Essays on Counterfactuals

Neil B. C. Scott

A thesis presented for the degree of Doctor of Philosophy

Department of Economics

Royal Holloway, University of London

September, 2019

I, Neil Scott, Declare that:

- The work in this thesis was composed by myself.
- Where the research was carried out alongside others, or where information has been derived from other sources, I confirm that this has been indicated in the thesis.
- This work has not been submitted for any other degree or professional qualification

To my grandad Victor, who after years jesting that he was hanging on for my graduation, sadly passed away a few short months afterwards, aged 92.

Acknowledgements

I would firstly like to thank my supervisors, Jonathan Wadsworth and Ija Trapeznikova, for their time, feedback, and support throughout the production of this thesis. I would also like to thank the various faculty members and PhD students at Royal Holloway and elsewhere who have provided helpful feedback at the various seminars, conferences, and workshops that were kind enough to invite me to present the work contained within these chapters.

I extend particular thanks to my co-authors, Celia Blanco Jiménez, Chris Howorth, and Michael Naef. Celia, for your much needed expertise in behavioural and experimental economics; Chris and Michael, for not only believing in my ideas but supporting them professionally, and also personally, through what became two valued friendships.

Additionally, I'd like to thank the various PhD colleagues with whom friendships developed through sharing office space, late nights, stress, and the occasional pint; only those who have put themselves through the process truly understand the various inevitable moments of plight.

Finally, I thank my parents for their support not only over the past four years, but over the past three decades; this thesis is as much a product of your efforts as mine.

Contents

1. Abstract

The construction of a definitive counterfactual is key for a properly rigorous assessment of the impacts of any policy change, whether in our field of Economics, or elsewhere. In simple terms, we do not just ask the question “what happened?”, but instead “what happened compared to what would have happened anyway?” A well-established predicament is that we cannot observe both treated and untreated outcomes for the same set of individuals over the same time period: you are either affected by the policy, or you are not.

The experimental ideal would be the evaluation of the effectiveness of policies using large-scale and carefully constructed randomised control trials (RCTs), though this is often prohibited by cost, practical, or ethical considerations. Researchers have therefore developed a variety of econometric techniques to derive statistically sound counterfactuals in the absence of the experimental ideal. A relatively recent development to this end has been the Synthetic Control Method (SCM), as developed by Abadie et. al. (2003) and Abadie et. al. (2010), which is a data-driven approach to deliver a ‘synthetic’ version of the treated observations post-treatment using a weighted combination of the untreated regions, based on observed pre-treated outcomes.

Chapter 2 utilises the Synthetic Control Method and LFS data to assess the impact of immigration from the Additional Eight (A8) accession countries to the European Union on domestic wages in the U.K. I also propose a new method for pooling unitary authorities in LFS data into stylised local labour markets. Consistent with the literature, I do not find evidence of a negative wage impact (coefficients are small but significant, and positive). I find some suggestive evidence that, consistent with Manacorda et. al. (2012), that wage impacts are most positive for native males.

In Chapter 3, I use extensions to the Synthetic Control Method, as developed by Cavallo et. al. (2013) and Galiani & Quistorff, (2017) to conduct a comparative case study on five expansions of the Oyster Pay-As-You-Go (PAYG) contactless payment system for public transport in London. I evaluate impacts on passenger demand, house prices, unemployment, traffic, and pollution. I find significant positive impacts on passenger demand where the price change is greatest for peak commuters; I find no significant impact on passenger demand where the price change is small, so we implicitly find no evidence that the convenience of contactless PAYG ticketing plays a role in increasing demand. I find no evidence of significant impacts on house prices, unemployment, traffic, nor pollution.

Deviating from SCM to construct a counterfactual in experimental conditions, Chapter 4 examines the effects of risky incentives in the workplace. We present a randomised experiment in which we compare the performance of subjects participating in a real effort task. Participants are allocated to groups of 5, but work individually. In the first treatment, subjects receive a bonus in the form of a piece rate for performing above a minimum performance threshold. In the second treatment, subjects receive a bonus that is 5 times higher than the value of the piece rate, but will only be received with a probability of 1 in 5. Lastly, in the third treatment, we create a group lottery by combining all the bonuses earned by the 5 players in the group into a single lottery prize, and allocating the group prize to one group member at random. In all treatments subjects are anonymous and able to observe the productivity of the other members or the group. Although we do not obtain significant difference between the three treatments in the point estimates, the group lottery always performs better than the other treatments, and, controlling for observables productivity growth over time in both lottery treatments is significantly higher than for the piece rate. We speculate that seeing someone earning a high prize incentivises subjects to exert more effort, although risk averseness and observing a high performance from their peers work as discouraging factors in the group lottery treatment.

2. Impact of A8 Migration on Domestic Wages in the UK

2.1 Introduction

This paper uses the Synthetic Control Method to assess the impact on wages in U.K. local labour markets most affected by post-accession A8 migration. There exists a general consensus in the literature that the overall effect of A8 migration on labour market outcomes is negligible; this paper contributes to that literature by comparing results obtained using the Synthetic Control Method to Spatial Correlation estimates, and testing whether local labour markets disproportionately affected by the policy were more affected by the policy. I also propose an alternative method for pooling LFS unitary authorities according to observed commuting characteristics, as opposed to geographical proximity.

In May 2004 the European Union enlarged to encompass ten additional nations. Eight of these nations, the "Accession Eight" (A8) nations, were formerly part of the Eastern Bloc and had significantly lower per-capita income than the rest of the EU (Dustmann, Casanova, Fertig, Preston, & Schmidt, 2003). Britain was one of just three member states not to exercise an option to apply initial restrictions on freedom of movement of labourers from the A8 nations, along with Sweden and Ireland (Lemos & Portes, 2008). As a result, large numbers of A8 nationals became economic migrants to the U.K.; 580,000 A8 nationals registered to work in the U.K. between May 2004 and December 2006, with an estimated additional 200,000 arrived to work on a self-employed basis over the same time period (Ruhs, 2007). This gross inflow was of a scale far exceeding that estimated before the policy change; a study commissioned by the Home Office (Dustmann, Casanova, Fertig, Preston, & Schmidt, 2003) bounded the estimates at 5,000 to 13,000 per year, basing their estimates on the assumption that all member states would lift inflow restrictions simultaneously.

Standard neoclassical theory dictates that this represents an exogenous labour supply shock, resulting (ceteris paribus) in an outward shift in the labour supply curve, a higher absolute level of employment and a lower wage rate. However this result does not account for general equilibrium effects; clearly a larger population has implications for aggregate demand within the economy, which has its own implications for labour demand. Taking such general equilibrium effects into account, the theoretical result become ambiguous, which is reflected in the empirical literature. An influential study by Card (1990) (discussed in more detail in the Literature Review below) argued that the stochastic nature of the supply shock to the Miami labour market resultant of Mariel Boatlift meant it could be treated as a 'natural experiment', as there was no potential for the Miami labour market to adjust in anticipation of the influx, and famously found no statistically significant wage impacts. Similarly, A8 migration to the U.K. is often treated as a natural experiment in the literature (Ruhs, 2007), (Lemos & Portes, 2008), as whilst the policy change was anticipated, its scale was not.

Lemos & Portes (2008) assess the impact of A8 migration to the U.K. by correlating the change in A8 shares in geographically defined regional cells with changes in labour market outcomes (known as spatial correlation), and find no statistically significant impact on wages. A criticism of this approach is that because migrants rationally select into higher-wage regions, there is a spurious positive correlation between migration levels and wages thus inducing positive bias in spatial estimators (Borjas, 2017); this is also known as simultaneity bias. The authors test for this by instrumenting with lags of A8 migrants, and find that their estimators do not change significantly. However Ruist, Stuhler, & Jaeger (2017) argue that instrumenting using migration lags does not generate unbiased estimators on the basis that the spatial distribution of migration inflows is relatively static over time, thus violating the strict exogeneity assumption required for Instrumental Variable estimators to be consistent. In his recent re-evaluation of the Mariel Boatlift, Borjas (2017) proposes the employment of the Synthetic Control Method (SCM); this data-driven method uses weighted averages of numerous untreated regions to generate a synthetic placebo region with a labour market characteristically similar to the treated region, against which post-treatment outcomes can be compared. Crucially, because the treated region and the synthetic placebo are similar in terms of observable labour market characteristics, the influx of migrants is argued to be a quasi-experiment, negating the possibility of selection on observables, in turn meaning simultaneity is unproblematic. This paper employs this methodology in the context of A8 migration to the U.K. to test the conclusion of Lemos & Portes (2008) that A8 immigration has little effect on local area labour market outcomes.

Whilst in the U.K. all regions were by definition affected by the enlargement of the EU, regions can be identified with similar labour market characteristics yet with very different patterns of net A8 inflows; indeed there is a large body of evidence to the effect that migrants do not select into regions entirely based on labour market conditions (Bartel, 1989), (LaLonde & Topel, 1991), (Dustmann, Fabbri, Preston & Wadsworth, 2003). Similarly to Card (1990) and Borjas (2017) in their analyses of the Mariels, I compare regions that were highly affected by the A8 influx to regions with similar labour market characteristics that were not. Given the consensus in the literature that the effect of A8 migration on U.K. labour market outcomes is negligible, focussing on the areas most affected by the change in this way provides a powerful test to this conclusion.

2.2 Literature Review

2.2.1 Evidence from the United States

There is a considerable body of evidence on the impact of immigration of wages in the U.S. using a variety of methodological approaches, with somewhat heterogeneous findings.

Card (1990) used matched difference-in-difference methodology to analyse the impact of the Mariel Boatlift on labour market outcomes (wages and unemployment) in Miami. In April 1980 Fidel Castro made a surprise announcement to the effect that Cuban citizens were permitted to emigrate should they wish to do so; as a result, between May and October of the same year 125,000 Cuban citizens emigrated from the port of Mariel to the nearest U.S. landmass, Miami. David Card used the unexpected nature of the Mariel Boatlift in 1980 to argue it constituted a 'natural experiment': because the policy change in Cuba was unexpected, it represented a significant exogenous shock to labour supply for which there was no scope for the Miami labour market to react in anticipation of the policy. Further, because Cuban migrants selected into settlement in Miami on the basis of its geographical proximity to Cuba as opposed to economic factors, this eliminated the potential for simultaneity (the difficulty in identifying the impact of migration on labour market outcomes where the spatial distribution of migration is dependent on the same variables on which those outcomes are measured). Matched difference-in-difference methodology was used to estimate the impact of the boatlift on Miami labour market outcomes, comparative to a counterfactual of outcomes in nearby control cities with characteristically similar labour markets. No significant impacts on outcomes for blacks, hispanics, nor existing Cubans in Miami were identified, and a small positive wage impact was identified for whites. These results were surprising to economists, as neoclassical theory dictates that (ceteris paribus) a large exogenous supply shock such as this should exert a significant negative impact on wages and/or a positive impact on unemployment.

Following Card (1990), numerous papers have found similar results when assessing the impact of immigration on wages in the U.S. using different methodological approaches. The difference-in-difference methodology employed by Card relies on the assumption that there is no regional selection by migrants on observables, which is arguably reasonable in the Mariel case but less so in more general contexts (Borjas, 2017). As such, spatial correlation became a popular tool to assess the impacts of immigration, e.g. LaLonde & Topel (1991), Altonji & Card (1991), Butcher & Card (1991), Borjas, Freeman, & Katz (1996), Card (2001). Spatial correlation analyses multiple regions, with each region essentially 'treated' to differing extents dependent on the extent of immigration to that region, and a treatment effect is identified from the correlation between the magnitude of migration to that region and changes in outcomes (controlling for other explanatories); in other words, the counterfactual becomes a region similar in other aspects but differing in the degree of immigration flows. The modal finding in this literature is that treatment effects are not significantly different from zero, or small and heterogeneous, with unskilled natives and existing migrants most likely to be adversely affected.

A criticism of both of these reduced form approaches is that they are uninformative as to the causal processes driving the results, in that they do not capture partial nor general equilibrium effects (Gibbons & Overman, 2010). Further, simultaneity is problematic in spatial correlation techniques: migrants are more likely to select into high-wage regions, resulting in a spurious positive correlation between high-immigration regions and wages (Borjas, 2017). As such, a variety of structural approaches have been proposed, some of which are described below.

Borjas, Freeman & Katz (1992) treat different types of worker, defined according to nationality and skill level, as different factors of production in a Cobb-Douglas setup, and compare returns to each factor to a counterfactual constructed using an estimate of the elasticity of substitution between each factor. Conversely to the papers described above, large and significant negative wage effects were found for unskilled labourers. The authors later repeated their study with the same data using spatial correlation methodology (Borjas, Freeman, & Katz, 1996), and found similar results. However perhaps the most important contribution of this paper was to spatial econometrics, as they found the size of their estimator was positively related to the geographical size of the region pools on which the spatial regressions were estimated.

A number of studies have similarly employed structural approaches to estimate partial equilibrium effects using estimates of the elasticity of substitution using U.S. data, finding similar results. Borjas (2003) finds large negative wage effects for unskilled natives, small negative wage effects for semi-skilled natives, and no significant effects for skilled natives. Borjas, Grogger & Hanson (2006) find the impact of immigration on the probability of an individual being unemployed is larger for blacks than for whites. Borjas (2006) shows that when there is an influx of migrant PhD graduates to a local economy there is a significant (though heterogeneous across discipline) negative impact on existing PhD graduate wages, indicating that wage effects are dependent on the skill group being competed with as opposed to being a phenomenon specific to the low skilled.

This approach, however, has been criticised on the basis that it is reliant on the assumption of perfect substitutability between units of labour within the categories into which the authors place them. Ottaviano & Peri (2012) replicate the simulations in Borjas (2003), but relaxing this assumption, allowing for imperfect substitution between natives and migrants using U.S. census data. Contrary to the findings of the original paper, they find a small positive impact on the wages for all native groups aside from unskilled natives, for which a small negative impact was found.

Weyerbrock (1995) calibrates a computable general equilibrium model to estimate the macroeconomic impacts of immigration to E.U. member states from the former Soviet Union. Despite very large inflows to the E.U. as a whole (of around 1.7 million per annum), small negative labour market impacts are identified only in experiments where annual immigration is large (between 3.5 and 7 million), and small positive labour market impacts are identified in member states where wages are more flexible.

Boeri & Brücker (2005) use a similar approach to predict the labour market impacts of the 2004 EU enlargement on pre-accession member states. They find that a 1% increase in the size of the labour force will result in only a 0.5% reduction in wages, owing to increases in GDP per capita in the host country.

2.2.2 Evidence from the United Kingdom

The first comprehensive study specifically on the impact of immigration on U.K. labour market outcomes is Dustmann, Fabbri, Preston, & Wadsworth (2003). The authors primarily use spatial correlations to establish the impact of migration on wages and unemployment in the U.K. between 1981 and 1991, using Labour Force Survey and census data. The study finds small positive wage effects on existing workers resultant of immigration, and no statistically significant effect on unemployment.

Dustmann, Fabbri, & Preston (2005) uses spatial correlations with LFS data between 1992 and 2000 to investigate the impact of immigration on native labour market outcomes; the authors find positive but insignificant wage effects for all education levels, and no significant effects on unemployment. Similarly, Dustmann, Frattini, & Preston (2007) use LFS data from 1997 to 2005 to look at the impact of immigration on native wages using spatial correlations, and find a small but significant positive effect for native males, though the authors note that results are heterogeneous across education levels, and that it is less than clear what demographic of natives migrants compete with in the labour market on the basis that skilled migrants more likely than natives to be employed in unskilled occupations.

Lemos & Portes (2008) is the first major analysis to focus specifically on the effects of A8 migration to the U.K. on native wages and unemployment rates. Unemployment is measured using Jobs Seekers Allowance administrative data; wage data is obtained from the Annual Survey of Hours and Earnings (ASHE); A8 inflows are measured using Worker Registration Scheme (WRS) data; control variables that explain native employment outcomes are derived from the LFS. Once again, the authors use spatial correlations to estimate the impact on A8 inflows on native outcomes, and no significant effect is identified.

Manacorda, Manning & Wadsworth (2012) is probably the most comprehensive analysis of the long run impacts of immigration in the U.K. Similarly to Ottaviano & Peri (2006), the authors construct a stylised model of labour demand which treats migrants and natives as separate factors of production within each age and education category, estimating the elasticity of substitution between migrants and natives, and using a simulation-based approach to estimate the impact of immigration on wages. Spatial cells were defined by both geography and skill level in a "skill-cell" approach, thus allowing for imperfect substitutability between migrants and natives with different skill levels. The paper finds that immigrants and natives are imperfect substitutes, new migrants are, on average, better educated than natives, and any negative wage effects of migration are thus borne by existing migrants, particularly more recent migrants.

A full review of the various methods employed, and the results yielded, is provided by Dustmann et. al. (2016).

2.3 Data and Descriptive Statistics

2.3.1 Data

British Labour Force Survey (LFS) data is used from Q3 2001 to Q4 2006. Analysis beyond 2006 is complicated by the further expansion of the EU into Romania and Bulgaria in January 2007, representing a second 'treatment'; unlike after the 2004 expansion the U.K. did this time exercise its right to exercise staggered restrictions on freedom of movement, and it is more difficult to reasonably argue the influx resultant of the 2007 expansion represents a natural experiment due to the recent experience of the 2004 expansion. Further, there is potential for regional impacts of the 2008 recession to be heterogeneous, such that (given this paper constructs counterfactuals geographically) inclusion of data in 2008 and the immediately ensuing years might compromise internal validity.

The LFS is a rolling panel survey of around 0.5% of the total population, submitting a questionnaire to 60,000 households, which amounts to around 140,000 respondents. Each household is followed for five consecutive quarters, with 'core' questions being asked each quarter, and 'non-core' questions which vary from quarter to quarter. The survey collects data on household characteristics, economic activity, and health status.

The LFS is used to construct the measure of migrant stocks in this paper. An A8 migrant is defined for the purposes of analysis in this paper as an individual who was born in one of the A8 nations, but now resides permanently in the U.K. Other, non-survey-based measures of migrant flows exist, such as National Insurance Number (NINo) registrations, and Worker Registration Scheme (WRS) data. Lemos & Portes (2008) employ WRS data in their analysis of the A8 impact on U.K. labour market outcomes, although the authors note the difficulty in drawing any inference as to migrant stocks over time using flow data, because A8 outflows (even for recent migrants) are nonzero; Pollard, Lattore, & Sriskandarajah (2008) find that A8 migrants are more likely to return to their home countries within a short timescale comparative to other migrants. Using LFS data to measure migration, however, is not without its drawbacks; Rendall, Tomassini, & Elliot (2003) compare unweighted LFS estimates of changes in migrant stocks to 1991 and 2001 Census data, and find that the LFS underrepresents growth in migrant stocks by around 26%. Rienzo & Vargas-Silva (2017) suggest this is due to the fact that the LFS only surveys those living in traditional households, and excludes people living in hotels, caravan parks and various other heteroclite arrangements. When aggregating regional data this paper uses person weights, which are calculated by the ONS to give higher weight in aggregation to underrepresented demographics[1].

The Labour Force Survey divides data geographically into 409 unitary authorities (local government constituencies), 49 counties and 12 regions. For the purposes of this study unitary authorities, as the most disaggregated measure, are combined according to observed commuting habits to generate representative local labour markets.

[1] PWT14 accounts for underrepresentation in terms of age, gender, and region in the LFS comparative to Census data.

2.3.2 Descriptive Statistics

Migrant Stocks

Figure 2.1 shows the evolution of the share of A8 migrants in the labour force over the time horizon studied. Prior to the 2004 accession migrants from A8 nations were subject to the same labour market restrictions as non-EU migrants: full time students were permitted to work up to twenty hours per week during term time, and others were permitted to work only with a work permit, obtainable by those possessing “key skills” which were deemed to be in short supply, or those who were provably sufficiently wealthy to support themselves independently (NOP Business & Institute for Employment Studies, 2002). The graph shows that A8s steadily accounted for approximately 0.25% of the total labour force until the EU enlargement in Q3 2004, after which this figure sharply rose to around 1.3% by the end of 2006.

Figure 2.2 shows the share of A8s in the labour force as a proportion of all migrants. It is clear that before the A8 accession A8s represented a relatively constant share of migrants (around 2.5%), but post-accession they represented a significantly increasing share, rising to around 14% in 2006 Q3.

This is roughly in line with the rapid increases in A8 workers estimated using the flow data described above. The Home Office (2007) estimated 444,000 A8 migrants registered to work in the U.K. between May 2004 and December 2006. This would represent roughly a 1.4 percentage point increase in labour force share over this time period, though this does not account for A8 outflows, nor registrants who did not actually settle in the U.K.. Figure 2.1 shows approximately a 1 percentage point increase in A8 stocks over the same period.

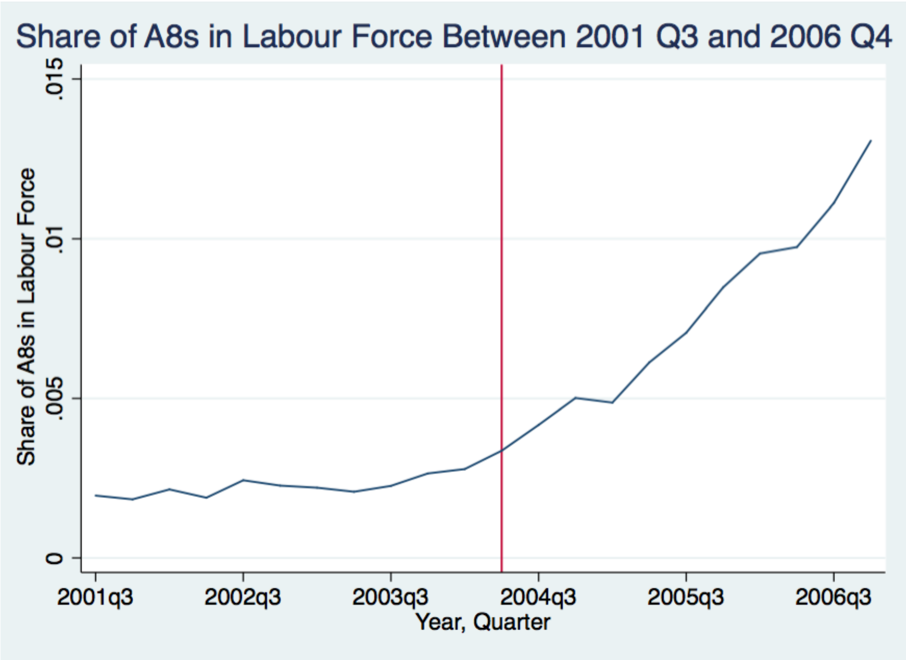


Figure 2.1

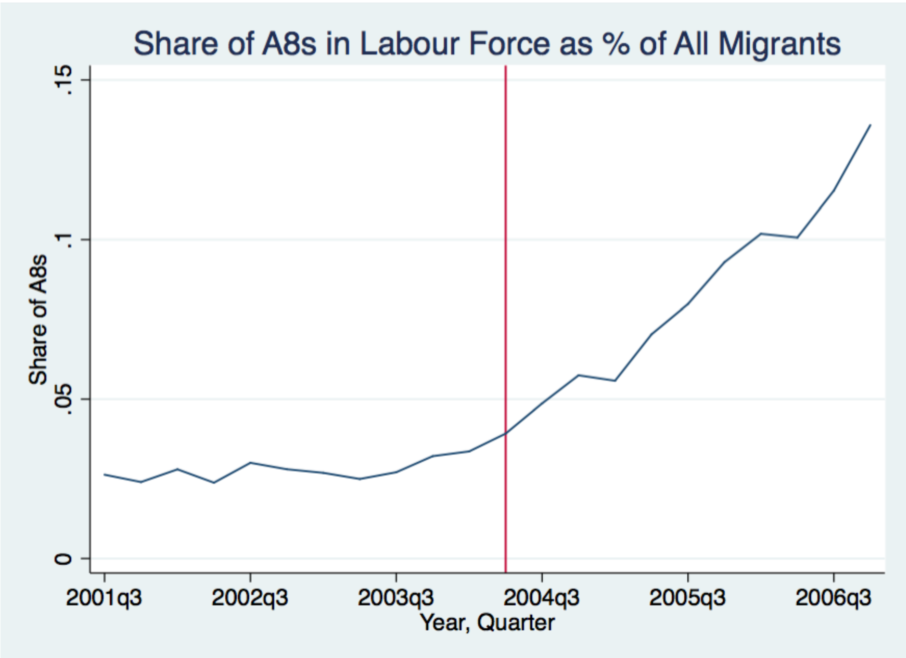


Figure 2.2

Occupations

Table 2.1 shows broad occupation groups for natives, A8s and non A8 migrants before and after the accession. For natives and non-A8 migrants these are relatively constant. For A8s the share in 'Professional', 'Associate Professional & Technical' and 'personal service' occupations decreased significantly, whilst the share of 'Process, Plant & Machine Operatives' and 'Elementary Occupations' doubled.

This is particularly interesting in the context of the education data presented in Table 2.3 below. Whilst on average new A8s have at least as many years of education as natives and existing migrants, they appear to select into lower skill work.

	Pre-Accession			Post-Accession		
	Natives	**A8s**	**Non-A8s**	**Natives**	**A8s**	**Non-A8s**
Managers & Senior Officials	14.1%	7.0%	15.9%	14.9%	4.5%	15.3%
Professional	11.3%	13.5%	17.0%	12.2%	5.6%	17. 8%
Associate Professional & Technical	13.4%	11.6%	15.4%	13.8%	4.9%	16.0%
Administrative & Secretarial	13.6%	6.8%	11.1%	12.9%	5.5%	9.8%
Skilled Trades	12.0%	15.5%	7.3%	11.5%	15.3%	7.2%
Personal Service	7.6%	15.3%	7.0%	8.0%	8.5%	7.9%
Sales & Customer Service	8.1%	4.6%	6.4%	8.0%	4.8%	6.8%
Process, Plant & Machine Operatives	8.2%	7.3%	7.5%	7.6%	14.7%	7.0%
Elementary Occupations	11.89%	18.45%	12.34%	11.12%	36.17%	12.26%

Table 2.1: percentage of natives, A8s, and non-A8 migrants in each major occupation group pre- and post-accession. Pre-Accession period Q3 2001-Q1 2004;Post-Accession period Q2 2004-Q4 2006.

Age

Post-accession A8 migrants are significantly younger than existing A8 migrants, other migrants and natives. Table 2.2 shows the age-groups of those of working age and in the labour force for all three groups before and after the A8 accession.

	Pre-Accession			Post-Accession		
	Natives	**A8s**	**Non-A8s**	**Natives**	**A8s**	**Non-A8s**
16-20	9.7%	8.8%	5.7%	9.9%	6.3%	5.3%
21-30	16.8%	36.2%	22.4%	16.1%	53.7%	22.2%
31-40	23.8%	22.2%	28.0%	22.2%	20.6%	28.2%
41-50	21.2%	18.0%	23.8%	22.3%	11.6%	23.4%
51-60	28.4%	14.7%	20.0%	29.5%	7.7%	20.8%

Table 2.2: percentage of natives, A8s, and non-A8 migrants in each age group pre- and post- accession. Pre-Accession period Q3 2001-Q1 2004;Post-Accession period Q2 2004-Q4 2006.

Native and non-A8 migrant ages are essentially unchanged before and after the A8 accession, however the population share of the 21-30 age-group among the A8 population living in the U.K. increased by 17.5 percentage points, indicating that the bulk of post-accession A8 migrants were within this range. It is not possible to accurately observe which observations are 'new' migrants using the LFS, however this result is consistent with the literature; Lemos & Portes (2008) find that 82% of new WRS registrations over the same time period were aged between 16 and 34.

Education Levels

A8 migrants are similarly educated to other migrants, and on average have around 2.5 years more education than natives. New A8 migrants appear to be at least as educated as existing A8 migrants. Mean education levels are shown in Table 2.3.

The LFS does contain data on the highest qualification attained, however most qualifications attained at foreign institutions are classified as "other", so it is not possible to carry out comparative analysis on highest educational attainment. See Manacorda, Manning & Wadsworth (2012) for a fuller discussion of this issue.

	Mean Years of Education		
	Natives	**A8s**	**Non-A8s**
Pre-Accession	12.35 (2.40)	14.84 (3.29)	14.31 (3.20)
Post-Accession	12.19 (2.46)	14.90 (3.33)	14.44 (2.88)

Standard deviations in parentheses

Table 2.3: Mean years of education of natives, A8s, and non-A8 migrants pre- and post- accession. Pre-Accession period Q3 2001-Q1 2004;Post-Accession period Q2 2004-Q4 2006.

Wages

Figure 2.3 shows gross hourly pay for natives, A8 migrants and non-A8 migrants. Non-A8 migrants earn around £1 per hour more than natives, though this is not a surprising result given migrants are on average more educated, as described above; this may also be driven by the fact that migrants settled disproportionately in London, where (partly due to London Weighting) there exists a wage premium. Both series are on an upward trend, with wages increasing at roughly the same rate; there are no obvious kinks around the A8 accession in Q2 2004.

Initially A8s are paid somewhere between natives and non-A8 migrants, but there is a decreasing trend across the series. It is not immediately clear what is driving this; one would expect to observe a decrease in mean A8 earnings post-accession, as the new migrants are younger and select into less skilled occupations, however it is not clear what is driving the decreasing pre-treatment A8 wage trend. It should be noted, however, that because A8s represented such a small share of the pre-accession labour force that the pre-accession subsample size is small.

It is important to note that a drop in mean A8 pay is not necessarily representative of new A8s competing with existing A8s in the labour market. Whilst new A8s appear to be educated to a similar level as existing A8s, it is clear from the statistics on the occupations of new A8s presented above that new entrants are more likely to select into low-skill occupations; the decrease in mean A8 wages might therefore simply be capturing this.

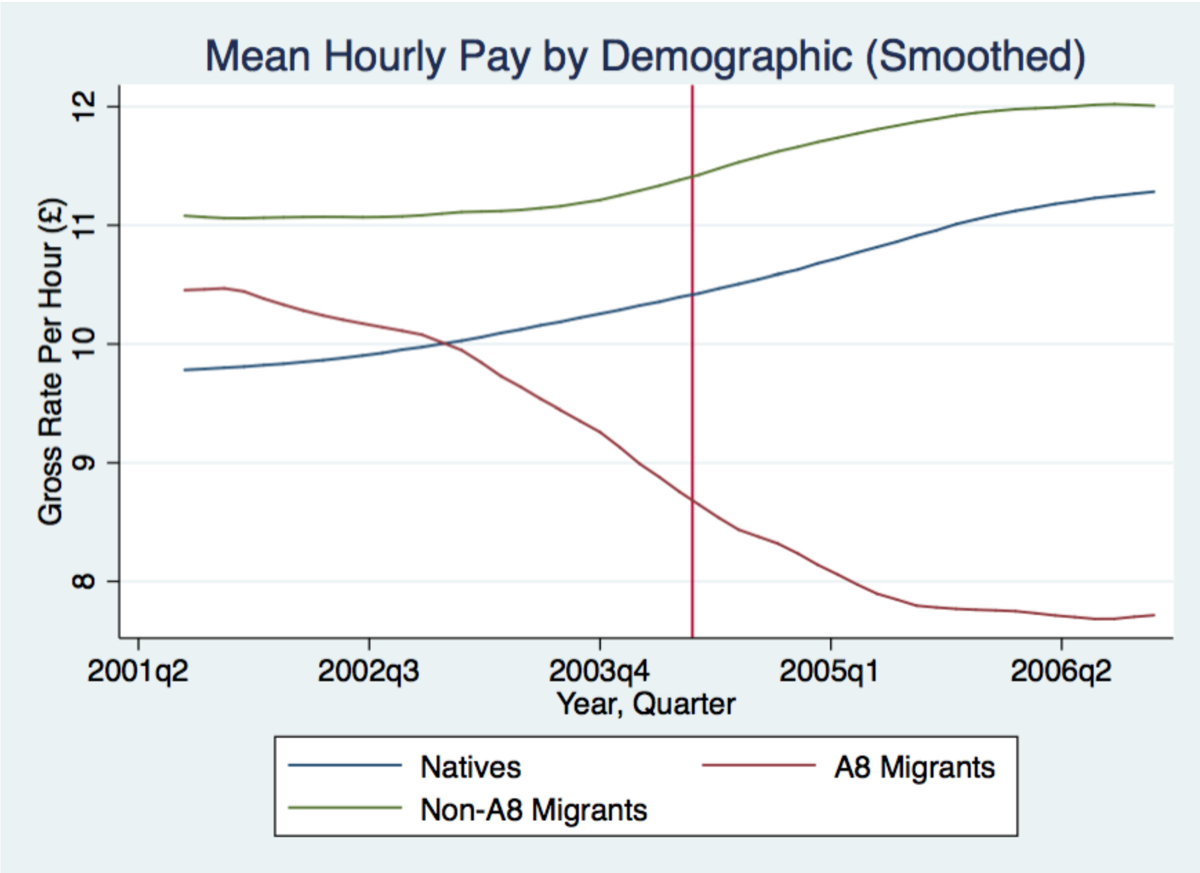


Figure 2.3

Unemployment

Figure 2.4[2] shows unemployment rate series for natives, A8 migrants and non-A8 migrants. Over the time horizon the native unemployment rate is relatively flat, at around 5%. Non-A8 migrant unemployment is also relatively constant, varying between around 7% and 8%.

[2] This graph, and others in this paper (where denoted as such) plot a local polynomial smooth of the outcome variable on time (using Stata's lpoly command). This is purely for visual clarity of the trends, net of seasonality.

Once again, there is significant variation in the aggregated A8 data owing to the relatively small number of A8s in the labour market. Unemployment is trending downwards for A8s before and after the accession, and there is no obvious structural break in this trend around the point of the 2004 accession.

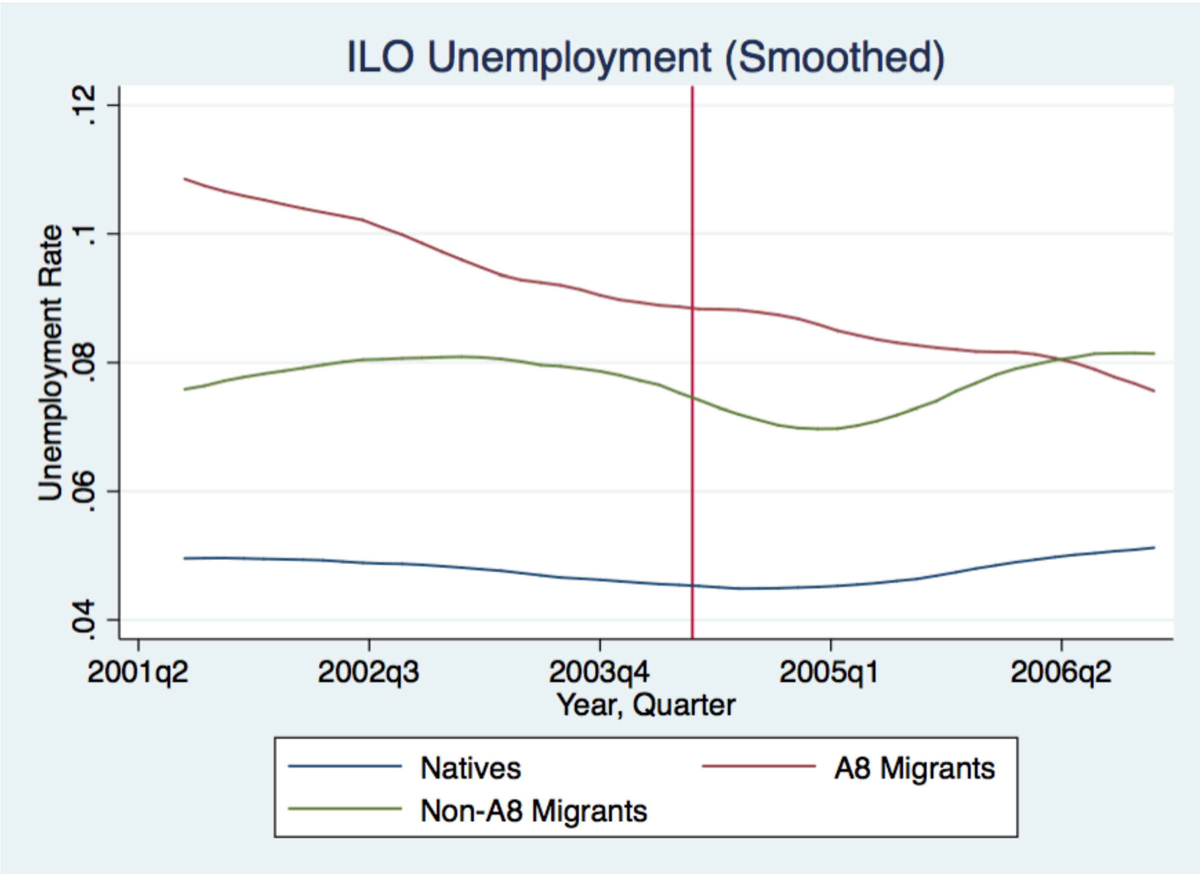


Figure 2.4

Dispersion of Settlement

Table 2.4 shows the percentage of all A8 migrants and all other workers living in each region before and after the 2004 A8 accession. The geographical dispersion of non-A8s is remarkably constant, whereas the dispersion of A8s changes quite significantly post-accession. Prior to 2004 the vast majority (80.9%) of A8s were settled in London and the South East. Post-accession London and the South East still had the largest concentration of A8s (53.9%), however large numbers settled in other regions, such as the East Midlands (8.3%, an increase of 6.7p.p.) and West Yorkshire (5.8%, an increase of 3.7p.p.).

Region	A8s Before	A8s After	Change (P.P.)	All Others Before	All Others After	Change (P.P.)
Tyne & Wear	0.3%	0.8%	0.5%	1.7%	1.7%	0.0%
Rest of Northern region	0.7%	0.7%	0.1%	3.1%	3.2%	0.1%
South Yorkshire	0.4%	1.6%	1.2%	2.0%	2.0%	0.0%
West Yorkshire	2.1%	5.8%	3.7%	3.5%	3.5%	0.0%
Rest of Yorks & Humberside	0.6%	2.6%	2.0%	2.7%	2.8%	0.0%
East Midlands	1.6%	8.3%	6.7%	7.2%	7.3%	0.1%
East Anglia	3.1%	4.7%	1.6%	3.8%	3.8%	0.0%
Inner London	28.0%	20.8%	-7.3%	4.9%	4.8%	-0.1%
Outer London	32.0%	19.1%	-12.9%	7.6%	7.5%	-0.1%
Rest of South East	20.9%	14.0%	-6.9%	19.9%	19.8%	-0.1%
South West	3.1%	3.9%	0.8%	8.5%	8.5%	0.0%
West Midlands (met county)	2.3%	3.2%	0.9%	4.1%	4.1%	0.0%
Rest of West Midlands	0.4%	2.2%	1.8%	4.8%	4.7%	0.0%
Greater Manchester	0.3%	2.4%	2.1%	4.2%	4.2%	0.0%
Merseyside	0.5%	0.3%	-0.2%	2.1%	2.1%	0.0%
Rest of North West	0.4%	1.7%	1.4%	4.0%	4.0%	0.0%
Wales	1.0%	1.4%	0.4%	4.6%	4.7%	0.1%
Strathclyde	0.7%	1.1%	0.4%	3.6%	3.6%	0.0%
Rest of Scotland	1.4%	3.0%	1.6%	5.0%	5.0%	0.0%
Northern Ireland	0.2%	2.2%	2.0%	2.6%	2.6%	0.0%
Total	100	100		100	100	

Table 2.4: regional settlements of A8 migrants and all other workers pre- and post-accession. Pre-Accession period Q3 2001-Q1 2004;Post-Accession period Q2 2004-Q4 2006.

2.4 Identification Strategy

This paper utilises the Synthetic Control Method (SCM) as developed in Abadie, Diamond, & Hainmueller (2010) in similar vein to Borjas (2017). The obvious difference between the A8 influx into the U.K. and the Mariel Boatlift is that the

Mariels settled almost exclusively in Miami due to its geographical proximity to Cuba, whereas A8 settlement in the U.K. was more widely dispersed (see Figure 2.4 for the geographical dispersion of A8 migrants pre and post accession). However the assumption Card (1990) and Borjas (2017) implicitly make is that there is no economically driven reason why Mariels settled in Miami as opposed to the control regions identified, that is, an individual migrant would have faced the same expected wage rate and probability of being unemployed infinitesimally before the policy change in any of the control regions as they would in Miami. Indeed, there is a significant body of evidence indicating that regional settlement decisions are based in part on non-economic factors, e.g. Bartel (1989), LaLonde & Topel (1991), Dustmann, Fabbri, Preston & Wadsworth (2003). This paper makes the same assumption for A8s in the UK: 'control' regions are identified where there is no increase in A8 settlement after the policy change, and these are matched to 'treated' regions, which did experience a significant increase, with similar labour market characteristics prior to the policy change.

2.4.1 Identification Challenges

Selection on Observables

Selection on observables is probably the most difficult challenge to overcome: clearly new A8 migrants do select their region of settlement. This assumption is required for the conditional independence assumption to hold, which is essential for causal interpretation of the estimators derived below (Angrist & Pischke, 2009). Regional selection by new migrants on the basis of wage rates would be particularly problematic here: if new migrants select into higher wage regions, and therefore comparatively lower wage regions are used to construct a synthetic control group, any estimator of the effect of migration on wages derived through double differencing between 'treated' regions and the synthetic control region would be positively biased (Dustmann, Fabbri, Preston & Wadsworth, 2003). This is known as simultaneity: the researcher wants to observe changes in wage rates resultant of changes in regional migration, where in practice regional migration is being driven by wage rates.

The same argument can be made around unemployment. There is a strong theoretical negative relationship between unemployment and wages (Blanchflower & Oswald, 1994), and so if new migrants select into regions where they perceive there to be a low probability of being unemployed, this would create an inflow of workers into regions where wages are likely to rise.

Using synthetic control methodology, it is possible to construct a single control group that is characteristically similar to the treated group(s) in terms of the explanatories on the outcome variable (here, hourly wage). This means pre-treatment trends in the treated regions and their respective synthetic control regions are almost exactly similar, the inference being that in the absence of the policy post-treatment trends should be similarly congruent. In the context of wages, regional unemployment rates are not specified as a control due to reverse causality; for this reason treatment and 'donor' control regions are matched according to unemployment rates immediately before A8 accession, as is described in more detail below. This does still allow for the possibility of post-treatment variation in wages between treated and synthetic control regions resultant of migration, and therefore the potential for simultaneity in periods after the initial treatment remains. However given the donor control regions are chosen because they are flat in terms of pre- and post-treatment A8 shares, this does not detract from estimation of the regional treatment effect of the policy.

Spillover Effects

Regional labour markets are not closed economies: trade occurs between those economies, and labourers can move between them. If A8 migration is found to have a negative impact on regional wages, natives or existing migrants might move to other regions thus biasing any estimator of the regional effect of immigration (Dustmann, Fabbri, Preston & Wadsworth, 2003).

A recent study for the U.K. (Home Office & BIS, 2014) finds very little evidence of this, which is consistent with evidence from elsewhere in Western Europe (e.g. Martins, Piracha, & Varejão (2012)). Andrews, Clark, & Whittaker (2008) note that interregional migration is particularly modest in the U.K. comparative to other developed economies, but is most prevalent amongst high-skill males owing to the fact that relocating is costly, and the cost of moving is less likely to be offset by the gains in employment outcomes in the low paid. Given the majority of A8 migrants compete in the low skill sector (Lemos & Portes, 2008), this is unlikely to be problematic for this study.

It should be noted, however, that workers do not necessarily have to relocate to take up a job in another region: they can choose to commute. This could be a source of spillovers for the same reasons described above. For this reason I define local labour markets in terms of areas within which there is a high propensity of commuting; this process is described below.

2.4.2 Defining Local Labour Markets

Ideally the regional data aggregation should represent local labour markets, as defined by the geographical areas within which individual workers commute to, and search for, work (Lemos & Portes, 2008). The Labour Force Survey fractionates data geographically into 409 unitary authorities (local government constituencies), 49 counties and 12 regions. In practice unitary authorities are unlikely to represent local labour markets, particularly in the case of large cities which are broken up into numerous unitary authorities but within which workers routinely commute. Lemos & Portes (2008) propose that aggregating at the county level provides a more realistic representation of local labour markets, assuming that workers are willing to commute to other unitary authorities within their county. However this fails to account for the large number of workers who live near county boundaries, and the increasing trend towards long distance commuting (Nielsen & Hovgesen, 2008). There are also a significant number of 'commuter towns' with high speed rail links to major cities such as London, in which a significant proportion of residents work in a city which is geographically distant, but very few work in unitary authorities in between that town and the city in question; further, very few workers who live those in-between unitary authorities work in that city.

I therefore construct stylised local labour markets (LLMs) by pooling unitary authorities according to the conditional probability of working in a particular unitary authority whilst living in another. The mathematical argument is as follows.

Let $\mathcal{R} = \{r_1, r_2, \dots, r_N\}$ be the set of all unitary authorities r, and $R \subseteq \mathcal{R}$ be the set of all local labour markets, so $R = \{R_1, R_2, \dots, R_N\}$. Each local labour market R_j consists of $K \geq 1$ unitary authorities r_i, such that $R_j = \{r_1, r_2, \dots, r_K\}$.

Now let w_{r_i} be a binary indicator for an individual working in unitary authority r_i. Let $l_{r_{-i}}$ be a binary indicator for an individual living in some other unitary authority $r_{-i} \neq r_i$.

$$R_j \coloneqq \left\{r_i : r_i \in \mathcal{R} \ and \ \Pr\left(w_{r_i} \middle| l_{r_{-i}}\right) \geq x \ for \ some \ r_i \in R_j\right\}$$
$$x = 7.5\%$$[3]

x is chosen such that it is small enough to generate pools with a reasonable number of unitary authorities, but large enough for the pooled unitary authorities to meaningfully represent local labour markets; 409 unitary authorities are pooled into 148 stylised local labour markets. In practice not all unitary authorities are pooled with others where there is very little interregional commuting, but these regions are primarily remote, rural locations, which were largely unaffected by A8 migration.

[3] Clearly the size of LLMs generated will depend on the choice of x. Ideally, results for a range of values of x would be presented, however this process is extremely computationally heavy. I propose that regardless of the chosen value of x, this represents a better method for pooling unitary authorities than on geography alone.

2.4.3 Treatment and Control Local Labour Market Identification

Treated LLMs, and 'donor' control LLMs which are used to construct the synthetic LLM, are identified from the set of local labour markets described above. Treated LLMs are identified by testing for a structural break in the LLM-specific simple linear regression of A8 stocks on time (measured in quarters) at the point of the policy, in a sample restricted to observations in the labour force (those aged between 16 and 65 and are either in employment or unemployed and active seeking work, as per the ILO definition). A Wald test for a structural break at the point of A8 accession is then performed using Stata's 'estat' command, the null hypothesis being no structural break at the point of the policy change. Regions with a p-value of greater than 0.1 are identified as potential 'donor' controls; these regions are then analysed graphically to ensure the trend in A8 stocks is relatively flat across time. Similarly, regions with a p-value close to zero are identified as potential treatment regions, and are analysed graphically to ensure trends in A8 stocks are relatively flat prior to Q2 2004, and that the structural break identified represents a notable increase in A8 stocks after Q2 2004.[4]

From the sets of LLMs identified above, LLMs with a significant post-accession A8 influx are chosen as 'treated' regions for analysis. In practice most LLMs were significantly impacted by the policy, so four LLMs with a large sample size and relatively flat pre-accession A8 stocks are selected for analysis. In addition, the sample size (when restricted to those in the labour force) needs to be sufficiently large to generate meaningful results where the sample is further restricted to age and skill groups; as such LLMs with sample size of less than 4,000 were rejected. Four suitable 'treated' LLMs are identified, and these are labelled A-D in Appendix B. Similarly, ten donor control LLMs are identified with flat pre- and post-accession A8 stocks, subject to the same sample size restrictions, and these are labelled 1-10 in the Appendix B.

[4] This approach lends itself to the comparison of the most treated to the least treated LLMs, which is the goal of this study.

2.4.4 Synthetic Control

The Synthetic Control Method (SCM) was first proposed by Abadie & Gardeazabal (2003). The authors investigate the effects of conflict in the Basque Country on GDP per capita. Problematically the region differed substantially from other Spanish regions in terms of the theoretical determinants of growth, such that there did not exist a valid counterfactual. The authors proposed a 'synthetic' control region, constructed using a weighted combination of other regions, which would resemble the Basque country in terms of pre-treatment characteristics. A negative impact on GDP per capita was identified, of a magnitude in keeping with the literature.

SCM was first employed in the context of assessing the impacts of immigration on local labour markets by Borjas (2017), where a synthetic control region is constructed using March-CPS data to reappraise the impact of the Mariel Boatlift, as first analysed by Card (1990). Contrary to Card's findings, a significant negative wage impact of between 10 and 30 per cent is identified for Miami high school dropouts, the group with whom the relatively unskilled Mariels were assumed to compete in the labour market. Borjas does, however, emphasise that this result is sensitive to the set of predictor variables chosen to construct the synthetic counterfactual. For this reason, as a robustness check in addition to the main results I present results of specifications with a variety of predictor variables, as discussed in more detail below.

Peri & Yasenov (2015) replicate the estimations of Borjas (2017) using ORG-CPS data, which contains a larger sample size and measures wages more precisely. The authors argue that the measurement error within the March-CPS data as used by Borjas is significant enough to account for the wage effects estimated, and find no significant difference between wage outcomes for unskilled Miamians and the synthetic counterfactual. Borjas (2016) contests that the differing results found by Peri & Yasenov (2015) are driven by a synthetic control group derived from a different set of control regions, and the inclusion of hispanics in the sample. Further, Peri & Yasenov include workers aged 16-61, whereas Borjas restricts the sample to those aged 25-59 to mitigate the possibility of those still in education being misclassified as high school dropouts.

The construction of the synthetic control regions in this paper closely follows Abadie, Diamond & Hainmueller (2010), and makes use of the 'Synth' extension package for Stata as written by the authors. An excellent overview of the mechanics of the method is provided in Abadie (2020), which is set out below in the context of the application in this paper.

There exist $J+1$ LLMs, which are denoted $j = 1,2,\ldots,J+1$, where the 'treated' LLM is denoted $j = 1$ and all $j > 1$ are donor control LLMs. The data span T periods, and periods denoted T_0 occur before A8 accession. For each LLM, j, and time, t, we observe log hourly wages $\log W_{jt}$, henceforth denoted Y_{jt} for ease of notation. For each LLM j, we also observe k predictors of the outcome, $X_{1j}, X_{2j}, \ldots, X_{kj}$ which are independent of A8 accession. The $k \times 1$ vectors $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{J+1}$ contain the values of predictors $X_{1j}, X_{2j}, \ldots, X_{kj}$, and the $k \times j$ matrix $\mathbf{X}_0 = [\mathbf{X}_2 \ldots \mathbf{X}_{j+1}]$ therefore collects the values of the predictors for the J untreated LLMs. Y_{jt}^N is defined as the expected post-treatment outcome in 'treated' LLMs in the absence of treatment, and Y_{jt}^I the observed post-treatment outcomes in treated LLMs. The effect of the A8 accession on treated LLM outcomes is therefore:

$$\tau_{1t} = Y_{jt}^I - Y_{jt}^N$$

Clearly Y_{jt}^N is unobserved. The SCM method estimates $\hat{Y}_{jt}^N$ using one or more untreated LLM that had similar characteristics to the treated LLM in the observed pre-treatment period. The synthetic control group is defined as a weighted average of the LLMs in the donor pool. A synthetic control can be represented by a $J \times 1$ vector of weights, $\mathbf{W} = (w_2, w_3, \dots, w_{J+1})'$, such that:

$$\hat{Y}_{jt}^{N} = \sum_{j=2}^{J+1} w_j Y_{jt}$$

and:

$$\tau_{1t} = Y_{1t} - \hat{Y}_{jt}^{N}$$

$\sum_{j=2}^{J+1} w_j = 1$, and $w_j \geq 0$, so synthetic control groups are weighted averages of donor LLMs. Given a set of non-negative constants $v_1, v_2, \dots, v_k$ (the derivation of which is discussed later in this section), optimal weights $\mathbf{W}^* = (w_2^*, w_3^*, \dots, w_{J+1}^*)'$ are chosen to minimise:

$$\|\mathbf{X}_1 - \mathbf{X}_0\mathbf{W}\| = \sqrt{\sum_{h=1}^{k} v_h \left(X_{h1} - w_2 X_{h2} - w_3 X_{h3} - \dots - w_{J+!} X_{hJ+1}\right)^2}$$

subject to $\sum_{j=2}^{J+1} w_j = 1$, and $w_j \geq 0$. The estimated treatment effect is therefore:

$$\hat{\tau}_{1t} = Y_{1t} - \sum_{j=2}^{J+1} w_j^* Y_{jt}$$

$\mathbf{V} = (v_2, v_3, \dots, v_k)$ is chosen such that the synthetic control $\mathbf{W}(\mathbf{V})$ minimises the mean squared prediction error of the synthetic control with respect to Y_{1t}^{N}:

$$\sum_{t \in \tau_0} \left[Y_{1t} - w_2(\mathbf{V})Y_{2t} - \dots - Y_{1t} - w_3(\mathbf{V})Y_{3t} - Y_{1t} - w_{J+1}(\mathbf{V})Y_{J+1t}\right]^2$$

for some $\tau_o \subseteq \{1,2,\ldots,T_0\}$.

2.4.5 Specifications

To benchmark the synthetic control estimates, I first estimate the impact of A8 migration on wages using spatial correlations, in line with the existing literature, by way of benchmarking the synthetic control specifications described below. Data are first-differenced to allow for the estimation of the impact of quarterly changes in local labour market A8 stocks. $\Delta\log(W_{Rt}) = \phi\omega_R\Delta A8_{Rt} + [\mathbf{Q}_{Rt}\omega_R]\psi + \varepsilon_{Rt}$ is estimated using weighted least squares, where $\Delta\log(W_{Rt})$ represents the change in mean hourly log wages in LLM R between quarter $t-1$ and quarter t, $\Delta A8_{Rt}$ is the change in LLM A8 stocks over the same time period, $\mathbf{Q}_{Rt}$ is a vector of first-differenced covariates, and ω_R is an analytical weight for each LLM equal to the reciprocal of that respective LLM's population (to account for the fact that absolute changes in A8 stocks are relative to differing LLM populations). ϕ is thus an estimate of the marginal impact of A8 inflows on log wages, subject to the assumption that the conditional independence assumption holds (as is discussed above). Analysis is restricted to the treated and donor control LLMs analysed in the synthetic control specifications so as to ensure the results in both relate to comparable samples. Results are reported in Table 2.5.

I then estimate a simple difference in difference estimator between the pooled 'treated' regions and the unweighted pool of donor control regions, controlling for covariates. $\log(W_{Rt}) = \hat{\beta}_0 + \hat{\beta}_1 t + \hat{\beta}_2 R^T + \hat{\beta}_3 R^T t + \mathbf{X}\beta + \hat{\varepsilon}_{Rt}$ is estimated by ordinary least squares, where log wages are on the left hand side, t is time specified continuously measured in quarters, R^T is a binary indicator variable equal to one if the local labour market is treated and the observation occurs in the post-treatment period and $\mathbf{X}$ is a $(N \times k)$ matrix of N observations and k covariates. $R^T t$ is therefore an interaction term where $\hat{\beta}_3$ represents the difference in difference estimator of trend hourly wage growth between treated and pooled donor control regions in the post-treatment period. Results are reported in Table 2.6.

The synthetic control specification uses an 'employment placebo' as per Borjas (2017); I additionally control for various explanatories on wage as this is found to reduce root mean squared prediction error in the nested optimisation process, and therefore better match treated regions to synthetic donors on pre-treatment trends. The $\mathbf{X}$ matrix therefore contains continuous measures of unemployment rates, age and years of education (and their quadratics), and binary indicator variables for gender, race, marital status, disability status, and public sector employment; in the aggregated data these binary indicators summate to represent population shares. These variables were chosen by regressing log wages on various specifications, and choosing the specification with the highest adjusted R-Squared.

For robustness I include two additional synthetic control specifications (results are presented in the appendix). Firstly, I include only the 'employment placebo', dropping all other predictor variables; in this specification the $\mathbf{Z}_{Rt}$ matrix controls only for local labour market pre-treatment employment rates. Treated local labour markets are therefore matched to donor controls on general macro-labour market conditions using a wage curve relationship (Blanchflower & Oswald, 1994).

As a second robustness check, employment rates are removed from the specification, as there is clearly potential for reverse causality due to the 'wage curve' relationship between employment and wages proposed by Blanchflower & Oswald (1994). Nonetheless it is important to control for employment in some way, as standard theory dictates that migrants should rationally select into local labour markets where their probability of employment is maximised. The pool of donor control LLMs for each treated region is therefore restricted to those with mean pre-treatment sample unemployment rates that are not statistically different from the respective treated LLM at the 95% level.

In each specification the impact on log hourly wages is estimated for the full population as well as the two groups with whom A8s primarily compete: under 30s and low-skilled workers. As discussed in the Data section, it is not possible to directly compare on educational attainment due to the way the LFS codes some foreign qualifications. I therefore restrict the low-skilled sample to those who left the education system below age 20. This is for two reasons: firstly because whilst A8s are more likely to select into low-skilled work, their education level does not appear to be different from existing migrants, so it would seem erroneous to impose that they compete with those who left school at 16; secondly, a cut-off closer to 16 would restrict the sample to a point where donor group sample sizes are too small to draw meaningful analysis.

Once the synthetic control regions are generated, deviations in post-treatment trends between treatment and control are identified using difference in differences. $\log(W_{Rt}) = \hat{\beta}_0 + \hat{\beta}_1 t + \hat{\beta}_2 R^T + \hat{\beta}_3 R^T t + \hat{\varepsilon}_{Rt}$ is estimated by OLS, where log wages are on the left hand side, t is time specified continuously measured in quarters, R^T is a binary indicator variable equal to one if the local labour market is treated and the observation occurs in the post-treatment period; $R^T t$ is therefore an interaction term where $\hat{\beta}_3$ represents the difference in difference estimator of trend hourly wage growth between treated and synthetic control regions in the post-treatment period.

2.5 Results

Spatial Correlation

	All Workers	Low Education	21-30s
Dep. Var.: $\Delta\log(W_{Rt})$			
ϕ	-0.0000125	-0.0000008	-0.0000369*
t-Stat	-1.57	-0.73	-1.65
Spatial Cells (*N*)	266	266	266
Adj R-sq	0.0640	0.0441	0.0610

*Significant at 10% level; **Significant at 5% level; ***Significant at 1% level

Table 2.5: results from spatial correlations specification (data aggregated to LLM level). Coefficient is interpretable as the log-point change in expected wages resultant of a one-unit regional change in A8 migrant stocks.

Difference in Differences

	All Workers	Low Education	21-30s
Dep. Var.: log (W_{Rt})			
$\hat{\beta}_3$	0.00194	0.00151	0.00000
t-Stat	0.00882	0.00820	0.01054
Spatial Cells (N)	42	42	42
Adj R-sq	0.9321	0.6363	0.9407

*Significant at 10% level; **Significant at 5% level; ***Significant at 1% level

Table 2.6: results from difference-in-differences specification (data aggregated to LLM level). Coefficient is interpretable as the log point change in expected wages in 'treated' regions compared to comparable regions that did not experience a significant increase in A8 stocks.

Consistent with the literature, the spatial correlation specification (see Table 2.5) estimates no statistically significant marginal impact of A8 migration on log hourly wages for the sample of all workers and those with lower education levels; a very small (0.000037 log point) negative marginal impact is identified for 21-30s, (just) significant (at 10% level).

Difference in difference results are presented in Table 2.6, and Figure 2.5. The unweighted difference in difference estimator similarly identifies no statistically significant impact for all workers, nor for 21-30s or the less educated, and the coefficients are very close to zero. It is important to note that the interpretation of the coefficients (including those in the synthetic control regressions) is slightly different to those of spatial correlation; the SCM regressions estimate the change in trend LLM wage growth post-accession comparative to the synthetic counterfactual, as opposed to the marginal unit impact. As such their polarity can be compared to the corresponding spatial correlation coefficients but not their magnitude.

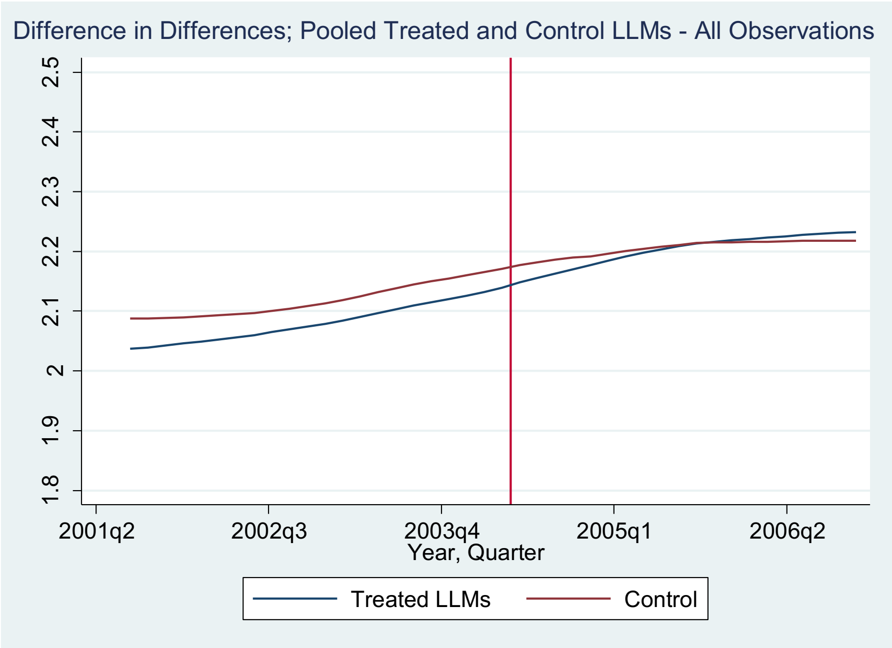

Difference in Differences; Pooled Treated and Control LLMs - All Observations
1.8
1.9
2
2.1
2.2
2.3
2.4
2.5
2001q2
2002q3
2003q4
2005q1
2006q2
Year, Quarter
Treated LLMs
Control

Figure 2.5

Synthetic Control – Employment Placebo with Controls

All Workers

Local Labour Market	A	B	C	D	Pooled
$\hat{\beta}_3$	0.00431***	0.00205	0.00361***	0.00112	.00194**
t-stat	3.20372	1.34509	3.26930	1.27382	2.19570
RMSPE	0.08179	0.04975	0.09680	0.05104	0.05991
Spatial Cells	42	42	42	42	42
Donor Labour Market Weights					
1	0	0	0	0.19	0.087
2	0	0	0	0	0.028
3	0.253	0	0.456	0.049	0.164
4	0.105	0.051	0	0	0
5	0.415	0	0.284	0.025	0.043
6	0.079	0.574	0.225	0.36	0.389
7	0.038	0.05	0	0	0
8	0.009	0.324	0	0.097	0.134
9	0.022	0	0.036	0.278	0.156
10	0.079	0	0	0	0

*Significant at 10% level; **Significant at 5% level; ***Significant at 1% level

Table 2.7: Synthetic control results for all workers, for treated LLMs a, b, c, and d, and all treated LLMs pooled. Coefficient interpretable as the log point change in expected wages in the respective 'treated' region(s) compared to that region's synthetic counterfactual. Treated LLMs A-D, and donor control LLMs 1-10, are defined in Appendix B.

Figure 2.6

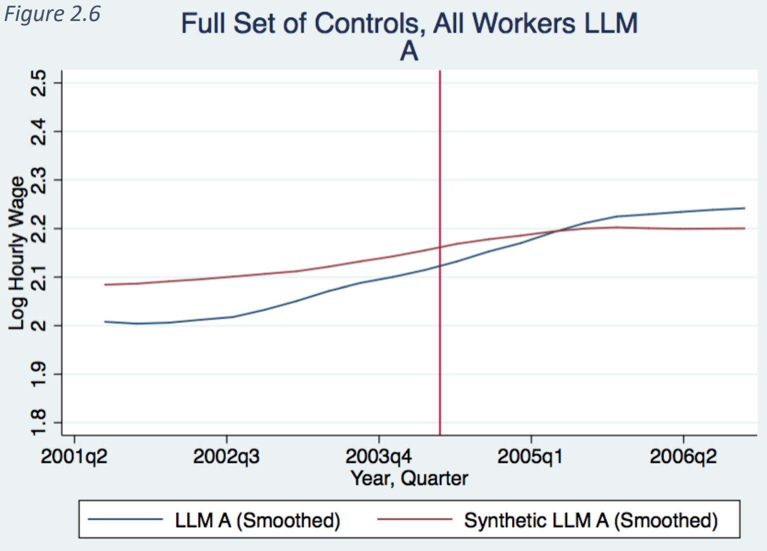


Figure 2.7

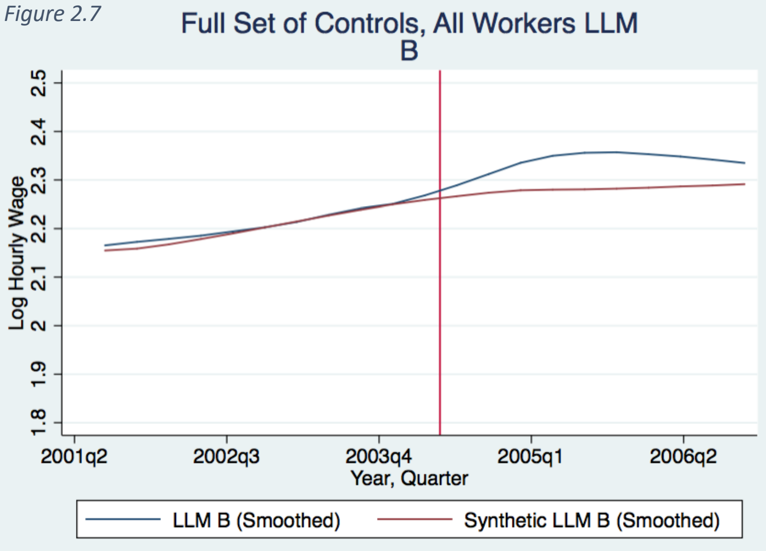


Figure 2.8

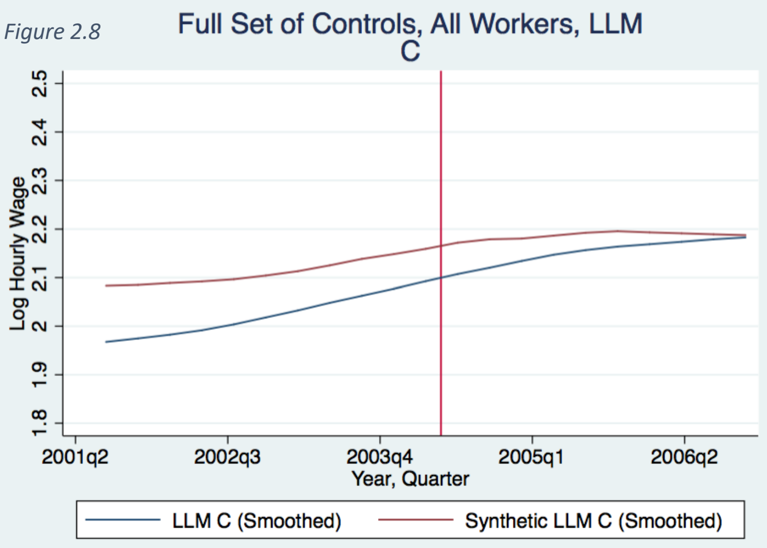
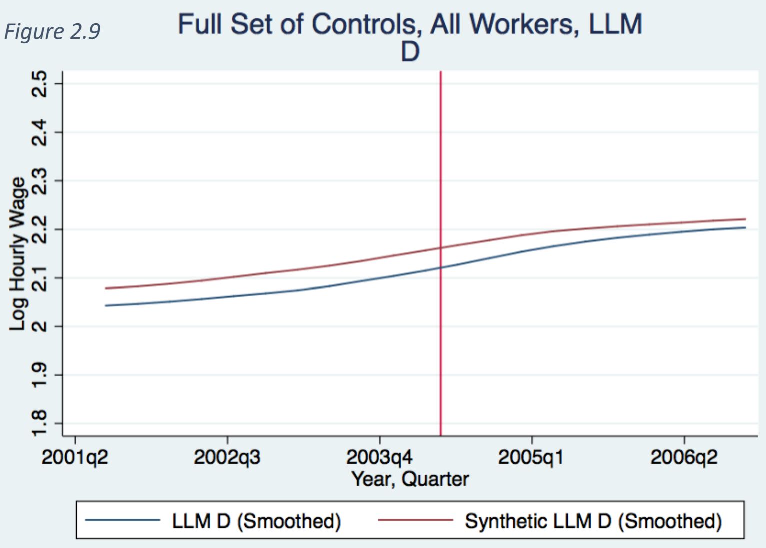


Figure 2.9



Low Education

Local Labour Market	a	b	c	d	Pooled
$\hat{\beta}_3$	0.00039	0.00113	0.00169	-0.00042	0.00049
t-stat	0.25680	0.71652	1.58704	-0.40306	0.63475
RMSPE	0.07468	0.06065	0.06133	0.05047	0.03675
Spatial Cells	42	42	42	42	42
Donor Labour Market Weights					
1	0.49	0	0.2	0.333	0.203
2	0	0	0.234	0	0.106
3	0	0	0.044	0	0
4	0	0.048	0	0	0
5	0	0	0	0	0
6	0	0.524	0	0.135	0.186
7	0	0.021	0	0	0
8	0.498	0.407	0.487	0.346	0.419
9	0.012	0	0.036	0.186	0.086
10	0	0	0	0	0

*Significant at 10% level; **Significant at 5% level; ***Significant at 1% level

Table 2.8: Synthetic control results for workers with low levels of education, for treated LLMs a, b, c, and d, and all treated LLMs pooled. Coefficient interpretable as the log point change in expected wages in the respective 'treated' region(s) compared to that region's synthetic counterfactual. Treated LLMs A-D, and donor control LLMs 1-10, are defined in Appendix B.

Figure 2.10

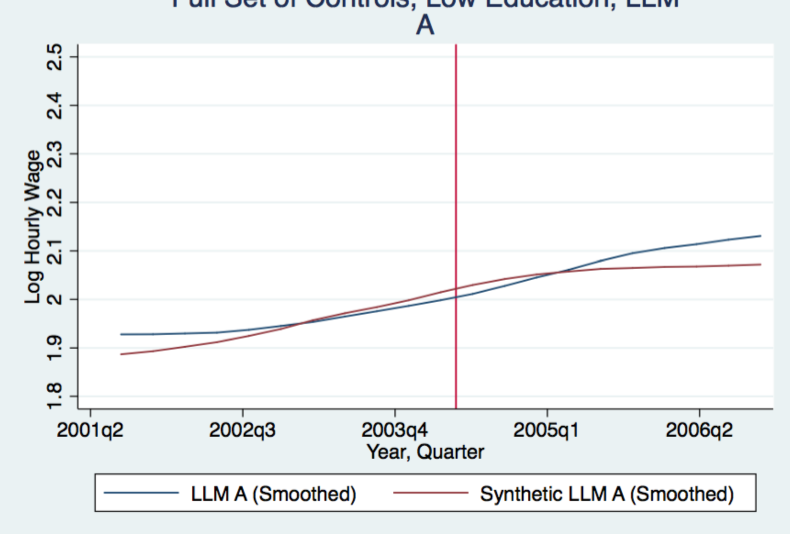


Figure 2.11

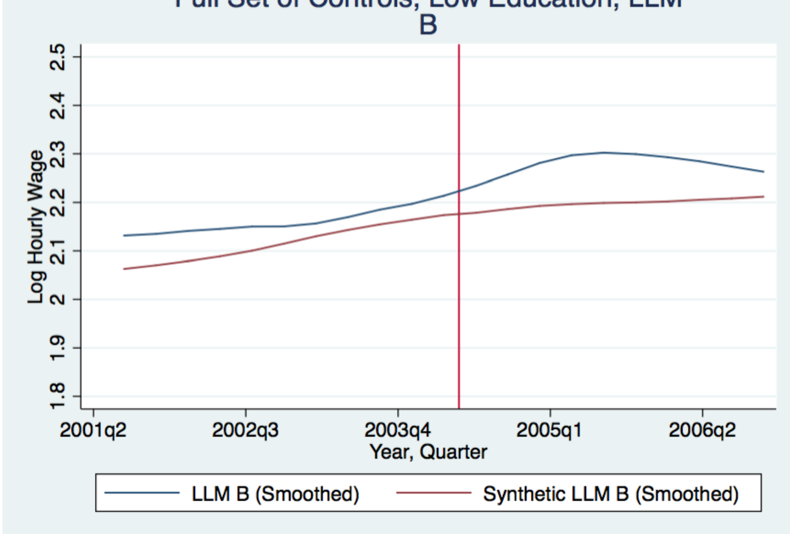


Figure 2.12

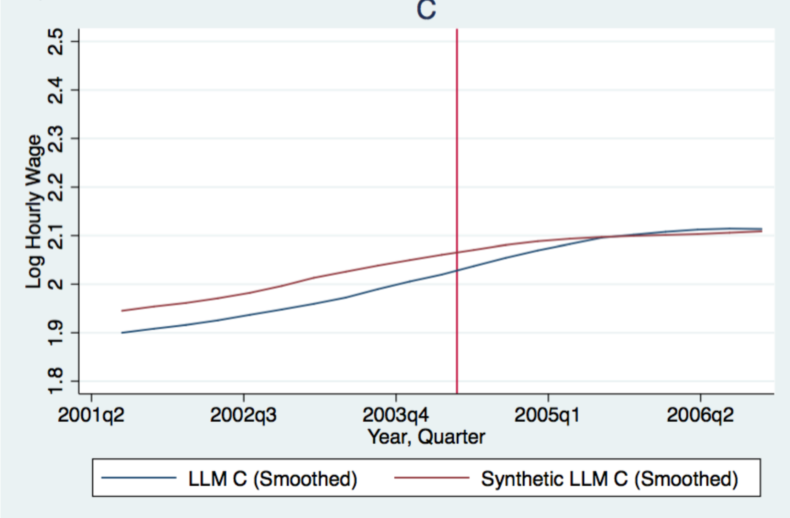
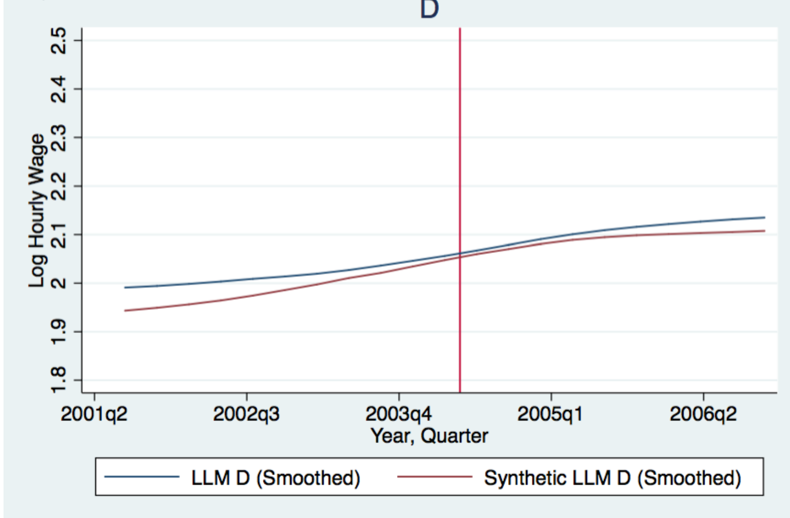


Figure 2.13



Under 30s

Local Labour Market	a	b	c	d	Pooled
$\hat{\beta}_3$	0.001183	-0.00349	-0.00029	-0.0003	-0.00114
t-stat	0.465353	-1.04993	-0.15066	-0.18393	-0.68738
RMSPE	0.12680	0.12022	0.15244	0.08592	0.09988
Spatial Cells	42	42	42	42	42
Donor Labour Market Weights					
1	0	0	0	0	0
2	0	0	0	0	0
3	0.456	0	0.342	0.186	0.234
4	0.285	0.136	0.051	0.13	0.143
5	0.108	0	0.217	0.068	0.076
6	0	0.795	0.39	0.273	0.365
7	0.061	0	0	0	0
8	0.09	0	0	0	0
9	0	0.069	0	0.343	0.182
10	0	0	0	0	0

*Significant at 10% level; **Significant at 5% level; ***Significant at 1% level

Table 2.9: *Synthetic control results for under-30s, for treated LLMs a, b, c, and d, and all treated LLMs pooled. Coefficient interpretable as the log point change in expected wages in the respective 'treated' region(s) compared to that region's synthetic counterfactual. Treated LLMs A-D, and donor control LLMs 1-10, are defined in Appendix B.*

Figure 2.14

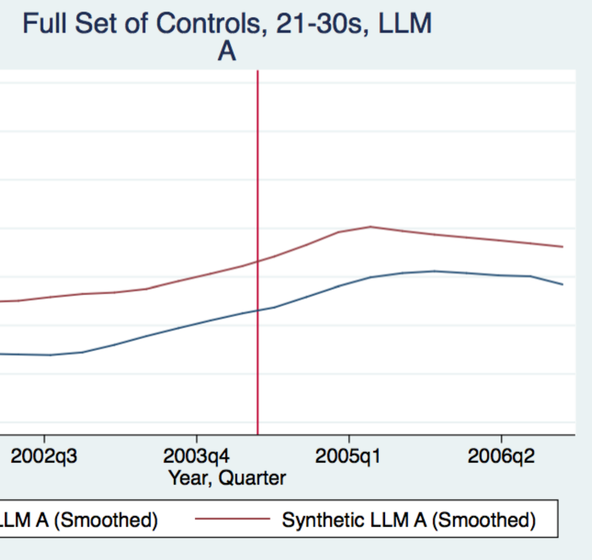


Figure 2.15

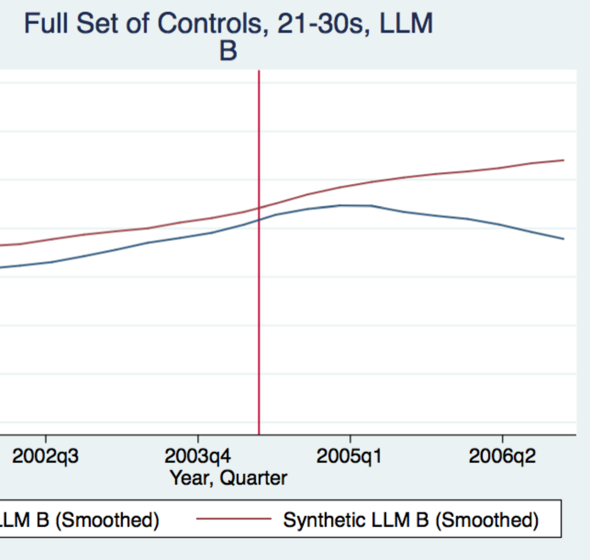


Figure 2.16

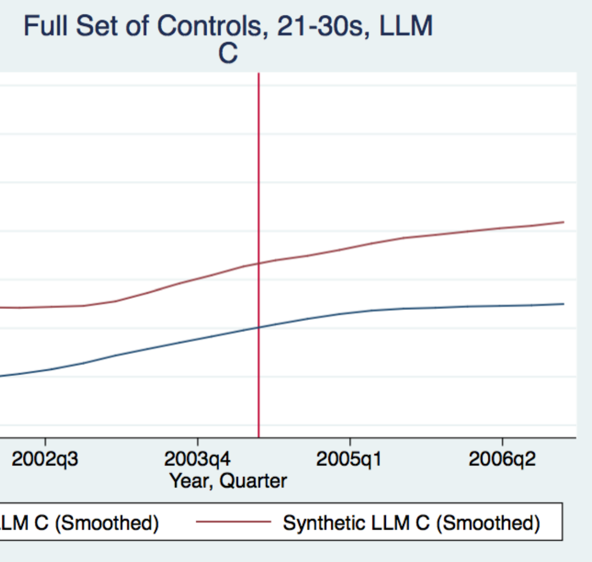
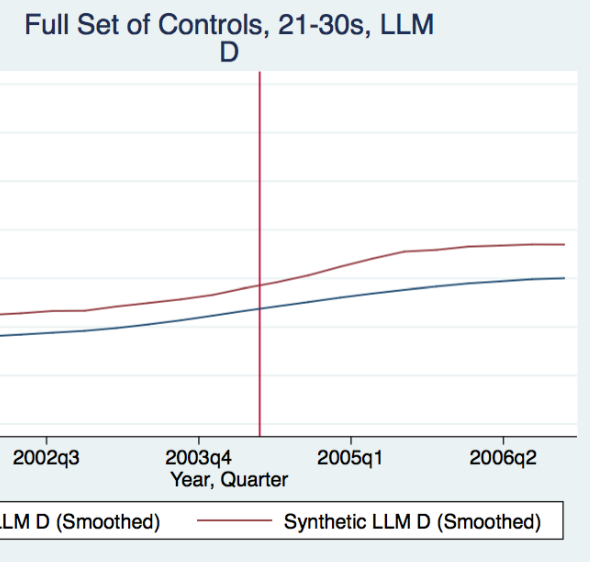


Figure 2.17



Synthetic Control results for employment placebo with controls are presented for All Workers in Table 2.7, and Figures 2.6-2.9. I find small positive wage growth impacts of between 0.36 and 0.43 percentage points, significant to the 1% level in two of the treated LLMs, and a positive growth impact of 0.19 percentage points at the 5% level in the pooled sample. Two treated LLMs, ‘b’ and ‘d’ yield positive growth estimates of 0.21 percentage points and 0.11 percentage points respectively, but are not statistically significant. Estimators in the low education sample (see Table 2.8; Figures 2.8-2.13) remain small and (mostly) positive, between 0.04 and 0.16 percentage points though none are found to be significant. No significant impact was identified for 21-30s (Table 2.9; Figures 2.14-2.17), though coefficients are negative in LLMs ‘b’, ‘c’ and ‘d’ and the pooled sample.

Results for the employment placebo without controls are presented in the Appendix A, Small positive but statistically significant wage impact for all workers (Table 2.11; Figures 2.18-2.21) in three out of the four ‘treated’ LLMs and the pooled sample, of between 0.17 and 0.31 percentage points. The employment placebo also identified a small but significant positive impact in two of the four treated LLMs in the low education sample and the pooled low education sample (Table 2.12; Figures 2.22-2.25), of between 0.14 and 0.25 percentage points. No statistically significant impact was identified for 21-30s (Table 2.13; Figures 2.26-2.29), though similarly to the spatial estimates the coefficients were small and negative in all but one LLMs and the pooled sample.

Results for the final SCM setup with donor LLMs matched to treated LLMs manually are presented in Appendix A. For the sample of all workers (Table 2.14; Figures 2.30-2.33), small and positive estimators of between 0.05 and 0.35 percentage points are identified, and are significant in two out of the four treated LLMs (at 1% and 5% levels respectively). The estimators on the low education sample (Table 2.15; Figures 2.34-2.37) remain positive, but are not significant; the estimated effect on 21-30s (Table 2.16; Figures 2.38-2.41) is close to zero in all four LLMs and none are significant.

The SCM estimates across all specifications are consistent with the literature, in that for the most part the null hypothesis of zero effect cannot be rejected, and where significant effects are identified they are very close to zero. Analogously to the spatial correlation estimates, the worst performing of the subpopulations studied in terms of post-accession wages is 21-30s, although the negative coefficients are not found to be significant. For the less educated subsample small positive coefficients are identified, comparative to no effect using spatial correlation, though these estimators switch from significance at the 1% or 5% level in the employment placebo to being insignificant when covariates are added in the matching process; this warrants some discussion.

A clear explanation is that where treated LLMs are matched to donor controls on LLM employment rates, e_R, alone, $\mathbf{V}$ simply becomes an identity matrix, and so minimising root mean square predictor error $||\mathbf{X}_1 - \mathbf{X}_o\mathbf{W}||_\mathbf{V}$ simply becomes $||\mathbf{e}_1 - \mathbf{e}_o\mathbf{W}||_{\mathbf{I}_1}$. In other words, instead of weighting donor control regions on similarity in terms of explanatory variables' relative explanatory power, the process weights employment rates as if they explain wages in their entirety. Adding (relevant) covariates to the $\mathbf{X}_0$ and $\mathbf{X}_1$ matrices will therefore inevitably reduce weights in the $\mathbf{W}$ matrix assigned to donor LLMs which are similar to the treated LLM in terms of employment rates, but dissimilar to the treated region in terms of other relevant explanatory characteristics. This can be directly observed in the results tables and graphs: for the low education subsample RMSPE is uniformly lower in the employment placebo with controls comparative to the employment placebo, and the graphs show a notably better pre-treatment fit.

This is not substantively dissimilar to the critique of Propensity Score Matching (PSM) found by Smith & Todd's (2005) replication of Dehejia & Wahba (1999). PSM is the statistical technique from which SCM was developed (Abadie & Gardeazabal, 2003), in which pairs of treated and untreated observations are matched and weighted in terms of pre-treatment characteristics. Dehejia & Wahba applied various non-experimental econometric techniques (including PSM) to survey data, comparing their results to those from the fully randomised NSW

Demonstration, concluding that bias in PSM estimates is negligible. Smith & Todd (2005) find that PSM is extremely sensitive to the set of control variables on which the propensity score is defined, and that results inconsistent with those from the experimental data can be obtained by PSM using a different (but equally reasonable) set of variables in the propensity scores.

However the results for the all-worker sample remain consistently positive, small and significant in all SCM specifications. This is supportive of the results of Manacorda, Manning, & Wadsworth (2012), who find evidence to suggest migrants and natives are gross compliments in the labour market, with a small positive wage effect of migration for natives. In Table 2.10, the employment placebo (without controls)[5] specification is repeated with a sample restricted to white British males.

Local Labour Market	A	B	C	D	Pooled
Whole Sample					
$\hat{\beta}_3$	0.00313***	0.00190	0.00204**	0.00169**	0.00180**
t-stat	2.67	1.60	2.51	2.10	2.54
RMSPE	0.09417	0.08044	0.11653	0.07915	0.05632
Spatial Cells	42	42	42	42	42
White British Males					
$\hat{\beta}_3$	0.00388***	0.00531***	0.00257**	0.00085	0.00216**
t-stat	2.82	3.60	2.14	0.80	2.15
RMSPE	0.077101	0.07528	0.15403	0.08937	0.13283
Spatial Cells	42	42	42	42	42

*Significant at 10% level; **Significant at 5% level; ***Significant at 1% level

Table 2.10: Synthetic control results for white British males, for treated LLMs a, b, c, and d, and all treated LLMs pooled. Coefficient interpretable as the log point change in expected wages in the respective 'treated' region(s) compared to that region's synthetic counterfactual. Treated LLMs A-D, and donor control LLMs 1-10, are defined in Appendix B.

[5] The specification without controls is chosen for comparison purposes as race and gender dummies are used as control variables.

With the exception of LLM ‘d’, for which the coefficient is very small and not significant, a more positive and significant wage effect is identified for white British males than the sample as a whole. Heterogeneity in treatment effects according to demography could therefore be driving the positive results in the all-worker sample, given no significant wage effect was identified for the groups A8s are thought to primarily compete with in the labour force. It should, however, be stressed that this is only an indicative result: ideally the process would be repeated for non-natives, but the small sample size is prohibitive.

2.6 Conclusion

In keeping with the findings in the spatial study by Lemos & Portes (2008), I find no evidence of a short run effect of A8 migration on hourly earnings for the groups A8s are believed to primarily compete with in the labour force. Further, I find a very small but statistically significant positive wage impact for the overall labour force in affected local labour markets, which is particularly strong for white British males; this finding is supportive of existing evidence from Manacorda, Manning, & Wadsworth (2012), which suggests migrants and natives are gross compliments in the labour force.

Three specifications have been used to construct synthetic placebo labour markets: firstly using an ‘employment placebo’ as per Borjas (2017), then adding control variables to the employment specification, and finally controlling only for these variables whilst restricting the sample to donor local labour markets with similar pre-accession unemployment rates. The makeup of the synthetic placebo, and the statistical significance of results, seems to be strongly dependent on the control variables specified by the researcher; this finding is similar to existing critiques of related Propensity Score Matching techniques.

In the context of a migration shock as nationally widespread as the A8s, synthetic control has obvious drawbacks in capturing the overall wage effect of migration in that it compares regions with a clear and significant influx to those with none; as such, it might be expected to deliver slightly larger treatment effect estimates than are externally valid to the country overall. Spatial correlation remains a suitable tool to assess the effects of such geographically diffused immigration, however the employment of synthetic control as a robustness check to spatial estimates seems a sensible precaution in view of the complications presented by simultaneity.

2.7 Appendix A

Synthetic Control – Employment Placebo

All Workers

Local Labour Market	A	B	C	D	Pooled
$\hat{\beta}_3$	0.00313***	0.00190	0.00204**	0.00169**	0.00180**
t-stat	2.66638	1.59512	2.50614	2.09838	2.53696
RMSPE	0.09417	0.08044	0.11653	0.07915	0.05632
Spatial Cells	42	42	42	42	42
Donor Labour Market Weights					
1	0.127	0.097	0.127	0.091	0.105
2	0.208	0.097	0.22	0.09	0.106
3	0.126	0.097	0.125	0.091	0.105
4	0.101	0.098	0.099	0.093	0.103
5	0.107	0.097	0.106	0.092	0.104
6	0.081	0.099	0.08	0.098	0.1
7	0.034	0.112	0.032	0.137	0.086
8	0.085	0.099	0.083	0.096	0.1
9	0.054	0.104	0.052	0.111	0.093
10	0.076	0.1	0.074	0.099	0.098

*Significant at 10% level; **Significant at 5% level; ***Significant at 1% level

Table 2.11: Synthetic Control (Employment Placebo) results for all workers, for treated LLMs a, b, c, and d, and all treated LLMs pooled. Coefficient interpretable as the log point change in expected wages in the respective 'treated' region(s) compared to that region's synthetic counterfactual. Treated LLMs A-D, and donor control LLMs 1-10, are defined in Appendix B.

Figure 2.18

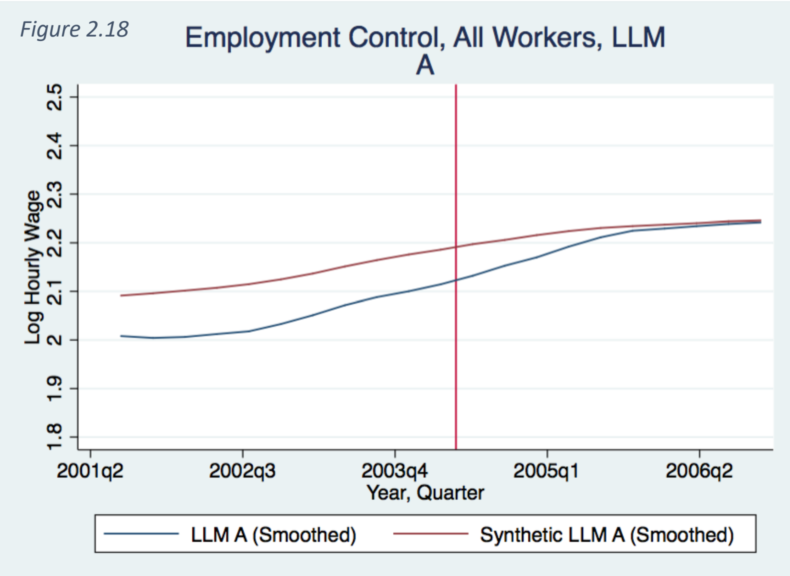


Figure 2.19

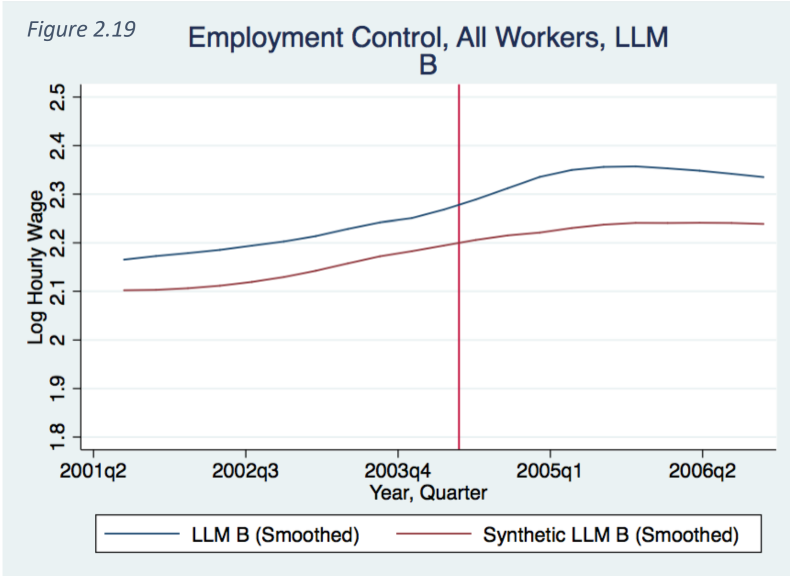


Figure 2.20

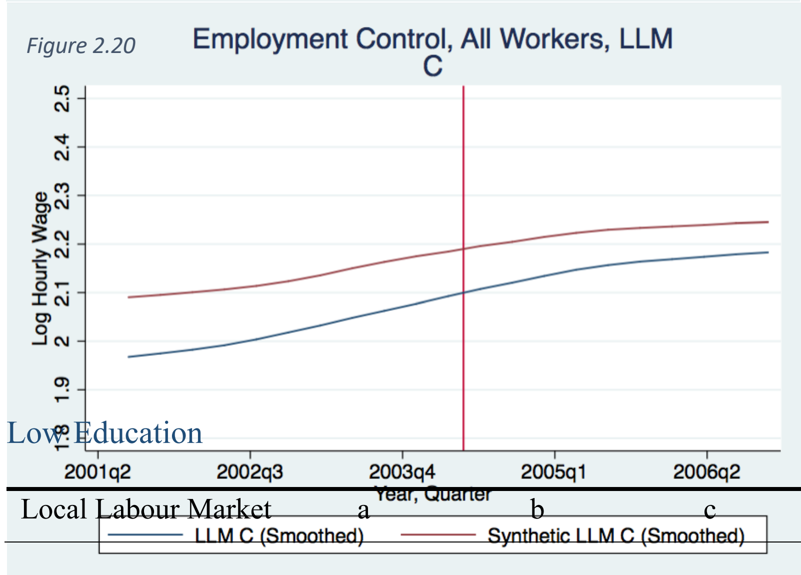
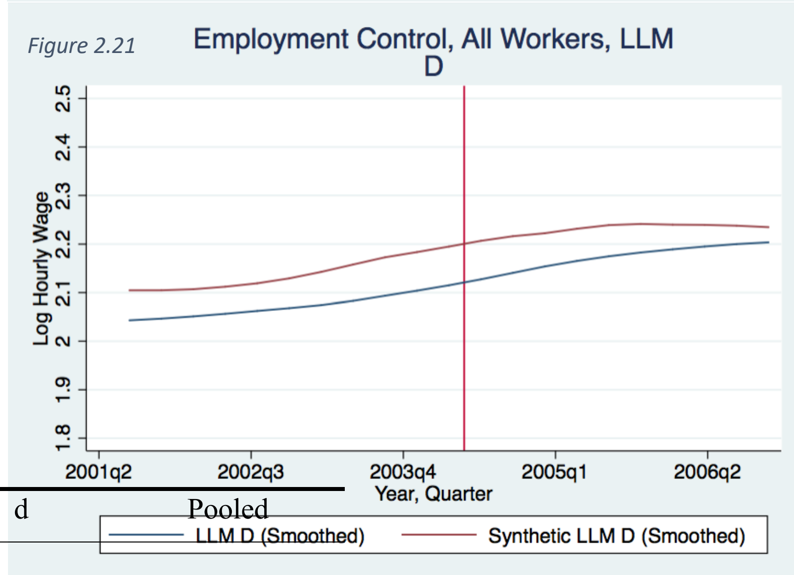


Figure 2.21



Low Education

Local Labour Market	a	b	c	d	Pooled

$\hat{\beta}_3$	0.00232**	0.00139	0.00249***	0.00101	0.00138**
t-stat	2.37624	1.17985	3.68509	1.32999	2.03426
RMSPE	0.11658	0.09752	0.10836	0.05851	0.04862
Spatial Cells	42	42	42	42	42
Donor Labour Market Weights					
1	0.119	0.107	0.132	0.098	0.112
2	0.111	0.105	0.108	0.098	0.108
3	0.136	0.11	0.251	0.098	0.119
4	0.104	0.103	0.094	0.099	0.104
5	0.1	0.101	0.087	0.099	0.101
6	0.1	0.101	0.086	0.099	0.101
7	0.059	0.081	0.027	0.108	0.071
8	0.1	0.101	0.087	0.099	0.101
9	0.081	0.092	0.056	0.102	0.087
10	0.091	0.097	0.071	0.1	0.095

*Significant at 10% level; **Significant at 5% level; ***Significant at 1% level

Figure 2.22

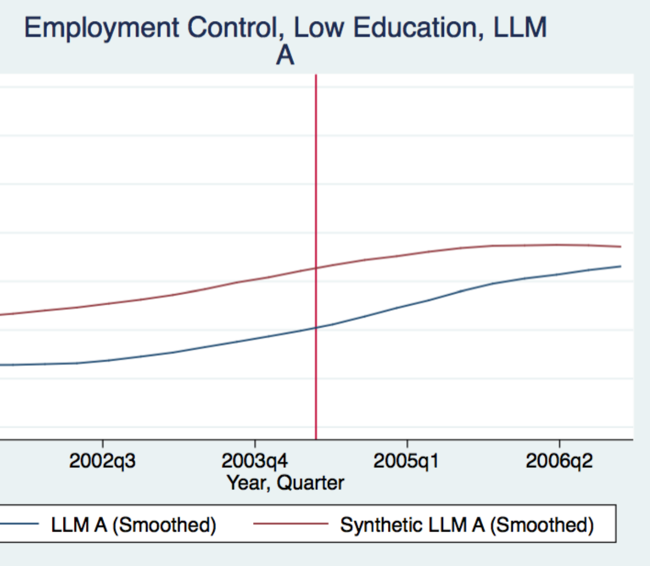


Figure 2.23

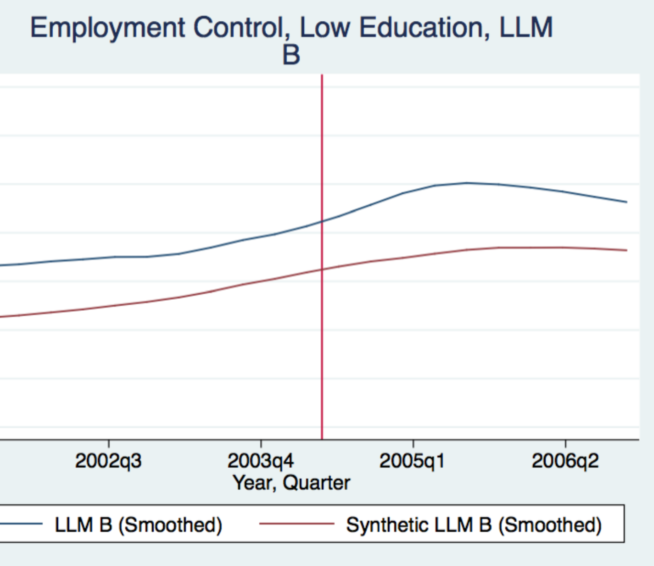


Figure 2.24

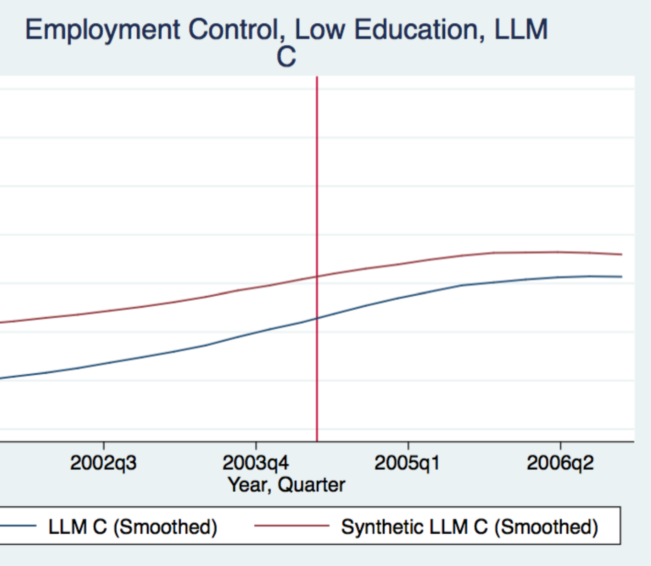
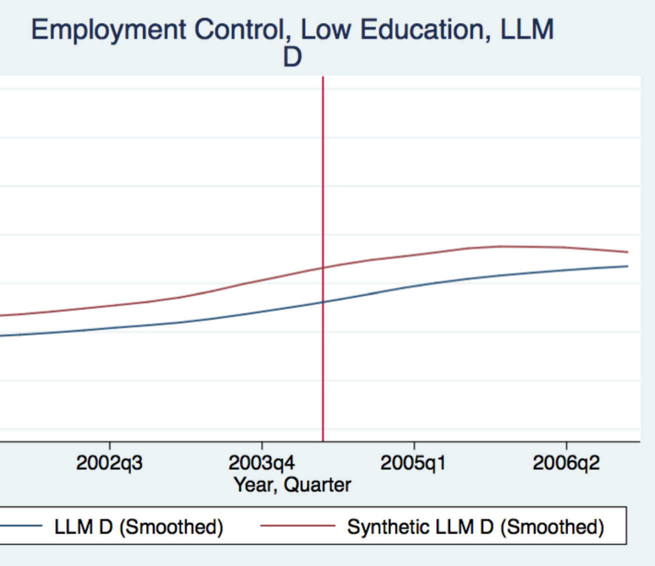


Figure 2.25



Under 30s

Local Labour Market	a	b	c	d	Pooled
$\hat{\beta}_3$	-0.00088	-0.00181	-0.00017	0.00081	-0.00056
t-stat	-0.33260	-0.81511	-0.11155	0.58729	-0.43331
RMSPE	0.18935	0.13761	0.15623	0.08689	0.07748
Spatial Cells	42	42	42	42	42
Donor Labour Market Weights					
1	0.08	0.099	0.11	0.089	0.105
2	0.116	0.098	0.135	0.085	0.111
3	0.081	0.099	0.11	0.089	0.105
4	0.492	0.098	0.212	0.082	0.116
5	0.056	0.1	0.09	0.096	0.099
6	0.05	0.1	0.084	0.099	0.098
7	0.011	0.104	0.043	0.141	0.084
8	0.045	0.1	0.079	0.101	0.096
9	0.022	0.103	0.056	0.117	0.089
10	0.048	0.1	0.082	0.1	0.097

*Significant at 10% level; **Significant at 5% level; ***Significant at 1% level

Table 2.13: Synthetic Control (Employment Placebo) results for under-30s, for treated LLMs a, b, c, and d, and all treated LLMs pooled. Coefficient interpretable as the log point change in expected wages in the respective 'treated' region(s) compared to that region's synthetic counterfactual. Treated LLMs A-D, and donor control LLMs 1-10, are defined in Appendix B.

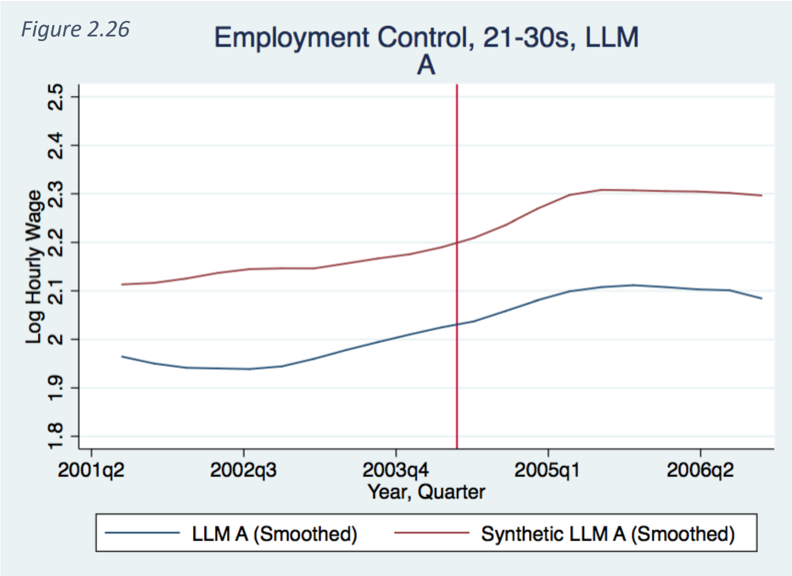


Figure 2.26
Employment Control, 21-30s, LLM A
Log Hourly Wage
1.8
1.9
2
2.1
2.2
2.3
2.4
2.5
2001q2
2002q3
2003q4
2005q1
2006q2
Year, Quarter
LLM A (Smoothed)
Synthetic LLM A (Smoothed)

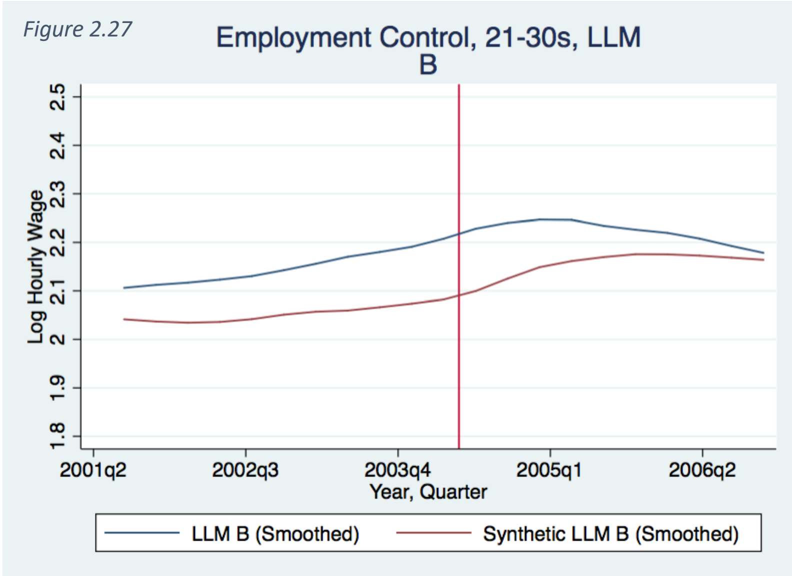


Figure 2.27
Employment Control, 21-30s, LLM B
Log Hourly Wage
1.8
1.9
2
2.1
2.2
2.3
2.4
2.5
2001q2
2002q3
2003q4
2005q1
2006q2
Year, Quarter
LLM B (Smoothed)
Synthetic LLM B (Smoothed)

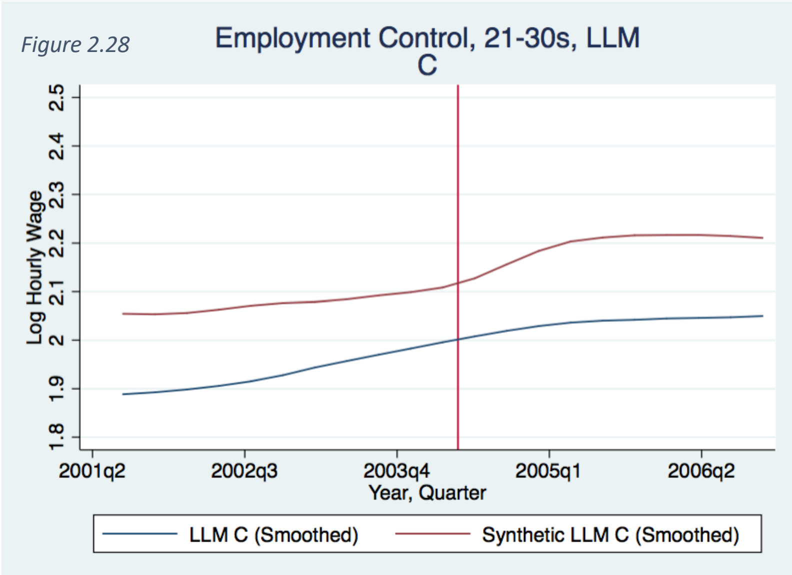


Figure 2.28
Employment Control, 21-30s, LLM C
Log Hourly Wage
1.8
1.9
2
2.1
2.2
2.3
2.4
2.5
2001q2
2002q3
2003q4
2005q1
2006q2
Year, Quarter
LLM C (Smoothed)
Synthetic LLM C (Smoothed)

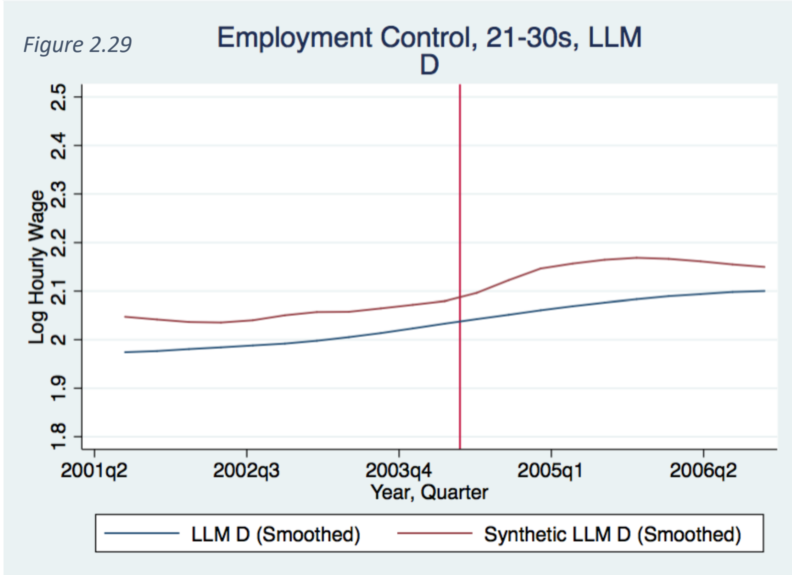


Figure 2.29
Employment Control, 21-30s, LLM D
Log Hourly Wage
1.8
1.9
2
2.1
2.2
2.3
2.4
2.5
2001q2
2002q3
2003q4
2005q1
2006q2
Year, Quarter
LLM D (Smoothed)
Synthetic LLM D (Smoothed)

Synthetic Control – Matching on Unemployment Rates Manually

All Workers

Local Labour Market	a	b	c	d
$\hat{\beta}_3$	0.00342***	0.00173	0.00277**	0.00048
t-stat	2.99566	1.20875	2.21652	0.54225
RMSPE	0.0444	0.04809	0.02640	0.03551
Spatial Cells	42	42	42	42
Donor Labour Market Weights				
1	0	N/A	N/A	N/A
2	0.108	N/A	NA	N/A
3	0.336	N/A	0.147	N/A
4	0	0.198	0	0.045
5	0	N/A	0.35	N/A
6	0.006	0.469	0.044	0.183
7	0.001	N/A	N/A	N/A
8	0	0.333	0.458	0.175
9	0.345	N/A	N/A	N/A
10	0.204	0	N/A	0.597

*Significant at 10% level; **Significant at 5% level; ***Significant at 1% level

Table 2.14: Synthetic Control (matching on unemployment rates manually) results for all workers, for treated LLMs a, b, c, and d, and all treated LLMs pooled. Coefficient interpretable as the log point change in expected wages in the respective 'treated' region(s) compared to that region's synthetic counterfactual. Treated LLMs A-D, and donor control LLMs 1-10, are defined in Appendix B.

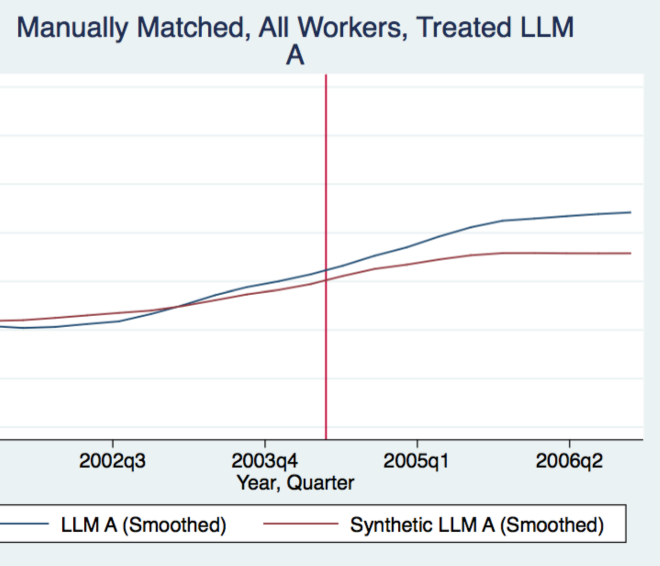


Figure 2.30

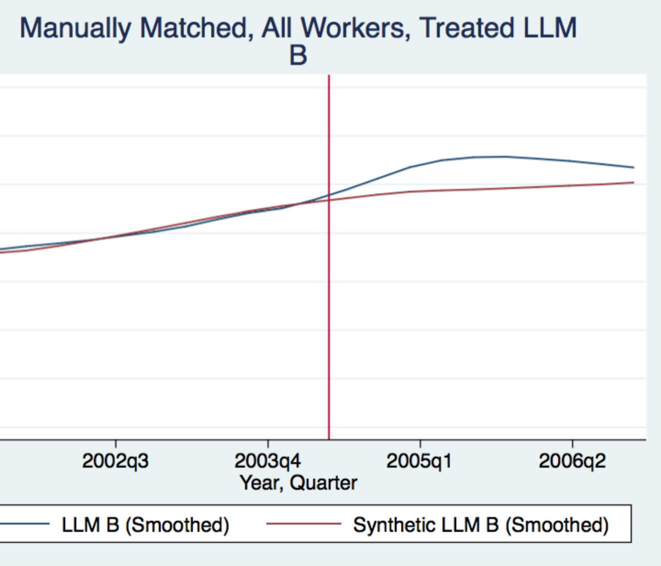


Figure 2.31

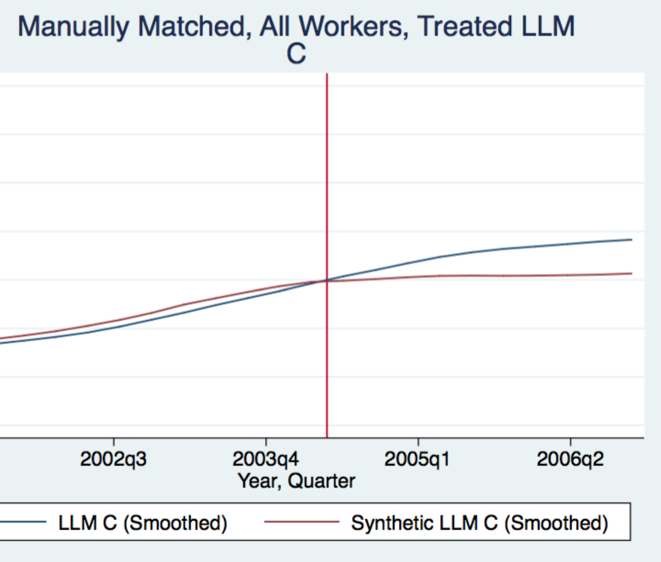


Figure 2.32

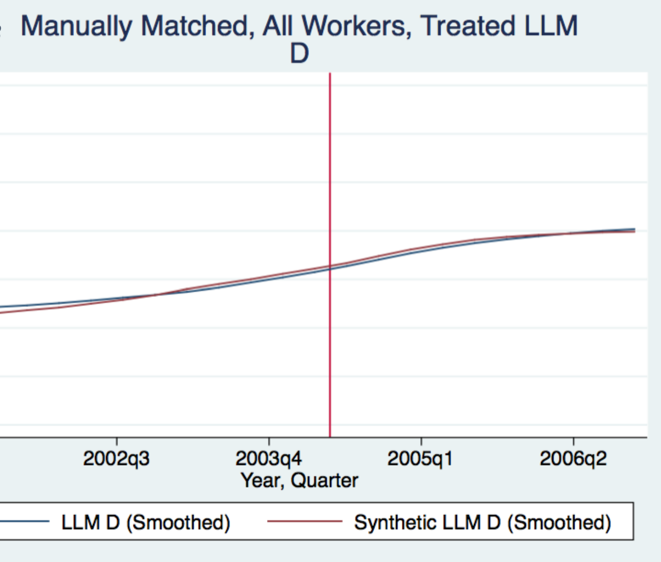


Figure 2.33

Low Education

Local Labour Market	a	b	c	d
$\hat{\beta}_3$	0.00112	0.00123	0.00221*	-0.00139
t-stat	1.05064	0.80649	1.81456	-1.53441
RMSPE	0.05056	0.03690	0.04146	0.09704
Spatial Cells	42	42	42	42
Donor Labour Market Weights				
1	N/A	N/A	N/A	N/A
2	N/A	N/A	0	N/A
3	N/A	N/A	N/A	N/A
4	0	0.13	0	N/A
5	0.347	0	0.502	N/A
6	0.02	0.595	0	N/A
7	N/A	N/A	N/A	N/A
8	0.204	0.276	0.498	0.274
9	N/A	N/A	N/A	N/A
10	0.43	0	N/A	0.726

*Significant at 10% level; **Significant at 5% level; ***Significant at 1% level

Table 2.15: Synthetic Control (matching on unemployment rates manually) results for those with low education, for treated LLMs a, b, c, and d, and all treated LLMs pooled. Coefficient interpretable as the log point change in expected wages in the respective 'treated' region(s) compared to that region's synthetic counterfactual. Treated LLMs A-D, and donor control LLMs 1-10, are defined in Appendix B.

Figure 2.34

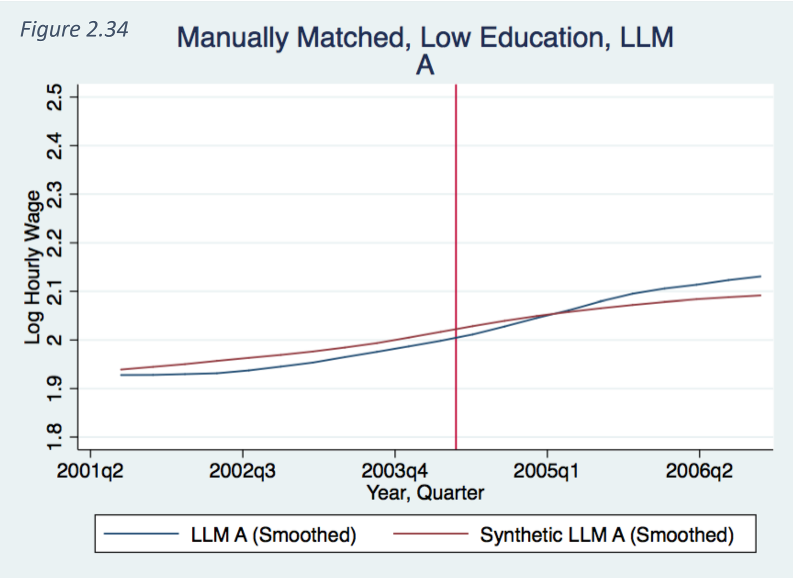


Figure 2.35

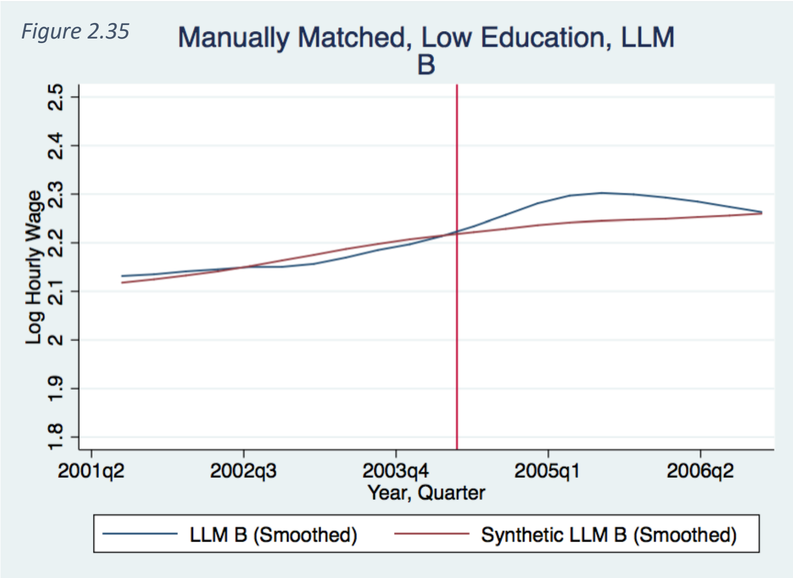


Figure 2.36

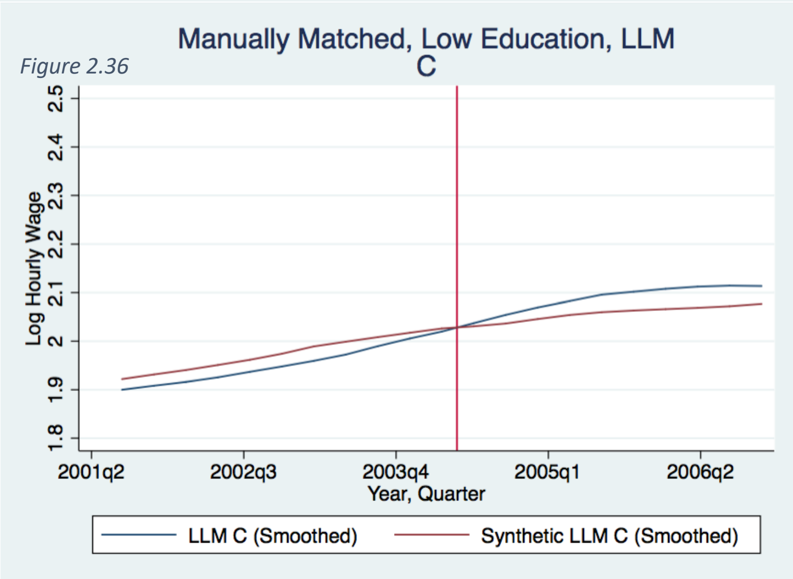
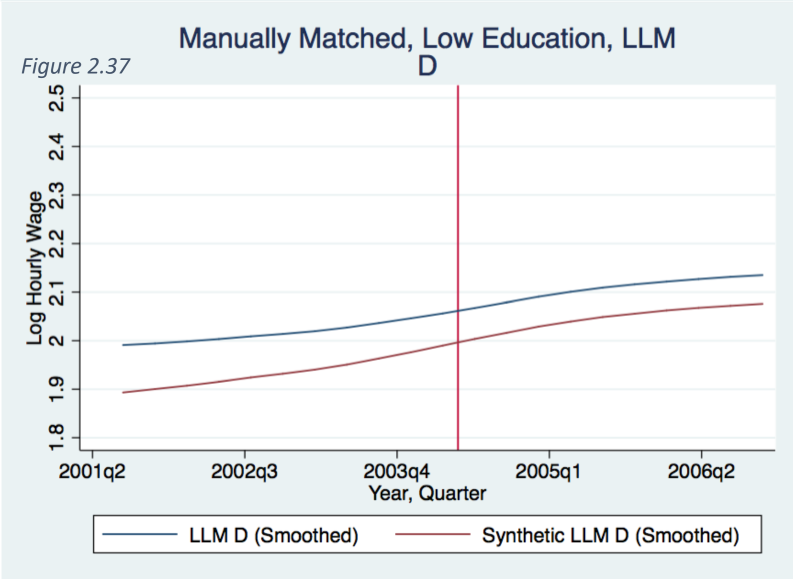


Figure 2.37



Under 30s

Local Labour Market	a	b	c	d
$\hat{\beta}_3$	0.00132	-0.00283	0.00107	-0.00171
t-stat	0.49531	-0.87399	0.58733	-1.07785
RMSPE	0.07900	0.11851	0.07556	0.04820
Spatial Cells	42	42	42	42
Donor Labour Market Weights				
1	0	0	0	0.273
2	0.059	N/A	N/A	N/A
3	0.67	0.132	0.587	N/A
4	0.021	N/A	N/A	N/A
5	0.251	0.022	0.202	N/A
6	N/A	0.744	N/A	0.234
7	N/A	N/A	N/A	N/A
8	0	0	0.211	0.149
9	N/A	N/A	N/A	N/A
131	N/A	0.102	N/A	0.344

*Significant at 10% level; **Significant at 5% level; ***Significant at 1% level

Table 2.16: Synthetic Control (matching on unemployment rates manually) results for under 30s, for treated LLMs a, b, c, and d, and all treated LLMs pooled. Coefficient interpretable as the log point change in expected wages in the respective 'treated' region(s) compared to that region's synthetic counterfactual. Treated LLMs A-D, and donor control LLMs 1-10, are defined in Appendix B.

Figure 2.38

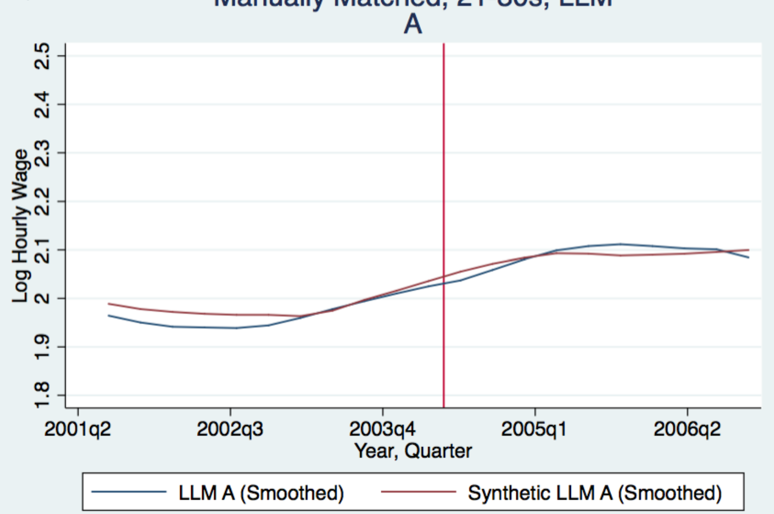


Figure 2.39

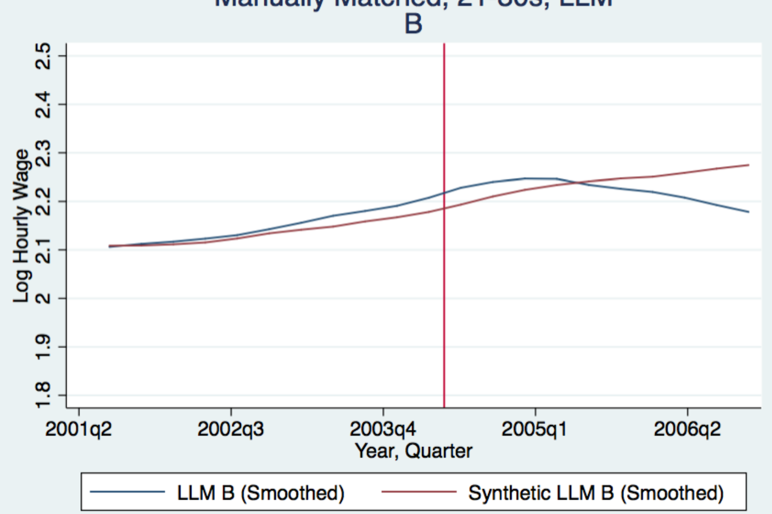


Figure 2.40

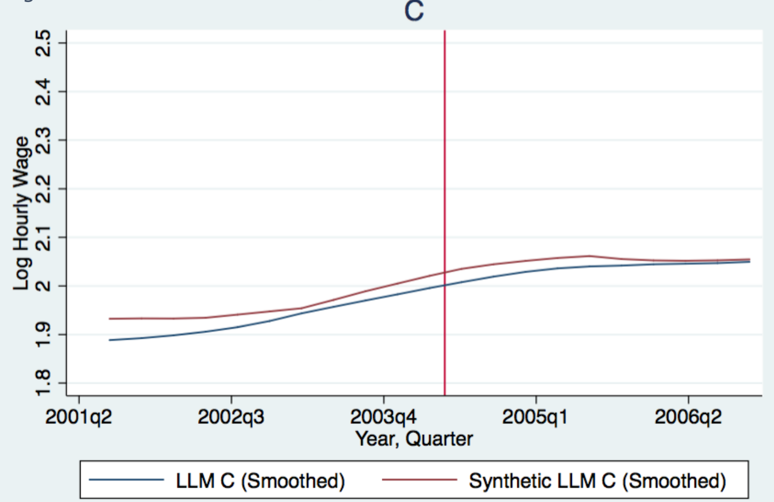


Figure 2.41

2.8 Appendix B

Treated LLMs

LLM	A	B	C	D
Unitary Authorities	Broxtowe Erewash Nottingham Rushcliffe	Barking & Dagenham Basildon Braintree Brentwood Castle Point Chelmsford Havering Maldon Newham Redbridge Rochford Southend-on-Sea Uttlesford Waltham Forest	Bradford Burnley Calderdale Craven Harrogate Kirklees Leeds Pendle Selby Wakefield	Babergh Birmingham Bromsgrove Cannock Chase Coventry Dudley East Cambridgeshire Forest Heath Ipswich Litchfield Mid Suffolk North Warwickshire Nuneaton & Bedworth Reddich Sandwell Solihull South Staffordshire St. Edmundsbury Stratford-Upon-Avon Suffolk Coastal Tamworth Warwick Wolverhampton
Sample Size (Labour force aged 16-65)	15,349	44,489	61,841	84,606
Unemployment Rate (at time of policy)	5.37%	4.91%	3.88%	5.29%
Average Hourly Wage (at time of policy)	£10.13	£11.72	£9.73	£9.65

Table 2.17

Figure 2.38

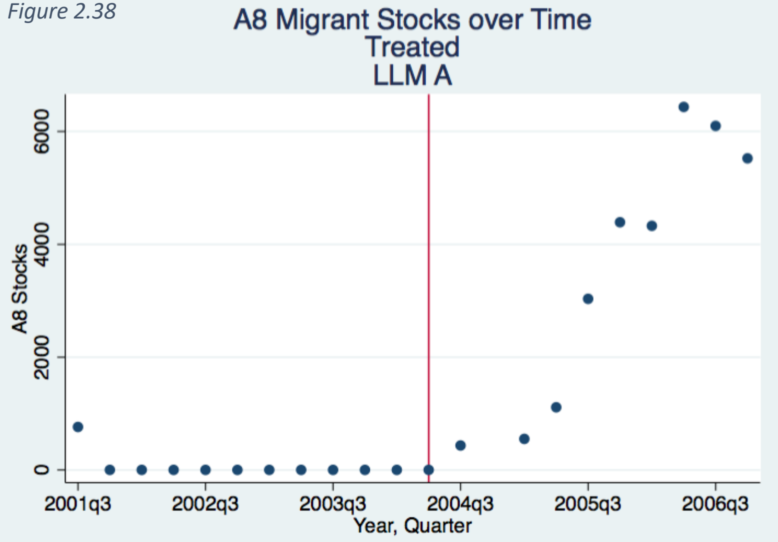


Figure 2.39

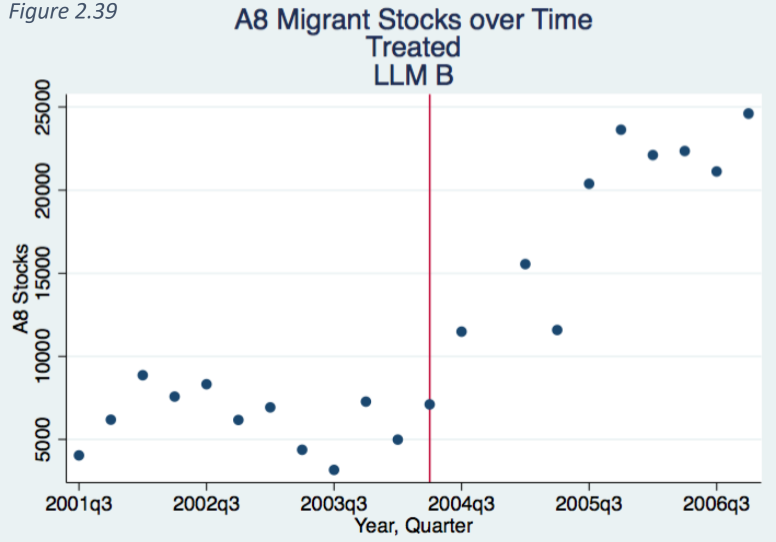


Figure 2.40

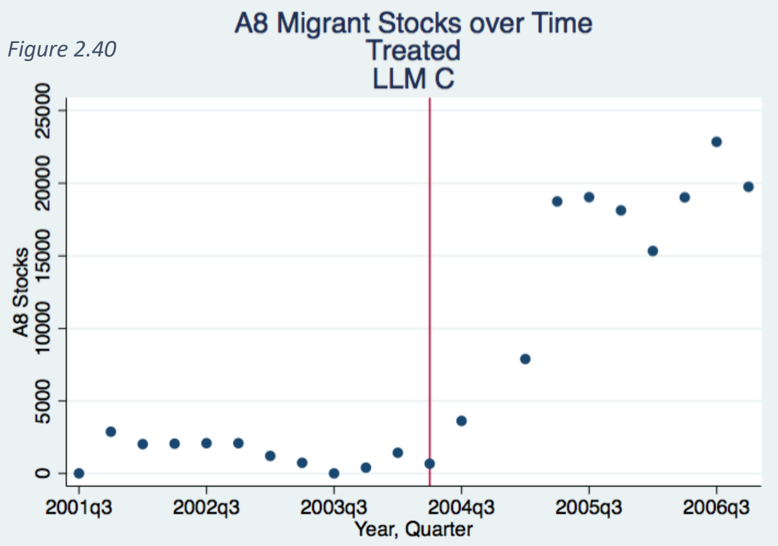
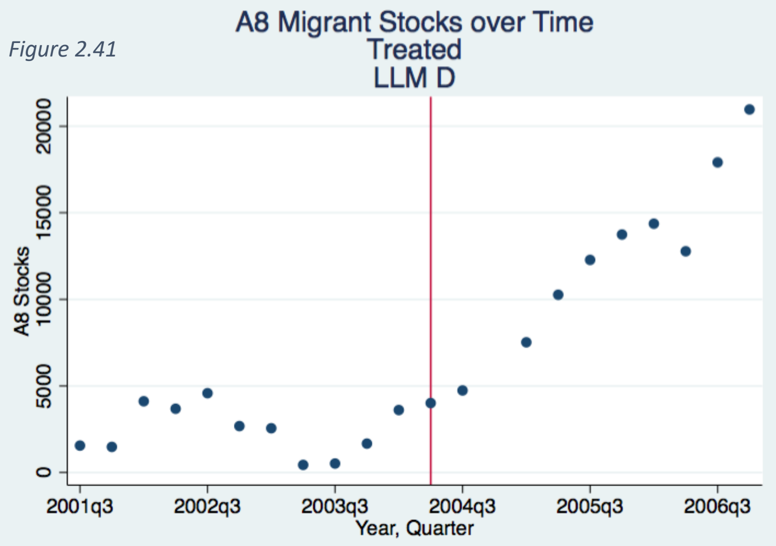


Figure 2.41



Donor Control LLMs

LLM	1	2	3	4	5	6	7	8	9	10
Unitary Authorities	Purbeck West Dorset Weymouth & Portland	Broxbourne East Hertfordshire Harlow	Fareham Gosport Havant Portsmouth	Three Rivers Watford	Dover Shepway	Croydon Tandridge	Southwark	Barnsley	Hartlepool Middlesbrough Redcar & Cleveland Stockton-on-Tees	Carmarthenshire North Port Talbot Swansea
Sample Size (Labour force aged 16-65)	4,131	7,054	11,395	4,382	4,451	9,631	4,486	4,822	11,831	10,551
Unemployment Rate (at time of policy)	2.37%	4.56%	4.51%	3.2%	3.5%	5.81%	16.22%	3.66%	7.45%	5.13%
Average Hourly Wage (at time of policy)	£9.44	£12.77	£9.69	£12.07	£10.56	£12.37	£13.32	£9.63	£9.59	£9.10

Table 2.17

Figure 2.42

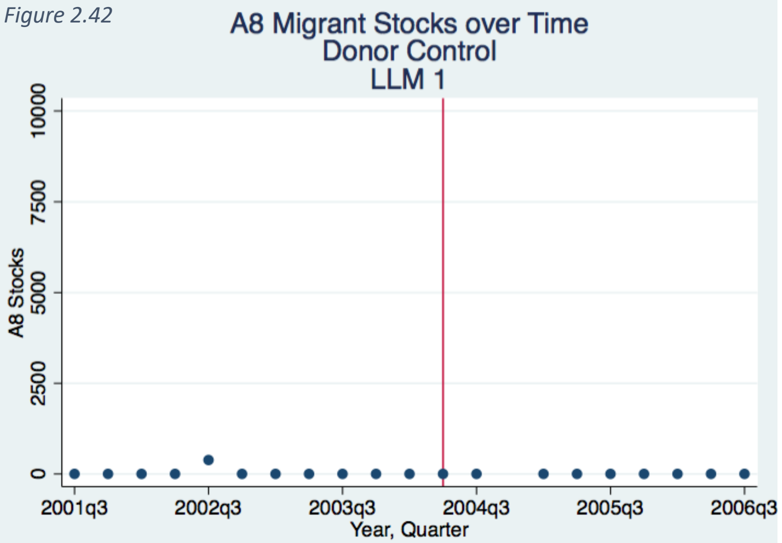


Figure 2.43

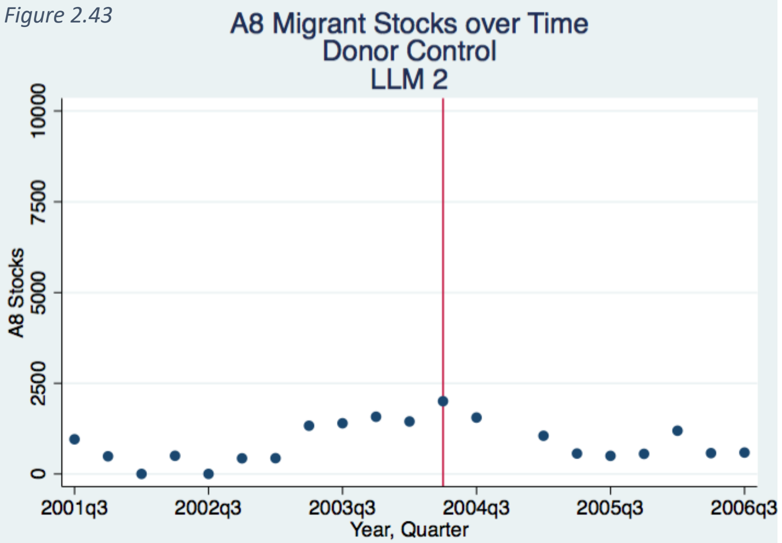


Figure 2.44

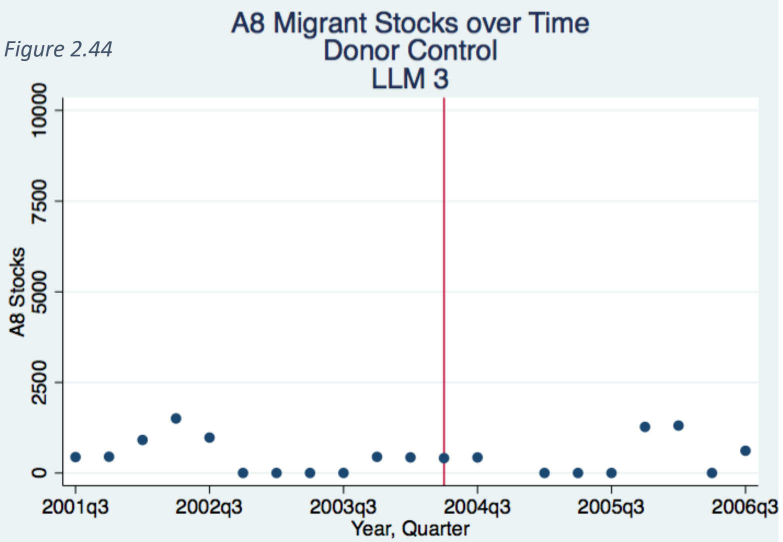


Figure 2.45

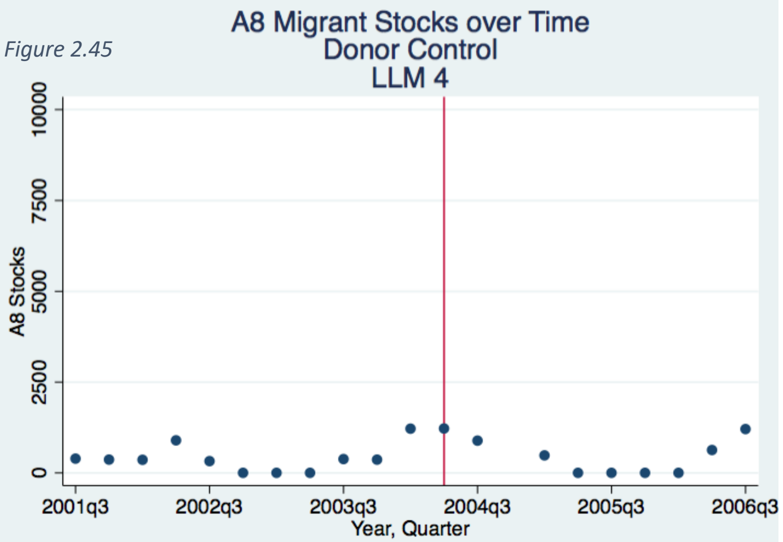


Figure 2.46

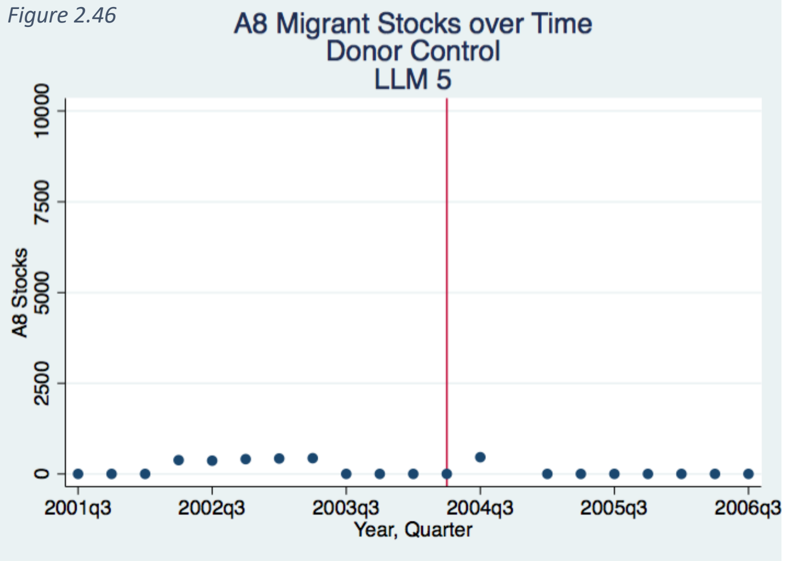


Figure 2.47

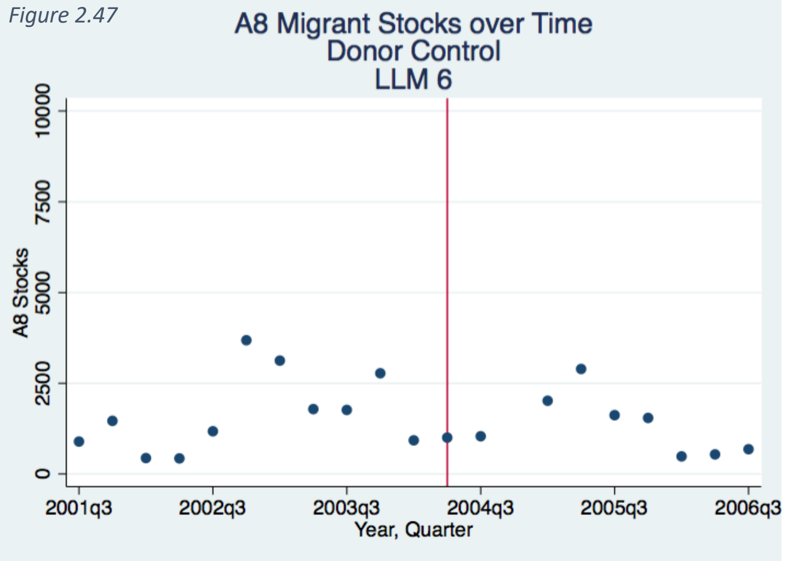


Figure 2.48

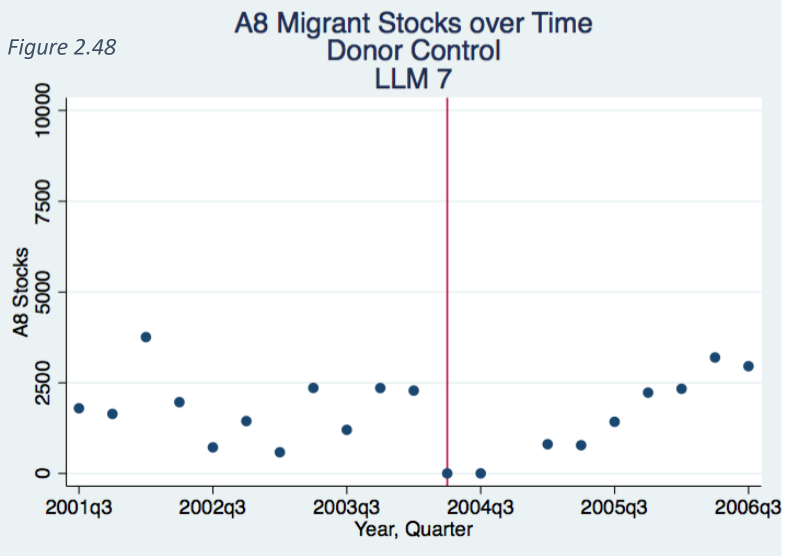


Figure 2.49

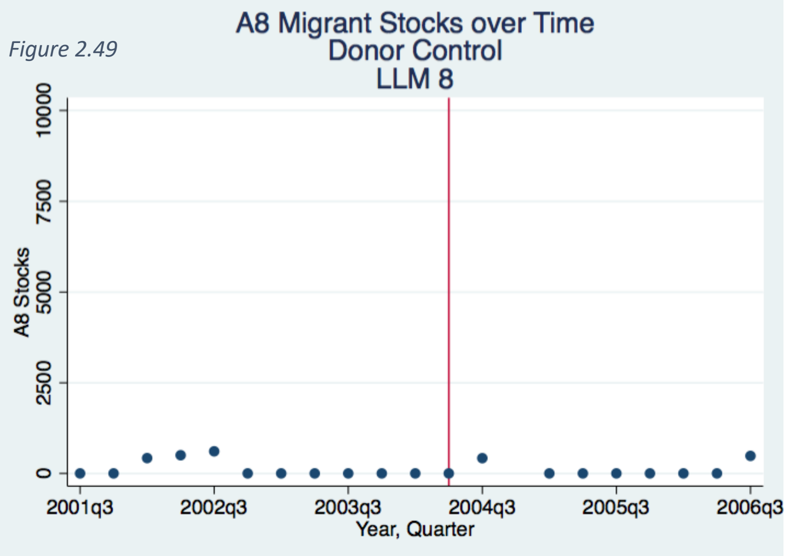


Figure 2.50

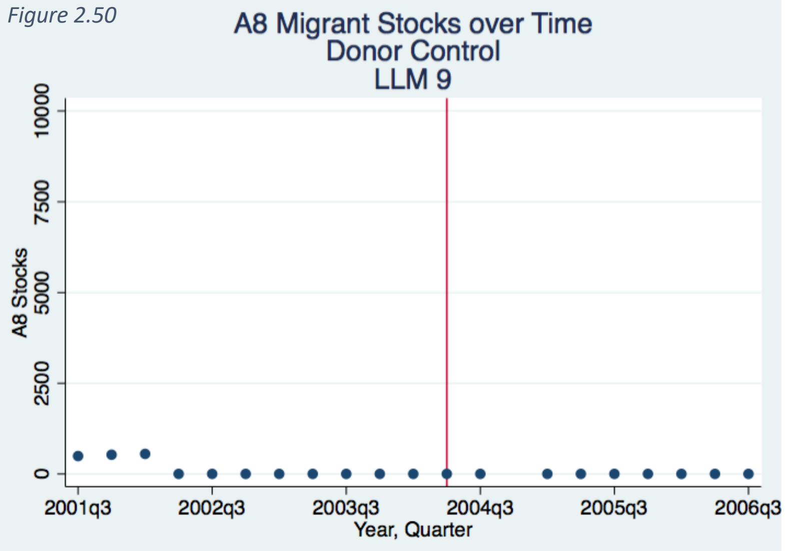
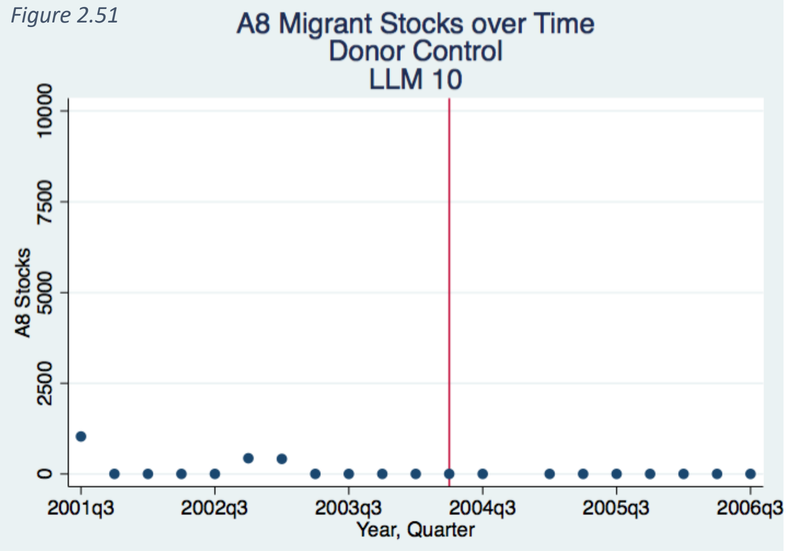


Figure 2.51



3. Impact of PAYG Ticketing on Rail Passenger Demand in London, and Wider Economic Impacts

With Chris Howorth* and Michael Naef**

*School of Management, Royal Holloway, University of London
**Department of Economics, Royal Holloway, University of London

3.1 Introduction

In this paper, we use the Synthetic Control Method (SCM) to evaluate the impact of Pay As You Go (PAYG) rail ticketing on rail passenger demand, as well as evaluating the secondary impacts on house prices, unemployment, and pollution in the locality of affected stations. We contribute to the literature by providing what we believe to be the on comparative case study of the impacts of PAYG ticketing in London, with what we believe to be the first application of SCM in the rail transport literature.

Oyster is a Pay-As-You-Go rail ticketing system, valid in the London Travel Zone (LTZ) on buses, trains, the London Underground, trams, and river boats; passengers travelling to or from stations outside of the LTZ must use traditional paper tickets. For passengers who travel to or from stations that are affected by LTZ extensions there are two potential driving factors for a change in travel behaviour: price, and convenience. We utilise SCM to conduct a comparative case study to identify the net change in passenger flows at the station level resultant of five LTZ expansions between 2010 and 2015; we find changes in passenger behaviour are driven entirely by price, and find suggestive evidence that the additional journeys are made by commuters. Further, we investigate the wider economic impacts to the immediate area surrounding stations affected by LTZ expansions where a change to passenger behaviour is identified; we find no evidence of wider impacts on house prices, unemployment, traffic, nor pollution.

Oyster was first introduced across the London transport network in 2004, as a contactless PAYG ticketing system. In its earliest form passengers pre-paid for travel by "topping up" the card at a convenient location, and accessed public transport by tapping their Oyster Card on an Oyster reader; at the end of each day and/or week, the system reconciles passengers' journeys, capping charges at the lowest possible fare for that day or week's travel (Day & Reed, 2010). In 2014 the system was simplified even further, allowing passengers to use their bank cards as an Oyster Card, with the appropriate daily/weekly cap being debited from that card

(Metro Report International, 2014). Initially, Oyster could only be used on the London Underground and buses; in 2010 it was rolled out onto National Rail services within the pre-existing London Travel Zone (Transport for London, 2011). Between 2010 and 2016, as part of the Mayor's Transport Strategy (Greater London Authority, 2010), the LTZ expanded to encapsulate various stations within the Greater London area and beyond into the Home Counties, allowing Oyster to be used at these stations for the first time; it is these extensions that are used for the purposes of analysis in this paper, as the fact that they were outside of the London Travel Zone altogether prior to expansion allows for the modelling of a clean counterfactual.

There exists a copious literature on the price elasticity of demand for rail travel in the UK, which is thoroughly reviewed by Wardman (2014). Studies typically find different results for the long and short runs, reflecting that it takes time for price changes to affect commuting habits and travel preferences. Using elasticity estimates from 167 studies, Wardman (2014) shows UK demand for rail travel is consistently price elastic in the Long Run across studies, and inelastic in the Short Run (PED of -1.11 vs -0.69); this is supported by Canavan et al. (2018), who find similar results for 32 Metro systems worldwide. Existing literature, however, posits a wide range of train fare elasticities. Owen and Phillips (1987) report inelastic London based inter-city short-run demand of -0.69 and long-run elastic demand of -1.08. Paulley et al. (2006)'s meta-analysis of 104 published and unpublished studies, providing 902 public transport fare elasticities, going back as far as 1951 but focusing on the period 1980 – 2002 (building on Webster and Bly 1980), finds inelastic short-run suburban rail demand of -0.6. Oxera (2005) finds inelastic short-run demand of -0.69 for rail travel between London and the rest of the country and long-run elastic demand of -2.03. For journeys in London and the South East, however, they report elastic demand for short-, medium- and long-run (1.47, -1.27 and -1.48 respectively). The studies summarised by Wardman (2014) suggest UK leisure travellers exert price elastic demand (PED of -1.05), whereas commuters exert price inelastic demand (PED of -0.82), and business passengers more inelastic still (PED of -0.62); this is supported by recent evidence from Chile (De Grange et. al., 2013), and China (Wang et. al., 2018). LeighFisher (2016) in a report for Department for Transport find Short Run price elasticities for Short-Distance journeys to London, and Short-Distance journeys to other metropolitan areas in the UK, to be similarly inelastic, with a PED of -0.69 in both cases. Mackett (1985), using the Leeds Integrated Land-use Transport Model (Mackett 1983) and focussing only on commuters traveling from Hertfordshire, however, found fares to be inelastic to central London (-0.4) but highly elastic for those traveling to Inner London[6] (-1.5) and marginally elastic for those traveling to neighbouring non-London locations (-1.1).

It is notable that much of the literature in this field is or refers to ageing material. Glaister (1983), Goodwin (1992) and Oum et al. (1992) are particularly well cited with many other venerable sources being cited as well. Button (2019) points out that rising incomes and changes in consumer taste may cause demand elasticities to change over time due to shifts in the demand function. Balcombe et al. (2004: 16) and Paulley et al. (2006: 297) also argue that fare elasticities can change over time and report that between the seminal Transport and Road Research Laboratory study undertaken in 1980 (Webster & Bly 1981) and 2004/06 short-run suburban rail fare elasticity increased from -0.5 to -0.58/-0.6. Short-run London bus fare elasticity, however, was fairly stable (-0.44 in 1980 and -0.43 on aggregate in 2006 (Paulley 2006). Short-run bus fare elasticity overall, however, rose from -0.30 in the 1980 study to -0.42 reported in Paulley (2006) and aggregate short-run UK metro fare elasticity doubled from -0.15 reported in Webster and Bly (1980) to -0.30 (see Table 3.1). In all cases, however, the range of reported values was wide suggesting a need for caution and careful interpretation.

Whilst the sheer range of recorded elasticities is difficult in itself, it is, as Hensher (2008), Oum et al (1992) point out (p.139), made still more challenging by the lack

[6] Inner London consists of the boroughs bordering Central London, as defined in the Local Government Act 1963

of consistent usage of the short- and long-run term. Owen and Phillips define the short-run as four weeks (p.242), and the long-run as within six months of a change (p.251). Oxera (2005) is consistent in using four weeks as the short-run period for all fare changes, but quote a nearly five years for long-run elasticity outside of London and the South East but only one year to reach 90% of the long term figure in London, Bresson et al. (2003) is vague in the timing of short- vs long-run (although short-run is longer than 1 month – p609). Paulley et al. (2006), on the other hand, assume the short-run to be between one and two years, medium run to be 5-7 years and the long-run to be twelve to fifteen years (and, possibly, as long as twenty). De Grange et al. (2013) defines short-run as under two years, medium-run as 2-5 years and long-run as anything over five years. Canavan et al. (2018) merely refers to short- and long-run without defining periods for either. Glaister (1983) does not engage with consideration of differences between short- and long-term elasticities. The lack of clarity and consistency over short- and long-run time frames is problematic in an industry sector in which franchises were, typically, and may still, be awarded for seven or eight years at a time (Preston 2016: 111; 2018: 187).

Measure	2006 Ave	2006 Lowest	2006 Highest	1980 Ave. elast.	# of 2006 Studies
Short-run UK bus fare elasticity	-0.42	-0.07	-0.86	-0.30	33
Short-run UK metro fare elasticity	-0.30	-0.15	-0.55	-0.15	15
Short-run UK Suburban rail fare elasticity	-0.58	-0.10	-1.02	-0.50	20
Short-run UK London bus fare elasticity	-0.43	0.14	-0.84	-0.44	15

Table 3.1. Comparison of fare elasticities between Paulley et. al. (2006), and the 1980 Black Book. Taken from Table 1 in Paulley et. al. (2006)

In terms of the impact of PAYG pricing on demand, the evidence is less clear. McCollom & Pratt (2004) find that PAYG ticketing appeals primarily to infrequent travellers; Graham & Mulley (2012) find PAYG is favoured by more wealthy passengers, whereas Cervero (1990) find it to be favoured by younger passengers.

Previous LTZ expansions have often been followed by claims from estate agents that there has been a resultant rise in demand for local housing stock, positively impacting prices (e.g. Evening Standard (2015), Savills (2017)), however serious authoritative evidence on the impact of Oyster, PAYG ticketing, or rail pricing in general, on house prices is surprisingly sparse. A literature does exist on the impact of new railway lines, which are typically found to increase property prices near those new lines, e.g. SuperTram in South Yorkshire (Henneberry, 1998), the Jubilee Line extension and Docklands Light Railway (Gibbons & Machin 2005), and Crossrail (Comber & Arribas-Bel, 2017). The magnitude of effects identified varies substantially between studies, and a minority of studies have reported a negative effect (e.g. Forrest et. al. 1996). Similar results can be found elsewhere in the world (see Debrezion, Pels & Rietveld (2007) for a review).

Similarly, the literature on the impact of rail on the labour market focusses largely on infrastructure investment as opposed to pricing. Combes et. al. (2008) notes that as new rail infrastructure is built, labour supply typically increases in affected areas at both intensive and extensive margins. Berechman & Paaswell (1983) analyse the wider economic impacts of the introduction of a light railway in Buffalo, New York, finding that there was a significant increase in retail employment in the local area after the introduction of the railway; in a follow-up study (Berechman & Paaswell, 2005) the authors note that this rise in employment was contingent on private sector investment occurring after the public funding for the railway was secured. In a case study on London's Crossrail project, Banister & Thurstain-Goodwin (2011) find that in addition to increasing labour force participation, the widening of the geographical labour market area will increase competition between firms, leading to innovation, in turn increasing labour productivity. The same study analyses the 1998 Jubilee Line extension, finding that employment near the new stations increased by 17% between 1998 and 2000, comparative to an 8% increase in Greater London more generally.

There does, however, exist a literature that explores the Economics of Commuting, which explains demand for rail travel in terms of the labour supply decision. Guglieminetti et al. (2017) found that unemployed workers will, initially, look for well-paid work within their locality with reservation wages increasing with commuting time. As periods of unemployment rise and remaining period on unemployment benefits (prior to unemployment assistance) dwindles job seekers will accept longer commutes and/or lower wages. Guglielminetti et al. (2017) also posit that in the case that unemployment benefits (from unemployment insurance) give way to less generous unemployment assistance in long term unemployment length of acceptable commute declines again as relative cost of job search increase. Nielsen and Hargesen (2008), using data from the 1991 and 2001 census and GIS mapping of commuter flows, show that the majority of commuting flow occurs within the corridor connecting London, Birmingham, Manchester and Leeds with concentration of commuter activity increasing with proximity to each of these metropolitan areas. Over time, however, Nielsen and Hargesen (2008) identify an increase in the width of the commuter corridor as greater numbers of workers choose to live in rural areas whilst taking employment based in urban areas.

Environmental impacts of rail pricing clearly depend on the impact of rail price changes on car usage. According to Paulley et. al. (2006) car usage and rail pricing are only weakly related in urban areas in the UK, with an estimated cross elasticity of 0.054. The study also estimates the cross elasticity between car cost and rail usage, finding it to be higher (0.59) for urban areas, and smaller (0.25) for interurban areas. Leape (2006) finds the introduction of congestion charging in London reduced the number of cars in the city by 34%. Transport for London estimated that about half of the fewer trips were diverted to some form of public transport, mostly buses. Effects on trains seem to be small and difficult to identify (Leape, 2006). This suggests that rail fare changes have little to no effect on people’s choice to use the car.

We contribute to this literature by presenting evidence on the impact of PAYG ticketing on demand for rail travel over a 3-6 year period, and the wider economic and environmental impacts where changes in demand are identified. We do not find any evidence that PAYG ticketing alone has an effect on demand, however we do find that there is a significant positive impact on demand where the change to fare structure results in lower fares for peak-time commuters. We do not find evidence that the expansion of Oyster has any effect on house prices, unemployment, traffic, nor pollution in the locality of the affected stations.

3.2 Data

Firstly, due to the differing spatial units reported in the various datasets used in our analysis, we utilise the ONS Postcode Directory (Office for National Statistics, 2018) to ensure stations are mapped to the correct CAS Ward, according to their postcode. 1 km^2 MAAQ spatial units are mapped to CAS Wards according to their centroid coordinates, and stations are mapped to their corresponding LSOA according to their full postcode. We also use these data to establish the geographical coordinates of stations, using which we calculate their distance from traffic count measures in the AADF data.

As a measure of passenger flows we utilise The Office for Rail and Road's Station Usage data (Office for Rail and Road, 2018), from 1997 to 2018; total annual passenger entries and exits are reported at the station level. Data in this series are predominantly sourced from ticket sales datasets LENNON and the revenue allocation mechanism MOIRA2, and are corrected with modelled estimates of journeys for which origin and/or destination are not recorded in ticket sales data (e.g. zonal products such as travelcards, and tickets sold at newsagents or by Train Operating Companies (TOCs) not covered by the Rail Settlement Plan) (Steer Davies Gleave, 2017). Due to various different firms being tendered to compile the annualised data over the history of the series, variables reported are not entirely consistent over time (e.g. some years include data on numbers of concessionary tickets, and sometimes types of concession, whereas other years not); however a measure of total annual entries and exits is consistently either reported or can be derived for all years apart from the 2003/04 financial year. We use these data to construct a panel dataset of annual passenger numbers at the station level. In most years there are very minor changes to estimation procedures, however these differences are marked for the 2006/07 financial year, and to a lesser extent 2007/08 and 2008/09 financial years, as various previously unobserved tickets, such as London Travelcards, were incorporated into the count through MOIRA2; this significantly affects observed passenger trends for many stations, particularly those in and around London. The 2006-09 changes do restrict the number of usable pre-treatment periods in our passenger demand analysis, particularly for the 2009/10 LTZ expansion to Grays, Ockendon and Purfleet, as it was immediately preceded by the changes in measurement. In all cases except the Zone A expansion, we therefore restrict the pre-treatment period in our analysis to the three years prior to the respective expansion, to exclude these inconsistent counts; for the Zone A expansion we restrict the pre-treatment period to the 2006/07 financial year onwards, as the change was most marked in 2006/07. In 2009/10 Oyster Pay as You Go (PAYG) fares were included for the first time, to coincide with the roll-out of Oyster PAYG onto the new London Overground (Steer Davies Gleave, 2017); this is unproblematic for our analysis.

Pollution is measured using Defra's Modelling of Ambient Air Quality (MAAQ), 2000-2018 data (Department for Environment and Rural Affairs, 2018). MAAQ utilises data from monitoring sites around the UK and local area characteristics to model specific pollutants at 1 km^2 concentration maps. Data validity is ensured by comparing modelled pollutants to primary data from measuring sites not included in the calibration process (Ricardo Energy & Environment, 2017). Aggregating 1 km^2 areas to CAS Ward level for comparability with spatial units in the other data in this study, we utilise annualised measures of Nitrogen Dioxide, Nitrogen Oxide, as well as PM_{10} and $\mathrm{PM}_{2.5}$ measures of particulate matter, to capture changes in pollutants commonly associated with vehicle usage, in line with the current literature.

Traffic flows are measured using The Department for Transport's Average Annual Daily Flows (AADF) data (both minor and major roads), from 2000 to 2018 (Department for Transport, 2018). Mean daily traffic at 31,844 locations across the UK are reported, with cars, light goods vehicles (LGVs), heavy goods vehicles (HGVs) and buses reported separately. Counts are estimated using a combination of manual counts, Automatic Traffic Counters (ATCs), and Automatic Number Plate Recognition (ANPR) cameras. In addition to the road name, the road type, longitude and latitude are also reported.

We utilise NOMIS claimant count panel data from 2004-2014 at the CAS Ward level (NOMIS, 2014) as a measure of unemployment. This is the most disaggregated measure of unemployment available in the UK; this is of particular importance to our identification strategy, given the large number of stations: if multiple stations exist within the same spatial unit, identification the effect of a change affecting only one of those stations is problematic. Data are reported monthly, however we aggregate to financial years such that data can be merged with ORR passenger data.

To measure house prices we use Land Registry cross sectional Price Paid data (HM Land Registry, 2018), which records quarterly residential property sales, and sale prices, for flats, terraced houses, semi-detached houses and detached houses at the postcode sector[7] level. We merge data across time to generate a panel dataset, and pool by financial year for compatibility with ORR data. Postcode sector does not perfectly predict CAS ward, so house price analysis is based on the postcode sector of the station.

[7] The UK is split into 9,550 postcode sectors. There are up to 390 postcodes per sector, and up to 80 households per postcode. Precise sizes of postcode sectors therefore varies; information on the number of households per sector are contained in Royal Mail PAF data, which are not freely available.

To facilitate the Synthetic Control approach, we utilise Index of Multiple Deprivation (IMD) (2010) data (Ministry of Housing, Communities and Local Government, 2015) to compare stations with similar local demographics. IMD data are available for 2007, 2010, and 2015; because the Oyster extensions analysed all occurred between 2010 and 2015, we utilise the 2010 measure in all analysis for direct comparability. The index is derived from a weighted combination of measures of income deprivation, employment deprivation, health deprivation, disability, human capital, barriers to housing and services, crime, and living environment deprivation, with weights allocated to each measure as determined by Dibben, et al. (2007) (Dibben, McLennan, Barnes, Noble, Davies, & Garratt, 2011). IMD data are disaggregated to the Lower layer Super Output Area (LSOA) level, which does not perfectly CAS Ward. This is unproblematic as we utilise IMD data as a matching variable, so we assign IMD indices to stations according to the LSOA in which the station resides.

3.3 Descriptive Statistics

3.3.1. Passenger Numbers

Table 3.2 details the stations that have been subject to Oyster extensions, the dates of those extensions, and the zone to which each station was allocated. Absent from this table is the 2016 Gatwick extension, which we exclude for two reasons: (a) Oyster remains more expensive than a paper ticket for some journeys (Oyster-Rail, 2019), and (b) Gatwick Airport itself was subject to significant infrastructure investment at a similar point in time to the Oyster extension (Gatwick Airport, 2018), which is likely to have impacted passenger numbers, thus impeding clean identification.

Figure 3.3.3 shows passenger numbers (total annual entries and exits) for stations that were subject to Oyster expansions and allocated into Zones 7-C respectively. In terms of trends, passenger numbers in the Zone 7 stations appear to follow a moderately positive trend prior to the extension, with much higher growth post-extension. For Zone 8, it is not clear visually that there is any deviation in trend for Cheshunt; there does, however, appear to be a significant positive deviation in passenger trends for Dartford. In terms of Zone A, the trend for Grays is (aside from the first data point) linear and slightly negative prior to the extension, linear and positive for the first four years post-extension, and then concave and positive thereafter; trends for Ockendon and Purfleet are linear and slightly positive, and there is no obvious deviation from trend around the point of the extension. The Zone B stations exhibit positive linear trends, with no clear deviation at the point of the extension; this is similarly true of the only Zone C extension, Shenfield.

Station Name	Financial Year of Oyster Extension	Zone
Waltham Cross	2012/13	7
Theobalds Grove	2012/13	
Cheshunt	2012/13	8
Dartford	2015/16	
Grays	2009/10	A
Ockendon	2009/10	
Purfleet	2009/10	
Brentwood	2012/13	B
Broxbourne	2012/13	
Ware	2015/16	
Hertford East	2015/16	
Rye House	2015/16	
St. Margaret's (Hertfordshire)	2015/16	
Shenfield	2012/13	C

Table 3.2. Rail stations subject Oyster zone extensions.

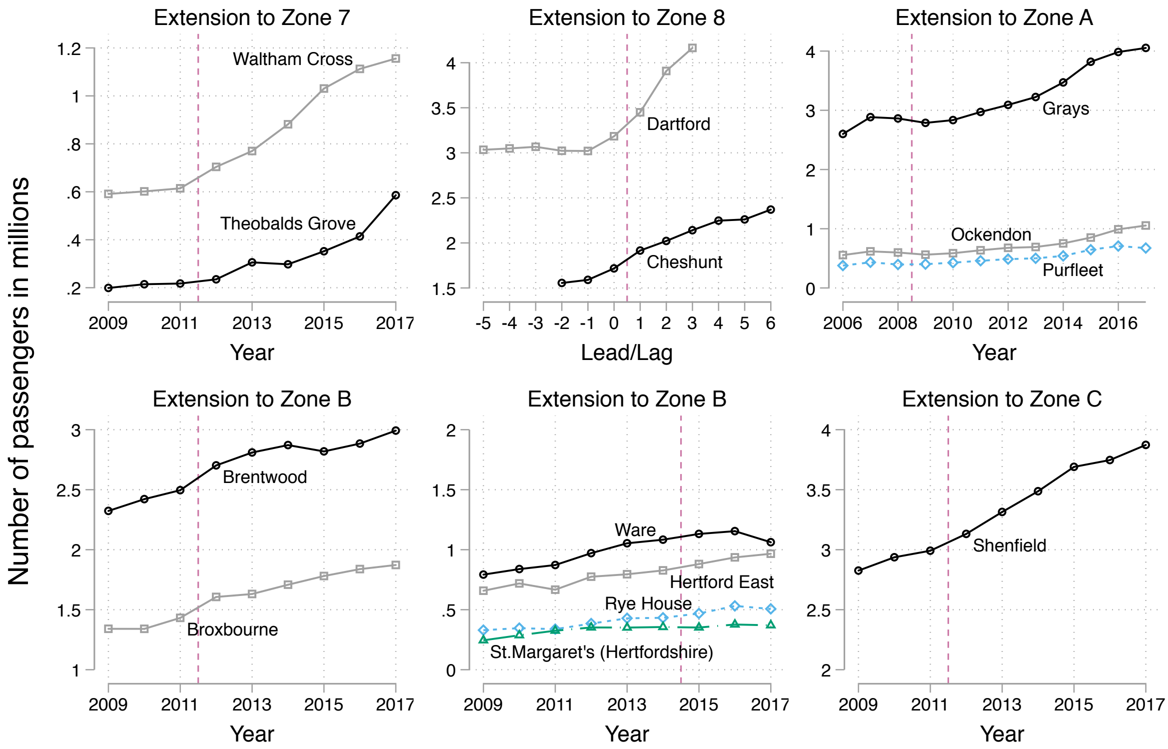


Figure 3.3. Number of passengers in stations subject to Oyster extensions. The vertical dashed line indicates the time the extension was implemented. The extensions of zones 8 and B were implemented in different years for the different stations. The Lead/Lag indicates the years after/prior to the extensions and a lead of 1 is the first year after the extension.

From the graphs, it therefore seems most likely that a positive impact on passenger numbers would be identified for Zones 7 and 8, which is what our counterfactual analysis will go on to demonstrate. Counterfactual analysis is necessary establish (a) whether deviations from trend are attributable to the extension of Oyster, and (b) that where no deviation from trend is observed, this would have been the case in the absence of the extension.

3.3.2. House Prices

Figure 3.4 shows the evolution in mean house price trends at the postcode sector level been 1997 and 2018 for stations encapsulated by Zones 7-C extensions respectively. Data represent house prices in the postcode sector of the respective station. Prior to extensions, house prices in all postcode sectors follow an upward, broadly linear trend, save for a sharp drop between 2008 and 2009 owing to the financial crash. For Zones 7 and 8, both of which occurred in 2012/13, there does appear to be an increase in house price growth in the period immediately after the Oyster extensions, however in each case this could be equally attributed to post-crisis recovery, as broadly prices seem to be returning to their pre-crisis trend. The Zone An extension, occurring in 2009/10, happened too close to the financial crisis to draw meaningful conclusions graphically: there was a sharp increase in house prices immediately after the extension, but much of this growth is likely to be attributable to the macroeconomic recovery. In Zones B and C there is very little deviation from trend; Broxbourne in Zone B does exhibit a spike in house prices immediately prior to, and after, the Oyster extension, however house prices return to trend in 2017/18, suggestive that this is anomalous.

Particularly given the temporal proximity of the various Oyster extensions to the financial crisis, the counterfactual analysis we present in the Methodology and Results section of this paper is of particular importance to net out any post-expansion changes to house prices that are not attributable to Oyster.

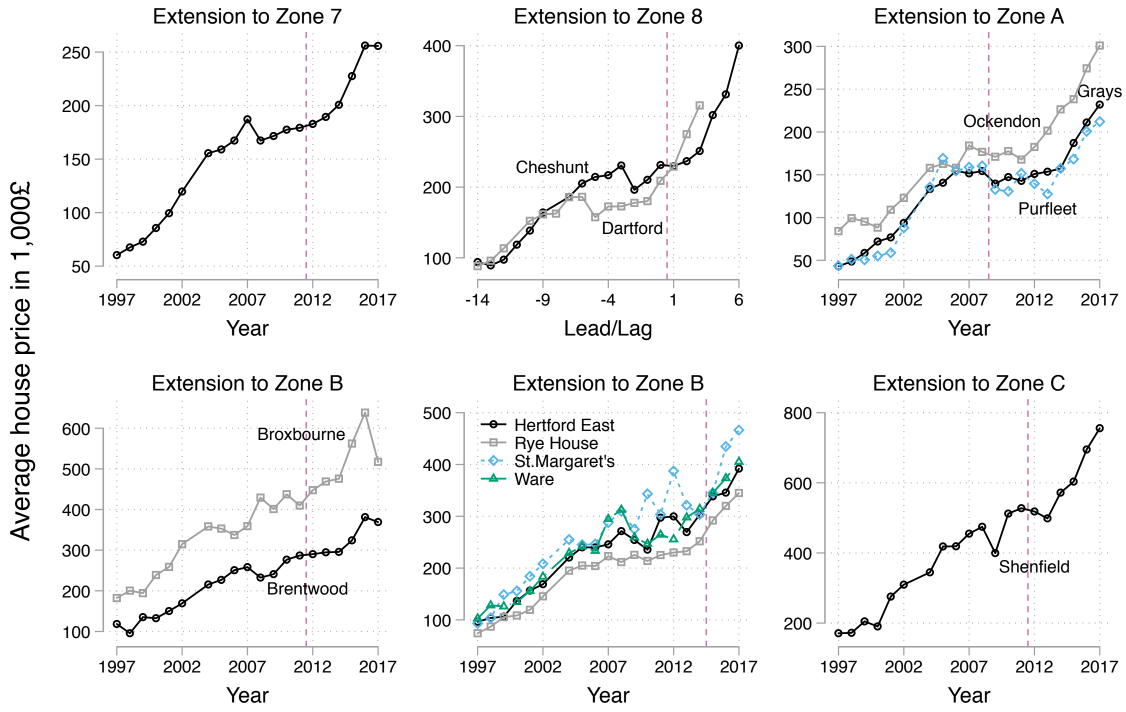


Figure 3.4. Average house price at the post code of the stations subject to Oyster extensions. The vertical dashed line indicates the time the extension was implemented. The extensions of zones 8 and B were implemented in different years for the different stations. The Lead/Lag indicates the years after/prior to the extensions and a lead of 1 is the first year after the extension.

3.3.3. Unemployment

Figure 3.5 shows Jobseeker's allowance (JSA) claimant count trends at the CAS Ward level between 2004 and 2014 for stations affected by Oyster expansions. Note that Dartford and all but one Zone B stations are excluded from this figure because these expansions occurred after this series ceased to be published by NOMIS. For all extensions observed, the claimant count was largely flat prior to the 2008 financial crisis. After 2008, claimants increased sharply, peaking between 2009 and 2011, before gradually decreasing. In Zones 7, 8 (Cheshunt), Zone 9 (Ockendon and Grays) and Zone C, the claimant count is observed to be monotonically decreasing post-expansion. In Zone A, the claimant count was increasing for three years post expansion, before trending downwards.

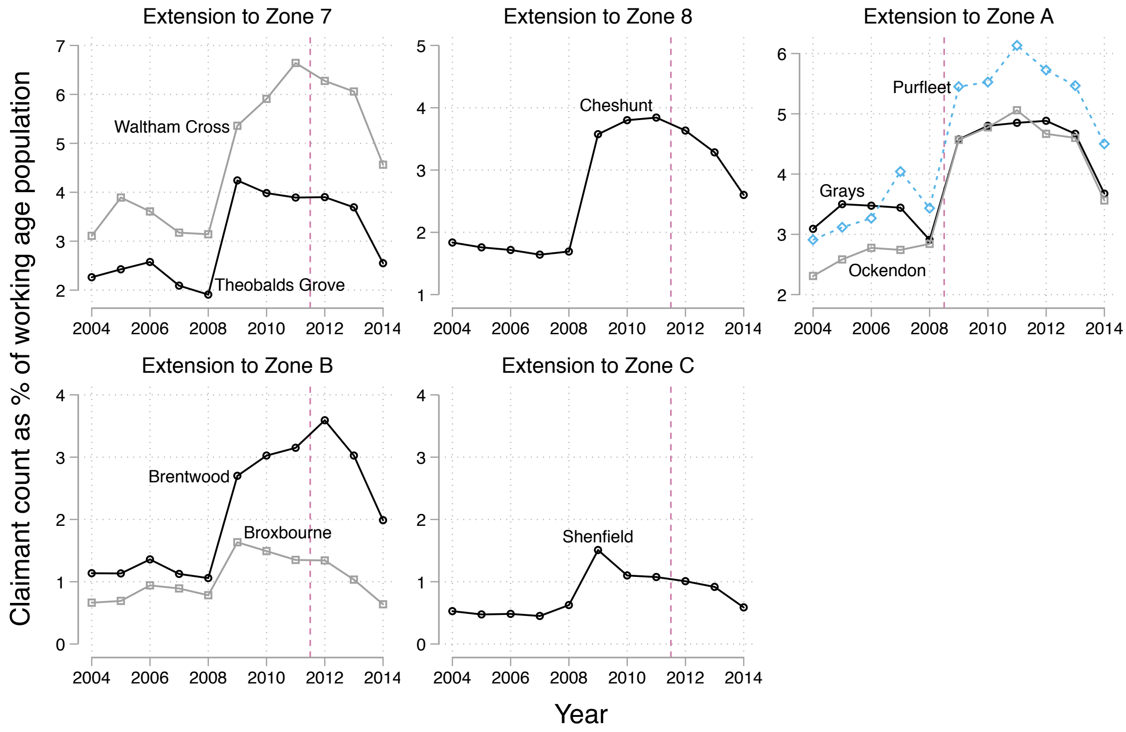


Figure 3.5. Jobseeker's allowance claimant count as percentage of the working population at the CAS ward level of the stations subject to Oyster extensions. The vertical dashed line indicates the time the extension was implemented.

Similarly to the house price data presented above, it is difficult to draw meaningful conclusions from these data alone, because of the temporal proximity of the various extensions to the financial crisis. This challenge is augmented when analysing unemployment: labour markets do not adjust immediately to output shocks due to the prevalence of employment contracts (Bachmann et. al., 2015), and there is observed persistence of impacts on some segments of the UK labour market following the financial crisis (Bell & Blanchflower, 2010), (Bell & Blanchflower, 2011), (Coulter, 2016), (Singleton, 2018).

All stations exhibit a sharp decrease in their respective CAS Ward claimant count between 2013 and 2014. It should be noted that whilst some portion of this decrease is likely attributable to macroeconomic recovery, it coincides with the introduction of the introduction of the Jobseekers Act 2013, which imposes sanctions on job seekers who exhibit imperfect compliance with the Back to Work Scheme, which ultimately resulted in temporary or permanent removal of JSA entitlement for around 11% of claimants in 2013/14 (Oakley, 2014). The close proximity of this national policy change to the 2012/13 Oyster expansions further precludes meaningful analysis of the raw trends.

Because both of these phenomenon occurred nationally, the counterfactual analysis we present in the latter sections of this paper allows us to cleanly identify effects on the claimant count measure of unemployment attributable to Oyster expansion; using this methodology we do not find any evidence of an effect of Oyster expansion on local unemployment.

3.3.4. Pollution

Consistent with the current literature, we assess air quality using measurements of Nitrogen Dioxide, Nitrogen Oxide, PM_{10}, and $PM_{2.5}$ at the CAS Ward level between 2001 and 2017 for stations affected by Zone 7 – Zone C extensions respectively. In most cases there is a notable decrease in both Nitrogen Dioxide and Nitrogen Oxide between 2001 and 2004, followed by a steady downward linear trend, with no obvious deviation around the point of the respective Oyster extension. Both measures of particulate matter, PM_{10}, and $PM_{2.5}$, exhibit notable increases between 2002 and 2004; between 2004 and 2010 there is a downward linear trend, preceding an increase in 2011 in all zones. Between 2011 and 2017 all zones exhibit a downward linear trend in particulate matter, save for a notable downward deviation in 2015. Once again, there is no obvious deviation attributable to the policy.

In the analysis section of this paper we carry out counterfactual analysis on pollution trends, and find no evidence of a change in pollution levels attributable to Oyster expansion. We seek to explain these results by analysing traffic flows near the most affected stations, and find some suggestive evidence of a temporary increase in flows of cars attributable to the policy.

3.4 Identification Strategy

Using the data available, deriving a before/after estimator by OLS in respect to passenger numbers, house prices, unemployment, traffic, and pollution would be straightforward. Drawing causal inference from such an estimator, however, would not be appropriate: variation in the regional measures of each dependent variable is likely determined mostly by observed and unobserved factors other than the Oyster expansion. We must therefore construct a counterfactual measure of what would have happened to treated stations (those that were subject to Oyster expansion) in the absence of treatment. One approach would be to use difference in difference estimation, using OLS to compare changes in the evolution of outcome variables in treated compared to untreated stations after each expansion. The key challenge with this method is the selection of stations to use as a control group: our full sample contains 14 treated and 2,541 untreated stations. Untreated stations would need to be selected such that they (a) are characteristically similar to those stations that received Oyster in terms of observables that may predict responsiveness to the policy (e.g. distance from Central London, rurality, socio-economic characteristics, etc.), and (b) exhibit pre-treatment trends in the outcome variable that are not statistically different from the respective treated station (Autor, 2003). This would involve designing a formal selection process for control stations, in turn leading to the risk that results are driven by the way in which controls are chosen (Abadie, Diamond, & Hainmueller, 2010), or specification searching (Caudill & Holcombe, 1999) (Vivalt, 2019). Instead, we follow Abadie & Gardeazabal (2003) and use the Synthetic Control Method (SCM) data-driven procedure to select a weighted combination of untreated stations to form a comparison group for each treated unit.

We firstly utilise the random nature of the timing of the various extensions to argue that the extensions analysed represent a natural experiment. During the initial 2009/10 introduction of Oyster onto National Rail within existing London Travel Zones, it was not possible to introduce Oyster onto other London National Rail services due pre-existing franchise agreements. Post 2009/10, all LTZ extensions were negotiated into the renewal of the franchise agreement for that respective line, in accordance with the then Mayor's Transport Strategy (Greater London Authority, 2010). Franchise agreement renewal dates varied by train line, and were predetermined before Oyster's introduction and/or expansion onto National Rail, in part dependent on the order in which the various lines were privatised some decades previously. The stations explored in this study receiving or not receiving treatment in 2009/10 was therefore quasi-random, as was the order of treatment assignment post 2009/10.

Formally, there exist $J + 1$ stations, which are denoted $j = 1,2,\dots,J + 1$, where the 'treated' station is denoted $j = 1$ and all $j > 1$ are donor control stations. The data span T periods, and periods denoted T_0 occur before the respective LTZ extension. For each station, j, and time, t, we observe the outcome variable Y_{jt}. For each station we also observe k predictors of the outcome, $X_{1j}, X_{2j}, \dots, X_{kj}$, which include pre-LTZ extension observations of Y_{jt}, which are independent of LTZ extension. The $k \times 1$ vectors $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_{J+1}$ contain the values of predictors $X_{1j}, X_{2j}, \dots, X_{kj}$, and the $k \times j$ matrix $\mathbf{X}_0 = [\mathbf{X}_2 \dots \mathbf{X}_{j+1}]$ therefore collects the values of the predictors for the J untreated stations. Y_{jt}^N is defined as the expected post-treatment outcome in 'treated' stations in the absence of treatment, and Y_{jt}^I the observed post-treatment outcomes in treated stations. The effect of the LTZ extension on treated station outcomes is therefore:

$$\tau_{1t} = Y_{jt}^I - Y_{jt}^N$$

Clearly Y_{jt}^N is unobserved. The SCM method estimates $\hat{Y}_{jt}^N$ using one or more untreated station that had similar characteristics to the treated station in the observed pre-treatment period. The synthetic control group is defined as a weighted average of the stations in the donor pool. A synthetic control can be represented by a $J \times 1$ vector of weights, $\mathbf{W} = (w_2, w_3, \dots, w_{J+1})'$, such that:

$$\hat{Y}_{jt}^{N} = \sum_{j=2}^{J+1} w_j Y_{jt}$$

and:

$$\tau_{1t} = Y_{1t} - \hat{Y}_{jt}^{N}$$

$\sum_{j=2}^{J+1} w_j = 1$, and $w_j \geq 0$, so synthetic control groups are weighted averages of donor stations. Given a set of non-negative constants $v_1, v_2, \dots, v_k$ (the derivation of which is discussed later in this section), optimal weights $\mathbf{W}^* = \left(w_2^*, w_3^*, \dots, w_{J+1}^*\right)'$ are chosen to minimise:

$$\|\mathbf{X}_1 - \mathbf{X}_0\mathbf{W}\| = \sqrt{\sum_{h=1}^{k} v_h \left(X_{h1} - w_2 X_{h2} - w_3 X_{h3} - \dots - w_{J+!} X_{hJ+1}\right)^2}$$

subject to $\sum_{j=2}^{J+1} w_j = 1$, and $w_j \geq 0$. The estimated treatment effect is therefore:

$$\hat{\tau}_{1t} = Y_{1t} - \sum_{j=2}^{J+1} w_j^* Y_{jt}$$

$\mathbf{V} = (v_2, v_3, \dots, v_k)$ is chosen such that the synthetic control $\mathbf{W(V)}$ minimises the mean squared prediction error of the synthetic control with respect to Y_{1t}^{N}:

$$\sum_{t \in \tau_0} \left[Y_{1t} - w_2(\mathbf{V}) Y_{2t} - \dots - Y_{1t} - w_3(\mathbf{V}) Y_{3t} - Y_{1t} - w_{J+1}(\mathbf{V}) Y_{J+1t}\right]^2$$

for some $\tau_o \subseteq \{1,2, \dots, T_0\}$.

A challenge with SCM is that, unlike regression-based approaches, standard errors are not reported, as only one data point exists for both treated and synthetic units per time period. We therefore use the technique developed by Cavallo et. al. (2013) to determine the significance of results. Significance is determined by carrying out a series of placebo tests, by applying SCM to every potential control station in the sample, and comparing the estimated effect for the treated station to the distribution of the estimated effects for the untreated placebo stations. This procedure generates a lead-specific $p - value_l = \frac{\sum_{j=2}^{J+1} I(|\hat{\alpha}_{1,l}^{PL}| \geq |\hat{\alpha}_{1,l}|)}{J}$, where $\hat{\alpha}_{1,l}^{PL}$ is the lead l specific placebo treatment effect, and $I(\cdot)$ is an indicator function. Notably $p - value_l$ has a non-standard interpretation: it reports the proportion of placebo effects identified that are higher than the effect identified for the treated unit. Also following Cavallo et. al. (2013), we carry out a diagnostic check on the quality of the pre-treatment match by reporting the proportion of placebo units which have a pre-treatment RMSPE at least as high as the treated unit.

To avoid the possibility of "cherry-picking" predictor variables, that is, researchers choosing predictor variables through trial and error which yield significant results (see Ferman, Pinto, & Possebom (2017) for a full discussion) we use the same predictor variables for the analysis of effects on all outcome variables throughout this paper (we also include a robustness check in which all but essential predictor variables are dropped, as discussed below). Stations are matched on IMD (2010) scores, to ensure the socio-economic characteristics of the LSOA in which the treated station resides are similar to the untreated comparison. CAS Ward population is used as a predictor, as a proxy for rurality. We also match on geographical distance from Central London, as Oyster users are likely to be commuters to Central London, so it is important that the comparison unit is similar to the treated station in this respect; further, we restrict our sample of untreated donor control stations to those less than 25 miles from Central London for the same reason; this leave 114 donor control stations to utilise for analysis. Additionally, we match on pre-treatment outcomes in two different ways, depending on the specification of the outcome variable. For passenger numbers, house prices, and traffic flows, we specify the outcome variable as a logarithm (base 10) of itself, to allow comparison of the percentage change in the outcome over time; where this is the case we match on pre-treatment trends as per Cavallo et. al. (2013), using the Stata package developed by Galiani & Quistorff (2017). In other cases the outcome variable is specified in levels, as matching on the absolute pre-treatment value as opposed to trends is important (in the cases of unemployment and measures of pollution); where the outcome variable is specified in levels we match on mean pre-treatment outcomes, following Abadie & Gardeazabal (2003), Abadie, Diamond, & Hainmueller (2015), and Kleven, Landais, & Saez (2013).

For robustness, we also report results where we only use distance from Central London, and either pre-treatment trends or mean pre-treatment outcomes (for outcome variables measured in logs or levels respectively), as predictor variables to demonstrate that results are not contingent on the specification (results from this specification are provided in Appendix A). For all outcome variables apart from pollution measures, the pre-treatment fit is better with the full set of predictors (Pre-Treatment Diagnostic P-Value is higher); in the case of unemployment, this is to the extent that the Pre-Treatment Diagnostic P-Value without a full set of predictors is significant (the specification fails the diagnostic test). In all cases, point estimates are slightly changed (in both directions), but level of significance is unchanged.

3.5. Results

3.5.1. Passenger Numbers

Table 3.3 and **Error! Reference source not found.**, compare evolution of passenger numbers for each zone comparative to the respective synthetic control zone. For all zones, the Pre-Treatment Diagnostic P-Value is insignificant (indicating that pre-treatment trends are not statistically different between treatment and synthetic controls). We identify a large positive and significant impact of Oyster on passenger numbers for Zone 7 stations (rising from 0.007 log points in year 1 to 0.057 log points in year 6), and a smaller positive but significant (in years 2 and 3) impact for Dartford in Zone 8. For Zone A, we identify significant impacts in years 4, 5, 7, 8 and 9, and a very small but significant positive impact for Zone C in year 6. We find a very small negative, but significant, impact in Zone C for years 1 and 2 only. Results for Zone 8 (Cheshunt) are of a slightly smaller magnitude than Dartford (e.g. 0.009 log points in year 3, compared to 0.016 in Dartford), however results are not significant. For Zones A and B, results are positive but smaller still, and are not significant.

The significant results for Zone 7 pass our robustness check (see Appendix A), remaining significant at the 5% level for years 2-6, and at the 10% level in year 1. Results for Zone 8 (Dartford) remain significant at the 10% level in years 2 and 3, and all other results are not significant. We therefore conclude that Oyster positively affected passenger numbers for Zone 7 stations, and for Zone 8 (Dartford); we cannot conclude that it had any significant effect on passenger numbers for other zones.

With the ORR data alone, we cannot compute price elasticities from these estimates because (a) we do not observe journey origins nor destinations, and (b) we do not observe what type of ticket each passenger purchases. However, we can draw meaningful inference by observing the percentage price changes resultant of switching from paper tickets to Oyster for various combinations of journeys. Tables 4 and 5 present the percentage discount (in 2019 prices) between Oyster and various ticket combinations of paper tickets, for Off-Peak and Peak Travel respectively. In both tables, column (1) represents passengers who travel one way, only between the stations affected by a particular LTZ extension and the Zone 1 terminus of that line; column (2) represents passengers who complete return journeys between the stations affected by a particular LTZ extension and the Zone 1 terminus of that line; column (3) represents passengers who complete return journeys between the stations affected by a particular LTZ extension and another Zone 1 station, and (4) represents passengers who complete return journeys between the stations affected by a particular LTZ extension and Zone 1, and make other journeys over the course of the day. Columns (3) and (4) are of particular importance empirically as there is no reason to assume a journey to/from a particular LTZ extension station will be between that station and another station on the same line.

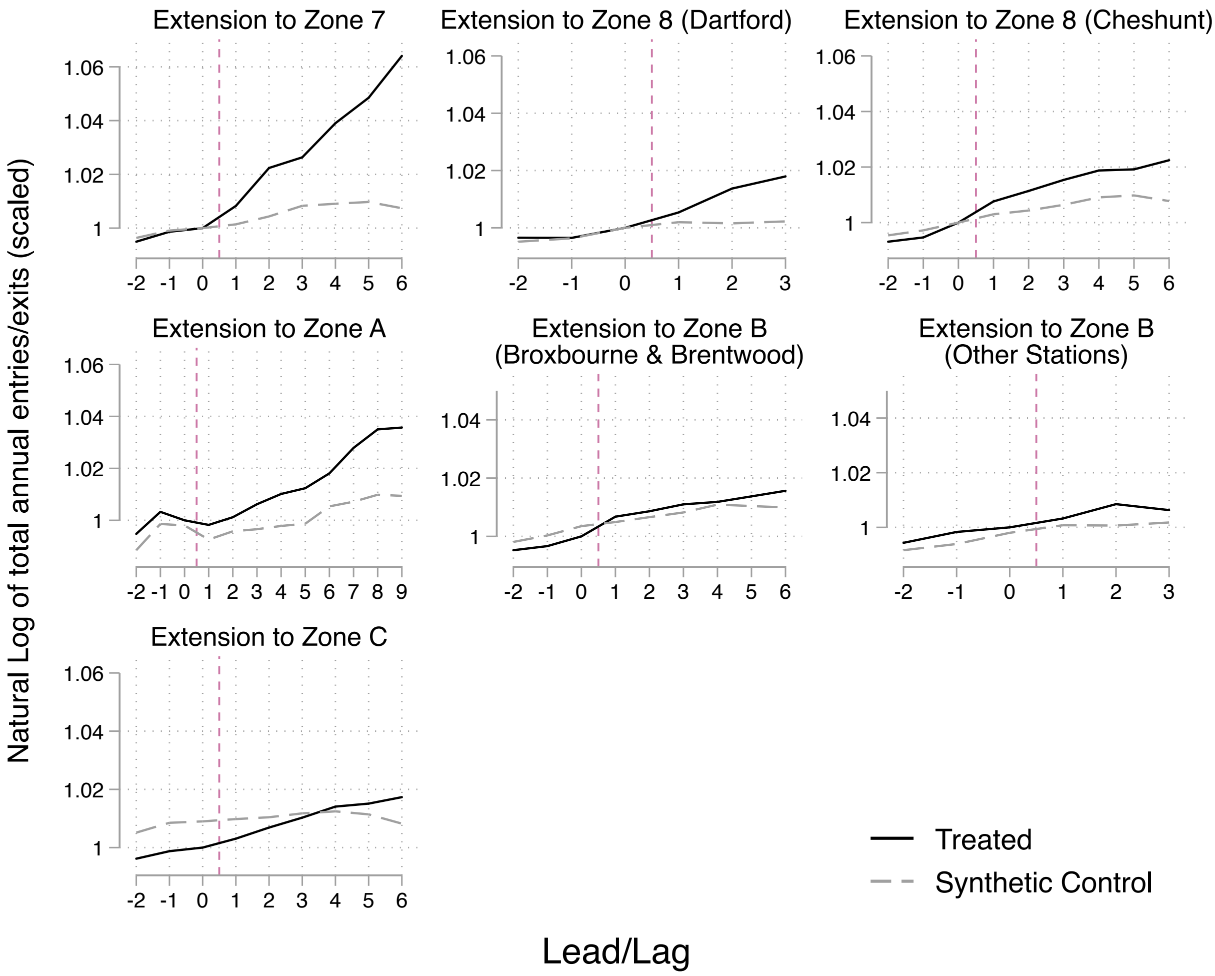


Figure 3.6. Evolution of passenger numbers for each zone comparative to the respective synthetic control zone. The vertical dashed line indicates the time the extension was implemented. The Lead/Lag indicates the years after/prior to the extensions and a lead of 1 is the first year after the extension.

	Pre-Treatment Diagnostic P-Value	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Year 8	Year 9
Zone 7	0.895	0.007**	0.018**	0.018**	0.030***	0.039**	0.057**			
		[0.029]	[0.018]	[0.015]	[0.000]	[0.017]	[0.017]			
Zone 8 (Dartford)	0.842	0.003	0.012**	0.016**						
		[0.175]	[0.012]	[0.044]						
Zone 8 (Cheshunt)	0.614	0.005	0.007	0.009	0.010	0.009	0.015			
		[0.272]	[0.184]	[0.184]	[0.211]	[0.272]	[0.184]			
Zone A	0.825	0.006	0.005	0.010	0.012*	0.014*	0.013	0.021**	0.025**	0.026**
		[0.181]	[0.240]	[0.107]	[0.094]	[0.070]	[0.129]	[0.047]	[0.037]	[0.029]
Zone B (Rye House – Hertford East)	0.551	0.002	0.008	0.005						
		[0.320]	[0.156]	[0.227]						
Zone B (Brentwood & Broxbourne)	0.959	0.002	0.002	0.003	0.001	0.003	0.006			
		[0.259]	[0.256]	[0.253]	[0.701]	[0.310]	[0.138]			
Zone C	0.947	-0.007***	-0.004**	-0.001	0.002	0.004	0.009**			
		[0.009]	[0.061]	[0.298]	[0.307]	[0.132]	[0.018]			

*Table 3.3. Evolution of passenger numbers for each zone comparative to the respective synthetic control zone. The table shows the difference in log passenger numbers between the treated stations/zones and its synthetic control for each year after the extension was implemented. The p-values (in parentheses) are standardised and based on placebo tests. *, **, and *** indicate significance at the 10%, 5%, and 1% level.*

.

From Table 3.4, there is very little to suggest the increase in passenger numbers in Zone 7 and Zone 8 (Dartford) is driven by off-peak travel, as the percentage change in ticket prices resultant of Oyster expansion is not higher in these stations comparative to others, in which we do not identify a significant increase. There is, however, suggestive evidence from Table 3.5 that the increases in passenger numbers identified are likely driven by peak return travellers (i.e. commuters). In Column (2) of Table 3.5, by far the largest price changes are for Zone 7 and Zone 8 (Dartford); this remains true in Columns (3) and (4) (peak return passengers who go on to travel elsewhere in London), though by a smaller margin. Our findings are consistent with the findings of Oxera (2005), who find demand for rail travel between London and the South East to be elastic in the short, medium, and long run; we cannot, however, compute precise elasticities for comparison without more granular data on journeys and prices.

A potential concern arises from spillover effects from stations treated by expansions to stations further down the same line: in particular, positive results could conceivably be driven by passengers travelling from further afield by means other than rail, to begin their rail journey at a treated station. We therefore propose a robustness check, whereby we utilise the same methodology as above to test for a treatment effect at untreated stations further down the same railway line. The placebo 'treated' stations for this purpose are Harlow Mill, Harlow Town, Roydon, and Sawbridgeworth; these are the four stations immediately outside of the LTZ on the same Greater Anglia service in which we identify a large and significant treatment effect for Zone 7. Results of this placebo analysis are presented in Table 3.6; whilst coefficients for Year 2 onwards are negative, coefficients are very close to zero, and none are statistically significant. We therefore conclude that the additional journeys for the LTZ extensions identified in the analysis above are not displaced passengers from elsewhere.

Zone	Off Peak One Way			Off Peak Return			Off Peak Return, plus Zone 1 return		Off Peak Return, plus Zone 1 Travelcard		
	Paper	Oyster	Percentage Difference	Paper	Oyster	Percentage Difference	Paper	Percentage Difference	Paper	Oyster Cap	Percentage Difference
Zone 7	£7.50	£4.00	-46.67%	£11.90	£8.00	-32.77%	£16.70	-52.10%	£18.90	£12.90	-31.75%
Zone 8 (Dartford)	£9.10	£4.40	-51.65%	£10.00	£8.80	-12.00%	£14.80	-40.54%	£17.00	£12.90	-24.12%
Zone 8 (Cheshunt)	£8.00	£4.00	-50.00%	£12.70	£8.00	-37.01%	£17.50	-54.29%	£19.70	£12.90	-34.52%
Zone A	£8.20	£5.20	-36.59%	£11.60	£10.40	-10.34%	£16.40	-36.59%	£18.60	£18.40	-1.08%
Zone B	£10.90	£5.90	-45.87%	£12.70	£11.80	-7.09%	£17.50	-32.57%	£19.70	£18.80	-4.57%
Zone C	£12.30	£6.70	-45.53%	£17.90	£13.40	-25.14%	£22.70	-40.97%	£24.90	£20.50	-17.67%

Table 3.4. Off Peak ticket prices when using a paper ticket or the oyster system to travel to the London Terminal stations in July 2019.

Zone	Peak One Way			Peak Return			Peak Return, plus Zone 1 return		Peak Return, plus Zone 1 Travelcard		
	Paper	Oyster	Percentage Difference	Paper	Oyster	Percentage Difference	Paper	Percentage Difference	Paper	Oyster Cap	Percentage Difference
Zone 7	£7.50	£5.60	-25.33%	£14.10	£11.20	-20.57%	£18.90	-40.74%	£21.10	£14.00	-33.65%
Zone 8 (Dartford)	£9.10	£7.60	-16.48%	£17.60	£15.20	-13.64%	£22.40	-32.14%	£24.60	£16.50	-32.93%
Zone 8 (Cheshunt)	£8.00	£6.60	-17.50%	£14.00	£13.20	-5.71%	£18.80	-29.79%	£21.00	£16.50	-21.43%
Zone A	£8.20	£6.20	-24.39%	£12.40	£12.40	0.00%	£17.20	-27.91%	£19.40	£24.60	26.80%
Zone B	£10.90	£8.30	-23.85%	£13.10	£16.60	26.72%	£17.90	-7.26%	£20.10	£24.60	22.39%
Zone C	£12.30	£9.60	-21.95%	£20.80	£19.20	-7.69%	£25.60	-25.00%	£27.80	£30.50	9.71%

Table 3.5. Peak ticket prices when using a paper ticket or the oyster system to travel to the London Terminal stations in July 2019.

	Pre-Treatment Diagnostic P-Value	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6
Placebo Stations	0.337	0.000 [0.847]	-0.001 [0.905]	-0.003 [0.639]	-0.005 [0.549]	-0.003 [0.651]	-0.006 [0.508]

Table 3.6: Placebo check, specifying Harlow Mill, Harlow Town, Roydon, and Sawbridgeworth, which did not receive treatment, as 'treated' stations in 2012/13.

Given the observed increase in commuting suggested by these results, we now go on to investigate whether there is any evidence that other economic outcomes have changed using the same synthetic control analysis that may, arguably, attribute any observed change to increased passenger flows.

3.5.2. House Prices

House Price results are presented in Table 3.7, and Figure 3.7; we do not find any evidence that inclusion in the London Travel Zone system has any effect on local house prices. In Year 3 for Zone 7 we identify a small negative coefficient of -0.011 log points, (just) significant at the 10% level. Graphically, there does appear to be some positive deviation from trend post-treatment in Zone 8 (Dartford), reaching 0.021 log points in the 3rd year post-treatment, but only significant in that one year at the 10% level. The Dartford result is also replicated in our robustness check (presented in Appendix A). Elsewhere, coefficients are both positive and negative, and not statistically significant, suggestive that there is no causal effect. It should be noted that in the main specification Zones A and C fail the pre-treatment diagnostic check, however Zone C passes diagnostics in the robustness specification and results are broadly unchanged.

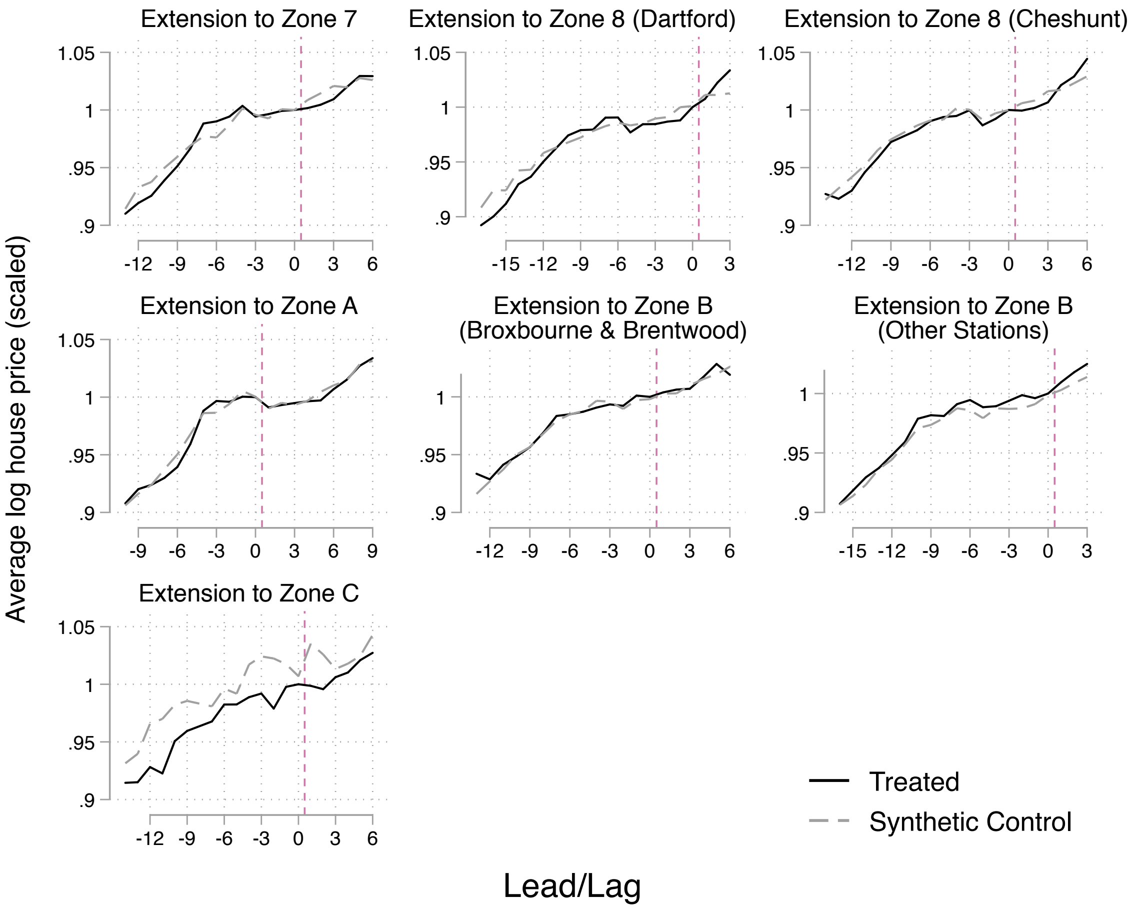


Figure 3.7. Evolution of average house prices for each zone comparative to the respective synthetic control zone. The vertical dashed line indicates the time the extension was implemented. The Lead/Lag indicates the years after/prior to the extensions and a lead of 1 is the first year after the extension.

	Pre-Treatment Diagnostic P-Value	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Year 8	Year 9
Zone 7	0.592	-0.007	-0.009	-0.011*	0.000	0.002	0.003			
		[0.427]	[0.168]	[0.093]	[0.986]	[0.830]	[0.708]			
Zone 8 (Dartford)	0.842	-0.003	0.011	0.021*						
		[0.699]	[0.265]	[0.071]						
Zone 8 (Cheshunt)	0.796	-0.006	-0.006	-0.010	0.004	0.006	0.015			
		[0.496]	[0.336]	[0.159]	[0.496]	[0.398]	[0.097]			
Zone A	0.097*	0.001	-0.002	0.002	0.000	-0.007	-0.003	0.001	-0.001	0.003
		[0.663]	[0.775]	[0.565]	[0.621]	[0.926]	[0.909]	[0.785]	[0.961]	[0.701]
Zone B (Rye House – Hertford East)	0.478	-0.002	0.003	-0.002	0.005	0.013	-0.011			
		[0.894]	[0.779]	[0.805]	[0.566]	[0.248]	[0.381]			
Zone B (Brentwood & Broxbourne)	0.538	0.001	0.003	-0.003	0.001	0.009	-0.007			
		[0.910]	[0.666]	[0.676]	[0.801]	[0.285]	[0.443]			
Zone C	0.009***	-0.037	-0.030	-0.007	-0.008	-0.004	-0.016			
		[0.265]	[0.265]	[0.761]	[0.726]	[0.912]	[0.584]			

Table 3.7. Coefficients in Log Points. Standardised P-Values in Parentheses.

3.5.3. Unemployment

As a clear effect on passenger numbers is identified for Zone 7, and suggestive evidence suggests the additional passengers are likely making return journeys at peak times (i.e. they are commuters), we also investigate the impact on unemployment in CAS Wards affected by the Zone 7 expansion[8]; results are presented in Table 3.8 and Figure 3.8. The pre-treatment diagnostic P-value is very high at 0.895, however no significant effect on the local claimant count is identified (coefficients are small and positive in Years 1 and 2 post-expansion, and small and negative in Year 3; none are statistically significant). This is suggestive that including an existing rail link in a London Travel Zone does not significantly improve the matching function between employers and workers.

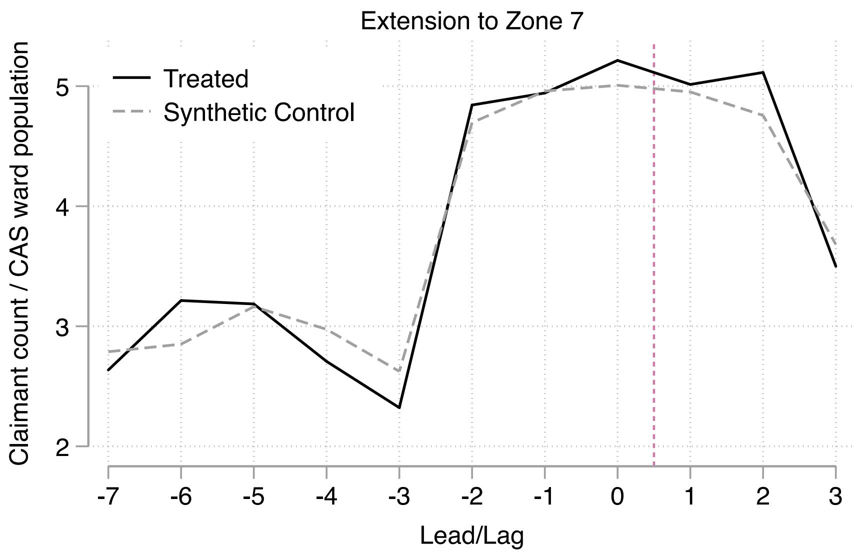


Figure 3.8. Evolution of jobseeker's allowance claimant count as percentage of the working population for zone 7 comparative to the respective synthetic control zone. The vertical dashed line indicates the time the extension was implemented. The Lead/Lag indicates the years after/prior to the extensions and a lead of 1 is the first year after the extension (2012).

[8] This would also logically be true of Zone 8 (Dartford), however the NOMIS Claimant Count data ends in 2014, prior to the Dartford expansion.

Coefficients are of the same polarity and larger but remain insignificant in our robustness specification presented in Appendix A, however the pre-treatment diagnostic P-value is significant is that specification.

Zone	Pre-Treatment Diagnostic P-Value	Year 1	Year 2	Year 3
Zone 7	0.895	0.062	0.358	-0.181
		[0.942]	[0.345]	[0.344]

Table 3.8. Standardised P-Values in Parentheses.

3.5.4. Traffic

As we find no evidence that the Zone 7 expansion decreased unemployment, we test the hypothesis that the additional Zone 7 passengers are displaced motorists, (i.e. commuters who would have otherwise used the car). We therefore compare annual traffic lows within a three mile[9] radius of treated stations to those stations' synthetic counterfactuals; we present the results of this analysis in Table 3.9 and Figure 3.9. Consistent with Paulley et. al. (2006), we find no strong positive relationship between rail pricing and car usage, identifying instead a small, temporary, increase in car (and all vehicle) usage in Waltham Cross in Years 2, 3, and 4 post-expansion, and a positive (but not significant) impact at Theobalds Grove over the same period. Once again, we find similar results in our robustness check in Appendix A: coefficients remain small and positive for both stations over the same years, but are significant for Theobalds Grove as opposed to Waltham Cross.

[9] The three mile radius was chosen due to the limitations of the AADF traffic flow data: any smaller radius and we have insufficient overlapping support; a higher radius decreases the feasibility that it is a policy impact we identify. We offer this in conjunction with the pollution analysis below, which are together supportive of our conclusions.

Station	Count	Pre-Treatment Diagnostic P-Value	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6
Waltham Cross	**All Traffic (3 mile radius)**	0.936	-0.003	0.005*	0.006**	0.009**	0.002	0.000
			[0.213]	[0.053]	[0.031]	[0.011]	[0.362]	[0.936]
	Cars Only (3 mile radius)	0.926	-0.003	0.005*	0.005**	0.007**	0.003	0.001
			[0.213]	[0.064]	[0.032]	[0.032]	[0.309]	[0.723]
Theobalds Grove	**All Traffic (3 mile radius)**	0.351	0.000	0.009	0.010	0.014	0.005	0.001
			[1.000]	[0.181]	[0.191]	[0.106]	[0.510]	[0.809]
	Cars Only (3 mile radius)	0.362	-0.001	0.008	0.009	0.012	0.006	0.003
			[0.702]	[0.213]	[0.160]	[0.138]	[0.351]	[0.660]

Table 3.9. Coefficients in Log Points. Standardised P-Values in Parentheses

The reasoning behind the small increase in traffic flows is unclear. It is of course feasible that these results are is driven by vehicles dropping passengers off at the station itself. Our test for spillovers to stations further down the line in Passenger Numbers analysis above rules out the possibility that these are passengers who are now driving for part of their journey to avoid having to purchase a paper ticket for that part of their journey. It is also the case that at a similar time, the nearby M25 motorway was undergoing major improvement works (Highways England, 2015), forcing traffic onto local roads; this is perhaps a more satisfactory explanation, in view of the identified increase in car usage being only temporary for both stations.

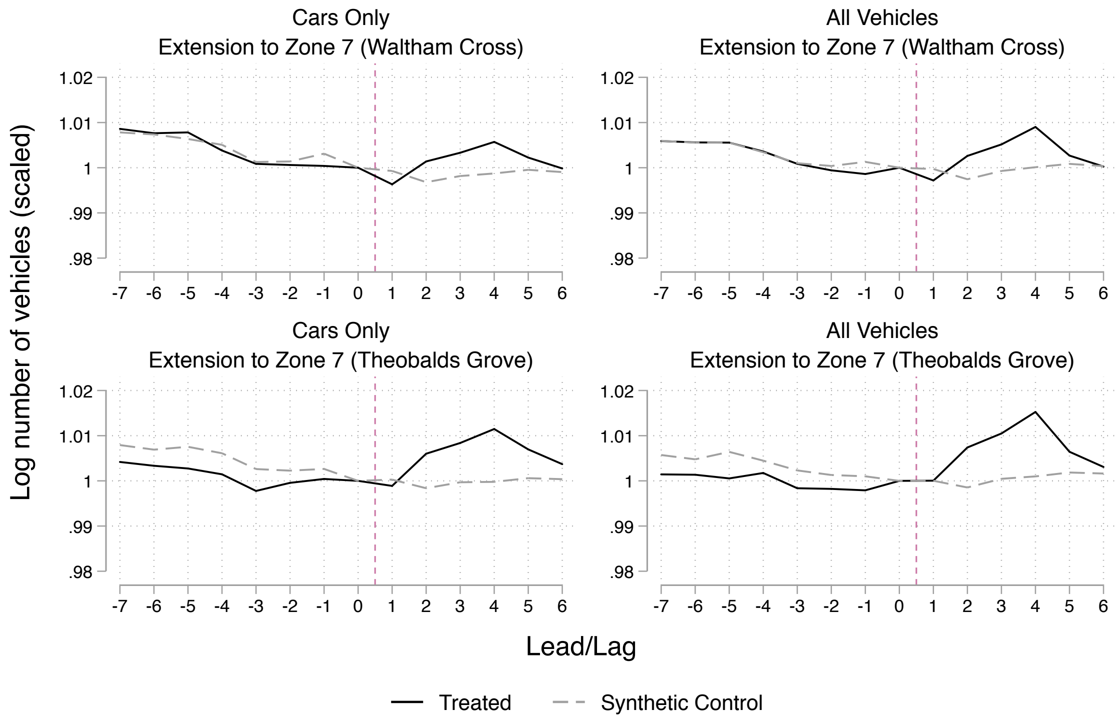


Figure 3.9. Evolution of the number of cars (left figures) and all vehicles (right figures) in a three miles radius around the stations for zone 7 comparative to the respective synthetic control zone. The vertical dashed line indicates the time the extension was implemented. The Lead/Lag indicates the years after/prior to the extensions and a lead of 1 is the first year after the extension (2012).

3.5.5. Pollution

Pollution results are presented in Table 3.10, and in Figures 3.10 and 3.11 for Zones 7 and 8 (Dartford) respectively. Pre-treatment diagnostic P-values are insignificant in all cases with the exception of $PM_{2.5}$ for Zone 7. Excluding $PM_{2.5}$, coefficients on all measured pollutants are negative in all years for Zone 7, and are not significant. For Zone 8 (Dartford), coefficients on Nitrogen Dioxide and Nitrogen Oxide are similarly negative and insignificant for all years. Small positive impacts are identified for $PM_{2.5}$ and PM_{10} measures of particulate matter in Year 3; these are not replicated in the robustness check presented in Appendix A (in which the pre-treatment diagnostic P-value is higher for these measures only), and as such should be treated as idiosyncratic.

This is consistent with our findings in the previous section, and the literature, in that no significant decrease in traffic flows can be attributed to the policy.

Zone	Pollutant	Pre-Treatment Diagnostic P-Value	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6
Zone 7	NO_2	0.278	-0.519	-0.632	-0.706	-0.406	-1.119	-0.303
			[0.595]	[0.644]	[0.627]	[0.729]	[0.487]	[0.711]
	NO_X	0.185	-2.519	-3.313	-3.834	-2.732	-5.170	-2.619
			[0.351]	[0.227]	[0.180]	[0.188]	[0.157]	[0.211]
	PM_{10}	0.138	-0.006	-0.153	0.007	-0.375	-0.600	-0.217
			[0.832]	[0.791]	[0.927]	[0.596]	[0.229]	[0.700]
	$PM_{2.5}$	0.028**	0.149	0.031	0.150	-0.211	-0.326	0.124
			[0.800]	[0.984]	[0.828]	[0.773]	[0.648]	[0.722]
Zone 8 (Dartford)	NO_2	0.105	-1.993	-2.944	-1.424			
			[0.184]	[0.228]	[0.491]			
	NO_X	0.105	-3.276	-6.408	-2.544			
			[0.263]	[0.236]	[0.552]			
	PM_{10}	0.614	0.219	0.849*	0.826**			
			[0.737]	[0.053]	[0.044]			
	$PM_{2.5}$	0.518	0.343	0.656	0.606**			
			[0.500]	[0.158]	[0.035]			

Table 3.10. Standardised P-Values in Parentheses

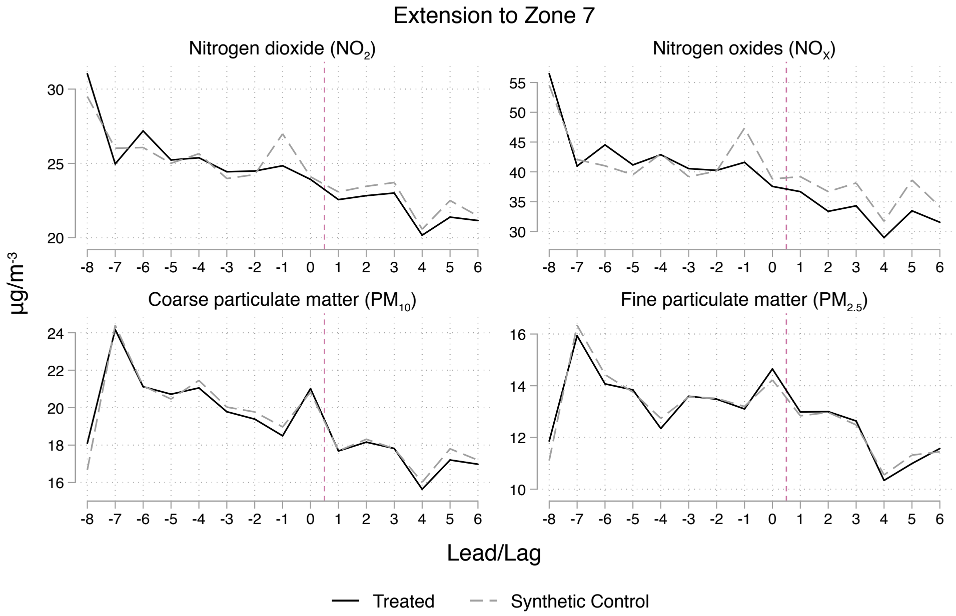


Figure 3.10. Evolution of air quality measures for zone 7 comparative to the respective synthetic control zone. The vertical dashed line indicates the time the extension was implemented. The Lead/Lag indicates the years after/prior to the extensions and a lead of 1 is the first year after the extension (2012).

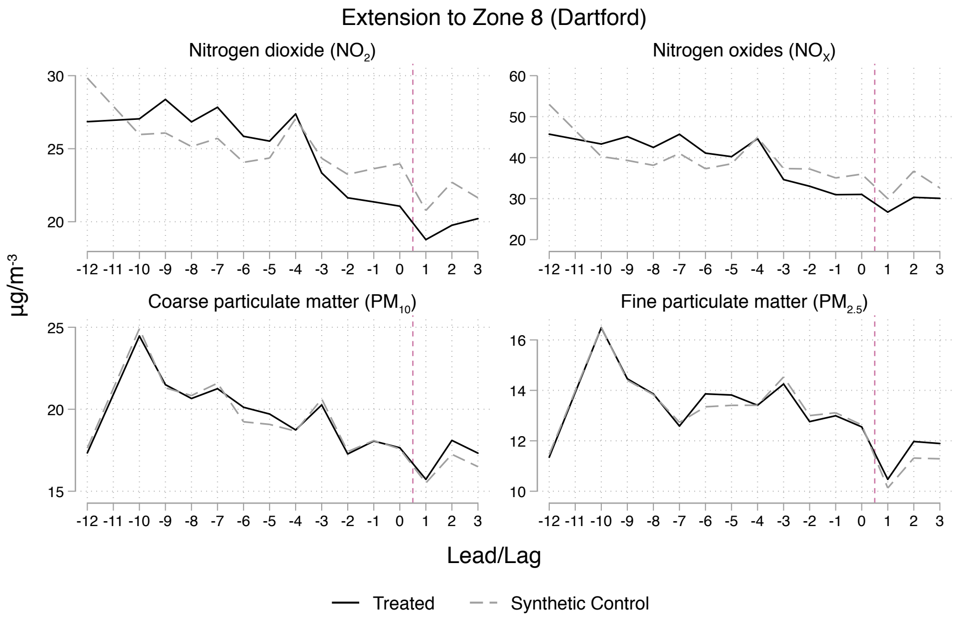


Figure 3.11. Evolution of air quality measures for zone 8 (Dartford) comparative to the respective synthetic control zone. The vertical dashed line indicates the time the extension was implemented. The Lead/Lag indicates the years after/prior to the extensions and a lead of 1 is the first year after the extension (2015).

3.6 Conclusion

Our findings suggest that PAYG rail ticketing positively impacts passenger demand only where it represents a meaningful reduction in pricing; suggestive evidence points to this effect being driven primarily by commuters. We do not find evidence that prior LTZ expansions have had any significant wider economic impacts on house prices, unemployment, traffic, nor pollution.

We utilise the Synthetic Control Method for its data-driven approach to the selection of control units against which to compare treated stations. The same set of predictor variables are used for the estimation of impacts on all outcome variables to avoid specification searching; results without predictor variables are reported for robustness, and do not suggest the choice of predictor variables meaningfully affects our findings.

The ORR Passenger Numbers data utilised by this study is limiting in two ways. First, changes in the methodology used to calculate passenger numbers in 2006/07, 2007/08, and 2008/09 financial years limits the useable pre-treatment period for the purposes of our estimation. Second, passenger entries and exits to and from a station are aggregated annually at the station level, meaning we do not see origin/destinations at the journey level. An extension of this study using the Rail Delivery Group's more granular LENNON data would therefore (a) increase the statistical power of results by extending the pre-treatment period by several years, and (b) allow for more detailed inference on the types of journeys, and characteristics of passengers, who are most likely to change their behaviour as a result of a policy change of similar nature. This, in turn, might allow for the estimation of ATT estimates of economic impacts on those affected by the policy.

3.7 Appendix A

3.8.1. Passenger Numbers

	Pre-Treatment Diagnostic P-Value	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Year 8	Year 9
Zone 7	0.903	0.010*	0.024**	0.026**	0.035**	0.047**	0.069**			
		[0.080]	[0.019]	[0.018]	[0.018]	[0.017]	[0.017]			
Zone 8 (Dartford)	0.816	0.001	0.011*	0.018*						
		[0.693]	[0.096]	[0.070]						
Zone 8 (Cheshunt)	0.518	-0.002	0.000	0.003	0.003	0.005	0.013			
		[0.675]	[1.000]	[0.693]	[0.781]	[0.605]	[0.316]			
Zone A	0.443	-0.007	-0.006	-0.004	-0.003	-0.002	0.002	0.008	0.017	0.018
		[0.284]	[0.351]	[0.531]	[0.702]	[0.733]	[0.944]	[0.596]	[0.271]	[0.262]
Zone B (Rye House – Hertford East)	0.752	0.001	0.007	0.003						
		[0.967]	[0.305]	[0.468]						
Zone B (Brentwood & Broxbourne)	0.974	0.008*	0.008*	0.009*	0.006	0.010	0.011*			
		[0.064]	[0.062]	[0.080]	[0.155]	[0.114]	[0.069]			
Zone C	0.702	0.005	0.007	0.009	0.010	0.011	0.012			
		[0.289]	[0.167]	[0.202]	[0.219]	[0.184]	[0.193]			

Table A 1. Coefficients in Log Points. Standardised P-Values in Parentheses.

3.8.2 House Prices

	Pre-Treatment Diagnostic P-Value	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Year 8	Year 9
Zone 7	0.168	-0.002	0.003	0.003	0.008	0.020	0.009			
		[0.852]	[0.747]	[0.770]	[0.451]	[0.131]	[0.551]			
Zone 8 (Dartford)	0.478	0.005	0.016	0.022*						
		[0.549]	[0.177]	[0.097]						
Zone 8 (Cheshunt)	0.628	-0.016	-0.013	-0.014*	-0004	0.003	0.010			
		[0.115]	[0.159]	[0.097]	[0.664]	[0.858]	[0.398]			
Zone A	0.024**	-0.010	-0.015	-0.014	-0.016	-0.017	-0.015	-0.012	-0.005	-0.004
		[0.481]	[0.309]	[0.187]	[0.305]	[0.280]	[0.388]	[0.445]	[0.842]	[0.951]
Zone B (Rye House – Hertford East)	0.759	0.004	0.006	0.008						
		[0.199]	[0.006]	[0.167]						
Zone B (Brentwood & Broxbourne)	0.380	0.001	0.001	-0.006	-0.001	0.005	-0.010			
		[0.955]	[0.878]	[0.537]	[0.917]	[0.662]	[0.398]			
Zone C	0.204	0.000	-0.005	-0.003	-0.004	0.000	0.001			
		[1.000]	[0.681]	[0.867]	[0.823]	[1.000]	[0.982]			

Table A 2. Coefficients in Log Points. Standardised P-Values in Parentheses.

3.8.3 Unemployment

Zone	Pre-Treatment Diagnostic P-Value	Year 1	Year 2	Year 3
Zone 7	0.065*	0.100	0.296	-0.244
		[0.853]	[0.453]	[0.201]

Table A 3. Standardised P-Values in Parentheses.

3.8.4 Traffic

Station	Count	Pre-Treatment Diagnostic P-Value	Year 1	Year 2	Year 3	Year4	Year5	Year6
Waltham Cross	**All Traffic**	0.617	-0.003	0.004	0.006	0.009*	0.001	-0.001
	(3 mile radius)		[0.404]	[0.255]	[0.149]	[0.096]	[0.830]	[0.830]
	Cars Only	0.681	-0.004	0.003	0.005	0.007	0.002	0.001
	(3 mile radius)		[0.351]	[0.330]	[0.191]	[0.138]	[0.511]	[0.840]
Theobalds Grove	**All Traffic**	0.191	0.019*	0.028**	0.030*	0.034**	0.024*	0.021
	(3 mile radius)		[0.074]	[0.042]	[0.053]	[0.043]	[0.096]	[0.160]
	Cars Only	0.223	0.018*	0.027**	0.029**	0.032**	0.026*	0.024
	(3 mile radius)		[0.064]	[0.043]	[0.021]	[0.043]	[0.053]	[0.128]

Table A 4. Coefficients in Log Points. Standardised P-Values in Parentheses.

3.8.5 Pollution

Zone	Pollutant	Pre-Treatment Diagnostic P-Value	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6
Zone 7	**NO2**	0.332	-0.766	-1.236	-1.269	-0.803	-1.523	-0.526
			[0.406]	[0.343]	[0.455]	[0.475]	[0.318]	[0.620]
	NOX	0.234	-2.497	-3.231	-3.400	-2.264	-4.277	-0.039
			[0.312]	[0.231]	[0.269]	[0.281]	[0.256]	[0.987]
	PM10	0.175	-0.546	-0.597	-0.368	-0.493	-0.244	-0.087
			[0.240]	[0.115]	[0.558]	[0.494]	[0.705]	[0.929]
	PM25	0.042**	-0.522	-0.525	-0.408	-0.201	-0.042	-0.767
			[0.204]	[0.165]	[0.459]	[0.863]	[0.937]	[0.704]
Zone 8 (Dartford)	**NO2**	0.096*	-2.120	-3.091	-0.502			
			[0.246]	[0.289]	[0.842]			
	NOX	0.105	-2.988	-5.704	-0.355			
			[0.351]	[0.368]	[0.930]			
	PM10	0.316	-0.392	0.564	0.227			
			[0.561]	[0.395]	[0.737]			
	PM25	0.105	0.009	0.858	0.419			
			[1.000]	[0.316]	[0.254]			

Table A 5. Standardised P-Values in Parentheses.

4. Risky Incentives in Labour Contracts: An Experiment

With Celia Blanco Jimenez*

*Department of Economics, University of Lancaster

4.1 Introduction

It is not uncommon for lotteries to be used as incentive mechanisms for participation in tasks; participation in online surveys is often incentivised at the extensive margin by entry to a prize draw to win an iPad, or a large face value in vouchers, for example. If would-be participants were offered the cash equivalent expected value of such lotteries for their time, it is unlikely they would participate: although we know that the expected value is infinitesimally small, we tend to neglect the cost of our effort at the prospect of winning a valuable prize Rachlin (1990).

But real life examples of such lottery incentive mechanisms are less easy to think of at the intensive margin; that is, lottery mechanisms being used to influence how much, or how hard, workers choose to work. It is not uncommon for workers in particular industries to be offered bonuses in share options, however whilst these involve risk they are not a lottery. The idea behind this paper was borne out of anecdotal evidence of a sales manager who believed his workers responded better to bonuses paid in scratchcards than the equivalent cash face value; was he right? In particular, can lotteries be used as an incentive mechanism to induce high effort (and thus productivity)? There is an extensive literature within personnel economics on mechanisms to induce optimal effort for the firm, which we summarise below (Lazear 1986, and Freeman, 1998 among others), though surprisingly little research exists on mechanisms through which lotteries might incentivise effort.

One of the most utilised theoretical models in the personnel economics literature is the agency theory model, which assumes that a worker (agent) experiences some disutility of effort, while the firm (principal) sets a wage that seeks to incentivise an optimal effort level such that profit is maximised (Ross, 1973; Stiglitz, 1989). If both parties' incentives are not aligned, this can result in the agent selecting to exert suboptimal effort, and thus the firm obtains suboptimal profits. Many solutions to this problem have been proposed, however it is often found to be prohibitively expensive for the firm to incentivise effort over and above the reservation level in equilibrium, as the cost to the firm of doing so exceeds the increase in revenue.

Various mechanisms have been proposed to optimally incentivise high effort in a manner that is profitable for the firm, most notably efficiency wages and piece rates (Prendergast, 1999; Ariely et al., 2009; Lazear, 2000; Lazear, 1986). However such mechanisms have been found to be empirically problematic for two main reasons. Firstly, measuring individual worker productivity is not always practically possible, or can be costly to the firm where it is (Freeman, 1998). Secondly, because effort is imperfectly correlated with productivity, and it is productivity that is ultimately incentivised by the firm, the worker takes on some risk when exerting additional effort; to incentivise high effort the firm must therefore compensate both the worker's disutility of effort and the risk the workers takes on (Pendergast, 2002). The costs to the firm of incentivising high effort are therefore often prohibitive when weighed against the productivity gains.

There is some suggestive evidence in the literature that workers' effort response may be higher where productivity incentives involve risk (Bandiera et al., 2005; Rula et al., 2014; Celis et al., 2013). We contribute by reviewing this evidence, and testing it in a controlled laboratory experiment. For that, we propose an experiment in which subjects are randomly allocated to three treatments, in each of which subjects receive a fixed payment for achieving a minimum standard of effort, plus a bonus for productivity over and above that minimum standard. In the first treatment this takes the form of a piece rate bonus, in which subjects are paid a fixed value for each unit of productivity over and above the minimum threshold. In the second treatment, the bonus is paid in the form of an individual lottery, in which subjects add a value five times higher than the piece rate to their bonus pot with each additional unit of productivity, but only receive the bonus with 20% probability (such that the expected value is the same as the piece rate). The third treatment takes the form of a group lottery bonus, where each member of a group receives a lottery ticket with the same expected value as the piece rate bonus for productivity over and above the threshold, and a winning ticket is randomly drawn such that one subject wins the entire group bonus pot. Each treatment is played across five rounds (with bonuses allocated each round), and subjects can observe other group members' productivity. The experiment utilises a real effort task to ensure a strong positive relationship between effort and productivity

We find that (i) lottery incentives are at least as effective as piece rates in incentivising effort; (ii) risk aversion does not seem to drive higher effort in the lottery settings (iii) other group members' performance in the previous period is negatively related to performance in the group lottery treatment, but positively correlated for individuals with prosocial preferences.

The rest of the paper is organised as follows. In Section 2 we review the literature on effort incentive mechanisms; Section 3 sets out the experimental design; Section 4. presents the results of the experiment; and Section 4 offers a discussion of the results, and the behavioural factors driving the overall effect.

4.2 Literature Review

According to contract theory, exerting high levels of effort is fatiguing to the worker, such that effort is costly in utility terms; the firm's output is, however, is positively related to effort. The incentives of employees and employers are clearly therefore not necessarily aligned, and so firms must design a compensation scheme that aligns these interests. This is the essence of principal agency theory.

Various compensation mechanisms exist in the personnel economics literature, which attempt to align these interests. Mechanisms such as delayed compensation and efficiency wages increase the value of retaining the job to the worker, while others such as piece rates and tournaments link compensation to absolute or relative performance. Incentives that directly link compensation to productivity are referred to as Pay for Performance, and there exists a substantial literature on the merits and demerits of such mechanisms (see Prendergast, 1999; Ariely et al., 2009; Lazear, 2000; Lazear, 1986; Gibbons, 1987).

There are a significant number of empirical studies on the effect of piece rate Pay for Performance mechanisms on worker effort. The main conditions for piece rates to be beneficial are low monitoring costs, accurate output measures, and heterogeneous ability amongst workers (Lazear, 1986). Freeman (1998) finds that a US shoe manufacturing firm, which switched from a piece rate payments to a rate per hour, increased profits by eliminating monitoring costs. He concludes that, when monitoring costs are high, piece rates may not always be the optimal payment method even if they do increase worker productivity. Conversely, Lazear (2000) analysed field data from an automotive glass installer which switched from an hourly payment regime to piece rate payment, finding a 44% increase in productivity under piece rates comparative to rate per hour (although half of this gain was found to be due to worker sorting). In this circumstance both wages and profit increased. This efficiency gain was largely due to a sophisticated computerised system that allowed the firm to directly measure productivity. Similarly, Paarsch (1996) finds that tree planters in British Columbia are more productive if paid piece rates comparative to fixed wages, although in this case all efficiency gains are attributed to increased effort, as there is no sorting effect. In a more recent experiment, Bandiera et al. (2005) compare a piece rate payment with a relative payment, where workers are paid according to their output comparative to the mean output that day; the authors find that piece rates lead to higher productivity.

A stream of the literature examines the impact on performance pay on educational outcomes. Lavy (2002) examines the effect of performance-based reward structure in a non-random sample of schools in Israel. He finds that providing both students and teachers with monetary incentives has a positive impact on student performance; positive results are also found where only teachers are incentivised, which proves more cost effective. Similarly, Muralidharan & Sundararaman (2009) find a positive and significant impact of providing performance-based incentives to teachers on student performance. Fryer et al. (2012) examine the effect of loss aversion on teachers, by paying teachers an advance for achieving a target student performance, and asking them to return the money if this is not fulfilled. The authors identify an increased students' performance on a standardised maths test.

In general, the literature concludes that Pay for Performance mechanisms lead to increased productivity in most circumstances. This result can, however, be reversed where incentive payments are set at a sufficiently high rate such that workers can achieve their target wage with minimal effort (Ariely et al., 2009). However, Bandiera, Barankay, & Rasul (2005) find no evidence of income targeting.

Although the literature shows that in most cases Pay for Performance mechanisms increase worker productivity, surprisingly little attention has been paid to pay per performance mechanisms that involve risk. This is perhaps surprising; as Zábojník (2002) notes, this has direct relevance to industries in which bonuses are paid in share options. Further, recent research shows some suggestive evidence that such contracts might deliver greater efficiency gains than piece rates.

Zábojník (2002) shows that if the assumption of global risk-averseness on the part of workers is dropped, and a Friedman-Savage quasi-convex utility function is instead assumed with respect to wages, the first best contract achieving full efficiency in a principal-agent setup is achieved using lottery payments. It is shown that in a locally convex section of the worker's utility function there exists a lottery payment with an expected wage lower than a fixed wage achieving the same worker utility.

Brune (2015) builds on this hypothesis by comparing the effect of a lottery bonus to a piece rate bonus with the same expected value on productivity and attendance of workers at a large agricultural firm in Malawi. The study finds a statistically significant increase in labour supply by workers at the intensive margin twice as large as that for piece rates. A small and marginally significant (at 10% level) increase in productivity is identified for the lottery bonus, though the piece rate bonus was not statistically different from the baseline. However notably this result could conceivably be dampened by worker fatigue: because workers are working more at the intensive margin, the effects of fatigue could conceivably reduce the treatment effect with respect to productivity (e.g. see Schor, 1991). There is clearly therefore scope for an experiment in which a productivity outcome alone is incentivised.

Levitt et al. (2016) examines the effect of financial incentives on student performance, comparing a fixed conditional transfer payment to a lottery payment, and with varied treatments in which both students and parents are incentivised on the child's exam performance. The authors find that a lottery payment with parents as recipients was most effective, whilst a small but significant effect for a fixed payment with students as the recipients is also identified.

Another stream of literature focuses on lottery payments for micro tasks. These are the papers most related to ours, although we will contribute with a real effort task in a lab setting. Rula et al. (2014) compare outcomes in effort and compliance in microtasks using piece rate micro-payments and a lottery, where payments were made in the form of coffee shop gift cards. They obtain that piece rates had higher compliance and user effort, while the lottery treatment achieved higher recruitment. However, there is the clear potential for these results to be driven by self-selection and heterogeneous preferences over the gift cards. Similarly, Rokicki et al. (2014) compare microtask speed and accuracy under competitive payments, piece rates, and lottery payments in Amazon Mechanical Turk. They obtained that an exponential piece rate payment outperforms a winner-takes-it all piece rate and lottery based payments in terms of accuracy, although the authors point out these results have potential to be driven by the relatively low value of rewards in the lottery treatment.

Celis et al. (2013) similarly compare a piece rate with a lottery payment for micro tasks in Amazon Mechanical Turk, where participants were tasked with the digitalisation of pieces of scrambled text. They find that lottery-based payments lead to more accurate digitalisation, and workers spent more time on these tasks. Furthermore, a third of participants reported to prefer the lottery payment comparative to the piece rate.

Our paper attempts to clearly test the effects of a lottery setting on workers productivity via a real effort task, and contribute to the literature on financial incentives on productivity, more concretely build on the aforementioned small stream of the literature on incentives involving risk.

4.3 Experimental Design

This experiment was conducted in the experimental lab at Royal Holloway, University of London between 29th November and 5th December 2017. Subjects were recruited from various undergraduate programmes at the university, including the Economics and Psychology programmes. Each session consisted of a single treatment, and lasted around one hour. Subjects were paid a show up fee of £4.

The task conducted is the Slider Task developed by Gill and Prowse (2012a), using zTree software (Fischbacher, 2007), whereby participants are tasked with moving a number of digital sliders to a particular position[10]. This task was chosen for most direct comparability of results with other real-effort tasks in the recent experimental literature (e.g. Gill & Prowse, 2012; Gill, Prowse and Vlassopoulos, 2013; Doerrenberg & Duncan, 2014; Gill and Prowse, 2014; Abeler & Jäger, 2015; Georganas, Tonin, & Vlassopoulos, 2015; Araujo et al, 2016; Buser & Dreber, 2016). The task was also chosen due to its advantages over other classic real effort

[10] See Appendix 1

tasks: it is simple to understand and communicate; unlike other tasks, the slider is exactly the same through repetitions; it involves little randomness and there is no scope for guessing. Also, there is a strong correlation between effort and productivity. The primary disadvantage of the slider task is that productivity is often found to be very tightly distributed, the lack of variation making significant results difficult to identify (Banuri & Keefer, 2015). This does, however, add power to significant results where they are identified.

At the beginning of each session, subjects were allowed into the lab in groups of five and each subject was randomly allocated to a workstation. Subjects were given around 15 minutes to read the task instructions fully and answer the control questions. Once finished, subjects' control questions were checked individually to ensure complete understanding of the task, and any doubts were clarified. A verbal summary of the instructions was then read to the entire group.

This real effort task consists of a screen with 51 sliders, all initially positioned at 0. Participants were tasked with dragging each slider to position 50 with their mouse; the slider could be adjusted to any position between 0 and 100, as many times as desired. The goal of the task is to correctly position as many sliders as possible at 50 within an allotted time of two minutes. Subjects could see a running total of correct sliders achieved in the current round, and the remaining time available. Five paid rounds were conducted for each treatment.

At the beginning of each treatment, subjects were assigned a subject number and randomly allocated into groups of five. Grouping subjects was necessary in order to identify peer effects. Each treatment consisted of 50 subjects.

All communication subjects received about their group members was provided in terms of subject numbers, thus keeping real identities anonymous. Two practice rounds were conducted prior to the paid rounds in order to familiarise subjects with the task; subjects' performance in these rounds did not affect the final payoff.

Our three treatments differed only in terms of the bonus payoff mechanism. In each, for achieving a minimum threshold of 15 correctly positioned sliders, subjects received a fixed payoff of 10 ECU (experimental currency units). The payoffs for each round were independent, and the subjects received feedback on their own performance and the performance of the other members of their group at the end of each round.

In the piece rate treatment, each correctly positioned slider over and above the minimum threshold of 15 earned the subject a 1 ECU bonus. Each ECU is equivalent to £0.30. The individual's payoff can be expressed as follows:

$$\Pi_{ir} = \begin{cases} 10 + 1 \cdot (N_{ir} - 15), & N_{ir} \geq 15 \\ 0, & N_{ir} < 15 \end{cases}$$

Where Π_{ir} is the payment for subject i in round r, and N_{ir} is the corresponding amount of correctly positioned sliders.

In the Individual lottery treatment, each correct slider over and above the minimum threshold added 5 ECU to the subject's bonus pool; at the end of each round the subject received the bonus pool with probability 1/5.

The individual's expected payoff can be described as follows:

$$E(\Pi_{ir}) = \begin{cases} 10 + [5 \cdot (N_{ir} - 15)] \cdot \frac{1}{5}, & N_{ir} \geq 15 \\ 0, & N_{ir} < 15 \end{cases}$$

$$\equiv \begin{cases} 10 + 1 \cdot (N_{ir} - 15), & N_{ir} \geq 15 \\ 0, & N_{ir} < 15 \end{cases}$$

In the group Lottery treatment, each correctly positioned slider over and above the minimum threshold earned the subject 1 lottery ticket with a face value of 1 ECU. At the end of each round, all lottery tickets in each group were added into a group bonus pool. The computer randomly selected one ticket in every group, and the winner earned the whole bonus pool. Hence, there was one winner of the entire bonus pool in each group per round. As before, earnings in one round did not affect the earnings in the following rounds.

The individual's expected payoff can be described as follows:

$$E(\Pi_{ir}) = \begin{cases} 10 + \left[\dfrac{(N_{ir} - 15)}{\sum_{i=1}^{5}(N_{ir} - 15)} \cdot 5 \cdot \sum_{i=1}^{5}(N_{ir} - 15)\right], & N_{ir} \geq 15 \\ 0, & N_{ir} < 15 \end{cases}$$

$$\equiv \begin{cases} 10 + 1 \cdot (N_{ir} - 15), & N_{ir} \geq 15 \\ 0, & N_{ir} < 15 \end{cases}$$

As is demonstrated above, each individual i's expected payoff Π for a given number of correctly positioned sliders in round r, $(E(\Pi_{ir}))$ is constant across treatments; behavioural differences across treatments are therefore interpretable as a treatment effect of the compensation mechanism alone.

There was a 35 second pause in between each round, during which time subjects received feedback on the number of sliders achieved by themselves and their group members. They also received their personal payoff information, and that of other members of their group (including the winner in the case of the group lottery).

At the end of each session, participants were asked to complete a short questionnaire, including measure of risk preferences as per Holt and Laury (2002), and social preferences as per Bartling et al. (2009)[11]. Choices in the questionnaire were not incentivised monetarily, and all subjects were paid in cash.

4.4 Theoretical predictions

We assume the subjects' utility is a separable function $U = u(w(x)) - c(x)$, where $u(\cdot)$ is an increasing continuous concave function, representing that the marginal utility is decreasing in wealth for the individual. $x \in \{0,1,\dots,102\}$ is the output produced by the agent, $w(.)$ is the compensation (or wage), which is increasing in x; $c(.)$ is the cost of the effort, which we assume to be increasing and convex in x.

In all treatments:

$$w(x) = \begin{cases} 0, & x < 15 \\ F + PR \cdot (x - 15), & x \geq 15 \end{cases}$$

Treatment 1

In our first incentive model subjects are paid a fixed bonus in the form of a piece rate for sliders correctly positioned over and above the minimum threshold. Since the piece rate value takes the value of 1 ECU in this treatment, we substitute this value into the wage function, which becomes the following:

$$w(x) = \begin{cases} 0, & x < 15 \\ F + (x - 15), & x \geq 15 \end{cases}$$

We can model the subjects' decision by the following problem:

$$Max_x U = u\big(w(x)\big) - c(x)$$

[11] See Appendix 2

Substituting the wage function:

$$Max_x U = u(x - 15) + u(F) - c(x)$$

If $u(F) < c(15)$, the solution is $x = 0$. Note that the subjects will never exert effort below 15, since the payment will be 0 and the cost of effort is positive.

If $u(x - 15) + u(F) \geq c(15)$ when $x \geq 15$, there exists a solution such that $x \geq 15$. Additionally, if there exists a value of x≥15 such that $u'(x - 15) + u(F) > c'(x)$, then there exists a solution such that $x > 15$. In such a case, the solution is given the following first order conditions:

$$u'(x^*) + u(F) - c'(x^*) = 0$$
$$u'(x^*) + u(F) = c'(x^*)$$

Where $x^*$is the equilibrium output.

Treatment 2

We propose a second setting where the piece rate bonus is higher but is only received with a fixed probability, such that the expected value of the bonus is the same as in the previous treatment. Here we can rule out group-effects from the next treatment. Since the piece rate value takes the value of 5 ECU in this treatment, we substitute this value into the wage function, which becomes characterised as follows:

$$w(x) = \begin{cases} 0, & x < 15 \\ F + 5 \cdot (x - 15), & x \geq 15 \end{cases}$$

We can model a subject's decision in to exert more effort as:

$$Max_x U = p \cdot u(w(x)) - c(x)\ ,$$

Substituting the values used in our experiment $p = \frac{1}{5}$ and $w = 1\ ECU$:

$$Max_x U = u(F) + \frac{1}{5} \cdot u\big(5 \cdot (x - 15)\big) - c(x)$$

If $u(F) < c(15)$, the solution is $x = 0.$ Note that, again, the subjects will never exert effort below 15, since the payment will be 0 and the cost of effort is positive.

If $\frac{1}{5} \cdot u(5 \cdot (x - 15)) + u(F) \geq c(15)$ when x≥ 15, there exists a solution such that $x \geq 15$. Additionally, if there exists a value of x≥15 such that $u'(5 \cdot x) + u(F) > c'(x)$, there exists a solution such that $x > 15$. In such a case, the solution has the following first order conditions:

$$\frac{1}{5} \cdot u'(5 \cdot x^{**}) + u(F) - c'(x^{**}) = 0$$

$$\frac{1}{5} \cdot u'(5 \cdot x^{**}) + u(F) = c'(x^{**})$$

Where $x^{**}$is the equilibrium output.

Treatment 3

The last setting is a winner-takes-it-all lottery, in which the total prize is the total pool of workers' accumulated bonuses, where the probability of achieving the prize is correlated to the rank, but even low performers have a positive chance of winning. Here, the expected value for every unit of output is the same as in the piece rate bonus. The performance of one subject can positively affect the total prize, while increasing her probability of winning at the same time.

Since the lottery ticket takes the value of 1 ECU in this treatment, we substitute its value and the wage function is the following:

$$w(x) = \begin{cases} 0, & x < 15 \\ F + p \cdot \left[\sum_{j \neq i}^{4} (x_j - 15) + (x_i - 15) \right], & x \geq 15 \end{cases}$$

We can model her decision in to exert more effort as:

$$Max_x U = u(F) + p \cdot u\big(x_i - 15 + \textstyle\sum_{j \neq i}^{4} x_j\,(x_j, x_i) - 15)\big) - c(x_i),$$

Here, with abuse of notation, we define $\sum_{j \neq i}^{4} x_j\,(x_j, x_i) - 15)$as the sum of each of the output of the 4 other players in the group, which simultaneously depends on the other player's output; x_i is the effort of player i, and $p = \frac{x_i}{\left[x_i + \sum_{j \neq i}^{4} x_j(x_j, x_i)\right]}$; hence substituting the wage function:

$$Max_x U = u(F) + \frac{x_i}{\left(x_i + \sum_{j \neq i}^{4} x_j(x_j, x_i)\right)} \cdot u\big(x_i - 15 + \textstyle\sum_{j \neq i}^{4} x_j\,(x_j, x_i) - 15)\big) - c(x_i)$$

If $u(F) < c(15)$, the solution is $x = 0$. Note that, again, the subjects will never exert an effort below 15, since the payment will be 0 and the cost of effort is positive. To simplify, we reduce $x_j(x_j, x_i)$ to x_j.

If $u(F) + \frac{x_i}{(x_i + \sum_{j \neq i}^{4} x_j)} \cdot u\big(x_i - 15 + \sum_{j \neq i}^{4} x_j - 15)\big) \geq c(x_i)$ being x$\geq$ 15, there exists a solution such that $x \geq 15$. Additionally, if there exists a value of x$\geq$15 such that $p'u'\big(x_i - 15 + \sum_{j \neq i}^{4} x_j - 15)\big) > c'(x)$, then there exists a solution such that $x > 15$.

Since we assume that all subjects have the same utility function, the equilibrium must necessarily be symmetric. Hence, $x_i = x_j \ \forall\ j \neq i$, and $x_i + \sum_{j\neq i}^{4} x_j = 5 \cdot x_i$.

The first order conditions are:

$$\frac{u(x_i^{***}+\sum_{j\neq i}^{4} x_j^{***})^2 + x_i^{***}\cdot u\prime(x_i^{***}+\sum_{j\neq i}^{4} x_j^{***})^2 - x_i^{*}\cdot u(x_i^{***}+\sum_{j\neq i}^{4} x_j^{***})\cdot(\sum_{j\neq i}^{4} x_j^{***}\cdot x_i^{***})\prime}{(x_i^{***}+\sum_{j\neq i}^{4} x_j^{***})^2} - c'(x_i^{***}) = 0$$

Simplifying:

$$\frac{u(x_i^{***}+\sum_{j\neq i}^{4} x_j^{***}) + x_i^{***}\cdot u\prime(x_i^{***}+\sum_{j\neq i}^{4} x_j^{***})}{(x_i^{***}+\sum_{j\neq i}^{4} x_j^{***})} - \frac{x_i^{***}\cdot u(x_i^{***}+\sum_{j\neq i}^{4} x_j^{***})\cdot(\sum_{j\neq i}^{4} x_j^{***\prime}+1)\prime}{(x_i^{***}+\sum_{j\neq i}^{4} x_j^{***})^2} - c'(x_i^{***}) = 0$$

Assuming the equilibrium is symmetric and all players maximize the same utility function, we substitute $x_i = x_j \ \forall\ j \neq i$, and $x_i + \sum_{j\neq i}^{4} x_j = 5 \cdot x_i$

$$\frac{u(5x_i^{***}) + x_i^{***}\cdot u\prime(5x_i^{***})}{5x_i^{***}} - \frac{x_i^{***}\cdot u(5x_i^{***})\cdot 5}{5x_i^{***2}} - c'(x_i^{***}) = 0$$

Simplifying:

$$\frac{1}{5}u'(5x_i^{***}) - c'(x_i^{***}) = 0.$$

This is the same result obtained in Treatment 2, hence $x_i^{**} = x_i^{***}$. If the equilibrium is symmetric, the equilibrium output in the Individual Lottery Treatment is equal to the equilibrium output in Group Lottery Treatment.

Now we will proceed to compare the equilibria obtained in each of the three treatments. We first analyse these results assuming that the subjects are risk neutral and $\mathrm{u}(x) = x$. We also assume the cost of effort to be a quadratic function $\mathrm{c(x)} = \mathrm{x}^2$, as the subject becomes tired over time and the cost of producing an additional unit increases.

Treatment 1: $u'(x^*) = c'(x^*)$, substituting we obtain $x^* = \frac{1}{2}$.

Treatments 2 and 3: $\frac{1}{5}u'(5x_i^{**}) - c'(x_i^{**}) = 0$, substituting we obtain $x_i^{**} = x_i^{***} = \frac{1}{2}$.

When the subjects are risk neutral, the equilibrium output will be the same in all three treatments.

Now we will assume that subjects are risk averse, and have a VNM utility function as previously considered. Since solving without assuming a functional form is not straightforward, we will assume a concave utility function $u(x) = lnx$, and a quadratic cost function $c(x) = x^2$.

Treatment 1: $u'(x^*) = c'(x^*)$, if we substitute we obtain $x^* = \sqrt{\frac{1}{2}}$.

Treatments 2 and 3: $\frac{1}{5} \cdot u'(w \cdot 5x_i) = c'(x_i)$, if we substitute we obtain
$x^* = \sqrt{\frac{1}{10}}$.

Hence, when the subjects are risk averse, the equilibrium output in the piece rate treatment is higher than the output generated in the either lottery treatment.

Note that we use output x as a proxy for the subject's effort. This is because we assume the effort is perfectly correlated with the output, hence the employment of a real effort task in the experimental design.

4.5 Results

	Periods				
Productivity (sliders)	1	2	3	4	5
Piece Rate	23.56	25.28	26.40	27.29	27.16
(T1)	(4.26)	(5.48)	(4.80)	(5.21)	(5.36)
Individual	24.88	26.14	26.40	27.50	28.71
Lottery (T2)	(7.11)	(7.15)	(7.13)	(6.93)	(6.25)
Group	24.00	26.22	26.84	27.56	27.92
Lottery (T3)	(6.03)	(6.52)	(6.22)	(6.02)	(6.84)

Standard deviations in parentheses

Table 11: Mean slider output in periods 1-5, under Piece Rate, Individual Lottery, and Group Lottery treatments.

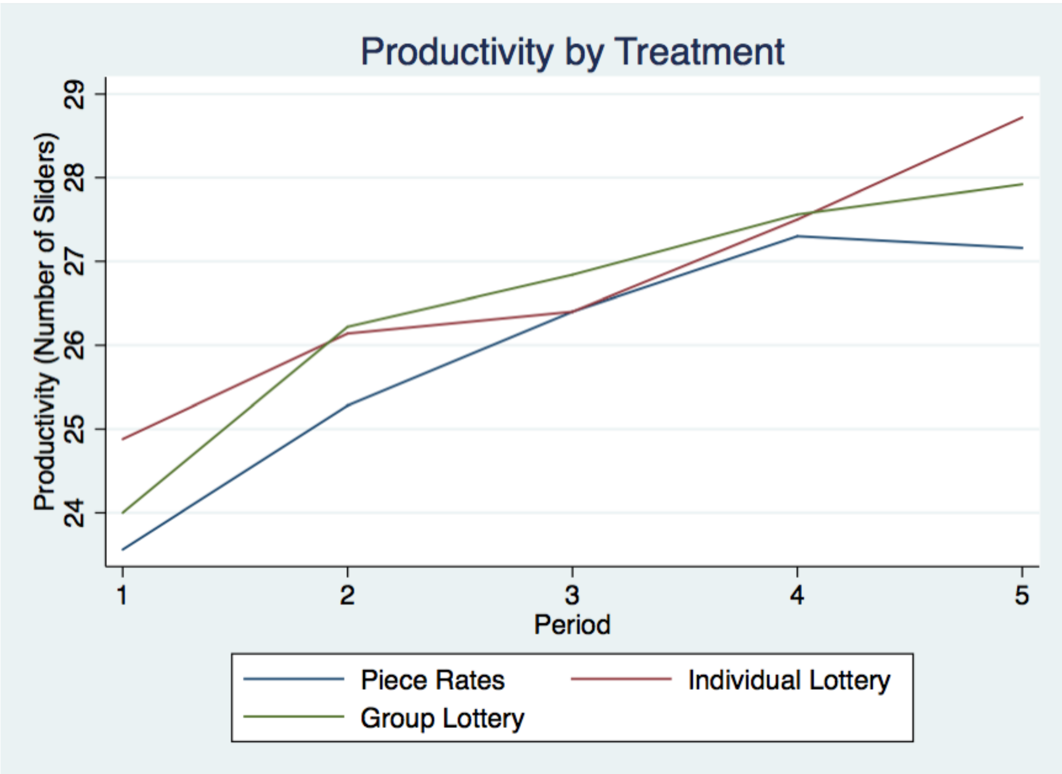


Figure 4.1:Mean slider output in periods 1-5, under Piece Rate, Individual Lottery, and Group Lottery treatments.

Figure 4.1 compares mean productivity in each treatment and its evolution over time; Table 1 shows the corresponding mean productivity for each treatment in each time period, and the associated standard deviations. The group lottery is uniformly more productive than piece rates (though it is shown below that we cannot reject the null that the two series are equal); productivity in the group lottery is uniformly more widely distributed than piece rates. Both the group lottery and piece rate productivity trends are concave, indicative of a decrease in the rate of learning over time, and/or fatigue. Participants in the individual lottery are at least as productive as the group lottery with the exception of Period 3.

We first conduct Mann-Whitney and Kolmogorov-Smirnov non-parametric tests on data aggregated at the group level for each period to test the null hypothesis that group productivity in both lottery treatments are drawn from the same distribution as the piece rate treatment. The p-values generated are presented in Table 4.2.

Productivity in the group lottery is consistently slightly above, but not is statistically different from piece rates in both non-parametric tests when observations are pooled by group. Productivity differentials are globally larger for the individual lottery, but the null hypothesis of no effect cannot be rejected for any period.

Null Hypothesis	Mann-Whitney	t-test	Kolmogorov-Smirnov
T1 = T2	0.7624	0.4411	0.976
T1 = T3	0.7623	0.5072	0.660
T2 = T3	0.9397	0.7918	0.660

Table 4.2: Mann-Whitney, t, and K-S tests of the null hypotheses of group equivalence.

It should be noted that an established characteristic of the slider task is that variation in productivity is typically very tightly distributed, such that significant results are hard to identify (e.g. Banuri & Keefer, 2015); this problem is compounded by the within-group variation lost when pooling across groups.

In the ensuing analysis we therefore further employ various parametric specifications at the individual level to establish the extent to which the lack of statistical significance in the non-parametric tests may be driven by this uncaptured variation; this also allows for the inclusion of relevant covariates to disseminate the factors driving the outcomes.

$$Y_{it} = \hat{\alpha}_0 + \hat{\alpha}_1\tau_t + \hat{\alpha}_2\tau_t^2 + \sum_{L=2}^{L=3}\hat{\beta}_L T_i^L + \left[\sum_{L=1}^{L=3}\delta_L\left(\bar{Y}_{gi(t-1)}^{-i}\cdot T_i^L\right)\right] + \left[\sum_{L=1}^{L=3}\hat{\theta}_L\left(R_i\cdot T_i^L\right)\right]$$
$$+\left[\sum_{L=1}^{L=3}(\mathbf{S}_i\cdot T_i^L)\hat{\vartheta}\right] + \left[\sum_{L=1}^{L=3}(\mathbf{S}_i\cdot T_i^L\cdot \bar{Y}_{gi(t-1)}^{-i})\hat{\rho}\right] + \mathbf{X}_i\eta + \hat{\varepsilon}_{it}$$

Equation 1

We estimate various iterations of Equation 1 by OLS. Y_{it} is output (or, equivalently, effort) by individual i in period t. τ is a continuous measure for time, measured in periods such that $\tau \in \mathbb{N}\,[1,5]$; τ^2 is specified to capture concavity. T_i^L is a set of treatment dummies, where $L = (1,2,3)$ for piece rate, individual lottery, and group lottery treatments respectively. $\hat{\beta}_2$ and $\hat{\beta}_3$ therefore estimates the differential impact of the individual lottery and group the group lottery, comparative to the piece rate baseline, respectively. $\bar{Y}_{gi(t-1)}^{-i}$ measures a one-period lag of mean productivity of individual i's group members, excluding individual i; δ_L therefore estimates at the margin the impact of higher group productivity on individual i's productivity in the following period, for each treatment (including the piece rate baseline). $R_i \in \mathbb{N}\,[0,10]$ is a continuous measure of risk aversion as per Holt and Laury (2002), such that preferences for risk are increasing in R_i; $\hat{\theta}_L$ therefore estimates the marginal increase in productivity resultant of a one unit increase in preferences for risky gambles, again for each treatment. $\mathbf{S}_i$ is a matrix of social preference dummies for revealed social preferences for group envy and pro-sociality, for each individual i; $\hat{\vartheta}$ is therefore a vector of estimators of the impact of each social preference classification on productivity, estimated separately for each treatment (including the piece rates). $\mathbf{S}_i\cdot T_i^L\cdot \bar{Y}_{gi(t-1)}^{-i}$ interacts social preferences with other group members' lagged productivity for each treatment, $\hat{\rho}$ is the coefficient on the interaction effect. $\mathbf{X}_i$ is a $(k\times N)$ matrix of k explanatory covariates, consisting of a sex dummy, dummies indicating whether the subject is reading Psychology or

Economics, and the (self-reported) mean number of hours the subject spends playing video games; it is important to control for this variable as whilst selection into treatment was fully random, there exist some observable differences between treatment groups in characteristics have potential to drive results (see Table 4.3). For example, the Group Lottery contains less economics students and no psychology students; the Individual Lottery contains a higher percentage of males, and fewer participants spend on average more time per week playing video games. $\mathbf{X}_i$ also contains an interaction between Psychology and Sex dummies, to prevent the relatively large proportion of female Psychology students biasing the Psychology dummy. $\hat{\varepsilon}_{it}$ is a residual error term, assumed to be mean zero and normally distributed. Across all specifications we estimate Equation 1 using cluster-robust standard errors to account for heteroscedasticity specifically arising from heterogeneous within-group variances. Results are presented in Table 4.4.

	Piece Rates	Individual Lottery	Group Lottery
Productivity	23.951	25.063	24.740
	(6.136)	(7.364)	(6.813)
Psychology	0.060	0.040	0.000
	(0.238)	(0.196)	(0.000)
Economics	0.220	0.260	0.040
	(0.415)	(0.439)	(0.196)
Male	0.300	0.500	0.380
	(0.459)	(0.501)	(0.486)
Videogames	2.796	4.540	2.786
	(5.839)	(7.585)	(5.028)
Risk	4.900	4.800	4.920
	(2.035)	(2.023)	(1.981)
Pro Soc	0.960	0.920	0.860
	(0.196)	(0.272)	(0.347)
Envy	0.680	0.740	0.700
	(0.467)	(0.439)	(0.459)

Standard deviations in parentheses

Table 4.3: Mean characteristics by treatment group.

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
T2	0.79	0.08	2.49	-3.51	-1.29	-1.78	0.47	4.60
	(1.11)	(1.04)	(1.93)	(2.60)	(2.99)	(2.88)	(4.64)	(7.35)

T3	0.57	0.51	8.428***	-4.72	3.44	2.72	2.76	14.18**
	(1.03)	(1.00)	(1.91)	(3.58)	(3.78)	(3.92)	(6.25)	(6.12)
Period	1.886***	1.928***	1.12	1.886***	1.12	1.08	1.13	1.13
	(0.35)	(0.35)	(0.82)	(0.35)	(0.81)	(0.81)	(0.81)	(0.84)
Period^2	-0.162***	-0.165***	-0.08	-0.162***	-0.08	-0.07	-0.07	-0.08
	(0.06)	(0.06)	(0.11)	(0.06)	(0.11)	(0.11)	(0.11)	(0.11)
Previous_ effort*T1			0.142***		0.141***	0.124***	0.118***	0.196**
			(0.04)		(0.04)	(0.04)	(0.03)	(0.08)
Previous_ effort*T2			0.07		0.06	0.05	0.04	0.02
			(0.06)		(0.06)	(0.06)	(0.06)	(0.20)
Previous_ effort*T3			-0.126***		-0.124**	-0.100**	-0.104**	-0.408***
			(0.05)		(0.05)	(0.04)	(0.05)	(0.12)
Risk_Pref *T1				-0.550**	-0.569**	-0.476*	-0.610**	-0.603**
				(0.27)	(0.27)	(0.25)	(0.25)	(0.25)
Risk_Pref *T2				0.34	0.25	0.27	0.24	0.24
				(0.37)	(0.35)	(0.32)	(0.33)	(0.34)
Risk_Pref *T3				0.53	0.43	0.42	0.43	0.40
				(0.58)	(0.56)	(0.61)	(0.55)	(0.54)
Male		1.997*				1.824*	2.209**	2.555**
		(1.07)				(1.02)	(1.06)	(1.07)
Econ		0.81				0.72	0.39	0.30
		(1.16)				(1.19)	(1.22)	(1.22)
Psycho		-1.04				-0.83	-0.58	-0.52
		(1.93)				(2.37)	(2.38)	(2.35)
Psycho* Male		-5.477*				-5.751*	-5.593*	-5.432*
		(2.96)				(3.19)	(3.14)	(3.08)
Vidogame _hours		0.187*				0.173*	0.13	0.11
		(0.10)				(0.10)	(0.10)	(0.11)
ProSocial* T1							3.932***	5.907***
							(1.40)	(1.67)
ProSocial* T2							1.18	1.85
							(1.87)	(5.14)
ProSocial* T3							2.58	-5.83
							(3.35)	(3.78)
Envy*T1							-1.33	-0.46
							(1.34)	(2.39)
Envy*T2							-1.13	-4.68
							(2.44)	(4.03)
Envy*T3							-0.30	-3.21
							(1.71)	(3.05)
Prosocial* Previous_ effort*T1								-0.06
								(0.05)

Prosocial* Previous_ effort*T2								-0.04
								(0.17)
Prosocial* Previous_ effort*T3								0.287***
								(0.10)
Envy* Previous_ Effort*T1								-0.03
								(0.07)
Envy* Previous_ Effort*T2								0.11
								(0.12)
Envy* Previous_ Effort*T3								0.10
								(0.09)
Constant	22.06***	20.87***	19.76***	24.76***	22.55***	21.59***	19.43***	16.92***
	(0.71)	(0.84)	(1.66)	(1.57)	(2.27)	(2.17)	(2.89)	(3.14)
Observations	750	740	600	750	600	592	592	592
R-squared	0.048	0.142	0.077	0.071	0.097	0.17	0.188	0.204

Cluster-Robust (at individual level) standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1
Table 4.4: Equation 1, results estimated by OLS

In specification (1) we estimate Equation 1 without covariates to establish raw differentials; similarly to the non-parametric specifications presented above, there is no statistically significant productivity difference between either lottery treatment and the piece rate. Though not significantly different from zero, point estimates for $\hat{\beta}_1$ and $\hat{\beta}_2$ and small and positive. Whilst the raw differentials are small and not significant, this result confutes the standard theory predictions: subjects are not found to be adversely disincentivised by risky incentives comparative to a riskless incentive with comparative expected value. We therefore add covariates to analyse the behavioural factors driving this result.

Specification (2) adds the characteristic control variables in $\mathbf{X}_i$. Conditional differentials $\hat{\beta}_1$ and $\hat{\beta}_2$ remain essentially unchanged, however R^2 increases from 0.048 to 0.142.

Specification (3) adds the lagged term $\bar{Y}_{gi(t-1)}^{-i}$ to the raw differential estimation in specification (1). A small but significant (0.142) marginal impact of other group members' productivity in the previous period on current performance is identified for the piece rate treatment; a small but significant negative effect of similar magnitude (-0.126) is identified for the group lottery. Point estimates of the effect of the individual lottery are close to zero and not statistically significant. Inclusion of $\bar{Y}_{gi(t-1)}^{-i}$ in (3) does induce a large and significant conditional differential of the Group Lottery treatment comparative to piece rates.

Specification (4) adds risk preferences to the raw differential specification. More risk-loving participants perform significantly less well in the piece rate treatment comparative to their peers (coefficient of -0.55); this is reversed for the individual and group lotteries respectively (positive coefficients of 0.335 and 0.527 respectively), though the lottery coefficients are not significant. Conditional differentials $\hat{\beta}_1$ and $\hat{\beta}_2$ become larger in magnitude and negative when controlling for risk preferences alone, but remain not significant. These results seem in line with standard theory: those with more appetite for risk perform better under the riskier incentive mechanisms, and when controlling for this alone the lottery treatments perform worse.

Specification (5) adds both $\bar{Y}_{gi(t-1)}^{-i}$ and risk preferences, and Specification (6) adds the full set of characteristic control variables in $\mathbf{X}_i$. Covariate coefficients remain broadly similar to those in (2), (3) and (4); the Individual and Group Lottery conditional differentials, $\hat{\beta}_1$ and $\hat{\beta}_2$, become negative and positive respectively, but neither are significant.

Specification (7) adds social preferences interacted with treatment dummies, $\mathbf{S}_i \cdot T_i^L$. We find that those with pro-social preferences are significantly more productive in the piece rate treatment than their peers by 3.932 sliders per round, significant at the 1% level. Pro-sociality coefficients are also positive for both lottery treatments, but remain insignificant. Coefficients on Envy are negative for all three treatments, but are not significant.

Finally, Specification (8) adds interactions between the social preference dummies for each treatment with lagged mean performance of other group members. This addition increases the size of the coefficient on Pro-Sociality in the piece rate treatment from 3.932 to 5.907, which remains significant at the 1% level; the Pro-Sociality coefficient in the Group Lottery treatment becomes large and negative, but remains insignificant. This seems to be driven by a significant (at 1% level) positive relationship between the prior performance of other group members and current performance of those with pro-social preferences (coefficient of 0.287). We hypothesize that those individuals with pro-social preferences behave as conditional co-operators, exerting higher effort when they observe their peers exerting high effort, hence maximizing the prize pot. Accounting for social preferences in this way, the conditional differential effect of the Group Lottery comparative to piece rates becomes very large (14.18), and significant at the 5% level.

4.6 Discussion & Conclusions

Contrary to the standard theoretical prediction that subjects will exert less effort where incentives involve risk comparative to riskless incentives, our experiment shows that lottery incentives perform at least as well as piece rates in inducing effort. Controlling for observable characteristics, risk preferences, social preferences, and prior group performance (including interactions between social preferences and prior group performance), we find a large and significant conditional differential in performance between the group lottery and piece rates. This is compatible with the findings obtained by Celis et al. (2013), Rula et al. (2014), Rokicki et al. (2014), and Levitt et al. (2016), where lottery incentives are found to have a positive effect on outcomes comparative to riskless alternatives.

Contrary to the theoretical suggestions in Zabojnik (2012), in which a quasi-convex utility is assumed on the part of the agent to assert that lottery contracts perform better because of a risk-seeking element in the locally convex portion of the workers utility function, we do not find that risk-seeking preferences significantly drive performance in the lottery treatments. We do, however, find that more risk-seeking agents perform slightly (but significant to 1% level) worse in the piece rate treatment comparative to their peers.

The prior effort of other group members is positively related to current period effort in the piece rate treatment. This is conceptually compatible with peer effects, where there is an imitation effect, or an increase of optimism, arising where a subject observes her peers performing well. Another possible explanation for this finding may competition effects, where the subjects may conceivably seek to retain or improve their group ranking. Nonetheless, the effect for the group lottery treatment is not significant. This could be due to the fact that, due to the risky nature of the payoff structure, higher effort does not necessarily correspond with higher payoff, hence risk aversion may counteract this peer effect. Lahno and Serra-Garcia (2014), find that peer effects are greater when subjects are given a choice between two lotteries, relative to the scenario where subjects are randomly allocated to one of two lotteries. We hypothesize that the illusion of control over the outcome increases the incidence of peer effects, which is consistent with both out and Serra-Garcia's findings. Nonetheless, more research would need to be carried out in this area.

We also find that individuals with prosocial preferences are significantly more productive than their peers in the piece rate treatment. The converse effect is observed in the group lottery treatment, where prosocial individuals perform worse than their peers, although whilst the negative coefficient is large it is not statistically significant. We hypothesise that this is compatible with prosocial individuals' preference for equitability: bonus payments in the piece rate treatment depend only on an individual's own effort, so there is no potential for shirkers to profit from the high effort of others.

Notably, by design, the total bonus pool in the Group Lottery is positively related to other group members' effort in any given period, though the probability of winning is negatively related to other group members' effort. In this context, our finding that prosocial individuals perform significantly better than their peers when they observe other subjects in their group exerting high effort in the Group Lottery is a powerful result. We hypothesise that prosocial individuals behave as conditional co-operators where contributing to the bonus pool is a group effort, exerting high effort only when they observe their peers doing similarly in previous periods.

It is noteworthy that neither of these results hold true of the individual lottery: the conditional differential for the individual lottery is smaller and insignificant; similarly the pro-sociality coefficient is positive, but small and not significant, and the interaction between pro-sociality and prior peer performance is very close to zero and not significant. In the context of the above, we argue that the design of the lottery plays an important role in whether a risky incentive mechanism induces effort in those likely to respond to it. This supports our findings: it does not appear to be risk in itself that induces an effort response that depends on group performance for prosocial participants, but rather a mechanism design that incentivises group co-operation.

Whilst the raw differentials between each lottery treatment and the piece rate are not statistically different between groups, productivity in the group lottery is higher than the piece rate across all periods. Given the slider task is often found to be characterised by tightly distributed productivity (Banuri & Keefer, 2015) there is scope for this research to be extended in a field experiment with a real-world task, and with a larger sample size. Whilst this means we cannot reject the null hypothesis that the raw differentials between each lottery and piece rates are zero, we do find strong evidence of heterogeneous effort responses, particularly for those with prosocial preferences.

4.8 Appendix 1: Experimental instructions

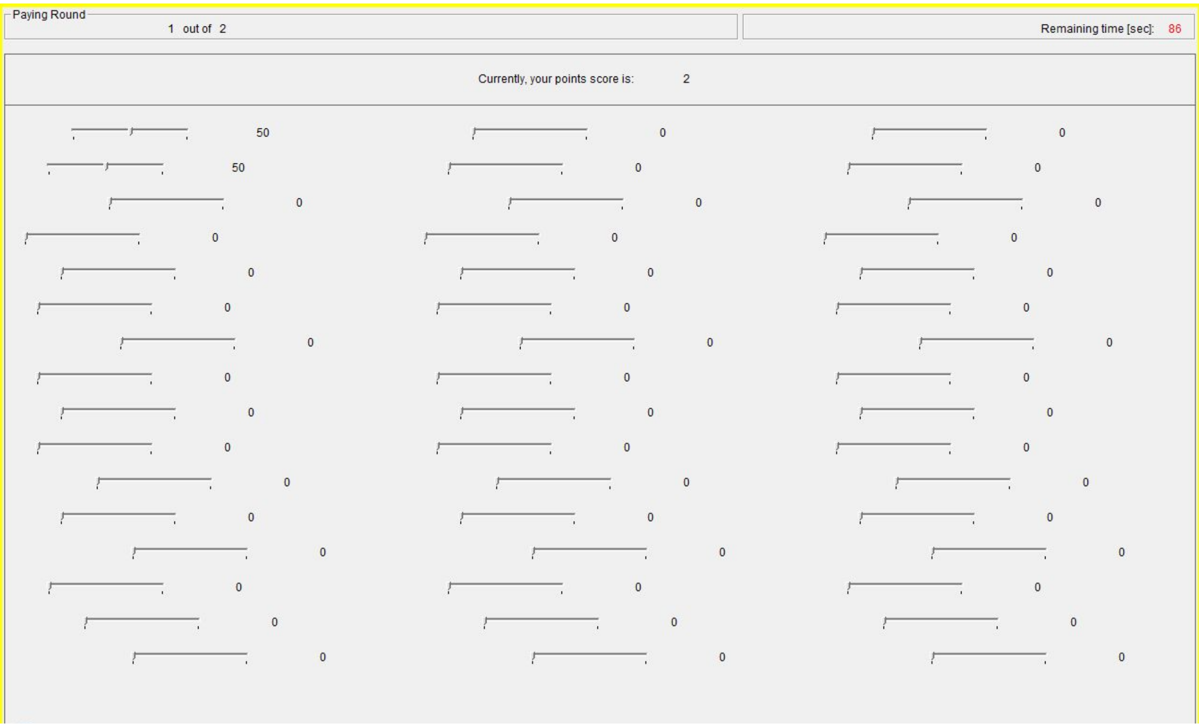
PIECE RATE TREATMENT

Welcome to the experiment!

You are about to take part in a study on economic decision-making. The study will last about one hour. We are very grateful for your participation. You will be paid a guaranteed show-up fee of £4 for participating in the experiment. This is independent of your actions during the experiment. You will also have the opportunity to earn additional money. The final amount you earn will depend on your actions.

THE TASK

In each round, you will see 48 sliders on the screen.



Your task is to set the position of each slider to **the centre**. Each slider is initially positioned at 0 and your task is to set as many sliders as possible **to 50**.
You have **3 minutes** to set as many sliders as possible to 50.
Each slider has a number to its right showing its current position. You should use your mouse to move each slider. You can readjust the position of each slider as many times as you wish. If you click on the slider, it will jump, hence it needs to be dragged.
At the top of the screen you see the time remaining and your points score in the task so far.

EARNINGS

Your payment in each round depends on the number of sliders you positioned. If you position 10 or more sliders at 50, you will receive a **fixed payment** of £0.60 per round. If you correctly position less than 10 sliders, you will not receive any payment for that round (save for the show-up fee).
After you correctly position 15 sliders, each extra correctly positioned slider will pay you an additional **bonus** of £0.06 each.
Therefore, if you position at least 10 sliders correctly in all rounds, you can earn a total fixed pay of £3 (5 * £0.60) plus your bonus in each round.

ROUNDS

You will play 5 rounds of this task. Before these commence, you will have 2 additional practice rounds that will allow you to familiarise yourself with the task. Your performance in the practice rounds will not influence your final payoff.

Please raise your hand if you have any questions.

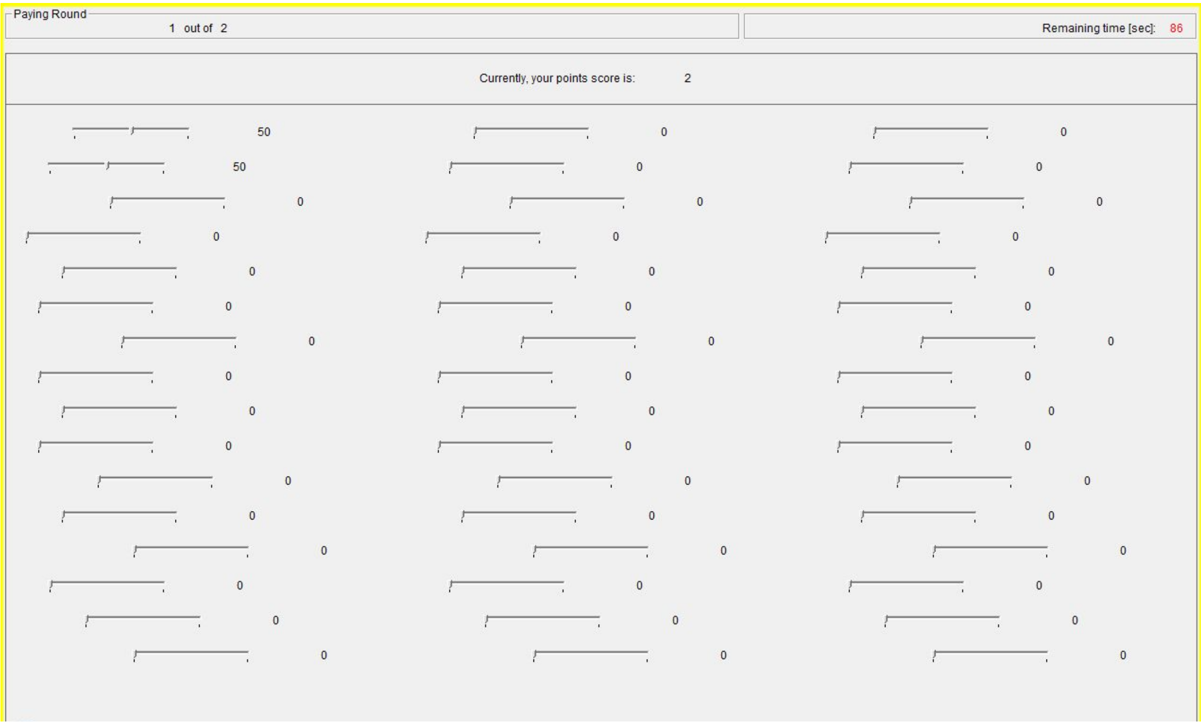
INDIVIDUAL LOTTERY TREATMENT

Welcome to the experiment!

You are about to take part in a study on economic decision-making. The study will last about one hour. We are very grateful for your participation. You will be paid a guaranteed show-up fee of £4 for participating in the experiment. This is independent of your actions during the experiment. You will also have the opportunity to earn additional money. The final amount you earn will depend on your actions.

THE TASK

In each round, you will see 48 sliders on the screen.



Your task is to set the position of each slider to **the centre**. Each slider is initially positioned at 0 and your task is to set as many sliders as possible **to 50**.
You have **3 minutes** to set as many sliders as possible to 50.

Each slider has a number to its right showing its current position. You should use your mouse to move each slider. You can readjust the position of each slider as many times as you wish. If you click on the slider, it will jump, hence it needs to be dragged.
At the top of the screen you see the time remaining and your points score in the task so far.

EARNINGS

Your payment in each round depends on the number of sliders you positioned. If you position 10 or more sliders at 50, you will receive a **fixed pay** of £0.60 per round. If you correctly position less than 10 sliders, you will not receive any payment for that round (save for the show-up fee).
After you correctly position 15 sliders, each extra correctly positioned slider add £0.30 to your bonus pool. At the end of each round, you may receive this bonus pool, with 20% probability.
Therefore, if you position at least 10 sliders correctly in all rounds, you can earn a total fixed pay of £3 (5 * £0.6) plus your bonus in each round.

ROUNDS

You will play 5 rounds of this task. Before these commence, you will have 2 additional practice rounds that will allow you to familiarise yourself with the task. Your performance in the practice rounds will not influence your final payoff.

Please raise your hand if you have any questions.

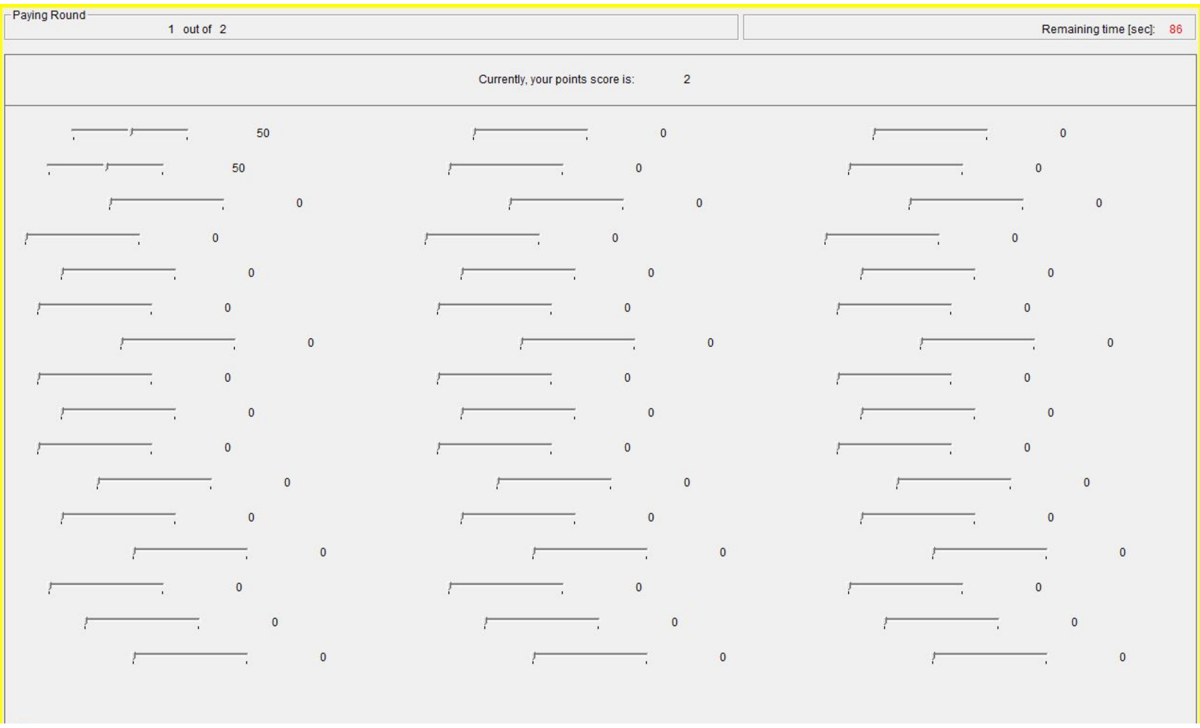
GROUP LOTTERY TREATMENT

Welcome to the experiment!

You are about to take part in a study on economic decision-making. The study will last about one hour. We are very grateful for your participation. You will be paid a guaranteed show-up fee of £4 for participating in the experiment. This is independent of your actions during the experiment. You will also have the opportunity to earn additional money. The final amount you earn will depend on your actions.

THE TASK

In each round, you will see 48 sliders on the screen.



Your task is to set the position of each slider to **the centre**. Each slider is initially positioned at 0 and your task is to set as many sliders as possible **to 50**.
You have **3 minutes** to set as many sliders as possible to 50.

Each slider has a number to its right showing its current position. You should use your mouse to move each slider. You can readjust the position of each slider as many times as you wish. If you click on the slider, it will jump, hence it needs to be dragged.
At the top of the screen you see the time remaining and your points score in the task so far.

EARNINGS

You will be working in a fixed group of 5 people.
Your payment in each round depends on the number of sliders you positioned. If you position 10 or more sliders at 50, you will receive a **fixed payment** of £0.60 per round. If you correctly position less than 10 sliders, you will not receive any payment.
After you correctly position 10 sliders, each extra correctly positioned slider will add £0.06 to the group bonus pool, and earn you a lottery ticket to win the entire pool. At the end of each round, one of the lottery tickets will be selected, and its owner will earn the total prize. Note that more lottery tickets implies a higher chance of winning the total prize.
Thus, if you position at least 15 sliders correctly in all rounds, you can earn a total fixed pay of £3 (5 * £0.60) plus your lottery earnings in each round.

ROUNDS

You will play 5 rounds of this task. Before these commence, you will have 2 additional practice rounds that will allow you to familiarise yourself with the task. Your performance in the practice rounds will not influence your final payoff.

Please raise your hand if you have any questions.

4.9 Appendix 2: Social preferences elicitation game

We use the social preference game developed in Bartling et al., 2009. Our participants are asked to imagine a hypothetical scenario where they are allocated with another anonymous participant, and have to choose between two binary allocations for each social preference principle. The first allocation corresponds to an egalitarian distribution in all games, while the second allocation favours the decision maker (other participant) in the prosociality (envy) and costly prosociality (costly envy) game.

All payoffs are in pounds.

self: other Distribution 2

self: other

(I) Pro-sociality 2: 2, 2: 1

(II) Costly pro-sociality 2: 2, 3: 1

(III) Envy 2: 2, 2: 4

(IV) Costly envy 2: 2, 3: 5

For our results, we consider the four principle variables as dummies that take value 1 when the first allocation is selected, and 0 when the second allocation is selected. This task is not monetarily incentivised.

4.10 Appendix 3: Risk Preferences elicitation

We use the task developed by Holt and Laury (2002) and measure it as the selected amount of non-safe options, hence a higher coefficient indicates a higher risk loving preference. This task is not monetarily incentivised.

Option A	Option B	Expected payoff difference
1/10 of \$2.00, 9/10 of \$1.60	1/10 of \$3.85, 9/10 of \$0.10	\$1.17
2/10 of \$2.00, 8/10 of \$1.60	2/10 of \$3.85, 8/10 of \$0.10	\$0.83
3/10 of \$2.00, 7/10 of \$1.60	3/10 of \$3.85, 7/10 of \$0.10	\$0.50
4/10 of \$2.00, 6/10 of \$1.60	4/10 of \$3.85, 6/10 of \$0.10	\$0.16
5/10 of \$2.00, 5/10 of \$1.60	5/10 of \$3.85, 5/10 of \$0.10	−\$0.18
6/10 of \$2.00, 4/10 of \$1.60	6/10 of \$3.85, 4/10 of \$0.10	−\$0.51
7/10 of \$2.00, 3/10 of \$1.60	7/10 of \$3.85, 3/10 of \$0.10	−\$0.85
8/10 of \$2.00, 2/10 of \$1.60	8/10 of \$3.85, 2/10 of \$0.10	−\$1.18
9/10 of \$2.00, 1/10 of \$1.60	9/10 of \$3.85, 1/10 of \$0.10	−\$1.52
10/10 of \$2.00, 0/10 of \$1.60	10/10 of \$3.85, 0/10 of \$0.10	−\$1.85

5. Bibliography

Abadie, A., & Gardeazabal, J. (2003). The Economic Costs of Conflict: A Case Study of the Basque Country. *American Economic Review, 93*(1), 113-132.

Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California’s Tobacco Control Program. *Journal of the American Statistical Association, 105*(490), 493-505.

Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. *Journal of the American Statistical Association, 105*, 493-505. doi:10.3386/w12831

Abadie, A., Diamond, A., & Hainmueller, J. (2015). Comparative Politics and the Synthetic Control Method. *American Journal of Political Science, 59*(2), 495-510

Abadie, A. (2020). Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects. *Journal of Economic Literature (forthcoming)*.

Abeler, J., & Jäger, S. (2015). Complex Tax incentives. *American Economic Journal: Economic Policy, 7*(3), 1-28.

Altonji, J. G., & Card, D. (1991). The Effects of Immigration on the Labor Market Outcomes of Less-skilled Natives. In N. B. Research, J. M. Abowd, & R. B. Freeman (Eds.), *mmigration, Trade and the Labor Market* (pp. 201-234). Chicago: Chicago University Press.

Andrews, M., Clark, K., & Whittaker, W. (2008). The Determinants of Regional Migration in Great Britain: A Duration Approach. *IZA Discussion Paper Series*(DP No. 3783).

Angrist, J. D., & Pischke, J. S. (2009). *Mostly Harmless Econometrics.* Princeton, New Jersey: Princeton University Press.

Araujo, F. A., Carbone, E., Conell-Price, L., Dunietz, M. W., Jaroszewicz, A., & Landsman, R. (2016). The slider task: an example of restricted inference on incentive effects. *Journal of the Econimic Science Association, 2*, 1-12.

Ariely, D., Gneezy, U., Loewenstein, G., & Mazar, N. (2009). Large Stakes and Big Mistakes. *The Review of Economic Studies, 76*(2), 451-469.

Asch, B. J. (1990). Do Incentives Matter? The Case of Navy Recruiters. *Industrial and Labor Relations Review, 43*(3), 89S-106S.

Autor, D. H. (2003). Outsourcing at Will: The Contribution of Unjust Dismissal Doctrine to the Growth of Employment Outsourcing. *Journal of Labor Economics, 21*(1), 1-42.

Bachmann, R., Bechara, P., Kramer, A., & Rzepka, S. (2015). Labour market dynamics and worker heterogeneity during the Great Recession – Evidence from Europe. *IZA Journal of European Labor Studies, 4*(19). doi:https://doi.org/10.1186/s40174-015-0043-0

Baker, G., & Gibbons, R. M. (1994). Subjective Performance Measures in Optimal Incentive Contracts. *The Quarterly Journal of Economics, 109*(4), 1125-1156.

Bandiera, O., Barankay, I., & Rasul, I. (2005). Social Preferences and the Response to Incentives: Evidence from Personnel Data. *The Quarterly Journal of Economics, 120*(3), 917-962.

Banister, D., & Thurstain-Goodwin, M. (2011). Quantification of the non-transport benefits resulting from rail investment. *Journal of Transport Geography, 19*, 212-223.

Banuri, S., & Keefer, P. (2015). Was Weber Right? Effects of Pay for Ability and Pay for Performance on Pro-Social Motivation, Ability and Effort in the Public Sector (Working Paper 7261). *World Bank Development Research Group, Macroeconomics and Growth Team*.

Bartel, A. P. (1989). Where Do the New U.S. Immigrants Live? *Journal of Labor Economics, 7*(4), 371-391.

Bartling, B., Fehr, E., Maréchal, M. A., & Schunk, D. (2009). Egalitarianism and competitiveness. *The American Economic Review, 99*(2), 93-98.

Bell, D. N., & Blanchflower, D. G. (2010). UK unemployment in the great recession. *National institute economic review2010, 214*(1), R3-R25. doi:https://doi.org/10.1177/0027950110389755

Bell, D. N., & Blanchflower, D. G. (2011). Young people and the Great Recession. *Oxford Review of Economic Policy, 27*(2), 241-267. doi:https://doi.org/10.1093/oxrep/grr011

Berechman, J., & Paaswell, R. (2005). Evaluation, prioritization and selection of transport investment projects in New York City. *Transportation, 32*(2), 223-249.

Berechman, J., & Paaswell, R. E. (1983). Rail Rapid Transit Investment and CBD Revitalisation: Methodology and Results. *Urban Studies, 20*(4), 471-486. doi:10.1080/00420988320080821

Blanchflower, D. G., & Oswald, A. J. (1994). *The Wage Curve.* MIT Press.

Boeri, T., & Brücker, H. (2005). Migration, Co-ordination Failures and EU Enlargement . *German Institute for Economic Research Discussion Papers*(481).

Borjas, G. J. (2003, June). THE LABOR DEMAND CURVE IS DOWNWARD SLOPING: REEXAMINING THE IMPACT OF IMMIGRATION ON THE LABOR MARKET. *NBER Working Paper Series*(w9755).

Borjas, G. J. (2006). IMMIGRATION IN HIGH-SKILL LABOR MARKETS: THE IMPACT OF FOREIGN STUDENTS ON THE EARNINGS OF DOCTORATES. *NBER Working Paper Series*(w12085).

Borjas, G. J. (2016). The Wage Impact of the Marielitos: Additional Evidence. *NBER Working Paper Series*(w21850).

Borjas, G. J. (2017). The Wage Impact of the Marielitos: A Reappraisal. *Industrial and Labor Relations Review (forthcoming)*.

Borjas, G. J., Freeman, R. B., & Katz, L. F. (1992). On the Labor Market Effects of Immigration and Trade. In G. J. Borjas, & R. B. Freeman, *Immigration and the Work Force: Economic Consequencesfor the United States and Source Areas.* (pp. 213-244). Boston: University of Chicago Press.

Borjas, G. J., Freeman, R. B., & Katz, L. F. (1996). Searching for the Effect of Immigration on the Labour Market. *American Economic Review, 86*(2), 246-251.

Borjas, G. J., Grogger, J., & Hanson, G. H. (2006). IMMIGRATION AND AFRICAN AMERICAN EMPLOYMENT OPPORTUNITIES: THE RESPONSE OF WAGES, EMPLOYMENT, AND INCARCERATION TO LABOR SUPPLY SHOCKS. *NBER Working Paper Series*(w12518).

Bresson, G., Dargay, J., Madre, J.-L., & Pirotte, A. (2003). The main determinants of the demand for public transport: a comparative analysis of England and France using shrinkage estimators. *Transportation Research Part A, 37*, 605-627.

Brune, L. (2016). The Effect of Lottery-Incentives on Labor Supply: A Firm Experiment in Malawi. *Economic Development and Cultural Change, 64*(2), 187-220.

Buser, T., & Dreber, A. (2016). The Flipside of Comparative Payment Schemes. *Management Science, 62*(9), 2626-2638.

Butcher, K. F., & Card, D. (1991). Immigration and Wages: Evidence from the 1980's. *American Economic Review, 81*(2), 292-296.

Canavan, S., Graham, D. J., Anderson, R. J., & Barron, A. (2018). Urban Metro Rail Demand: Evidence from Dynamic Generalized Method of Moments Estimates using Panel Data. *Transportation Research Record, 2672*(8), 288–296.

Card, D. (1990). The Impact of the Mariel Boatlift on the Miami Labour Market. *Industrial and Labor Relations Review, 43*(2), 245-257.

Card, D. (2001). Immigrant Inflows, Native Outflows, and the Local Market Impacts of Higher Immigration. *Journal of Labor Economics, 19*(1), 22-64.

Caudill, S. B., & Holcombe, R. G. (1999). Specification Search and Levels of Significance in Econometric Models. *Eastern Economic Journal, 25*(3), 289-300.

Cavallo, E., Galiani, S., Noy, I., & Pantano, J. (2013). Catastrophic Natural Disasters and Economic Growth. *The Review of Economics and Statistics, 95*(5), 1549-1561.

Celis, L. E. (2013). Lottery-Based Payment Mechanism for Microtasks. *Human Computation and Crowdsourcing: Works in Progress and Demonstration Abstracts., AAAI Technical Report CR-13-01*, 12-13.

Cervero, R. (1990). Transit pricing research. *Transportation, 17*(2), 117–139.

Chevalier, J., & Ellison, G. (1999). Are Some Mutual Fund Managers Better Than Others? Cross-Sectional Patterns in Behaviour and Performance. *Journal of Finance, LIV*(3), 875-899.

Comber, S., & Arribas-Bel, D. (2017). "Waiting on the train": The anticipatory (causal) effects of Crossrail in Ealing. *Journal of Transport Geography, 64*, 13-22.

Combes, P.-P., Thisse, J.-F., & Mayer, T. (2008). *Economic Geography: The Integration of Regions and Nations.* Princeton, New Jersey: Princeton University Press.

Coulter, S. (2016). The UK labour market and the ‘great recession'. In M. Myant, S. Theodoropoulou, & A. Piasna, *Unemployment, Internal Devaluation and Labour Market Deregulation in Europe* (pp. 197-224). Brussels: European Trade Union Institute. Retrieved from http://eprints.lse.ac.uk/65615/

Day, J. R., & Reed, J. (2010). *The Story of London's Underground* (11 ed.). London: Capital Transport Publishing .

De Grange, L., González, F., Muñoz, J. C., & Troncoso, R. (2013). Aggregate estimation of the price elasticity of demand for public transport in integrated fare systems: The case of Transantiago. *Transport Policy, 29*, 178-185.

Debrezion, G., Pels, E., & Rietveld, P. (2007). The impact of railway stations on residential and commercial property value: A meta-analysis. *Journal of Real Estate Finance and Economics, 35*(2), 161-180.

Dehejia, R., & Wahba, S. (1999). Causal effects in nonexperimental studies: reevaluating the evaluation of training programs. *Journal of the American Statistical Association, 94*(448), 1053-1062.

Department for Environment and Rural Affairs. (2018). *Modelling of Ambient Air Quality (MAAQ) Data.* Retrieved November 2018, from https://uk-air.defra.gov.uk/data/pcm-data

Department for Transport. (2018). *Average Annual Daily Flows.* Retrieved November 2018, from https://roadtraffic.dft.gov.uk/downloads

Dibben, C., Atherton, I., Cox, M., Watson, V., Ryan, M., & Sutton, M. (2007). *Investigating the Impact of Changing the Weights that Underpin the Index of Multiple Deprivation 2004.* Department for Communities and Local Government.

Dibben, C., McLennan, D., Barnes, H., Noble, M., Davies, J., & Garratt, E. (2011). *The English Indices of Deprivation 2010.* Department for Communities and Local Government.

Doerrenberg, P., & Duncan, D. (2014). Experimental evidence on the relationship between tax evasion opportunities and labor supply. *European Economic Review, 68*(C), 48-70.

Dustmann, C., Casanova, M., Fertig, M., Preston, I., & Schmidt, C. (2003). *The impact of EU enlargement on migration flows.* Home Office, Research Development and Statistics Directorate, London.

Dustmann, C., Fabbri, F., & Preston, I. (2005, November). THE IMPACT OF IMMIGRATION ON THE BRITISH LABOUR MARKET. *The Economic Journal, 115*, F325-F341.

Dustmann, C., Fabbri, F., Preston, I., & Wadsworth, J. (2003). *The local labour market effects of immigration in the UK.* Home Office.

Dustmann, C., Frattini, T., & Preston, I. (2007). Immigration and Wages: New Evidence for Britain. . *Department of Economics and Centre for Research and Analysis of Migration (CReAM), UCL.*

Dustmann, C., Schönberg, U., & Stuhler, J. (2016). The Impact of Immigration: Why Do Studies Reach Such Different Results? *Journal of Economic Perspectives, 30*(4), 31-56.

Evening Standard. (2015). *Will Oyster cards give Epsom the edge over Ewell as the top commuter homes hotspot?* Retrieved from https://www.homesandproperty.co.uk/area-guides/uk-areas/will-oyster-cards-give-epsom-the-edge-over-ewell-as-the-top-commuter-homes-hotspot-45791.html

Ferman, B., Pinto, C., & Possebom, V. (2017). Cherry Picking with Synthetic Controls. *Munich Personal RePEc Archive*.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economics experiments. *Experimental Economics, 10*(2), 171-178.

Forrest, D., Glen, J., & Ward, R. (1996). The impact of a light rail system on the structure of house prices - A hedonic longitudinal study. *Journal of Transport Economics and Policy, 30*(1), 15-29.

Freeman, R. B., & Kleiner, M. M. (1998). The Last American Show Manufacturers: Changing the Method of Pay to Survive Foreign Competition. *NBER Working Paper Series (w6750)*.

Fryer Jr, R. G., Levitt, S. D., List, J., & Sadoff, S. (2012). Enhancing the Efficacy of Teacher Incentives Through Loss Aversion: A Field Experiment. *NBER Working Paper Series (w18237)*.

Galiani, S., & Quistorff, B. (2016). *The synth_runner Package: Utilities to Automate Synthetic Control Estimation Using synth∗*. University of Maryland.

Galiani, S., & Quistorff, B. (2017). The synth runner package: Utilities to automate synthetic control estimation using synth. *Stata Journal, 17*(4), 834-849.

Gatwick Airport. (2018). *2018 Gatwick Capital Investment Programme.* Retrieved from https://www.gatwickairport.com/globalassets/publicationfiles/business_an d_community/all_public_publications/transforming_gatwick/2018-cip---final---issued-june-2018.pdf

Gaynor, M., Rebitzer, J. B., & Kleiner, M. M. (2004). Physician Incentives in Health Maintenance Organizations. *Journal of Political Economy, 112*(4), 915-931.

Gelman, A. (2006). Multilevel (Hierarchical) Modeling: What it Can and Cannot Do. *TECHNOMETRICS*, 432-435.

Georganas, S., Tonin, M., & Vlassopoulos, M. (2015). Peer pressure and productivity: The role of observing and being observed. *Journal of Economic Behavior & Organization, 117*, 223-232.

Gibbons, R. (1987). Piece-Rate Incentive Schemes. *Journal of Labor Economics, 5*(4), 413-429.

Gibbons, R., & Murphy, K. J. (1992). Optimal Incentive Contracts in the Presence of Career Concerns: Theory and Evidence. *Journal of Political Economy, 100*(3), 468-505.

Gibbons, S., & Machin, S. (2005). Valuing rail access using transport innovations. *Journal of Urban Economics, 57*(1), 148-169. doi:10.1016/j.jue.2004.10.002

Gibbons, S., & Overman, H. G. (2010). Mostly Pointless Spatial Econometrics? *SERC Discussion Paper, 61*.

Gill, D., & Prowse, V. (. (2014). Gender differences and dynamics in competition: The role of luck. *Quantitative Economics, 5*, 351-376.

Gill, D., & Prowse, V. (2011). A Novel Computerized Real Effort Task Based on Sliders. *IZA Discussion Papers, 5801*.

Gill, D., & Prowse, V. (2012). A Structural Analysis of Disappointment Aversion in a Real Effort Competition. *American Economic Review, 102*(1), 469-503.

Gill, D., Prowse, V. L., & Vlassopoulos, M. (2013). Cheating in the Workplace: An Experimental Study of the Impact of Bonuses and Productivity. *Journal of Economic Behaviour and Organization*, forthcoming.

Graham, P., & Mulley, C. (2012). Public transport pre-pay tickets: Understanding passenger choice for different products. *Transport Policy, 19*(1), 69-75.

Greater London Authority. (2010). *Mayor's Transport Strategy.* London: Greater London Authority. Retrieved from https://infrastructure.planninginspectorate.gov.uk/wp-content/ipc/uploads/projects/TR010021/TR010021-000409-Mayor's%20Transport%20Strategy.pdf

Guglielminetti, E., Lalive, R., Ruh, P., & Wasmer, E. (2017). Home Sweet Home? Job Search with Commuting and Unemployment Insurance.

Henneberry, J. (1998). Transport investment and house prices. *Journal of Property Valuation and Investment, 16*(2), 144-158. doi:10.1108/14635789810212913

Hensher, D. A. (2008). Assessing systematic sources of variation in public transport elasticities: Some comparative warnings. *Transportation Research Part A, 42*, 1031-1042.

Highways Agency. (2014). *Post Opening Project Evaluation, M25 Junction 16-23 Widening- One Year After Study.* Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/387327/POPE___M25_J16-23_OYA___Final.pdf

Highways England. (2015). *Planned roadworks in and around the M25*. Retrieved from https://www.gov.uk/government/news/planned-roadworks-in-and-around-the-m25-weekly-summary-for-monday-29-june-sunday-5-july

HM Land Registry. (2018). *Price Paid Data.* Retrieved November 2018, from https://www.gov.uk/government/statistical-data-sets/price-paid-data-downloads

Holmström, B. (1999). Managerial Incentive Problems: A Dynamic Perspective. *The Review of Economic Studies, 66*(1), 169-182.

Holmström, B., & Milgrom, P. (1991). Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design. *Journal of Law, Economics and Organisation, 7*, 24-52.

Home Office & BIS. (2014). *Impacts of migration on UK native employment: An analytical review of the evidence.* London.

Home Office. (2007). *Accession Monitoring Report May 2004 – December 2006.* London.

Ichniowski, C., Shaw, K., & Prennushi, G. (1997). The Effects of Human Resource Management Practices on Productivity: A Study of Steel Finishing Lines. *The American Economic Review, 87*(2), 197-313.

Karthik, M., & Sundararaman, V. (2009). Teacher Performance Pay: Experimental Evidence From India. *NBER Working Paper Series (w15323)*.

Kézdi, G. (2004). ROBUST STANDARD ERROR ESTIMATION IN FIXED-EFFECTS PANEL MODELS. *Hungarian Statistical Review, Special Number 9*, 95-116.

Kleven, H. J., Landais, C., & Saez, E. (2013). Taxation and International Migration of Superstars: Evidence from the European Football Market. *American Economic Review, 103*(5), 1892-1924.

LaLonde, R. J., & Topel, R. H. (1991). Labor Market Adjustments to Increased Immigration. In N. B. Research, J. M. Abowd, & R. B. Freeman (Eds.), *Immigration, Trade and the Labor Market* (pp. 167-199). Chicago: University of Chicago Press.

LaLonde, R. J., & Topel, R. H. (1991). Labor Market Adjustments to Increased Immigration. In &. R. J. M. Abowd, *mmigration, Trade and the Labor Market* (pp. 167-199). Chicago: University of Chicago Press.

Lavy, V. (2002). Evaluating the Effect of Teachers' Group Performance Incentives on Pupil Performance. *Journal of Political Economy, 110*(6), 1286-1317.

Lazear, E. P. (1986). Salaries and Piece Rates. *The Journal of Business, 59*(3), 405-431.

Lazear, E. P. (1996). Performance Pay and Productivity. *NBER Working Paper Series (w5672)*.

Lazear, E. P. (2000). Performance Pay and Productivity. *The American Economic Review, 90*(5), 1346-1361.

Lazear, E. P., & Oyer, P. (2007). Personnel Economics. *NBER Working Paper Series*.

Leape, J. (2006). The London congestion charge. *Journal of Economic Perspectives, 20*(4), 157-176. doi:10.1257/jep.20.4.157

LeighFisher. (2016). *Rail Demand Forecasting Estimation.* London: Department for Transport. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/610059/phase2-rail-demand-forecasting-estimation-study.pdf

Lemos, S., & Portes, J. (2008, October). New Labour? The Impact of Migration from Central and Eastern European Countries on the UK Labour Market. *IZA Discussion Paper Series*(DP No. 3756).

Levitt, S. D., List, J. A., & Sadoff, S. (2016). The Effect of Performance-Based Incentives on Educational Achievement: Evidence From a Ramdomized Experiment. *NBER Working Paper Series (w22107)*.

Mackett, R. L. (1983). *The Leeds integrated land-use transport model (LILT).* UK Transport & Road Research Laboratory, Supplementary Report.

Mackett, R. L. (1985). Forecasting the long term effects of increases in the cost of travel. *Optimization and Discrete Choice in Urban Systems*, 327-342.

MacLeod, B. W. (2003). Optimal Contracting with Subjective Evaluation. *American Economic Review, 93*(1), 216-240.

Manacorda, M., Manning, A., & Wadsworth, J. (2012). The Impact of Immigration on the Structure of Wages. *Journal of the European Economic Association , 10*(1), 120-151.

Martins, P., Piracha, M., & Varejão, J. (2012). *Do Immigrants Displace Native Workers? Evidence from Matched Panel Data.* The Institute for the Study of Labour.

Mason, W., & Watts, D. J. (2010). Financial Incentives and the "Performance of Crowds". *ACM SigKDD Explorations Newsletter, 11*(2), 100-108.

McCollom, B. E., & Pratt, R. H. (2004). Chapter 12 - Transit Pricing and Fares. In *Traveler Response to Transportation System Changes.* Washington D.C.

Metro Report International. (2014). *London goes contactless.* Retrieved from https://www.railwaygazette.com/fileadmin/user_upload/railwaygazette.com/PDF/metroreportinternational-1409-tfl-contactless-payment.pdf

Ministry of Housing, Communities and Local Government. (2015). *English Indices of Deprivation 2010.* Retrieved November 2018, from https://data.gov.uk/dataset/4c4c4e50-0925-4ae6-b45c-ca8a350bd5c6/english-indices-of-deprivation-2010

Murphy, K. J. (1999). *Executive Compensation.* University of Southern California, Marshall Business School.

Nielsen, T. A., & Hovgesen, H. H. (2008). Exploratory mapping of commuter flows in England and Wales. *Journal of Transport Geography, 16*, 90-99.

NOMIS. (2014). *Jobseeker's Allowance with rates and proportions.* Retrieved November 2018, from https://www.nomisweb.co.uk/query/select/getdatasetbygeog.asp

NOP Business & Institute for Employment Studies. (2002). *Knowledge Migrants: The Motivations and Experiences of Professionals in the UK on Work Permits.* Department of Trade and Industry & Home Office, London.

Oakley, M. (2014). *Independent review of the operation of Jobseeker's Allowance sanctions validated by the Jobseekers Act 2013.* Department for Work and Pensions, London.

Office for National Statistics. (2018). *ONS Postcode Directory (May 2018).* Retrieved November 2018, from https://data.gov.uk/dataset/596462f6-1115-401b-a8be-a66bc03ed2c7/ons-postcode-directory-may-2018

Office for National Statistics. (n.d.). Labour Force Survey, July 2001 - December 2006. Retrieved April 2016, from ukdataservice.ac.uk

Office for Rail and Road. (2018). *Station Usage Estimates.* Retrieved November 2018, from https://orr.gov.uk/statistics/published-stats/station-usage-estimates

Osborne. (2014). *Osborne Complete £10million Upgrade of Dartford Station for Network Rail.* Retrieved from https://www.osborne.co.uk/2014/04/07/osborne-complete-a10million-upgrade-of-dartford-station-for-network-rail/

Ottaviano, G. I., & Peri, G. (2012). RETHINKING THE EFFECTS OF IMMIGRATION ON WAGES. *Journal of the European Economic Association, 10*(1), 152–197. doi:https://doi.org/10.1111/j.1542-4774.2011.01052.x

Oum, T. H., Waters, W. G., & Yong, J.-S. (1992). Concepts of Price Elasticities of Transport Demand and Recent Empirical Estimates: An Interpretative Survey. *Journal of Transport Economics and Policy, 26*(2), 139-154.

Owen, A. D., & Phillips, G. D. (1987). The characteristics of railway passenger demand. *Journal of Transport Economics and Policy, 21*(3), 231-253.

Oxera. (2005). *Consumer responses: how quick are they?* Agenda.

Oyster-Rail. (2019). *When not to use Oyster.* Oyster-Rail.org.uk. Retrieved May 2019, from https://www.oyster-rail.org.uk/when-not-to-use-oyster/

Paarsch, H. J., & Shearer, B. (1997). Fixed Wages, Piece Rates, and Intertemporal Productivity: A Study of Tree Planters in British Columbia. *CIRANO Scientific Series*.

Paulley, N., Balcombe, R., Mackett, R., Titheridge, H., Preston, J., Wardman, M., . . . White, P. (2006). The demand for public transport: The effects of fares, quality of service,income and car ownership. *Transport Policy, 13*(4), 295-306. doi:10.1016/j.tranpol.2005.12.004

Peri, G., & Yasenov, V. (2015, December). The Labor Market Effects of a Refugee Wave: Applying the Synthetic Control Method to the Mariel Boatlift. *NBER Working Paper Series*(w21801).

Pollard, N., Lattore, M., & Sriskandarajah, D. (2008). Floodgates or turnstiles? Post- EU enlargement migration flows to (and from) the UK. *Institute for Public Policy Research*.

Prendergast, C. (1999). The Provision of Incentives in Firms. *American Economic Association, 37*(1), 7-63.

Preston, J. (2016). Deja Vu all over again? Rail franchising in Britain. *Research in Transportation Economics, 59*, 107-115.

Rendall, M. S., Tomassini, C., & Elliot, D. J. (2003). Estimation of annual international migration from the Labour Force Surveys of the United Kingdom and the continental European Union. *Statistical Journal of the United Nations, ECE 20*, 219-234.

Ricardo Energy & Environment. (2017). *Technical report on UK supplementary assessment under The Air Quality Directive (2008/50/EC), The Air Quality Framework Directive (96/62/EC) and Fourth Daughter Directive (2004/107/EC) for 2015.* Didcot.

Rienzo, C., & Vargas-Silva, C. (2017). Migrants in the UK: An Overview. *Migration Observatory Briefing*.

Rogers, W. H. (1993). Regression standard errors in clustered samples. *Stata Technical Bulletin, 13*, 19-23.

Rokicki, M., Chelaru, S., Zerr, S., & Siersdorfer, S. (2014). Competitive Game Designs for Improving the Cost Effectiveness of Crowdsourcing. *L3S Research Centre*, 1469-1478.

Rosenthal, M. B., Frank, R. G., Li, Z., & Epstein, A. M. (2005). Early Experience With Pay-For-Performance: From Concept to Practice. *JAMA, 294*(14), 1788-1793.

Ruhs, M. (2007). *Changing status, changing fortunes? The impact of acquiring EU status on the earnings of East European migrants in the UK.* University of Oxford, Centre on Migration, Policy and Society (COMPAS).

Ruist, J., Stuhler, J., & Jaeger, D. A. (2017). *Shift-Share Instruments and the Impact of Immigration.* Gothenburg University.

Rula, J. P., Navda, V., Bustamante, F. E., Bhagwan, R., & Guha, S. (2014). No "One-Size Fits All": Towards a Principled Approach for Incentives in Mobile Crowdsourcing. *Proceedings of the 15th Workshop on Mobile Computing Systems and Applications.* New York.

Savills. (2017). *How has the London Overground affected property prices over 10 years?* Retrieved from https://www.savills.co.uk/blog/article/225758/residential-property/how-has-the-london-overground-affected-property-prices-over-10-years.aspx

Schor, J. B. (1991). *The Overworked American: The Unexpected Decline of Leisure.* New York: New York Basic Books.

Shapiro, C. &. (1984). Equilibrium Unemployment as a Worker Dicipline Device. *The American Economic Review, 74*(3), 433-444.

Singleton, C. (2018). Long-Term Unemployment and the Great Recession: Evidence from UK Stocks and Flows. *Scottish Journal of Political Economy, 65*(2), 105-126. doi:https://doi.org/10.1111/sjpe.12139

Smith, J. A., & Todd, P. E. (2005, March-April). Does matching overcome LaLonde's critique of nonexperimental estimators? *Journal of Econometrics, 125*(1-2), 305-353.

Steer Davies Gleave. (2017). *Estimates of Station Usage 2016-2017.* London.

Transport for London. (2011). *OYSTER ON NATIONAL RAIL- PROGRESS UPDATE.* London: TfL Rail and Underground Panel. Retrieved from http://content.tfl.gov.uk/Item08-Oyster-NR.pdf

Vivalt, E. (2019). Specification Searching and Significance Inflation Across Time, Methods and Disciplines. *Oxford Bulletin of Economics and Statistics*. doi:https://doi.org/10.1111/obes.12289

Wardman, M. (2014). Price Elasticities of Surface Travel Demand: A Meta-analysis of UK Evidence. *Journal of Transport Economics and Policy, 48*(3), 367-384.

Weyerbrock, S. (1995, April). Can the European community absorb more immigrants? A general equilibrium analysis of the labor market and macroeconomic effects of east-west migration in Europe. *Journal of Policy Modeling, 17*(2), 85-120.

Zabojnik, J. (2002). The employment contract as a lottery. *University of Southern California*.

Zi-jia, W., Feng, C., B, o. W., & Jian-ling, H. (2018, September). Passengers' response to transit fare change: an ex post appraisal using smart card data. *Transportation, 45*(5), 1559-1578.