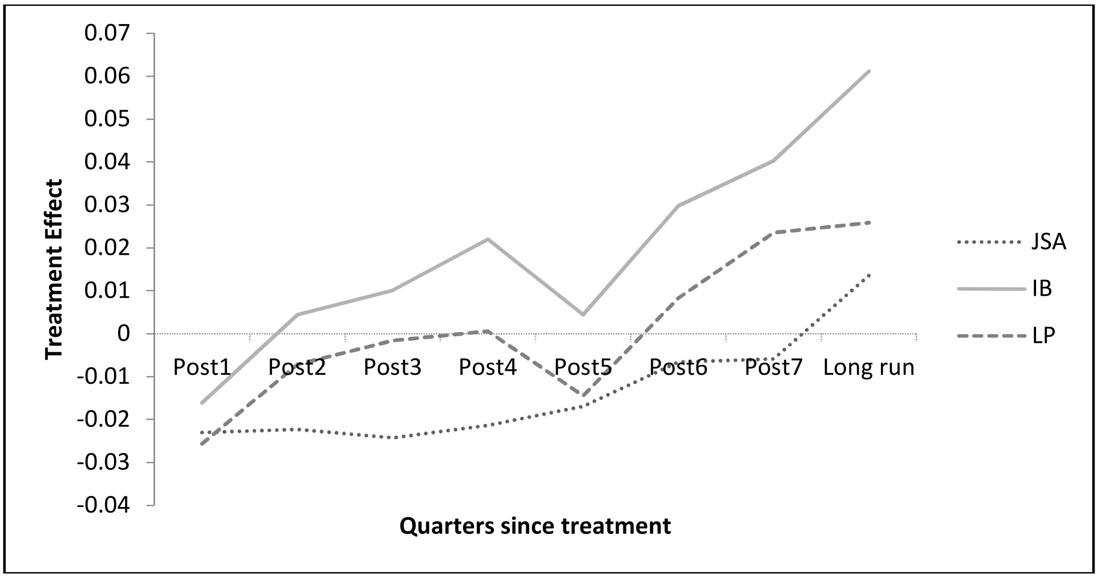
Notes: These are the coefficients in Table 6 column (5) and column (6). The dependent variable is ln (outflow) for IB. Standard errors are clustered at the district level. The middle line denotes the estimated coefficients and the top and bottom dashed lines denote the 95% confidence intervals.

Figure 7: Cost-benefit analysis of Jobcentre Plus: Dynamic evaluation



Notes: We consider an immediate roll-out of the policy. The solid horizontal line represents set-up costs of the policy. The dashed lines represent the cumulative benefit of the policy each in each quarter (increase in wage bill, lower deadweight costs of taxation and lower administrative running costs). See text for details.

Figure 8: Comparing Dynamic Effects on IB, JSA and LP Outflow



Notes: The sample is a panel of 406 districts from 1999Q3 to 2008Q2. The dependent variable is ln (outflow) for IB for the top line. The middle line represents ln (outflow) for LP and the bottom line for

JSA. The three lines denote the estimated coefficients for the dynamic specification in equation (2), plotted together for comparison across the three. Standard errors are clustered at the district level.

Table 1: Summary Statistics

	Quarterly Aggregate		Unweighted average across all district-age-quarter cells		
	Mean	SD	Mean	SD	Obs
	(1)	(2)	(3)	(4)	(5)
JSA Outflow	653,819	78,049	805	860	29,232
JSA Stock (t-1)	939,267	115,578	1,156	1,650	29,232
JSA Outflow rate (outflow(t)/stock(t-1))	0.698	0.057	0.871	0.266	29,232
JSA Inflow	648,957	58,156	799	843	29,232
JSA Outflow to Work	288,037	61,225	353	384	29,343
IB Outflow	56,166	11,267	70	106	29,232
IB Stock (t-1)	2,045,210	356,417	2,567	4,259	29,232
IB Outflow rate (outflow(t)/stock(t-1))	0.028	0.0027	0.048	0.043	29,232
IB Inflow	148,318	12,125	181	241	29,232

Notes: These are descriptive statistics across all districts in our sample for the UK over the period 1999Q3-
2008Q2. The first two columns aggregate stocks and flows to the year-quarter level and then average over the 36 quarters in our sample. The last 3 columns present the unweighted average of the district-age-quarter unit of observations used in our analysis. JSA= Job Seekers Allowance (unemployment insurance in the UK); IB=Incapacity Benefit, main form of disability benefit in UK.

Table 2: Policy Effects on outflow from unemployment (JSA) and disability (IB) benefits

	Log(Total Outflow)					
	(1)	(2)	(3)	(4)	(5)	(6)
Benefit:	**JSA**	**JSA**	**JSA**	**IB**	**IB**	**IB**
$\boldsymbol{Post_{it}}$	-0.0153***	-0.0152***	-0.0152***	0.0166*	0.0158*	0.0151*
	(0.0054)	(0.0055)	(0.0055)	(0.0089)	(0.0087)	(0.0087)
$\boldsymbol{lnU_{ait-1}}$	0.7249***	0.6355***	0.6323***	0.2495***	0.1462***	0.3475***
	(0.0095)	(0.0085)	(0.0100)	(0.0290)	(0.0195)	(0.0314)
$\boldsymbol{lnU_{a'it-1}}$	-0.0820***	0.0072	0.0105	0.1705***	0.2251***	0.0502*
	(0.0102)	(0.0086)	(0.0097)	(0.0246)	(0.0196)	(0.0256)
Observations	29,168	29,168	29,168	26,450	26,450	26,450
District*Age FE	NO	YES	YES	NO	YES	YES
Age*Time FE	NO	NO	YES	NO	NO	YES

Notes: Each column estimates equation (1) with the dependent variable as the log of the outflow from benefit during a year-quarter. All regressions control for district, time and age fixed effects. The sample is a panel of 406 districts from 1999Q3 to 2008Q2 and two age groups (18-25 and 26-60). In columns 1, 2 and 3, the dependent variable is ln (outflow) for JSA. In columns 4, 5 and 6, the dependent variable is ln (outflow) for IB. Standard errors are clustered at the district level. "Post" is a dummy equal to 1 in the post policy period and zero otherwise. U_{ait-1} is the lagged stock of individuals on benefits in the same age group (and $U_{a'it-1}$ the same for the other age groups). Time effects are a separate dummy for each quarter by year pair. *** p<0.01, ** p<0.05, * p<0.1

Table 3: Effects of policy allowing for dynamics

	Dependent Variable: Log(Total Outflow)					
Quarters after	(1)	(2)	(3)	(4)	(5)	(6)
Policy intro	**JSA**	**JSA**	**JSA**	**IB**	**IB**	**IB**
$Post_1$ **(t+1)**	-0.0234***	-0.0229***	-0.0230***	-0.0203*	-0.0154	-0.0162
	(0.0061)	(0.0062)	(0.0062)	(0.0119)	(0.0117)	(0.0117)
$Post_2$ **(t+2)**	-0.0230***	-0.0230***	-0.0230***	0.0052	0.0047	0.0044
	(0.0068)	(0.0068)	(0.0068)	(0.0119)	(0.0115)	(0.0115)
$Post_3$ **(t+3)**	-0.0249***	-0.0249***	-0.0249***	0.0112	0.0115	0.0101
	(0.0064)	(0.0064)	(0.0064)	(0.0140)	(0.0139)	(0.0139)
$Post_4$ **(t+4)**	-0.0208***	-0.0208***	-0.0208***	0.0254*	0.0223*	0.0220*
	(0.0061)	(0.0062)	(0.0062)	(0.0130)	(0.0132)	(0.0131)
$Post_5$ **(t+5)**	-0.0166**	-0.0166**	-0.0166**	0.0044	0.0045	0.0044
	(0.0068)	(0.0069)	(0.0069)	(0.0143)	(0.0139)	(0.0139)
$Post_6$ **(t+6)**	-0.0066	-0.0067	-0.0067	0.0370**	0.0309**	0.0298**
	(0.0065)	(0.0066)	(0.0066)	(0.0144)	(0.0138)	(0.0138)
$Post_7$ **(t+7)**	-0.0077	-0.0076	-0.0076	0.0430***	0.0415***	0.0403***
	(0.0098)	(0.0099)	(0.0099)	(0.0150)	(0.0150)	(0.0149)
$Post_{8LR}$ **(t>7)**	0.0117	0.0117	0.0117	0.0646***	0.0622***	0.0612***
	(0.0104)	(0.0105)	(0.0105)	(0.0150)	(0.0144)	(0.0145)
lnU_{ait-1}	0.7237***	0.6344***	0.6312***	0.2589***	0.1552***	0.3572***
	(0.0095)	(0.0084)	(0.0100)	(0.0288)	(0.0189)	(0.0308)
$lnU_{a'it-1}$	-0.0832***	0.0060	0.0092	0.1808***	0.2351***	0.0595**
	(0.0101)	(0.0085)	(0.0096)	(0.0239)	(0.0191)	(0.0255)
Observations	29,168	29,168	29,168	26,450	26,450	26,450
District*Age FE	NO	YES	YES	NO	YES	YES
Age*Time FE	NO	NO	YES	NO	NO	YES
F Test	4.5560	4.4980	4.4920	4.6310	4.3260	4.2410
p-value	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001

Notes: Post indicates the treatment effect each quarter after the policy is introduced). Each column estimates equation (2) with the dependent variable as the log of the outflow from benefit during a year-quarter. All regressions control for district, time and age fixed effects. The sample is a panel of 406 districts from 1999Q3 to 2008Q2. In columns 1, 2 and 3, the dependent variable is ln (outflow) for JSA. In columns 4, 5 and 6, the dependent variable is ln (outflow) for IB. The last row contains the p-value of the F test for the joint significance of the post-treatment dummies. Standard errors are clustered at the district level. U_{ait-1} is the lagged stock of individuals on benefits in the same age group (and $U_{a'it-1}$ the same for the other age groups). Time effects are a separate dummy for each quarter by year pair. *** p<0.01, ** p<0.05, * p<0.1

Table 4: Pre-treatment Trends in Benefit Outflow

	Log(Total Outflow)			
	(1) JSA	(2) JSA	(3) IB	(4) IB
$Post_{-4}$		0.0013		0.0162
		(0.0065)		(0.0109)
$Post_{-3}$		0.0039		-0.0135
		(0.0063)		(0.0114)
$Post_{-2}$		-0.0021		0.0053
		(0.0068)		(0.0119)
$Post_{-1}$		-0.0134*		-0.0010
		(0.0079)		(0.0123)
$Post_{it}$	-0.0152***	-0.0168**	0.0151*	0.0160
	(0.0055)	(0.0069)	(0.0087)	(0.0100)
lnU_{ait-1}	0.6323***	0.6323***	0.3475***	0.3473***
	(0.0100)	(0.0100)	(0.0314)	(0.0314)
$lnU_{a'it-1}$	0.0105	0.0105	0.0502*	0.0502*
	(0.0097)	(0.0097)	(0.0256)	(0.0256)
Observations	29,168	29,168	26,450	26,450
District*Age FE	YES	YES	YES	YES
Age*Time FE	YES	YES	YES	YES
F Test		1.8830		1.3560
p-value		0.1130		0.2480

Notes: Post indicates the treatment effect each quarter before the policy is introduced. Each column estimates equation (3) with the dependent variable as the log of the outflow from benefit during a year-quarter. All regressions control for district, time and age fixed effects. The sample is a panel of 406 districts from 1999Q3 to 2008Q2. In columns 1 and 2, the dependent variable is ln (outflow) for JSA. In columns 3 and 4, the dependent variable is ln (outflow) for IB. The last row contains the p-value of the F test for the joint significance of the pre-treatment dummies. Standard errors are clustered at the district level. U_{ait-1} is the lagged stock of individuals on benefits in the same age group (and $U_{a'it-1}$ the same for the other age groups). Time effects are a separate dummy for each quarter by year pair *** p<0.01, ** p<0.05, * p<0.1

Table 5: Analysis of Benefit Inflows

	Log(Total Inflow)			
	(1) JSA	(2) JSA	(3) IB	(4) IB
$\boldsymbol{Post}$	-0.0063		-0.0054	
	(0.0085)		(0.0072)	
$\boldsymbol{Post_1}$		-0.0155*		-0.0175**
		(0.0081)		(0.0080)
$\boldsymbol{Post_2}$		-0.0327***		-0.0179*
		(0.0081)		(0.0094)
$\boldsymbol{Post_3}$		-0.0170**		-0.0068
		(0.0086)		(0.0088)
$\boldsymbol{Post_4}$		-0.0267***		-0.0064
		(0.0086)		(0.0093)
$\boldsymbol{Post_{5LR}}$		0.0164		0.0043
		(0.0129)		(0.0094)
Ln(population)	0.1441	0.1340	-0.0072	-0.0134
	(0.1278)	(0.1254)	(0.0610)	(0.0609)
Observations	29,096	29,096	26,727	26,727
District*Age FE	YES	YES	YES	YES
Age*Time FE	YES	YES	YES	YES

Notes: All columns estimate equation (4) with the dependent variable as the inflow into benefits during a year-quarter. All regressions control for district, time and age fixed effects. The sample is a panel of 406 districts from 1999Q3 to 2008Q2. In columns 1-4, the dependent variable is ln (inflow). Standard errors are clustered at the district level. *** p<0.01, ** p<0.05, * p<0.1

Table 6: Dynamic Policy Effects on JSA and IB Outflows controlling for inflows

	Log(Total Outflow)					
	(1) JSA	(2) JSA	(3) JSA	(4) IB	(5) IB	(6) IB
$\boldsymbol{Post_1}$	-0.0235***	-0.0267***	-0.0254***	-0.0156	-0.0174	-0.0191
	(0.0054)	(0.0053)	(0.0061)	(0.0118)	(0.0118)	(0.0119)
$\boldsymbol{Post_2}$	-0.0149***	-0.0176***	-0.0256***	0.006	0.0048	0.0014
	(0.0057)	(0.0057)	(0.0068)	(0.0117)	(0.0119)	(0.0119)
$\boldsymbol{Post_3}$	-0.0145***	-0.0178***	-0.0278***	0.0089	0.0074	0.0039
	(0.0056)	(0.0055)	(0.0064)	(0.0140)	(0.0139)	(0.0139)
$\boldsymbol{Post_4}$	-0.0131***	-0.0119**	-0.0237***	0.0291**	0.0283**	0.0255*
	(0.0049)	(0.0046)	(0.0062)	(0.0133)	(0.0135)	(0.0134)
$\boldsymbol{Post_5}$	-0.0073	-0.006	-0.0196***	0.0075	0.0062	0.0036
	(0.0059)	(0.0061)	(0.0068)	(0.0141)	(0.0142)	(0.0141)
$\boldsymbol{Post_6}$	-0.0021	0.0022	-0.0098	0.0334**	0.0312**	0.0293**
	(0.0053)	(0.0051)	(0.0066)	(0.0138)	(0.0139)	(0.0140)
$\boldsymbol{Post_7}$	-0.0018	-0.0015	-0.0123	0.0401***	0.0362**	0.0363**
	(0.0082)	(0.0083)	(0.0099)	(0.0148)	(0.0150)	(0.0152)
						0.0547**
$\boldsymbol{Post_{8LR}}$	0.0064	0.0028	0.0062	0.0589***	0.0503***	*
	(0.0068)	(0.0057)	(0.0103)	(0.0144)	(0.0141)	(0.0145)
$\boldsymbol{ln(Inflow)_{-1}}$	0.4252***	0.3765***		(0.0244)	(0.0286)	(0.0285)
	(0.0200)	(0.0147)		0.0863***	0.0764***	
$\boldsymbol{ln(Inflow)_{-2}}$		0.0773***		(0.0133)	(0.0132)	
		(0.0109)			0.0544***	
$\boldsymbol{ln(Inflow)_{-3}}$		0.0457***			(0.0130)	
		(0.0137)			0.0344***	
$\boldsymbol{ln(Inflow)_{-4}}$		0.0827***			(0.0119)	
		(0.0108)			0.0359***	
$\boldsymbol{lnU_{ait-1}}$	0.3679***	0.3323***	0.6350***		(0.0121)	
						0.3761**
	(0.0149)	(0.0117)	(0.0105)	0.3312***	0.2775***	*
$\boldsymbol{lnU_{a'it-1}}$	0.0069	-0.0109*	0.0033	(0.0307)	(0.0344)	(0.0328)
	(0.0069)	(0.0059)	(0.0092)	0.0493**	0.0231	0.0346
Observations	28352	25915	25915	24402	22304	22304
District*Age FE	YES	YES	YES	YES	YES	YES
Age*Time FE	YES	YES	YES	YES	YES	YES

Notes: All columns estimate equation (2) with the dependent variable as the outflow from benefits during a year-quarter. All regressions control for district, time and age fixed effects. The sample is a panel of 406 districts from 1999Q3 to 2008Q2. In columns 1-3, the dependent variable is ln (outflow) for JSA and the dependent variable in columns 4-6 is the ln (outflow) for IB. Standard errors are clustered at the district level. U_{ait-1} is the lagged stock of individuals on benefits in the same age group (and $U_{a'it-1}$ the same for the other age groups). Time effects are a separate dummy for each quarter by year pair. *** p<0.01, ** p<0.05, * p<0.1

Table 7: Dynamic Effects on the Outflow to Work

	Log(Outflow by destination)			
	(1) To Work	(2) To work	(3) Not to work	(4) Not to work
$\boldsymbol{Post}$	-0.0244***		-0.0169***	
	(0.0065)		(0.0065)	
$\boldsymbol{Post_1}$		-0.0320***		-0.0402***
		(0.0074)		(0.0118)
$\boldsymbol{Post_2}$		-0.0163**		-0.0386***
		(0.0078)		(0.0125)
$\boldsymbol{Post_3}$		-0.0375***		-0.0355***
		(0.0090)		(0.0120)
$\boldsymbol{Post_4}$		-0.0218**		-0.0403***
		(0.0090)		(0.0119)
$\boldsymbol{Post_5}$		-0.0351***		-0.0253**
		(0.0107)		(0.0125)
$\boldsymbol{Post_6}$		-0.0139		-0.0238*
		(0.0095)		(0.0130)
$\boldsymbol{Post_7}$		-0.0187*		-0.0122
		(0.0107)		(0.0154)
$\boldsymbol{Post_{8LR}}$		-0.0082		0.0084
		(0.0107)		(0.0145)
$\boldsymbol{lnU_{ait-1}}$	0.6213***	0.6262***	0.6305***	0.5488***
	(0.0156)	(0.0157)	(0.0121)	(0.0215)
$\boldsymbol{lnU_{a'it-1}}$	-0.0313**	-0.0278**	0.0240**	0.0319*
	(0.0131)	(0.0129)	(0.0110)	(0.0193)
Observations	28,019	28,019	28,075	28,075
District*Age FE	YES	YES	YES	YES
Age*Time FE	YES	YES	YES	YES

Notes: Column 1 estimates equation (1) and column 2 estimates equation (2), both with the dependent variable as the log of the outflow from benefit to work during a year-quarter. All regressions control for district, time and age fixed effects. The sample is a panel of 406 districts from 1999Q3 to 2008Q2. In columns 1, 2 and 3, the dependent variable is log (outflow to work) for JSA. The last row contains the p-value of the F test for the joint significance of the post-treatment dummies. Standard errors are clustered at the district level. U_{ait-1} is the lagged stock of individuals on benefits in the same age group (and $U_{a'it-1}$ the same for the other age groups). Time effects are a separate dummy for each quarter by year pair. *** p<0.01, ** p<0.05, * p<0.1

Table 8: Cost-benefit analysis: Steady-state evaluation

	Re-employment earnings		
	Lower bound (min wage)	Mean re-employment earnings after benefits	Upper Bound (median wage)
1. Administration cost in old regime (£m)	3552	3552	3552
2. Administration cost in new regime (£m)	3314	3314	3314
3. Annual saving in administrative costs (£m)	238	238	238
4. Weekly earnings (ASHE 2000)	122.00	250.0	360.0
5. Increase in GDP from wage income (£m)	552.17	1131.50	1629.37
6. Deadweight gain (£m) (30% net exchequer saving)	113.68	204.34	121.79
7. Annual social benefit (£m)	903.85	1573.84	1989.16
8. PDV of social benefit (£m)	25824.34	44966.84	56833.08
9. Total JCP Setup Cost (£m)	1859	1859	1859
10. Net benefit (£m)	23965.34	43107.84	54974.08

Notes: This is a cost benefit analysis where we assume that the policy in introduced nationally and we immediately jump to the steady state effects (i.e. we ignore transitional dynamics). The administrative costs are from NAO (2008) and the benefits based on our econometric analysis discussed in the text.

Table 9: Treatment effect in individual waves

WAVE	1	2	3	4	5	6	Baseline
			Panel A: Log(Total Outflow) from JSA				
$\boldsymbol{Post_{it}}$	-0.0196	-0.0162*	-0.0147*	-0.0121	-0.0103	-0.0107*	-0.0152***
	(0.0135)	(0.0093)	(0.0078)	(0.0080)	(0.0073)	(0.0062)	(0.0055)
$\boldsymbol{lnU_{ait-1}}$	0.3253***	0.3407***	0.3602***	0.3835***	0.3916***	0.3949***	0.6323***
	(0.0182)	(0.0188)	(0.0160)	(0.0158)	(0.0156)	(0.0149)	(0.0100)
$\boldsymbol{lnU_{a'it-1}}$	0.3373***	0.2825***	0.2594***	0.2272***	0.2248***	0.2219***	0.0105
	(0.0195)	(0.0187)	(0.0180)	(0.0210)	(0.0224)	(0.0226)	(0.0097)
Observations	9727	12967	16207	19448	22688	25928	29168
			Panel B: Log(Total Outflow) from IB				
$\boldsymbol{Post_{it}}$	0.0114	0.0319	0.0315**	0.0295**	0.0254**	0.0171*	0.0151*
	(0.0279)	(0.0202)	(0.0159)	(0.0124)	(0.0108)	(0.0096)	(0.0087)
$\boldsymbol{lnU_{ait-1}}$	0.2221***	0.1511***	0.1800***	0.1624***	0.1534***	0.1626***	0.3475***
	(0.0493)	(0.0407)	(0.0355)	(0.0320)	(0.0308)	(0.0284)	(0.0314)
$\boldsymbol{lnU_{a'it-1}}$	1.1741***	0.7325***	0.3921***	0.4031***	0.4081***	0.3202***	0.0502*
	(0.2219)	-0.1647	(0.1307)	(0.1118)	(0.0980)	(0.0836)	(0.0256)
Observations	7635	10435	13256	16070	18844	21637	26450

Notes: The dependent variable is the log of the outflow from benefit during a year-quarter. All regressions control for interacted district-age and interacted age-time fixed effects. The regressions restrict the post-treatment period to 4 quarters after each wave. The sample is a panel of 406 districts for each wave. In the upper panel the dependent variable is ln (outflow) for JSA. In the lower panel, the dependent variable is the ln (outflow) from IB. Standard errors are clustered at the district level. U_{ait-1} is the lagged stock of individuals on benefits in the same age group (and $U_{a'it-1}$ the same for the other age groups). Time effects are a separate dummy for each quarter by year pair. *** p<0.01, ** p<0.05, * p<0.1

Table 10: Treatment Effect on Lone Parents

	Log(Total Outflow)				
	(1)	(2)	(3)	(4)	(5)
	LP	LP	LP	LP	LP
$\boldsymbol{Post_{-4}}$		-0.0120	0.0014		
		(0.0112)	(0.0111)		
$\boldsymbol{Post_{-3}}$		-0.0302***	-0.0166		
		(0.0103)	(0.0106)		
$\boldsymbol{Post_{-2}}$		-0.0140	-0.0003		
		(0.0119)	(0.0117)		
$\boldsymbol{Post_{-1}}$		-0.0401***	-0.0255**		
		(0.0119)	(0.0121)		
$\boldsymbol{Post_{it}}$	-0.0131*	-0.0270***	-0.0159*	-0.0096	
	(0.0067)	(0.0087)	(0.0095)	(0.0073)	
$\boldsymbol{Post_{1}}$					-0.0265**
					(0.0110)
$\boldsymbol{Post_{2}}$					-0.0067
					(0.0116)
$\boldsymbol{Post_{3}}$					-0.0016
					(0.0107)
$\boldsymbol{Post_{4}}$					-0.0020
					(0.0118)
$\boldsymbol{Post_{5}}$					-0.0166
					(0.0131)
$\boldsymbol{Post_{6}}$					0.0083
					(0.0115)
$\boldsymbol{Post_{7}}$					0.0211
					(0.0143)
$\boldsymbol{Post_{8LR}}$					0.0247**
					(0.0122)
$\boldsymbol{lnU_{ait-1}}$	0.4819***	0.4845***	0.5529***	0.5523***	0.5535***
	(0.0634)	(0.0632)	(0.0895)	(0.0895)	(0.0898)
$\boldsymbol{lnU_{a'it-1}}$	-0.1921***	-0.1907***	-0.1304*	-0.1307*	-0.1302
	(0.0553)	(0.0550)	(0.0790)	(0.0790)	(0.0793)
Observations	26,378	26,378	26,378	26,378	26,378
District*Age FE	YES	YES	YES	YES	YES
Age*Time FE	YES	YES	YES	YES	YES
District Trend	NO	NO	YES	YES	YES
F-Test		3.894	1.757		2.137
P-value		0.00408	0.137		0.0389

Notes: The dependent variable is the log of the outflow from benefit during a year-quarter. All regressions control for district, time and age fixed effects. The sample is a panel of 406 districts from 1999Q3 to 2008Q2. The dependent variable is ln (outflow) for LP. The last row contains the p-value of the F test for the joint significance of the pre-treatment dummies. Standard errors are clustered at the district level. U_{ait-1} is the lagged stock of individuals on benefits in the same age group (and $U_{a'it-1}$ the same for the other age groups). Time effects are a separate dummy for each quarter by year pair. *** p<0.01, ** p<0.05, * p<0.1

Table 11: Spillover effects of the policy into neighbouring districts

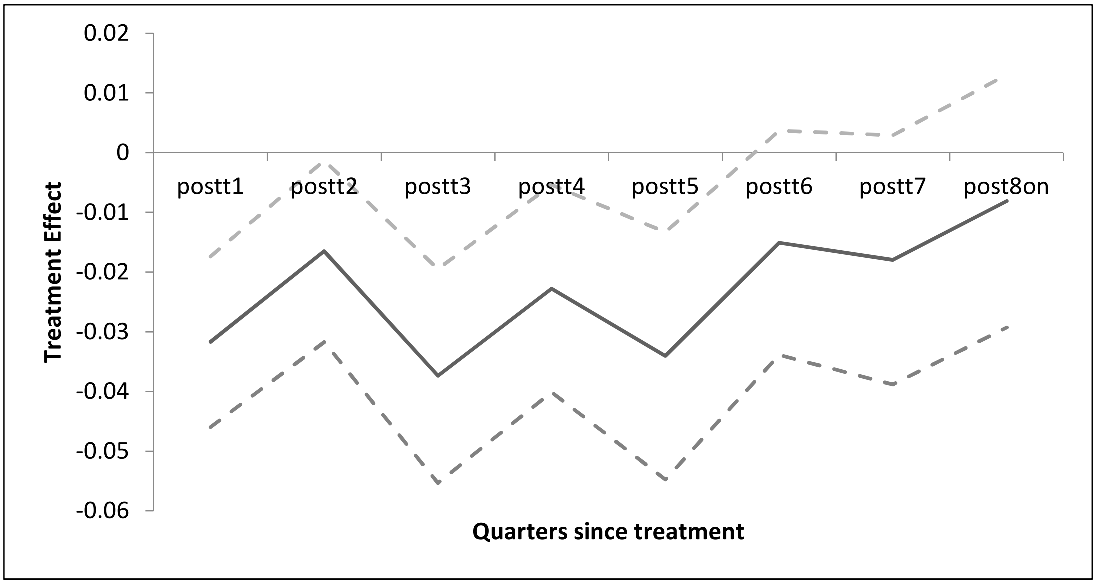
	Log (Total Outflows)			
	(1) JSA	(2) JSA	(3) IB	(4) IB
$\boldsymbol{Post_{it}}$	-0.0113	-0.0123*	0.0323***	0.0311***
	(0.0074)	(0.0073)	(0.0106)	(0.0106)
$\boldsymbol{NBR_{it}}$		0.0150		0.0146
		(0.0092)		(0.0126)
$\boldsymbol{lnU_{ait-1}}$	0.6262***	0.6242***	0.3514***	0.3522***
	(0.0108)	(0.0110)	(0.0383)	(0.0383)
$\boldsymbol{lnU_{a'it-1}}$	0.0099	0.0079	0.0673**	0.0683**
	(0.0101)	(0.0102)	(0.0297)	(0.0297)
Observations	22,688	22,688	20,374	20,374

Notes: Each column estimates equation (5) with the dependent variable as the log of the outflow from benefit during a year-quarter. All regressions control for district, time and age fixed effects. The sample is a panel of 406 districts from 1999Q3 to 2006Q2. In columns 1 and 2, the dependent variable is ln (outflow) for JSA. In columns 3 and 4, the dependent variable is ln (outflow) for IB. Standard errors are clustered at the district level. U_{ait-1} is the lagged stock of individuals on benefits in the same age group (and $U_{a'it-1}$ the same for the other age groups). Time effects are a separate dummy for each quarter by year pair. *** p<0.01, ** p<0.05, * p<0.1

Appendix

A. Appendix Tables and Figures

Figure A1: Dynamic Effects on JSA Outflow to Work



Notes: The sample is a panel of 406 districts from 1999Q3 to 2008Q2. The dependent variable is ln (outflow to work) for JSA. Standard errors are clustered at the district level. The middle line denotes the estimated coefficients for the dynamic specification in Table 7 Column (2). The top and bottom lines denote the 95% confidence intervals.

Table A1: Job Entry Target points (2002-03)

Client Group	Points Awarded
Disabled People and inactive benefits (IB)	12
Lone Parents (LP)	12
New Deal 50+, 25+, Young People	8
Other long term JSA	8
Short term unemployed JSA	4
Employed job-entries	1
Area-based points	1

Notes: The second column lists the number of points awarded to a benefit officer for placing a claimant from the corresponding benefit category in column (1) into jobs.

Table A2: Policy Effects on the JSA and IB Outflows from Offices

Log(Total Outflow)		
	(1)	(2)
	JSA	JSA
$\boldsymbol{Post_{-4}}$		-0.0025
		(0.0070)
$\boldsymbol{Post_{-3}}$		-0.0134*
		(0.0079)
$\boldsymbol{Post_{-2}}$		-0.0182**
		(0.0086)
$\boldsymbol{Post_{-1}}$		-0.0550***
		(0.0099)
$\boldsymbol{Post_{it}}$	-0.0545***	-0.0710***
	(0.0056)	(0.0117)
$\boldsymbol{lnU_{ait-1}}$	0.7410***	0.7413***
	(0.0144)	(0.0155)
$\boldsymbol{lnU_{a'it-1}}$	0.1418***	0.1421***
	(0.0167)	(0.0227)
Observations	48,351	48,351
District*Age FE	YES	YES
Age*Time FE	YES	YES
F Test		17.8300
p-value		0.0000

Notes: The above table is limited to the JSA due to data availability. The first column estimates equation (1) and column (2) estimates equation (3) with the dependent variable as the log of the outflow at the office level from JSA during a year-quarter. All regressions control for district, time and age fixed effects. The sample is a panel of 695 offices districts from 1999Q3 to 2008Q2. The last row contains the p-value of the F test for the joint significance of the pre-treatment dummies. Standard errors are clustered at the district level. U_{ait-1} is the lagged stock of individuals on benefits in the same age group (and $U_{a'it-1}$ the same for the other age groups). Time effects are a separate dummy for each quarter by year pair. *** p<0.01, ** p<0.05, * p<0.1

Table A3: Treatment effect in Individual waves (dropping previous waves)

WAVE	1	2	3	4	5	6	Stacked
	Panel A: Log(Total Outflow) from JSA						
$\boldsymbol{Post_{it}}$	-0.0196	-0.0314	-0.024	-0.0234	-0.0373	-0.0323	-0.0333
	(0.0135)	(0.0200)	(0.0129)	(0.0183)	(0.0129)	(0.0185)	(0.0073)
$\boldsymbol{lnU_{ait-1}}$	0.3253	0.341	0.3688	0.388	0.3924	0.387	0.3816
	(0.0182)	(0.0199)	(0.0160)	(0.0167)	(0.0167)	(0.0166)	(0.0159)
$\boldsymbol{lnU_{a'it-1}}$	0.3373	0.2838	0.2582	0.2219	0.2127	0.205	0.2182
	(0.0195)	(0.0197)	(0.0174)	(0.0188)	(0.0194)	(0.0202)	(0.0181)
Observations	9727	11943	13848	14889	15801	15817	20216
	Panel B: Log(Total Outflow) from IB						
$\boldsymbol{Post_{it}}$	0.0114	-0.0336	-0.0015	-0.0073	-0.0052	-0.0288	-0.0118
	(0.0279)	(0.0376)	(0.0245)	(0.0202)	(0.0228)	(0.0213)	(0.0137)
$\boldsymbol{lnU_{ait-1}}$	0.2221	0.1464	0.1601	0.1575	0.132	0.1214	0.1593
	(0.0493)	(0.0419)	(0.0384)	(0.0361)	(0.0364)	(0.0354)	(0.0331)
$\boldsymbol{lnU_{a'it-1}}$	1.1741	0.5537	0.356	0.4087	0.4283	0.4151	0.3894
	(0.2219)	(0.1658)	(0.1341)	(0.1219)	(0.1203)	(0.1072)	(0.1055)
Observations	7635	9556	11209	12016	12594	12331	16533

Notes: The dependent variable is the log of the outflow from benefit during a year-quarter. All regressions control for interacted district-age and interacted age-time fixed effects. The regressions restrict the post-treatment period to 4 quarters after each wave and drop districts treated in previous waves. In the upper panel the dependent variable is ln (outflow) for JSA. In the lower panel, the dependent variable is the ln (outflow) from IB. Standard errors are clustered at the district level. U_{ait-1} is the lagged stock of individuals on benefits in the same age group (and $U_{a'it-1}$ the same for the other age groups). Time effects are a separate dummy for each quarter by year pair. *** p<0.01, ** p<0.05, * p<0.1

Table A4: Specification test for dynamic structure

	Log(Total Outflow from Benefit)					
	(1)	(2)	(3)	(4)	(5)	(6)
	JSA	JSA	JSA	IB	IB	IB
$\boldsymbol{Post_1}$	-0.0274***	-0.0260***	-0.0238***	-0.0209*	-0.0190*	-0.0144
	(0.0065)	(0.0064)	(0.0063)	(0.0114)	(0.0114)	(0.0115)
$\boldsymbol{Post_2}$	-0.0282***	-0.0265***	-0.0240***	-0.0041	-0.0019	0.0033
	(0.0072)	(0.0071)	(0.0069)	(0.0115)	(0.0115)	(0.0115)
$\boldsymbol{Post_3}$	-0.0308***	-0.0289***	-0.0261***	-0.0027	-0.0002	0.0058
	(0.0067)	(0.0066)	(0.0065)	(0.0131)	(0.0132)	(0.0133)
$\boldsymbol{Post_4}$		-0.0255***	-0.0222***		0.0079	0.0147
		(0.0065)	(0.0063)		(0.0121)	(0.0123)
$\boldsymbol{Post_5}$			-0.0182***			0.0011
			(0.0068)			(0.0138)
$\boldsymbol{Post_6}$			-0.0085			0.0286**
			(0.0065)			(0.0134)
$\boldsymbol{Post_{4LR}}$	-0.0044			0.0353***		
	(0.0066)			(0.0105)		
$\boldsymbol{Post_{5LR}}$		0.0001			0.0413***	
		(0.0073)			(0.0113)	
$\boldsymbol{Post_{7LR}}$			0.0078			0.0576***
			(0.0089)			(0.0127)
$\boldsymbol{lnU_{ait-1}}$	0.6474***	0.6473***	0.6469***	0.1511***	0.1521***	0.1547***
	(0.0084)	(0.0084)	(0.0084)	(0.0180)	(0.0179)	(0.0177)
$\boldsymbol{lnU_{a'it-1}}$	-0.0013	-0.0015	-0.0019	0.2198***	0.2209***	0.2239***
	(0.0085)	(0.0085)	(0.0085)	(0.0184)	(0.0184)	(0.0183)
Observations	30,788	30,788	30,788	28,074	28,074	28,074
District*Age FE	YES	YES	YES	YES	YES	YES
Age*Time FE	YES	YES	YES	YES	YES	YES
P-value	0.0000	0.0000	0.0000	0.0003	0.0003	0.0000

Notes: The dependent variable is the log of the outflow from benefit during a year-quarter. All regressions control for district, time and age fixed effects. The sample is a panel of 406 districts from 1999Q3 to 2008Q2. In columns 1, 2 and 3, the dependent variable is ln (outflow) for JSA. In columns 4, 5 and 6, the dependent variable is ln (outflow) for IB. The last row contains the p-value of the F test for the joint significance of the post-treatment dummies. Standard errors are clustered at the district level. *** p<0.01, ** p<0.05, * p<0.1

Table A5: Robustness to Weighting by District Level Benefit Caseload

	Log(Total Outflow from Benefit)			
	(1)	(2)	(3)	(4)
	JSA	JSA	IB	IB
$\boldsymbol{Post_{it}}$	-0.0251***		0.0091	
	(0.0060)		(0.0088)	
$\boldsymbol{Post_1}$		-0.0288***		-0.0074
		(0.0077)		(0.0089)
$\boldsymbol{Post_2}$		-0.0317***		0.0029
		(0.0079)		(0.0116)
$\boldsymbol{Post_3}$		-0.0340***		-0.0096
		(0.0080)		(0.0107)
$\boldsymbol{Post_4}$		-0.0280***		0.0188
		(0.0067)		(0.0132)
$\boldsymbol{Post_5}$		-0.0283***		-0.0100
		(0.0065)		(0.0128)
$\boldsymbol{Post_6}$		-0.0203***		0.0316**
		(0.0067)		(0.0140)
$\boldsymbol{Post_7}$		-0.0243**		0.0240
		(0.0095)		(0.0167)
$\boldsymbol{Post_{8LR}}$		0.0031		0.0592***
		(0.0087)		(0.0148)
$\boldsymbol{lnU_{ait-1}}$	0.5940***	0.5898***	0.4055***	0.3950***
	(0.0144)	(0.0139)	(0.0522)	(0.0502)
$\boldsymbol{lnU_{a'it-1}}$	0.0560***	0.0561***	0.0106	0.0384
	(0.0135)	(0.0135)	(0.0265)	(0.0239)
Observations	29,159	29,159	26,450	26,450
District*Age FE	YES	YES	YES	YES
Age*Time FE	YES	YES	YES	YES
P-value		0.0002		0.0000

Notes: Columns 1&3 estimate equation (1) and columns 2&4 estimate equation (2), both with the dependent variable as the log of the outflow from benefit during a year-quarter. All regressions are weighted by the particular benefit caseload in the district-age group in 1999Q3 (prior to treatment). All estimations control for district, time and age fixed effects. The sample is a panel of 406 districts from 1999Q3 to 2007Q4. In columns 1 and 2 the dependent variable is ln (outflow) for JSA. In columns 3 and 4, the dependent variable is ln (outflow) for IB. The last row contains the p-value of the F test for the joint significance of the post-treatment dummies. Standard errors are clustered at the district level. U_{ait-1} is the lagged stock of individuals on benefits in the same age group (and $U_{a'it-1}$ the same for the other age groups). Time effects are a separate dummy for each quarter by year pair. *** p<0.01, ** p<0.05, * p<0.1

Table A6: Robustness to Different Age Groups (18-54 year olds instead of 18-59)

	Log(Total Outflow from Benefit)			
	(1)	(2)	(3)	(4)
	JSA	JSA	IB	IB
$\boldsymbol{Post_{it}}$	-0.0155***		0.0143*	
	(0.0056)		(0.0087)	
$\boldsymbol{Post_1}$		-0.0223***		-0.0109
		(0.0062)		(0.0118)
$\boldsymbol{Post_2}$		-0.0241***		0.0081
		(0.0070)		(0.0119)
$\boldsymbol{Post_3}$		-0.0265***		0.0068
		(0.0064)		(0.0134)
$\boldsymbol{Post_4}$		-0.0210***		0.0177
		(0.0062)		(0.0125)
$\boldsymbol{Post_5}$		-0.0166**		0.0044
		(0.0070)		(0.0140)
$\boldsymbol{Post_6}$		-0.0069		0.0276*
		(0.0067)		(0.0141)
$\boldsymbol{Post_7}$		-0.0078		0.0434***
		(0.0100)		(0.0155)
$\boldsymbol{Post_{8LR}}$		0.0123		0.0515***
		(0.0106)		(0.0176)
$\boldsymbol{lnU_{ait-1}}$	0.6303***	0.6291***	0.3474***	0.3554***
	(0.0101)	(0.0100)	(0.0314)	(0.0309)
$\boldsymbol{lnU_{a'it-1}}$	0.0112	0.0099	0.0502*	0.0581**
	(0.0098)	(0.0097)	(0.0256)	(0.0256)
Observations	29,159	29,159	26,450	26,450
District*Age FE	YES	YES	YES	YES
Age*Time FE	YES	YES	YES	YES
P-value		0.0000		0.0094

Notes: Columns 1&3 estimate equation (1) and columns 2&4 estimate equation (2), both with the dependent variable as the log of the outflow from benefit during a year-quarter. The age groups considered in these regressions are young (18-24) and old (25-54) as opposed to the definition of the older group as 25-59 year olds in all previous tables. All estimations control for district, time and age fixed effects. The sample is a panel of 406 districts from 1999Q3 to 2007Q4. In columns 1 and 2 the dependent variable is ln (outflow) for JSA. In columns 3 and 4, the dependent variable is ln (outflow) for IB. The last row contains the p-value of the F test for the joint significance of the post-treatment dummies. Standard errors are clustered at the district level. U_{ait-1} is the lagged stock of individuals on benefits in the same age group (and $U_{a'it-1}$ the same for the other age groups). Time effects are a separate dummy for each quarter by year pair. *** p<0.01, ** p<0.05, * p<0.1

Table A7: Treatment effects on Monthly Outflows from JSA

Log(Total Outflow)		
	(1)	(2)
	JSA	JSA
$\boldsymbol{Post_{it}}$	-0.0162***	
	(0.0056)	
$\boldsymbol{Post_1}$		-0.0251***
		(0.0067)
$\boldsymbol{Post_2}$		-0.0205***
		(0.0072)
$\boldsymbol{Post_3}$		-0.0215***
		(0.0068)
$\boldsymbol{Post_4}$		-0.0187***
		(0.0063)
$\boldsymbol{Post_5}$		-0.0217***
		(0.0076)
$\boldsymbol{Post_6}$		-0.0064
		(0.0070)
$\boldsymbol{Post_7}$		-0.0061
		(0.0088)
$\boldsymbol{Post_{8LR}}$		0.0067
		(0.0097)
$\boldsymbol{lnU_{ait-1}}$	0.7099***	0.7090***
	(0.0108)	(0.0109)
$\boldsymbol{lnU_{a'it-1}}$	-0.0620***	-0.0629***
	(0.0103)	(0.0102)
Observations	84,202	84,202
District*Age FE	YES	YES
Age*Time FE	YES	YES

Notes: Column 1 estimates equation (1) and column 2 estimates equation (2), both with the dependent variable as the log of the outflow from benefit during a year-month. All regressions control for district, time and age fixed effects. The sample is a panel of 406 districts from January 1999 to December 2008. The dependent variable is ln (outflow) for JSA. Standard errors are clustered at the district level. U_{ait-1} is the lagged stock of individuals on benefits in the same age group (and $U_{a'it-1}$ the same for the other age groups). Time effects are a separate dummy for each quarter by year pair. *** p<0.01, ** p<0.05, * p<0.1

Appendix A: Data Description

The empirics are based on three primary sources of data. First, the design of the policy and the list of districts covered under each wave of the rollout was provided by the Department of Work and Pensions. Second, data on other welfare benefits including the key disability benefit, Incapacity Benefit (IB) was sourced directly from the Department of Work and Pensions Tabulation Tool – www.tabulation-tool.dwp.gov.uk/WorkProg/tabtool.html. This provides only quarterly data on the stocks, inflows and outflows of benefit recipients. The data is available for Great Britain (i.e. including England, Wales and Scotland) at the Local Authority Districts level from 1999 Q3 onwards only. The 4 quarters in the dataset are defined as February-April, May-July, August-October and November-January. The data can be disaggregated by age and duration.

The second dataset, the Job Seeker's Allowance database was downloaded from www.nomisweb.co.uk. It provides monthly information from 1983 on the stocks, inflows and outflows of recipients' unemployment benefits. The data is available at various geographical levels. We use the data at the Local Authority Districts across Great Britain and there are 406 districts defined on a consistent basis. The data can be disaggregated by age, duration as well as both.

To be consistent across the two datasets, the monthly JSA information is aggregated to the quarterly level. The quarters are defined as February-April, May-July, August-October and November-January. In order to create a quarterly dataset using the monthly information, the flows of each month in a quarter were added up to get the total flows in a quarter. The stocks at the end of a quarter are measured as the stock in the last month of the quarter.

In addition, we use the quarterly micro individual-level panel of the UK Labour Force Survey from 1998 to 2008 to get estimates on the household composition of benefit claimants, mean wages, origins of benefit inflows and destination of benefit leavers. The data is securely provided by the UK Data Service under Special Access License. The quartiles of the weekly earnings distribution were taken from the 2000 New Earnings Survey.

We used digitized boundary datasets and geographic look-up tables corresponding to the census geography of Great Britain, provided by the UK Data Service. We used the boundary data in ArcGIS to illustrate the policy rollout and to define the neighbours of districts.

Finally, the IFS had generously provided benefits estimates using the IFS tax and benefit micro-simulation model, TAXBEN. In order to estimate the net exchequer cost of benefit claimants, their estimates assumed that the house rent is £44 per week (the average among families receiving income support, jobseeker's allowance or incapacity benefit) and that all disposable income is spent on items subject to the standard rate of VAT and no excise duties.

Appendix B

B1. Steady-state change in the IB rate for the Cost-benefit analysis

In the cost-benefit analysis we have to consider translating our estimates of flow changes into changes in unemployment and welfare stocks. We obtain the steady state change in the IB rate, based on permanent changes in the IB outflow rate following the introduction of Jobcentre Plus.

Assume there are only two states, employment and IB, and denote by s the inflow rate from employment into IB, and by f the outflow rate from IB into employment. In steady state the IB rate is constant, and flow equilibrium implies that the IB rate (as a fraction of the total population) is given by:

$$u = \frac{s}{s+f}$$

The policy has an impact on f, leaving s unaffected. The resulting change in the (log) IB rate is given by

$$d \ln u = -(1-u) d \ln f.$$

The implied change in the number of jobs in steady state is given by:

$$\Delta e = -u \, d \ln u = u(1-u) d \ln f \qquad \text{(B1)}$$

According to our estimates, $d\ln f = \beta - (1-\alpha)d\ln u$, where β is the treatment effect estimated by diffs-in-diff, and α is the coefficient on the log IB stock. The terms in u on the right-hand side of (B1) are evaluated using the actual IB rate in the pre-policy period.

B2. Off steady-state approximation

At each point in time the unemployment rate evolves according to

$$\frac{du_t}{dt} = s_t(1-u_t) - f_t u_t. \qquad \text{(B2)}$$

Solving (B2) forward one period gives:

$$u_t = \gamma_t u_t^* + (1-\gamma_t)u_{t-1}, \qquad \text{(B3)}$$

where u_t^* denotes steady state unemployment and γ_t denotes the rate of convergence to it:

$$\gamma_t = 1 - \exp(f_t + s_t).$$

Using a log-linear approximation to (B3) it can be shown that:

$$d\ln u_t = -\gamma_{t-1}(1-u_{t-1}^*)d\ln f_{t,}$$

where, as above, $d\ln f = \beta - (1-\alpha)d\ln u$, u is evaluated using the actual IB rate in the previous quarter and u^* is evaluated using the (constant) pre-policy inflow rate into benefits obtained from the Labour Force Survey and the time varying outflow rate from benefits as obtained from our estimates.

While the steady-state result stated above is only used for IB predictions (as the steady-state impact of policy on the JSA outflow is close to zero), the off-steady state results are used to obtain predictions for both the IB and JSA rate during the transition to a new steady state.

Chapter 3: Team Incentives in the Public Sector

Abstract: This chapter explores the impact of introducing team based rewards in a public sector organization. In my context, the new rewards switch from being individually based to being based in a team at the district level. Districts are organized into spatially dispersed sub-teams (i.e. offices within districts). In this setting, I explore whether it is really the case that incentives are weakened when rewards are based on team output rather than individual output. I find compelling evidence that this indeed occurs. With the introduction of team rewards, employees are more likely to free-ride on each other's efforts in large teams. I find support for team size determining the degree to which incentives are weakened. Peer monitoring, may, however, limit free-riding in teams. I formalize the impact on output under two benchmark models to ascertain the relative impact of peer monitoring and free-riding. To investigate the model's predictions empirically, I exploit the national rollout of an incentive structure in the UK public employment service (Jobcentre Plus) in the mid-2000s. Using difference-in-differences estimators, I find that the team based measures dilute incentives. Consistent with a degree of peer monitoring, however, the dilution effect is smaller when peer monitoring is easier (as captured by the concentration of teams across offices within a district).

1 Introduction

The economic theory of team incentives states that the effectiveness of group rewards depends on the type of the organization where it's implemented and the characteristics of its production process (Dixit, 2002). The measurability of performance, the size of the team, the multi-dimensional nature of tasks are all elements that need to be considered in designing the optimal incentive structure. This chapter contributes to the literature by specifically examining team incentives in a public sector organization with a unique production process. The production in the organization is divided among teams. Each team is further comprised of sub-groups (or sub-teams) and each sub-group is in a different geographical location. The team dynamics within and across sub-teams is explicitly modelled. The production process is similar to a decentralized manufacturing process and thus is likely to have wider relevance.

In the empirics, I study the incentives in Jobcentres in the UK. Jobcentres deliver active labor market programs in the UK and are part of the Department of Work and Pensions. The main "output" of the Jobcentres is job placements of benefit claimants. I investigate the effects of redesigning performance measurement from individual based to team based achievements in the Jobcentres. In January 2005, "Job Outcome Targets" were introduced. The introduction of the Job Outcome Targets (JOT) was a major shift in the existing explicit performance standards structure. Since 2005, the JOT scheme rewarded benefit officers according to job placements achieved by their *team* which is the district, in contrast to being previously based on their individual placement record. The unique feature of this scheme is that while the team was defined at the district level, each district had spatially dispersed offices in which employees worked.

Under the new scheme, employees could potentially free-ride and the new incentive scheme would dilute pre-existing incentives. However, members working in the same location/office could also monitor each other. Such peer pressure would counterbalance the impact of free-riding. I formalize the degree to which peer monitoring affects the impact of JOT on output under two benchmark models of perfect monitoring and no peer monitoring, to ascertain the relative impact of monitoring and free-riding.

For identification, I exploit the staggered roll-out of Job Outcome Targets across UK districts and offices. I identify the policy impact by comparing the change in exit rates for benefit claimants in districts treated at a point in time to that in districts treated at an earlier or later date. Information on benefit claimants at the district and office level is provided by the Department for Work and Pensions.

Empirically, I find support for the presence of peer monitoring in teams. I find three main results. First, and most striking, the shift from individual to team based incentives does dilute pre-existing incentives. This suggests that front line staff free-ride on their team members' efforts. Second, the degree to which incentives are unraveled depends on the size of the district (aka the team size). The larger the number of district employees, the more the initial policy is diluted. Lastly, conditional on a district of a given team size, the dilution is absent in districts with employees concentrated in a few large offices. This effect of the concentration of employees is the result of peer monitoring among employees.

This chapter is related to several strands of literature. First, the literature on team incentives is ambiguous as to the net effect of introducing team contracts. Holmstrom's (1982) seminal contribution shows that when there are complementarities in production and if all the output of the team is shared among team members, team members are induced to free-ride. Alchian and Demsetz (1972) also show that employees have more incentive to shirk while working as part of a team since only a small percentage of the losses from shirking is borne by them.

However, there are contexts in which individual incentives could lead to perverse results compared to team incentives. Individuals have less incentive to share common working knowledge in the absence of team incentives and reduce productivity. Teamwork can also help reduce individuals' exposure to the risk of poor outcomes beyond their control by pooling risk at the team level and as such performance will be better under team contracts. Further, as Kandel and Lazear (1992) argue, team based incentives facilitate informal peer monitoring and pressure from colleagues to perform. When reward contracts are individual based, this type of external monitoring is less likely.

This chapter extends this literature by incorporating the scenario where a team comprises of sub-teams and monitoring is complex across sub-teams. I build a simple model to disentangle the effect of free-riding and peer monitoring on output. Using a simple functional form, I show how the marginal effect of the team based incentives on output varies by whether or not I allow for the presence of peer monitoring.

Secondly, this chapter relates closely with empirical papers that have studied agency issues in groups such as Burgess et al (2010). They find that even in quite large teams, a team based incentive scheme in the UK Customs and Excise can raise the productivity of agency workers. In addition to free-riding, Encinosa, Gaynor and Rebitzer (2007) highlight how employees engage in non-contractual and informal interactions while working in groups. The presence of this peer monitoring counterbalances the free-riding effect. Hamilton, Nickerson and Owan (2003) also find support for mutual team learning that outweighs free-riding. This chapter addresses a gap in understanding public sector employee behaviour - How do nuanced features of the production process affect the answer to whether it is always the case that incentives are weakened when rewards are based on team output rather than individual output?

Finally, this chapter contributes to a growing empirical literature that has highlighted how incentive systems in the public sector have aimed to align staff behaviour to organizational goals and to achieve higher efficiency in employee performance (see Bloom and Van Reenen, 2011, for a survey). Heckman et al (2011) investigate incentive structures in the US that reward officers based on measured outcomes to find that these systems lead to cream-skimming of program participants. One mechanism for these impacts has been front line staff altering their behaviour in response to incentive structures and performance measures. An incentives system could lead front line staff in government agencies to re-allocate resources which may be socially suboptimal (Bagaria et al, 2014).

In section 2, I describe the institutional framework in the UK. In sections 3 and 4, I discuss the hypotheses to test and the empirical approach. I outline the data used in the empirical analysis in section 5. I report the results of how the treatment impacts outflow from different benefit categories in section 6. I examine the robustness of my results to different specifications in section 7 and conclude in section 8.

2 Institutional Framework

The UK has, historically, introduced performance benchmarking and linking targets to resources allocation since early 1980s (Propper and Wilson 2003; Bagaria et al, 2014). They had been designed to relate to targets embodied in the Public Service Arrangements (PSAs) of different government agencies[22]. During 2001-2008, there were major infrastructure changes in the delivery of public employment and benefits services. The change was part of the policy emphasis towards welfare-to-work initiatives[23] that sought to make sure that those on unemployment benefits were active in the labor market. To ensure that they were actively seeking job opportunities and had not dropped out of the labor market, the government introduced conditionalities to benefit payments since the mid-1900s.

As part of the organizational restructuring, since October 2001, the public job placement services (formerly run by the Employment Service) and benefit services (run by the Benefits Agency) were integrated into one single organization called the Jobcentre Plus. It became the key provider of job placement services and benefit advice in the UK. Final output in Jobcentre Plus consisted of putting people into jobs.

A Jobcentre office provides support for people of working age in the UK by helping them in their job search and in claiming for benefits. A wide variety of customers approach Jobcentres, with different needs: young people, lone parents, disabled people. Different programs and activities have to be undertaken in order to match their different needs. Jobcentre staff are engaged in multiple activities, which are very difficult to measure. There is not a single indicator of overall performance comparable to profit in a private company. Assessing performance is thus complicated and requires that all the different activities undertaken are taken into consideration. In addition to this multi-tasking feature, lags in information availability can also exacerbate output measurement.

[22] Makinson (2000) describes the performance standards in the Employment Service, The Benefits Agency, HM Customs and Excise, and, Inland Revenue.

[23] The guiding principle of welfare reform was “work for those who can and security for those who cannot” - DWP.

A striking change, as part of the Jobcentre Plus package in 2001, was the introduction of explicit performance targets called Job Entry Targets. Since October 2001, Jobcentre Plus officers faced a regime called Job Entry Targets in which every benefit officer who helped a benefit claimant into a job was awarded a certain number of explicit Job Entry Target points. As opposed to the previous system of national-level targets for the number of job placements, under the regime of Job Entry Targets, every bureaucrat who placed a benefit claimant into a job was awarded a certain number of explicit Job Entry Target (JET) points varying by the category of the benefit claimant. They were designed to reflect organizational priorities towards the disabled i.e. Incapacity Benefit claimants against the unemployed Jobseeker Allowance claimants. These performance standards acted like a performance benchmark for the managers and mattered for the career prospects of the officers (Bagaria et al, 2014).

Thereafter, in January 2005, Job Outcome Targets (JOT) was introduced to replace the JET (see Figure 2). JOT marked a clear shift in Jobcentre Plus' approach to performance management and was used by managers to monitor and reward performance. Most importantly, as opposed to the existing JET points which were attributed to individual staff members on a daily basis, under JOT reward points were only measurable at the aggregated team level and with a lag of up to six months[24]. Reward points could no longer be attributed to individual staff. With free-riding at play, employees had lower incentives to help the priority group of disabled/incapacity claimants and higher incentives to free-ride by helping the easy-to-help Jobseekers Allowance claimants.

Teams were defined at the district level. Each district comprised of multiple offices in different locations with different catchment areas. Figure 1 illustrates the structure of the Jobcentres. At the helm was the Department of Work and Pensions that set the mandates and targets of the Jobcentres. Below them, Jobcentres were operated by districts. These districts were the effective "team" for JOT reporting purposes. Each team comprised of multiple sub-teams or offices in different locations and of different

[24] The DWP Benefits system was now linked to the tax administration, HMRC, databases. This resulted in a lag of six months in obtaining information on benefit outflows to work due to the six month window of setting up a tax account upon entering employment.

sizes. Finally, at the lowest level, employees worked in a particular office. While the effort of all the employees in a district mattered for team performance, the office manager as well as any employee could only monitor employees and peers in his/her own office. For example, the office manager is more aligned to the district manager and has greater ability to monitor his/her own employees. Monitoring across offices was unfeasible. It is this feature of the production process that calls for a new theoretical model.

The JOT represented a stark shift in performance measurement and management. The relationship between inputs and outcomes became less clear due to the team aggregation as well as the information lags. It is important to highlight that the aggregation was not a matter of administrative reporting alone. For instance, it was not the case that individual points were still available on the system but reported to the district managers at the team level. The information system changed such that individual level data no longer existed. Thus, while my identification strategy relies on a quasi-experiment, this change in the very existence of information allays concerns of information manipulation or informal information transfers.

Anecdotal evidence[25] suggests that the JOT impacted employee motivation and behaviour. One District Manager described the situation as “It’s really hard with regards to JOT, we haven’t got a clue; we used to know with JET because you had your daily placing list which detailed each placement inputted the day before and you knew how many points each customer is worth and so you knew where you were in terms of your target. [..] As a manager trying to manage performance it’s really difficult.”

3 Model

Based on Holmstrom’s model (1982), the baseline marginal effect of moving from individual based to team based incentives in Jobcentre Plus offices is that pre-existing incentives will be diluted after the introduction of team based incentives. This effect will, in turn, depend on the team size as well as the degree of monitoring taking place.

[25] Based on interviews in “Qualitative assessment of Jobcentre Plus Delivery of Jobseeker’s Allowance and New Deal Interventions”, DWP Research Report 445.

Theoretically, I distinguish between two opposing factors determining the effect of moving from individual based to team based incentives in Jobcentre Plus. Firstly, team based incentives increase the probability of free-riding in a team environment. In a setting where team members depend on each other to produce final output, if all the output of the team is shared among team members, team members are induced to free-ride. Larger the number of team members, greater is the incentive to free-ride.

However, counteracting this is a facilitation of communication and peer monitoring in teams. If members of a team work in the same location and the organization of work is such that they are able to observe each other and if their reward is linked to the team performance, they are more inclined to monitor how their peers are performing. This can help in enforcing proper levels of effort and tackle the free-rider problem. The effect would be to reinforce the baseline incentives and not dilute incentives in an environment where peers can monitor each other. It is akin to an internalization of incentives. This counterbalances the free-riding effect.

The opposing forces described above will determine the marginal effect of JOT. To disentangle the two effects described above, I model employee behaviour formally in the next subsection. I lay out the theoretical predictions to test in section 3.2 and I describe the empirical strategy in section 4.

3.1 Theoretical Model

I construct two models based on the extreme assumptions of no peer monitoring and perfect monitoring. In the first model, I assume that there is no peer monitoring in teams. This is the simply the Holmstrom (1982) model. Each agent works independently. In the second model, I assume there is perfect monitoring among workers in the same location i.e. same office, but not across offices within a district. The idea is similar to Mas and Morreti's (2009) results where they find peer effects only among workers who can see each other even within the same supermarket. Under the extreme assumption of perfect monitoring within an office, the model describes each office to be acting like a single agent.

In reality, the truth will sit somewhere in between the two extremes. While there is likely to be peer monitoring only within offices, not across, the effectiveness of monitoring is unlikely to be the same in all offices. Larger offices may have scale economies of monitoring such that monitoring is stronger in large offices. The contrary could also be true – large offices may face increasing costs of monitoring. Thus, the assumption of equal and perfect monitoring in offices is only a theoretical benchmark. I describe the two models below along with their theoretical predictions.

3.1.1 Model A: No Peer Monitoring

The first model is intuitively based on the incentive to free-ride in teams and is exactly Holmstrom (1982)'s $\frac{1}{N}$ problem. Each employee i acts like a single agent when choosing effort e_i. As Holmstrom showed, in this benchmark case, the private marginal benefit $(\frac{Y}{N})$ will be less than the social marginal benefit (Y) since the reward is shared among all members of the team.

Thus, the level of effort chosen by the individual will be lower than the Pareto efficient level. The chosen effort will depend inversely on the number of reward-sharers i.e. the total team size, $\frac{1}{N}$, where N is comprised of sub-teams of size $N_1, N_2 \ldots N_s$ such that $N = N_1 + N_2 \ldots + N_s$. Total team output also, thus, only depends on total team size. Larger teams have lower output due to a higher incentive to free-ride. All employees in a district choose the same level of effort, irrespective of the size of the office they work in. There is no peer monitoring at all.

Empirically, this model implies that the total team size i.e. the number of employees in a district matters in determining the impact of JOT on output (job placements). The JOT policy effect will vary by district size measured in terms of number of employees. The larger a district, the stronger are incentives to free-ride and hence greater dilution of incentives in the event of introduction of team incentives. The distribution of teams across sub-teams does not matter in this case. And, the size of one's immediate office also does not matter in determining effort.

3.1.2 Perfect Peer Monitoring

At the other end of the spectrum, I assume there is perfect and costless peer monitoring within an office. Monitoring is easier within an office than between offices. Intuitively, imagine there is an office-specific monitoring fixed cost. This upfront fixed cost of monitoring another office prevents employees from monitoring across offices. This fixed cost could simply be the telephony technology of contacting and monitoring another office or the transport cost of being physically present to monitor. This structure of monitoring implies that the employees in any office choose effort as if an office-level social planner was maximizing at the office level. Each office in a district, thus, acts like a single agent.

For simplicity, I assume a standard concave production function in the district (or team) i.e.

$$f(e) = Ae^{\alpha}$$

where e is the effort chosen by team members, $0 < \alpha < 1$ and A is a positive constant representing technology. The only input into the production function is the effort of the N employees in the team; $e = \sum_{j=1}^{N} e_j$. There is a cost to putting in effort. The cost is represented by a quadratic function: $c(e_j) = \frac{e_j^2}{2}$.

Consider a district with S benefit offices in its jurisdiction. Each office has a different number of employees, given by $N_1, N_2 \ldots . N_s$. The total number of employees across the S offices is N i.e. $N = N_1 + N_2 \ldots . + N_s$. In other words, the total team size in the district is N.

The equilibrium effort of an employee in office k with N_k employees is a solution to the office social planner's maximization problem given by (see Appendix B for details):

$$max_{e_{k1}, e_{k2} \ldots e_{kN_k}} \frac{N_k}{N} A \left(\sum_{i=1}^{N_k} e_{ki} + \sum_{N_t, t=1, t \neq k}^{S} \sum_{i=1}^{N_t} e_{ti} \right)^{\alpha} - \sum_{i=1}^{N_k} \frac{e_{ki}^2}{2}$$

The office planner's output is the proportion of total output that is allocated to his/her whole office, not to its individual employees. Each employee in an office makes a symmetric choice of effort in equilibrium. Finally, the planner faces a cost equal to the total cost of effort of all its employees.

The solution to the above problem (see Appendix B for details) is an effort level chosen by all employees in an office k given by:

$$e_{ki}^{*} = N_k \left[\frac{A\alpha}{N}\right]^{\frac{1}{2-\alpha}} (N_1^2 + N_2^2 + \cdots . + N_s^2)^{\frac{\alpha-1}{2-\alpha}}$$

And, the equilibrium total output in the district is:

$$f(e^{*}) = \left[\frac{N_1^2 + N_2^2 + \cdots . + N_S^2}{N}\right]^{\frac{\alpha}{2-\alpha}} A^{\frac{2}{2-\alpha}} \alpha^{\frac{\alpha}{2-\alpha}}$$

It depends positively on A, α and the ratio $\frac{N_1^2+N_2^2+\cdots.+N_S^2}{N}$. This ratio is nothing but a measure of the concentration of office sizes in the district. The ratio $\frac{N_1^2+N_2^2+\cdots.+N_S^2}{N} = N * HI$ where *HI* is the Herfindahl index of office size concentration in the district.

This suggests that output in a district is positively correlated with the concentration of office sizes. In fact, $\frac{\partial f(e)}{dN_k} \geq 0$ if $N_k \geq \frac{N}{S}$. In other words, conditional on a district team size, the marginal effect of JOT decreases if the employees in a big office are increased, rather than being evenly spread across all the *S* offices. (See Appendix B for solution with one, two and multiple offices).

Intuitively, in this model, the office social planner is the decision-maker and hence, an office completely internalizes the negative externality of free-riding within an office by implementing perfect monitoring. Thus, the more concentrated a district, the lower is the dilution of incentives.

Figure 1 helps illustrate how the concentration of employees matters using the case of two districts. Both districts depicted in the figure have the same average size $\mu = 4$. However, Team A (or District A) has two identical sized sub-teams (or offices), each with 4 employees. Team B (or District B) has one large sub-team (or office) with 7 employees and two small sub-teams (or offices), such that the mean team size is still 4 employees. According to the perfect peer monitoring model, the equilibrium output will be higher in the case of Team B with its unequally distributed sub-teams. The workings are also shown in Appendix B.2.

3.2 Hypotheses

The first hypothesis, based on the prediction of the effects of moving from individual based to team based incentives in Jobcentre Plus offices, is that pre-existing incentives will be diluted after the introduction of team based incentives (JOT). As highlighted, this effect will depend on both the team size and the degree of peer monitoring.

To estimate the degree to which free-riding and peer monitoring affect the marginal impact of JOT, the next two hypotheses are based on the models that allow for the presence and absence of peer monitoring. The second hypothesis which is based on the model of pure free-riding predicts that the total size of the team will drive a differential impact of the JOT policy. Districts with a greater number of total employees will face larger incentive dilution than districts with a small number of employees. The baseline effect of JOT will, thus, be reinforced in districts with more employees.

The third hypothesis, based on the assumption of perfect and costless monitoring within an office, is that the distribution of employees in a team across the sub-teams matters for JOT's impact on output. As the model shows, a district with a higher Herfindahl index of team concentration will internalize the negative externality of free-riding to a greater extent. The third hypothesis, thus, tests the differential impact of the policy by team size distribution. In more concentrated teams, the pre-existing incentives will be intact.

In an earlier paper, Bagaria, Petrongolo and Van Reenen (2014) find that the introduction of JET earlier in 2001 incentivized the front line staff to focus on placing Incapacity Benefit (IB, henceforth) claimants, as was the objective of the policy. The reader can skip the next two paragraphs if they have already read Bagaria et al (2014).

The authors highlight the effect of an introduction of explicit incentives for Jobcentre employees to move certain targeted welfare recipients into jobs and the consequent unintended substitution effects due to multi-tasking when employees substitute effort away from less targeted groups. They examine the impacts before and after a major UK policy change ("Job Centre Plus") that shifted relative incentives of benefit officers to place disabled people into jobs compared with the unemployed.

The officers in charge of helping welfare claimants find jobs had changed incentives with an explicit points system (illustrated in Table A1) introduced with three times as many points given for placing a disabled person into a job than for the unemployed[26]. Points fed into career progression. Bagaria et al find that the policy significantly increased outflows from disability benefits (IB), but reduced the outflow from unemployment benefits in the short run. In the long-run, however, both groups benefit but the disabled outflows rose by more (6%) than for the unemployed (1%). These patterns are consistent with a simple model where, given a quarterly target number of points and a fixed stock of inputs, the points adversely affect the job placement efforts and outcomes for those on Job Seeker's Allowance claimants. Thus, JET is likely to have a larger impact on IB recipients than JSA recipients. Note that JOT was introduced in offices that were already treated with JET.

In this chapter, I find similar policy effects of JET in section 6.1. My first hypothesis can thus be restated as –the introduction of JOT should have a positive impact on JSA and a negative effect on IB claimants. The second hypothesis is that the effect on JSA will increase as team size increases and vice versa for IB. Finally, the third hypothesis states that the JOT effect on JSA will decrease as team concentration increases and vice versa for IB.

4 Empirical Strategy

After describing the empirical strategy used to test the hypothesis in this section, I describe the data sources in section 5. I summarize my findings from this analysis for job placements in section 6.

[26] As shown in Table A1, a benefit officer was awarded fifty per cent more points if he/she placed a person on Incapacity Benefit into a job than a long-term JSA beneficiary, and three times more points relative to a short-term JSA beneficiary.

4.1 Identification

I exploit the staggered roll-out of the Job Outcome Targets (JOT) across Local Authority Districts[27] across the UK to identify its causal impact. It was phased-in in 2 staggered waves. The first wave was started on 10th January, 2005 in 50 offices across 18 districts. The second wave began in April 2006 when it was introduced in the remaining offices and local authority districts. Figure 2 shows a clear timeline of the introduction of JET as well as JOT.

I use a difference-in-differences framework to identify the causal impact of the Job Outcome Targets (JOT) incentive structure. Since all districts are treated eventually, effectively I am comparing districts which are treated with JOT in a particular year and quarter to those who are treated at a later stage. Of course one concern is that districts were selected into treatment and were not randomly assigned into treatment. This would confound my causal interpretation. To examine this, I check for the robustness of my results to differences in pre-trends between the treated and non-treated, as discussed further in section 7.1.

In my empirical analysis, I measure performance outcomes by the number of exits from each benefits register each quarter in each district, for the 2 benefit registers – Incapacity Benefit (IB) and Job Seekers Allowance (JSA). I control for the past stock of benefit recipients so the analysis is effectively looking at changes in the outflow rate. To measure the impact of JOT, I estimate standard benefit outflow equations for these two key groups: IB and JSA. I disaggregate this further by age groups - young and old. The young are defined as being aged between 18 to 24 years. The older group consists of people aged 25-59 years[28].

[27] For local government, England is divided into areas with a two-tier structure of counties and districts governed by two Local Authorities, and unitary authority areas where there is one local authority. In England, local authorities are divided into single-tier and two-tier authorities. The Local Authorities that I use includes the numerous districts within counties, Greater London as well as the unitary authorities. Due to missing information on office sizes, I base my analysis on 338 Local Authority districts.

[28] The specification is robust to a re-classification of age groups into 18-24 and 25-54 years.

4.2 Job Outcome Targets and Benefit Outflows

I estimate standard benefit outflow equations in a difference-in-differences framework. I start by estimating the effect of introducing team based employee incentives in a district. The baseline specification is pooled across age groups and is given by equation (1):

$$\ln(y_{adt}) = \alpha_1 T_{dt}^{base} + \beta_1 T_{dt}^{base} * T_{dt}^{JOT} + \gamma_1 lnU_{adt-1} + \gamma_2 lnU_{a'dt-1} + \delta_d + \delta_t + \delta_a + \varepsilon_{adt} \quad (1)$$

where y_{adt} is the output measure in terms of the number of people in age group *a* leaving the benefit register in district *d* at time *t* (which is a quarter-year). I estimate equation (1) separately for each benefit category – IB and JSA.

T_{dt}^{base} denotes a treatment dummy which turns on in the quarter when the district had introduced Jobcentre Plus (JET). This effect has also been analysed in Bagaria et al (2014) and forms the pre-existing scenario. They found α_1 was positive and significant for IB claimants and significantly negative for JSA.

I compare this to the marginal effect of JOT, β_1. β_1 is identified by comparing districts which are treated with JOT, conditional on being treated with JET, in a particular time period to those who are treated at a later stage. In other words,

$$\beta_1 = \frac{\partial \ln(y_{adt})}{\partial T^{JOT}}\bigg|_{T_{dt}^{base}=1}$$

My first hypothesis relates to the dilution of existing incentives due to the introduction of team based contracts. The hypothesis translates into β_1 being in the opposite direction of α_1. In other words, β_1 should be negative for IB and positive for JSA.

I include as controls the stock of people claiming benefit in age group *a* (young/old) at the end of the previous quarter, U_{ait-1} as well as the stock of people in the other age group, *a'* (old/young respectively), claiming benefit at the end of the previous quarter. This enables me to account for substitution effects between job applicants in the two age groups in the labor market. To control for district specific demographic, economic and scale factors, I include district by age fixed effects. I control for time fixed effects which are quarter by year[29] to eliminate any time specific factors that may affect the outflow from benefits. Because I have multiple age groups per local authority district, I also cluster standard errors at the district level.

I estimate equation (1) for district level outcomes since the entire district was described as the team. The district output is, however, in essence an aggregation of outputs from individual offices. To check the robustness of the findings at the district level, I re-estimate equation (1) for JSA[30] at the office level with office fixed effects in section 6.4.

It is, however, plausible that the results are driven by district managers endogenously choosing optimal office sizes after the policy. To address this in my estimations, I fix the size of teams at the pre-policy level of June 2004. However, using annual data on the size for the years 2004 and 2007, I check whether there is any policy effect on the size of an office post policy (see Section 7.3), but do not find any effects.

5 Data

I use two sources of data. The first is a rich and detailed dataset containing information on the number of employees in every Jobcentre office and the number of offices in a local authority district. The information is available for the years 2004 and 2007 and provides rich data on local administrative units. This is proprietary data of the Department of Work and Pensions (DWP) and was obtained for research purposes.

The second database is a set of administrative datasets provided publicly by the DWP. Two different panel datasets from the Department are used. The first is the Incapacity Benefits dataset that provides panel data on quarterly stocks and flows only at the Local Authority level from 1999 quarter 3 onwards. Ideally, the outcome measure should be job placements. But, the dataset does not contain any information on the reasons for leaving the benefit register. Benefit claimants could leave a particular caseload register for various reasons. These include finding work, increasing work hours

[29] For instance, there is a control dummy for the first quarter of 2000 as well as separate dummies for the three other quarters of 2000. Similarly for 1999-2007.
[30] Information on IB outflows at office level is not made available by the DWP.

beyond 16 hours per week, claiming a different benefit, entering full time education, joining training programmes, defective claims and failure to sign. I expect Jobcentre Plus to impact the outflow to work, but, due to data limitations, I use total outflow from Incapacity Benefits, irrespective of destination.

The second panel provides information on the monthly stocks and flows for those on Job Seeker's Allowance (JSA, henceforth) at the Local Authority as well as office level. It contains information at the monthly level from 1995. To be consistent across the two datasets, I aggregate the monthly JSA data to the quarterly level. I use the monthly information as a robustness check in Table A2. Due to the limited data availability, the period of analysis is from the fourth quarter of 1999 to the last quarter of 2007 (before the Great Recession set in). I have the total outflow and total stock for each of the benefit registers and office information in 338 local authority districts and 33 quarters in the final panel.

Due to data limitations, office level analysis can only be conducted for the JSA group. Outflows data on IB is not available at a disaggregated level of the office.

5.1 Summary Statistics

The summary statistics for the outcome variable i.e. total outflow from both benefit categories- IB and JSA- are reported in Table 1 Panel A. The differences-in-differences analysis is based on data from 851 offices across 338 districts. The summary statistic of the district employment and office sizes is reported in Panel B of Table 1.

Alternatively, one can visually see the distribution of the number of offices in a district in Figure 1. While the full distribution is shown in the Appendix (Figure A1), Figure 3 plots the region below the 95th percentile of the distribution (i.e. 10 offices per district or less) to get a clearer picture without the extreme values. The mode is at 1 office per district and the average is 3.97. About 25.45% of the districts have just one office while 24% have 2 offices.

To get a better understanding of the distribution of the number of employees across the 338 districts, Figure 4 plots the distribution of employees across districts. The plot is for the number of employees below the 95th percentile i.e. 607 employees. There is considerable variation in the number of employees or the team size across districts.

Each district is made up of several offices. Figure 5 plots the histogram of the office sizes in the 851 offices, in terms of the number of employees in an office. Again, while the full distribution is shown in the Appendix (Figure A3), Figure 3 plots the region below the 95th percentile of the distribution i.e. 252 employees to get a clearer sense of the distribution. There is considerable variation in the number of employees in an office as well. The mean of the distribution is 68.38 while the median is about 40 employees.

6 Main Results

In the first section below, I present the estimated marginal impact of the JOT policy on the outflow from benefits, and then turn to examining the effect based on team size and distribution.

6.1 Baseline

In Table 2 column (1) I start by estimating the pre-existing scenario, with β_1 set to zero, where the dependent variable is the log (total outflow) from JSA with controls for JSA caseloads last period. All estimations in this table include district by age fixed effects as well as age by time period (quarter by year) fixed effects.

I find a significant coefficient of -0.0101 for the previous JET policy impact on JSA, very similar[31] to Bagaria et al (2014), suggesting that a district treated with the Jobcentre Plus experiences, on average, about a 1% decrease in total outflow from JSA caseload, compared to the non-treated districts.

In Table 2 column (2) I present the first major new result. The estimated marginal effect of introducing team incentives under JOT on the log (total outflow) from JSA is positive and highly significant. It is in the opposite direction of α_1 (which is -

[31] The estimate is exactly not the same in magnitude due to a slightly smaller sample because of missing information on office sizes.

0.118). This seems to support the first hypothesis that conditional on individual incentives under JET ($T_{dt}^{base} = 1$) which led to a fall in JSA outflows, the introduction of team incentives ($T_{dt}^{JOT} = 1$) reverses this[32]. This is consistent with the idea that team incentives dilute individual incentives.

Similarly, in column (3) of Table 2, I estimate the pre-existing scenario for the log (total outflow) from IB and find a coefficient of 0.143, though not significant[33]. Column (4) now includes the policy dummy for JOT. The estimated coefficient β_1 is now negative, just as expected from a dilution of baseline incentives (though it is not significantly different from zero). This suggests that JOT is harmful for IB claimants and a treated district experiences, on average, approx. 1% decrease in total outflow from IB as a result of introducing JOT.

Hence, in summary I find strong evidence in Table 2 that the introduction of team based rewards leads to a weakening of incentives. After the introduction of JOT, employees had higher incentives to free-ride by helping the easy-to-help JSA.

6.2 Team Size

Next, based on the no peer monitoring model's prediction that the number of employees in a district determines the marginal effect of JOT, I estimate the effect of total team size on benefit outflows. To test this hypothesis, I estimate:

$$\ln(y_{adt}) = \beta_1[T_{dt}^{base} * T_{dt}^{JOT} * \ln(S_d)] + \beta_2(T_{dt}^{base} * T_{dt}^{JOT}) + \alpha_1 T_{dt}^{base} + \alpha_2(T_{dt}^{base} * \ln(S_d)) + \gamma_1 lnU_{adt-1} + \gamma_2 lnU_{a'dt-1} + \delta_d + \delta_t + \delta_a + \varepsilon_{adt} \quad (2)$$

where S_d represents the total number of employees in district *d*. Again, I estimate equation (2) for both benefit categories – IB and JSA at the district level. The key hypothesis is that β_1, the coefficient on the triple interaction is positive for JSA and negative for IB. The idea is that the dilution effect of team incentives (compared to individual incentives) is particularly strong when district employment is larger (i.e. more

[32] The extent to which JOT unravels the effect of JET is discussed in the robustness section 7.5.
[33] The estimate is exactly not the same in magnitude due to a slightly smaller sample because of missing information on office sizes.

benefit officers, so easier free-riding). Thus, IB outflows will fall by more in large districts and JSA outflows will increase by more.

In Table 3, I estimate the coefficient on the interaction term, β_1. All estimations include district by age fixed effects as well as age by time (quarter by year) fixed effects, in addition to district, age and time dummies.

Columns (1) and (3) replicate columns (2) and (4) of Table 2, respectively, for comparison. In column (2) of Table 3, the coefficient,β_1, is estimated to be positive (as predicted) and highly statistically significant. The introduction of JOT increased the outflow from JSA, presumably because individual incentives were diluted when group incentives where introduced.

The estimate suggests that the effect of JOT on JSA increases with the team size in the district. Larger offices face a greater dilution of incentives. For example, for a district with 210 employees (the mean of the distribution of the number of employees), the introduction of JOT increased the outflow of JSA by 3.4%.

By contrast, the effect of JOT in districts where there was a smaller team size (and hence more peer monitoring), the effect of JOT was more muted. My estimates suggest that in a district with 96 employees (median) the rise in JSA outflows after JOT was only 2.2%. The JOT policy effect, thus, increases monotonically with team size and becomes positive overall in teams with more than 20 employees (Figure 6A). Districts with team sizes less than 20 employees account for only 17% of districts and only 4.8% of the pre-policy (i.e. 1999 quarter 3) stock of JSA benefit claimants.

I plot the marginal change in JSA and IB outflows when moving to team based JOT by office size, conditional on already being treated with JET (i.e. $T_{dt}^{base} = 1$) in Figures 6A and 6B, respectively. Comparing the two, we can see that the effect of increasing team size is the opposite for JSA and IB.

Column (4) of Table 3 estimates the policy effect of JOT by team size on IB outflows. The estimate of the coefficient β_1 is negative and significant at the 5% level. It goes in the direction suggested by the second hypothesis. After JOT was introduced, outflows from IB decreased, probably because pre-existing individual incentives were diluted. For instance, for a district with 210 employees (mean), the introduction of JOT reduced the IB outflow by 1.7%.

By contrast, the effect of JOT in districts with a smaller team size was less harmful. In a district with 96 employees (median) the fall in IB outflows was only 0.8%. As Figures 6B and 4 show, the marginal effect of JOT on IB continues to decline with team size and is negative for more than two-thirds of the districts. Districts in the far left of the distribution where JOT preserves the baseline positive effect of JET (i.e. districts less than 47 employees) account for only one-third of all districts and for just 12 % of the pre-policy (i.e. 1999 quarter 3) IB caseload. The evidence that team incentives under JOT could replicate (or even add to) the positive effect of the individual incentives under JET for small districts (<47 employees) may suggest the emergence of information sharing or even risk pooling in teams under JOT. In large teams, the net effect is however governed by the perverse $\frac{1}{N}$ problem (Holmstrom, 1982).

The above estimates support the presence of free-riding in teams, confirming the second hypothesis. Further, I find that α_2 i.e. the baseline JET effect by team size is not only statistically insignificant in columns (2) and (4), but it is also in the opposite direction of β_1. This reconfirms that my estimates are not merely picking up scale effects.

6.3 Team Size Distribution

I now investigate the third hypothesis which is based on the model with perfect peer monitoring. To test the third hypothesis that the pre-existing incentives are intact i.e. the impact of JOT is lower in highly concentrated teams, I estimate the following specification:

$$\ln(y_{adt}) = \beta_1[T_{dt}^{base} * T_{dt}^{JOT} * \ln(S_d * H_d)] + \beta_2(T_{dt}^{base} * T_{dt}^{JOT}) + \alpha_1 T_{dt}^{base} + \alpha_2[T_{dt}^{base} * \ln(S_d * H_d)] + \gamma_1 lnU_{adt-1} + \gamma_2 lnU_{a'dt-1} + \delta_d + \delta_t + \delta_a + \varepsilon_{adt} \quad (3)$$

The policy is now interacted with the product of the district size, S_d, and its Herfindahl index, H_d. This effect is captured by β_1 in equation (3). And the third hypothesis can be restated as β_1 is positive (negative) if α_1 is positive (negative).

The results are shown in Table 4. Columns (1) and (4) replicate columns (2) and (4) of Table 3 for comparison. In column (2), the effect of the distribution of the team is in the first row, given by the coefficient on $T_{dt}^{base} * T_{dt}^{JOT} * \ln(S_d * H_d)$. The estimate for JSA shows that the effect of the team size concentration is the absence of any dilution of incentives. The coefficient is negative and significant. It is in the same direction as the baseline – negative for JSA. This is evidence in support of the third hypothesis and suggests that there is peer monitoring going on at the office level. It is difficult to say with certainty whether this is the minimum level of dilution possible. In other words, if there is perfect monitoring, whether this is the maximum level of output achievable. Even if there is lower than the maximum level of monitoring going on, the result definitely suggests that there is peer monitoring taking place in offices. And there is sufficient peer monitoring going on to preserve the pre-existing incentives such that the total effect of JOT is in the same direction as the negative effect of JET. Also note that there is no effect of team size concentration on the pre-existing JET policy.

An equivalent specification is to separately interact the policy with the Herfindahl Index and the district size. This is equivalent to equation (3) since the log of the product is equal to the log of each separately. The result for this equivalent specification is shown in column (3). Though the two coefficients are no longer significant, they (-0.0332, -0.0581) are in the same direction as the effect of -0.0117 in column (2).

Surprisingly, comparing columns (1) and (3), the coefficient on the total district size reverses. It becomes negative in column (3) while it was positive in column (1). This suggests that the total district size and the Herfindahl index are negatively correlated and that the Herfindahl index is an omitted variable in column (1). Thus, the estimates obtained in column (1) are upwardly biased. In summary, this is evidence in support of the second model of the presence of peer monitoring and the strength of peer monitoring in preserving baseline incentives even when moving to a team based rewards system.

In column (5), the estimated coefficient on $T_{dt}^{base} * T_{dt}^{JOT} * \ln(S_d * H_d)$ for IB shows that the effect of the team size concentration is the absence of any dilution of incentives. The coefficient is positive and in the same direction as the pre-existing JET (0.0121). It is also statistically different from zero at the 10% level of significance. This supports the third hypothesis and suggests that there is peer monitoring going on within offices.

The equivalent specification is estimated in column (6). The estimated coefficients on district size and Herfindahl Index are no longer significant. Similar to the bias in the estimates for JSA, the estimates obtained in column (4) are upwardly biased (-0.0115 in column 4 compared to -0.0360 in column 6). Thus, including the Herfindahl Index is essential to get unbiased estimates. Peer monitoring is present and drives the marginal impact of JOT on team output.

6.4 Extension: Office Level Regressions

This section considers office level output. The district output is in essence an aggregation of outputs from individual offices. To check the robustness of the findings at the district level in section 6.1 and 6.2, I re-estimate equations (1) and (2) for office level outcomes[34]. To account for common shocks or unobservables across all offices in a district, I cluster standard errors at the district level for the office level specifications.

Column (1) of Table 5 estimates the pre-existing scenario when JET was in place (i.e. $T_{dt}^{base} = 1$), for the log (total outflow) from an office for JSA. I find a highly significant negative coefficient of -0.137, similar to the district level coefficients. More interestingly, in column (2), I find the estimated marginal effect of JOT on the total outflow from JSA is positive and highly significant. The effect is almost double the district level effect. This again is evidence in support of the first hypothesis that the introduction of team incentives dilutes initial incentives.

Next, I translate the second hypothesis to the office level. The model predicts that the size of one's own office should not affect policy impacts, only the total team size in the district should affect incentives and output. I estimate the following equation at the office level for JSA:

[34] Information on IB outflows at office level is not made available by the DWP.

$$\ln(y_{aot}) = \beta_1[T_{dt}^{base} * T_{dt}^{JOT} * \ln(S_d)] + \beta_2[T_{dt}^{base} * T_{dt}^{JOT} * \ln(S_o)] + \beta_3(T_{dt}^{base} * T_{dt}^{JOT}) + \alpha_1 T_{dt}^{base} + \alpha_2 T_{dt}^{base} * \ln(S_d) + \alpha_3 T_{dt}^{base} * \ln(S_o) + \gamma_1 lnU_{aot-1} + \gamma_2 lnU_{a'ot-1} + \delta_o + \delta_t + \delta_a + \varepsilon_{aot} \quad (4)$$

where S_d represents the total number of employees in district d and S_o represents the total number of employees in office o. The second hypothesis states that β_1 be positive and significant, while β_2 is insignificant.

Table 6 reports the estimates of equation (4). Column (1) replicates column (2) of Table 5 for comparison. Column (2) shows that the estimate of the coefficient,β_1, is positive and highly significant. This suggests that the size of the district team size matters even for office level outcomes. The magnitude is slightly higher than the district level estimates in Table 3. I also find that α_3 is in the opposite direction of β_1. This suggests that my estimates are not merely picking up scale effects.

Column (3) adds the policy interaction with the own office size S_o i.e. the number of employees in one's immediate office. I find that after controlling for the district team size, own office size does not matter. This confirms the second hypothesis and suggests that there is perfect monitoring wherein officers in a particular office act symmetrically. They only take the size of other offices as given controls in their decision making. The result at the office level reinforces the evidence in favour of the presence of peer monitoring.

7 Robustness

In this section, I show different specifications to check the robustness of my baseline results on the marginal impact of JOT. I start by testing for pre-treatment effects that may be confounding the causal interpretation of my estimates.

7.1 Pre-trends

I include pre-treatment dummies in the basic specification of equation (1) to test for the presence of selection into treatment. I estimate the following outflow equation for districts:

$$\ln(y_{adt}) = \alpha_1 T_{dt}^{base} + \beta_1 T_{dt}^{base} * T_{dt}^{JOT} + + \sum_{\tau=1}^{4} \beta_{1\tau} T_{dt-\tau}^{base} * T_{dt-\tau}^{JOT} + \gamma_1 lnU_{adt-1} + \gamma_2 lnU_{a'dt-1} + \delta_d + \delta_t + \delta_a + \varepsilon_{adt} \quad (5)$$

For simplicity, I include four pre-dummies, but the results are robust across other pre-dummies. In Table 7, I report the coefficients of interest $\beta_{1\tau}$ of equation (5) for JSA and IB. (The office level equivalent is reported in Table A3 of Appendix A[35]). Columns (1) and (3) replicate the baseline results for comparison.

In column (2), I estimate the total outflow from JSA at the district level and see that the coefficients on pre-treatment dummies indicate the absence of any significant pre-policy effects in the JOT treated districts. The p-value for the F-test of the joint significance of the pre-treatment dummies shows that they are jointly insignificant as well. Similarly, for outflows from IB, I can reject the joint significance of the pre-treatment dummies in column (4). The estimated coefficients on the policy effect are slightly smaller than baseline but are in the same direction.

Hence, a robust result from Table 7 is that moving from individual reward contracts to incentive structures based on team output leads to a dilution of initial incentives, even after controlling for any pre-policy differences across districts.

[35] In the office level outcome specifications in Table A3, the pre-treatment dummies are significantly negative for the year before the policy. The pre-treatment dummies are jointly insignificant however. Most importantly, the estimated baseline specifications are very robust even after I control for these pre-treatment differences in columns (2) and (4).

7.2 Robustness to Monthly Frequency Data

Given data constraints, I am also able to estimate outflow equations at the monthly, rather than quarterly, frequency only for JSA. The dependent variable in Table A2 is now the monthly outflow from JSA, having included the JSA stock at the end of the previous month as a control. Column (2) in Table A2 shows a marginal effect of the JOT policy on outflows of 5.4%, which is almost double of the quarterly impact of 2.5%. The results for the effect of team size are also similar to the quarterly estimations in Table 3.

7.3 Endogeneity of Office Size

So far in the analysis, the size of an office and district has been fixed at the pre-policy level of June 2004. One concern is that an Office Manager could endogenously choose the size of the office under the new incentive structure. This could confound some of the results. To test this, I use annual data on the size of each office in the years 2004 and 2007 to check whether there is any policy effect on the size of an office post policy.

Results reported in Table 8 are based on a pooled regression of the logs as well as levels of the office employee numbers in the years 2004 and 2007. It re-estimates equation (1) at the annual, rather than quarterly level, with the outcome as employee numbers. It includes district fixed effects as well as year fixed effects.

Column (2) shows that the JOT policy did not have any impact on the total number of employees in an office after the policy in 2007. When using the natural log of the number of employees in an office as the dependent variable as in columns (3) and (4), I again find an absence of any policy effect on the office employee numbers. This suggests that the size of an office is not driven by the JOT policy but is determined by other factors such as the size of population or density in a district. I explore this in the next section.

7.4 Determinants of Team Size

What determines the size of a district's team? The answer to this matters since one could argue that while team size is not effected by the policy, it is fundamentally determined by certain characteristics of the district. And, my estimations of the marginal effect of JOT are simply picking up the effect of these district characteristics.

To test this, I first determine what characteristics of a district are correlated with the team size in a district. Based on data from the 2001 census in England, Wales and Scotland, Table 9 estimates the correlation between team size and district characteristics such as the size of working age population, the number of lone parent families, economic activity, population density, gender composition and historical benefit caseloads. I looked at various other characteristics of a district such as historical benefit exit rates, size of the student population and family composition. But, they were not significantly correlated with team size and have been omitted only for brevity.

The dependent variable in column (1) of Table 9 is the absolute team size in a district while in column (2) it is the log of the team size – the variable that I use in baseline estimations. Among the variables presented in Table 9, the percentage of the population that is of working age is positively and significantly correlated with average office size in both columns. Districts with a larger working age population tend to have larger offices. Similarly, districts with a higher percentage of economically active labor force have bigger offices. The greater the share of unemployed people and smaller the share of self-employed people in the labor force, the district has more officers in benefit offices. This is intuitive as this implies a greater number of benefit claimants and job searchers to help. Finally, the higher the historical[36] caseload of IB claimants in a district, the larger is the team.

Next, I explore the robustness of the baseline results to the inclusion of the district characteristics. The above results suggest that the size of the working age population, economically active labor force, unemployment, self-employment and historical IB caseload in a district are correlated with the size of offices in a district. One concern is, however, that my estimate of the team size effect in Table 3 is not really the effect of the team size but the effect of the demographics in a district. If this were the case, my estimates would vanish if I include the interaction of the policy with these demographic variables. To test this, I estimate equation (2) adding the demographic correlates interacted with the policy. Columns (1) and (3) of Table 10 replicate columns

[36] The number of historical claimants is measured as the average caseload in a district before 2001.

(2) and (4) of Table 3 for comparison. Estimates presented in columns (2) and (4) of Table 10 show that the baseline results of Table 3 are robust to the inclusion of demographic variables.

7.5 Total Policy Effect - JET and JOT

A question that may arise is to what degree JOT reverses the pre-existing JET effect in large offices i.e. whether the two effects are statistically equal. The p-value for the F-test of equality of the two sets of coefficients $(\beta_1 + \beta_2)$ versus $(\alpha_1 + \alpha_2)$ in equations (2) and (3), is reported in the third to last row in Tables 3 and 4. In most cases, even after accounting for team size (Table 3), the p-value suggests that the two effects are equal for both JSA as well as for IB. This reinforces the baseline prediction that the marginal effect of JOT just overturns the pre-existing JET effects and is statistically equivalent in magnitude. It suggests that the JOT policy completely setbacks the incentive boost provided by JET to help the disabled.

8 Conclusions

I estimate the marginal effect of introducing team based performance contracts in the job placement services of the UK. The new system marked a major shift in the performance measurement and benchmarking processes. The first striking result in my empirical analysis is that a dilution of incentives takes places if performance measurement is based at the team level, instead of the individual level. Employees shift effort towards easier tasks and free-ride on their peers' efforts. Thus, I see a decrease in benefit outflows from IB but an increase in the easy-to-help JSA outflows.

Second, the degree to which the dilution takes place depends on the incentives to free-ride and monitor peers. My results show that the marginal impact of JOT depends on the size of the district team. The larger the team size, the greater is the incentive to free-ride. But, most surprisingly, the distribution of the team size across spatially separated sub-teams also determines policy impact. I find support for the presence of peer monitoring when team based rewards are introduced. In more concentrated teams (across sub-teams), peer monitoring allows some of the negative externalities to be internalized. This counterbalances the free-riding incentives and preserves the baseline incentives. While it is difficult to say with certainty whether this is the minimum level of dilution possible, the results definitely suggest that there is peer monitoring taking place in offices. And there is sufficient peer monitoring going on to preserve the pre-existing incentives such that the total effect of JOT is in the same direction as the negative effect of JET. These results are robust to a wide range of controls including fixed effects and using different specifications.

This chapter highlights that when a weaker incentive scheme is used in the public sector, output can decrease due to the incentives to free-ride in teams. In my context, the new team based contracts are unique in the way that multiple sub-teams combine to form the relevant team and the sub-teams are spatially dispersed. In this unique setting, I find that despite free-riding in teams, if there is sufficient peer monitoring going on within sub-teams then weaker incentive schemes could increase output even though rewards are based on team output rather than individual output.

While this unique feature of the policy is the starting point for the model and analysis, the production process is similar to a decentralized manufacturing process and is likely to have wider relevance beyond the public sector. An example of an analogous set-up is of a car manufacturer that has distribution agents (for example, car distributors) for each region of the market and who compete against each other to achieve a certain benchmark (or target). Each distributor, in turn, has many local dealers working for him towards the target. Thus, the model as well as the analysis may be generalizable to other public sector as well as private sector production contexts. In summary, nuanced features of the production process such as the spatial organization of teams can affect the power of incentives, ceteris paribus.

There are many directions I want to take this work. In particular, I am concerned about the general equilibrium impacts of the JOT policy on the labor market. To what extent does the lower labor supply lead to higher equilibrium wages? How does the composition of the benefit caseloads impact public finance? These are areas I am actively engaged in exploring.

References

Aghion, Philippe, Tim Besley, John Browne, Francesco Caselli, Richard Lambert, Rachel Lomax, Nick Stern and John Van Reenen (2013) *Investing for Prosperity: Report of the LSE Growth Commission* http://www2.lse.ac.uk/researchAndExpertise/units/growthCommission/documents/pdf/LSEGC-Report.pdf

Alchian, Armen A. and Harold Demsetz (1972), "Production, Information Costs and Economic Organization", *American Economic Review*, 62(5): 777-795

Bagaria, Nitika, Novella Bottini and Miguel Coelho (2013), "Human Capital and Economic Growth: A Focus on Primary and Secondary Education in the UK", *Investing for Prosperity: Report of the LSE Growth Commission* in Tim Besley and John Van Reenen (eds), Chapter 3: 91-151

Besley, Tim and Maitreesh Ghatak (2005), "Competition and Incentives with Motivated Agents", *American Economic Review* 95(3): 616-636

Bloom, Nick and John Van Reenen (2011), "Human Resource Management and Productivity", *Handbook of Labor Economics Volume 4B* in Orley Ashenfelter and David Card (eds), Chapter 19 1697-1769

Blundell, Richard, Monica Costa Dias, Costas Meghir and John Van Reenen (2004), "Evaluating the Employment Impact of a mandatory job search assistance programme", *Journal of the European Economics Association,* 2(4): 569-606

Blundell, Richard, David Card and Richard Freeman (2004), "Active Labor Market Policies and the British New Deal for Youth in Context", *Seeking a Premier Economy*

Crepon, Bruno, Esther Duflo, Marc Gurgand, Roland Rathelot and Philippe Zamora (2012), "Do Labour Market Policies have Displacement effects? Evidence from a clustered randomized experiment", CEPR Discussion Paper No. 9251

Dixit, Avinash (2002), "Incentives and Organizations in the Public Sector: An interpretative Review", *Journal of Human Resources,* 37(4): 696-727

Encinosa, William, Martin Gaynor and James B. Rebitzer (2007), "The Sociology of Groups and the Economics of Incentives: Theory and Evidence on Compensation Systems", *Journal of Economic Behavior and Organization*, 62(2): 187-214

Frank, Richard G. and Martin Gaynor (1995), "Incentives, Optimality, and Publicly Provided Goods: The Case of Mental Health Services", *Public Finance Quarterly*, 23(2): 167-192

Gaynor, Martin (1986), "Misperceptions, Moral Hazard, and Incentives in Groups", *Managerial and Decision Economics*, 7(4): 279-282

Gaynor, Martin and Paul R. Kleindorfer (1987), "Misperceptions, Equilibrium and Incentives in Groups and Organizations", *Agency Theory, Information, and Incentives*, Gunter Bamberg and Klaus Spremann (eds.): 389-414

Gregg, Paul (2008), "UK Welfare Reform 1996 to 2008 and beyond: A personalised and responsive welfare system?", CMPO Working paper No. 08/196, Department of Economics, University of Bristol

Gregg, Paul and Jonathan Wadsworth (2011), "The Labour Market in Winter – The 2008-2009 Recession", OUP, Oxford

Hamilton, Barton H., Jack A. Nickerson, and Hideo Owan (2003), "Team Incentives and Worker Heterogeneity: An Empirical Analysis of the Impact of Teams on Productivity and Participation", *Journal of Political Economy*, 111(3): 465-497

Heckman, James, Carolyn J. Heinrich, Pascal Courty, Gerald Marschke, and Jeffrey Smith (2011), "The Performance of Performance Standards", published by W.E. Upjohn Institute of Employment Research

HMT, "Opportunity for all: The strength to take the long-term decisions for Britain" (2004), Pre-Budget Report

Holmström, Bengt (1982), "Moral Hazard in Teams", *Bell Journal of Economics,* 13(2): 324-340

Holmström, Bengt and Paul Milgrom (1991). "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design", *Journal of Law, Economics, and Organization,* 7(0): 24–52

Hyde, Mark, John Dixon and Melanie Joyner (2002). "Work for Those Who Can and Security for Those Who Cannot: The United Kingdom's New Social Security Reform Agenda", *International Social Security Review,* 52(4): 69-86.

Kandel, Eugene and Edward P. Lazear (1992), "Peer Pressure and Partnerships", *Journal of Political Economy,* 100(4): 801-817

Lazear Edward P. and Paul Oyer (2012), "Personnel Economics", *Handbook of Organizational Economics* in Robert Gibbons and D. John Roberts (eds), Chapter 12: 479-519

Makinson, John (Chair), (2000), "Incentives for Change: Rewarding Performance in National Government Networks", Public Services Productivity Panel, HMT.

Manning, Alan (2009), "You Can't Always get What You Want: The Impact of the UK Jobseeker's Allowance", *Labour Economics*, 16: 239-250

Mas, Alexandre and Enrico Moretti (2009), "Peers at Work", *American Economic Review*, 99:112-145

National Audit Office (2008), "The roll-out of the Jobcentre Plus Office network", TSO.

Petrongolo, Barbara (2009), "What Are the Long-Term Effects of UI? Evidence from the UK JSA Reform", *Journal of Public Economics,* 93: 1234-1253.

Propper, Carol and Deborah Wilson (2003) "The Use and Usefulness of Performance Measures in the Public Sector", CMPO Working Paper 03/073

Riley, Rebecca, Helen Bewley, Simon Kirby, Ana Rincon-Aznar and Anitha George (2011), "The introduction of Jobcentre Plus: An evaluation of labor market impacts", DWP Research Report No 9781

Tables and Figures

Figure 1: Jobcentre Hierarchy

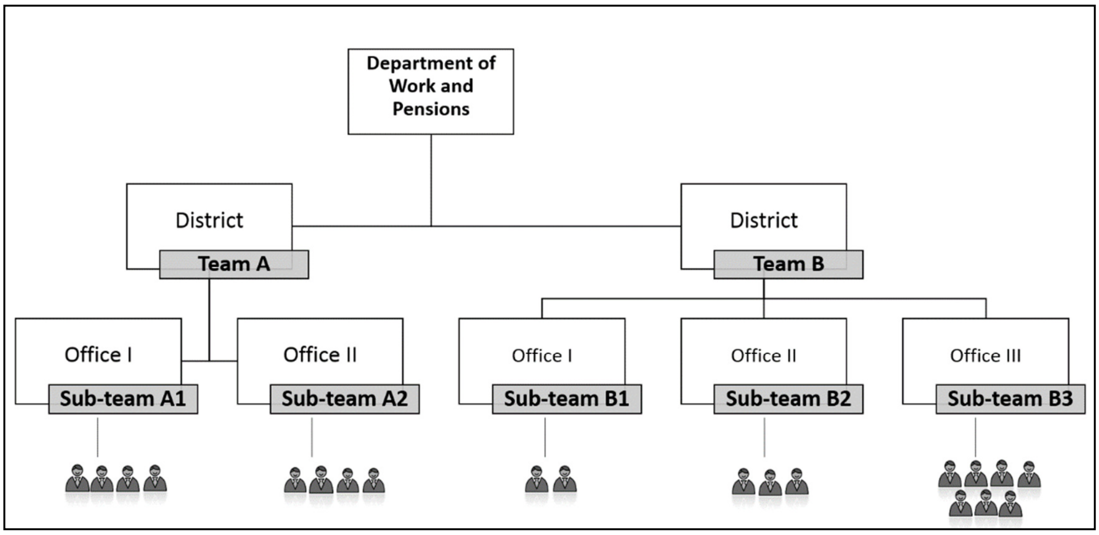


Figure 2: Rollout of the Job Entry Target (JET) and Job Outcome Target (JOT) Points

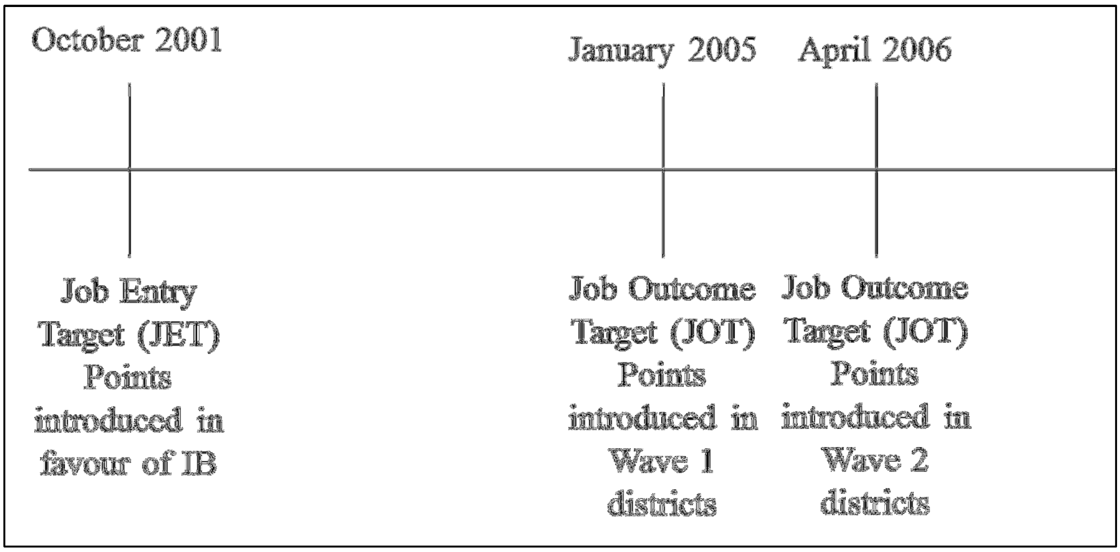
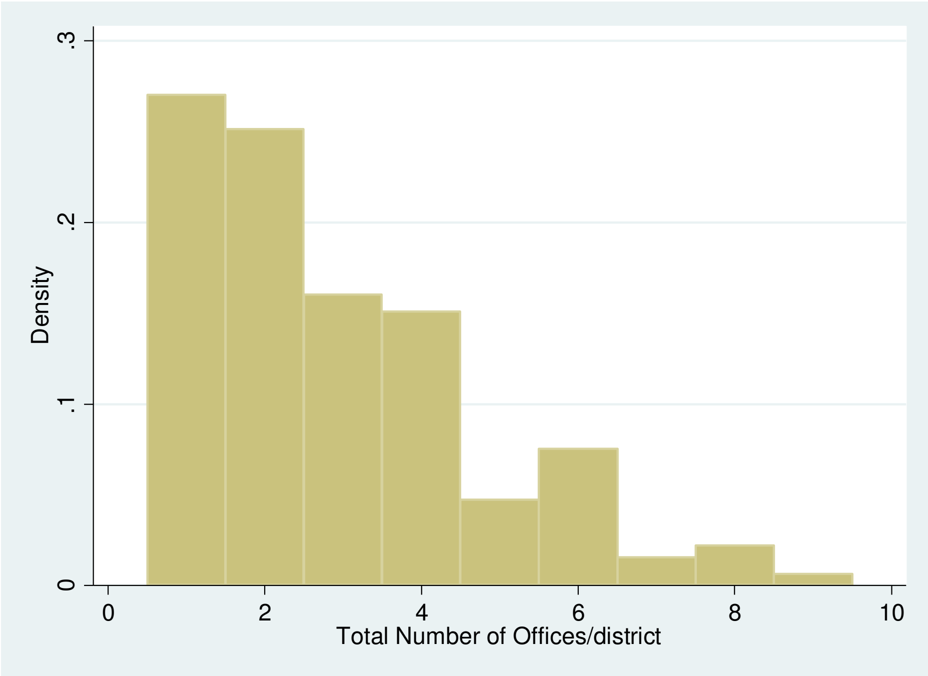
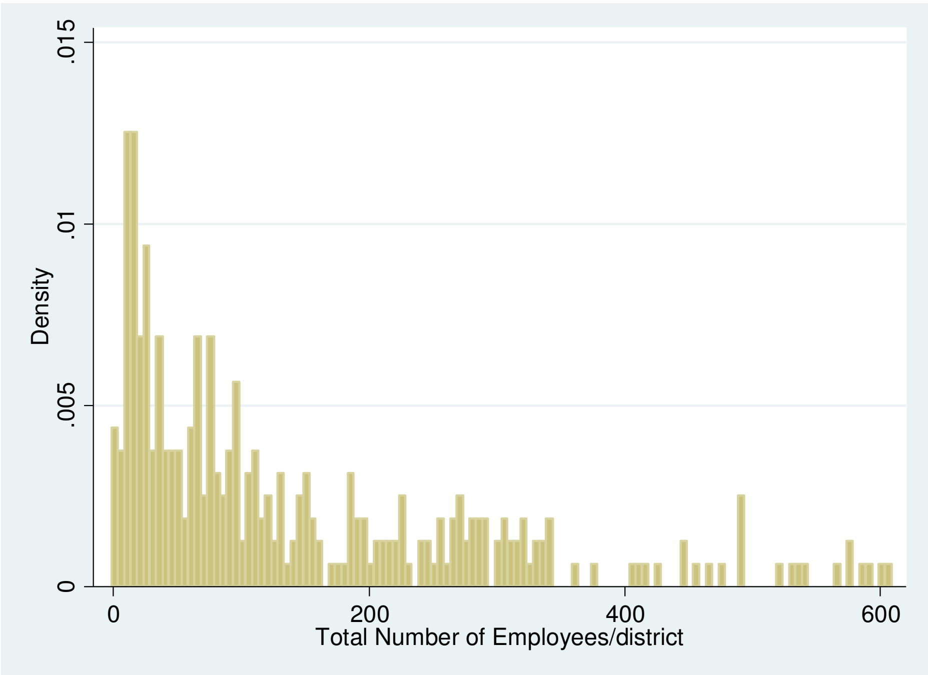


Figure 3: Distribution of Offices in Districts



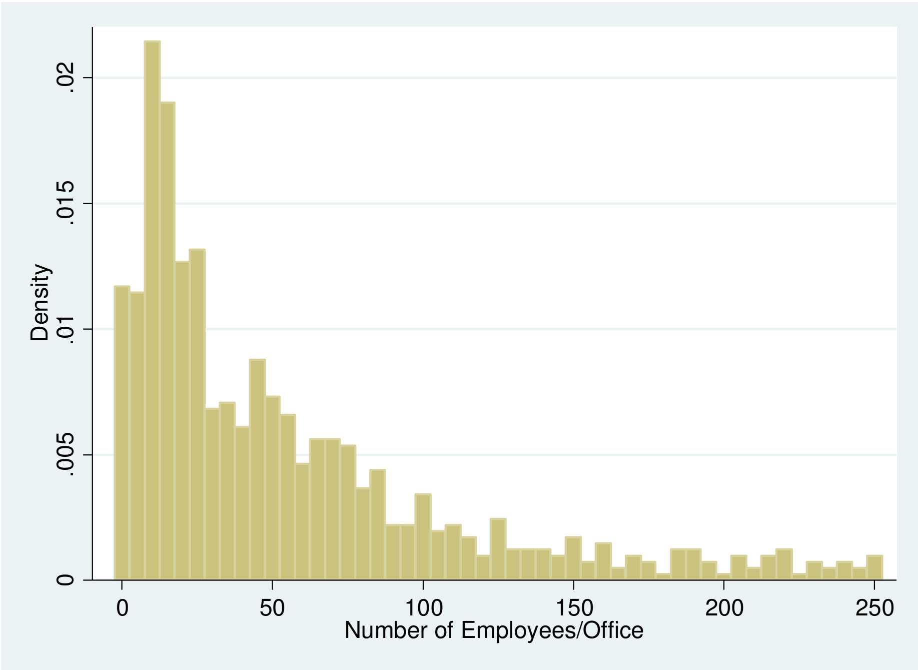
Note: This histogram plots the distribution of the number of offices per district in the 338 districts. The upper bound used is the 95th percentile of the distribution i.e. 10 offices.

Figure 4: Distribution of Employees in Districts



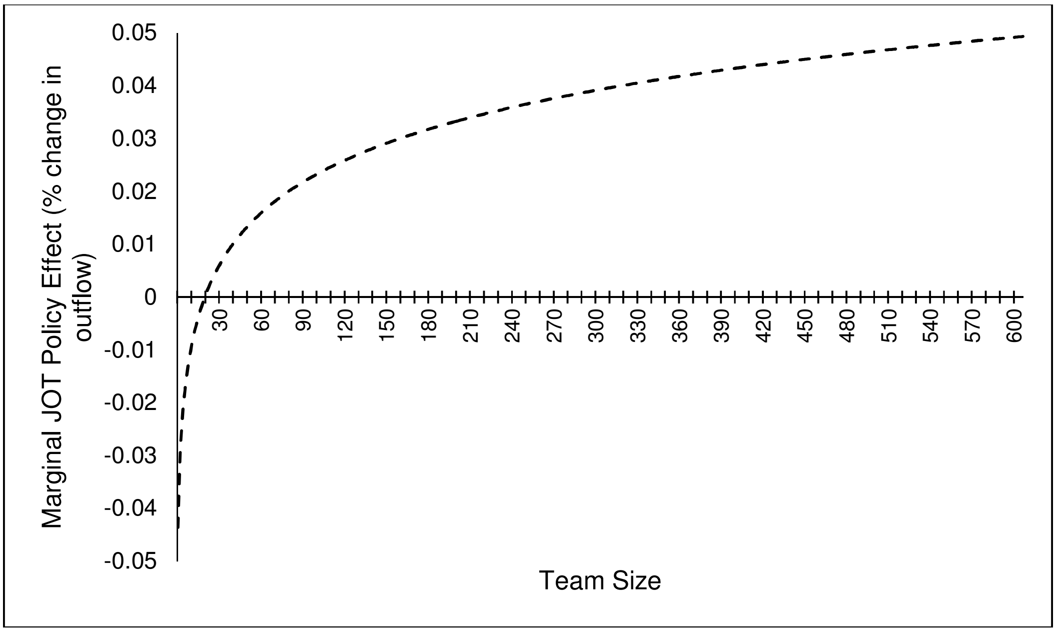
Note: This histogram plots the distribution of the number of employees per district in the 338 districts. The upper bound used is the 95th percentile of the distribution i.e. 607 employees.

Figure 5: Distribution of number of employees across Offices



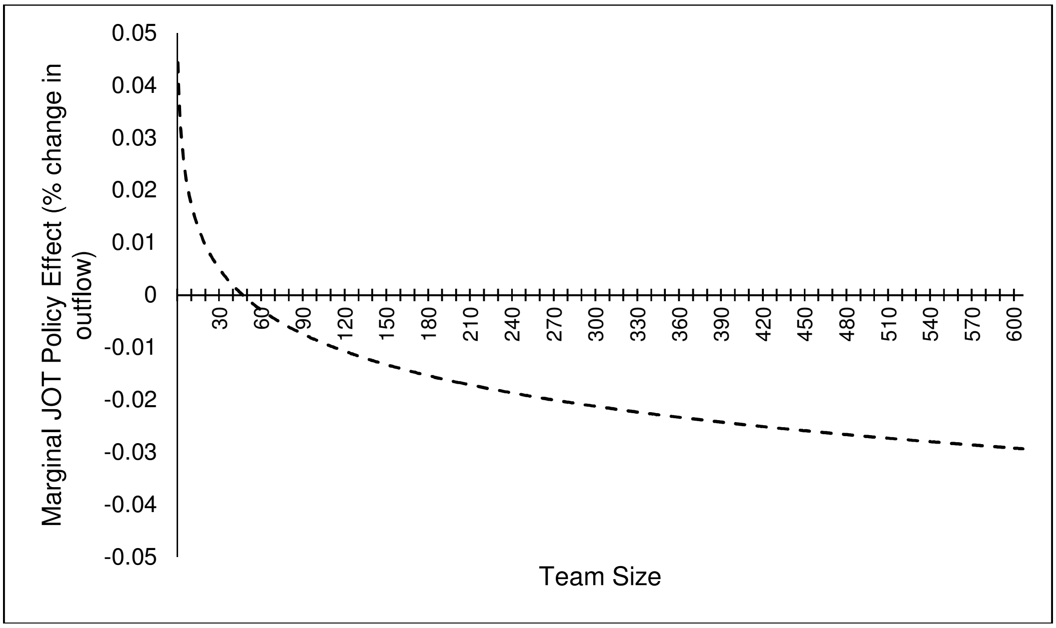
Note: This histogram plots the distribution of the number of employees per office in the 851 offices. The upper bound used is the 95th percentile of the distribution i.e. 252 employees.

Figure 6A: Marginal Policy Effect of JOT on JSA



Note: The graph plots the marginal effect of the JOT policy on outflow rates from JSA. The effect is based on regression estimates of column 2, Table 3. The upper bound of team size used is the 99th percentile of the distribution i.e. 607 employees.

Figure 6B: Marginal Policy Effect of JOT on IB



Note: The graph plots the marginal effect of the JOT policy on outflow rates from IB. The effect is based on regression estimates of column 4, Table 3. The upper bound of team size used is the 99th percentile of the distribution i.e. 607 employees.

Table 1: Summary Statistics

Panel A: Outcome variables - Quarterly Aggregate	Mean	SD
	(1)	(2)
JSA Outflow	653,819	78,049
JSA Stock*	939,267	115,578
JSA Outflow rate	0.698	0.057
IB Outflow	56,166	11,267
IB Stock*	2,045,210	356,417
IB Outflow rate	0.028	0.0027

* Based on the stock at the end of the previous quarter.

Panel B: Office Size variables	Districts = 338		Offices = 851
	No. of Employees	*No. of Offices*	*No. of Employees*
Mean	210.41	3.97	68.38
Standard Deviation	459.94	7.33	85.4
Median	96	3	40
Mode	14	1	1
25th Percentile	31	1	15
95th Percentile	607	10	252
99th Percentile	1992	30	392
Average per Office	45.40		

Table 2: Baseline effects of JOT on JSA and IB

Dependent Variable: Log (Total Outflow)				
	(1)	(2)	(3)	(4)
	JSA	**JSA**	**IB**	**IB**
$\boldsymbol{T_{dt}^{base} * T_{dt}^{JOT}}$		0.0258**		-0.0104
(Marginal effect of JOT policy)		(0.0103)		(0.0132)
$\boldsymbol{T_{dt}^{base}}$	-0.0101*	-0.0118*	0.0143	0.0161
(Baseline JET policy effect)	(0.0057)	(0.0064)	(0.0099)	(0.0098)
$\boldsymbol{lnU_{adt-1}}$	0.6383***	0.6379***	0.3542***	0.3546***
	(0.0119)	(0.0119)	(0.0344)	(0.0345)
$\boldsymbol{lnU_{a'dt-1}}$	0.0252**	0.0247**	0.0375	0.0379
	(0.0108)	(0.0108)	(0.0262)	(0.0261)
N	22,109	22,109	20,416	20,416
p-value		0.0061		0.118

Note: The dependent variable is the log of the outflow from benefit during a quarter-year. All regressions control for district by age fixed effects and age by time fixed effects. T_{dt}^{base} is a dummy which switches on in a district when the previous JET policy was introduced. $T_{dt}^{base} * T_{dt}^{JOT}$ is a dummy that switches on in a district when JOT was introduced (this was introduced only in those areas that had switched to JET and at a later time period). The sample is a panel of 338 districts from 1999Q4 to 2007Q4. In columns 1 and 2, the dependent variable is log (outflow) for JSA in an age group and district in a particular quarter t. In columns 3 and 4, the dependent variable is log (outflow) for IB in an age group and district in a particular quarter t. The p-value row contains the p-value of the F test for the equality between the effect of JET and marginal JOT effect. Standard errors are clustered at the district level. *** p<0.01, ** p<0.05, * p<0.1

Table 3: Effects of Team Size on JSA and IB

Dependent Variable: Log (Total Outflow)				
	(1)	(2)	(3)	(4)
	JSA	**JSA**	**IB**	**IB**
$\boldsymbol{T_{dt}^{base} * T_{dt}^{JOT} * \ln(S_d)}$		0.0145***		-0.0115**
(Marginal effect of JOT policy by team size)		(0.0052)		(0.0057)
$\boldsymbol{T_{dt}^{base} * T_{dt}^{JOT}}$	0.0258**	-0.0436*	-0.0104	0.0444
(Marginal effect of JOT policy)	(0.0103)	(0.0248)	(0.0132)	(0.0300)
$\boldsymbol{T_{dt}^{base}}$	-0.0118*	-0.0103	0.0161	0.0060
(Baseline JET policy effect)	(0.0064)	(0.0067)	(0.0098)	(0.0355)
$\boldsymbol{T_{dt}^{base} * \ln(S_d)}$		-0.0004		0.0019
(Baseline JET policy effect by team size)		(0.0029)		(0.0067)
$\boldsymbol{T_{dt}^{JOT} * \ln(S_d)}$		0.0031		0.0037
		(0.0029)		(0.0035)
N	22,109	22,109	20,416	20,416
p-value	0.0061	0.429	0.118	0.603

Note: The size variable, S_d, used in these regressions is the total number of employees in a district. The dependent variable is the log of the outflow from benefit during a quarter-year. All regressions control for district by age fixed effects and age by time fixed effects. T_{dt}^{base} is a dummy which switches on in a district when the previous JET policy was introduced. $T_{dt}^{base} * T_{dt}^{JOT}$ is a dummy that switches on in a district when JOT was introduced (this was introduced only in those areas that had switched to JET and at a later time period). The sample is a panel of 338 districts from 1999Q4 to 2007Q4. In columns 1 and 2, the dependent variable is log (outflow) for JSA in an age group and district in a particular quarter t and includes controls for the stock of JSA in the previous quarter. In columns 3 and 4, the dependent variable is log (outflow) for IB in an age group and district in a particular quarter t and includes controls for the stock of IB in the previous quarter. The p-value row contains the p-value of the F test for the equality between the effect of JET and total effect of JOT. Standard errors are clustered at the district level. *** p<0.01, ** p<0.05, * p<0.1

Table 4: Effects of Team Distribution on JSA and IB

Dependent Variable: Log (Total Outflow)	(1) JSA	(2) JSA	(3) JSA	(4) IB	(5) IB	(6) IB
$T_{dt}^{base} * T_{dt}^{JOT} * \ln(S_d * H_d)$ *(Marginal effect of JOT policy by team distribution)*		-0.0117**			0.0074*	
		(0.0047)			(0.0043)	
$T_{dt}^{base} * T_{dt}^{JOT} * \ln(H_d)$ *(Marginal effect of JOT policy by Herfindahl index)*			-0.0332			-0.0150
			(0.0232)			(0.0152)
$T_{dt}^{base} * T_{dt}^{JOT} * \ln(S_d)$ *(Marginal effect of JOT policy by team size)*	0.0145***		-0.0581	-0.0115**		-0.0360
	(0.0052)		(0.0498)	(0.0057)		(0.0349)
$T_{dt}^{base} * T_{dt}^{JOT}$ *(Marginal effect of JOT policy)*	-0.0436*	-0.0371	-0.0354*	0.0444	0.0291	0.0280
	(0.0248)	(0.0258)	(0.0187)	(0.0300)	(0.0273)	(0.0335)
T_{dt}^{base} *(Baseline JET policy effect)*	-0.0103	-0.0102	-0.0098	0.0060	0.0121	0.0050
	(0.0067)	(0.0067)	(0.0068)	(0.0355)	(0.0307)	(0.0356)
$T_{dt}^{base} * \ln(S_d * H_d)$		0.0008			-0.0006	
		(0.0025)			(0.0049)	
$T_{dt}^{base} * \ln(H_d)$			0.0232			0.0045**
			(0.0181)			(0.0020)
$T_{dt}^{base} * \ln(S_d)$	-0.0004		0.0494	0.0019		0.0017
	(0.0029)		(0.0391)	(0.0067)		(0.0067)
$T_{dt}^{JOT} * \ln(S_d * H_d)$		-0.0025			-0.0030	
		(0.0026)			(0.0029)	
$T_{dt}^{JOT} * \ln(H_d)$			0.0190			0.0324
			(0.0222)			(0.0295)
$T_{dt}^{JOT} * \ln(S_d)$	0.0031		0.0444	0.0037		0.0786
	(0.0029)		(0.0476)	(0.0035)		(0.0655)
N	22,109	22,109	22,109	20,416	20,416	20,416
p-value	0.429	0.243	0.0779	0.603	0.674	0.583

Note: The size variable, S_d, is the total number of employees in a district and H_d is the Herfindahl index of team size distribution in a district. All regressions control for district by age fixed effects and age by time fixed effects. T_{dt}^{base} is a dummy which switches on in a district when the previous JET policy was introduced. $T_{dt}^{base} * T_{dt}^{JOT}$ is a dummy that switches on in a district when JOT was introduced (this was introduced only in those areas that had switched to JET and at a later time period). The sample is a panel of 338 districts from 1999Q4 to 2007Q4. In columns 1-3, the dependent variable is log (outflow) for JSA in an age group and district in a particular quarter t and includes controls for the stock of JSA in the previous quarter. In columns 4-6,the dependent variable is log (outflow) for IB in an age group and district in a particular quarter t and includes controls for the stock of IB in the previous quarter. The p-value row contains the p-value of the F test for the equality between the effect of JET and total effect of JOT. Standard errors are clustered at the district level. *** p<0.01, ** p<0.05, * p<0.1

Table 5: Effects of JOT on JSA – outflow from offices

Dependent Variable: Log (Total Outflow)		
	(1) **JSA**	(2) **JSA**
$T_{dt}^{base} * T_{dt}^{JOT}$ *(Marginal effect of JOT policy)*		0.0444** (0.0187)
T_{dt}^{base} *(Baseline JET policy effect)*	-0.0137** (0.0067)	-0.0172*** (0.0066)
lnU_{aot-1}	0.8045*** (0.0233)	0.8039*** (0.0232)
$lnU_{a'ot-1}$	0.1326*** (0.0217)	0.1322*** (0.0217)
N	53,729	53,729
Dist.*Age FE	YES	YES
Age * Time FE	YES	YES

Note: The dependent variable is the log of the outflow from benefit during a quarter-year in an office. All regressions control for office by age fixed effects and age by time fixed effects. T_{dt}^{base} is a dummy which switches on in a district when the previous JET policy was introduced. $T_{dt}^{base} * T_{dt}^{JOT}$ is a dummy that switches on in a district when JOT was introduced (this was introduced only in those areas that had switched to JET and at a later time period). The sample is a panel of 851 offices from 1999Q4 to 2007Q4. The p-value row contains the p-value of the F test for the equality between the effect of JET and marginal JOT effect. Standard errors are clustered at the district level. *** p<0.01, ** p<0.05, * p<0.1

Table 6: Effects of Team Size on JSA – outflow from offices

Dependent Variable: Log (Total Outflow)			
	(1)	(2)	(3)
	JSA	**JSA**	**JSA**
$\boldsymbol{T_{dt}^{base} * T_{dt}^{JOT} * \ln(S_d)}$		0.0399***	0.0470***
(Marginal effect of JOT policy by district team size)		(0.0058)	(0.0128)
$\boldsymbol{T_{dt}^{base} * T_{dt}^{JOT} * \ln(S_o)}$			-0.0095
(Marginal effect of JOT policy by own office size)			(0.0207)
$\boldsymbol{T_{dt}^{base} * T_{dt}^{JOT}}$	0.0444**	-0.1230***	-0.1466***
(Marginal effect of JOT policy)	(0.0187)	(0.0271)	(0.0297)
$\boldsymbol{T_{dt}^{base}}$	-0.0172***	-0.0207***	-0.0206***
(Baseline JET policy effect)	(0.0066)	(0.0068)	(0.0068)
$\boldsymbol{T_{dt}^{base} * \ln(S_d)}$		-0.0147**	-0.0267*
		(0.0066)	(0.0145)
$\boldsymbol{T_{dt}^{base} * \ln(S_o)}$		0.0104	-0.0145
		(0.0087)	(0.0150)
$\boldsymbol{T_{dt}^{JOT} * \ln(S_d)}$			0.0233
			(0.0208)
$\boldsymbol{T_{dt}^{JOT} * \ln(S_o)}$			0.0364
			-0.0238
N	53,729	53,729	53,729

Note: The dependent variable is the log of the outflow from benefit during a quarter-year in an office. All regressions control for office by age fixed effects and age by time fixed effects, as well as, for the stock of JSA in the previous quarter. T_{dt}^{base} is a dummy which switches on in a district when the previous JET policy was introduced. $T_{dt}^{base} * T_{dt}^{JOT}$ is a dummy that switches on in a district when JOT was introduced (this was introduced only in those areas that had switched to JET and at a later time period). The sample is a panel of 851 offices from 1999Q4 to 2007Q4. The size variable, S_d, used in these regressions is the total number of employees in a district and S_o refers to the number of employees in a particular office. The p-value row contains the p-value of the F test for the equality between the effect of JET and total effect of JOT. Standard errors are clustered at the district level. *** p<0.01, ** p<0.05, * p<0.1

Table 7: Pre-treatment Trends in Benefit Outflow

Dependent Variable: Log(Total Outflow)				
	(1)	(2)	(3)	(4)
	JSA	**JSA**	**IB**	**IB**
$\boldsymbol{T_{dt}^{base} * T_{dt}^{JOT}}$	0.0258**	0.0530***	-0.0104	-0.0034
(Marginal effect of JOT policy)	(0.0103)	(0.0179)	(0.0132)	(0.0154)
$\boldsymbol{T_{dt}^{base}}$	-0.0118*	-0.0153**	0.0161	0.0177*
(Baseline JET policy effect)	(0.0064)	(0.0066)	(0.0098)	(0.0099)
$\boldsymbol{T_{dt-4}^{base} * T_{dt-4}^{JOT}}$		0.0092		0.0081
		(0.0065)		(0.0202)
$\boldsymbol{T_{dt-3}^{base} * T_{dt-3}^{JOT}}$		0.0098		0.0113
		(0.0066)		(0.0185)
$\boldsymbol{T_{dt-2}^{base} * T_{dt-2}^{JOT}}$		0.0072		0.0104
		(0.0078)		(0.0217)
$\boldsymbol{T_{dt-1}^{base} * T_{dt-1}^{JOT}}$		0.0124*		0.0125
		(0.0075)		(0.0202)
$\boldsymbol{lnU_{adt-1}}$	0.6379***	0.7561***	0.3546***	0.3235***
	(0.0119)	(0.0109)	(0.0345)	(0.0457)
$\boldsymbol{lnU_{a'dt-1}}$	0.0247**	-0.0964***	0.0379	0.1008***
	(0.0108)	(0.0125)	(0.0261)	(0.0378)
N	22,109	22,109	20,416	20,416
p-value		0.367		0.963

Note: The dependent variable is the log of the outflow from benefit during a quarter-year. All regressions control for district by age fixed effects and age by time fixed effects. T_{dt}^{base} is a dummy which switches on in a district when the previous JET policy was introduced. $T_{dt}^{base} * T_{dt}^{JOT}$ is a dummy that switches on in a district when JOT was introduced (this was introduced only in those areas that had switched to JET and at a later time period). The sample is a panel of 338 districts from 1999Q4 to 2007Q4. In columns 1 and 2, the dependent variable is log (outflow) for JSA in an age group and district in a particular quarter t. In columns 3 and 4, the dependent variable is log (outflow) for IB in an age group and district in a particular quarter t. The p-value row contains the p-value for the F-test of the joint significance of the pre-treatment dummies. Standard errors are clustered at the district level. *** p<0.01, ** p<0.05, * p<0.1

Table 8: Endogeneity of Office Size

Dependent Variable: Number of Employees in an Office				
	(1) **Levels**	(2) **Levels**	(3) **Logs**	(4) **Logs**
$\boldsymbol{T_{dt}^{base} * T_{dt}^{JOT}}$		289.5959		-0.0149
(Marginal effect of JOT policy)		(430.8118)		(0.8699)
$\boldsymbol{T_{dt}^{base}}$	283.6716	265.1177	0.4922	0.4931
(Baseline JET policy effect)	(388.5852)	(371.7877)	(0.3523)	(0.3340)
N	1,571	1,571	1,571	1,571
Office FE	YES	YES	YES	YES
Time FE	YES	YES	YES	YES

Note: The dependent variable is the number of employees in an office. T_{dt}^{base} is a dummy which switches on in a district when the previous JET policy was introduced. $T_{dt}^{base} * T_{dt}^{JOT}$ is a dummy that switches on in a district when JOT was introduced (this was introduced only in those areas that had switched to JET and at a later time period). The sample is a panel of 851 offices for the years 2004 and 2007. In columns 1 and 2, the dependent variable is the absolute number of employees in an office. In columns 3 and 4, the dependent variable is the log (number of employees) in an office. *** p<0.01, ** p<0.05, * p<0.1

Table 9: District Characteristics correlated with Team Size in a District

Dependent Variable: Number of Employees in a District		
	(1)	(2)
	Level	**Logs**
Percentage of working age population	56.2616***	0.1421***
	(17.7970)	(0.0521)
Total Lone Parent Families	-0.0053	-0.0000
	(0.0105)	(0.0000)
Percentage of Active Working Age population	-0.0014	0.0000***
	(0.0018)	(0.0000)
Percentage of fulltime employed	15.4307	-0.1194
	(28.6752)	(0.0837)
Percentage of part-time employed	-1.4997	-0.0162
	(34.2117)	(0.0998)
Percentage of self employed	46.8351	-0.1645*
	(30.9013)	(0.0902)
Percentage of unemployed	62.6983*	0.1808*
	(37.2059)	(0.1088)
Historical JSA caseload	0.0820**	-0.0000
	(0.0343)	(0.0001)
Historical IB caseload	0.0114	0.0001*
	(0.0112)	(0.0000)
Population Density	-1.3171	0.0072
	(2.5966)	(0.0076)
Percentage of Males	78.4820*	-0.1787
	(44.1164)	(0.1289)
N	338	338

Note: The dependent variable is the total number of employees in a district. T_{dt}^{base} is a dummy which switches on in a district when the previous JET policy was introduced. $T_{dt}^{base} * T_{dt}^{JOT}$ is a dummy that switches on in a district when JOT was introduced (this was introduced only in those areas that had switched to JET and at a later time period). The sample is a cross-section of 338 districts in 2001. The district variables are based on data from the Census 2001. In column 1, the dependent variable is the total number of employees in a district. In column 2, the dependent variable is the log (number of employees) in a district. *** p<0.01, ** p<0.05, * p<0.1

Table 10: Robustness to District Demographic Characteristics

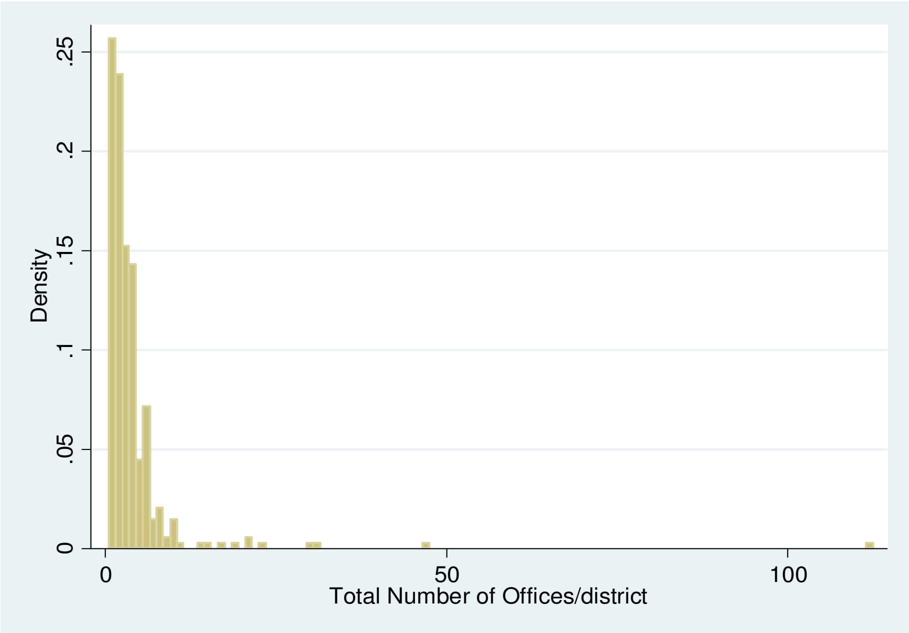
Dependent Variable: Log (Total Outflow)				
	(1)	(2)	(3)	(4)
	JSA	**JSA**	**IB**	**IB**
$\boldsymbol{T_{dt}^{base} * T_{dt}^{JOT} * \ln(S_d)}$	0.0145***	0.0167**	-0.0115**	-0.0253***
(Marginal effect of JOT policy by team size)	(0.0052)	(0.0082)	(0.0057)	(0.0094)
$\boldsymbol{T_{dt}^{base} * T_{dt}^{JOT}}$	-0.0436*	-0.7794***	0.0444	0.1057
(Marginal effect of JOT policy)	(0.0248)	(0.2573)	(0.0300)	(0.1825)
$\boldsymbol{T_{dt}^{base} * T_{dt}^{JOT} * \ln(WorkingAge_d)}$		0.0111***		-0.0022
		(0.0039)		(0.0034)
$\boldsymbol{T_{dt}^{base} * T_{dt}^{JOT} * \ln(Active_d)}$		-0.0000		-0.0000
		(0.0000)		(0.0000)
$\boldsymbol{T_{dt}^{base} * T_{dt}^{JOT} * \ln(Unemp_d)}$		0.0144***		0.0180*
		(0.0054)		(0.0103)
$\boldsymbol{T_{dt}^{base} * T_{dt}^{JOT} * \ln(Selfemp_d)}$		0.0088***		0.0025
		(0.0031)		(0.0034)
$\boldsymbol{T_{dt}^{base} * T_{dt}^{JOT} * \ln(IBclaimant_d)}$		0.0000		0.0000
		(0.0000)		(0.0000)
$\boldsymbol{lnU_{adt-1}}$	0.6388***	0.6291***	0.3499***	0.3662***
	(0.0117)	(0.0125)	(0.0349)	(0.0396)
$\boldsymbol{lnU_{a'dt-1}}$	0.0256**	0.0134	0.0328	0.0440
	(0.0106)	(0.0111)	(0.0265)	(0.0291)
N	22,109	20,129	20,416	17,994

Note: The size variable, S_d, used in these regressions is the total number of employees in a district. The dependent variable is the log of the outflow from benefit during a quarter-year. All regressions control for district by age fixed effects and age by time fixed effects. T_{dt}^{base} is a dummy which switches on in a district when the previous JET policy was introduced. $T_{dt}^{base} * T_{dt}^{JOT}$ is a dummy that switches on in a district when JOT was introduced (this was introduced only in those areas that had switched to JET and at a later time period). The sample is a panel of 338 districts from 1999Q4 to 2007Q4. In columns 1 and 2, the dependent variable is log (outflow) for JSA in an age group and district in a particular quarter t and includes controls for the stock of JSA in the previous quarter. In columns 3 and 4, the dependent variable is log (outflow) for IB in an age group and district in a particular quarter t and includes controls for the stock of IB in the previous quarter. The district variables are based on data from the Census 2001. Standard errors are clustered at the district level. *** p<0.01, ** p<0.05, * p<0.1

Appendix

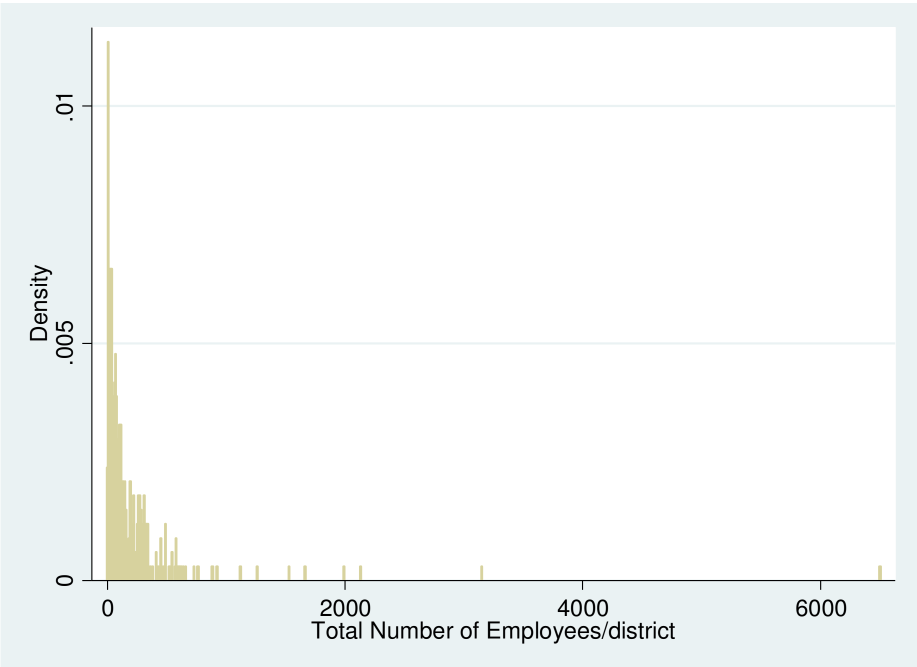
Appendix A – Tables and Figures

Figure A1: Distribution of Offices in Districts



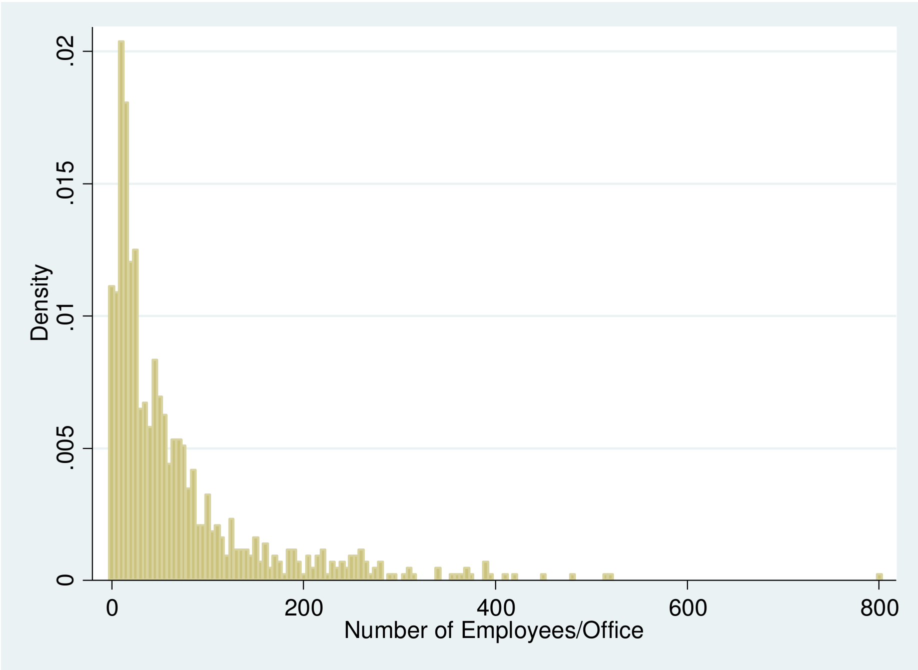
Note: This histogram plots the distribution of the number of offices per district in the 338 districts.

Figure A2: Distribution of Employees in Districts



Note: This histogram plots the distribution of the number of employees per district in the 338 districts.

Figure A3: Distribution of number of employees across Offices



Note: This histogram plots the distribution of the number of employees per office in the 851 offices.

Table A1: Job Entry Target points – Introduced in 2001

Client Group	Points Awarded
Disabled People and inactive benefits (IB)	12
Lone Parents (LP)	12
New Deal 50+, 25+, Young People	8
Other long term JSA	8
Short term unemployed JSA	4
Employed job-entries	1
Area-based points	1

Table A2: Robustness to Monthly Frequency Data for JSA

Dependent Variable: Log (Total Outflow)			
	(1) **JSA**	(2) **JSA**	(3) **JSA**
$\boldsymbol{T_{dt}^{base} * T_{dt}^{JOT} * \ln(S_d)}$ *(Marginal effect of JOT policy by team size)*			0.0171*** (0.0038)
$\boldsymbol{T_{dt}^{base} * T_{dt}^{JOT}}$ *(Marginal effect of JOT policy)*		0.0543*** (0.0075)	0.0088 (0.0181)
$\boldsymbol{T_{dt}^{base}}$ *(Baseline JET policy effect)*	-0.0334*** (0.0052)	-0.0600*** (0.0067)	-0.0609*** (0.0067)
$\boldsymbol{T_{dt}^{base} * \ln(S_d)}$ *(Baseline JET policy effect by team size)*			-0.0084*** (0.0019)
$\boldsymbol{T_{dt}^{JOT} * \ln(S_d)}$			-0.0018 (0.0026)
$\boldsymbol{lnU_{adt-1}}$	0.6551*** (0.0141)	0.6360*** (0.0144)	0.6496*** (0.0147)
$\boldsymbol{lnU_{a'dt-1}}$	0.0789*** (0.0115)	0.0903*** (0.0113)	0.0782*** (0.0111)
N	66,086	66,086	66,086

Note: The dependent variable is the log of the outflow from JSA during a month-year. All regressions control for district by age fixed effects and age by time fixed effects. T_{dt}^{base} is a dummy which switches on in a district when the previous JET policy was introduced. $T_{dt}^{base} * T_{dt}^{JOT}$ is a dummy that switches on in a district when JOT was introduced (this was introduced only in those areas that had switched to JET and at a later time period). The sample is a panel of 338 districts from September 1999 to December 2007. Standard errors are clustered at the district level. Standard errors are clustered at the district level. *** p<0.01, ** p<0.05, * p<0.1

Table A3: Pre-treatment Trends in Benefit Outflow – Outflow from offices

Dependent Variable: Log (Total Outflow)		
	(1)	(2)
	JSA	**JSA**
$\boldsymbol{T_{dt}^{base} * T_{dt}^{JOT}}$	0.0444**	0.0535**
(Marginal effect of JOT policy)	(0.0187)	(0.0247)
$\boldsymbol{T_{dt}^{base}}$	-0.0172***	-0.0187***
(Baseline JET policy effect)	(0.0066)	(0.0065)
$\boldsymbol{T_{dt-4}^{base} * T_{dt-4}^{JOT}}$		-0.0245*
		(0.0127)
$\boldsymbol{T_{dt-3}^{base} * T_{dt-3}^{JOT}}$		-0.0000
		(0.0220)
$\boldsymbol{T_{dt-2}^{base} * T_{dt-2}^{JOT}}$		0.0148
		(0.0194)
$\boldsymbol{T_{dt-1}^{base} * T_{dt-1}^{JOT}}$		0.0332
		(0.0249)
$\boldsymbol{lnU_{aot-1}}$	0.8039***	0.8037***
	(0.0232)	(0.0232)
$\boldsymbol{lnU_{a'ot-1}}$	0.1322***	0.1319***
	(0.0217)	(0.0217)
N	53,729	53,729
p-value		0.233

Note: The dependent variable is the log of the outflow from benefit during a quarter-year. All regressions control for office by age fixed effects and age by time fixed effects. T_{dt}^{base} is a dummy which switches on in a district when the previous JET policy was introduced. $T_{dt}^{base} * T_{dt}^{JOT}$ is a dummy that switches on in a district when JOT was introduced (this was introduced only in those areas that had switched to JET and at a later time period). The sample is a panel of 851 offices from 1999Q4 to 2007Q4. The p-value row contains the p-value for the F-test of the joint significance of the pre-treatment dummies. Standard errors are clustered at the district level. *** p<0.01, ** p<0.05, * p<0.1

Appendix B – Model

The output in a district is given by a concave production function: $f(e) = Ae^{\alpha}$ where e is the effort chosen by team members, $0 < \alpha < 1$ and A is a positive constant representing technology. The only input into the production function is the effort of the N employees in the team; $e = \sum_{i=1}^{N} e_i$. The cost of effort is represented by a quadratic function: $c(e_i) = \frac{e_i^2}{2}$.

Consider a district with *S* benefit offices in its jurisdiction. Each office has a different number of employees, given by $N_1, N_2 \dots . N_s$. The total number of employees across the *S* offices is *N* i.e.$N = N_1 + N_2 \dots . + N_s$. In other words, the total team size in the district is *N*.

B.1 Case of One Office (S=1)

The office social planner maximizes:

$$max_{e_{11}, e_{12} \dots e_{1N_1}} \frac{N_1}{N} A \left(\sum_{i=1}^{N_1} e_{1i} \right)^{\alpha} - \sum_{i=1}^{N_1} \frac{e_{1i}^2}{2}$$

Since there is a single office, $N = N_1$. The First Order Condition w.r.t e_{1i} is given by:

$$A\alpha \left(\sum_{i=1}^{N_1} e_{1i} \right)^{\alpha - 1} - e_{1i} = 0$$

$$A\alpha (N_1 e_{1i})^{\alpha - 1} - e_{1i} = 0 \text{ (by symmetric } e_{1i}'s)$$

$$e_{1i}^{*} = (A\alpha)^{\frac{1}{2-\alpha}} N_1^{\frac{\alpha-1}{2-\alpha}}$$

$$f(e) = A^{\frac{1}{2-\alpha}} \alpha^{\frac{\alpha}{2-\alpha}} N_1^{\frac{\alpha}{2-\alpha}}$$

And, $\frac{\partial f(e)}{\partial N_1} > 0.$

B.2 Case of Two Offices (S=2)

Suppose the district now has two offices, each with size N_1 and N_2. Therefore, $N = N_1 + N_2$.

Employees in the first office choose effort levels e_{1i} as the solution to the maximization problem facing its social planner:

$$max_{e_{11},e_{12}....e_{1\ 1}} \frac{N_1}{N} A\left(\sum_{i=1}^{N_1} e_{1i} + \sum_{i=1}^{N_2} e_{2i}\right)^{\alpha} - \sum_{i=1}^{N_1} \frac{e_{1i}^2}{2}$$

The social planner takes the effort choices of employees in the other office ($e_{2i}'s$) as exogenous in its decision-making process. The First Order Condition w.r.t e_{1i} is given by:

$$\frac{N_1}{N} A\alpha\left(\sum_{i=1}^{N_1} e_{1i} + \sum_{i=1}^{N_2} e_{2i}\right)^{\alpha-1} - e_{1i} = 0$$

$$\left(\sum_{i=1}^{N_1} e_{1i} + \sum_{i=1}^{N_2} e_{2i}\right) = e_{1i}^{\frac{1}{\alpha-1}} \left(\frac{N}{N_1 A\alpha}\right)^{\frac{1}{\alpha-1}} \tag{B1}$$

The maximization problem facing the social planner of the second office is:

$$max_{e_{21},e_{22}....e_{2N_2}} \frac{N_2}{N} A\left(\sum_{i=1}^{N_1} e_{1i} + \sum_{i=1}^{N_2} e_{2i}\right)^{\alpha} - \sum_{i=1}^{N_2} \frac{e_{2i}^2}{2}$$

The social planner takes the effort choices of employees in the other office ($e_{1i}'s$) as exogenous in its decision-making process. The First Order Condition w.r.t e_{2i} is given by:

$$\frac{N_2}{N} A\alpha\left(\sum_{i=1}^{N_1} e_{1i} + \sum_{i=1}^{N_2} e_{2i}\right)^{\alpha-1} - e_{2i} = 0$$

$$\left(\sum_{i=1}^{N_1} e_{1i} + \sum_{i=1}^{N_2} e_{2i}\right) = e_{2i}^{\frac{1}{\alpha-1}} \left(\frac{N}{N_2 A\alpha}\right)^{\frac{1}{\alpha-1}} \tag{B2}$$

Solving equations B1 and B2 yields $e_{2i} = \frac{N_2}{N_1} e_{1i}$. Substituting this into equation B1 gives equilibrium effort choice of e_{1i}:

$$e_{1i}^* = N_1 \left[\frac{A\alpha}{N}\right]^{\frac{1}{2-\alpha}} (N_1^2 + N_2^2)^{\frac{\alpha-1}{2-\alpha}}$$

Using $e_{2i} = \frac{N_2}{N_1} e_{1i}$, yields the equilibrium effort choice of e_{2i}:

$$e_{2i}^* = N_2 \left[\frac{A\alpha}{N}\right]^{\frac{1}{2-\alpha}} (N_1^2 + N_2^2)^{\frac{\alpha-1}{2-\alpha}}$$

Equilibrium production in the district is:

$$f(e^*) = \left[\frac{N_1^2 + N_2^2}{N}\right]^{\frac{\alpha}{2-\alpha}} A^{\frac{2}{2-\alpha}} \alpha^{\frac{\alpha}{2-\alpha}}$$

Next, I compute the comparative statics of the output w.r.t. N_1 , given a total team size of N.

$$\frac{\partial f(e)}{dN_1} = f(e) \left(\frac{\alpha}{2-\alpha}\right)\left[\frac{4N_1 - 2N}{(N_1^2 + (N - N_1)^2)}\right]$$

Now, $\frac{\partial f(e)}{dN_k} \geq 0$ if $(4N_1 - 2N) \geq 0$ i.e. $N_1 \geq \frac{N}{2}$. Thus, output is increasing the more the larger offices grow i.e. the more concentrated office size becomes.

Example

To get the intuition behind the impact of office size concentration, let me illustrate using the case of two districts – A and B - with just 2 offices. Suppose they both have the same average size $\mu = \frac{N}{2}$.

However, District A has two identical sized offices, each with μ employees. Then,

$$f(e^*)_A = \left[\frac{N}{2}\right]^{\frac{\alpha}{2-\alpha}} A^{\frac{2}{2-\alpha}} \alpha^{\frac{\alpha}{2-\alpha}}$$

And, district B has one large office with $N_1 (> N_2)$ employees, such that the mean office size in the district is still μ employees. To get the equilibrium output in district B, let $N_1 = B\mu$, where *B>1*. Equilibrium output was:

$$f(e^*) = \left[\frac{N_1^2 + (N - N_1)^2}{N}\right]^{\frac{\alpha}{2-\alpha}} A^{\frac{2}{2-\alpha}} \alpha^{\frac{\alpha}{2-\alpha}}$$

$$f(e^*)_B = \left[\frac{(B\mu)^2 + (2\mu - B\mu)^2}{2\mu}\right]^{\frac{\alpha}{2-\alpha}} A^{\frac{2}{2-\alpha}} \alpha^{\frac{\alpha}{2-\alpha}}$$

Now $f(e^*)_B \geq f(e^*)_A$ if $\left[\frac{(B\mu)^2 + (2\mu - B\mu)^2}{2\mu}\right] \geq \mu$ i.e. if *B>1*, which is true. Thus, output is higher in the more concentrated district B.

B.3 General Case of S Offices

Suppose the district now has *S* benefit offices in its jurisdiction. Each office has a different number of employees, given by $N_1, N_2 \ldots . N_s$. The total number of employees across the S offices is N i.e.$N = N_1 + N_2 \ldots . + N_s$. In other words, the total team size in the district is N.

Employees in the first office choose effort levels e_{1i} as the solution to the maximization problem facing its social planner:

$$max_{e_{11}, e_{12} \ldots e_{1N_1}} \frac{N_1}{N} A \left(\sum_{i=1}^{N_1} e_{1i} + \sum_{N_t = N_1; \neq N_1}^{N_s} \sum_{i=1}^{N_t} e_{ti} \right)^{\alpha} - \sum_{i=1}^{N_1} \frac{e_{1i}^2}{2}$$

The social planner takes the effort choices of employees in the other office as exogenous in its decision-making process. The First Order Condition w.r.t e_{1i} is given by:

$$\frac{N_1}{N} A\alpha \left(\sum_{i=1}^{N_1} e_{1i} + \sum_{N_t = N_1; \neq N_1}^{N_s} \sum_{i=1}^{N_t} e_{ti} \right)^{\alpha - 1} - e_{1i} = 0$$

$$\left(\sum_{i=1}^{N_1} e_{1i} + \sum_{N_t = N_1; \neq N_1}^{N_s} \sum_{i=1}^{N_t} e_{ti} \right) = e_{1i}{}^{\frac{1}{\alpha - 1}} \left(\frac{N}{N_1 A \alpha} \right)^{\frac{1}{\alpha - 1}} \tag{B3}$$

The maximization problem facing the social planner of the second office is:

$$max_{e_{21}, e_{22} \ldots e_{2N_2}} \frac{N_2}{N} A \left(\sum_{i=1}^{N_2} e_{2i} + \sum_{N_t = N_1; \neq N_2}^{N_s} \sum_{i=1}^{N_t} e_{ti} \right)^{\alpha} - \sum_{i=1}^{N_2} \frac{e_{2i}^2}{2}$$

The social planner takes the effort choices of employees in the other office as exogenous in its decision-making process. The First Order Condition w.r.t e_{2i} is given by:

$$\frac{N_2}{N} A\alpha \left(\sum_{i=1}^{N_2} e_{2i} + \sum_{N_t = N_1; \neq N_2}^{N_s} \sum_{i=1}^{N_t} e_{ti} \right)^{\alpha - 1} - e_{2i} = 0$$

$$\left(\sum_{i=1}^{N_2} e_{2i} + \sum_{N_t = N_1; \neq N_2}^{N_s} \sum_{i=1}^{N_t} e_{ti} \right) = e_{2i}{}^{\frac{1}{\alpha - 1}} \left(\frac{N}{N_2 A \alpha} \right)^{\frac{1}{\alpha - 1}} \tag{B4}$$

Similarly, in office k with N_k employees, the maximization problem facing the social planner of the k^{th} office is:

$$max_{e_{k1},e_{k2....}e_{N_k}} \frac{N_k}{N} A\left(\sum_{i=1}^{N_k} e_{ki} + \sum_{N_t=N_1;\neq N_k}^{N_s} \sum_{i=1}^{N_t} e_{ti}\right)^{\alpha} - \sum_{i=1}^{N_k} \frac{e_{ki}^2}{2}$$

The First Order Condition w.r.t e_{ki} is given by:

$$\frac{N_k}{N} A\alpha \left(\sum_{i=1}^{Nk_1} e_{ki} + \sum_{N_t=N_1;\neq N_k}^{N_s} \sum_{i=1}^{N_t} e_{ti}\right)^{\alpha-1} - e_{ki} = 0$$

$$\left(\sum_{i=1}^{N_k} e_{ki} + \sum_{N_t=N_1;\neq N_k}^{N_s} \sum_{i=1}^{N_t} e_{ti}\right) = e_{ki}^{\frac{1}{\alpha-1}} \left(\frac{N}{N_2 A\alpha}\right)^{\frac{1}{\alpha-1}} \quad \text{(B5)}$$

Solving equations B3, B4 and similarly for all the k offices, $e_{ki} = \frac{N_k}{N_1} e_{1i}$. Substituting this into equation B3 gives equilibrium effort choice of e_{1i}:

$$e_{ki}^* = N_k \left[\frac{A\alpha}{N}\right]^{\frac{1}{2-\alpha}} (N_1^2 + N_2^2 + \cdots. + N_s^2)^{\frac{\alpha-1}{2-\alpha}}$$

And, the equilibrium total output in the district is:

$$f(e^*) = \left[\frac{N_1^2 + N_2^2 + \cdots. + N_s^2}{N}\right]^{\frac{\alpha}{2-\alpha}} A^{\frac{2}{2-\alpha}} \alpha^{\frac{\alpha}{2-\alpha}}$$

The ratio $\frac{N_1^2+N_2^2+\cdots.+N_s^2}{N} = N * \left\{\frac{N_1^2}{N^2} + \frac{N_2^2}{N^2} + \cdots.. + \frac{N_s^2}{N^2}\right\}$ where $\left\{\frac{N_1^2}{N^2} + \frac{N_2^2}{N^2} + \cdots.. + \frac{N_s^2}{N^2}\right\} = Herfindahl\ Index$. Thus, $f(e^*) = P * (N * HI)^{\frac{\alpha}{2-\alpha}}$, where P is a positive constant.

Chapter 4: Fiscal consolidation during a depression

Abstract: In 2009-10, the UK's budget deficit was about 11 per cent of GDP. A credible plan for fiscal consolidation was introduced in the UK over the fiscal years 2011-12 to 2016-17. In this chapter, we assess the impact of the scale and timing of this fiscal consolidation programme on output and unemployment in the UK. During a prolonged period of depression when unemployment is well above most estimates of the NAIRU, the impact of fiscal tightening may be different from that in normal times. We contrast three scenarios: the consolidation plan implemented during a depression; the same plan, but with implementation delayed for three years when the economy has recovered; and no consolidation at all. The modelling confirms that doing nothing was not an option and would have led to unsustainable debt ratios. Under both our "immediate consolidation" scenario and the "delayed consolidation", the necessary increases in taxes and reductions in spending reduce growth and increase unemployment, as expected. But our estimates indicate that the impact would have been substantially less, and less long-lasting, if consolidation had been delayed until more normal times. The impact is partly driven by the heightened magnitude of fiscal multipliers, and exacerbated by the prolongation of their impact due to hysteresis effects. The cumulative loss of output over the period 2011-21 amounts to about £239 billion in 2010 prices, or about 16 per cent of 2010 GDP. And unemployment is considerably higher for longer - still 1 percentage point higher even in 2019.

Acknowledgements: We are grateful to Simon Kirby, who provided the details of the UK budget plans in Table 1 that underlie all the scenarios in this note, as well as the UK forecast baseline reported in Figures 9-11.

The financial crisis and resulting recession led to sharp rises in government deficits in almost all major industrialized countries, primarily because of falls in tax receipts. This was further increased by fiscal stimulus packages and emergency financial sector support. This in turn has led to a sharp rise in global government debt, giving rise to concerns about long-term fiscal sustainability. Despite this, long-term interest rates remain low in virtually all major developed economies outside the Euro Area, reflecting the fact that growth is weak and short-term interest rates are expected to remain low. However, many of the major economies have introduced fiscal tightening measures in recent years despite the widespread slowdown in GDP growth, and a level of GDP that remains well below that of 2007. The IMF estimates that the overall global fiscal position tightened by 1 per cent of GDP in 2011 (IMF, 2012a). Meanwhile, in the Euro Area, where countries can neither finance their deficits through quantitative easing nor adjust via the exchange rate, market pressures on some countries have been intense, and austerity programs have been introduced in a number of countries in an attempt to stem the rise in sovereign debt and ease the pressure on bond yields.

Although the long-term government borrowing rates are at historic lows in the UK, it is clearly the case that over the medium to long term fiscal consolidation is essential for debt sustainability. The UK has announced fiscal consolidation measures amounting to a total of 7.4 per cent of GDP over the fiscal years 2011–12 to 2016–17. Table 1 details the current plans by period and instrument.

In this chapter we assess the impact of the scale and timing of this fiscal consolidation programme on output and unemployment in the UK. We begin by using the National Institute's model, NiGEM, to analyze the impact of the ongoing policy on the UK economy using the standard version of the model, which would reflect the impact in 'normal' times. However, we do not appear to be in 'normal' times but in a prolonged period of depression, which we define as a period when output is depressed below its previous peak. As Delong and Summers (2012), Auerbach and Gorodnichenko (2012) and others point out, the impact of fiscal tightening during a depression may be different from that in normal times.

Table 1: Fiscal consolidation plans

Ex-ante, % of GDP

	2011-12	2012-13	2013-14	2014-15	2015-16	2016-17	Cumulative
Spending							
Consumption	-0.44	-0.76	-0.46	-0.78	-0.81	-0.34	-3.58
Investment	-0.27	-0.28	-0.36	-0.04	-0.22	0.00	-1.16
Transfers to households	-0.09	-0.20	-0.37	-0.19	-0.03	0.02	-0.85
Subsidies	-0.05	0.01	-0.02	0.00	-0.01	0.00	-0.07
Revenue							
Direct tax, households	0.10	0.40	0.20	0.33	-0.11	0.01	0.92
Direct tax, business	0.15	0.01	0.04	-0.12	-0.02	0.02	0.08
Indirect tax	0.70	0.00	0.09	0.03	-0.06	-0.02	0.76
Total	1.80	1.64	1.54	1.24	0.87	0.33	7.42

Note: Here we define the fiscal impulse as the ex-ante expected change in revenue/spending (as a % of GDP) as a result of announced policy changes. Tax credit policy changes are classified as changes to direct taxes in this analysis. The impact on GDP will depend on the fiscal multipliers in each country, and cannot be read directly from this table. The *ex-post* impact on government balances will depend on the response of GDP and the endogenous response of government interest payments, and so also cannot be read directly from this table.

There are a number of channels that the differences may feed through; for each we modify NiGEM to take account of the differential impacts. First, there is the interest rate response. Under normal circumstances a tightening in fiscal policy can be expected to be accommodated by a relaxation in monetary policy. However, with interest rates already at exceptionally low levels, further tightening of fiscal policy is unlikely to result in such an offsetting monetary policy reaction. While quantitative easing/credit easing measures have been introduced, the effects of these measures are also limited by low interest rates on 'risk-free' assets. It is less clear that monetary easing measures have a significant impact on the risk premia attached to assets that bear a greater risk of default.

Second, during a downturn, when unemployment is high and job security low, a greater percentage of households and firms are likely to find themselves liquidity constrained. This is likely to be particularly acute when the downturn is driven by an impaired banking system, as lending conditions will tighten beyond what would be expected in a normal downturn. There is less scope to smooth consumption in response to short-term income losses through an adjustment in savings.

Finally, long spells of depressed output and high unemployment can lead to 'hysteresis' which keeps the productive capacity of the economy persistently or even permanently lower (for example through the 'scarring' effect of unemployment which we discuss below). The economy may converge to the steady state levels of output and employment in the very long run, but in the medium term output levels could be substantially lower due to hysteresis effects. The time the economy takes to converge to the long-run steady state is also prolonged.

In this note we consider the potential impact on the economy, both in the short and long term, of postponing the planned consolidation measures that were introduced from 2011–12 onwards until the UK economy has emerged from the current period of depression and the output gap has narrowed significantly. While our analysis is not strictly dependent on the length of this delay, NiGEM-based estimates suggest that, in the absence of fiscal tightening, the output gap in the UK would be approaching balance by 2014. In the absence of deeper and more prolonged financial distress driven by events in the Euro Area, we would then have anticipated a 'normal' response to the fiscal consolidation measures after 2014, rather than the rather larger response that may result in the current period of depressed output and high unemployment.

In order to decompose the channels of transmission, we present four separate scenarios. In the first scenario, we illustrate the expected impact on output and employment of the fiscal programme detailed in table 1, had it been introduced in normal times, rather than during a period of depression. We then consider, one at a time, three channels that may differ during a period of depression: the impacts of an impaired interest rate channel; the impacts of heightened liquidity constraints; and the impacts of hysteresis, all of which exacerbate the impact of the consolidation programme on output and unemployment. In the final section, we construct a combined scenario that cumulates the effects of all three channels, and illustrates our estimate of the impact of the consolidation programme as it has been put forward, during a period of depression with limited downward flexibility in interest rates, heightened liquidity constraints and rising levels of long-term unemployment. We compare this to a scenario with no fiscal consolidation, and one where the same consolidation programme is introduced with a delay (2014–20), when the economy is expected to have returned to normal conditions. This allows us to estimate the cumulative impact that may be associated with the early introduction of the consolidation programme.

1 Scenario I - Impact of fiscal programme in normal times

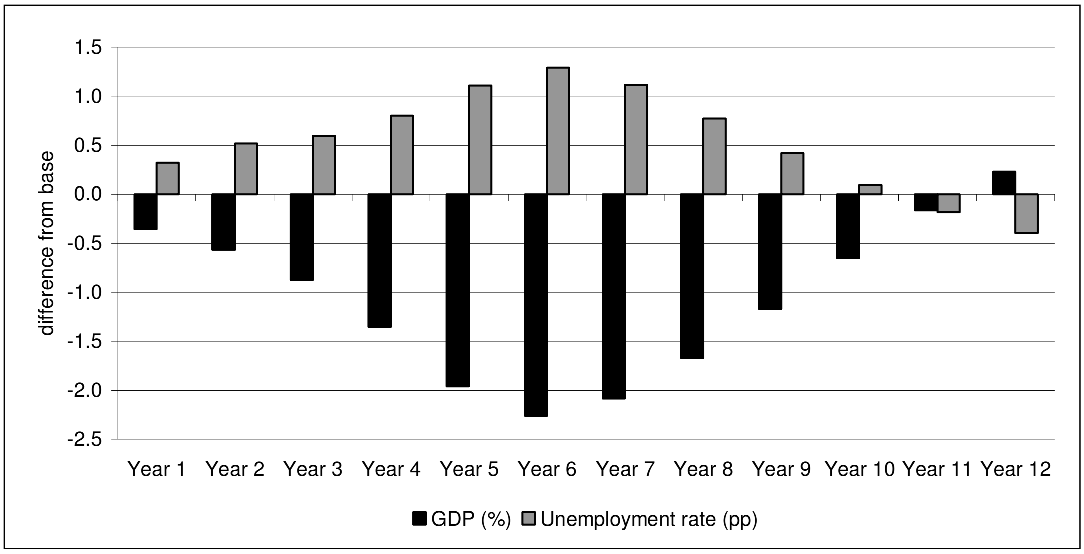
Fiscal multipliers[37] are not uniform either across countries (e.g. Ilzetzki *et al*., 2010), across time or across instruments (e.g. tax vs. spending). Barrell *et al.* (2012) provides an overview of NiGEM and compares estimates of fiscal multipliers across instruments for a set of seventeen OECD economies. In general, spending multipliers tend to be larger than tax multipliers in the first year, as tax adjustments are partially offset through savings and feed in more gradually. For the UK, they find a direct spending multiplier of about 0.5–0.7 per cent in the first year, while tax multipliers averaged about 0.1–0.2 per cent[38]. Much of the current consolidation plan is spending based, and so can be expected to have a more significant impact on GDP in the short term.

[37] The fiscal multiplier is generally defined as the expected impact on output in the first year, following a policy innovation that raises spending or cuts taxes by 1 per cent of GDP (*ex ante*).

[38] Fiscal multipliers tend to be less than 1, primarily due to import leakages, the anticipated monetary policy response, and an offset through the consumption channel through savings.

In Figure 1, we illustrate the impact on the level of GDP and the unemployment rate that we would expect in response to the current fiscal programme outlined in table 1, were it introduced in 'normal' times, e.g. when the output gap is close to balance and unemployment is close to its equilibrium level. We hold the exchange rate fixed in this scenario, as exchange rate behavior depends not just on the policy adopted in the UK, but on the relative stance of UK fiscal policy in a global context. Where many major economies are consolidating simultaneously, the assumption of a neutral impact on the exchange rate is probably justified. If the UK is tightening relatively more than its trading partners, we would expect to see a modest depreciation of the exchange rate, whereas if it is tightening relatively less than its partners the exchange rate would appreciate, holding all other risk factors constant.

Figure 1: Impact of fiscal consolidation in normal times



Note: Impact of policies described in table 1 on the level of GDP and the unemployment rate, if introduced when the output gap is close to 0 and the unemployment rate is close to its long-run equilibrium.
Source: NiGEM simulations

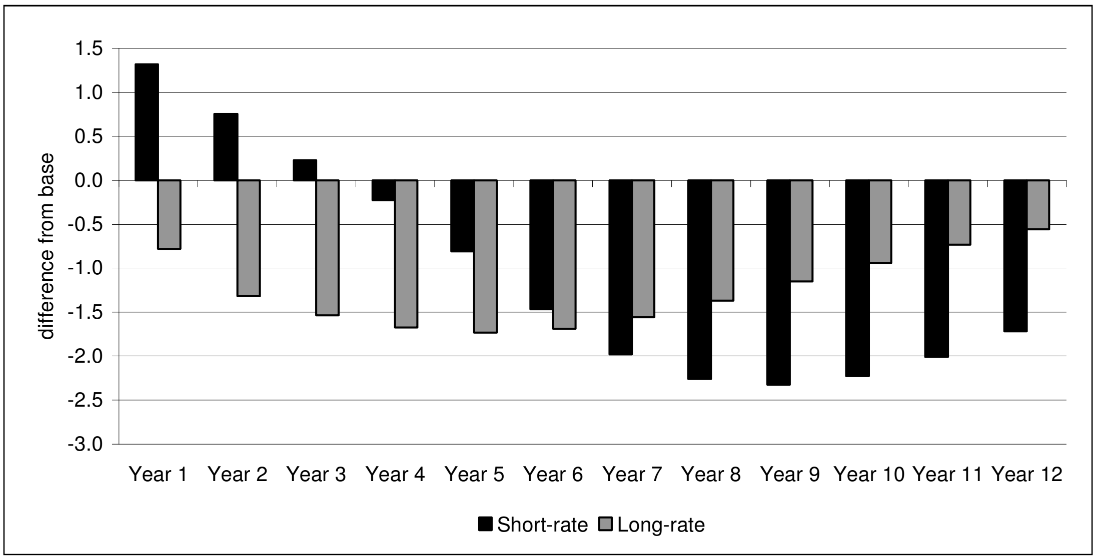
We would expect the level of output to decline by 0.4 per cent relative to the baseline in the first year, reaching a peak of 2.3 per cent below base after six years. Over the longer term, we would expect both GDP and unemployment to return to levels that would have been anticipated in the absence of fiscal consolidation. The normal cyclical behavior of the model suggests that output would rise slightly above base and unemployment fall slightly below base after year 11, although these effects would not persist over the longer term. The loss of government investment can be expected to have a negative impact on the productive capacity of the economy in the longer term, but these effects are relatively small. Unemployment is brought back towards base levels as output recovers, and through an adjustment in real wages.

In general, a fiscal tightening can be expected to be accompanied by a monetary loosening, as an inflation targeting central bank maintains a given inflation target with lower rates of interest. However, not all fiscal instruments have the same impact on inflation. One of the instruments employed in the fiscal consolidation programme outlined in table 1 is the indirect tax, or VAT, rate. A rise in the VAT rate will initially put upward pressure on inflation, as it is a direct shock to the price level. This may induce an inflation targeting central bank to raise interest rates in the short term. After the first year or so, the jump in the price level would fall out of the inflation rate, and we would expect inflation to be somewhat below what it would have been in the absence of the VAT rise, allowing a lower interest rate over the medium term.

Our preliminary scenario reflecting the response in 'normal' times allows an endogenous response in short-term interest rates[39]. In normal times, the fiscal programme described in table 1 would initially put upward pressure on interest rates, as the indirect tax rate rises by 250 basis points, with a direct impact on inflation in the first year of the shock. As the effects of the VAT rise dissipate, this is followed by an extended period of short-term policy interest rates below base. With forward-looking financial markets, the long-term interest rate, which determines the borrowing costs of firms for investment, is driven by the expected path of short-term interest rates over a 10-year forward horizon. As such, despite the initial rise in the short-term rates, long-term interest rates fall immediately, stimulating investment and offsetting part of the fiscal contraction. The expected impact on short-term and long-term interest rates in response to the policy, were it to be introduced during 'normal' times, is illustrated in Figure 2. Long-term interest rates would be expected to fall by about 150 basis points for an extended period, allowing a strong boost to investment.

[39] The policy rule followed is the standard two-pillar rule in NiGEM, which is described in Barrell *et al.* (2012).

Figure 2: Impact of fiscal consolidation on interest rates in normal times

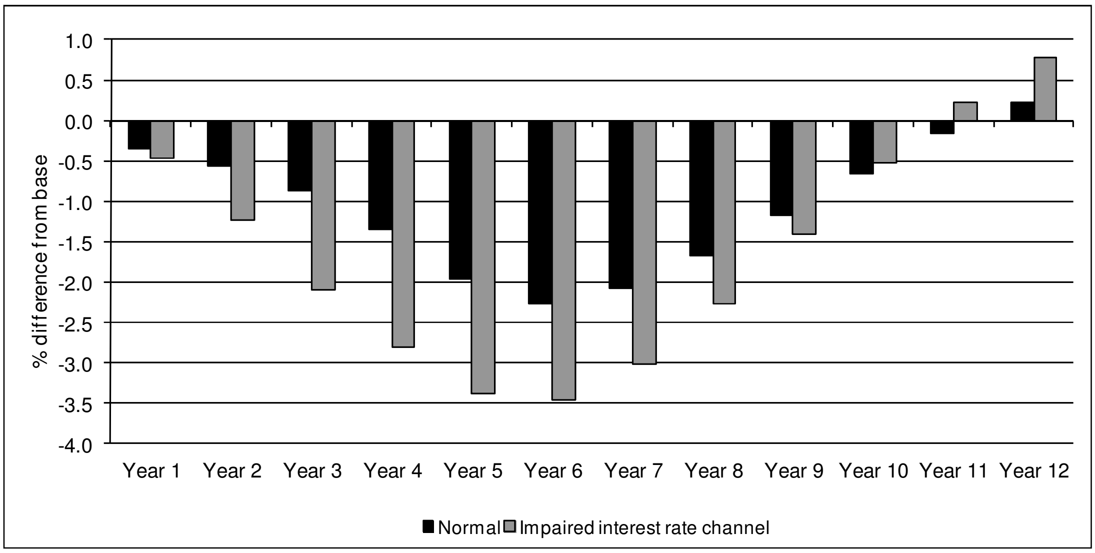


Note: Impact of policies described in Table 1 on interest rates, if introduced in 'normal' times. Short-term interest rates are determined by a central bank policy rule that targets inflation; long-term interest rates allow for 'rational' or out-turn consistent expectations in financial markets.
Source: NiGEM simulations

2 Scenario II - Impact of fiscal programme in a depressed economy (Impaired interest rate channel)

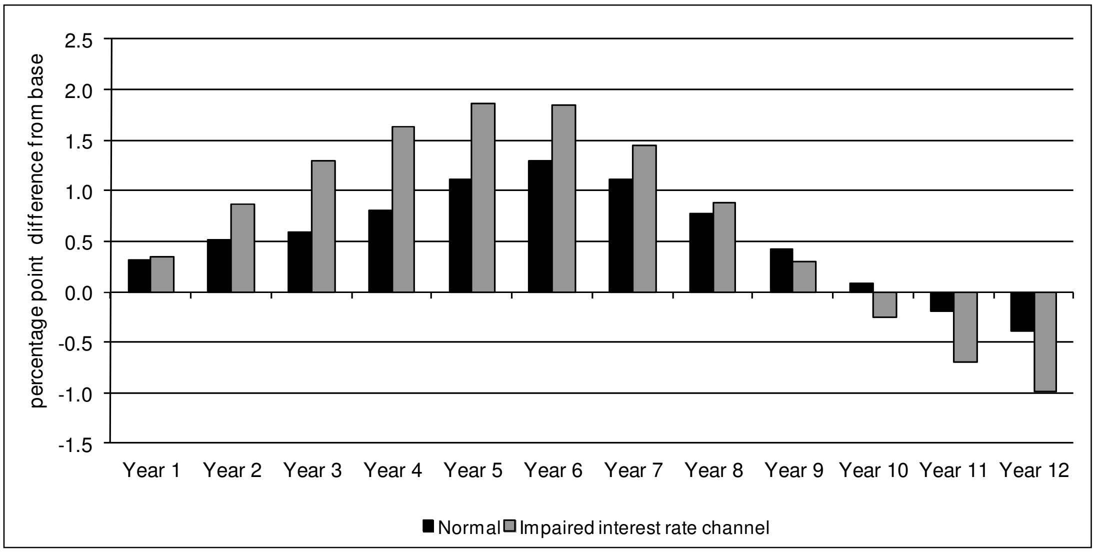
In the previous section we considered the impact of a fiscal consolidation in normal times, and demonstrated that, under normal circumstances, the consolidation programme detailed in table 1 would be expected to reduce long-term interest rates by about 150 basis points for several years. However, when interest rates are close to zero, their downward flexibility may be restricted (the 'zero lower bound'). With no offsetting stimulus from lower interest rates, the impact of the fiscal consolidation programme on GDP would be somewhat higher. Ten-year government bond yields in the UK are not at zero, but are exceptionally low, suggesting that there may be little scope for further reductions. If we hold long-term interest rates fixed, rather than allowing them to decline as in the first scenario, the negative effects on output and unemployment would be amplified. Figures 3 and 4 compare the impact on GDP and the unemployment rate under normal times with an endogenous interest rate response, to the same consolidation programme in an environment where there is no downward flexibility of interest rates. The impact on GDP would be about 1½ per cent greater after four years if the interest rate adjustment channel is impaired, while the unemployment rate would be expected to rise by a further ¾ percentage point.

Figure 3: Impact of an impaired interest rate adjustment on GDP



Note: Impact on the level of GDP from Figure 1 and under the same scenario with the interest rate adjustment impaired.
Source: NiGEM simulations

Figure 4: Impact of an impaired interest rate adjustment on unemployment rate



Note: Impact on the unemployment rate from Figure 1 and under the same scenario with the interest rate adjustment impaired.
Source: NiGEM simulations

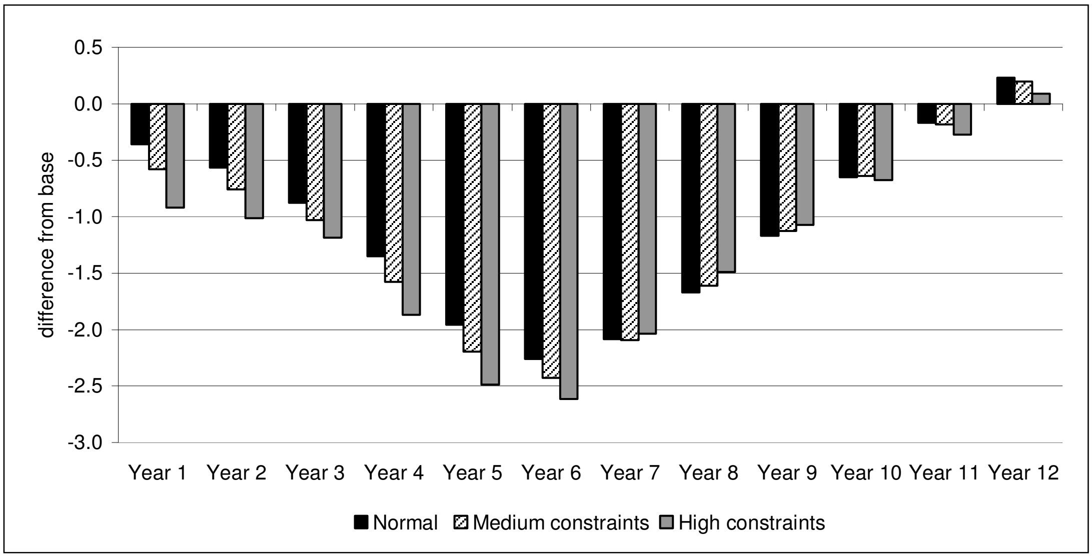
3 Scenario III - Heightened liquidity constraints

In the presence of perfect capital markets and forward-looking consumers with perfect foresight, households will smooth their consumption path over time, and consumer spending will be largely invariant to the state of the economy or temporary fiscal innovations. In the extreme example of a fully Ricardian world, the fiscal multiplier is effectively zero, as fiscal policy will simply be offset by private sector adjustments to savings behavior. However, at any given time, some fraction of the population and of firms is liquidity constrained; that is, they have little or no access to borrowing, so that their current spending is largely restrained by their current income. In the first scenario, we make the assumption that savings behavior and the number of liquidity constrained consumers and businesses are as in normal times. However, in a prolonged period of depressed activity, this is unlikely to be the case, especially when the downturn has at its roots an impaired banking system. In this section we consider the effects of an increase in the share of consumers and firms that are liquidity constrained. We operationalize this effect in the NiGEM model through an adjustment to the short-term income elasticity of consumption and investment. If liquidity constraints are not important, households and firms can borrow when incomes or profits are low in order to smooth their spending path. In this case, the path of consumption and investment will be less sensitive to short-term fluctuations in income or profits. However, when liquidity constraints are high, there is less scope to borrow to smooth spending, and consumption and investment will be much more reliant on current revenue streams. A detailed illustration of the sensitivity of the scenarios to assumptions on the short-term income elasticity parameters is given in the Appendix.

In the standard version of NiGEM, the short-term income elasticity of consumption in the UK is given by 0.17, suggesting a relatively low level of liquidity constraints. Barrell, Holland and Hurst (2012) put this into an internationally comparative context, which suggests that UK liquidity constraints are on the low side, but not out of line with other advanced economies. The short-term elasticity of investment to GDP is between 1 and 2 per cent, with business investment more sensitive to the state of the economy than housing investment.

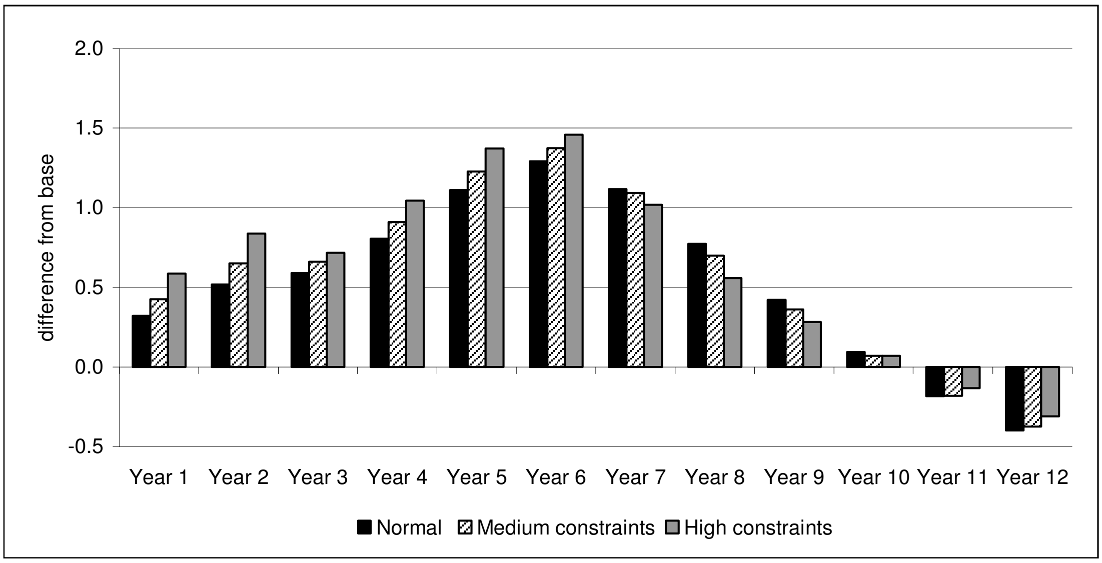
We now consider the impact on output and unemployment that we would expect when liquidity constraints are heightened. Figures 4 and 5 illustrate the expected impact on output and the unemployment rate of the consolidation programme detailed in table 1 if it were introduced in ‘normal’ times (scenario 1), and compares this to a scenario with moderately heightened liquidity constraints (model 4 in the Appendix) and high liquidity constraints (model 7 in Appendix table A1). The moderate scenario can be interpreted as representing an environment where the number of liquidity constrained consumers is roughly double that in normal times, while the high liquidity constraints scenario reflects an environment where the number of liquidity constrained consumers is twice that in the moderate scenario. In all three scenarios we allow an endogenous adjustment of both short-term and long-term interest rates. Under high liquidity constraints, we would expect output to decline by ½ per cent more in the first year than it would in normal times. The unemployment rate can be expected to increase by 0.25 percentage points more in the first year compared to the first normal times scenario. By year 7, the differences between the three scenarios are largely eliminated.

Figure 5: Impact of liquidity constraints on GDP



Note: Impact of policies described in table 1 on GDP, if introduced in 'normal' times, and with heightened liquidity constraints. See models 4 and 7 in the Appendix for details on the parameter assumptions.
Source: NiGEM simulations

Figure 6: Impact of liquidity constraints on unemployment rate



Note: Impact of policies described in table 1 on unemployment rate, if introduced in 'normal' times, and with heightened liquidity constraints. See models 4 and 7 in the Appendix for details on the parameter assumptions.
Source: NiGEM simulations

4 Scenario IV - Presence of hysteresis

Extended periods of depressed output and high unemployment can have long-term implications for the productive capacity of the economy. A host of mechanisms could be responsible for these hysteresis effects. These include reduced capital investment, premature capital scrapping, reduced labor force attachment on the part of the long-term unemployed resulting in lower wage pressures, scarring effects on young workers who have trouble beginning their careers and changes in managerial attitudes. In particular, the incidence of long-term unemployment may reduce the downward pressure on wages exerted by a high general unemployment rate and thus lead to unemployment hysteresis or persistence long after the shocks have dissipated. We focus on this labor market channel of hysteresis in this chapter. This does not mean that the other potential channels of hysteresis are unimportant[40].

A potential explanation of hysteresis effects is that a decrease in aggregate demand initially causes a rise in short-term unemployment, but this turns into long-term unemployment if the depression continues. As the survival rate (in

[40] IMF's recent report, 'United Kingdom 2012 Article IV Consultation', IMF Country Report No. 12/190, also focuses on the labor market channel of hysteresis to explain changes in the NAIRU. It stresses a slightly different channel, namely, labor employment protection laws as the driver of hysteresis impacts.

unemployment) for the long-term unemployed is higher[41], they put less downward pressure on wages and inflation and so can contribute to the persistence of unemployment into the medium term. Machin and Manning (1999) model this in an efficiency wage framework. Similar results are found in Blanchard and Diamond (1994) in a matching model context, Calmfors and Lang (1995) and Manning (1993) in the context of a union bargaining model. Thus, high long-term unemployment has been argued to be a cause of high unemployment itself. However, it is still possible that the unemployment rate returns to its steady state NAIRU in the very long run.

Alternatively, it is highly likely that the long-term unemployed may cease to participate in the labor market altogether. There is sparse evidence on the decline in participation rate of those who have been unemployed for a prolonged period. More recently, it has been observed that in the US, the labor force participation rate plummeted during the Great Recession. It declined from a peak of 66.5 per cent in 2007 to 62 per cent in 2012[42]. The demographic trend relating to the retirement of the 'baby boom' generation, which has been ongoing since the turn of the century, is a slow-moving generational trend and cannot explain this substantial recent decline. This seems to suggest that this decline is at least in part a result of the labor market pressures arising from the 2008 crisis[43]. By contrast, in the UK, labor force participation has held up relatively well compared with previous recessions, although long-term unemployment has risen sharply.

The standard model for wages within NiGEM is based around a profit maximizing condition that sets the marginal product of labor equal to the real wage. The price and wage equations are determined by the first order profit maximizing conditions. Using a CES-style of production function, this can be described as:

$$ln\left(\frac{w}{p}\right) = \alpha + \frac{1}{\sigma} ln\left(\frac{ycap}{l}\right) - \frac{1-\sigma}{\sigma} techl \tag{1}$$

[41] Comparing the short-term and long-term unemployed, evidence shows that the outflow rates for the long-term unemployed have always been lower than that for the short-term unemployed. The lower outflow rate for the long-term unemployed, compared to the short-term unemployed, is called negative duration dependence. The most natural interpretation is that the long-term unemployed have a lower chance of finding a job.

[42] Authors' calculations based on data from The U.S. Bureau of Labor Statistics.

[43] Holland (2012) assesses the impact of labour force withdrawal in the US on potential output.

where $\frac{w}{p}$ is the real wage, $ycap$ is potential capacity output, l is labour input, $techl$ is labour augmenting technical progress, σ is the elasticity of substitution between labour and capital, and, α is a constant term.

This forms the long-run relationship and the firm side of the wage bargain. The unemployment rate acts as the bargaining instrument to bring labor demand in line with labor supply. We embed this into a dynamic equation of the form:

$$\Delta \ln(w) = \omega_1 + \omega_2 \left\{ ln\left(\frac{w}{p}\right)_{-1} - \frac{1}{\sigma} ln\left(\frac{ycap}{l}\right)_{-1} + \frac{1-\sigma}{\sigma} techl_{-1} \right\} + \omega_3 \Delta \ln(p) +$$
$$(1-\omega_3)\Delta \ln(p^e) + \omega_4(U_{-1}) \qquad (2)$$

where U is the unemployment rate, Δ is the difference operator, $\omega_1 - \omega_4$ are parameters and superscript *e* denotes expectations.

When the unemployment rate rises, this puts downward pressure on real wage growth. Firms can then afford to employ more workers, which brings labor demand in line with labor supply, and pushes unemployment back towards its equilibrium.

Arguably, those who have been unemployed for an extended period of time begin to search for work less intensively, or because of 'scarring' effects on skills or motivation, may simply not be regarded as suitable potential workers by employers. They may thus exert less pressure on wages than those who have been unemployed for only a short period. A more sophisticated model would, therefore, differentiate the unemployed by their duration out of work, and allow the wage elasticity to decline as the duration rises. In order to allow for this form of hysteresis we consider what we define as the long-term unemployed (LTU) – those who have been unemployed for twelve months or longer – separately from total unemployment.

It is difficult to identify empirically differences in the wage elasticities of different groups of unemployed, given the very strong correlation among the duration groups and unobserved heterogeneity between groups. In order to calibrate the differences in wage pressure, we draw on the study by Elsby and Smith (2010), who calculate the unemployment-to-employment transition rate by duration for the UK (see Figure 9, p. R35 in Elsby and Smith, 2010). Those unemployed for longer face markedly lower job-finding rates. Job seekers with more than twelve months duration find jobs at an average rate of just over 4 per cent per month, whereas the total pool of unemployed find jobs at an average rate of 10 per cent per month, using a sample that covers 1992–2010. This would suggest that long-term unemployed exert about 60 per cent less pressure on wages than the total pool of unemployed.

We, thus, construct an augmented wage equation, which incorporates wage-bargaining that is less sensitive to the long-term unemployment rate, using an equation of the form:

$$\Delta \ln(w) = \omega_1 + \omega_2 \left\{ ln\left(\frac{w}{p}\right)_{-1} - \frac{1}{\sigma} ln\left(\frac{ycap}{l}\right)_{-1} + \frac{1-\sigma}{\sigma} techl_{-1} \right\} + \omega_3 \Delta \ln(p) + (1-\omega_3)\Delta \ln(p^e) + \omega_4(U_{-1}) - 0.6\omega_4(LTU)_{-1}$$

(3)

where LTU is the long-term unemployment rate. We assume $\omega_4 < 0$ to reflect the bargaining process.

Some older studies, for example Nickell (1987), find a somewhat stronger feedback from LTU to wages. The sample used for estimation in his paper covers 1953–83, and so may be less relevant for today, given the significant changes to the labor market that have occurred since 1979. Nonetheless, we consider an alternative scenario, where the long-term unemployed have essentially stopped searching altogether, and so put no pressure on wages:

$$\Delta \ln(w) = \omega_1 + \omega_2 \left\{ ln\left(\frac{w}{p}\right)_{-1} - \frac{1}{\sigma} ln\left(\frac{ycap}{l}\right)_{-1} + \frac{1-\sigma}{\sigma} techl_{-1} \right\} + \omega_3 \Delta \ln(p) + (1-\omega_3)\Delta \ln(p^e) + \omega_4(U_{-1} - LTU_{-1})$$

(4)

This can be viewed as an upper limit to the potential effects through this channel. However, it should not be interpreted as an upper limit to the effects of hysteresis overall. Hysteresis may set in earlier than we allow for here – for example after six months rather than after twelve months. And the potential for labor market withdrawal could lead to significantly more prolonged effects on the productive capacity of the economy.

The impact of LTU on wages will also depend on how we model the rate of long-term unemployment itself. OECD (2009) estimates a simple relationship between the total unemployment rate and the long-term unemployment rate. For the UK, the relationship they identify is:

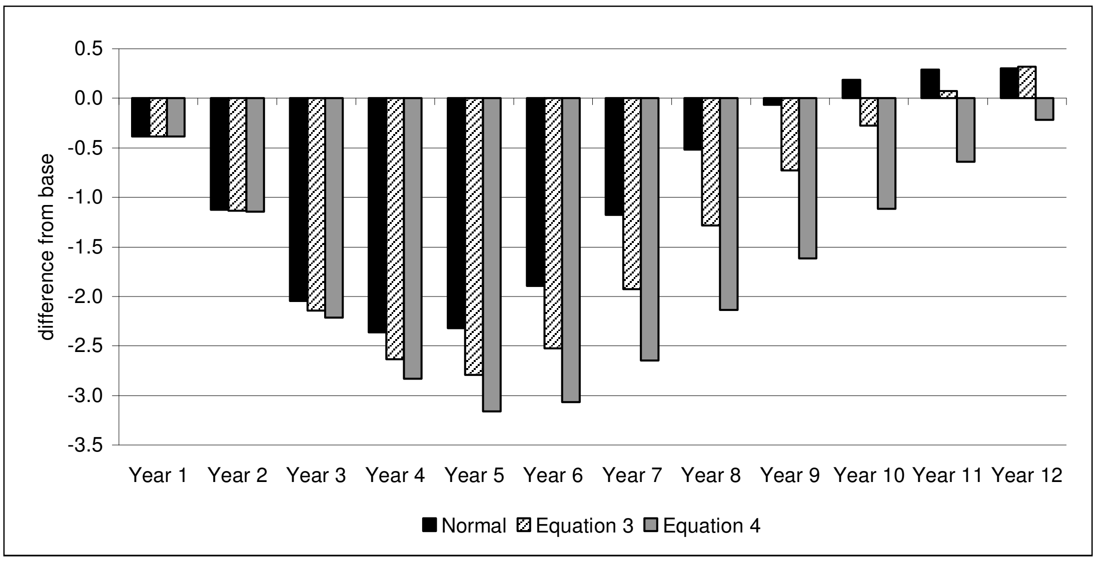
$$LTU = 0.76\, LTU_{-1} - 0.29\, LTU_{-2} + 0.34\, U \quad (5)$$

We use this relationship, rewritten in error correction format, to model LTU in the revised NiGEM model. The equation can be written as:

$$\Delta LTU = 0.29\, \Delta LTU_{-1} + 0.34\, \Delta U - 0.53\, \{LTU_{-1} - 0.6\, U_{-1}\} \quad (6)$$

Figures 7 and 8 illustrate the expected impact on output and the unemployment rate in the presence of labor market hysteresis effects, and compares our 'normal times' scenario to the two augmented wage equations discussed above – where the long-term unemployed exert 60 per cent less pressure on wages than shorter-term unemployed, and where the long-term unemployed exert no pressure on wages. In order to decompose the effects, we assume the interest rate channel is not impaired and liquidity constraints are not important. An important point of comparison with the baseline (scenario 1) is the much slower speed with which output returns to supply equilibrium; in other words, hysteresis not only magnifies the negative impacts of fiscal consolidation on output and employment, but means that they are much more long-lasting.

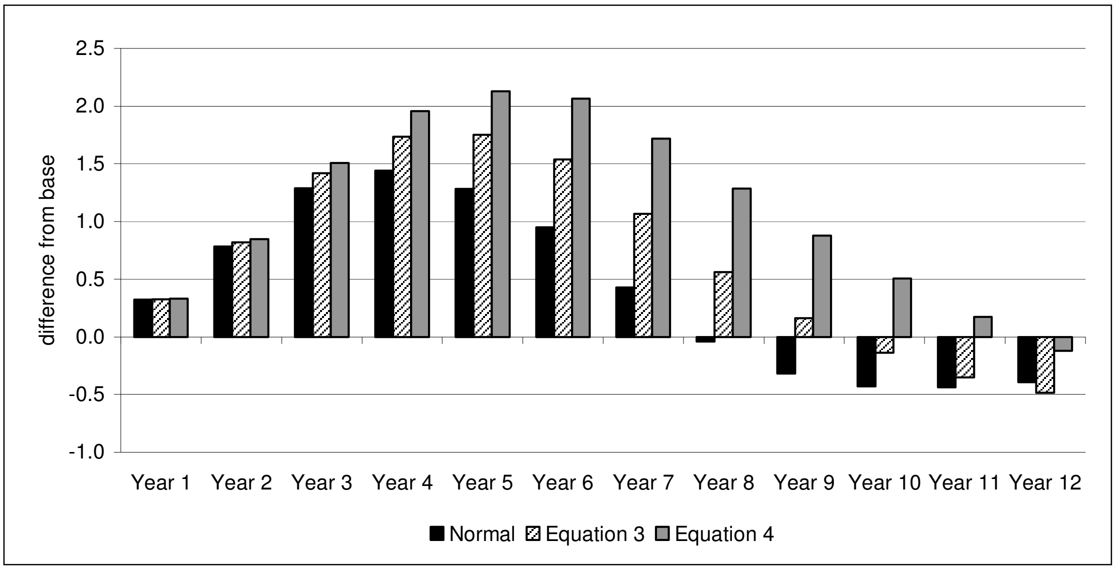
Figure 7: Impact on GDP



Note: Impact of policies described in table 1 on GDP, if introduced in 'normal' times under the standard version of NiGEM and with the augmented wage equations (3) and (4) described above.

Source: NiGEM simulations

Figure 8. Impact on unemployment rate



Note: Impact of policies described in table 1 on unemployment rate, if introduced in 'normal' times under the standard version of NiGEM and with the augmented wage equations (3) and (4) described above.
Source: NiGEM simulations

By introducing tightening during a period of high unemployment and large output gap, the negative impacts of the consolidation programme can be expected to persist for 2–4 years longer than they would have if the policy had been postponed until the level of unemployment had reverted to its long-run equilibrium.

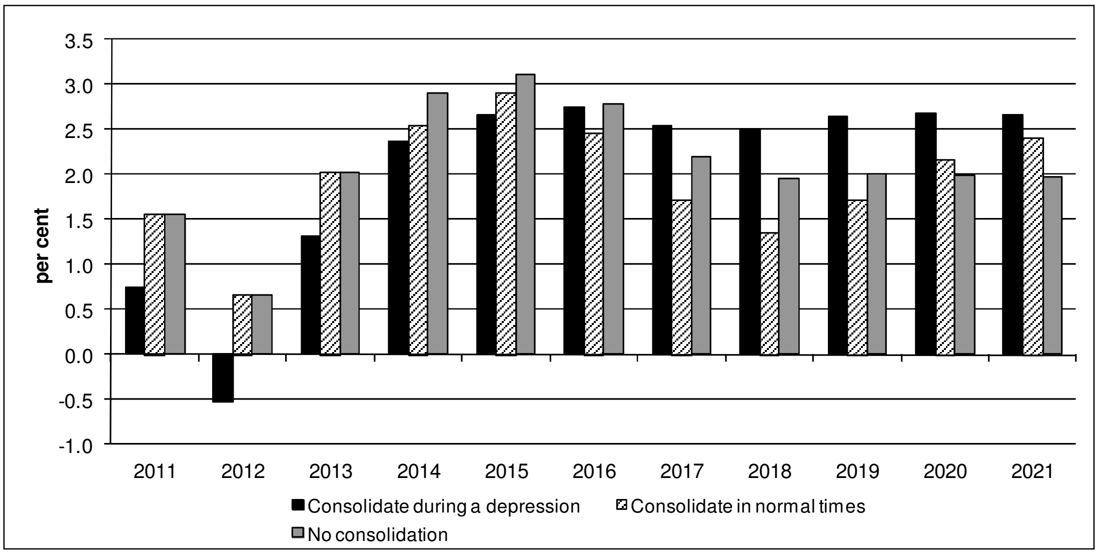
5 Cumulative impacts

Based on the results of the scenarios presented above, we can calibrate an estimate of the cumulative impacts on the economy from introducing fiscal tightening starting in 2011, rather than postponing the measures until output and unemployment had recovered from the downturn. The impact is partly driven by the heightened magnitude of fiscal multipliers, and exacerbated by the prolongation of their impact due to hysteresis effects. As an illustrative scenario, we assume that the interest rate response is impaired, with no adjustment in the long-term interest rate. We allow for moderately high liquidity constraints, so assume that the number of liquidity constrained agents is roughly double what it is in normal times (model 4 in the

Appendix), and model wages as in equation 3 above, with the long-term unemployed exerting 60 per cent less pressure on wages than total unemployment. Changing this set of assumptions could lead to a stronger or weaker impact on the economy than shown here, as demonstrated by the sensitivity of the results to the scenarios reported above.

Figures 9–11 illustrate projections for GDP growth, the unemployment rate and government debt as a ratio to GDP that we would anticipate under three different scenarios. The first reflects our assessment of the fiscal consolidation programme for 2011–17 as reported in table 1, introduced during the current environment of a depressed economy with moderately high liquidity constraints. This is consistent with the baseline forecast for the UK presented in this *Review*, and we designate this scenario as 'consolidate during a depression'. The second scenario illustrates the path that we would have expected had the consolidation programme been delayed until economic recovery was well underway, which model-based estimates suggest would have been by about 2014 in the absence of early fiscal tightening. The programme detailed in table 1 is implemented, but the timing is shifted so that it is enacted over the period 2014–20, with no consolidation measures introduced 2011–14. We designate this scenario as 'consolidate during normal times'. Finally we illustrate a scenario that shows the economic path that would have been expected in the absence of any consolidation programme, which we designate as 'no consolidation'. Scenarios 2 and 3 are identical for the first three years.

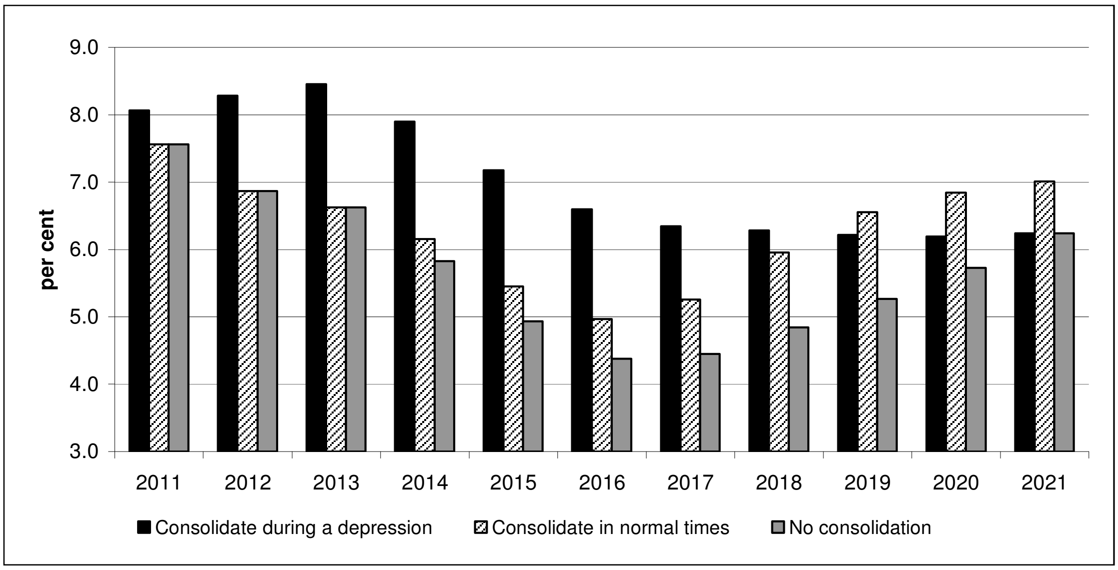
Figure 9: GDP growth under three consolidation scenarios



Note: Consolidation starting in 2011 during a depression, consolidation starting in 2014 when the economy has returned to 'normal', no consolidation.

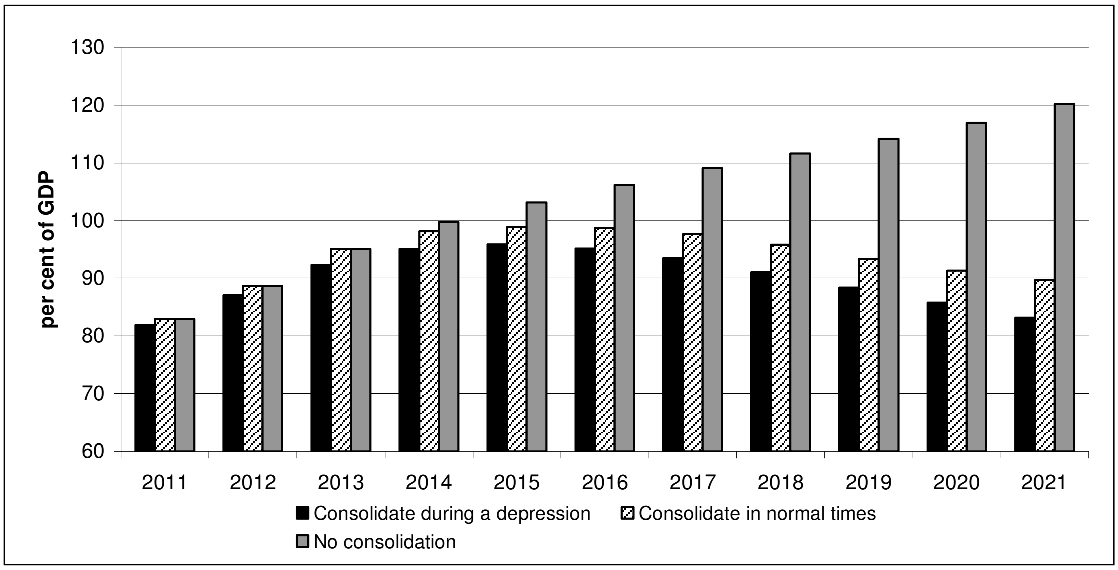
Source: NiGEM simulations

Figure 10: Unemployment rate under three consolidation scenarios



Note: Consolidation starting in 2011 during a depression, consolidation starting in 2014 when the economy has returned to 'normal', no consolidation.
Source: NiGEM simulations

Figure 11: Government debt under three consolidation scenarios



Note: Consolidation starting in 2011 during a depression, consolidation starting in 2014 when the economy has returned to 'normal', no consolidation.
Source: NiGEM simulations

A number of studies have looked at the links between the risk premium on government borrowing and fiscal sustainability, captured by current or expected values of the general government deficit or the stock of government debt (Laubach, 2009; Baldacci and Kumar, 2010; Schuknect et al, 2010; Bernoth and Erdogan, 2012 and others). These studies suggest that rising government debt is likely eventually to put upward pressure on interest rates, so that fiscal tightening is likely to be necessary at some point. Figure 11 indeed illustrates that in the absence of any fiscal tightening, the stock of government debt would have been on a steadily rising and almost certainly unsustainable path over the next decade. The option not to consolidate at all, therefore, was and is not a viable one. However, the differences between the debt profiles reflecting early consolidation and delayed consolidation are relatively modest, and the likely impact on interest rates is therefore small. Empirical estimates, on average, point to a 2–4 basis point rise in interest rates for a 1 per cent of GDP rise in the government debt to GDP ratio. A 10 percentage point differential, therefore, would be expected to induce at most a 40 basis point rise in borrowing costs. Even this may overstate the impacts for non-Euro Area countries. IMF (2012b) points out that, "fiscal indicators such as deficit and debt levels appear to be weakly related to government bond yields for advanced economies with monetary independence".

The scenarios suggest that the recession in 2012 could have been avoided had fiscal tightening measures been delayed. Table 2 details the differences between the two scenarios in level terms. Our estimates indicate that the cumulative loss of output from early consolidation accumulated over the period 2011–21 amounts to £239 billion in constant 2010 prices. This is equivalent to 16½ per cent of 2010 GDP (or about 1.3 per cent of total output over the entire period). These losses are sustained despite the fact that the growth rate of GDP is expected to be higher after 2016 under the early consolidation scenario compared to the delayed consolidation scenario, as consolidation measures in the latter are ongoing until 2020. In the long run, the level of GDP in the three scenarios should converge to a common level. Figure 1 indicates that the negative impact on output of the fiscal consolidation programme initiated in normal times can be expected to dissipate by eleven years after the onset of the programme, so that by 2025 the growth rate of GDP should converge in all three scenarios. A substantial permanent deadweight loss associated with the early consolidation programme will persist, as the amplified losses in the early years will not be fully offset by amplified gains once recovery sets in.

Similarly, the unemployment rate is expected to be higher until 2018 under the early consolidation programme than it would have been with a delayed fiscal tightening, as shown in Figure 10. In the long run, the level of the unemployment rate can be expected to converge to the same level in all three scenarios. It may take 10–11 years for these effects to feed through. The 'consolidate in a depression' scenario sees the unemployment rate falling below that of the 'consolidation in normal times' scenario over the period 2019–21. This reflects the fact that the delayed consolidation programme comes to an end only in 2020, whereas in the early consolidation scenario the recovery has been ongoing for three years, and the differences can be expected to dissipate by 2024. More importantly, the unemployment rate in the delayed scenario would never be expected to exceed 7 per cent.

Table 2: GDP in £billion, 2010 prices under two scenarios

	Consolidate during a depression	Consolidate in normal times	Difference	% 2010 GDP
2011	1478	1489	11	0.8
2012	1476	1505	29	2.0
2013	1495	1535	40	2.7
2014	1531	1575	44	3.0
2015	1572	1622	49	3.4
2016	1614	1660	45	3.1
2017	1654	1686	33	2.2
2018	1694	1708	14	1.0
2019	1738	1737	-1	-0.1
2020	1785	1775	-10	-0.7
2021	1832	1817	-15	-1.1
Sum 2011-2021	17869	18109	239	16.3

Source: NiGEM simulations

6 Conclusion

The concern today is that the Great Recession starting in 2008 and the consequent early fiscal tightening policies may lead to significant losses in output and a protracted period of high unemployment. The analysis presented in this note indicates that these concerns are well-founded. Under current policy plans the unemployment rate is expected to remain above 7 per cent until 2016. Had tightening measures been delayed until economic recovery was well underway, cumulative output on the period 2011–21 would have been significantly higher, and the unemployment rate would have been expected to rise no higher than 7 per cent over the next decade. In light of the above results, it can be argued that fiscal policy choices have to be considered in the light of the monetary policy response function. When monetary policy is constrained by the zero lower bound on interest rates, the impact of fiscal policy (the fiscal multiplier) will be magnified compared to *normal* times. The health of the banking sector is also an important determining factor. When unemployment is high or job security low, a greater percentage of households and firms are likely to find themselves liquidity constrained. This is likely to be particularly acute when the downturn is driven by an impaired banking system, as lending conditions will tighten beyond what would be expected in an ordinary downturn. Heightened liquidity constraints amplify the effects of any contractionary policy on output and unemployment.

This study is necessarily narrow, and does not take into account a number of factors that may also cause the impacts of a policy innovation introduced in normal times to differ from that observed during a prolonged downturn. For example, there may be additional effects on savings behavior, hysteresis effects may also be deeper and more prolonged, and interest rates may respond more significantly if the link between the magnitude of government debt and government borrowing premia is important.

Ball (1996) finds that inadequate responses to recessions have contributed to hysteresis in some countries. A corollary conclusion is that policies of deficit reduction in the presence of substantial output shortfalls will have adverse impacts in both the short and long run. The standard policy prescription – to delay deficit reduction until after recovery is clearly under way and the output shortfall significantly reduced – remains valid.

References

Auerbach, A.J. and Gorodnichenko, Y. (2012), 'Fiscal multipliers in recession and expansion", *American Economic Journal*: *Economic Policy*, 4(2), pp. 1–27.

Baldacci, E. and Kumar, M. (2010), 'Fiscal deficits, public debt and sovereign bond yields', IMF Working Paper 10/184.

Ball, L.M. (1996), 'Disinflation and the NAIRU', NBER *Reducing Inflation: Motivation and Strategy*, pp. 167–94.

Barrell, R., Fic, T. and Liadze, I. (2009), 'Fiscal policy effectiveness in the banking crisis', *National Institute Economic Review*, 207.

Barrell, R., Holland, D. and Hurst, A.I. (2012), 'Fiscal consolidation Part 2: fiscal multipliers and fiscal consolidations', OECD Economics Department Working Paper No. 933

Bernoth, K. and Erdogan, B. (2012), 'Sovereign bond yield spreads: a time-varying coefficient approach', *Journal of International Money and Finance*, 31, pp. 639–56.

Blanchard, O.J. and Diamond, P. (1994), 'Ranking, unemployment duration and wages', *Review of Economic Studies*, 61(3), pp. 417–34.

Calmfors, L. and Lang, H. (1995), 'Macroeconomic effects of active labour market programmes in a union wage-setting model', *Economic Journal*, 105(430), pp. 601–19.

DeLong, J.B. and Summers, L.H. (2012), 'Fiscal policy in a depressed economy', *Brookings Papers on Economic Activity 2012*.

Elsby, M.W.L. and Smith, J.C. (2010), 'The great recession in the UK labour market: a transatlantic perspective', *National Institute Economic Review*, 214, p. R26.

Holland, D. (2012), 'Reassessing productive capacity in the United States', *National Institute Economic Review*, 220.

Ilzetzki, E., Mendoza, E.G. and Végh, C.A. (2010), 'How big (small?) are fiscal multipliers?', Centre for Economic performance Discussion Paper 1016.
IMF (2012a), Fiscal Monitor Update, July.

— (2012b), United Kingdom 2012 Article IV Consultation, Country Report No. 12/190.

Laubach, T. (2009), 'New evidence on the interest rate effects of budget deficits and debt', *Journal of the European Economic Association*, 7, pp. 858–85.

Machin, S. and Manning, A. (1999), 'The causes and consequences of long term unemployment in Europe', *Handbook of Labour Economics*, Vol. 3.

Manning, A. (1993), 'Wage bargaining and the Phillips curve: the identification and specification of aggregate wage equations', *Economic Journal*, 103(416), pp. 98–118.

Nickell, S.J. (1987), 'Why is wage inflation in Britain so high?', *Oxford Bulletin of Economics and Statistics*, 49(1), pp. 103–28.

OECD (2009), 'Adjustment to the OECD's method of projecting the NAIRU', OECD Economics Department.

Schuknecht, L., von Hagen, J. and Wolswijk, G. (2010), 'Government bond risk premiums in the EU revisited. The impact of the financial crisis', European Central Bank Working Paper, No. 1152.

Appendix

Appendix A: Fiscal multipliers and liquidity constraints

In this appendix we illustrate the sensitivity of the estimated fiscal multipliers to assumptions on the short-term income elasticity of consumption and investment. In the presence of perfect capital markets and forward-looking consumers with perfect foresight, households will smooth their consumption path over time, and consumer spending will be largely invariant to the state of the economy or temporary fiscal innovations. However, some fraction of the population at any given time is liquidity constrained with little or no access to borrowing, so that their current consumption is largely restrained by their current income. The share of the population that is liquidity constrained will affect the short-term income elasticity of consumption, given by parameter b_1 from equation (A1) below:

$$\Delta \ln(C_t) = \gamma\{\ln(C_{t-1}) - [a + b_0 \ln(TAW_{t-1}) + (1 - b_0)\ln(RPDI_{t-1})]\} + b_1\Delta \ln(RPDI_t) + b_2\Delta \ln(NW_t) + b_3\Delta \ln(HW_t) \quad \text{(A1)}$$

Where C is consumption, TAW is total asset wealth, which is the sum of net financial wealth (NW) and tangible wealth (HW), $RPDI$ is real personal disposable income, Δ is the difference operator, and the remaining symbols are parameters.

Cross-country differences in the average short-term income elasticity of consumption have a strong correlation with the tax multipliers, as highlighted by Barrell, Holland and Hurst (2012). However, access to credit is dependent both on credit history and on current income, and so is necessarily sensitive to the state of the economy. As unemployment rises, a greater share of the population will be unable to access credit at reasonable rates of interest – at precisely the moment when they are in need of borrowing to smooth their consumption path. This means that consumption is likely to be cyclical, and that b_1 is likely to be time varying and dependent on the position in the cycle. Following a banking crisis the effects can be expected to be particularly acute, as banks tighten lending criteria, as discussed by Barrell, Fic and Liadze (2009). This also suggests that fiscal multipliers are dependent on the state of the economy – especially tax innovation multipliers – and this is consistent with recent studies such as Delong and Summers (2012) and Auerbach and Gorodnichenko (2012).

Investment is always more cyclically sensitive than consumer spending, but these effects may be particularly amplified when the banking system is impaired. We model investment as an adjustment towards a desired capital stock. The stock of capital is one of the factors of production underlying the supply-side of the economy, and a profit maximizing condition that sets the marginal product of capital equal to its price (the user cost of capital) leads to the following long-run relationship:

$$ln\left(\frac{K}{ycap}\right) = \alpha_1 - \sigma \ln(user) \tag{A2}$$

where K is the capital stock, $ycap$ is potential GDP, $user$ is the tax adjusted user cost of capital and σ is the elasticity of substitution between labor and capital.

Embedded within a dynamic framework, the standard equation to model capital demand in NiGEM is given by:

$$\Delta \ln(K_t) = \delta_1 - \delta_2[\ln(K_{t-1}) - \ln(ycap_{t+12}) + \sigma \ln(user_{t-1})] + \delta_3 \Delta \ln(K_{t-1}) + \delta_4 \Delta \ln(y_t) + \delta_5 \Delta \ln(y_{t-1}) \tag{A3}$$

where y is real GDP.

From this we determine investment through the identity relationship:

$$I_t = K_t - (1 - dep)K_{t-1} \tag{A4}$$

where I is gross investment and dep is the depreciation rate.

We distinguish between housing and business investment as the dynamics of behavior are significantly different for the two. The parameters δ_4 and δ_5 may be sensitive to the position of the cycle and particularly to the health of the banking sector.

In order to assess the sensitivity of fiscal multipliers to the magnitude of liquidity constraints, we run our consolidation scenario under a series of eleven different models, allowing the parameters b_1, δ_4 and δ_5 to rise incrementally. The models allow b_1 to rise from 0, which implies perfect capital markets with no liquidity constraints, to 1, which implies that all current income is spent on consumption, with no scope for saving and smoothing consumption. In our standard model, the estimated parameter for b_1 is given by 0.17056, suggesting a relatively low level of liquidity constraints historically. Barrell, Holland and Hurst (2012) put this into an internationally comparative context, which suggests that UK liquidity constraints are on the low side, but not out of line with other advanced economies. Choosing appropriate values for δ_4 and δ_5 is somewhat less straightforward, as a 1 per cent increase in the capital stock is equivalent to a 50–100 per cent increase in the investment flow. The estimated parameters of the standard NiGEM model are 0.042 (δ_{4b}) and 0.013 (δ_{5b}) for business capital and 0.015 (δ_{4h}) and 0.01 (δ_{5h}) for housing capital. We calibrate the parameters by centering so that the NiGEM standard model is between model 2 and 3 in the table below. The δ_5 parameters are set to maintain the ratio of δ_4/δ_5 in the standard version of NiGEM.

The estimated impact on GDP of the consolidation scenario, under different assumptions on the short-run income elasticity of consumption and investment are reported in table A1 below. With no liquidity constraints, we would expect the policy to reduce output by just 0.2 per cent in the first year, while with no options for borrowing to smooth consumption we would expect output to decline by 1.4 per cent. Our standard model predicts that the fiscal policy would reduce output by 0.4 per cent in the first year, under normal conditions with limited liquidity constraints. Differences between the different models dissipate by year 7.

Table A1: Impact of consolidation programme on UK GDP, under different short-term income elasticities of consumption and investment

Model	1	2	3	4	5	6	7	8	9	10	11
Short-run income elasticity of consumption (b_1)	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Short-run capital-output elasticity (business) (δ_{4b})	0.035	0.042	0.049	0.057	0.064	0.071	0.078	0.086	0.093	0.100	0.107
Short-run capital-output elasticity (housing) (δ_{4h})	0.012	0.015	0.018	0.020	0.023	0.026	0.029	0.031	0.034	0.037	0.040
Year 1	-0.22	-0.30	-0.39	-0.48	-0.58	-0.68	-0.80	-0.92	-1.05	-1.20	-1.36
Year 2	-0.44	-0.51	-0.59	-0.67	-0.76	-0.84	-0.93	-1.01	-1.10	-1.18	-1.27
Year 3	-0.77	-0.84	-0.90	-0.97	-1.03	-1.09	-1.14	-1.19	-1.23	-1.25	-1.26
Year 4	-1.20	-1.29	-1.39	-1.48	-1.58	-1.67	-1.77	-1.87	-1.97	-2.08	-2.19
Year 5	-1.80	-1.90	-2.00	-2.10	-2.19	-2.29	-2.39	-2.49	-2.59	-2.69	-2.79
Year 6	-2.13	-2.21	-2.29	-2.36	-2.43	-2.49	-2.56	-2.62	-2.67	-2.72	-2.76
Year 7	-2.04	-2.06	-2.08	-2.09	-2.09	-2.08	-2.07	-2.04	-2.00	-1.95	-1.89
Year 8	-1.66	-1.66	-1.65	-1.64	-1.61	-1.58	-1.54	-1.49	-1.43	-1.36	-1.28
Year 9	-1.16	-1.16	-1.15	-1.14	-1.12	-1.11	-1.09	-1.07	-1.05	-1.04	-1.03
Year 10	-0.63	-0.63	-0.64	-0.64	-0.64	-0.64	-0.66	-0.67	-0.70	-0.74	-0.80
Year 11	-0.14	-0.14	-0.16	-0.17	-0.18	-0.20	-0.23	-0.27	-0.32	-0.38	-0.47
Year 12	0.26	0.26	0.23	0.22	0.20	0.17	0.13	0.09	0.04	-0.02	-0.08

Chapter 5: Human Capital and Growth - A Focus on Primary and Secondary Education in the UK

Abstract: Economic theory and evidence shows that in the long-run human capital is the critical input for growth. While indicators of average educational outcomes at the secondary level, in the UK, tend to show significant improvements over time, they mask the fact that the UK has a long tail of poor (secondary) education performance compared to other countries. This holds back growth and social mobility. The incentives for schools to focus on the performance of children from disadvantaged backgrounds are weak. Dissemination of high quality teaching through the school system depends fundamentally on school incentives - performance measures, autonomy and competition. We propose a flexible system for education, which gives schools greater autonomy and the ability to grow within a national accountability framework that places a premium on radically raising the standards of the bottom ability group. Together with improved choice for parents, better quality information (across the entire distribution of achievement) and more effective incentives for teachers and schools, this will improve the quality of teaching.

Acknowledgements: We are grateful to a large number of people for kindly submitting written or oral evidence to the Secretariat and the LSE Growth Commission. In no specific order, we are particularly grateful to Simon Burgess, Oliver Clifton-Moore, officials at the Department for Business, Innovation and Skills, Peter Dolton, Eric Hanushek, Sandra McNally, Stephen Machin, David Mardsen, Richard Murphy, the OECD, Olmo Silva, Anna Vignoles, Dylan Wiliam, Ludger Wößmann, and Gil Wyness. We are also thankful to Anna Valero for providing useful comments on previous versions of this paper, to Jo Cantlay and Linda Cleavely for the logistical support.

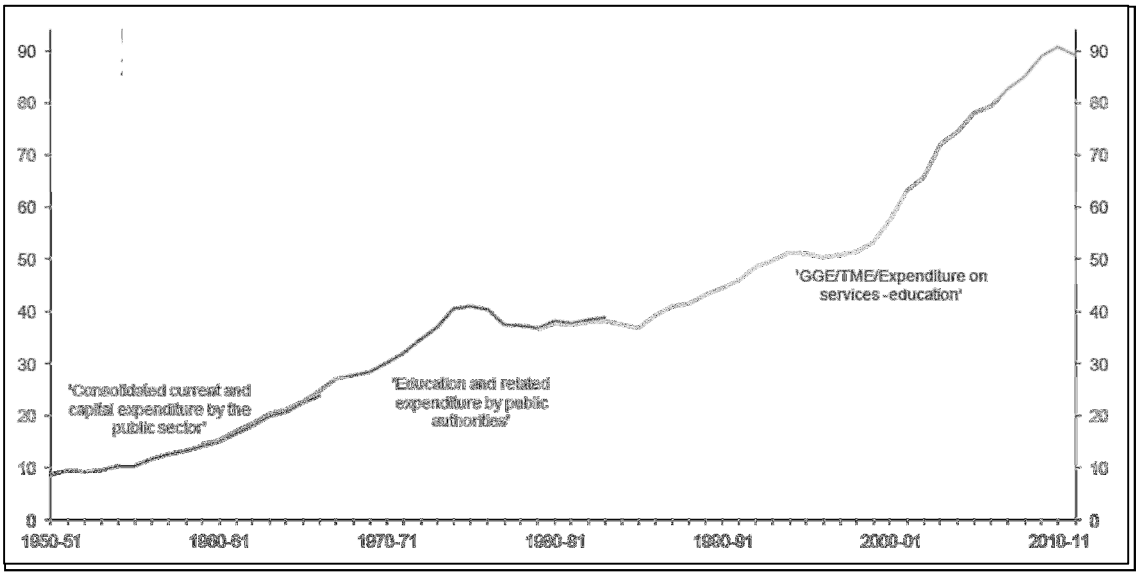
1 Human Capital in the UK

School spending per UK pupil has risen sharply over the last ten years. While national indicators of average educational outcomes show significant improvements, these indicators mask the fact that the UK is performing poorly at the lower end of the educational distribution.

1.1 Educational Expenditure in the UK

Figure 1 tracks education spending per pupil in real terms in the UK since 1950. Real spending levels have increased steadily from the early 1950s to the mid-1970s. After this period of continuous increase –the longest so far - spending fell during the periods 1976-77 and 1979-80. The first half of the 1980s was characterized by flat levels of real spending, and mid-1970s levels of expenditure were only returned to in the late 1980s. As Figure 1 shows the largest annual increases occurred during the 2000s. The real increase in the 11 years to 2010-11 was just over two-thirds (Bolton, 2012).

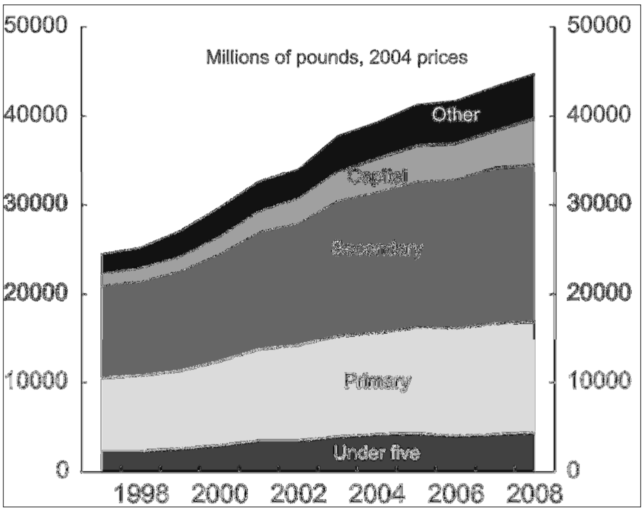
Figure 1: UK Real Public Expenditure on Education, £ 2011-12 prices



Note: Between 1950-66 the series used is the 'Consolidated current and capital expenditure by the public sector'. It excludes spending on school meals and milk. Between 1960-86 the series used is the Education in the UK series 'Education and related expenditure by public authorities'. This series includes expenditure on teacher training and the youth service. Between 1980-current, the series used are the General Government Expenditure (until 1982-83), Total Managed Expenditure (until 1987-88); expenditure is calculated on a resource basis from 2000-01.

Looking closely at the primary and secondary school sectors in Figure 2, we find that real spending per pupil in these sectors in the UK has increased by 4.8 per cent per annum between 1997-98 and 2009-10, leaving spending per pupil significantly above the OECD average. As Figure 2 shows, since 2000, total real school spending has increased by about 40 per cent in real terms for both primary and secondary schools.

Figure 2: Education Expenditure by Sector in England



Source: OECD (2012)

1.2 Educational outcomes in the UK

In this section, we document the level of educational attainment in the UK in the recent years. Outcomes of different groups within the UK are compared – specifically, the educational attainment of disadvantaged children is compared to that of their wealthier counterparts. In addition, we benchmark to international comparators where possible.

Secondary education performance in UK is commonly measured by the percentage of pupils attaining five or more GCSEs (Key Stage 4) at grades A*-C at the end of compulsory schooling. Taken at face value, national indicators suggest that performance in GCSEs has been improving. However, there is a concern that at least some of this could be due to students taking easier subjects, grade inflation or ‘teaching to the test’. This prompts us to look at how UK performs internationally.

There are three international tests for evaluating performance in education[44]: Progress International Reading Literacy (PIRLS) that is available for the years 2001 and 2006 for pupils of about 10 years old; the Programme for the International Student Assessment (PISA- which measures cognitive skills of 15-year olds), conducted in 2000, 2003, 2006 and 2009; and Trends in International Mathematics and Science Study (TIMSS) available in 1999, 2003 and 2007 for pupils of about age 10 and age 14 (that is, years 5 and 9 in England).The latest 2009 PIRLS results put England significantly above the international average of 10 year olds in terms of their reading abilities. But England's performance is below some major European countries (including Italy and Germany) and it has worsened since 2001.The PISA study places the UK close to the OECD average, behind strong performers such as Finland and the Netherlands in 2009.

With regards to measures of secondary school performance, TIMSS is closer to what is checked in national key stage tests and more curriculum-based, while PISA measures the application of knowledge in everyday situations. In TIMSS, England is one of the highest performers and there has been an increase in test scores over time. Thus, international tests attest the improvement in overall performance at the secondary school level in the UK.

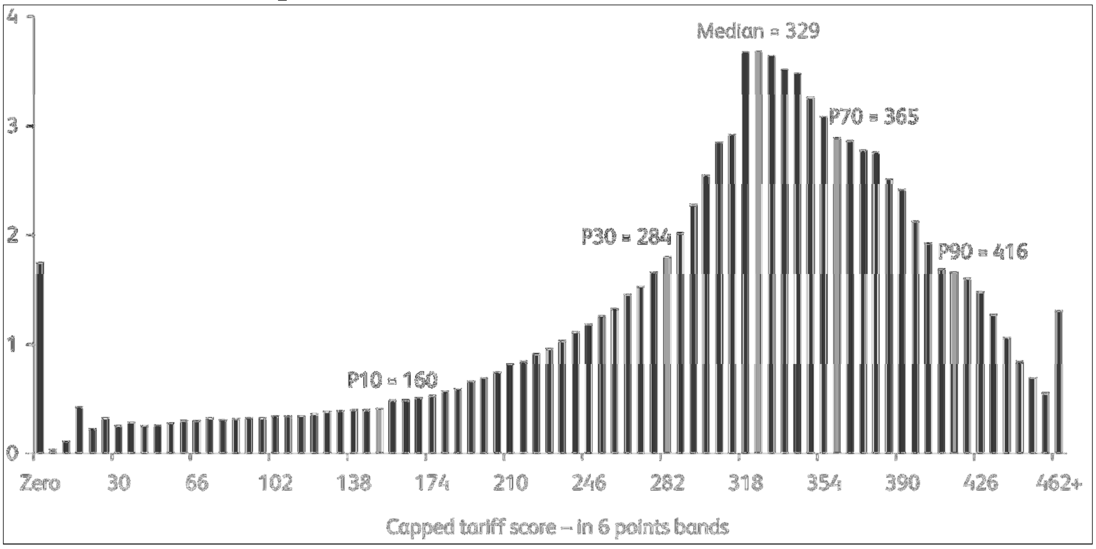
1.2.1 The UK's long tail of poor achievement

One of the most striking features of educational outcomes in the UK is the high frequency of low performers. GCSE results show a 'long tail' of low achievement amongst 16 year-olds (Figure 3). This has been a persistent feature of the UK education system and a continuing policy concern. The UK also does worse than other countries in the proportion of the population aged 24-35 with upper level secondary qualifications - equivalent to GCSE passes at A*- C or above (McNally, 2012).

---

[44] A new indicator has been recently published by the Economist Intelligent Unit: "The Learning Curve". This measure is based on the existing indices but adopt a wider prospect by adding new criteria such as graduation rates, adult literacy and the effect of years in school on productivity. The new results don't change the ranking at the top: Finland, South Korea and Hong Kong are the best performing countries, followed by Japan and Singapore. But other countries sharply changed their ranking position. Britain, for example, gained sixth place in comparison to the PISA 2011 classification. At the same time, this result put in light the large quality gap between compulsory and higher education and the potential negative effect for students that would not attend universities (The Economist, 2013).

Figure 3: Key stage 4 (GCSE) results, England (2008): % with results in each band of total GCSE points



Source: DfE. Results are for maintained (state) schools only.
Note: Total points 'capped' by DCSF to show those from a pupil's best 8 GCSE (or equivalent) passes at age 16. The system awards 16 points for a pass at G, 22 for an F, up to 52 for an A and 56 for an A*.

This distribution is confirmed by the PISA data which shows significant performance variability within the UK (OECD 2010). Table 1 shows the proportion of pupils at each level of performance compared to other countries. High-performing countries such as Korea and Finland have a narrower range of scores overall. The OECD finds that in the UK the gap between the bottom performers and middle performers is bigger than the gap between the middle performers and the top performers (OECD 2010)[45]. In other words, there is a bigger gap created by students falling behind the average score than there is by students pulling away at the top.

[45] The gap between the 10th percentile and the median is larger than the gap between the median and the 90th percentile.

Table 1: % of students at each level of the PISA proficiency scale for reading, 2009

	Below Level 2	Level 2	Level 3	Level 4	Level 5	Level 6
United Kingdom	18	25	29	20	7	1
Key Competitors Average*	10	18	30	28	12	2
OECD Average	19	24	30	21	7	1

Note: *Key competitors defined as Australia, Canada, Finland, Korea and Singapore. These were chosen as countries that traditionally score well on PISA and are frequently cited in comparison to UK performance.
Source: Clifton and Cook (2012) using data from OECD 2010

Of particular note in this table is the sheer volume of UK students failing to achieve basic proficiency (level 2). Around a fifth of students failed to reach basic proficiency in reading and maths, which translates to around 113,000 students in England. This group is more than twice as big as the group of students that reached the top two performance levels. Unsurprisingly, high-performing countries do not just have lots of students at the highest levels, but also relatively few students at the lower levels.

A picture therefore emerges of a large pool of 'poor performers' that contributes to UK's relative weak performance in international rankings. The UK therefore faces a *two-pronged challenge* - both to stretch those at the top as well as to raise the performance of those falling behind. In terms of quantity of pupils, the latter is the bigger challenge, with around a fifth of pupils failing to get the basic skills required to succeed in life.

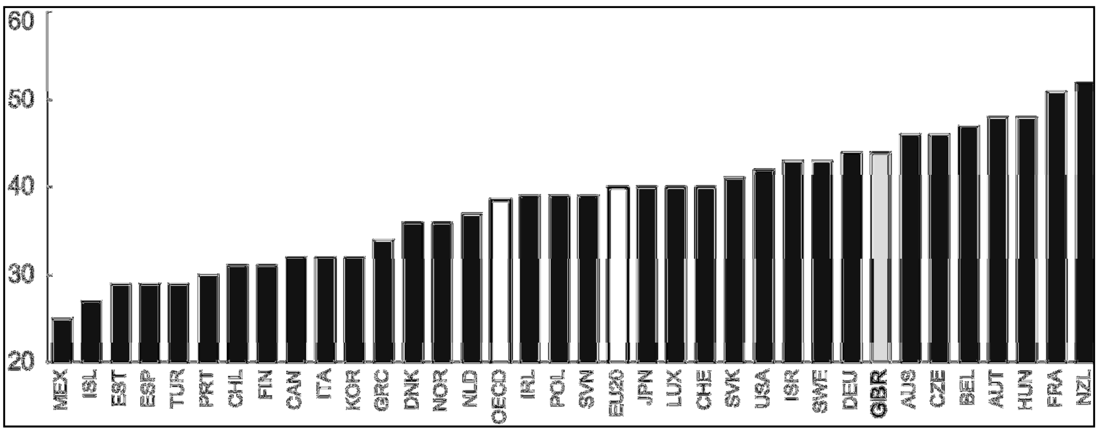
1.2.2 The socio-economic gradient in UK educational attainment

In this section, we discuss the extent to which the long tail discussed above is driven by socio-economic disadvantage. The relationship between socio-economic background and educational attainment is called the socio-economic gradient of education. It is a well-established empirical fact that children from disadvantaged backgrounds are over-represented in the group of poor performers while children from wealthier families are over-represented in the group of high performers.

According to the OECD PISA 2009 results (the most recently published), 14 per cent of the variation in student performance in the UK is explained by students' socio-economic background. This is in line with the OECD average but contrasts with Canada or Japan, where the explained variation is 9 per cent. If we consider a wider family context (including, for example, the immigrant background or the language spoken at home) differences in family background characteristics explain 25 per cent of the differences in performance across UK students – versus 22 per cent in the typical OECD country and 19 per cent in Canada, Finland, Japan, and Korea. These numbers suggest a weaker relationship between socio-economic background and educational inequality in other OECD countries compared to the UK.

The same message from a different perspective is given by Figure 4 which measures the impact of socio-economic background on PISA scores. It shows that UK has one of the highest impacts among the OECD countries. Moreover the share of UK students from weak socioeconomic backgrounds performing well is low: the average PISA score of the worst performing 10 per cent of UK students is below the average for the same group in other OECD countries (OECD, 2012).

Figure 4: Impact of socio-economic background on PISA 2009 reading score



Note: Score point difference associated with one unit increase in the PISA index of social and cultural status. **Source:** OECD (2012)

The PISA study also shows the percentage of "resilient students" - those who come from the lower quartile of the distribution of socioeconomic background but go on to score in the top quartile of PISA test results. On this measure, the UK trails both the OECD average and key competitors, with only 6 per cent of students meeting considered “resilient” according to this criteria.

Schuetz et al. (2005) relate family background to student test scores across countries using TIMSS. Although the gradient is present in most countries, the estimated effect is higher in England than in any other country for this particular survey. Achievement gaps between children from rich and poor backgrounds are evident from a very early age and continue to widen as children grow up. As a result the achievement gap between rich and poor is really significant at age 16 in GCSE results[46].

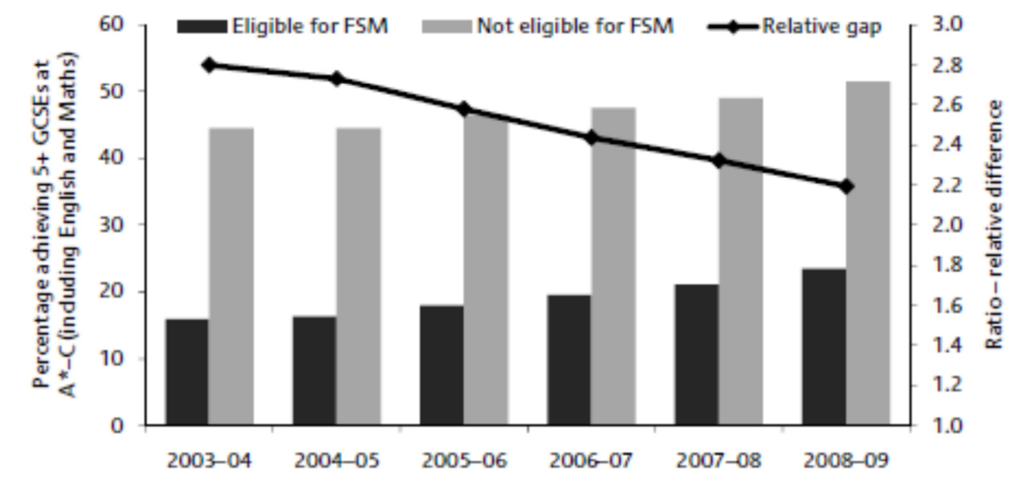
One way to analyse this pattern is to look at the school attainment of children eligible for free school meals (FSM)[47]. In 2008, while half of all children from higher-income families (not eligible for free school meals) achieved five good GCSEs (A*–C) inclusive of English and Maths, less than a quarter of FSM children achieved these grades (Figure 5). However, Figure 5 also shows that this gap has narrowed in relative terms in recent years[48]. Despite the improvement, the current achievement gap is still large and negatively impacts on later-life income and earnings inequalities with the potential risk of being passed on to future generations (Chowdry et al, 2010).

Figure 5: Percentage of children achieving 5+ GCSEs at A*–C (including English and Maths) by FSM eligibility

[46] The fact that these achievement gaps are present even before school suggests that the educational system cannot take sole responsibility.

[47] Children are classified as ‘eligible’ for FSM in administrative data only if they are both eligible for and claiming FSM. Only families with a low income and no adults in full-time paid work are eligible for FSM. A priori therefore, we would expect FSM ‘eligibility’ to identify children in the lowest income households.

[48] It is worth noting that only one part of this achievement reflects a real improvement in pupils’ performance owing to the ‘grade inflation’ phenomenon.

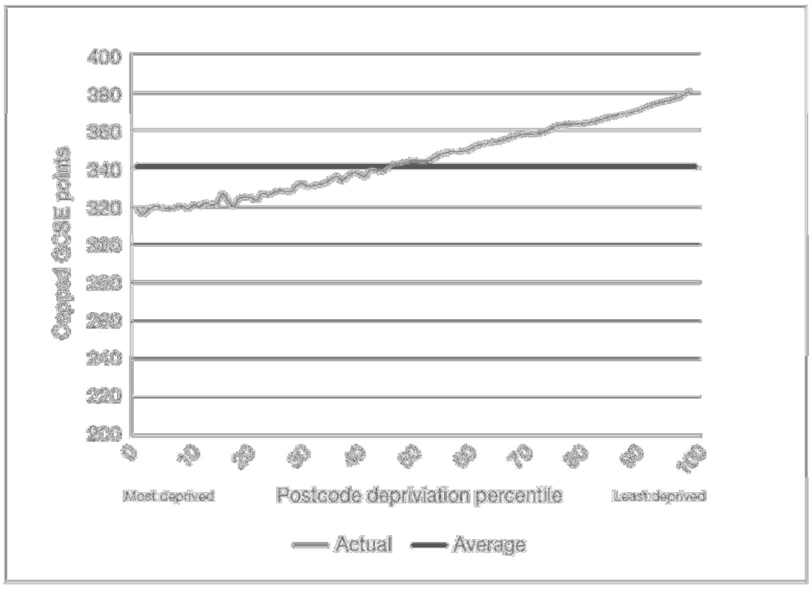


Sources: 2003–04 to 2007–08 – Department for Children, Schools and Families (2009a); 2008–09 – Department for Children, Schools and Families (2009b).

However, the problem is much broader and goes beyond FSM pupils. Indeed, there is a clear and consistent link between deprivation and academic achievement wherever you are on the scale (Figure 6). With deprivation measured at the neighbourhood level, those pupils living in the most disadvantaged postcodes score on average 320 points at GCSE (or the equivalent of about eight Cs), and the results gradually improve as you move towards better (least disadvantaged) postcodes. Pupils living in the wealthiest postcodes score on average 380 points or the equivalent of just over eight Bs. It is therefore not possible to identify a particular indicator of deprivation at which performance falls. This challenges the assumption that programmes targeted towards pupils who are eligible for free school meals will be sufficient to close the gap, as the problem is much wider than just this group of pupils.

Another striking feature of the "long tail" is the higher variability in GCSE results of poorer pupils compared to wealthier pupils. The highest-achieving pupils from disadvantaged postcodes score almost as well as the highest-achieving pupils from wealthier areas (about 40 points less at GCSE). However, low-achieving pupils from disadvantaged neighbourhoods score much worse than low-achieving pupils from wealthier areas (about 120 points less at GCSE).

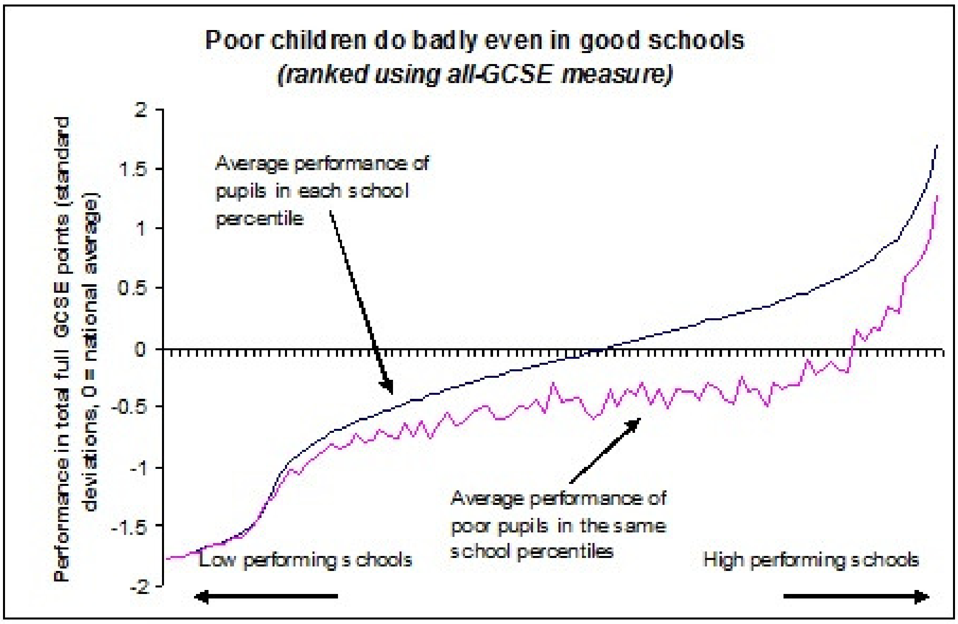
Figure 6: Capped GCSE points by postcode deprivation



Source: Clifton and Cook (2012)

More importantly, in contrast with popular belief, disadvantaged children are not all concentrated in a small number of poorly performing schools (though of course, generally disadvantaged areas will tend to have larger shares of disadvantaged children than wealthier areas). Disadvantaged children are spread across the entire school system, and perform poorly compared to their wealthier peers in the vast majority of schools. Figure 7 illustrates this problem. The darker line shows the average point score for all pupils in each percentile of schools. The pink line shows the average point score only for poorer children within those schools (those living in the bottom fifth of households, as measured by the deprivation level of the household postcode). While the darker line's slope increases rapidly, the pink one is flatter. In other words, the problem is not just that there are a few schools which have all the disadvantaged children in them performing poorly. The problem is also that disadvantaged children perform poorly (compared their wealthier peers) in a vast majority of schools.

Figure 7: Attainment of poor pupils vs. (within school) average by school deprivation



Source: Cook, 2012

In the next two paragraphs, we discuss what attributes can be used to identify disadvantaged children. Low income is the attribute of disadvantage that, over the years, has attracted most attention from academics, politicians and the general public. The simple correlation between low income and poor educational outcomes (the so-called socio-economic gradient of education) is long established[49]**.** Income is often used as a measure of disadvantage for three main reasons: (i) the income gradient gives a measure of educational inequality in its own right; (ii) some other features of disadvantage discussed below are associated with income[50]; and (iii) the relationship between family income and education is one of the key drivers of intergenerational income mobility across time (e.g. Blanden, Gregg and Macmillan, 2007).

A multitude of possible reasons explain why the children of low income families do less well at school; some of these are causal (family income actually influences a child's educational attainment), while others are non-causal (for example, income simply acts as an indicator for many other aspects of disadvantage, such as parental education level or social class). Different studies[51] suggest that sustained income shocks do impact on child educational outcomes, and that low

[49]Rowntree, 1901, Glennerster, 1995.

[50]This could be because those features are a causal driver of income; are they caused by income; or both those features and income are commonly determined by a third factor.

[51] E.g. Steven and Schaller (2011) and Gregg et al. (2012) analyse fathers' job displacement; Dahl and Lochner (2012) examine the Earned Income Tax Credit's introduction in the US, and Milligan and Stabille (2011) study the variation in child tax benefits across time and Canadian provinces.

household income is likely to be one of the primary drivers behind the developmental deficits of poor children. Adults in low income families may have characteristics that negatively impact on children's' educational achievement, such as poorer innate ability; a lower emphasis on educational achievement in parenting; a reduced ability to translate parenting time into educational development; or lower ambitions (Blanden and Gregg, 2004).

2 The Importance of Human Capital

A country's educational outcome and human capital formation has an important bearing on economic growth. Theoretical models of economic growth have considered different determinants of economic growth and a range of theoretical approaches. The standard neoclassical growth model of Solow (1957) considers the output of the macro economy as a direct function of just its capital and labour (and technological level) in a given period. Augmented neoclassical growth theories, developed by Mankiw, Romer and Weil (1992), include also human capital among the factors of production: a change in the education level induces growth. However, in these models, technology, the key driver of growth rates, is exogenous. A different view comes from the 'endogenous growth' models developed by Lucas (1988), Romer (1990) and Aghion and Howitt (1998) whereby technology is now endogenous to the model. This literature stresses on the role of education in increasing the innovation capacity of the economy through developing new ideas and new technologies. Other macro models on technological diffusion (such as Nelson and Phelps (1966), Welch (1970) and Benhabib and Spiegel (1994)) argue that education and training facilitate the adoption and implementation of new technologies with positive effect on growth and development.

Empirical analysis based on growth and development accounting models have generally shown that education accounts for a large share of economic growth and development. Growth accounting models such as Griliches (1970) using US data; Barro and Lee (1993, 2001); Mankiw et al (1992); Barro and Sala-i-Martín (2004) and more recently Sala-i-Martin, Doppelhofer, and Miller (2004); De La Fuente and Domenech (2006) and Cohen and Soto (2007) relate educational attainment to economic growth (measured as GDP or GDP per capita) while development accounting models seek to explain cross-country differences in income levels and education's role in this[52](see for example Lagakos et all, 2012; Gennaioli et al 2011).

Human capital has usually been measured as educational attainment, in terms of years of education. More recently, with the development of new datasets, authors have been able to consider different measures, such as making a distinction between entrepreneurs/managers and worker education (Bloom and Van Reenen, 2007, 2010; La Porta and Shleifer, 2008; Syverson, 2011); or consider the quality of education (or cognitive skills) using the PISA, TIMSS results (see Hanuschek and Woessmann, 2008 for a review of these studies). These new empirical approaches confirm the critical role of human capital in the growth and development process of countries, regions and firms.

Hanushek and Woessmann (2007) and Hanushek and Kimbo (2000) emphasise the positive effect of the quality of education, rather than quantity alone on economic growth. Hanushek and Schultz (2012) state that a one standard deviation difference on test performance (100 points on the PISA assessment) is related to a 2 percentage point difference in annual growth rates of gross domestic product per capita. OECD (2010b) estimated that UK would gain more than $US 6 trillion by increasing its average performance by only 25 PISA points (or by ¼ standard deviation). These are obviously long-run calculations, but they do give some indication of the huge potential prize if the UK could attain the quality of education in Germany or Australia this would put us on a path that would more than double our income per person. Following the more ambitious goal of reaching the Finland's PISA performance, the UK would record a GDP improvement of more than $US 7 trillion or 3 times the current GDP.

The major criticism of these cross-country analyses is that they show associations between human capital and growth but not necessarily causation – i.e. they do not address issues of endogeneity: estimates of school attainment could reflect reverse causality since improved growth could lead to more schooling rather than the reverse (Bills and Klenow, 2000). Whether or not there is a casual relationship is a very important issue from a policy point of view. However, this problem is not easily solved by using standard econometric techniques since the potential instruments for education are often correlated with institutional features (Glaeser et al, 2004). Nonetheless, Hanusheck and Woessman (2009) and Gennaioli

[52] See Hall and Jones (1999) and Caselli (2005) for a theoretical foundation.

et al (2011) have tried to account for the endogeneity problem using different estimation methods, although none of these approaches address all the critical issues, they provide some assurance that the results are not biased by any of the most obvious problematic issues.

2.1 Human capital inequality and growth

In addition to the overall accumulation of human capital, inequality of human capital within a country is important for growth. Reducing human capital inequality practically means to improve the educational attainments of pupils at the bottom of the distribution, hence focusing on worst-performing students. OECD (2010b) estimates how GDP would improve if OECD countries addressed education inequality by ensuring all students reach a level of minimal proficiency (i.e. a PISA score of 400). Under this scenario that brings up the bottom of the distribution, the present value for UK improvement is more than 6 trillion USD or 2.7 times its current GDP.

Recent literature has pointed at human capital inequality and its influence on demographic variables to explain the negative relationship between human capital inequality and growth. De la Croix and Doepke (2003) and Moav (2005) study the link between human capital inequality and increasing fertility (with negative outcome on growth), while Castelló-Climent and Doménech (2008) focus on how human capital inequality may dampen growth by reducing life expectancy and investment in education.

Some authors analyse the effect of inequality on growth under imperfect credit markets (e.g., Mookherjee and Ray, 2003) or at different level of country's development. Galor and Zeira (1993) show that in the presence of credit market imperfections and indivisibilities in investment in human capital, the initial distribution of wealth affects investment in education both in the short and in the long run. With regards to development studies, Castello-Climent (2010) shows that a greater degree of human capital inequality discouraged the per capita income growth rates in most developing countries during the period 1965–2005, where the life expectancy and fertility channels seem to play a prominent role. However the effect disappears in higher income economies.

2.2 Human capital inequality and social mobility

As discussed above, disadvantaged students, on average, perform worse than pupils from rich families. Poor educational outcomes for disadvantaged students reduce social mobility later in life, which in turn perpetuates low intergenerational social mobility[53]. Two facts help us understand the size of the problem of educational inequality. Firstly, 24 per cent of UK disadvantaged students are resilient, in the sense that they come from the 25 per cent of the socio-economically most disadvantaged students but perform much better than expected on the base of their socio-economic background (31 per cent is the average in the OECD) (OECD, 2010a). Secondly, it is also true that the odds of a young person from a family with low levels of parental education attaining higher education is at 0.61 in the UK, well above the OECD average of 0.44. This suggests that if socio-economic disadvantages were not allowed to hold back educational attainment, social mobility could be high in the UK.

A series of reports have highlighted the UK's low levels of social mobility, showing how children from poorer backgrounds struggle to gain access to university, enter professional jobs and earn decent wages (see Milburn 2012, Sutton Trust 2011, Blanden et al 2005). This, in turn, means that disadvantages can become entrenched across generations.

A high level of education has become more important for getting a good job over the past 30 years, meaning those families which are unable to invest in education are left further behind (Lindley and Machin 2012). Education can provide access to many opportunities later in life, and schools can help to create a level playing field for young people as they start out (Clifton and Cook, 2012). Research has identified a causal relationship between high levels of education and a number of outcomes in later life, including higher earnings (Dickson 2009), lower teenage pregnancy (Black et al 2008), healthier behaviours and a lower likelihood of serving a prison sentence (Heckman et al 2006).

The link between intra-generational income inequality and intergenerational social mobility is rather complex. Recent evidence (OECD, 2012) shows that higher

[53] It is important to remember that raising achievement alone is not enough. There also needs to be sufficient demand for these skills and qualifications in the labour market, so that young people can put their education to good use. Recent cuts to post-16 education, a weak youth labour market and the prevalence of low-quality jobs will also have to be tackled for improvements to social mobility to be realised (Lawton and Lanning 2012, Keep et al 2006). Other factors can also be important to social mobility, such as having access to social networks and inherited wealth. Raising achievement in schools is therefore just one piece of a much bigger jigsaw (Clifton and Cook, 2012).

inequality is associated with lower intergenerational mobility. First of all this can be explained by the fact that with higher wage dispersion allows for higher returns to education and this may benefit individuals whose don't have any constraint to invest in education. Secondly, if income inequality increases the severity of credit constraints, mobility decreases. Thirdly, large differences in educational outcomes raise income inequality and lower intergenerational social mobility.

Wide empirical evidence strongly supports the fact that education is one of the major drivers of intergenerational social mobility, particularly income mobility. The UK has recorded a rise in intergenerational mobility between the cohorts born at the end of the 1950s and those born in the 1970s. (Blanden et al, 2004). This was mainly due to a disproportional increase in educational opportunities biased towards individuals from better-off backgrounds.

3 Drivers of Educational Attainment: A Review of the Literature

In this section we will briefly discuss the factors that the economics of education has identified as driving educational attainment. A complex inter-play of factors contribute to and cause attainment gaps between advantaged and disadvantaged children. These factors include:

- Broad contextual drivers such as the socioeconomic background of a child (e.g. family income and parental education) and their knock-on effect on home learning environment;
- Pupil-level factors e.g. having been in care at some stage, having English as another language (EAL) status, Special Educational Needs (SEN)[54] status, mobility and ethnicity. These have a complex relationship with material disadvantage;

[54]SEN is a multifaceted classification which brings together children with innate cognitive/learning difficulties and children who are underperforming for reasons other than their innate ability (e.g. strong negative impact of family background and/or poor teaching quality and/or unsupportive peer effects). The first sub-group is defined by a characteristic that puts it at disadvantage. The second sub-group is defined by its (poor) performance level, and may or may not be at disadvantage (depending on drivers of poor performance are family background or other factors). In that sense, SEN conflates discretionary inputs (e.g. teaching quality); non-discretionary inputs (e.g. unsupportive family background); and outputs (e.g. low attainment).

- School-level factors that determine the quality of the child's formal learning environment such as teaching, peer composition, resources and the general effectiveness of individual schools in overcoming material barriers.

Recent evidence (Kramarz, Machin and Ouazad, 2009) on the relative contributions of pupils, schools and peers shows that the variance of test scores is mostly explained by the pupil effect[55]. The standard deviation of pupil effects is between 4 to 5 times larger than the standard deviation of school effects[56] - the second largest source of variance in the results. Many other studies suggest that families are much more important[57] than schools and peers in explaining the variance in results (Teddlie and Reynolds, 2000; Todd and Wolpin, 2007).

We now consider in more detail the literature on the key factors driving educational outcomes. We start with pupil effects, and then school and teacher effects, and peer effects. Finally, we consider the effects of expenditure.

3.1 Pupil Effects

The finding that pupil effects account for the majority of the variance in test scores implies that the influence of home environment and socioeconomic background on schooling outcomes is very important.

Even before pupils start school, there is a large gap in cognitive ability between children from high and low socio-economic backgrounds. Feinstein (2003) finds significant gaps between children from a high and low socio-economic background in an index of development. Another way to illustrate pre-school gaps is to look at vocabulary skills by gender and ethnic group at the time of school entry. Table 2 shows gaps in the vocabulary skills of five year olds in the Millennium Cohort Study (MCS).

[55] In this context, pupil effects consist of a range of educational experiences pupils carry with them, reflecting parental background, the quality of the schools previously attended, innate ability, etc. This research measures the relative contributions of pupils, schools and peers without restricting the analysis to observable proxies for peers' characteristics or school quality.

[56] However, any assessment of the relative merits of various policy alternatives (e.g. targeted at individual effects vs. school effects) needs to allow for that fact that school quality has an impact on multiple pupils.

[57] This may also explain the low attainment of disadvantaged pupils in all the school without regards to the quality level that we highlighted above.

Table 2: Age 5 Differences in Vocabulary by Gender and Ethnicity, Millennium Cohort Study

Ethnic Group	Boys	Girls
White British	55.9	56.5
Black, Caribbean	48.4*	51.0*
Black, Other	44.2*	47.2*
Bangladeshi	40.4*	41.7*
Pakistani	40.6*	40.7*
Indian	49.8*	50.3*
Chinese	41.2*	55.2
Number of Children	*4,587*	*4,452*

Notes: Based on Table 3 of Dustmann, Machin and Schonberg (2010). The vocabulary test is standardised to have mean 50 and a standard deviation of 10. A * denotes statistically significant differences relative to White British boys or girls respectively.
Source: McNally (2012)

This illustrates that human capital acquisition is not something that begins at school and that inequality is evident even at an early stage. Breaking the link between family background and educational attainment (and improving educational attainment generally) seems to require policies directed at families before the start of formal schooling. This might involve close attention to the quality of early childcare and pre-school settings. However, if part of the issue is poverty and worklessness, then the policy solutions may also lie in other areas of social policy such as housing, employment benefit, childcare provision (McNally, 2012).

As discussed in section 2, low income is the attribute of disadvantage that, over the years, has attracted most attention from academics, politicians and the general public. The simple correlation between low income and poor educational outcomes (the so-called income gradient of education) has been long established (see Appendix 1 for evidence on the relationship between family income and educational attainment). However, the most significant social background characteristic is parental education, which has been shown to account for between a quarter and two-fifths of the deficits of low income children[58]. While a range of other family background characteristics (for example parents' employment status or family

[58]Gregg, Propper and Washbrook (2008). Chowdry et al., 2009 also found that differences in parental education between young people from different socioeconomic backgrounds provide a major explanation for differences in their educational attainment, and some of the evidence presented points towards the relationship being causal.

structure) have occasionally been linked to child attainment, the evidence of their effects, conditional on other economic circumstances is still mixed.

3.2 School and Teacher Effects

It is well acknowledged in the theoretical and empirical literature that the key driver of school quality (defined as value added of the school) is the quality of their teaching staff. There are a large number of anecdotes about the positive impact of excellent teaching on pupil's performance. However, trying to quantify this effect is difficult, principally because of the data requirements. Slater, Davies, and Burgess, (2009) use a unique dataset to estimate the effect of individual teachers on student outcomes, linking over 7000 pupils to the individual teachers who taught them in each of their compulsory subjects and to the results of the exams they take at age 16. Their results suggest that being taught by a high quality (75th percentile) rather than low quality (25th percentile) teacher adds 0.425 of a GCSE point[59] per subject (25 per cent of the standard deviation of GCSE points).

Rivkin et al (2005) relate the teacher quality measure to the socioeconomic gap in outcomes. They measure the gap in GCSE points between a poor and non-poor student (equal to 6.08 GCSE points) and suppose that this gap arises over 8 subjects that they both take. If the poor student had good (75^{th} percentile) teachers for all 8 subjects and the better off student had poor (25^{th} percentile) teachers for all 8, this would account for 3.4 GCSE points. This is a powerful effect which is not typically addressed in explaining the socio-economic educational gap.

Similar studies for the US suggest that having a teacher at the 25th percentile versus the 75th percentile of the quality distribution would imply a difference in learning attainments of roughly 0.2 standard deviations in a single year. This would induce a move of a student at the middle of the achievement distribution to the 59th percentile. The magnitude of such an effect is large relative to the estimated effects of a ten student reduction in class size, which is of 0.1-0.3 standard deviations (Hanushek and Rivkin, 2010).

The academic literature has consequently sought markers for high quality teachers, looking in particular at pay, teachers' experience and academic level. There is a large literature on the impact of teacher pay. One strand investigates the effect of teacher salaries on school performance. Although initial evidence on this was

[59] An increase from one grade to the next, says a B to an A, is one point.

mixed[60] more recent work has mostly found mixed results (see Appendix 2 for a more detailed review of the literature). The evidence suggests that teaching staff respond to pecuniary and market incentives aimed at increasing their effort and 'output' (i.e. learning). However, it is worth making a distinction between a general increase in teachers' wages (due, for example, to a general increase in the national pay scheme) and an improvement in pay linked to geographical disparities in costs of living or teachers' outcomes. Only the latter types of interventions seem to have a positive impact on teacher's performance and pupils' achievement (Propper and Britton, 2012).

In particular evidence indicates that relative salaries in alternative employment opportunities are important influences on the attractiveness of teaching as a profession (Santiago, 2004). As OECD (2005) discussed, teachers' salaries relative to those in other occupations influence: (i) the decision to become a teacher after graduation; (ii) the decision to return to teaching after a career interruption as returning rates are generally higher among those teaching subjects that provide the fewest opportunities for employment elsewhere; and (iii) the decision to remain a teacher as, in general, the higher teachers' salaries are the fewer people leave the profession. McKinsey (2007) suggested that while raising salaries in line with other graduate salaries is important, raising them above the graduate market average level would not lead to substantial further increases in the quality or quantity of applicants. Relative earnings seem to be less important when the decision is whether to enrol in teacher education or another college course (Hanushek and Pace, 1995). So, relative salaries matter during career choice, not choice of education stream.

Using data on university graduates in UK, Dolton (1990) showed that increasing teacher salaries by a small amount (10 per cent) resulted in a large rise in applications (30 per cent). The "wage elastic" teacher supply could be explained by the comparatively low level of teacher's wages. Wolter and Denzler (2003) run a similar analysis for Switzerland and showed that since salaries were already high, further increases in salary had little impact on the number or quality of applicants to teaching. While starting salaries in general are high in England[61], low top wages at higher career levels discourage the more experienced teachers from remaining in the profession and also deter good graduates from starting a teaching career. There is

[60] For example, Hanushek (1986) highlights that only nine out of sixty teacher salary studies found a positive effect of teacher wages on school performance

[61] Working hours in teaching are also fairly long compared to many other OECD countries.

also a growing body of work investigating the impact of performance related pay. Whilst again there is some mixed evidence, the general consensus appears to be that performance pay for teachers does improve student attainment in a variety of settings[62].

However, a good salary is not necessarily the main or the only motivation for teaching. The status of the teaching profession, the career opportunities and the decisional power given to them are all important factors in explaining their performance. For example, Hoxby (2002b) provides evidence that school choice affects the teaching profession by increasing demand for staff with higher qualifications (especially in mathematics and science).

While a strong relationship between teacher experience and performance has historically been established, recent studies have consistently found that the impact of experience is concentrated in the first one or two years of teaching with little impact of any additional one (Hanushek, 2008). Teachers' education also tends not to be correlated with quality[63].

Given the critical role of teacher quality on pupil's performance and evidence on the correlation between teachers' education and learning, recruiting and maintaining the most efficient teachers should be prioritised. The issue is how to attract and select good teachers. This is not a straightforward process since it is difficult to assess ex-ante if a candidate would be a good teacher. Qualitative research suggests that top-performing school systems manage to attract better people into the teaching profession, leading to better student outcomes. They do this by introducing highly selective teachers training, developing effective selection processes for identifying the right candidates and paying good (but not great) starting compensation. Conversely, lower-performing school systems rarely attract the right people into teaching. The success in attracting talented people into teaching is linked to specific country features such as history, culture and status of teaching profession. However, there are some policies that can be implemented to attract the best graduates, such as effective mechanisms for selecting teachers, good teacher training programmes, good starting compensations and increasing professional autonomy in

[62]See Appendix 2 for a review of the literature.

[63] However, some argue that the education system as a whole could benefit from a teacher's higher education level in other forms. First, more educated teachers may increase the success of school autonomy, by providing better inputs in the curriculum design and in developing new teaching methods. Secondly, the perception of the teaching profession is linked to the anticipated level of education and training teachers are required to undertake to become teachers (Day et al, 2006).

schools. All these policies could contribute to increase the status of the teaching profession, the attractiveness of teaching as a career and hence attract the best graduates.

3.3 Peer Effects

Another driver of educational attainment is believed to be peers' behaviour and characteristics. This has been documented empirically (Coleman, 1966) as well as theoretically (Angrist and Lang (2004), Hoxby (2000) and Lavy and Schlosser (2011), Gould et al. (2009)). The main rationale is that group actions or attributes might influence individual decisions and outcomes. However, the estimation of peer effects is empirically challenging. Manski (1993) highlights the pitfalls of endogenous peer selection and the difficulty of distinguishing between average school effects and peer effects.

Recent empirical evidence based on better data and better identification strategies has reached consensus that to capture peer effects, analyses should not focus on the average students but should consider pupil distributions. There is little conclusive evidence suggesting that studying with high ability peer group leads to better outcomes (Atkinson et al, 2008; Bradley and Taylor, 2008; Dills, 2005; Summers and Wolfe, 1977) for all pupils while the presence of low ability peer groups can decrease general outcomes (Lavy et al, 2012; Gibbons and Telhaj, 2008; Winston and Zimmerman, 2004; Zimmerman, 2003 and Handerson et al, 1978, Summers and Wolfe, 1977)[64]. Lavy et al (2012) show that it is only the very bottom 5 per cent students that (negatively) affect average outcome and not "bad" peers in any other part of the ability distribution. They also find evidence that the presence of students in the top 5 per cent of the ability distribution does not impact average outcomes. Henderson et al (1978) show that mixing weak and strong students lowers educational attainment for higher achievers. Similar results are also found by Bradley and Taylor (2007), who use pupils moving between schools to address the problems inherent with estimating peer effects, and find the effects of a more able peer group are stronger for low ability students than for higher ability students. On the other hand, Betts and Shkolnik (2000) find little evidence of differential effects of ability grouping for high or low ability pupils.

[64] For a slightly different message see Carrell et al 2011.

The negative impact of low ability students on the outcome of other students has been explained by some academics by the fact that more homogeneous groups of students might be taught more effectively (Duflo et al, 2010) or by pointing at the classroom disruption and decrease in attention paid by the teachers (Lavy et al. 2012). Some studies suggest that these general findings mask some market heterogeneity along the gender dimension by showing that girls are significantly affected from interactions with peers (Lavy et al, 2012 and Stinebrickner and Stinebrickner, 2006).

To overcome the difficulties of endogenous peer selection, a number of studies use the random allocation of accommodation within higher education in the US. Sacerdote (2001) finds that peers have an effect on grade point average. In a similar framework, Zimmerman (2003) and Winston and Zimmerman (2004) find no credible effect on the top of the SAT ability distribution, but do find evidence of a negative impact on students in the middle of the SAT distribution when grouped with students in the bottom 15 per cent of the SAT distribution.

Taking a step further, Carrell et al (2011) use a random experiment to determine whether student academic performance can indeed be improved through systematic sorting of students into peer groups. They design peer groups at the United States Air Force Academy (USAFA) and using an experimental design, sort the incoming college freshman cohorts at USAFA into these peer groups. The objective was to improve the grades of the bottom one-third of incoming students by academic ability. The actual outcomes from the experiment yielded unexpected results. For the lowest ability students there is a negative and statistically significant treatment effect; for the middle ability students, who were expected to be unaffected, there is a positive and significant treatment effect of 0.067. High ability students were unaffected by the treatment.

Finally, Gibbons and Telhaj (2008) offer an alternative interpretation of the peer effect. They suggest that peer effects may impact other factors different from school attainment such as subsequent educational decisions and may provide other immediate and long-run benefits – such as life-time friendship networks- which make schools with good peer groups desirable commodities.

3.4 How significant is expenditure?

Assessing the role of educational expenditure on attainment, existing research has struggled to show a clear causal relationship between the amount that schools spend and student achievement. How money is spent is typically much more important than how much is spent (see Hanushek, 2008 for a review of the literature). Analysing the effect of spending on reduced pupil–teacher ratios, most studies find no significant relationship with achievement.

Levacic and Vignoles (2002) find that in the British context, the impact of school resources is small. Holmlund et al (2010) find that after controlling for the range of pupil and school-level characteristics, the estimated effect of an increase of £1,000 in average expenditure per pupil would raise standardised test scores by about 5 per cent of a standard deviation. They find evidence of a consistently positive effect of expenditure across subjects.

The studies looking at resource effects for primary schools (Gibbons et al, 2011; Holmlund et. al. 2010) find that effects are substantially higher for economically disadvantaged students. These findings are encouraging for policy because they suggest that despite large imperfections, mechanisms can be designed to ensure that disadvantaged students benefit from increasing school resources (see the discussion about the pupil premium in the UK in section 4.4). This provides some support to the recommendation for increasing targeted resources for the disadvantaged.

There is also evidence to suggest that targeted investments, which address problems in specific areas or subjects and are specifically designed for pupils with learning disadvantages, deliver larger benefits. A case in point is the ‘Excellence in Cities’ programme (Machin et al, 2010) and the ‘Literacy Hour’ policy (Machin and McNally, 2004).

4 The UK Institutional Framework

Having considered the drivers of success and failure in educational systems, we turn to a critical appraisal of the UK institutional framework, highlighting areas which are working or improving and pointing to problem areas that still need to be addressed.

4.1 Accountability

An important feature of an education system is the way in which its performance is held to account. A growing body of literature posits that the key to improving education outcomes lies in altering the incentives structure, so that it promotes strong schools with high quality teachers (Hanushek, 2008; Hanushek and Woesmann, 2006, 2011a). For example, there is empirical evidence suggesting that schools that face external exit exams tend to have better results than schools that face no such exam. The same literature reports a negative link between accountability and autonomy – i.e. in the absence of central accountability frameworks, schools with greater autonomy tend to underperform (Woessman, 2012).

The UK accountability system is based on two pillars: (i) school performance (or 'league') tables, which have traditionally focused on schools' average GCSE results; and (ii) inspection reports from the statutory agency responsible for monitoring schools' performance, the Office for Standards in Education (Ofsted). Both have significant limitations.

School performance (or 'league') tables are useful tools for parents and government to evaluate school performance and educational outcomes[65]. Allen and Burgess (2011) use seven years of pupil census to show that using the performance tables to select schools does on average lead to better choices than choosing at random. However, test scores and value-added as published in the league tables are not an accurate measure of school quality (Kramarz, Machin and Ouazad, 2009).

Furthermore, league tables may encourage behavioural distortions. For example, in order to improve the average exam results, individual teachers would focus their effort more towards exam preparation ("teaching to the test"); schools may also decide to develop a more selective intake approach or change the mix of subjects offered to students so that examination success is more probable. The consequences of such distortions are grade inflation, focusing on the average

[65] The use of benchmarking is more widespread in UK than in virtually any other OECD country (Gonand et al, 2007).

student's performance and not at the entire distribution. It also distorts funding allocations within the school.

The second pillar of the accountability system is the role played by Ofsted. Recent empirical evidence suggests Ofsted's inspections are effective in improving poor school performance (Hussein, 2012). Allen and Burgess (2012) show that schools only just failing to reach the minimum standards expected by school inspectors do indeed see an improvement in scores over the following two to three years, over and above those schools that only just make it above the threshold. The effect size is moderate to large at around 10 per cent of a pupil-level standard deviation in test scores. Moreover, this improvement occurs in core compulsory subjects which suggest that schools are not altering their subject mix.

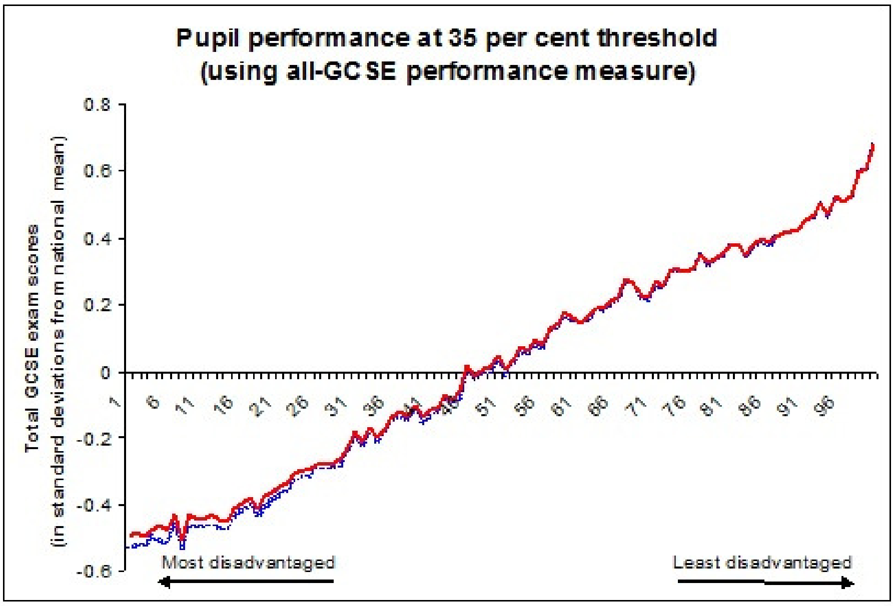
The results mentioned above, however, indicate little positive impact on lower ability pupils, with equally large effects for those in the middle and top end of the ability distribution. This raises doubts about the effectiveness of the incentives placed upon schools to improve the performance of disadvantaged children. These doubts are exacerbated by the fact that the performance of disadvantaged children appears to be diluted in the criteria Ofsted applies while judging the overall effectiveness of schools.

4.1.1 Floor targets and the Academies Act

One of the government's flagship policies to tackle poor school quality is based on the definition of a "floor target". This sets an expectation that a minimum of 35 per cent of children at every secondary school should get five A* to C including English and maths. A primary school will be below the floor if less than 60 per cent of pupils achieve the 'basics' standard of level four in both English and mathematics and fewer pupils than average make the expected levels of progress between key stage one and key stage two. Schools that fail to meet this target (and a few other criteria) are at risk of having their management replaced (the so-called "sponsor academy" conversion). Where there has been long-term underperformance, little sign of improvement and serious Ofsted concern, the Government converts schools into Academies, partnering them with a strong sponsor or outstanding school.

Unfortunately, the impact of this program on the socio-economic gradient is likely to be rather limited as is illustrated by figure 8 below. The blue line in the figure gives every 16-year-old who took GCSEs at a state school in 2010 a point score for their exam performance: 8 points for an A* down to 1 point for a G. It standardises the lot, and divides them up by the poverty of their neighbourhoods. Children in disadvantaged postcodes are at the left of the graph and the richest are at right. The red line strips out the failing schools (according to the floor target mentioned above) and assumes the children who previously attended those schools are dispersed into the rest of the school system in a way that does not damage the performance of those other schools. The resulting improvement in the gradient is very limited.

Figure 8: GCSE exam scores (pupil level performance) by neighborhood deprivation level



Source: Cook, 2012

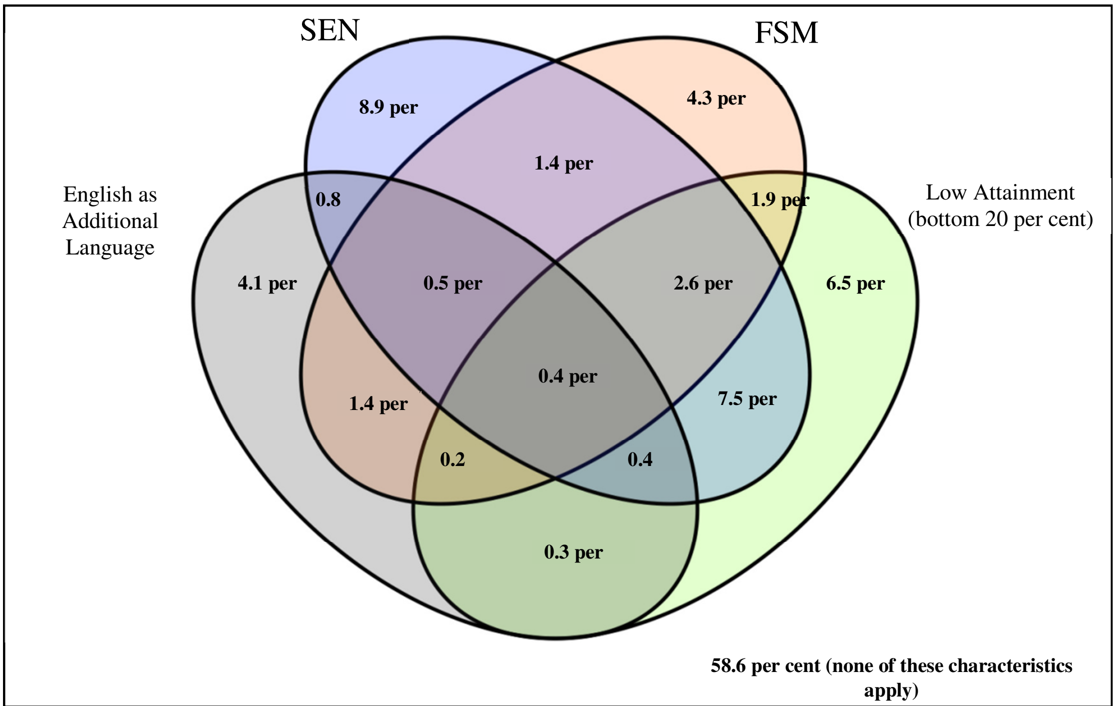
This is yet another reminder of the point we highlighted in section 1.2.2: the problem is not that there are a few schools which have all the disadvantaged children in them performing poorly. The problem is that disadvantaged children perform poorly (compared their wealthier peers) in a vast majority of schools.

4.1.2 Targeting symptoms, not causes

Over the years, most of central government's policy interventions have not been systematically targeted at economically disadvantaged/ FSM children but have instead focused on a number of pupil-characteristics that are (imperfectly) associated with economic disadvantage (such as Special Education Needs (SEN) [66] status; ethnic minorities; and low attainment)[67].

Although there are significant overlaps between these groups (see figure 9 below), there are drawbacks to this approach. First, it leaves out a multitude of cases of socio-economically disadvantaged children who are not income disadvantaged. Second, some of these groups conflate pupil deprivation with poor teaching performance (low attainment could be simply driven by poor teaching). Third, it provides mixed messages to schools and blurs their priorities - anecdotal evidence suggests that schools/teachers do struggle to understand why/ how to target needs of Free School Meals (FSM) pupils compared to more visible types of need (e.g. SEN, EAL).

Figure 9: GCSE targeting symptoms - a Venn diagram



Source: Data from the 2009 National Pupil Database

---

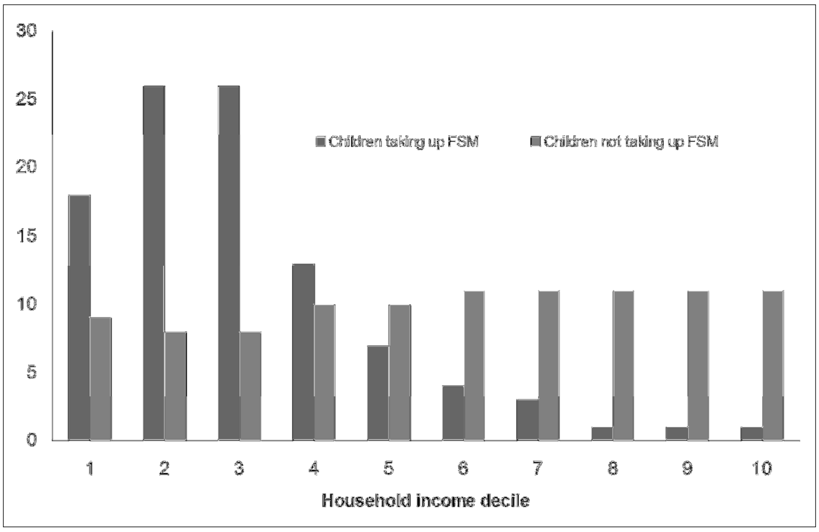
66 SEN is a multifaceted classification which brings together children with innate cognitive/learning difficulties and children who are underperforming for reasons other than their innate ability (e.g. strong negative impact of family background and/or poor teaching quality and/or unsupportive peer effects). The first sub-group is defined by a characteristic that puts it at disadvantage. The second sub-group is defined by its (poor) performance level, and may or may not be at disadvantage. In that sense, SEN conflates discretionary inputs (e.g. teaching quality); non-discretionary inputs (e.g. unsupportive family background); and outputs (low attainment).

67 Under the previous government, there were only a handful of interventions designed to directly target disadvantaged/ FSM children [e.g. apart from additional funding, there was a two year old childcare pilot; extended services subsidy and an Educational Maintenance Allowance (EMA)].

4.1.3 Main measure of deprivation subject to substantial limitations

Free School Meals (FSM) status is widely used as the main measure of deprivation. The Pupil Premium is also based on this indicator. FSM is a crude indicator of parental income. Hobbs et al (2010) have examined the relationship between children's FSM 'eligibility' and equivalent net household income (Figure 10) and find that there is considerable overlap between the range of household incomes of children taking up FSM and those not taking up FSM. In other words, many children taking up FSM are in households with higher incomes than children not taking up FSM. This makes it likely that many children eligible for and claiming FSM are not in the lowest income households.

Figure 10: Distribution of household income by children's FSM take-up status



Source: Hobbs and Vignolet (2010)

Other well established limitations of the FSM measure include the fact that: a) Dropping out of FSM category could simply mean children are not claiming FSM although they would be entitled to them (which is known to be particularly significant in the later stages of secondary education); b) Changes in FSM status may reflect increases in income beyond the thresholds defining the FSM category, but not to the extent of having meaningful impacts on attainment; c) Even if reductions in the proportion claiming FSM eligibility from one year to the next reflect significant improvements in household income, we would not expect this to lead to an instant improvement in pupil attainment. The effects of earlier poverty are likely to persist;

d) it is well established in the empirical literature that it is being eligible for FSM at any point in the pupil's academic career that is most strongly associated with attainment rather than the number of years a pupil is eligible.

4.1.4 Recent reforms

The government has partially addressed the issue of targeting the right groups by re-defining the target group for the Pupil Premium. From April 2012 the Pupil Premium was extended to include children who have been eligible for free school meals (FSM) at *any* point in the last 6 years.

Since January 2012, the government has also started to publish new league tables that report GCSE results by groups of pupils (*within* schools) defined by their prior attainment at key stage 2 (KS2)). Specifically, for each school the tables will report the percentage of pupils attaining at least 5 A* – C grades (including English and maths) separately for low-attaining pupils, high attaining pupils and a middle group. This is a change for the better as the main differences between schools in the performance of different group of pupils within the school will tend to emerge from variation in schools' teaching effectiveness.

However, a particular disadvantage of the new measure is that it uses very broad pupil bands. The groups are defined to cover the entire pupil population: the low attaining group are students below the expected level (Level 4) in the KS2 tests; the middle attaining are those at the expected level, and the high attaining group comprises students above the expected level. The disadvantage is that the broad groups (about 45 per cent are counted to be in the middle) hide the significant variation in average ability within that group across schools. This implies that differences in league table performance between schools will still reflect differences in intake in addition to effectiveness - even within the group, thus partly undermining the aim of group-specific reports.

4.2 Autonomy

Schools are only as good as their teachers. Since it is hard to find good *ex-ante* predictors of teaching quality, it is likely to be important to give schools the tools and incentives to hire and reward high performing teachers, and to remove low performing ones. The case for giving schools more freedom is based on the notion that this will allow them to take advantage of local knowledge to operate more efficiently and become more innovative.

Several countries have enabled a certain proportion of state funded schools to operate with greater autonomy than the norm within the state system. The structure and rules differ between (and sometimes within) countries but they also have much in common – for example, 'charter schools' in the US; 'free schools' in Sweden and 'academies' in England (see Appendix 3 for a detailed description of academies in the UK). In an international context, English schools are high up in the autonomy rankings, second only to the Netherlands according to OECD (2012).

The empirical evidence (both for the UK and other countries) provides support for the hypothesis that increasing school autonomy can lead to improvements in pupil performance and might also have positive effects on neighbourhood schools (see Appendix 4 for a review of the literature). In the UK, recent studies that have investigated the conversion of disadvantaged schools into academies have noted an improvement in pupils' performance compared to pupils in similar schools.

The important discussion for policy, though, is not so much whether autonomy is a good idea in general but in what spheres and contexts schools should be made more autonomous. Hanushek et al. (2011b) provide a good discussion on where 'autonomy' may and may not be desirable. In their view some decisions – such as hiring and budget allocations – require significant local knowledge and are more appropriately made at the school level. In contrast, where standardisation is important (for example in setting course offerings and requirements) decision should be made at a higher level[68]. Furthermore, the impact of autonomy may vary with other elements of the schools system - for example, whether there is a strong system of accountability in place.

In the UK, community schools (which still represent a large portion of the schools system) enjoy some autonomy[69] compared to the other types of school such as academies and voluntary aided schools. Localising hiring and making pay conditions more flexible would put these schools on a more similar footing to

[68] In a cross-country analysis, Woessman (2003) found that school autonomy in setting educational standards and the size of the school budget was negatively related to pupil performance. The opposite was true of school autonomy in personnel management and process decisions, for example, hiring teachers and setting salaries.

[69] The 1988 Education Reform Act gave community schools the option to become "grant-maintained" community schools where they were free from local authority control. The Act also gave community schools greater local management rights wherein schools could control their budget.

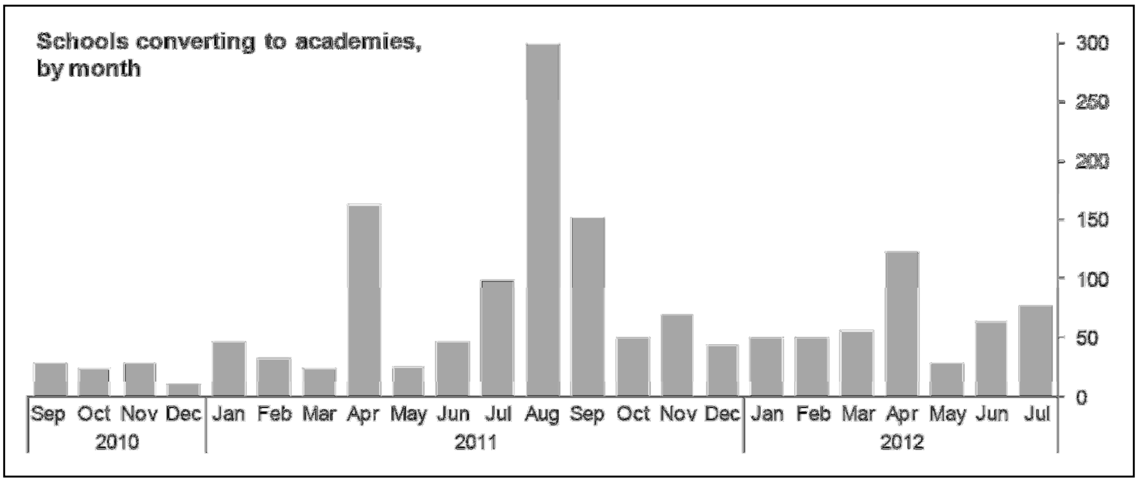
independent schools, academies, free schools and faith schools. It could also help overcome the problem of regional disparities in the real salary linked to the national pay scale (see Appendix 2 for a discussion of teacher's payment in UK).

In practice, this movement towards greater school autonomy in the UK is taking place through a piecemeal *academisation* of the schools system. In the Academies Act of 2010, the coalition government specified that any primary, secondary or special school that has been rated outstanding by the Office for Standards in Education (Ofsted) should be allowed to become an academy on a fast-track route.

From November 2010, all other primary and secondary schools that wished to benefit from Academy status will be able to apply to convert, provided they work in partnership with a high performing school that will help support improvement, or another sponsor – such as larger charities or small federation of schools. Over time, the government has taken out many of the requirements from the Academy funding agreement. It has removed prescriptions on curriculum and qualifications, target setting and the production of rigid plans.

The share of academies is rapidly increasing. Figure11 below shows the number of new converters each month. 29 schools converted in the first month of the start of the program, September 2010. Numbers remained below 50 per month for each of the next two terms. More than 150 converted at the start of the summer and autumn terms 2011, but the peak number of 300 converted during August 2011. Moreover, as shown in Table 3, the phenomenon of academies' chain is also emerging since some sponsors control more than one school. For example, the Academies Enterprise Trust (AET) is the largest sponsor and administers more than 60 schools. This would raise the challenging issue – both for academics and policymakers- to identify the optimal structure of the academies' system. Drawing from the US charter's schools system could provide useful hints. For example, KIPPS one of the major US charter schools, has adapted the *franchise model* to manage its expansion. Each KIPP school pays 1 per cent of its annual revenues to the KIPPS foundations; teachers and school leaders are carefully selected and trained; KIPPS schools are subject to annual inspections on financial, academic, real estate, and legal personnel issues and schools who fail to maintain the system's quality would lose the KIPPS brand and support. However, unlike the typical business franchisor, KIPP grants its new schools considerable freedom in deciding how they will earn and keep the brand. However, the majority of charter management organisation opt for a greater control over each school and adopt *a corporate-style growth* approach that assure that each new site replicates their own standards for building design, staffing and programs (Bennett, 2008).

Figure 11: Schools converting to academies, by month



Source: House of Common Library, SNSG/6233, July 2012

Table 3: Academies' chain in the UK

Sponsor	Prim, open	Prim, upcoming	Sec, open	Sec, upcoming	Total, inc schools not shown
Academies Enterprise Trust (AET)	18	15	20	12	66
E-ACT	7	3	18	1	30
Oasis	3	10	11		25
United Learning Trust (ULT)			17		21
Ormiston Trust	1		18		19
School Partnership Trust	4	6	5	3	19
Kemnal Academies Trust	9	5	2		16
ARK schools	3	1	7	1	15
Harris Federation	2	1	11	1	15
Greenwood Dale Foundation Trust	4	1	5		11
Academies Transformation Trust (ATT)	1	4		5	10

Source: Department of Education data and Guardian Datablog.

Another initiative in the direction of academisation has been the creation of Free Schools, which are being set up in response to real parental demand within a local area for a greater variety of schools. The first such schools opened in September 2011. Free schools are non-profit making, independent, state-funded schools. The model is flexible: free schools may be primary or secondary schools; located in traditional school buildings or appropriate community spaces; and may be set up by a wide range of proposers.

Like academies, Free Schools are to be funded on a comparable basis to other state-funded schools. The groups running these schools cannot make a profit, and are subject to the same Ofsted inspections as state schools. Free Schools cannot be academically selective and must take part in their local coordinated admissions process.

However, free schools have additional freedoms compared to academies, a key example being that teachers in Free Schools do not necessarily need to have Qualified Teacher Status.

4.3 School Choice and Competition

Increasing parental choice is often one of the front-runners amongst the policies proposed to promote competition and improve school outcomes (e.g. OECD, 2012). The rationale is that as schools compete to attract students, parental demands will create strong incentives for schools to improve performance.

In the UK, parents' ability to choose schools is limited. While parents can (since the 1980s) apply to any state school, schools are allowed to discriminate in case of over-subscription according to an enforced Code of Practice. The most important oversubscription criterion is usually proximity to the school. This means that some people have greater empowerment to exercise choice than others, and this tends to work against lower income families and those with difficulties in accessing and understanding school performance information provided through league tables and Ofsted reports.

There is evidence from England and other countries that many parents act on available information when they are purchasing a home (for England: see Rosenthal, 2003; Gibbons and Machin, 2003; Gibbons et al, 2013; and Burgess et al, 2009). Higher income parents move to locations with better schools and this is reflected in a high correlation between house prices and the quality of the school in the neighbourhood. The consequence is that parents from lower income households are not able to exercise meaningful choice because they cannot afford to live very close to a popular school. West and Pennell (2000) also show that higher socioeconomic households have better information about and understanding of school performance. Thus, 'school choice' as desirable as it is, is not an effective instrument for addressing attainment gaps by household background (McNally 2012).

As discussed in Section 4.2, there are some types of schools which are more autonomous and their admission criteria are not linked to residence criteria. These include faith-based schools, academies and independent schools. The ability to make effective choices is thus highly influenced by whether families can afford independent schools, have access to faith-based schooling, have children with specific aptitudes, or are able to move close to attractive maintained schools (Braconier, 2012).

International evidence on choice and competition is voluminous but its findings are still mixed. Evidence on competition and choice focusing specifically on the UK is very limited[70], and mainly focused on secondary education. Gibbons, Silva, and Machin (2008) are the first (pupil-level) analysis that investigates the effects of choice and competition on academic achievement in primary schools in England. The study reports little evidence of a causal link between either choice or competition and achievement. Encouragingly, they find some positive effects of competition for children in the tail of the performance distribution, in primary schools.

4.4 Funding System

Central government provides additional funding per disadvantaged student to local authorities (equivalent to roughly £4,000 per year in 2010). Local Authorities use their own individual funding formulas to transfer funds to schools. In 2010, on average, LAs passed through roughly £3,000 per disadvantaged student to schools, with the difference spread across all schools within the LA. A complex funding system makes it difficult for LAs to understand the share of deprivation funding in their total grants. The partial pass through may also reflect that LAs disagree with central government priorities. This may be one reason why LAs sometimes do not express support for extensive deprivation funding.

At the school level, where funding for deprivation is lower than the perceived costs, the school may engage in "cream skimming", try to dissuade disadvantaged

[70] See Allen and Vignoles (2009) for a good review of the existing literature.

children from admission, and recruit more able students. The lag in receiving deprivation funding also incentivises schools not to retain disadvantaged students (Sibieta et al, 2008).

The current government introduced a "pupil premium" to help mitigate these incentive biases and make funding directly tied to the disadvantaged (DfE, 2010). In 2011/12 schools received a premium amounting to £488 per child entitled to free school meals on top of base funding and for pupils in care who had been continuously looked after for six months. The premium amount increased to £600 per pupil for 2012–13. The Premium increases central government's notional funding for deprivation. But, it is still unclear whether the level of funding proposed is sufficient, whether schools will use the funds to tackle the performance of disadvantaged pupils, whether LAs will divert their grants away from deprivation funding, and whether schools will continue to "cream skim".

Recent research by the Sutton Trust also casts doubt on the impact of the premium. Less than 2 per cent in a survey[71] of primary and secondary school teachers said it would be used to improve feedback between teachers and pupils and less than 1 per cent said they will introduce peer-to-peer tutoring schemes. The Trust argues that these two schemes, if implemented well, could indeed boost recipients' performance by the equivalent of an extra eight or nine months in a school year.

Similar results are obtained in a recent Ofsted survey[72] that aimed at identifying how schools were using the pupil premium to raise achievement and improve outcomes for its recipients. The qualitative survey found that in the more disadvantaged areas, only one in 10 school leaders thought it had significantly changed the way they worked. Schools often failed to disentangle the pupil premium from their main budget, and said that they were not using the funding to put in place new activity. The pupil premium funding was most commonly used to pay for teaching assistants. In summary, there seems to be a significant risk that the pupil premium will not benefit the students who need it most, and that it will be used to fund existing programmes with no real impact in terms of additionality.

Ways of alleviating this problem include making schools directly accountable for the achievement of pupil premium recipients. One particular policy option is to

[71] NFER Omnibus Survey February 2012.
[72] The survey, conducted by Ofsted, is based on the views of 262 school leaders gathered from additional survey questions during routine inspections and telephone interviews.

publish school level information on the attainment of pupils eligible for the pupil premium. This transparency and accountability may incentivise schools to use the funding to improve the attainment of the target group. The coalition government has recently taken steps in that direction, by requiring schools to publish how they spend the pupil premium on their websites and by asking Ofsted to survey how the money is spent, and introducing the performance of pupil premium eligible students in performance tables of schools. Although these initiatives are arguably going in the right direction, it is doubtful they will be strong enough to counteract the bias in incentives created by the complexity and opacity of the way in which schools are funded in the UK.

4.5 Teacher Recruitment and Training

In the UK, the prestige (or lack thereof) of the teaching profession is reflected by the fact that only 3.7 per cent of graduates enter teaching; the average from Russell Group universities was 2.7 per cent, and for Cambridge, Bristol, Imperial College, UCL and LSE it was less than 2 per cent, with Oxford only just over 2 per cent.

Teachers are not civil servants in the UK, but are employed directly by the individual school. In order to teach in maintained schools, teachers must hold Qualified Teacher Status (QTS). There are a number of different routes available. Initial Teacher Training is a complex system, involving both undergraduate and postgraduate programmes in university-led, school-centred and employment-based provision. The question is how should this be set up to produce the most effective teachers who will have the greatest impact on pupil progress?

Traditionally teachers were trained either on undergraduate (BEd or BA QTS) or postgraduate (PGCE) courses run by higher education institutions. From 1994, School Centred Initial Teacher Training (SCITT) was introduced; this is a fulltime postgraduate training based in a school or a group of schools. Employment-based routes into teaching were first introduced in 1990. These are designed for qualified mature people who needed to earn a living while they were in training. They included the Graduate Teacher Programme (GTP), the Registered Teacher Programme (RTP) and the Overseas Trained Teacher Programme.

Initial teacher training, as a route to the teaching profession, plays two roles for the profession – training and selection with the emphasis typically placed on the former. Allen and Burgess (2012b) argue that selection seems to be the most important and it should be made at the point where evidence on ability is strongest. According to these authors, the final decision on who can become a teacher should be made at a stage when there is enough evidence on the candidate's teaching effectiveness i.e. after completing the training. Given that variations in teacher effects on pupil progress are very substantial, and that the future effectiveness of a potential teacher is hard to judge from their own academic record, a broader group (with a relatively low academic entry requirement) should be allowed to try out teaching. But, towards the end of the program, a much stricter probation policy should be enforced.

*T*he Coalition Government has proposed significant changes to the teacher training landscape wherein under the current operation of selection in ITT, the selection is tight at the beginning but negligible thereafter. The current policy[73]direction of tightening of academic entry requirements into teaching is not helpful: it will restrict the quantity of recruits and have no impact at all on average teaching effectiveness. The key decision on final certification should be made after a significant probation period (e.g. three years), and ideally, the probation should involve classes of varying ability and year group (Allen and Burgess, 2012a).

One of the successful recruitment routes has been the Teach First Program started in 2003 in London. Teach First is a charitable organisation. Teach First introduced a training programme for graduates who can commit to teaching for two years in challenging London secondary schools. According to an early stage evaluation[74] by Hutchings et al. (2006), after the first two years of its operation it had been successful in recruiting star graduates.

The key attractions for potential participants were keeping career options open, gaining qualifications and making a social contribution and the prestige surrounding the programme: participants were encouraged to view themselves as a privileged group. They underwent both teacher training (in the first year) and a programme of leadership training. Qualified Teacher Status (QTS) is normally gained at the end of the first year. Teaching in challenging London schools gave a sense of mission to the graduates.

[73] http://www.publications.parliament.uk/pa/cm201012/cmselect/cmeduc/1515/1515.pdf

[74]The Institute for Policy Studies in Education was commissioned by the Teacher Training Agency in 2003 to conduct an evaluation of innovative practice on the Teach First programme, with the aim of ensuring that ITT as a whole is able to benefit from innovative practice developed in the programme. The evaluation was conducted between September 2003 and September 2005, the first two years of the Teach First programme when it operated only in London.

Once teachers have been carefully recruited and trained, mechanisms for teachers, schools leaders and LAs to share best practice should be more strongly encouraged. The 'London Challenge' and the more general 'City Challenge' programmes have shown how successful this could be. The 'City Challenge' was launched in April 2008 building on the success of the London Challenge 2003-08. Its aim was to improve the educational outcomes of young people and 'to crack the associated cycle of disadvantage and underachievement' in the Black Country, Greater Manchester and London' (DfES, 2007). In particular, its goals were to reduce the number of underperforming schools, especially in English and maths; increase the number of good and outstanding schools; and improve the educational outcomes of disadvantaged children. City Challenge was based in a different approach than other government's interventions. First of all, it was built on the belief that educational problems should be addressed at local level, with Local Authorities and schools working together. Secondly it focused on all aspects of education (leadership, accountability through a better data collection; pupil's attainments; school-to-school collaboration) and involving all the parties (LAs, school leaders, teachers, parents and pupils). Thirdly it was characterized by a great flexibility that allowed modifying activities on the base of changing school's needs. Finally, there was not a single approach but the support package was bespoke for each school and agreed by schools' leaders, LAs, civil servants and the local team of Advisors. Based on the findings of a mixed methods evaluation, City Challenge areas achieved the majority of their initial targets.

Indeed, London schools in each quintile of 2008 attainment improved significantly than in areas not included in the City Challenge programme (with the exception of the highest quintile of secondary schools). In Greater Manchester and the Black Country, the picture was less clear since only schools in the lowest quintiles of attainment (and in some other quintiles) improved significantly more than those outside City Challenge areas. The attainment of pupils eligible for Free School Meals (FSM) increased by more than the national figure in all areas (with the exception of Greater Manchester primary pupils) and the attainment gap between pupils eligible for FSM narrowed for London primary and secondary pupils, and Greater Manchester primary pupils. Also the proportion of Good and Outstanding schools increased in all three areas, despite the introduction of a more challenging Ofsted inspection framework (DfE, 2012).In addition, 'London Challenge', thanks to the high involvement of schools and their staff in the decision and implementation process, had also a positive impact on inspiring teachers already in the system and attracting new one into the profession (Brighthouse, 2007).

Additional evidence on the importance of teacher's satisfaction to attract better teachers is provided by Green, Machin and Murphy (2008). They show an increasing outflow of teachers from the state to the private schools: the net annual flow of teachers from public to private has quadrupled over the last 15 years, rising from 400 in 1993 to 1,600 in2008. Moreover private schools employ more teachers with a postgraduate degree and the gap has grown over time. In the period since 2000, 60 per cent of male teachers in the private sector had a higher degree compared with 45 per cent in the state sector. Given that the wage gap between these two groups of schools is negligible (at least in non 'shortage subjects') what really matters in explaining the teachers flow are the better working conditions and in particular the higher level of satisfaction in the private schools.

4.6 Individual policies have not been properly evaluated

Evaluating educational reforms and identifying efficient policies is often difficult. Firstly, evaluations of long term labour market and social outcomes cannot be performed immediately after the programs have been initiated. Secondly, education systems are very context-specific; different countries perform well under different institutional settings. This means that policy evaluations have to be interpreted in a context-specific institutional framework.

In the UK, there has been a lack of rigorous and independent evaluation of policies implemented over the years. Even where they exist, they are not always considered in the policy-making process. The abolition of the Education Maintenance Allowance is a case in point. The Education Maintenance Allowance was launched nationally in the UK in 2004. It provided low-income 16 to 19-year-olds with payments of up to GBP 30 per week if they stay on at school or college. This policy was independently and rigorously evaluated, yet that evaluation seems to have been ignored when the policy was scrapped in 2010. The Department for Education cited research by the National Foundation for Educational Research which showed that 90 per cent of students who receive EMA would still continue with their education without the payment. However, this was a gross misrepresentation of the evaluation evidence and research. Extensive quantitative and econometric evaluations of the EMA by the Institute for Fiscal Studies (IFS) in 2005 showed that the scheme significantly improved both staying-on rates and qualifications for students from poorer backgrounds. The government had chosen to ignore this rigorous and independent evidence, and had instead argued that the abolition of EMA is justified by high levels of "deadweight".

There are nevertheless notable exceptions. The more recent and rather encouraging one is the Education Endowment Foundation (EEF), established in 2010 to look at what interventions work to overcome educational disadvantage. The EEF aims to build a rigorous evidence base of what works to raise the attainment of the lowest performing, and most disadvantaged children. It has very generous resourcing (Department for Education have given a grant of £125m over 10 years to the winner of a tender process). The creation of such an independent[75] organisation is a positive step towards rigorous policy evaluation and a similar approach should be encouraged more widely to inform the debate around education policy making.

There are many advantages to using this kind of platform, i.e. an independent, well -resourced organisation with a very clear remit to focus on the evidence around what works: (i) dedicated and focussed team without the distractions of the normal business of government; (ii) insulation from the demands of other government departments, to rule options out before they had a chance to be considered; (iii) insulation from “political” vetoes; iv) research continuity and strong institutional memory; and (v) ability to bring in multi-disciplinary expertise.

5 Policy Recommendations

5.1 Core recommendations on education

Our proposals go with the grain of the academies movement. But the system needs to deal more squarely with the UK’s failure to develop the talents of disadvantaged pupils. We therefore propose some direct steps, particularly financial and non-financial incentives, to address this fundamental problem.

The ‘academisation’ of the school system should deepen into a ‘flexible ecology’, building on aspects of the higher education system (see below). There are four integral parts: greater school autonomy, strengthened central accountability

[75] The EEF has no one from DfE on the Board, there are no politicians on the board; it is an independent organisation supported by charities.

(transparent information and inspection), wider parental choice and more flexibility for successful schools and their sponsors to expand.

To improve school governance, leadership and management, it must become easier for outstanding sponsored academies to grow. Ideally this operates at the school level by making physical expansion easier. But there may be spatial limitations, which is why expansion through the growth of networks of sponsored academies is also an important way to spread better practices. By the same token, it should be made easier for underperforming schools to shrink and, if they do not improve, to be taken over or, in extreme cases, closed down.

Changes to help to develop the talent of **disadvantaged pupils** include:

- Information on school performance needs to be changed to also reflect the performance of disadvantaged children within the school. Such changes should apply to league tables and targets and they should be more closely reflected in Ofsted's inspection regime. Improving the performance of disadvantaged children should be given a central role when Ofsted awards an 'outstanding' grade to a school.
- 'Floor targets' must be redesigned to become effective in addressing poor school performance and should be aligned with the guidelines defined in the framework for schools inspection. This should involve moving away from undifferentiated average performance targets (such as the current target, which requires 40 per cent of A* to C passes at GCSE level). These are 'blind' targets that distort schools' incentives to target resources and support towards those children who can more readily be expected to reach the pre-defined threshold.
- Contextual value added (school exam results adjusted for intake quality) should be published by school for pupil premium children and for the medium-performing Key Stage 2 group.

The expansion of new sponsored academies should be focused on underperforming schools serving disadvantaged children. The original programme was shown to be very successful in doing this (Machin and Vernoit, 2011). But the post-2010 academies are less focused on this group of schools.

Teacher quality needs to be improved through better conditions for both entry and exit. Teacher recruitment and training could be improved by:

- Teach First (which is renowned for its outstanding track record in recruiting high quality graduates) should expand until it becomes one of the main routes into school teaching.
- Mainstream teacher recruitment should become more concentrated in the best universities and schools, following a national recruitment process.
- The probation period for teachers should be extended in length – for example, by doubling it from two to four years.
- Policies that insist on grades, qualifications and backgrounds should be relaxed to encourage a wider range of applications to reflect the fact that teacher effectiveness is not highly correlated with crude background indicators.
- Mechanisms for teachers and schools to share best practice should be more strongly encouraged. The 'London Challenge' programme has shown how successful this could be.

Our proposed measures would, we believe, work together to increase the skills that are needed to make the UK economy a more competitive and dynamic place to do business and directly tackle the longstanding problem of poor intermediate and low-level skills. Together they would ensure that fewer of our children leave school ill-equipped to work in the competitive international environment that we now face. These proposals would also reduce disadvantage without compromising the achievements of other children.

5.2 Further recommendations for schools

To provide additional support for disadvantaged pupils, the criteria for receiving the pupil premium should be expanded to reflect a wider measure of disadvantage than simply free school meals. This need has now been acknowledged by making eligibility for the pupil premium dependent on whether a family has ever been eligible for free school meals in the last six years. But available databases could expand the definitions of eligibility further.

The pupil premium is planned to increase from £600 to £900 in 2014/15. We recommend that part of the premium should be given in cash to the pupils and families to provide an individual incentive. This should be conditional on improvements in performance after age 14, such as attendance and grade improvement beyond pre-agreed baseline expectations. This kind of ‘conditional cash transfer’ programme has proved to be effective in a wide variety of programmes (in welfare reform, for example, re-employment vouchers are usually more effective if the bonus is kept by the jobseeker rather than the firm). The precursor to this approach was the Educational Maintenance Allowance, which evaluations show was effective in encouraging children from disadvantaged backgrounds to remain in school. We recommend that the bursary scheme that replaced Educational Maintenance Allowance should be wrapped back into this.

More resources should be made available for programmes that provide better information to low income children and parents on the economic returns to different subjects. In the spirit of encouraging better teaching, a more flexible system of rewards should be introduced for pay and promotion. This would include ending automatic increments; basing pay on performance and local market conditions; and extra rewards for teachers of core subjects in tough schools. We need swifter action on improved professional development and movement out of the classroom for underperforming teachers. Some of these changes are starting to happen and we expect this process to accelerate under the flexible education system that we are recommending, which should give head-teachers the incentives and capabilities to make these reforms.

UK education policy has traditionally lacked rigorous, independent evaluations. Positive steps have been taken in this direction with the creation of the Education Endowment Foundation, but much more could be done. For example, we recommend piloting the release of teacher-level information on performance (in similar vein to NHS data available on surgeons).

References

Abdulkadiroglu, A., Angrist, J.D., Dynarski, S.M., Kane, T.J., and Pathak, P. (2011): “Accountability and Flexibility in Public Schools: Evidence from Boston's Charters and Pilots”, Quarterly Journal of Economics, 2011, 126(2): 699-748.

Aghion, P. and Howitt, P. (1998): “Endogenous Growth Theory”, MIT Press, Cambridge, MA.

Allen, R. and Burgess, S. (2011): “Can school league tables help parents choose schools?” Fiscal Studies, 32(2)245-261.

Allen, R. and Burgess, S. (2011): "Evaluating the provision of school performance information for school choice", DoQSS Working Papers11-10, Department of Quantitative Social Science - Institute of Education, University of London.

Allen, R. and Burgess, S. (2012): “How should we treat under-performing schools? A regression discontinuity analysis of school inspections in England”, CMPO working paper No. 12/87 and DoQSS working paper No. 12/02.

Allen, R. and Burgess, S. (2012 b): “Reforming teacher training”, CMPO Viewpoint.

Allen, R., Burgess, S. and McKenna, L. (2012): “The short-run impact of using lotteries for school admissions: early results from Brighton and Hove’s reforms”, Transactions of the Institute of British Geographers, forthcoming.

Allen, R. and Vignoles, A. (2009): "Can school competition improve standards? The case of faith schools in England," DoQSS Working Papers 09-04, Department of Quantitative Social Science - Institute of Education, University of London.

Ananat, E. O., Gassman-Pines, A., Francis, D. V., and Gibson-Davis, C. M. (2011): “Children Left Behind: The Effects of State-wide Job Loss on Student Achievement”, NBER Working Paper No. 17104.

Angrist, J. D., Dynarski S.M., Kane, T.J., Pathak P.A., and Walters, C. (2010): “Inputs and Impacts in Charter Schools: KIPP Lynn", American Economic Review Papers and Proceedings 100(2).

Angrist, J.D. and Lang, K. (2004): “Does School Integration Generate Peer Effects? Evidence from Boston's Metco Program", American Economic Review, American Economic Association, vol. 94(5), pages 1613-1634, December.

Atkinson, A., Burgess, S., Croxon, B., Gregg, P., Propper, C., Slater, H. and Wilson, D. (2009): “Evaluating the Impact of Performance-related Pay for Teachers in England", Labour Economics Volume 16, Issue 3, June 2009, Pages 251–261.

Atkinson, A., Burgess, S., Gregg, P., Propper, C. and Proud, S. (2008): “The impact of classroom peer groups on pupil GCSE results”, Working Paper 08/187, Centre for Market and Public Organisation, University of Bristol.

Barro, R. J. and Lee, J-W. (1993) “International Comparisons of Educational Attainment”, Journal of Monetary Economics, Vol. 32, No.3 pp. 363-394.

Barro, R. J. and Lee, J-W. (2001), “International Data on Educational Attainment: Updates and Implications”, Oxford Economic Papers, Vol. 53, No.3, pp. 541-563.

Barro, R. J. and Sala-i-Martin, X. (2004), Economic Growth (second edition), The MIT Press, Cambridge, MA.

Becker, G. (1974):“A theory of social interaction”, Journal of Political Economy. 82, 1063–1093

Bennett, J. (2008) Brand-Name Charters – The franchise model applied to schools, Education Next, Summer 2008.

Besley, T. and Machin, S. (2008): “Are Public Sector CEOs Different? Leadership Wages and Performance in Schools”, LSE, mimeo.

Bettinger, E. P. (2005): "The effect of charter schools on charter students and public schools", Economics of Education Review, Elsevier, vol. 24(2), pages 133-147, April.

Bettinger, E., P. (2010): “Paying to Learn: The Effect of Financial Incentives on Elementary School Test Scores." NBER Working Paper No. 16333

Betts, J.R. and Shkolnik, J.L. (2000): “The Effects of Ability Grouping on Student Achievement and Resource Allocation in Secondary Schools”, Economics of Education Review. 19(1): 1-15.

Bils, M. and Klenow, P.J. (2000): “Does Schooling Cause Growth?”, American Economic Review, Vol. 90, No.5, pp. 1160-1183.

Black, S. E., Devereux, P. J. and Salvanes, K. G. (2008): “Staying in the Classroom and out of the maternity ward? The effect of compulsory schooling laws on teenage births”. The Economic Journal, 118:1025–1054.

Blanden, J. (2004):"Family Income and Educational Attainment: A Review of Approaches and Evidence for Britain," Oxford Review of Economic Policy, Oxford University Press, vol. 20(2), pages 245-263, Summer.

Blanden, J. and Gregg, P. (2004): “Family Income and Educational Attainment: A Review of Approaches and Evidence for Britain”, CMPO Working Paper Series No. 04/101.

Blanden, J; Gregg, P., and Macmillan, L. (2007): "Accounting for Intergenerational Income Persistence: Non-cognitive Skills, Ability and Education", Economic Journal, Royal Economic Society, vol. 117(519), pages C43-C60, 03.

Blanden, J, Gregg, P., and Machin, S. (2005): “Educational Inequality and Intergenerational Mobility” in Stephen Machin and Anna Vignoles (eds.) ‘What's the Good of Education? The Economics of Education in the UK’, Princeton University Press, 2005.

Blanden, J. and Machin, S. (2004): “Educational inequality and the expansion of UK higher education”, Scottish Journal of Political Economy, Special Issue on the Economics of Education, vol. 54, pp. 230–49.

Bloom, N., and Van Reenen, J. (2007): “Measuring and Explaining Management Practices across Firms and Countries.” Quarterly Journal of Economics 122(4): 1351-1408.

Bloom, N., and Van Reenen, J. (2010): “Why Do Management Practices Differ across Firms and Countries?” Journal of Economic Perspectives, 24(1): 203-224.

Böhlmark, A. and Lindahl, M. (2008): "Does School Privatization Improve Educational Achievement? Evidence from Sweden's Voucher Reform", IZA Discussion Papers No. 3691, Institute for the Study of Labor (IZA).

Bolton, P. (2012): “Education spending in the UK”, House of Commons Library, Standard Note SN/SG/1078.

Booker, K., Gilpatric, S., Gronberg, T. and Jansen, D (2007): “The Impact of Charter School Student Attendance on Student Performance,” Journal of Public Economics, Vol. 91, Nos. 5–6, June 2007, pp. 849–876.

Braconier, H. (2012): “Reforming education in England”, OECD Economics Department Working Papers, no.939.

Bradley, S. and Taylor, J. (2008): “Do Peers Matter? Estimation of Peer Effects from Pupil Mobility Between Schools”, Department of Economics, Lancaster University.

Branch, G.F., Rivkin, S.G., Hanushek, E.A. (2013): “School Leaders Matter: Measuring the impact of effective principals”, Education Next, Vol.13 (1), pp. 62-69, Winter 2013.

Brighouse, T. (2007): “The London challenge: a personal view”, in T. Brighouse and L. Fullick (eds) Education in a Global City: Essays from London, Bedford Way Papers.

Burgess, S., Greaves, E., Vignoles, A. and Wilson, D. (2009): "What parents want: school preferences and school choice", DoQSS Working Papers 09-01, Department of Quantitative Social Science - Institute of Education, University of London.

Carrell, S. E., Sacerdote, B. and West, J.E. (2011): “From Natural Variation to Optimal Policy? The Lucas Critique Meets Peer Effects”, National Bureau of Economic Research Working Paper 16865.

Caselli, F. (2005): “Accounting for Cross-Country Income Differences.” In Philippe Aghion and Steven Durlauf (eds.), Handbook of Economic Growth, 1 (9): 679-741 Amsterdam: Elsevier.

Castello-Climent, A. (2010):"Channels through Which Human Capital Inequality Influences Economic Growth," Journal of Human Capital, University of Chicago Press, vol. 4(4), pages 394 - 450.

Castelló-Climent, A. and Doménech, R. (2008): "Growth, “Economic, Royal Economic Society, vol. 118(528), pages 653-677, 04.

Chevalier, A., Dolton, P. and McIntosh, S. (2007): “Recruiting and retaining teachers in the UK: an analysis of graduate occupation choice from the 1960s to the 1990s”, Economica, 74, 69-96.

Clifton, J.; and Cook, W. (2012): “A long division: Closing the attainment gap in England’s secondary schools”, IPPR Report September 2012.

Chowdry, H., Crawford, C. and Goodman, A. (2009): “Drivers and barriers to educational success: evidence from the Longitudinal Study of Young People in England”, DCSF Research Report No. RR102.

Chowdry, H.; Greaves, E. and Sibieta, L. (2010): “The pupil premium: assessing the options”, Institute for Fiscal Studies.

Clotfelter, C. T., Glennie, E., Ladd, H. F., Vigdor, J. L. (2006): “Teacher Bonuses and Teacher Retention in Low Performing Schools: Evidence from the North Carolina $1,800 Teacher Bonus Program.” Public Finance Review.

Cohen, D. and Soto, M. (2007), “Growth and Human Capital: Good Data, Good Results”, Journal of Economic Growth, Vol. 12, No.1, pp. 51–76.

Coleman, J.S. et al. (1966): “Equality of Educational Opportunity, U.S. Department of Health, Education, and Welfare”, Office of Education.

Cook C. (2012) FT data – blogs; http://blogs.ft.com/ftdata/author/christophercook/
CREDO (2009): “Multiple Choice: Charter Performance in 16 States”, Stanford University: Center for Research on Education Outcomes.

Dahl, G.B. and Lochner, L. (2012):"The Impact of Family Income on Child Achievement: Evidence from the Earned Income Tax Credit", American Economic Review, American Economic Association, vol. 102(5), pages 1927-56, August.

De La Croix, d. and Doepke, M. (2009): "To Segregate or to Integrate: Education Politics and Democracy", Review of Economic Studies, Wiley Blackwell, vol. 76(2), pages 597-628, 04.

De La Fuente, A. and Domenech, R. (2006): “Human Capital in Growth Regressions: How Much Difference Does Data Quality Make?”, Journal of the European Economics Association, 4(1):1-36.

Demirgüç-Kunt, A. and Levine, R. (2009): "Finance and Inequality: Theory and Evidence," Annual Review of Financial Economics, Annual Reviews, vol. 1(1), pages 287-318, November.

DfES (2007): “City Challenge for World Class Education”, DfES.

DfE (2012): “Evaluation of the City Challenge programme”, Research Report DFE-RR215, June 2012.

Dickson, M. (2009): "The Causal Effect of Education on Wages Revisited," The Centre for Market and Public Organisation09/220, Department of Economics, University of Bristol, UK.

Dills, A.K. (2005): “Does cream-skimming curdle the milk? A study of peer effects”, Economics of Education Review, 24(1): 19-28.

Dobbie, W. and Fryer, R.G.(2009): “Are High Quality Schools Enough to Close the Achievement Gap? Evidence from a Social Experiment in Harlem", NBER Working Paper 15473.

Dolton, P. (1990): “The economics of UK teacher supply: the graduate’s decision”, Economic Journal, 100, conference papers, 91-104.

Dolton, P. and Van Ver Klaauw, W. (1999): "The Turnover of Teachers: A Competing Risks Explanation", The Review of Economics and Statistics, MIT Press, vol. 81(3), pages 543-550, August.

Dolton, P. and Gutierrez, O.M. (2011):"Teachers' Pay and Pupil Performance", CentrePiece - The Magazine for Economic Performance352, Centre for Economic Performance, LSE.

Duflo, E., Dupas, P. and Kremer, M. (2010): “Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya”, Massachusetts Institute of Technology, mimeo.

Duncan G. J., and Murnane, R. J. (2011): “Introduction: The American Dream, Then and Now,” in ‘Whither Opportunity: Rising Inequality, Schools, and Children’s Life Chances’, edited by Greg J. Duncan and Richard Murnane. New York: Russell Sage Foundation. 2011.

Dustmann, C. and Machin, S. and Schönberg, U. (2010): "Ethnicity and Educational Achievement in Compulsory Schooling", Economic Journal, Royal Economic Society, vol. 120(546), pages F272-F297, 08.

Feinstein, L. (2003): “Inequality in the early cognitive development of British children in the 1970s cohort”, Economica, 70: 73-97.

Galor, O. and Zeira, J. (1993):”Income Distribution and Macroeconomics," Review of Economic Studies, Wiley Blackwell, vol. 60(1), pages 35-52, January.

Gennaioli, N., La Porta, R., Lopez-de-Silanes, F., and Shleifer, A. (2011): "Human Capital and Regional Development", NBER Working Papers 17158, National Bureau of Economic Research, Inc.

Gibbons, S. and Machin, S. (2003): "Valuing English primary schools," Journal of Urban Economics, Elsevier, vol. 53(2), pages 197-219, March.

Gibbons, S., Machin, S. and Silva, O. (2008): "Choice, Competition, and Pupil Achievement, "Journal of the European Economic Association, MIT Press, vol. 6(4), pages 912-947, 06.

Gibbons, S., Machin, S. and Silva, O. (2013): “Valuing school quality using boundary discontinuity”, Journal of urban economics, Online ISSN 0094-1190

Gibbons, S. and Silva, O., (2008): "Urban density and pupil attainment, "Journal of Urban Economics, Elsevier, vol. 63(2), pages 631-650, March.

Gibbons, S. and Telhaj, S. (2006): "Peer Effects and Pupil Attainment: Evidence from Secondary School Transition", CEE Discussion Papers No. 0063, Centre for the Economics of Education, LSE.

Gibbons, S. and Telhaj, S. (2008): “Peers and achievement in England's secondary schools”, SERC Discussion paper, SERCDP0001. Spatial Economics Research Centre, London, UK

Gibbons, S., McNally, S. and Viarengo, M. (2011):“Does additional spending help urban schools?: an evaluation using boundary discontinuities”, SERC discussion paper, 90. Spatial Economics Research Centre, London School of Economics and Political Science, London, UK

Glennerster, Howard (1995): “British social policy since 1945”, Blackwell, Oxford, UK. ISBN 0631189629

Glennerster, Howard (1995): “Opportunity, costs, education: finance needs change if we are to build a high skill Britain”, New economy, 2 (2). pp. 110-114. ISSN 1070-3535.

Gonand, F., I. Joumard and R. Price (2007): “Public Spending Efficiency in Primary and Secondary Education: Institutional Indicators”, OECD Economics Department Working Papers, No. 543.

Gordon B. D. and Lochner, L. (2012): "The Impact of Family Income on Child Achievement: Evidence from the Earned Income Tax Credit", American Economic Review, American Economic Association, vol. 102(5), pages 1927-56, August.

Gould E. D., Lavy V, and Paserman M. D. (2009):“Does Immigration Affect the Long-Term Educational Outcomes of Natives? Quasi-Experimental Evidence”, The Economic Journal, 119, 1243-1269

Green, F., Machin, S. and Murphy, R. (2008): “Competition for Private and State School Teachers”, Journal of Education and Work 21(5):383-404 (November 2008).

Gregg, P. and Macmillan, L (2010):.”Family income and Education in the Next Generation: Exploring income gradients in education for current cohorts of youth”, Longitudinal and Life Course Studies Vol. 1(3) (2010)

Gregg, P., Macmillan, L. and Nasim, B. (2012): “The Impact of Fathers’ Job Loss during the 1980s Recession on their Child’s Educational Attainment and Labour Market Outcomes” CMPO Working Paper No. 12/288

Gregg, P, Propper C and Washbrook E. (2008): “Understanding the Relationship between parental income and multiple child outcomes: a decomposition analysis”, CMPO working paper 08/193.

Griliches, Z. (1970): "Notes on the Role of Education in Production Functions and Growth Accounting", NBER Chapters, in: Education, Income, and Human Capital, pages 71-128 National Bureau of Economic Research, Inc.

Hall, R., and Jones, C. (1999): “Why Do Some Countries Produce So Much More Output per Worker than Others?”, Quarterly Journal of Economics 114(1): 83-116.

Hanushek, E. A. (2003): "The Failure of Input-Based Schooling Policies." Economic Journal 113(485): F64-F98

Hanushek, E. A. (2008): “Education Production Functions”, in Durlauf, S. and L. Blume (eds.), ‘The New Palgrave Dictionary of Economics’, Basingstoke: Palgrave Macmillan.

Hanushek, E.A. and Kain, J.F. and O'Brien, D.M. and Rivkin, S.G. (2005): "The Market for Teacher Quality", NBER Working Papers11154, National Bureau of Economic Research, Inc.

Hanushek, E. A. and Kimko, D.D. (2000), “Schooling, Labour Force Quality, and the Growth of Nations”, American Economic Review, Vol. 90, No.5, pp.1184-1208.

Hanushek, E. A., Rivkin, S. J. (2010): “Generalizations About Using Value-Added Measure of Teacher Quality”, American Economic Review. May, 2010.

Hanushek, E. A., and Shultz, G.P. (2012): “Education Is the Key to a Healthy Economy”, The Wall Street Journal, 30th April 2012.

Hanushek, E. A. and Woessmann, L. (2006): "Does Educational Tracking Affect Performance and Inequality? Differences- in-Differences Evidence Across Countries", Economic Journal, Royal Economic Society, vol. 116(510), pages C63-C76, 03.

Hanushek, E. A. and Woessmann, L. (2007): “The Role of Education Quality for Economic Growth”, World Bank Policy Research Working Paper No. 4122, Feb. 2007.

Hanushek, E. A. and Woessmann, L. (2008), “The Role of Cognitive Skills in Economic Development”, Journal of Economic Literature, Vol. 46, No.3, pp. 607-668.

Hanushek, E. A. and Woessmann, L. (2009), “Do Better Schools Lead to More Growth? Cognitive Skills, Economic Outcomes, and Causation”, NBER Working Paper No. 14633, National Bureau of Economic Research, Cambridge, MA.

Hanushek, E.A, Link, S. and Woessmann, L. (2011b): “Does School autonomy make sense everywhere? Panel estimates from PISA”, Discussion Paper series, Forschungsinstitut zur Zukunft der Arbeit, No. 6185.

Hanushek, E. A. and Woessmann, L. (2011a): “Overview of the symposium on performance pay for teachers”, Economics of Education Review, Elsevier, vol. 30(3), pages 391-393, June.

Heckman J, Stixrud J and Urzua S (2006): ‘The effects of non-cognitive abilities on labour market outcomes and social behaviour’, Journal of Labor Economics, 24(3): 411–482.

Henderson, V., Mieszkowski, P. and Sauvageau, Y. (1978): “Peer Group Effects and Educational Production Functions”, Journal of Public Economics. 10(1): 97-106.

Hobbs, G. and Vignoles, A. (2010): “Is children's free school meal 'eligibility' a good proxy for family income?”, British Educational Research Journal, pages 673–690,August 2010

Holmlund, H., McNally, S. and Viarengo, M. (2010): “Does money matter for schools?”, Economics of education review, 29 (6). pp. 1154-1164. ISSN 0272-7757

Hoxby, C.M. (2002a): "School Choice and School Productivity (or Could School Choice be a Tide that Lifts All Boats?)", NBER Working Papers No. 8873, National Bureau of Economic Research, Inc.

Hoxby, C.M. (2002b): "Would School Choice Change the Teaching Profession?", Journal of Human Resources, University of Wisconsin Press, vol. 37(4), pages 846-891.

Hoxby, C.M., and Murarka, S. (2009): "Charter Schools in New York City: Who Enrols and How They Affect Their Students' Achievement", NBER Working Papers 14852, National Bureau of Economic Research, Inc.

Hussain, I. (2012): “Subjective Performance Evaluation in the Public Sector: Evidence from School Inspections” Centre for Economics of Education Discussion Paper No. 135

Hutchings, M., Maylor, U., Mendick, H., Menter, I. and Smart, S. (2006): “An evaluation of innovative approaches to teacher training on the Teach First programme: Final report to the Training and Development Agency for Schools”, Institute for Policy Studies in Education.

Jackson, C. (2010): “The Effects of an Incentive-Based High-School Intervention on College Outcomes", NBER Working Paper 15722

Jerrim, J. (2013): “The reliability of trends over time in international education test scores: is the performance of England's secondary school pupils really in relative decline?”, Journal of Social Policy

Keep, E. Mayhew, K. and Payne, J (2006): “From Skills Revolution to Productivity Miracle – Not As Easy As It Looks?”, Oxford Review of Economic Policy, Vol. 22, No. 4, 539-559

Kounali, D., Robinson, T., Goldstein, H., and Lauder, H. (2008): “The probity of free school meals as a proxy measure for disadvantage”, online at www.bristol.ac.uk/cmm/publications/fsm.pdf

Kramarz, F. and Machin, S. and Ouazad, A. (2009): "What Makes a Test Score? The Respective Contributions of Pupils, Schools and Peers in Achievement in English Primary Education", CEE Discussion Papers No. 0102, Centre for the Economics of Education, LSE.

Lagakosy, D., Mollz, B., Porzio, T. and Qian, N. (2012): “Experience Matters: Human Capital and Development Accounting, Princeton University.

La Porta, R. and Shleifer, A. (2008): “The Unofficial Economy and Economic Development.” Brookings Papers on Economic Activity, 2008: 275-352.

Lawton, K. and Lanning, T. (2012): “No train, no gain: Beyond free-market and state-led skills policy”, IPPR April Report 2012.

Lavy, V. (2002): “Evaluating the Effect of Teacher Group Performance Incentives on Students Achievements”, Journal of Political Economy, December 2002.

Lavy, V. (2009): “Performance Pay and Teachers’ Effort, Productivity, and Grading Ethics”, American Economic Review 2009, 99:5, 1979–2011.

Lavy, V. And Paserman, M.D. and Schlosser, A. (2012):"Inside the Black Box of Ability Peer Effects: Evidence from Variation in the Proportion of Low Achievers in the Classroom", Economic Journal, Royal Economic Society, vol. 122(559), pages 208-237, 03.

Lavy, V. and Schlosser, A. (2011):"Mechanisms and Impacts of Gender Peer Effects at School", American Economic Journal: Applied Economics, American Economic Association, vol. 3(2), pages 1-33, April.

Lazear, E. (2003): "Output-Based Pay: Incentives, Retention or Sorting?", IZA Discussion Papers761, Institute for the Study of Labor (IZA).

Levacic, R. and Vignoles, A. (2002):"Researching the Links between School Resources and Student Outcomes in the UK: A Review of Issues and Evidence", Education Economics, Taylor and Francis Journals, vol. 10(3), pages 313-331.

Lindley, J. and Machin, S. (2012): "The Quest for More and More Education: Implications for Social Mobility", Fiscal Studies, Institute for Fiscal Studies, vol. 33(2), pages 265-286, 06.

Loeb, S., and Page, M., (2000): “Examining the Link between Teacher Wages and Student Outcomes: The Importance of Alternative Labor Market Opportunities and Non-Pecuniary Variation”, The Review of Economics and Statistics, 82(3), 393-408.

Lucas, R. E., Jr. (1988) On the Mechanics of Economic Development, Journal of Monetary Economics (22 July), pp. 3-42.

Machin, S. and McNally, S (2004): “The Evaluation of English Education Policies”, CEE discussion papers, CEE DP 131, Centre for the Economics of Education, London School of Economics and Political Science, London, UK

Machin, S., McNally, S. and Meghir, C. (2010):"Resources and Standards in Urban Schools", Journal of Human Capital, University of Chicago Press, vol. 4(4), pages 365 - 393.

Machin, S. and Murphy, R. (2010): “The Social Composition and Future Earnings of Postgraduates’, interim report of a research project commissioned by the Sutton Trust.

Machin, S. and Vernoit, J. (2011): “Changing School Autonomy: Academy Schools and their Introduction to England's Education”, CEE Discussion Papers 0123, Centre for the Economics of Education, LSE.

Machin, S. and Wilson, J. (2009): “Public and Private Schooling Initiatives in England”, in Rajashri Chakrabarti and Paul Peterson (eds) ‘School Choice International: Exploring Public-Private Partnerships’, MIT Press.

Mankiw, N.G., Romer, D. and Weil, D.N. (1992) A Contribution to the Empirics of Economic Growth, The Quarterly Journal of Economics, Vol. 107, No. 2. (May, 1992), pp. 407-437.

Manski, C. (1993): “Identification of endogenous social effects: The reflection problem”, Review of Economic Studies, 60, 531–542

McKinsey (2007): “How the World’s best-performing school systems come out on top”, McKinsey Report September 2007.

McNally (2012): “Education & Skills: Report to the LSE Growth Commission”.

McNally (2010): “Evaluating Education Policies: The Evidence from Economic Research”, CEP Election Analysis Series.

Milligan, K. and Stabile, M. (2011): "Do Child Tax Benefits Affect the Well-Being of Children? Evidence from Canadian Child Benefit Expansions," American Economic Journal: Economic Policy, American Economic Association, vol. 3(3), pages 175-205, August.

Milburn, A. (2012): “The Independent Reviewer on Social Mobility and Child Poverty”

Moav, O. (2005): "Cheap Children and the Persistence of Poverty", Economic Journal, Royal Economic Society, vol. 115(500), pages 88-110, 01.

Moffitt, Robert. (2001): “Policy Interventions, Low-Level Equilibria, and Social Interactions”, in Social Dynamics, edited by Steven Durlauf and Peyton Young, MIT press.

Mookherjee, D and Ray, D. (2003): "Persistent Inequality," Review of Economic Studies, Wiley Blackwell, vol. 70(2), pages 369-393, 04.

Muralidharan, K. and Sundararaman, V. (2009): “Teacher Performance Pay: Experimental Evidence from India", NBER Working Papers15323, National Bureau of Economic Research, Inc.

Murnane, R. and Olsen, R. J. (1989): "Will There Be Enough Teachers?," American Economic Review, American Economic Association, vol. 79(2), pages 242-46, May.

Murnane, R. and Olsen, R. J. (1990): "The Effects of Salaries and Opportunity Costs on Length of Stay in Teaching: Evidence from North Carolina", Journal of Human Resources, University of Wisconsin Press, vol. 25(1), pages 106-124.

Nelson, R. R. and Phelps, E. (1966): “Investment in Humans, Technology Diffusion and Economic Growth”, American Economic Review, Vol. 56, No.2, pp. 69-75.

OECD (2005): “Teachers Matter – Attracting, Developing and Retaining Effective Teachers”, Education and Training Policy, OECD Publishing, Paris.

OECD (2010a): “Viewing the United Kingdom school system through the prism of Pisa”, OECD Publishing, Paris.

OECD (2010b): “The High Cost of Low Educational Performance”, OECD Publishing, Paris.

Ostinelli, G (2009): “Teacher Education in Italy, Germany, England, Sweden and Finland”, European Journal of Education - Special Issue: The Training of Adult Education Professionals in Europe - Volume 44, Issue 2, pages 291–308, June 2009

Propper, C. and Britton, J. (2012): “Does Wage Regulation Harm Kids? Evidence from English Schools”, CMPO, Working Paper No. 12/293.

Rivkin, S. G., Hanushek, E.A. and Kain, J.F. (2005): “Teachers, Schools, and Academic Achievement.” Econometrica. Vol. 73, No. 2, pp. 417-458.

Romer, P. (1990a) Endogenous Technological Change, Journal of Political Economy, Vol. 99, No.5, part. II, pp. S71-S102.

Rosenthal, L. (2003): "The Value of Secondary School Quality", Oxford Bulletin of Economics and Statistics, Department of Economics, University of Oxford, vol. 65(3), pages 329-355, 07.

Ross A and Hutchings M (2003): “United Kingdom Country Background Report: Attracting, Developing and Retaining Effective Teachers”, OECD

Rowntree, B. S. (1901): “Poverty: A study of town life”, London: Macmillan

Sacerdote, B. (2001): "Peer Effects with Random Assignment: Results for Dartmouth Roommates," Quarterly Journal of Economics, v116: 681-704.

Sala-i-Martin, X., Doppelhofer, G. and Miller, R.I. (2004), “Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach”, American Economic Review 94, No.4, pp. 813-835.

Santiago, P. (2004): "The labour market for teachers". International Handbook on the Economics of Education, edited by G. Johnes and J. Johnes. Cheltenham, Edward Elgar: 522-578.

Schuetz, G., Ursprung, H. and Woessmann, L. (2005): "Education Policy and Equality of Opportunity", CESIFO Working Paper No. 1518

Shultz, G.P. and Hanushek, E.A. (2012): “Education is the Key to a Healthy Economy: If we fail to reform K-12 schools, we will have slow growth and more income inequality”, Wall Street Journal, May 1, 2012

Sibieta, L., Chowdry, H. and Muriel, A. (2008): “Level Playing Field? The implications of school Funding”, Reading: CfBT Educational Trust

Slater, H., Davies, N. and Burgess, S. (2009): “Do teachers matter? measuring the variation in teacher effectiveness in England”, Bristol Centre for Market and Public Organisation, Working Paper no 09/212.

Solow, R. M. (1957) Technical Change and the Aggregate Production Function, Review of Economics and Statistics, Vol.39, No.3, pp.312-320.

Stevens, A.H and Schaller, J. (2011): “Short-run effects of parental job loss on children's academic achievement”, Economics of Education Review, Volume 30, Issue 2, April 2011, Pages 289–299.

Stinebrickner, R. and Stinebrickner, T.R. (2006): “What Can Be Learned About Peer Effects Using College Roommates? Evidence from New Survey Data and Students from Disadvantaged Backgrounds”, Journal of Public Economics,.90(8-9): 1435-54.

Summers, A.A. and Wolfe, B.L. (1977): “Do Schools Make a Difference?” American Economic Review, 67(4): 639-52.

Sutton Trust (2011): “Toolkit of Strategies to Improve Learning - Summary for Schools, Spending the Pupil Premium, Research Report May 2011.

Syverson, C. (2011): “What Determines Productivity?”, Journal of Economic Literature 2011, 49:2, 326–365.

The Economist (2013): “Pisa envy”, January 19th 2013

Teddlie, C., and Reynolds, D. (2000): “The international handbook of school effectiveness research”. London: Falmer Press.

Todd, P. E., and Wolpin, K.I. (2003): “On the Specification and Estimation of the Production Function for Cognitive Achievement.” The Economic Journal, 113 (485):F3–F33.

Walcher, P. (2013): “Teachers’ pay rises to be based on performance, Michael Gove confirms”, the Guardian, 15th January 2013.

Welch, F. (1970): “Education in Production”, Journal of Political Economy, Vol. 78, No.1, pp. 35-59.

West, A. and Pennell, H. (2000): “Publishing school examination results in England: incentives and consequences”, Educational studies, 26 (4). pp. 423-436. ISSN 0305-5698.

Winston, G.C. and Zimmerman, D.J. (2004): “Peer Effects in Higher Education, in College choices: The economics of where to go, when to go, and how to pay for it”, NBER Conference Report series, Edited by C.M. Hoxby, University of Chicago Press: Chicago and London: 395-421.

Woessmann, L. (2003): "European education production functions: what makes a difference for student achievement in Europe?, "European Economy - Economic Papers 190, Directorate General Economic and Monetary Affairs (DG ECFIN), European Commission.

Wolter, S.C. and Denzler, S. (2004):"Wage elasticity of the teacher supply in Switzerland", Brussels Economic Review, ULB –Universite Libre de Bruxelles, vol. 47(3-4), pages 387-408.

Zabalza, A, Turnbull, P. and Williams, G. (1979): “The Economics of Teacher Supply", Cambridge University Press: Cambridge.

Zimmerman, D.J. (2003): “Peer Effects in Academic Outcomes: Evidence from a Natural Experiment”, Review of Economics and Statistics, 85(1): 9-23.

Appendix

Appendix 1: Family Income and Educational Attainment

Blanden (2004) gives evidence of a significant impact of family income on educational attainment in the UK. The results suggest that a reduction of one third in income from the mean increases the probability of a child getting A-C GCSEs by on average 3 to 4 percentage points, and reduces the probability of getting a degree by a similar magnitude. These results imply that the probability of a young people at the 90th percentile of the income distribution of getting a degree is 42 per cent, compared to 21 per cent for students at the 10th percentile.

The result of Gregg and Macmillan (2009) show that a unit change in the log of income predicts a gap of over one-tenth of a standard deviation in both IQ and Key Stage 1 scores. In this analysis, the magnitude of this effect is much larger than the contributions of both adverse family structures and poor parental labour market outcomes, and is also double the importance of disadvantaged local neighbourhood for IQ. Only low parental education is a more important predictor of low income children's cognitive deficits.

Chowdry et al (2009) estimate that differences in the availability of material resources for educational purposes play a key role in explaining why teenagers from poor families tend to make less progress between Key Stage 3 and Key Stage 4 than teenagers from rich families. After accounting for differences in material resources, the gap in Key Stage 4 test scores between young people from the richest and poorest fifths of their sample falls by 37 per cent compared to its value after controlling for parental education, and demographic and other family background characteristics.

Gregg et al (2012) have shown how, in England, a child's educational progress suffered if their father lost his job in the recession of the 1980s, something that did not happen for children whose parents remained in work. Similar results have been found after spikes in job losses in the United States (Ananat et al, 2011). Dahl and Lochner (2012) estimated the effect of income on children's maths and reading achievement in the US using data from the Earned Income Tax Credit. Their estimates suggest that a \$1,000 increase in income raises combined maths and reading test scores by 6 per cent of a standard deviation in the short run. Test gains are larger for children from disadvantaged families.

Appendix 2: Teachers and Headmaster' payment incentives

The empirical and theoretical literature on the functioning of the labour market for teachers (and headmasters) has increased in recent years. Overall, it suggests that teachers and headmasters respond to monetary and market incentives aimed at increasing their effort and 'output' (i.e. learning). Using different methodologies and data, Dolton and Van Der Klaauw (1999), Hanushek (2003), Murnane and Olsen (1989, 1990), Chevalier et al (2007) show that individuals respond to (relative) wage incentives in their decision to start teaching or leave the occupation. Loeb and Page (2000) find that teacher wages are a significant determinant of their performance and decision to stay in the profession - a 10 per cent increase in teacher wages would reduce quit rates among US teachers by 3-6 per cent. Dolton et al (2011) (using a panel data on 39 countries) show how both relative and absolute levels of teacher salaries strongly impact on pupil performance. Propper and Britton (2012) provide further evidence favouring the argument that teacher wages are important for school performance in England.

There is also a growing body of work investigating the impact of performance related pay. Whilst again there is some mixed evidence, the general consensus appears to be that performance related pay for teachers does improve student attainment in a variety of settings. Examples include Lavy (2009) in Israel, Muralidharan and Sundararaman (2009) in India, Jackson (2010) in Texas, Bettinger (2010) in Ohio and Atkinson et al (2004) in England. Hanushek et al (2003) and Lavy (2002) show that teacher performance related pay schemes could effectively attract good teachers and improve their motivation with positive outcomes on pupils' attainment. Woessman (2011) use cross-country data to show that the introduction of performance related pay is significantly associated with mathematics, science, and reading achievements across countries. In particular, countries that adopt this type of teacher compensation record about one quarter standard deviations higher scores. Atkinson et al (2009) evaluate the impact of a performance-related pay scheme for teachers in England, using teacher level data matched with pupil's test scores and value-added. They show that the introduction of a payment scheme based on pupil attainment improved test scores and value added, on average by about half a grade per pupil. They also find heterogeneity across subjects, with maths teachers showing no improvement. Green, Machin and Murphy (2008) show that private schools in the UK, that are characterized by an higher education level of their staff and attract each year a lot of teachers from the state school, are used to pay a premium for teaching shortage subjects', such as maths or science. They use pay flexibility as an effective strategy to attract more and better teachers in these subjects.

Looking at teachers' decisions to stay in the profession, Lazear (2003) argues that a reduction in teacher pay in the US and Sweden has caused an adverse selection and induces highest quality teachers leaving the job; the author further suggests that linking compensation to performance would improve teacher quality and school effectiveness. Clotfelter et al (2006) report that a monetary bonus given to qualified teachers in North Carolina greatly reduced their probability of leaving high-poverty schools. This incentive was especially effective for teachers with more years of experience, who are usually associated with better pupil outcomes (Hanushek et al, 2005).

With regards to headmasters, Besley and Machin (2008) investigate the link between the pay and performance of school principals. They show that, in line with the evidence on pay and performance of private sector CEOs, school principals' payment is linked to publicly observable performance measures and poorly performing principals face a higher chance of being replaced. The results of Branch, Rivkin, and Hanushek (2013) show that highly effective principals increase the performance of a typical student by between two and seven months of learning in a single school year; ineffective principals lower achievement by the same amount.

Appendix 3: What are Academies in the UK?

Academies are publicly-funded independent schools. They benefit from greater freedoms to innovate and raise standards. These include: a) freedom from local authority control; b) the ability to set their own pay and conditions for staff; c) freedoms around the delivery of the curriculum; and d) the ability to change the lengths of terms and school days. Head teachers are given the freedom to innovate with the curriculum, pay staff more, extend school hours and develop a personal approach to every pupil. Academy schools enjoying direct funding and full independence from central and local bureaucracy. The principles of governance are the same in academies as in maintained schools, but the governing body has greater autonomy. Academies are required to have at least two parent governors.

Funding: Academies cannot charge fees and receive the same level of funding per pupil from the local authority as a maintained school, plus additions to cover the services that are no longer provided for them by the local authority. However, academies have greater freedom over how they use their budgets. Funding comes directly from the Education Funding Agency (EFA) rather than from local authorities.

Staffing: When a school converts from a local authority (LA) maintained school to a new academy, staff from the predecessor school must be transferred to the new Academy school under the 1981 Transfer of Undertakings (Protection of Employment) or TUPE regulations in which case their existing terms and conditions of employment are upheld. Once open, the academy trust may consult with staff and their union representatives on changes to these terms and conditions, for example to enable the academy to operate over different term times or change the length of the school day. Thus, the governing body is able to authorize changes to the terms and conditions of employment and approve personnel practices regarding staff development and discipline.

Admission: Academies are also required to give priority to children 'who are wholly or mainly drawn from the area' in which the school is located. This means that the majority of pupils admitted must live close to the school. All schools, whether maintained or academy are required to comply with the 'Greenwich Judgement' which requires schools to not treat pupils living outside the LA area less favourably than those living in the same LA. To simplify, the LA boundary cannot be used as the admission catchment area. Academies will need to take part in their local coordinated admissions process, and so parents apply for places for their child in the same way as any other local school. Maintained schools which have previously selected some or all of their pupils by ability are able to continue this practice when they become academies, but schools becoming academies cannot decide to become selective schools if not previously selective. Independent selective schools joining the academies sector are not legally able to continue to select by ability. However, any school with a relevant specialism can select 10 per cent of its intake by aptitude in sport, modern foreign languages, visual arts or performing arts.

Expansion: While they were earlier restricted in their scope of expansion, a little-known reform in 2011 allowed academy and voluntary aided schools to expand in size without the permission of the local authority.

Appendix 4: School Autonomy and Educational Outcomes

Machin and Wilson (2009) provide some early evidence on academies, comparing the impact on GCSE performance for schools that turn into academies with a comparison group of similar schools. There was an improvement in the GCSE performance of schools that became academies, but it was no different from the improvement for schools in the comparison group. Wilson (2010) finds that intake into academies over the period 1997 to 2007, has consisted of a lower proportion of pupils from relatively disadvantaged backgrounds (measured by those FSM).

Machin and Vernoit (2011) evaluate the schools that became academies up to 2008/09. Their main findings were: Firstly, schools that became academies started to attract higher ability students. Secondly, there was an overall improvement in performance at GCSE exams. These results were strongest for schools that have been academies for longer and for those who experienced the largest increase in their school autonomy. Thirdly, schools in the neighbourhood of academies started to perform better as well. This might either be due to more competition or the sharing of school facilities and expertise with the local community.

Gibbons and Silva (2008) investigate the effects of the emergence of the private sector in education on the performance of public-sector schools and find no evidence that a higher concentration of privately managed schools improves the performance of neighbouring public-sector schools in England. However, the authors find that certain types of state-schools (Voluntary Aided schools) which have autonomous governance and admission procedures react positively to greater competition with local schools- their students' value-added attainment score improves by about 1.6 points for each additional competitor.

The evidence for other countries is in line with the UK. Bohlmark and Lindahl (2008) look at the long-term as well as short-term effects of academies in

Sweden. They find evidence of only small positive effects in the short-term, which do not persist. Other studies adopting non-experimental methods tend to produce more mixed results.

In the US, there are charter schools that are similar in mandate and autonomy to Academies in the UK. In the US, some charter schools use lotteries to allocate places when the school is oversubscribed. Abdulkadiroglu et al (2011) exploit this randomisation to estimate the impact of charter attendance on student achievement in Boston. They find that charter school attendance leads to significant increases in pupils' English language and maths scores compared to students not attending charter schools. Interestingly, they find that the highest achievement gains are for students who performed poorly before they attended the charter school.

Similarly, using data from New York City, Hoxby and Murarka (2009) find that pupils who won the lottery to attend charter schools experience significant improvements in both maths and reading scores between the third and eighth grade, compared to those pupils who lost the lottery and remain in traditional public schools.

Angrist et al (2010) evaluate the impact of a specific Charter School that is targeted at low income students that qualify for free school meals. They find significant increases in the math and reading scores in students who attend this Academy - increasing by 0.35 standard deviations and 0.12 standard deviations respectively for each year they spent enrolled at the Academy, compared to pupils not attending the Academy. Most importantly, they find that pupils with limited English proficiency, special educational needs or lower baseline scores achieve the highest gains in both scores.

The spill-over effects of charter schools in the US has been studied by Bettinger (2005) who looks at the spill-over impact of charter schools in Michigan, Hoxby (2002a) who evaluates the effect of charter schools in Michigan as well as Arizona and Booker et al. (2007) who look at the impact of charter schools in Texas. All three studies find improvements in the traditional public schools that can be attributed to the introduction of charter schools.