Tackling the UK’s regional economic inequality: Binding constraints and avenues for policy intervention

Anna Stansbury
MIT Sloan

Dan Turner
Harvard Kennedy School

Ed Balls
King’s College London
Harvard Kennedy School

March 2023

M-RCBG Associate Working Paper Series | No. 198
Tackling the UK’s regional economic inequality: Binding constraints and avenues for policy intervention

Anna Stansbury (MIT Sloan)
Dan Turner (Harvard Kennedy School)
Ed Balls (King’s College London and Harvard Kennedy School)

JEL Codes: O47, R12, R58

With enormous appreciation to Nyasha Weinberg and Esme Elsdon, co-authors on our companion paper, for thought partnership throughout this project and extensive comments, and to Kyra Rodriguez for excellent research assistance. We thank Jack Britton, Guilherme Rodrigues, Andrea Conte and Davide Auteri, and Nick Wilson and Marek Kacer for their help in providing data. For helpful comments, we thank Anthony Breach, Diane Coyle, Cullen Hendrix, Ron Martin, Philip McCann, Marcus Noland, Jonathan Portes, Guilherme Rodrigues, Paul Swinney, Greg Thwaites, John Van Reenen, and seminar participants at the Centre for Cities, Harvard Kennedy School, King’s College London, the National Institute for Economic and Social Research, the Peterson Institute for International Economics, and the Resolution Foundation. We gratefully acknowledge financial support from the Gatsby Foundation, the Harvard Kennedy School’s Mossavar-Rahmani Center for Business and Government, and Kings’ College London Department of Political Economy. All errors are our own. Corresponding author: Anna Stansbury amms@mit.edu.
1. INTRODUCTION

For most of the 20th century, inequality in GDP per capita between UK regions – while not insignificant – was relatively low by European standards (Rosés and Wolf 2018). In the 1980s and 1990s, however, regional economic inequality began to rise in most industrialised economies (IMF 2019). The UK stands out for how far this has developed: by the 2010s the UK had become one of the most regionally unequal of the world’s industrialised economies in terms of GDP per capita, productivity, and disposable income (McCann 2020).¹

There are three reasons to be concerned about the UK’s large and rising regional inequalities. First, inequalities in GDP per capita and productivity across regions are mirrored with inequalities in several other socio-economic indicators, including disposable incomes, life expectancy, and educational attainment (Farquharson, McNally, and Tahir 2022; Marmot 2020). Second, as in other countries, rising regional economic inequalities have been linked to a changing political geography – dubbed a ‘geography of discontent’ (McCann 2020) – which may feed political instability. Third, the UK has a large and growing national productivity problem, with a fall in productivity growth post-2007 larger than any other G7 country except Italy (Office for National Statistics 2022). If productivity can be boosted in lagging regions, this opens the possibility for gains in both equity and efficiency.²

What are the defining features of the UK’s regional economic inequality problem? In Section 2, we present five stylised facts. These illustrate that the UK’s regional economic inequality problem today is best characterised by productivity differentials between London and the greater South East of England vs. the rest, largely driven by the underperformance of non-London cities. Earlier differences in employment rates and in industrial structure across regions have narrowed, leaving within-industry productivity differentials the prime driver of divergent regional economic performance. And, while different education mixes across regions mechanically drive a large portion of average productivity and earnings differentials, there is still a large regional productivity gap even controlling for education.

These stylised facts inform the animating question for the rest of our analysis: how can policy most effectively boost productivity in the UK’s lagging regions? For this analysis, we focus on four key policy levers – education, infrastructure, support for Research and Development (“R&D”), and access to finance – analysing each in turn in sections 3 through 6.³ For each of these policy areas, we attempt to identify not just whether and to what degree each input for productivity growth is present, but also to identify whether or not that input is in relative shortage in each region as compared to its demand: whether it is a particularly

---

¹ Calculating 28 measures of regional economic inequality across industrialised economies, McCann shows that the UK is in the top half for all 28 measures, is in the top quarter for 21 of the 28 measures, and is the most unequal on five of the 28 measures. He concludes that the UK is “almost certainly the most interregionally unequal large high-income country”, and that the UK’s interregional inequality is particularly notable given that it occurs over small geographic distances.

² Moreover, localised aggregate demand externalities may lead to persistently and unnecessarily depressed regional economies; and regionally-concentrated economic shocks (like plant closures) combined with slow adjustment processes can lead to inefficiently low economic activity for long transition periods.

³ The UK government’s recent Levelling Up White Paper set out a framework of “six capitals”. Four of those are the levers we study in this paper: human capital (education), physical capital (which includes infrastructure), knowledge capital (which can be provided through R&D), and financial capital (HM Government 2022). We do not examine the remaining two capitals - social capital and institutional capital – due to the difficulty of doing so quantitatively. In our companion paper, (Turner et al. forthcoming), we analyse institutional and governance aspects of the UK’s regional economic inequality through a qualitative, contemporary historical lens.
Identification of which constraints are more or less binding is crucial for effective policy interventions. Since there are likely a large number of factors associated with growth which are lacking, and a large number of possible models through which to interpret a region’s growth problem and possible remedies to fix it, identification of which policy interventions will be most effective requires identifying not just a laundry list of all the factors that are lacking, but also a diagnosis of which of these constraints are the most binding (Rodrik 2010). And a policy intervention to increase a region’s supply of an input which is not a binding constraint will fail to generate productivity growth in that region unless the other binding constraints are also alleviated at the same time. Ideally, the process of identifying binding constraints and tackling them is iterative: the work of policy becomes to target its efforts at addressing particular bottlenecks in succession, and identifying new bottlenecks that may emerge as a result.

To identify whether an input is a more or less binding constraint, we build a collage of evidence from a range of indicators, comparing within the UK across regions and over time, and with other countries in Western Europe where possible. We then determine whether this collage of evidence is consistent with the input being a binding constraint on growth: this includes considering whether the input is in particularly scarce supply, whether there is evidence of economic actors being willing to pay a high price (in terms of money or time) to access this input, and whether there is reason to believe that increasing the provision of this input would boost productivity – or whether some other factor would still prevent growth from occurring. This collage approach to identifying more and less binding constraints on growth across an entire country can necessarily only be indicative: it must be done at a high level, across regions and indicators, with judgment required to identify the key messages from the collage of evidence. But we believe even a necessarily imprecise comparison of the available evidence on which constraints are more or less binding is important to inform – as best as possible – the prioritisation of scarce policy time and resources.

Our binding constraint analysis focuses on policy interventions to boost productivity growth in lagging regions. But even if policy fails to achieve this, one might expect regional earnings differentials to converge over time (net of moving costs and amenities) as people move from low-earning to high-earning regions. In section 7, we therefore study the UK’s internal migration flows, attempting to disentangle why regional earnings inequalities have persisted even in the face of free labour mobility across regions.

In our analyses we draw on a range of data sources, including the EU’s ARDECO database, the OECD regional and national statistics databases, the University of Gothenburg Quality of Government EU regional dataset, UK government statistics from the Office for National Statistics, the Department for Transport, the Office of Rail and Road, the Department for Education, the British Business Bank, and OFCOM, individual-level survey data from the UK’s Labour Force Survey, and data from private sector sources including UK Finance, BVA BDRC, TomTom, and INRIX. We also draw on data compiled and
analysed by other researchers: educational outcomes data from the Department for Education, compiled and analysed by Britton, Waltmann, Xu, and van der Erve (2021), travel time data from TravelTime, compiled and analysed by Rodrigues and Breach (2021), travel time data from Google Maps, compiled and analysed by Conwell, Eckert, and Mobarak (2022), private equity funding data from Beaufurst, compiled and analysed by Wilson, Kacer, and Wright (2019), peer to peer lending data from FundingCircle, compiled and analysed by Ekpu, Wright, Prashar, and Ri (2020) and Xu, Su, and Celler (2021), and academic publication data from Web of Science, compiled and analysed by György (2018). Where possible, we always conduct our analyses for the entire UK. However, some data sources are available for England only, or for England and Wales only. We make it clear where this is the case.

Our analyses in Sections 3-7 challenge a number of common arguments about the UK’s regional economic inequality problem. We find little evidence consistent with the hypotheses (i) that low shares of university graduates remain the primary constraint on growth for the UK’s regions; (ii) that there is a generalised issue with access to finance for firms outside the South East; or (iii) that low or falling regional migration rates are to blame for the persistence of the UK’s regional economic inequalities. Instead, we find evidence consistent with (i) a specific relative shortage of STEM degrees; (ii) binding transport infrastructure constraints within major non-London conurbations; (iii) a failure of public innovation policy to support clusters beyond the South East, in particular through the regional distribution of public support for Research and Development (R&D); and (iv) missed opportunities for higher internal mobility due to London’s overheating housing market. We also find some suggestive evidence consistent with constraints on access to early-stage equity financing for high-growth-potential SMEs in some of the UK’s regions, particularly the North West. We discuss each of these in more detail below.

We analyse education in section 3. While the UK’s non-London regions have lower graduate shares than London, it is possible that low population education levels are in part a result of low regional productivity, rather than its cause. We therefore use university wage premia to estimate the degree to which different graduate-level skills are in scarce supply relative to demand in each region over time. We find that in the 1990s, a low university graduate share in many UK regions seems to have been a more binding constraint on productivity growth: outside London, university wage premia were high and university graduate shares were low. But over the last three decades, the massive increase in university attainment has come alongside a substantial decline in the university wage premium in almost all regions outside London, suggesting that the expansion of higher education has alleviated the shortage of general university-level skills relative to demand. On the other hand, we find that the wage premium for university-level STEM skills has hardly fallen even as STEM attainment has risen rapidly. The STEM wage premium is now higher than the wage premium for formerly highly rewarded degrees in law, finance, and management. Together, this evidence suggests that (while there is an important mechanical role for increased educational attainment of all kinds in boosting productivity) the expansion of university-level STEM skills is likely to have higher returns for the UK’s lagging regions, as compared to a generic focus on increasing the number of university graduates. In the context of free labour mobility, we
also caution that increasing enrolment alone cannot be the solution. Net graduate outmigration from even low-enrolment regions suggests that the dominant problem for most regions is a lack of demand for graduates, more than a lack of supply of graduates.

We examine infrastructure in section 4. We show that UK cities’ transport infrastructure stands out in international context: the UK’s non-London cities have more congested roads and smaller road networks than peer US cities, and are less accessible by public transport than peer Western European cities. High levels of road and rail congestion at peak commuting times in several non-London cities suggest high demand for more and better transportation infrastructure to access city centres. And unreliable trains outside the South of England means latent demand for rail may be even higher than our congestion analysis would suggest. Improved transportation infrastructure brings benefits in terms of increasing firms’ access to larger labour markets and workers’ access to high productivity jobs: the UK’s large non-London cities’ “effective size” is therefore substantially limited by their poor transport infrastructure (alongside low-density housing). Spending data suggests more can be done: UK infrastructure spending has been very disproportionately concentrated on rail in London, and overall spending as a share of GDP, particularly on roads, has been low in international context. Transportation infrastructure spending has been disproportionately directed toward London in order to alleviate London’s high congestion and to expand the effective size of the greater London economy: by this logic, much more could be done in other cities as well.

We turn to innovation in section 5. While data directly enabling us to discern whether or not a lack of support for innovation is a binding constraint on growth is hard to come by, we show that public and higher education R&D expenditure – a key lever for policy in generating information spillovers and in coordinating high-productivity economic activity – has been relatively low in international context. Moreover, public support for R&D has been heavily biased toward the already-productive “Golden Triangle” area around London, Oxford, and Cambridge; in fact, even more biased to rich regions than private sector R&D spend (which may be seen as an indicator of a region’s absorptive capacity for innovative activity). In Germany, in contrast, public sector and higher education R&D spend – while still directed toward richer regions on average – are substantially less spatially biased than private sector R&D spend, thus helping generate convergence across places.

In section 6, we analyse access to finance. We find limited evidence that a lack of access to finance is a barrier to productivity-enhancing investment in non-London regions. Across a wide range of indicators, we find no evidence of differential access to bank lending for SMEs outside London as compared to SMEs in London. And while there is a large equity funding gap between London and non-London based SMEs, the vast majority of this gap can be explained by differential business characteristics across regions. But there is some indication of constraints on access to equity financing for high growth SMEs in certain regions: SMEs in the North of England and the Midlands are substantially less likely to receive equity financing than observably equivalent firms in London. The extent to which this is driven by differential growth prospects, as opposed to investor bias or missed opportunities, is not clear from available data. More research is warranted to understand the degree to which, and the places in which, this is the case.
Finally in section 7, we consider the role of migration in equalizing incomes across regions. Under most models of spatial equilibrium, labour mobility reduces the degree of regional economic inequality caused by differential shocks across place, as people leave lower-earning regions for higher-earning regions. We find no evidence to suggest that the persistent income differentials between the UK’s regions are a result of a low propensity for UK residents to move. In fact, UK interregional migration is relatively high in international context and has not declined over time (unlike in the US). Instead, evidence indicates that investing in housing in London and the South East would contribute to easing regional divides by easing the barriers to labour mobility. Interregional mobility in the UK goes in the wrong direction: people on net move away from high-productivity London to other regions, as high housing costs in London, and to a lesser extent the South East, erode the London wage premium for most of the income distribution, making the net return to migration to London small or negative (and therefore limiting the opportunity to benefit from London’s productivity to either the highly educated, or to those who happened to own property there already). This suggests a clear role for policy in alleviating London’s housing crisis.

In Section 8 we conclude by reviewing the contemporary policy debate in the UK in light of our findings. Arguments in this debate tend to fall into three camps – boosting the inputs for neoclassical growth (education, infrastructure, access to finance); boosting government’s role in endogenous growth (R&D, coordination of economic activity, new industrial policy), or boosting internal migration. Our analysis suggests that each of these views has merits. Yet our analysis also makes clear that, outside education policy, which has seen very large boosts to education and skills, the large-scale, systematic, and consistent policy action necessary for any of these has not taken place in the UK in recent decades. In a world of uncertainty about the true dominant underlying economic model, and in the absence of silver bullet evidence for any single binding constraint on productivity growth in lagging regions, policymakers seeking to tackle regional economic inequality would therefore do well to seek to simultaneously alleviate each of the constraints we have identified: increasing attainment in specific skills which are in short supply in certain regions, particularly STEM; increasing transport investment across the UK outside London and the South East, with a priority on highly congested urban areas with potential to benefit from agglomeration economies; increasing government R&D expenditure outside London and the South East, in areas which are or have the potential to become clusters of excellence in particular fields; and increasing efforts to reduce housing costs and increase housing availability in London and the greater South East.

---

4 The neoclassical view is inspired by the neoclassical growth model, and posits that government’s role is to provide certain core inputs to growth: specifically, human capital, infrastructure, and financial capital. Conditional on these inputs being present, growth should occur and regions should converge to similar levels of output and productivity (Solow 1956; Mankiw, Romer, and Weil 1992; Holtz-Eakin and Schwartz 1995). The endogenous growth view is inspired by endogenous growth theory and new economic geography (Romer 1994; Krugman 1991), and emphasizes a central role for information and coordination externalities, as well as path dependence, in determining a region’s growth trajectory. The role of government, beyond ensuring access to human, physical, and financial capital, is to generate knowledge spillovers and coordinate economic development activity. The migration view relies on it being too difficult for policy to stimulate growth in lagging regions (perhaps because path dependence is too important and new agglomerations too hard to develop) – in this case, the appropriate role for government is to enable people to move to the productive places (and alleviate hardship for those who remain).
How much could this help? Regional economic inequalities have been rising in most industrialised economies. The UK had a particularly steep legacy of deindustrialisation, and was particularly affected by London’s rise as a global centre of financial and professional services. The effects of major global economic trends are unlikely to be able to be countered fully even by major policy efforts. But our analyses of UK and international policy decisions suggest that much more can be done. Boosting productivity in the UK’s lagging regions is not a sufficient condition to fix regional economic inequality: inequality in incomes and living standards is high even in the UK’s richest regions, and policymakers concerned with poverty and inequality must ensure that any boost to productivity growth is broadly shared. But while it may not be sufficient, boosting productivity is a necessary step to create a sustained, meaningful improvement in the economic prospects of the UK’s lagging regions.

2. WHERE ARE WE, AND HOW DID WE GET HERE? FIVE STYLISED FACTS

In this section, we document five stylised facts about regional economic inequality in the UK today, where we are now, and how we got here. In the past, the UK’s regions differed substantially in their industrial structures, making the UK particularly exposed to the global decline of heavy industry and rise of finance and business services. Rapid deindustrialisation then led to large differences in employment and unemployment rates across regions. Today, however, the UK’s regional economic inequality is best characterised by large and persistent productivity differentials between London and the greater South East of England vs. the rest of the country, largely driven by the underperformance of cities. This characterisation of the situation informs our approach in the rest of the paper: an analysis of factors with the most potential to boost productivity in regions outside London and the South East.

Stylised Fact 1: The UK’s regional economic inequality problem is primarily a productivity problem, not an employment problem. In the 1980s, the defining regional economic disparity in the UK was in employment and unemployment (Balls, Katz, and Summers 1991; Evans and McCormick 1994). As of the 2010s, while there is still evidence of employment rate scarring from the 1970s/80s (Rice and Venables 2021; Beatty and Fothergill 2017), UK regional inequality in employment rates is now relatively low in international context and has been falling.\(^5\) In contrast, regional inequality in productivity, as measured by GVA per worker, is high by international standards and has been rising, as shown in Figure 1. Regional inequality in the UK today is less about fewer people working in different regions, and more about people working at similar rates and generating less value added.

\(^5\) The US in comparison has seen large and widening regional employment differentials (Austin, Glaeser, and Summers 2018).
Stylised Fact 2: The UK’s regional economic inequality is best characterised as London and the South East vs. the rest of the country (McCann 2020; Carrascal-Incera et al. 2020; Harris and Moffat 2022). As Figure 2 illustrates, since 1980 the rise in regional inequality as measured either by GDP per capita or GVA per worker (productivity) has been driven by London and the South East, which started off substantially richer and more productive than the rest of the country and grew faster than almost any other region. Among the rest of the UK’s regions, if anything, there was a small trend towards convergence over time. Today, the gap between London and the South East, vs. the rest of the UK, is extremely large in international context: larger than the gaps between East and West Germany – despite the legacy of the GDR – or North and South Italy. In 2019, GVA per worker in East Germany was 80% of that of West Germany, and GVA per worker in South Italy was 78% of that of North Italy. In the same year, GVA per worker in the rest of the UK was 71% of that of London and the South East. This is visualised in Figure 3: in 2019, London stands out with particularly high productivity; the South East also has productivity a little higher than the national average; the East of England and Scotland follow with productivity at 93% and 95% of the national average respectively, and the rest of the UK regions have productivity between 80% and 90% of the national average.

Source: ARDECO. Note: Regions defined at NUTS 1 level. We do not include Germany data pre-reunification.

---

6 We calculate these figures using the EU’s ARDECO database, which provides GDP and population statistics for every EU NUTS region since 1980. We define the regions as follows. North Italy: ITH, ITI, and ITC (Nord-Est, Nord-Ovest, and Centro). East Germany: DED, DEE, DEG, DE4, and DE8 (Sachsen, Sachsen-Anhalt, Thuringen, Brandenburg, Mecklen-Vorpommern). (We exclude Berlin from both East and West Germany due to its split status prior to German reunification). We define London and the Greater South East as NUTS regions UKH, UKI, and UKJ: London, the South East, and the East of England.
Figure 2. Growth in GDP per capita and GVA per worker, 1980-2019, UK regions

Source: ARDECO

Figure 3: GVA per worker relative to national average, 2019

Source: ARDECO; Authors’ calculations. Note: Regions are red if GVA per worker is below the national level and blue if GVA per worker is above the national level.
Stylised Fact 3: The underperformance of regions outside London and the South East is largely driven by underperforming cities. The UK’s cities – outside of a handful of Southern exceptions, most notably London – do not appear to benefit from the agglomeration economies seen in other industrialised countries, where scale and population density are strongly associated with higher productivity (Özgüzel 2020a; T. Forth 2017; OECD 2015; McCann and Yuan 2022). This is illustrated in Figure 4, which shows that Manchester and Birmingham have particularly low productivity for their population size. As a result, cities outside the South East contribute less to the economy than we would otherwise expect. The underperformance of large cities outside London and the South East also means that the growing urban/non-urban divides seen in other countries are less stark in the UK: in fact, small towns and rural areas are often more prosperous than nearby cities (McCann 2020; Martin et al. 2021). While underperforming cities are a central part of the story, for maximum comparability across datasets and countries in our analysis we primarily focus on regions (defined at the NUTS1/TL2 level). Large metropolitan clusters or ‘city-regions’ make up a majority of the population in many of the UK’s regions.8

Figure 4. Productivity and city populations, Western Europe

Source: Recreated from Rodrigues and Breach (2021). Original data source: OECD.

Stylised Fact 4: Education mix is important – but it’s not all about education.

Different education and skill mixes play a central role in productivity differentials across regions. Since more highly educated residents earn more, different education levels across regions can mechanically explain a large share of the regional inequality in incomes and productivity in the UK.9 Boosting education levels in non-London regions, both in terms of

---

7 Özgüzel 2020a finds an elasticity of urban area productivity with respect to employment density of 0.01, which is at the low end of estimates from other countries (including Spain, France, Italy, and the Netherlands).
8 58% of the North West’s population is in the Greater Manchester and Merseyside metropolitan counties, 49% of the West Midlands’ population is in the West Midlands metropolitan county (containing Birmingham), 43% of the North East’s population is in the Tyne and Wear metropolitan county (made up of Newcastle, Gateshead, Tyneside, and Sunderland), and 35% of Yorkshire and the Humber’s population is in either the West Yorkshire metropolitan county (containing Leeds), or the South Yorkshire metropolitan county (containing Sheffield). Population figures from the Office for National Statistics’ 2021 Census data release.
9 Gibbons, Overman, and Polkön (2014) find that over two-thirds of the spatial variance of wages in the UK is explained by worker characteristics and occupations. See also McCann and Vorley (2020) and Overman and Xu (2022).
tertiary and secondary qualifications, would therefore almost certainly boost productivity mechanically (see e.g. Machin and Vignoles 2018). But wage differentials suggest that regional productivity disparities are not solely explained by education mix. To illustrate this, we estimate region wage premia, relative to the North East, for private sector workers aged 25-59 in each UK region over 2010-19 using the Labour Force Survey. Our coefficient estimates are shown in Figure 5, left panel. Specifically, we regress the log of hourly wages on indicators for each region of workplace, and fixed effects for age, sex, and year in our baseline specification (blue), and run a further regression with fixed effects for highest educational qualification (red). As the figure illustrates, even conditional on education, wages in the 2010s were somewhat higher than average in the East and the South East, and substantially higher in London, with workers in London earning 29% more than workers of the same age, sex, and education level in the North East.\(^\text{10}\) Notably, these wage differentials across regions have been relatively consistent for each region over the last two decades (illustrated by the comparison with Figure 5, right panel, which shows estimates for 1997-99), suggesting the factors driving the productivity disparities across regions are persistent.

Figure 5. Region wage premia, relative to the North East

![Figure 5](image_url)

**Fixed effects:** Age, sex, education, industry, occupation

Source: Labour Force Survey. Note: Dots represent coefficient estimates (and bars 95% confidence intervals) in a regression of log hourly wage on region dummies. All specifications have sex, age, and year fixed effects. Sample: 25-59 year olds, private sector. Each specification progressively adds more fixed effects: blue shows coefficient estimates for a specification with age, sex, and year fixed effects, red adds fixed effects for highest level of education fixed effects, green adds 2-digit industry fixed effects, and orange adds 4-digit occupation fixed effects. Robust standard errors clustered at region level.

\(^\text{10}\) Note: throughout this paper we use the percentage approximation for log point differentials. Note also that region wage differentials may partly be driven by sorting of individuals with high unobservable productivity into high wage places. While we are unable to estimate this in our data, Overman and Xu (2022) estimate Mincer-type wage premia at the TTWA level, alongside estimating TTWA wage effects when controlling for individual worker fixed effects (which controls for unobservable individual characteristics which may be correlated with earnings potential). They find that 1/3 of the max-min difference in region wage premia for TTWAs, as estimated in a Mincer regression, is actually explainable by sorting, and the remaining 2/3 represents a true region effect. Applying this to our estimates would suggest a roughly 20%, rather than 29%, region wage effect for London as compared to the North East.
Stylised Fact 5: Industry is important, and was a major driver of the UK’s regional economic inequalities - but it’s no longer primarily about industry mix. The UK’s regional inequalities today are a result of particularly extreme exposure to two global economic trends: deindustrialisation, and the rise of the knowledge economy. In 1980, the UK had some of the most industrialised regions in Western Europe: the East and West Midlands had the highest employment shares in manufacturing and mining, with three other UK regions – Yorkshire and the Humber, Wales, and Northern Ireland – in the top sixteen. At the same time, London, the South East, and the East of England were second, third, and sixth across Western European regions in terms of their employment shares in finance and business services. As finance and business services became the engine of productivity growth across the industrialised world (Ehrlich and Overman 2020a), London and the greater South East were uniquely poised to benefit; meanwhile, as automation and globalisation led to deindustrialisation, the Midlands and North of England, Wales, and Northern Ireland were uniquely exposed.

Moreover, deindustrialisation in the UK was particularly rapid compared to peer regions, even given its initial exposure (Figure 6). The West Midlands, for example, saw its employment share in manufacturing and mining fall by 30 percentage points, and the East Midlands and Yorkshire and the Humber by around 25 percentage points, over 1980-2018. The five other Western European regions with comparably high manufacturing and mining employment shares in 1980 – North-Rhine Westphalia, Baden-Wuerttemberg, and the Saarland in Germany, North-Western Italy, and North-Eastern Spain – saw much smaller falls of 14-20 percentage points (Figure 7). Indeed, the only other EU regions which have ever seen a ten-year period of deindustrialisation as fast as the experience of the Midlands or Yorkshire in the 1980s were regions in formerly Communist countries in the decade directly after the transition to capitalism. These large, regionally-concentrated collapses in employment led to long-lasting scarring in labour market outcomes (Martin et al. 2021), and appear to have led to a loss in economic complexity and external trade, which likely had long-lasting productivity effects (Rice and Venables 2021; 2022).

---

11 London, the East, and the South East of England were all also in the lowest third by employment share in manufacturing and mining.
12 In 1980 there was little correlation between a region’s employment share in finance and business services, and its productivity; by 2018, this correlation was very strong across Western European regions, with an R-squared of 55% (see Appendix Figure 1).
13 The full list is in Appendix Table 1. The other European regions which saw a decade of deindustrialisation as fast as the Midlands or Yorkshire were four former East German regions (1991-2001), three Romanian regions (1990-2000), and Lithuania (1990-2000).
Figure 6. Deindustrialisation in Western European regions, 1980-2018

Source: ARDECO. Note: Gray denotes line of best fit across all regions; red line denotes line of best fit for UK only. “Other” includes NUTS1 regions in Austria, Belgium, Denmark, Finland, Greece, Ireland, Luxembourg, Netherlands, Norway, Portugal, Spain.

Figure 7. Deindustrialisation in Western Europe's sixteen most industrialised regions, 1980-2018

Source: ARDECO. Note: Figure shows the sixteen Western European regions which had the highest employment share in manufacturing and mining in 1980.
The UK’s regional economic inequalities are thus in large part a result of regions’ different industry mix and the experience of these industries over recent decades. At the same time, one of the effects of deindustrialisation and the rise of the knowledge economy has been a partial convergence of the sectoral structure of the economy across UK regions, as illustrated in Figure 8. This convergence means that differential within-sector productivity, particularly in finance and business services, is now a much bigger contributor to the productivity gap between London and non-London regions than differential industry composition. Two exercises support this conclusion. First, large region wage differentials persist even when we add fixed effects for 2-digit industry codes in our estimates of region wage premia from the Labour Force Survey, illustrated in the green dots in Figure 5, left panel. The addition of industry fixed effects has almost no effect on the estimated region wage differentials once we control for education, suggesting that differential industry composition can explain only a small fraction of region wage differentials. Second, a simple counterfactual exercise where each region is given London’s 2018 industry composition, but keeps its own region’s industry productivity, does almost nothing to close the regional productivity gap, as illustrated in Figure 9. This is because, while the remaining manufacturing in formerly industrialised regions is highly productive, the finance and business services sector in these regions is not.\footnote{\par Indeed, manufacturing and mining productivity is particularly high in the UK regions which saw the most deindustrialization, as illustrated in Appendix Figure 4. Differential within-industry productivity in the knowledge economy across regions fits with evidence that the share of firms belonging to the national “productivity frontier” is the highest in Greater London, followed by the South East, with lower values for the rest of the regions (Haldane 2016; Kierzenkowski, Gal, and Fulop 2017). The limited role for industry composition in explaining regional productivity differences remains true if you disaggregate into smaller industry categories (Zymek and Jones 2020; Oguz 2018), or if you analyse cities rather than regions (R Martin et al. 2019). Some, but not all, of the gap in within-sector productivity can be closed by accounting for differential occupational mix across regions (Beatty and Fothergill 2019).}

\textbf{Figure 8: Regional employment shares in manufacturing/mining and finance/business services, 1980 and 2018}

\begin{center}
\includegraphics[width=\textwidth]{Figure8.png}
\end{center}

Source: ARDECO. Regions sorted by 1980 manufacturing & mining employment share.
Figure 9. Regional GVA per worker, 2018 - actual and counterfactuals

Source: ARDECO; Authors’ calculations. Note: This reports three counterfactuals for regional GVA per worker in 2018. (1) Red bars: each region is assigned London’s employment shares for each sector, but retains its sector productivity. (2) Green bars: each region is assigned London’s productivity for each tradable sector – defined as manufacturing and mining, and finance and business services – but retains its employment shares. (3) Orange bars: each region is assigned London’s productivity for all sectors, but retains its employment shares. GVA per worker is reported in thousands of 2015 EUR.
3. EDUCATION

Informed by the stylised facts in section 2, we now begin our analysis of binding constraints on productivity growth in the UK’s lagging regions, and possible areas for promising policy intervention. We start with education. Education measures – both the share of the population with a tertiary education, and the share with a good secondary qualification – are extremely strong correlates of regional productivity, and can explain mechanically much of the average income differences between regions (see e.g. Overman and Xu 2022). But, as we illustrate in Figure 5, there are large regional wage differentials even conditional on education level. While there has been a rapid increase in educational attainment across all UK regions over recent decades, both in terms of secondary and tertiary qualifications, this increase has happened in London as well as outside London, meaning that in relative terms UK regions outside of London still have less educated populations. This has led to calls for upskilling, or for the reversal of the “graduate drain” (trying to convince graduates to remain in non-South East parts of the UK) as a solution to regional economic disparities. But, while increasing educational attainment will surely boost productivity mechanically, its effects may be weaker than anticipated if a lack of education and skills is not the most binding constraint on firms’ growth and expansion. Low population education levels may in part be an outcome of low regional productivity – if a lack of well-paid graduate jobs, for example, reduces the incentive to go to university or pushes graduates to leave the region to seek employment outside – rather than a cause of it. If this is the case, increasing education levels in lower-productivity regions without other reforms alongside may simply lead either to underemployment or outmigration for highly-educated people from non-London regions.

One indicator of whether or not a factor of production is a particularly binding constraint on firms’ productive economic activity is its price: a high price suggests the factor of production is in high demand relative to its supply, and a rising price suggests the factor is becoming increasingly scarce relative to demand (or increasingly in demand relative to its supply). In the case of education, we can infer the market price for skills, and how they have changed over time, by estimating wage premia for people with different types and levels of education. For example, by comparing the university wage premium across regions (the premium paid for a university graduate of a given age and gender, relative to someone with a school leaving qualification), we can infer the relative scarcity of university graduates in some regions relative to others. All else equal, a higher university wage premium in one region than in another would indicate that the demand for university graduates from employers in that region is particularly high relative to their supply. Similarly, by studying the evolution of the university wage premium over time, we can infer how this relative scarcity has changed across regions over time (See Box 1).

---

15 See Appendix Figure 2. In the UK, much of this correlation is driven by London and the South East vs the rest. Outside London and the South East, there is wide dispersion in education levels across regions, but relatively small dispersion in GDP per capita.
16 Over 2000-2018, the share of the 25-64 year old population with a tertiary qualification increased not just in every large region but also in every smaller (NUTS2) region by between 9 and 19 percentage points, and the share without a good secondary qualification decreased by between 8 and 21 percentage points (Appendix Figure 3). Education spending is also relatively evenly distributed across UK regions, according to Office for National Statistics data.
17 The university wage premium is an outcome of the relative supply of and demand for university graduates, as well as labour market institutions (like unions) which can alter the pay distribution. Differences in the university wage premium across regions and over time can be used to infer the relative scarcity of university graduates under the assumption that labour market institutions are constant.
Box 1: Inferring the relative price of university skills from the university wage premium

The logic of our education wage premia analysis is based on the “race between education and technology” model of education wage premia, as found in Goldin and Katz (2010) and Autor, Goldin, and Katz (2020). We augment their model in two minor ways: (i) adding a quality parameter for the efficiency per unit of labour of each group, to allow for changes in the average unobserved ability or education quality within each education group over time, and (ii) adding a wedge between the wage and the marginal product, to allow for market power or labour market rules or institutions to play a role in wage-setting. Our augmented version of their model features output as a CES aggregate of two factors, university-educated workers (A) and non-university-educated workers (B), who perform imperfectly substitutable tasks:

\[ Q_t = [\alpha_t (\alpha_t e_{At} L_{At})^\rho + (1 - \alpha_t (b_t e_{Bt} L_{Bt})^\rho)]^{\frac{1}{\rho}} \]

where \( L_{At} \) and \( L_{Bt} \) are the quantities of university-educated labour and non-university-educated labour employed in period \( t \), \( e_{At} \) and \( e_{Bt} \) represent the efficiency of each unit of university-educated and non-university-educated labour respectively, which can come either from intrinsic ability or from the quality of the education received, \( \alpha_t \) and \( b_t \) represent university-educated and non-university-educated labour augmenting technological change, and \( \alpha_t \) represents the share of work activities allocated to university-educated labour. We can express the aggregate elasticity of substitution between university-educated and non-university-educated labour, \( \sigma_{AB} \), as a function of CES parameter \( \rho \) as follows: \( \sigma_{AB} = \frac{1}{\rho} \).

We assume that the wage for each group of labour \( i \in \{A, B\} \) is set as the product of its marginal product \( MPL_i \) and some markdown \( \mu_i \) (which might reflect product or labour market power of the firm, or labour market institutions like unions):

\[ w_{At} = \mu_{At} MPL_{At} \quad , \quad w_{Bt} = \mu_{Bt} MPL_{Bt} \]

such that the log ratio of the wages of the two skill groups is given by

\[ \ln \left( \frac{w_{At}}{w_{Bt}} \right) = \ln \left( \frac{\alpha_t}{1 - \alpha_t} \right) + \left( \frac{\sigma_{AB} - 1}{\sigma_{AB}} \right) \ln \left( \frac{\alpha_t}{b_t} \right) + \left( \frac{\sigma_{AB} - 1}{\sigma_{AB}} \right) \ln \left( \frac{e_{At}}{e_{Bt}} \right) - \ln \left( \frac{L_{At}}{L_{Bt}} \right) + \ln \left( \frac{\mu_{At}}{\mu_{Bt}} \right) \]

where \( \alpha_t/(1 - \alpha_t) \) and \( \alpha_t/b_t \) represent the relative skill demand created by the technology in the aggregate production function, \( e_{At}/e_{Bt} \) the relative efficiency/quality of a unit of labour from each group, \( L_{At}/L_{Bt} \) the relative quantity of each group, and \( \mu_{At}/\mu_{Bt} \) the relative wedge between wages and marginal product for each skill group. We can therefore interpret differences in the university wage premium across regions (within a time period), or over time (within a region), based on differences in these three terms. The below expression illustrates differences over time within a region using a log approximation to percentage changes:

\[ \%\Delta \left( \frac{w_{At}}{w_{Bt}} \right) = \%\Delta \frac{\alpha_t}{1 - \alpha_t} + \left( \frac{\sigma_{AB} - 1}{\sigma_{AB}} \right) \%\Delta \frac{\alpha_t}{b_t} + \left( \frac{\sigma_{AB} - 1}{\sigma_{AB}} \right) \%\Delta \frac{e_{At}}{e_{Bt}} - \left( \frac{1}{\sigma_{AB}} \right) \%\Delta \frac{L_{At}}{L_{Bt}} + \%\Delta \frac{\mu_{At}}{\mu_{Bt}} \]

This gives us four reasons why the (private sector) university wage premium in a given region might change over time:

1. \%\Delta \frac{\alpha_t}{b_t}: a change in the relative demand for university versus non-university educated labour (caused for example by changes in technology or industry structure)
2. \%\Delta \frac{L_{At}}{L_{Bt}}: a change in the relative supply of university vs non-university educated labour across regions (to compare levels of wage premia across regions), and that these labour market institutions change in the same way across regions over time (to compare changes in wage premia over time).
3. $\% \Delta \frac{\mu_{A}}{\epsilon_{B}}$: a change in the relative quality of university vs non-university educated labour (caused for example by changes in the quality of degrees or high school, or changes in sorting by underlying ability into university).

4. $\% \Delta \frac{\mu_{A}}{\mu_{B}}$: a change in the relative wedge between wage and marginal product.

If we take a double difference – the difference in the change over time in the university wage premium across regions – we can argue that differences are given by changing supply and demand patterns under two assumptions. (1) The change in $\mu_{A}/\mu_{B}$ was the same across regions – there was no differential change in the relative wedge between wage and marginal product affecting university-educated vs non-university-educated workers across regions. This is consistent with policy around unions, minimum wages, etc being set nationally in the UK, and low levels of private sector unionization and collective bargaining coverage. (2) The change in $\epsilon_{A}/\epsilon_{B}$, the average quality of university educated vs. non-university-educated workers (where quality is defined in its relation to productivity), was the same across regions. While it is very possible that there are differences across regions in this parameter (given London’s high wage premia and high concentration of elite jobs in the private sector and government, the highest productivity university graduates are particularly likely to sort into London), and it is possible that this parameter has changed over time (as the expansion of university education disproportionately increased enrolment of students in lower-ranked universities), it seems quite plausible to expect the double difference in this variable to be zero. This would imply that there has been no change in the degree to which the top graduates sort into London over the 1990s–2010s, and that there was no differential change in the average quality of university graduates across regions over the 1990s-2010s. If this double difference is plausibly zero, we can then use the change in university wage premia over time across regions to infer the different relative demand shifts for university graduates across regions.

We use the Labour Force Survey to estimate the university wage premium within each region for 25-59 year old private sector employees, relative to someone with A-levels only.\(^\text{18}\)

In Figure 10, we plot these estimated university wage premia for each region against the share of private sector employees in that region with each of these types of degree (on a log scale). The figure shows that all regions substantially increased their university graduate share from 1997-2019, and – in all regions except London – this came alongside a large decline in the university wage premium. Most regions see a very pronounced downward-sloping line, suggestive of a demand curve: as the relative supply of university graduates increases in a region, the relative price of a university graduate as compared to a non-graduate falls. London is a stark exception: it saw a large increase in its university graduate share, but no decline at all in its university wage premium over the period.\(^\text{19}\) In most UK regions in 1997, the

---

\(^{\text{18}}\) Specifically, we regress the log of the gross hourly wage on an indicator for whether or not someone has a university degree interacted with their region of workplace, with dummies for age and gender, and region-by-year fixed effects. We also incorporate dummies for highest level of educational attainment, where the omitted category is A-levels (i.e. dummies for non-university tertiary education, GCSEs, other qualifications, or no qualifications). Robust standard errors are clustered at the region level. Note that all individuals are included in these regressions, including immigrants. Non-UK born individuals made up over 14% of the UK population by 2019 (Wadsworth 2019); running these regressions with only the UK-born gives us very similar results.

\(^{\text{19}}\) This is interesting in the context of work by Blundell, Green, and Jin (2022), who find that the ratio between median log hourly wages for university graduates and those with one GCSE A*-C grade in the UK did not fall over 1993-2016, controlling for age and sex. They argue that the absence of any decline in the graduate wage premium is a puzzle in light of the large increase in the UK’s
university wage premium was higher than it was in London; in contrast, by 2019 the university wage premium was much higher in London than in any other UK region, at about 40% in London compared to 30% in most other regions. This pattern of university wage premia, holding all else constant, would suggest that (i) in the 1990s university graduates were relatively scarce across the UK, (ii) that the mass expansion of university education in the last two decades alleviated this scarcity relative to demand for graduates for all regions outside of London, and (iii) that even despite London’s very high graduate share, by the late 2010s university graduates were scarcer, relative to demand, in London than in the rest of the UK.

Figure 10. University wage premia across regions, 1997-2019

Source: Analysis of UK Labour Force Survey. Note: “5yma” refers to five year moving average. Each point is the estimated wage premium for university graduates relative to A-level recipients, and the share of university graduates among private sector employees in a region in each year 1997-2019.

We then use the same data and methodology to estimate the university wage premium separately for three groups of degrees – STEM (science, technology, engineering, and mathematics), LFM (law, finance, and management), and all other degree subjects. We do this for two reasons. First, the skillsets conferred by STEM and LFM degrees may be differently in demand from employers. Second, we may be concerned that the average value of work skills conferred by university degrees may have fallen as university education
expanded, both in terms of numbers and in terms of types of subjects offered. The average quality of a STEM or LFM degree may be expected to have remained more constant over the period. Figure 11 illustrates the results, plotting the estimated university wage premium for each year 1997-2019 for each of STEM, LFM, and other degrees, against the share of 25-59 year old private sector employees in each region who have those degrees. For other degrees (neither STEM nor LFM, shown in blue), we see the same kind of pattern across regions and over time as we saw for all degrees in Figure 10: almost all regions outside London saw a large increase in university attainment and a large decline in university wage premia, while London saw a large increase in university attainment but no decline in the university wage premium. Indeed, the university wage premium outside London for non-STEM, non-LFM degrees is now quite low at around 20%. For LFM degrees, the pattern over time is similar: while the estimates are noisier as these represented only around 15% of degrees in 2019, most regions saw an increase in LFM degree attainment and a decline in the LFM wage premium, again seeming to trace out demand curves (in green). LFM degree wage premia in regions outside London fell from around 50-60% in the late 1990s to between 35% and 45% by the late 2010s.

For STEM degrees, in contrast, the pattern looks different. First, the STEM wage premium is higher than the LFM wage premium in most regions, and much higher than the wage premium for other non-STEM and non-LFM degrees. Moreover, as STEM degree attainment has increased the STEM wage premium has fallen much more slowly than it has for LFM degrees – in fact, it rose in London and hardly fell at all in the East or West Midlands, the North West, or Yorkshire and the Humber. Using the same logic as in the previous analysis, this suggests that STEM degrees are scarcer relative to their supply than other degrees, and that – unlike for LFM degrees – the increase in supply over the last twenty years has not succeeded much at alleviating this shortage. This suggests that demand for STEM degrees has been increasing just as fast as supply in most regions. We can illustrate this by estimating region-specific average annual growth in relative demand for these different degree categories by regressing the wage premium for each qualification on the workforce share with that qualification, with region-specific time trends. Following Goldin and Katz (2010), we can think of the coefficient on the region-specific time trend as a proxy for the average growth in relative demand for each of these skills, as compared to high school leavers. Figure 12 illustrates the coefficients on these region-specific time trends, suggesting that in most regions, the relative demand growth for STEM degrees has been much greater than for LFM degrees or other degrees.

22 Specifically, we regress the 5-year moving average of the wage premium for (e.g.) STEM degrees on the 5-year moving average of the log of the share of 25-59 year old private sector employees with a STEM degree in that region, with region dummies and region-specific time trends.
**Figure 11. University wage premia for STEM, LFM and other degrees**

Source: Analysis of UK Labour Force Survey. Note: STEM stands for Science, Technology, Engineering, and Mathematics. LFM stands for Law, Finance, and Management. Other degrees are all other subjects that are not STEM or LFM. Each point is the estimated wage premium for (STEM/LFM/other) degrees relative to A-levels, and the share of employees who have that type of degree in a region in each year 1997-2019. The most recent year, 2019, is marked with an X.

**Figure 12. Estimated annual demand growth for different degrees, by region, 1997-2019**

Source: Analysis of UK Labour Force Survey. Note: Points represent coefficient estimates (and lines 95% confidence intervals) on the region-specific time trends from a regression of the estimated university wage premium for a (STEM/LFM/other) degree in the region in question on the log of the share of private sector employees with that degree in that region, and region-specific time trends. Under certain assumptions this coefficient can be interpreted as the annual increase in relative demand for workers with (STEM/LFM/Other) university degrees.
Overall, therefore, the facts (i) that university wage premia are so much lower outside than in London on average and (ii) that university wage premia outside London have fallen so much since the 1990s as the supply of university graduates rose, together suggest that the massive expansion of university education in the UK helped alleviate the degree to which a lack of university graduates in general was a binding constraint on growth in most UK regions outside London. However, the facts (iii) that STEM-specific wage premia are consistently much higher than for the average degree subject and (iv) that STEM wage premia have fallen much less over time (even compared to other high-wage degrees in law, finance, and management) suggests that a shortage of STEM-specific degree-level skills may still be a binding constraint on growth across most UK regions (including London, which has seen a large increase in the STEM wage premium since the 1990s).

Note that our interpretation above, of differentials across regions and over time in university wage premia, relies on the average relative quality of the skills not changing differentially across regions over time (as we lay out in Box 1 in more detail). To the extent that rising educational attainment pulls in new students to get degrees who had lower intrinsic ability than the average for university graduates in prior years, or to the extent that rising educational attainment has lowered the average quality of the education provided, this – rather than shifts in relative supply and demand for skills – could be causing the observed university wage premium to fall (see e.g. O’Leary and Sloane 2005). We think it is unlikely that this is driving our results, for three reasons.

First, to explain the fact that university wage premia declined in non-London regions but not in London, sorting on unobserved ability or changes in the quality of degrees would have to be happening differentially in vs. outside London, since the expansion of university attainment was of a similar magnitude across all regions.

Second, this change in ability must happen differentially across the two groups considered. When estimating the university wage premium relative to A-level recipients, expansion of university education may be expected to reduce the average intrinsic ability of university graduates, but by the same logic it would also be expected to reduce the average intrinsic ability of those who leave school after A-levels (since university expansion would tend to remove higher ability people from this pool). In their analysis of the national UK university wage premium, Blundell, Green, and Jin (2022) find that the 50-10 wage differential among university graduates in the UK has not changed over time, while a deterioration in the quality of university graduates caused by increasing numbers of lower-ability individuals with degrees would be expected to widen this differential. They also carry out a detailed bounding exercise on the degree to which sorting on unobserved ability would have been expected to reduce the measured university wage premium and conclude that even under quite extreme assumptions, it would be small. Moreover, to the extent that improved primary and secondary schooling increased students’ preparation for university or increased opportunities for intrinsically talented students to access university, one might expect the observed university wage premium to rise.
Third, our estimate of university wage premia for LFM degrees – where one might expect the quality to have remained more constant over time – also finds large decreases in the wage premium as the share of the workforce with these degrees rose, with the slope of the relationship strikingly similar to that of other degrees.

Nonetheless, to directly estimate the pure effect of changes in relative supply and demand holding the skills of individuals and quality of education constant would require a detailed analysis of returns to individual university courses across regions and over time, with high quality controls for prior academic preparation and ability. We leave this for future research.

Next, we perform a similar wage premium analysis for advanced further education – a non-university tertiary qualification – once again estimating the wage premium relative to someone with an A level or equivalent. Our estimates of region-specific wage premia for advanced further education, and the share of private sector 25-59 year old employees in each region with advanced further education as their highest qualification, are shown in Figure 13. In every region the wage premium for advanced further education has declined substantially, by around 10 percentage points, even though in most regions there has been very little increase in the share of the workforce with these qualifications. (Scotland and Northern Ireland are exceptions).

The uniform decline of the advanced further education wage premium across regions, without much increase in the supply of advanced further educated workers, suggests a decrease in private sector demand for these qualifications since the late 1990s. This could be a result of a decline in the relative usefulness (productivity) of the skills provided by advanced further education, or it could be the result in a decline of the quality of the education provided. Alternatively, it could be that as university education expanded, the signalling value of an advanced further education qualification fell in the eyes of employers. In any case, it suggests that the average value private sector firms attach to advanced further education qualifications has fallen substantially in recent years, suggesting that a shortage of these skills as they are currently taught is unlikely to be an important constraint. Note that it is still possible that other non-university tertiary skills may be important – and we are unable to examine whether there is a relative shortage for certain specific advanced further education skills within this broad category. We leave this question for future research.

---

23 Advanced further education includes all qualifications which are above the level of A-levels, NVQ level 3, or equivalent: the most common qualifications in 2014-19 in this category were Higher Nationals/BTECs, Diploma in higher education, nursing qualifications, NVQ level 4, and teaching qualifications. It corresponds to the category called “higher education” but no university degree in the Labour Force Survey.
Figure 13. Estimated wage premium for advanced further education, by region

Source: Analysis of UK Labour Force Survey. Note: “5yma” refers to five year moving average. Each point is the estimated wage premium for advanced further education recipients relative to A-level recipients, and the share of advanced further education recipients among private sector employees in a region in each year 1997-2019.

Our analysis of wage premia focused on the returns to skills within a region. Since the UK has free mobility between regions, we next use patterns of graduate mobility between regions to help us understand the degree to which regional education differentials are in part an outcome, not just a cause, of regional productivity differentials. If a region has a low graduate share because local residents are not going to university – but when they do go to university, they find productive graduate jobs in the region – this might illustrate that the region has a shortage of university graduates caused by some bottleneck in the education system (e.g. access to university, or preparation at school). On the other hand, if a region has a low graduate share because local residents with degrees leave to work elsewhere, this suggests that the supply of local graduates is not the most binding constraint on growth – rather, local demand for graduates (availability of productive graduate jobs) is relatively low.

Recently, Britton, Waltmann, Xu, and van der Erve (2021) combined new administrative data from the Longitudinal Educational Outcomes (LEO) database, which includes school records and higher education records of English students, with geographic information after they leave education, to analyse the geographic mobility of university graduates. We use these data, kindly provided to us at an aggregate level by Britton et al. (2021), to analyse a snapshot of English regions’ “graduate gain” vs “graduate drain”: the difference between the graduate share among 27-year olds who grew up in the region, vs. the graduate share among 27-year
olds currently living in the region (Figure 14).24 The LEO data only includes people who went to school in England, so we cannot estimate this for Scotland, Wales, and Northern Ireland). First, we note that in all regions except London and the South East, the share of children growing up in the region who go on to get university degrees is lower than the England average (blue bars). This share is particularly low in the North East, North West, and Yorkshire and the Humber. Yet despite their low shares of students going to university to begin with, there is still substantial net “graduate drain” to London for all English regions: the share of 27-year olds with degrees who live in each non-London region is lower than the share of 27-year olds with degrees who came from that region. The combination of low shares of students going to university, and net outmigration of those who do get university degrees, is consistent with the hypothesis that the low graduate shares in the North of England and the Midlands are the result of a lack of high-paying jobs for graduates – and not a result of a bottleneck on the supply of graduates (see also McCann 2016). This suggests in turn that increasing the graduate shares from these regions without simultaneously working to increase the availability of graduate jobs may do little to boost regional productivity, and may simply lead to underemployment or further outmigration of university graduates.

The supply of university graduates in a region is not only determined, of course, by the mobility patterns of those born and raised in the UK. 14% of the UK population was born outside the UK; immigrants to the UK are more likely to be working age and to have a university degree than the UK population (Wadsworth 2017; 2019). Since immigrants also very disproportionately move to London, regional immigration patterns exacerbate the interregional mobility patterns documented above. Given the disproportionate flow of immigrant university graduates to London, with less immigration, the divergence between the graduate wage premium in London and the rest of the country may have been even greater.

Overall: It is important to underscore the role of education in regional disparities. Education differentials across regions are large and strongly correlated with productivity. The massive increases in educational attainment in the UK in the last two to three decades appear to have been important in significantly reducing the role of skills as a constraint on growth both regionally and nationally: private sector university wage premia fell substantially from the 1990s to the 2010s in almost every region (with the major exception being London). Increasing educational attainment further would still increase incomes and productivity both in lagging regions and nationwide, as individuals with more education tend to be more productive and earn more. But our analysis points up a cautionary note around orienting regional policy around boosting the supply of university graduates: how much effect this would have at the margin, both for the individuals in question and in terms of any spillovers, is less clear and depends on the extent to which tertiary education remains a binding constraint on regional growth. Low and falling wage premia for university graduates in most regions suggest that investments in tertiary education alone in non-London regions may have lower than expected (and falling) returns. At the same time, high graduate outmigration even from

24 People who grew up in a region are defined as people who took their GCSEs there in 2002, 2003, 2004, or 2005. These cohorts each are observed again at age 27. These estimates likely somewhat overstate the degree of “graduate drain”, since younger graduates are particularly likely to move to London and older graduates may be more likely to leave London (Swinney and Williams 2016).
regions with the lowest shares of students going to university suggests that many of these returns may, in fact, accrue to London. These facts together suggest that a lack of highly productive jobs which require graduate skills is as much – if not more – of a constraint on regional productivity than the lack of university-educated workers.

Figure 14. Graduate shares in and from English regions (27-year olds)

Source: Authors’ analysis of data provided by Britton et al (2021), which was drawn from the Department for Education Longitudinal Educational Outcomes dataset

Do specific skills constraints remain? Our analysis does point to a role for increased attention on university-level STEM skills. The massive expansion of tertiary education in the UK over the last three decades appears to have alleviated the scarcity of general university skills, including skills relevant for professional services like law and finance. On the other hand, STEM degree skills remain relatively scarce throughout the country: indeed, in several regions STEM wage premia have barely fallen or have risen despite the expansion of STEM education, suggesting that demand is rising faster than the supply of STEM skills can keep up with.

We note three caveats to this analysis. First, our estimates focus on formal educational qualifications. The extent to which these measures of educational attainment translate into business relevant skills is an open question, and it may be that premia for these types of education have fallen because the skills imparted for this education are not in demand – even as other skills may well be a binding constraint on growth. Second, our estimates suggest that general graduate education levels are not currently the most binding constraint on growth in lagging regions, but this does not necessarily mean they will remain that way – in fact, the sustained policy effort to increase educational attainment over recent decades seems to have worked to alleviate the bindingness of the education constraint, and continued action on
education may be necessary to ensure this remains the case. Finally, even though education disparities across regions may not be the most binding constraint on the growth of the UK’s lagging regions, improved education and skills are still a good in and of themselves, both from the perspective of individuals’ welfare and to improve national productivity.

4. INFRASTRUCTURE

Theory suggests that good transportation infrastructure matters for economic growth; it connects businesses to pools of potential workers, clients and customers, and suppliers. Evidence on agglomeration effects suggests that areas with larger or denser populations tend to have higher productivity even controlling for the composition of the population (Graham and Gibbons 2019; Özgüzel 2020b; Rice, Venables, and Patachini 2006), suggesting that improved transportation infrastructure – by creating a larger effective population for a given area – can generate these productivity boosts. And while empirical evidence on the direct causal effect of transport infrastructure on local economic outcomes is relatively scarce, since isolating plausibly exogenous transportation investments is difficult, there is evidence to suggest beneficial local economic effects of improvements to both road and rail accessibility (Gibbons, Heblich, and Pinchbeck 2018; Bernard, Moxnes, and Saito 2019; Gibbons et al. 2019; Heuermann and Schmieder 2019; Özgüzel 2020b).

The UK’s transport infrastructure investment, as a share of GDP, has been relatively low by international standards. Road investment, in particular, has been among the lowest of any industrialised economy over the last three decades according to OECD data: the UK spent an average of 0.3% of GDP per year on road infrastructure investment over 1995-2020, compared to 0.4% in Italy, 0.5% in Germany and the United States, and 0.65% in France. By contrast, in recent years the UK has spent one of the highest shares of GDP on rail investment of any large industrialised economy. Within the UK, this infrastructure spending has been heavily tilted toward London: per capita transport infrastructure spending in London and the South East was nearly twice as high as in other English regions over 1999-2019, with the gap diverging even more in more recent years (Appendix Figure 5, Appendix Figure 6).

The low infrastructure spending in non-London regions of the UK does not, however, necessarily tell us that non-London regions lack sufficient transport infrastructure or that it is a binding constraint on future economic development. To understand whether infrastructure is a binding constraint on growth – and therefore, whether policy aimed at increasing connectivity would boost productivity growth – we need indicators of whether the demand for transport infrastructure outstrips the supply. Unlike in the case of education, there are few market prices which can help us infer relative private sector demand for transportation investment.

25 Transportation investments may not benefit all regions equally: Crescenzi and Rodríguez-Pose (2012) argue that additional roads only yield benefits where R&D is high.

26 Part of this discrepancy is likely a result of the UK’s relatively small landmass, meaning fewer kilometres of road are required to connect the population in an equivalent fashion. Germany for example has about 50% larger landmass than the UK, as well as around 50% more kilometres of road (data from statistikportal.de and the UK government road length statistics).

infrastructure. But an analogous indicator of whether the existing infrastructure is sufficient to meet the needs of the existing regional economy is the degree of congestion. Road congestion is particularly analogous to a price: higher degrees of congestion impose a direct cost on commuters in terms of their lost time. For rail traffic, crowded trains impose a smaller cost (discomfort, but not lost time); nonetheless, the degree of crowding on trains can indicate the degree to which there is latent demand for trains which is not met by the existing supply. Due to data availability, and because road and rail infrastructure is typically oriented around access to and between cities, we focus here on cities/metropolitan areas, rather than regions.

We start by measuring road congestion. This is the most relevant congestion for most UK commuters: 76% normally commute by road (mostly by car), according to the ONS 2017 Commuting to Work database. Even in large non-London cities like Manchester, Birmingham, or Liverpool, more than 70% of commuters in the 2011 census reported commuting by road (with over 50% by car and less than 20% by bus). To investigate road congestion, we use three indicators, from the UK’s Department for Transport Road Congestion Statistics and from two private companies (TomTom and INRIX). Each of these measures a slightly different concept of congestion. The Department for Transport’s statistics measure the average delay on locally managed “A” roads in seconds per vehicle mile in 2019, for English cities only. TomTom uses real-time traffic data to calculate a congestion index, defined as the average percentage increase in travel time for a thirty-minute trip in a given city over the year, as compared to free-flow conditions (weighted across roads by their total traffic) (TomTom 2022). INRIX estimates the total hours lost per driver per year driving at peak times relative to a free traffic flow scenario (INRIX 2022). Using the three measures, we create a composite congestion index for 17 UK cities by taking the first principal component (which captures 85% of the variance overall) from 2019 data. The index is shown, standardised to a mean of 0 and standard deviation 1, in Figure 15. As can be seen, while London is the most congested city, levels of congestion close to those in London are found in Edinburgh, Belfast, Manchester, and Bristol.

Next, we look at rail congestion. To measure rail congestion, we create a composite congestion index for the morning and evening peak hours from four measures provided in the Department for Transport’s Rail Passenger Use Statistics: passengers in excess of capacity as a share of load, and passengers in excess of capacity as a share of services, estimated at both the 1-hour and 3-hour peaks. These statistics only cover some cities in England and Cardiff in Wales. To create our generalised congestion measures, for each of the morning and evening peak we estimate the first principal component across these four measures (which captures 95% of variance for the morning peak and 83% of variance for the evening peak). Figure 16 illustrates our generalised congestion measures for morning and evening peaks, alongside

---

28 The ONS 2017 Commuting to Work database estimates commute patterns from the Labour Force Survey. 2011 figures are from the ONS Method of Travel to Work database which uses the 2011 Census.
29 We impute values for the four cities which are missing the Department for Transport measure: Edinburgh, Belfast, Cardiff, and Glasgow, by taking the mean of the standardised values of the TomTom and INRIX measures.
30 Passengers in excess of capacity (PiXC) is calculated for each service as the difference between the standard class critical load (the number of passengers on a typical weekday at the point in a service where the passenger load is highest on route into (AM) or out of (PM) the city centre) and the standard class capacity (which includes allowances for passengers standing on journeys less than 20 minutes). It is zero if the critical load is below capacity. PiXC as a share of load is calculated by aggregating the service-level figures up to the city level and expressing this as a percentage of the total standard class critical load for the city. PiXC as a share of services is calculated as the share of services on which the critical load exceeds capacity.
each of the individual measures (standardised to mean 0 and standard deviation 1) for each city in the data over 2018-2019. While it is well known that London has substantial rail congestion, several other cities also suffer from serious rail congestion, including Birmingham, Cardiff, Leeds, Manchester, Bristol, and Cambridge – with Birmingham’s rail congestion actually more severe than London’s on both the morning and evening generalised congestion measures.

Rail congestion at peak times can illustrate the degree of demand from firms and workers for good public transport options to access city centres – giving an indication of the degree to which improvements in these public transport options may help stimulate productive economic activity. Realised congestion, however, may not consistently indicate the latent demand for rail travel: in particular, it will be affected by the frequency and reliability of rail services, the price of rail commuting, and the structure of the rail network. The Office of Rail and Road provides data on the average punctuality of trains, by operator, over 2014-19. Our analysis of this data suggests that trains outside London and the South of England are, on average, substantially more likely to be late (Figure 17). The disparity is large: There were seven operators where more than one in every twenty-five trains was over fifteen minutes late, all of which primarily serve places outside London and the South. In contrast, there were five operators where fewer than one in every two hundred trains were more than fifteen minutes late, four of which primarily serve London and the South of England. These differences in reliability across regions become even more striking when seen in international context: according to the EU’s 2021 Rail Market Monitoring report (EU 2021), the UK is substantially below the EU27 average in terms of both the punctuality of its trains and the share of services cancelled. If trains are unreliable, commuters may not use them – making measured congestion low – even if there is latent demand for reliable public transport options that could boost economic activity by connecting workers and firms more efficiently.

31 In contrast, there is no significant difference in the likelihood of train cancellation between operators that primarily serve London and the South, and operators that primarily serve the rest of the UK (Appendix Figure 8. Train cancellations by operator, Great Britain 2014-19). We categorise the train operators into those which primarily serve London and the South of England, vs. the rest of the country, based on their route maps. We can also analyse bus punctuality: according to Department for Transport statistics on waiting time for frequent bus services, and punctuality for non-frequent bus services, performance is similar in most cities outside London as it is in London.

32 Specifically, over 2015-2018, 88% of the UK’s regional and local passenger rail services were punctual (arriving with a delay of five minutes or less), compared to 92% on average for EU27; 74% of long-distance and high-speed passenger services were punctual, compared to 82% on average for EU27; 3% of regional and local passengers services were cancelled, compared to 2% for EU27; and 4% of long-distance and high-speed passenger services were cancelled, compared to 1% for EU27.
Figure 15. Road congestion in UK cities, 2019

Source: TomTom, INRIX, Department for Transport Congestion Statistics

Figure 16. Rail congestion in cities in England and Wales, 2018-2019

Source: Department for Transport Rail Passenger Use Statistics
Figure 17. Train punctuality in Great Britain, by operator, 2014-19

Source: Office of Rail and Road passenger rail performance, Tables 3123 and 3133. Note: Quarterly data is averaged over 2014-2019 inclusive. We categorized operators primarily serving London and the South of England as Chiltern Railways (CR), Govia Thameslink (GT), Greater Anglia (GA), Heathrow Express (HE), London Overground (LO), South Western (SW), Southeastern (SE), TL Rail (TL), and c2c (c2c). The remaining operators are Hull Trains (HT), TransPennine Express (TPE), London North Eastern (LNE), Avanti West Coast (AWC), CrossCountry (CC), Grand Central (GC), Caledonian Sleeper (CS), Northern Trains (NT), Great Western (GW), East Midlands Railway (EM), West Midlands Trains (WM), TfW Rail (TfW), and ScotRail (SR).

To understand the quality of UK cities’ infrastructure in absolute terms, we next compare statistics on road and rail infrastructure across countries. First, we analyse road congestion, comparing all US, UK, and Western European cities whose metropolitan areas had a 2018 population over 500,000 (according to the OECD). We find that UK cities have much higher road congestion than comparable sized American cities, and somewhat higher congestion than comparable sized Western European cities. Specifically, on the TomTom measure UK cities have 48% higher road congestion levels than similarly-sized US cities, and 15% higher road congestion levels than similarly-sized Western European cities (Figure 18).33

Next, we can examine the reach of the public transport network. To do this, we use data kindly provided to us by Rodrigues and Breach (2021), who estimate the number of people who can reach the city centre within 30 minutes by public transport across 48 Western European cities (using data from TravelTime). This data shows that UK cities have systematically less extensive public transport networks than other Western European cities: conditional on population, the share of the total city population that can reach the city centre

33 These estimates are obtained from a regression of the log of the congestion measure on the log of city population and a dummy for the UK and for Western Europe. On the INRIX measure, the differences are even starker: UK cities have 101% higher congestion than US cities and 31% higher congestion than Western European cities (Appendix Figure 7). The set of cities used is all cities in Western Europe and the US with metropolitan area populations greater than 500,000 in 2018 according to the OECD, for which data on congestion is available. This includes 160 cities for TomTom and 145 cities for INRIX. Older studies similarly suggest particularly high congestion in the UK (Christidis and Rivas 2012).
within 30 minutes by public transport is 23 percentage points lower in the UK’s cities than in Western Europe. We visualise this in Figure 19. As Forth (2017) and Rodrigues and Breach (2021) illustrate, poor public transport accessibility – combined with low housing density – means UK cities have much smaller “effective sizes” than their population would suggest. They argue that this disparity between actual and effective city population sizes can go some way to resolving the UK’s apparent lack of agglomeration economies.

Finally, we use data collected by Conwell, Eckert, and Mobarak (2022) on the geographic area accessible from the city centre of US and European cities, by car and by public transport, at rush hour. Analysing 103 large US and European cities, they find that on average US cities are better-served by car and European cities are better-served by public transport. But the UK looks poor on both counts: for UK cities, the area accessible by car within 30 minutes at rush hour is much smaller than for almost any American city analysed, and the area accessible by public transport within 30 minutes at rush hour is much smaller than for most other European cities analysed (illustrated in Figure 20).

Overall, the evidence collected above suggests that the UK’s large non-London cities are constrained by their limited transport infrastructure. The UK’s cities are less well-served by roads than US or Western European cities, and less well-connected by public transport than Western European cities. The limited scope of the UK’s road and public transport networks makes the UK’s cities outside London systematically smaller in terms of their “effective size” than peer cities with similar total populations. In several non-London cities, high congestion on roads, and a combination of high crowding and poor reliability on trains, suggest a high economic value of commuter travel, and therefore that improving road and rail infrastructure in congested cities would likely bring significant economic returns.34 These facts in particular make the case for intra-city transport improvements to enable greater commuting flows and increase effective city sizes.

We note that we have focused here on intra-city, and not inter-city, transport links. However, the arguments above about the “effective size” of UK cities would also apply to some extent to building more, quicker, and more reliable inter-city transport links between geographically proximate cities. Clear candidates would be cities in the North and Midlands of England, which are not far apart but are very limited in their “effective size”.35 Note that this would suggest a different prioritisation of high speed rail links than that adopted under the HS2 plan, where benefits likely disproportionately accrue to London relative to its share of the UK population (New Economics Foundation 2019).

We have focused here on transport infrastructure. Two other aspects of infrastructure are important to discuss. First, the limited effective city size of UK cities is only partly a result of

---

34 As Coyle and Sensier (2020) have argued, a line-by-line consideration of, for example, rail projects in cost benefit analysis also neglects the possibility of complementarities of efforts which together could expand effective city sizes and generate a productivity boost larger than the sum of its parts. Note that there may still be a strong rationale for investing in transportation infrastructure in cities, towns, and rural areas which are not congested – for example, expanding rural bus services may be important on equity grounds. Our focus in this paper is on constraints to productivity growth, hence our focus on alleviating congestion.
35 Public transport links are much less frequent between major UK cities in the “Northern Powerhouse” region of the North East, North West, and Yorkshire and the Humber, than they are between the major cities of the similarly-sized Rhine-Ruhr region in Germany for example (Swinney 2016).
the limited scope of transport networks - it is also partly driven by low housing density relative to peer European cities (Rodrigues and Breach 2021). This suggests that improvements to transport infrastructure may need to come alongside increased housing density to fully reap the benefits of agglomeration. Second, another key aspect of physical infrastructure in the modern economy is access to broadband internet. Analysis of OFCOM’s fixed broadband coverage data shows no evidence of meaningful differences in broadband speeds or in access to ultrafast or full fibre broadband between London and the UK’s other major cities. However, it is important to note that the UK as a whole lags behind peer countries in its access to ultrafast internet (download speeds of 100Mbps or more): according to the OECD in 2021 the UK had fewer than 10 ultrafast fixed broadband subscriptions per 100 inhabitants, a figure among the lowest in the OECD and substantially lower than Germany (14), France (20), the USA (25), or Korea (40).36 Thus while the regional disparity in access to high speed internet within the UK is minimal, the UK’s regions may well be held back by this as compared to peer countries.

Figure 18. Road congestion in UK, US and Western European cities

Source: TomTom 2021 Congestion Index, OECD. Note: All metropolitan areas in the US or Western Europe, with population greater than 500,000 in 2018, for which there is both congestion data from TomTom and population data from the OECD, are included. Note that TomTom uses its own city definition that may not overlap directly with the OECD metropolitan area definitions. UK metropolitan areas with population > 750,000 as of 2018 are labelled.

36 Similarly, across three different rankings of average experienced download speed of fixed broadband connections across countries 2020–21 (Ookla, M-Lab, and Steam), the UK ranked in the bottom quarter of OECD countries across all three, according to the OECD Broadband Statistics database.
Figure 19. Public transport accessibility in UK and Western European cities

Source: Rodrigues and Breach (2021). These data are from Centre for Cities’ calculations based on data from TravelTime, ONS, and Eurostat. London and Paris are not shown.

Figure 20. Area accessible by road and public transport, UK, US and Western European cities

Source: Conwell, Eckert, and Mobarak (2022). These estimates are calculated from Google Maps, using a start time of Wednesday 8:30am.
5. INNOVATION, RESEARCH AND DEVELOPMENT

If knowledge and information spillovers play a major role in firms’ economic activity, the existence of a highly-educated population and high-quality infrastructure may not be enough on their own to provide the necessary conditions for rapid productivity growth (as in e.g. Rodrik and Stantcheva 2021). Why? Knowledge has characteristics of a public good: it is non-rivalrous in its use and is to some extent non-excludable (as information spills over to others either through communication by individuals, flows of workers between firms, or embedded in new goods or services). Since private sector actors cannot capture the full return on the knowledge they produce, there will typically be underinvestment in innovative activity relative to the socially efficient level, suggesting an important role for the public sector (Jones and Williams 2000).

In this section, we analyse government support for research and development (“R&D”). We focus on R&D partly because of its central role in productivity-enhancing innovation: A substantial share of high-productivity economic activity is concentrated on sectors which benefit from innovative activity (especially in pharmaceuticals, software development, and the automotive and aerospace sectors), and a large body of evidence suggests the importance of R&D activity in boosting productivity (see e.g. Crepon, Duguet, and Mairessec 1998; Griffith et al. 2006). Consistent with this, empirical research suggests that public R&D spending tends to “crowd in” private sector R&D, particularly for smaller firms (Aitken et al. 2021; Becker 2015; Azoulay et al. 2019; Moretti, Steinwender, and Van Reenen 2019). It is also important to note, however, that there is much innovative activity which is not captured by R&D measures and therefore will be excluded from our analysis (C. I. Jones 2022).

The argument that knowledge is a public good generates an argument for public support for R&D. But why should its regional distribution matter? First, to the extent that the knowledge spillovers generated by R&D activity are local, the regional distribution of this activity matters. Second, R&D activity can help overcome coordination externalities, which arise where there are benefits to firms of colocation because of shared pools of workers, customers, suppliers, or because of knowledge spillovers: in these cases, it may not be profitable for any individual firm to locate in a given area, even if it would be profitable for a collection of similar firms to co-locate there. Publicly-supported R&D activity, whether carried out directly or via public support for research by universities, is therefore often cited as a policy tool to support the development of a cluster.38

---

37 According to the Office for National Statistic’s Business enterprise research and development, UK: 2020 data release.
38 There is substantial evidence of localised spillover effects of public support for R&D on private sector innovation and economic activity, starting with Jaffe (1998). Several recent papers provide credible, large estimates: Andrews (forthcoming) finds that the establishment of new universities in the US, 1839-1984, led to increased patenting in their county (relative to counties which were considered but not chosen for new universities), most of which was done by individuals who were not faculty or alumni of the university - consistent with large spillover effects on innovation. Kantor and Whalley (2014), using shocks to university endowments as instruments for university research expenditure, find statistically significant and persistent local productivity spillovers from increased university spending, particularly in industries which are closer to universities in terms of their propensity to hire graduate students and/or their history of citing university-produced patents. Hausman (2022) studies the implementation of the Bayh-Dole Act which increased universities’ incentives to connect to industry and disseminate technology, finding that firms which were close to a university, and in industries which were related to the specific university’s technological strengths, substantially increased their employment and wages, as well as patenting rates. Myers and Lanahan (2022) estimate spillovers from US Department of Energy R&D grants to small firms, finding that for every new patent produced by grant recipients, an additional three patents are produced by others who benefit from spillovers, with a very disproportionate share of these additional patents generated locally.
Two aspects of the UK’s national R&D expenditure patterns stand out relative to peer countries. First, OECD data shows that total gross expenditure on R&D is relatively low in the UK at 1.7% of GDP in 2019, compared to an OECD average of 2.5% of GDP (see also R. A. Jones 2022). Second, while total public support for R&D is relatively similar to peer countries, the UK’s public support for R&D is tilted more heavily toward public support for business R&D (particularly through tax relief), as opposed to direct expenditure on R&D by the public sector.

How does the UK’s R&D policy affect regional economies? The fact that the bulk of public innovation expenditure is place-blind – simply subsidising business R&D activity – serves to amplify existing regional disparities in private sector innovation activity. Business R&D expenditure is, as would be expected, heavily concentrated in higher productivity, higher income locations. The more discretionary components of government R&D funding – direct public R&D expenditure, and higher education R&D – are also heavily skewed to richer regions (Forth and Jones 2020). Per capita, direct government R&D expenditure in 2016 was £60 in London and the South East of England, compared to only £21 in the North of England, £14 in the Midlands, and £7 and £5 in Northern Ireland and Wales respectively. For higher education R&D expenditure, the pattern was similar, with per capita figures in London (as well as Scotland) more than twice their level in the North of England, the Midlands, Wales, or Northern Ireland. This in itself may be efficient: government R&D may be directed where there is the maximal impact per £, and the most effective research institutions may be disproportionately concentrated in these already-richer regions. To understand whether increased public support R&D outside London and the South East of England would help boost growth and productivity, we would ideally be able to assess the extent to which firms in these regions have the absorptive capacity for the knowledge spillovers generated by this R&D. In practice, there are few good measures of this.

One possible metric for assessing whether the regional distribution of public sector or higher education sector R&D expenditure is acting to boost lagging regions is to study its distribution relative to business sector R&D expenditure. Business R&D is (presumably) undertaken in places with productive research environments, where the expected return on R&D investment is high. In Figure 21, left hand panel, we illustrate government R&D expenditure across UK regions relative to business R&D expenditure. Our metric of interest is each region’s share in national government R&D expenditure minus that region’s share in business R&D expenditure, taken as an average over 2000-2018, with a positive share meaning that the region receives a larger share of government R&D expenditure than it does of business R&D expenditure. As can be seen, even with this metric the government R&D skew toward richer regions persists, with a clear divide between the South of England and Scotland, which are richer than average and which receive a higher share of government R&D than they do of business R&D, compared to the North of England and Midlands, Wales, and

39 The R&D share of GDP is 2.2% in France, and 3.2% in Germany and the United States. Note: recent research by the Office of National Statistics suggests that the UK R&D figures may be undercounted by 50%, which would bring the UK to the OECD average.

40 Per capita, direct government R&D expenditure in 2016 was £60 in London and the South East of England, compared to only £21 in the North of England, £14 in the Midlands, and £7 and £5 in Northern Ireland and Wales respectively.

41 Moreover, it is not clear that this is an exogenous parameter: places seeing high public or private R&D investments also tend to see a growth in firms’ capacity for innovation (Van den Bosch, Volberda, and de Boer 1999, Lau and Lo 2015).
Northern Ireland, which are poorer than average and which receive a lower share of government R&D even than they do of business R&D. In the right-hand panel, we carry out the same exercise for higher education R&D expenditure relative to business R&D expenditure. Higher education sector R&D has less of a general Southern skew than direct public R&D expenditure, but is still heavily skewed toward the richest region, London, relative to business R&D.

It is not inevitable that public support for R&D spend should be even more skewed toward rich regions than business R&D spend. Figure 22 illustrates the same metric – each region’s share of public or higher education R&D expenditure minus that region’s share of business R&D expenditure – against the region’s GVA per worker in 2016 for the UK and Germany. Note that there are two differences from the maps above: we use smaller regions for increased granularity (OECD TL3 level), and data only from 2009-2016, which is the longest time period for which data on all these smaller regions is available. In the UK, a positive slope illustrates that richer regions receive an even larger share of public or higher education R&D spending than they do of business R&D spending. In Germany, the situation is the opposite: the downward-sloping patterns illustrate that poorer regions (particularly those in the former East Germany) receive substantially larger shares of public and higher education R&D spending than they do of business R&D spending. That is: German public sector and higher education R&D to some extent counterbalance regional economic inequality in business R&D spending, while UK public sector and higher education R&D spending exacerbate regional economic inequalities even more than business R&D does. Forth (n.d.) and Forth and Jones (2020) similarly show that the UK’s regional allocation of public sector R&D spend is more skewed to London than its business R&D spend, while Germany’s public sector R&D spend flows more disproportionately to its poorer regions.

In sum, evidence on the distribution of government support for R&D vs. private sector R&D spend suggests that several lower-income regions which may have the absorptive capacity for further innovative activity (as proxied by their level of business R&D), are held back by a lack of public sector innovation support. Indeed, unlike some OECD peers such as Germany, public policy in the UK may be making regional inequalities in R&D worse by prioritising regions with academic expertise over those with existing private sector R&D strengths.

The UK’s limited public R&D spend outside the “Golden Triangle” area around London, Oxford, and Cambridge is also particularly striking when considering that fact that many non-Southern UK cities perform very highly in international measures of academic research. For example, ranking cities by the number of highly-cited academic articles they produced between 2014 and 2016 using Web of Science data, we find 9 UK cities in the global top 100; this compares to only 4 French cities, 5 German cities, 3 Italian cities, and 2 Spanish cities.

---

42 These data are from the Gothenburg Quality of Government EU Regional data set. The pattern of results looks similar for analyses at the large (TL2) region level.
43 Surveys of firms’ innovative activity across regions also suggest that there are large numbers of innovative firms outside London and the South East, for example the UK government’s Innovation Survey (BEIS 2021).
Of course, our analysis here is at a high level: the effectiveness of public sector R&D support depends not only on the level of R&D spend but also the way the spending occurs, and more broadly the way regional innovation ecosystems are supported and developed (McCann and Ortega-Argilés 2013).

Figure 21. Government/higher education R&D spending share relative to business R&D spending share, by region (2001-2018 average)

Source: OECD Regional Statistics.

44 The UK cities in the top 100 (and their rankings) were London (3), Oxford (10), Cambridge (21), Edinburgh (52), Manchester (57), Birmingham (89), Bristol (94), Glasgow (96.5), and Southampton (99). Forth and Jones (2020) provide suggestive evidence of “home bias” in research funding allocation by large funders like the Wellcome Trust or Innovate UK, relative to the research strength of UK regions outside the greater South of England.

45 Innovation is one example of a wider set of coordination externalities. We raised another – industrial structure and clustering – in Section 2. Hidalgo et al (2007), Hausmann et al (2013) - and in the UK context Mealy and Coyle (2022) and Rodrigues and Breach (2021) – highlight adjacent high value added industries that existing regional economies could branch towards, given their existing industrial structure, which a coordination infrastructure (such as innovation or financial support) could facilitate.
6. ACCESS TO FINANCE

In Sections 3, 4, and 5, we explored the potential for education, infrastructure, and innovation policy to boost productivity in poorer regions of the UK. A poorly educated workforce, poor infrastructure, or limited access to knowledge and innovative ideas, can reduce the potential return on private sector investment. It is possible, however, that the growth of lagging regions in the UK is held back not by an absence of high-return economic opportunities, but by an absence of financing for those opportunities — if access to finance is limited or expensive in these regions. Indeed, London receives a disproportionate share of lending and, in particular, a disproportionate share of equity investment relative to its share of the population of Small and Medium Sized Enterprises (“SMEs“). London-based SMEs are more likely to use external finance, and have higher credit balances as a % of turnover, than SMEs in most other regions (BVA BDRC 2020). But from these facts alone it is not possible to conclude that access to finance is a binding constraint on economic development for regions outside London and the greater South East. Total financing is an outcome of both supply and demand conditions: financial investment in a region may be low because the returns to this investment are low, and not because of supply constraints. Below, we analyse

---

46 Unless otherwise noted, we define SMEs as firms with less than 250 employees.
a number of indicators which can help us distinguish whether limited access to finance is a binding constraint on investment, rather than a symptom of poor prospects for returns on investment. We find little evidence that finance in general is a more binding constraint on growth for SMEs outside than in London, but some evidence that a lack of access to equity financing in particular may be constraining a subset of high growth potential SMEs.

First, survey evidence provides no indication that SMEs outside London feel more constrained by access to finance than those in London. If anything, the opposite appears to be the case. In Figure 24, we show nine indicators of financing constraints from BVA BDRC’s SME Finance Monitor surveys, 2016-19. On all indicators, SMEs within London report being more constrained than the average across other regions. SMEs in London are more likely than average to report that it “feels quite difficult to get external finance” and less likely than average to report that they “never think about using (more) external finance”. On six of the nine indicators, London’s SMEs report being more financially constrained than those in any other region, including: the share reporting that access to external finance and/or cashflow or late payment would be an obstacle in the next 12 months, the share reporting that they do not have confidence a bank would say yes if they requested a loan, and the share reporting that they injected personal funds into their business because they felt they had no other choice.

Second, SME bank lending appeals rates are similar across regions. In 2011, the Banking Taskforce established a right for SMEs to appeal if they had been refused a loan by a bank. If SMEs outside London are more likely to be unfairly denied finance, or are more likely to be constrained in their growth plans by lack of access to finance, one would expect to see a higher appeals rate for SMEs in non-London regions. Figure 23 illustrates that there is no evidence for this: the share of total appeals from each region to the SME Banking Taskforce is almost exactly the same as the share of total SME lending received by each region.

Third, SME bank rejection rates, margins, and collateral requirements are similar for SMEs in London and outside London. To investigate this, we draw on data from three papers which study bank lending at the firm level across regions. Armstrong et al (2013) study SME loan applications over 2001-2012. In regressions of the rejection rate for SME bank loan applications on a range of firm-level characteristics, they find that there is no statistically significant difference in rejection rates across regions when controlling for business characteristics (including industry, sales, age, and risk level) and owner qualifications. If anything, rejection rates are higher in London than outside London. Similarly, they find no significant evidence of a higher margin over base rate or higher collateral requirements for SMEs outside of London than for SMEs in London. Cowling et al (2020) similarly find no differences in credit application outcomes between observably equivalent firms across UK regions. They do find evidence of a higher price of credit for firms in the North West, Scotland, and Wales as compared to London, but no differences between London and the other regions. Zhao and Jones-Evans (2017) find evidence for differences in credit rationing across UK regions, but with London one of the more financially constrained regions.

47 This data comes from around 18,000 interviews of SMEs in each year 2016-2019. The SME Finance Monitor data is the most comprehensive data on small business finance in the UK, and was established and funded via the Business Finance Taskforce.
48 In recent years peer-to-peer (P2P) lending has become more significant as a source of funds, equivalent to 4% of gross bank lending flows by 2018 (British Business Bank 2019). We find no evidence that P2P lending is differentially difficult to access across regions.
The above discussion focuses on bank lending, which is the primary source of finance for the majority of SMEs. Of the roughly half of SMEs who used any source of external finance in 2019, 87% were using traditional banking products (overdraft, loan, mortgage, or credit card). Only 4% of SMEs reported either using or planning to use equity finance (BVA BDRC 2020). But while equity investments are only relevant for a small subset of high growth potential SMEs, these SMEs are disproportionately responsible for economic growth, and early stage equity investments in these companies are important for stimulating innovation and rapid, transformative growth, particularly in new sectors (see e.g. Müller and Zimmermann 2009; Lerner and Nanda 2020). We therefore now analyse regional disparities in equity investments.

London is both the recipient and the source of the lion’s share of equity financing in the UK: there were nine equity deals per 10,000 SMEs in London in 2021, compared to two per 10,000 in the rest of the UK (British Business Bank 2022). and nearly half of all equity investment in UK firms over 2011-2017 was from London-based investors (Wilson et al 2019). Most of this disparity across regions, however, is explained by differences in the investment opportunities available to equity investors: Wilson, Kacer, and Wright (2019) analyse all equity investments in SMEs in the UK 2011-17, and find that 90% of variation in deal flow can be explained by three differences in regions’ economic structure (and therefore the investable opportunities in the region): the number of SMEs, the share of “high-growth firms”, and the share of firms in high-tech manufacturing or knowledge-intensive services.

The remainder of the regional disparity in equity investment, however, may still reflect differential access to financing. When controlling for a large range of firm-level characteristics, Wilson et. al find that there was no statistical difference in the likelihood of receiving equity funding for SMEs located in the North East or Northern Ireland, as compared to London, and Scotland-based SMEs were actually more likely to receive equity funding than observably equivalent London-based firms. However, they do find a statistically significant difference in the likelihood of receiving funding for SMEs in other regions, with a particularly large gap in the East and West Midlands and in Yorkshire and the Humber: SMEs in these regions were only 50% as likely to receive equity funding as observably equivalent firms in

---

Analysis of FundingCircle data over 2010-17 (the first, and one of the largest, UK P2P lenders) shows that the median interest rate on P2P loans is lower for most non-London regions than for London (Ekpu et al. 2020 Table 4), and that the default rate on P2P loans across all UK regions is lower in London (Xu, Su, and Celler 2021). Note that our analysis focuses on the level of large regions, comparing for example the West Midlands or North West to London. At a more disaggregated level, there is evidence that some peripheral regions far from major cities, for example in parts of the South West and Scotland, do face difficulties accessing bank lending (Hutton and Lee 2012; Lee and Brown 2017), and that this may in part be due to the centralised nature of the UK banking system (Mayer, McCann, and Schumacher 2021).

This is illustrated particularly clearly in the US, where only 0.5% of firms receive venture capital finance, but venture capital backed firms represented 47% of non-financial IPOs over 1995-2018 (Lerner and Nanda 2020). Early stage equity investors can stimulate firm growth not only in terms of providing access to finance but also in terms of the capacity building and support experienced investors can provide to new firms (e.g. Gonzalez-Uribe et al. 2021; Bernstein, Giroud, and Townsend 2016).

Similarly, London, the South East, and the East of England received 67% of all equity investments in SMEs (and 75% of all invested funds in these deals) over 2011-2017 (Wilson et al 2019).

Moreover, in 2019, there were 3.4 unique venture capital investors per 10,000 SMEs in London, compared to an average of 0.3 for the rest of the UK (British Business Bank 2022). In contrast, bank branches are relatively evenly divided across the country: London hosts only 16% of bank branches and 19% of the SME population (British Business Bank 2022).

“High-growth firms” are defined as companies with more than 10 employees and average annual employment growth of 20% or more over 3 years. To conduct this analysis, they match data on 17,431 equity investments from Beauhurst with firm-level data from Companies House and other sources, defining SMEs per the European Commission as firms with less than 250 employees and either turnover lower than 50 mil. EUR or total assets lower than 43 mil. EUR.
London, and, conditional on receiving funding, received 27-35% less.\textsuperscript{53} We visualise these results in Figure 25, with the blue dots showing the relative likelihood of any SME in a given region receiving an equity deal, as compared to London-based SMEs, the red triangles showing the relative likelihood of a high-growth SME receiving an equity deal, and the green squares showing the estimated regression coefficients (the relative likelihood that an SME in a given region receives an equity deal, as compared to an observably equivalent SME in London).

This disparity in equity funding even conditional on firm observables could be consistent with the hypothesis that access to early stage equity financing for high growth potential SMEs is more constrained outside of London, perhaps because the concentration of equity investors in London reduces their ability to identify good investments outside London, and reduces their ability to monitor and support growth in portfolio companies which are physically more distant or harder to get to. This is relevant at both the pre-investment screening stage (see e.g. Howell 2020), as well as the post-investment capacity building and support stage: quicker and easier access to portfolio companies improves venture capitalists’ ability to monitor companies and leads to improved business outcomes (Bernstein, Giroud, and Townsend 2016). On the other hand, it could also be consistent with the hypothesis that the growth prospects of a firm of a given size and in a given industry are better in London than they are in other regions – for example because of agglomeration economies.

There are two additional pieces of evidence we can bring to bear to try to distinguish between these two explanations for the equity funding gap across regions. First, if it is the case that distance means London-based investors overlook profitable investment opportunities in other parts of the country, this should provide particularly attractive investment opportunities for local investors who can identify, fund, monitor, and support the growth of these businesses. In this situation, we would expect to see non-London-based equity investors disproportionately directing funding to these missed opportunities in their own regions, and directing only few of their investments to London and the South East of England (which are already, under this hypothesis, relatively over-funded by London-based investors).

To see whether this is happening, we analyse data on the destination regions of private sector equity investments, by the region the investor is based in (using data kindly provided to us by Wilson, Kacer, and Wright (2019)). Overall, UK investors based outside London and the South East direct 32% of their funding to London/South East based companies, and 31% to their home regions. The large share of funding going to firms based in London and the South East, and relatively low share going to their home region, would not at first glance suggest a large number of highly constrained, potentially profitable investment opportunities in regions outside London and the South East. However, there are regional differences, which we illustrate in Figure 26. In some regions, local private sector equity investors invest the large

\textsuperscript{53} The equivalent figures on the likelihood of receiving equity funding, for other regions, are 65% for the North West, 70% for the South West, 77% for the South East, and 80% for the East of England. The observables controlled for are: variables related to financial situation of the firms (logarithm of total assets, intangible assets to fixed assets, profit and loss account reserve to total assets, cash to total assets, bank overdraft and long-term liabilities to total assets trade creditors to total liabilities and net worth to total assets, number of charges on assets, age), directors-related variables (board size, directors’ age, directors’ age diversity, directors’ tenure, directors’ experience, founding directors’ experience, proportion of female directors, proportion of foreign directors, number of directorships, proportion of non-institutional directors, indicator of family firm and indicator of previous experience with equity funding) variables related to industry sector (21 industry sectors based on NICE 2007 – Hirschman-Herfindahl competition index and indicators of industry sectors), and variables representing national-level macroeconomic changes (year indicators).
majority of their funds in local SMEs, and very little of their funds in London and the South East. The North West is a particularly stark example, with 57% of private sector equity investments going to North West based SMEs and only 11% going to SMEs based in London and the South East.

Second, we analyse data from the British Business Bank, who use Beauhurst data to study all UK firms who received equity investments during 2011-13, and their progress over seven subsequent years (British Business Bank 2022). These data show that, amongst firms who received equity investment, firms inside and outside London had similar initial investment sizes – in fact if anything, the initial round was slightly larger for non-London firms – and had a similar chance of follow-on funding rounds, with an average of one additional round completed by seven years after the first investment. However, the size of follow-on funding rounds was much larger for firms in London: seven years after the first investment, London-based firms had raised almost twice as much as non-London-based firms on average (illustrated in Figure 27). These data are more consistent with the interpretation that non-London firms’ growth prospects are more constrained, than they are with the interpretation that investor bias or lack of awareness of opportunities outside London is driving equity funding disparities: among the firms investors are already aware of (because they have already chosen to invest a first round), the London-based firms attract bigger second rounds, suggesting that investors see more promise in the growth prospects of their London-based than their non-London-based investments over time. However, given compelling evidence from the US that proximity to venture investors improves not only start-ups’ likelihood of receiving funding, but also their growth prospects as they are able to receive more intensive advice and assistance (Bernstein et al (2016)), it is possible that these smaller follow-on rounds are still in part a function of a lack of access to the knowledge and opportunities early stage investors provide.

A final possibility is that firms’ access to finance is limited not because they lack intrinsically profitable investment opportunities, but because they lack the entrepreneurial or management capacity to take advantage of them. A growing body of research documents the existence of a set of management practices which are strongly associated with higher productivity across countries, industries, and firms (Bloom and Van Reenen 2007; Bloom, Sadun, and Van Reenen 2016). While the average UK firm’s uptake of these high-productivity management practices is lower than that in US or German firms, suggesting that there may be scope to improve UK productivity overall by improving the quality of management, there is little evidence to date to suggest that management practices differ in a large, systematic, and consistent way across UK regions. It is harder to evaluate whether there is a lack of entrepreneurial knowhow or capacity across regions. It is clear that there is more start-up

54 The 2019/2020 wave of the Management and Expectations Survey, for example, finds average management scores of 0.62 (out of 1) for Scotland and the South East, 0.61 for the East Midlands, 0.6 for London and the West Midlands, 0.61 for the East Midlands, 0.59 for the North East, North West, Yorkshire and Humber, and the East of England, 0.58 for the South West, and 0.57 for Wales. While these differences are somewhat correlated with regional productivity, the magnitudes of these differences are very small. Moreover, average management practices do not seem particularly stable within region over time: in the 2016 wave of the same survey, Scotland had the lowest score of any UK region at 0.44 (Schneebacher 2021). Analyses of a 2016 pilot version of this survey found no statistically significant differences in management practices across regions when adjusting for firm characteristics (Awano 2017). On the other hand, Forth, Askenazy, and Bryson (2019) find that firms in most regions outside London have a substantially and significantly lower likelihood of using modern human resource management practices, even when controlling for firm characteristics.
activity in London, as well as more “creative destruction” as businesses fail at higher rates: over 2015-19, on average each year there were 16 new businesses started and 11 business deaths per 1,000 working-age residents (age 15-64). These figures were lower for every other region: for example, in the West Midlands these figures were 9 and 8 respectively, and in Northern Ireland (the lowest region) it was 5 new businesses and 3 business deaths per 1,000 working-age residents (according to OECD Business Demography data). But – particularly in light of our analysis on constraints on growth – these disparities could be a function of a lack of profitable business opportunities, rather than a lack of entrepreneurial capacity.

Overall, then, the evidence suggests that SMEs outside London on average are not particularly constrained in their overall access to funding: if anything, London-based SMEs are more financially constrained relative to their demand for funding. In terms of equity financing for high-growth-potential SMEs, the evidence is less clear. There are large differences in the likelihood of receiving equity investments, and the size of deals, for observably equivalent SMEs in London vs. in (for example), the North of England or the Midlands. Further evidence is needed to evaluate whether this gap arises from difficulty evaluating the prospects of non-London-based firms in order to decide whether to invest, or from difficulty monitoring and supporting non-London-based firms in order to maximise the chances of success of the investment – as the appropriate role for government intervention would differ in each case.

Figure 23. SME appeals 2011-2016 against total lending 2015
Figure 24. Self-reported access to finance, SMEs 2016-19

Source: SME Finance Monitor, BVA BDRC. Note: All variables are averaged over 2016-2019 unless otherwise noted. “Difficult to get ext fin” is the share of SMEs who agreed or strongly agreed with the statement “My impression is that it is quite difficult for businesses like ours to get external finance” (2018 & 2019 only). “Think about more ext fin” is 1 minus the share of SMEs who agreed or strongly agreed with the statement that they “Never think about using (more) external finance”. “Need more ext fin” is the share of SMEs who reported that they had a need for more external funding in the past 12 months (2018 & 2019 only). “Access ext fin obstacle” and “Cashflow obstacle” are, respectively, the share of SMEs who identified “Access to external finance” or “Cash flow/late payment” as an obstacle to them running the business as they would wish in the next 12 months. “Discouraged” is the number of SMEs who report being discouraged from pursuing bank lending, as a share of the SMEs who do not plan to apply/renew external finance because there are barriers that stop them from applying. “No confidence (1)” is the share of SMEs who do not plan to apply for finance in the next 12 months, but would have liked to, who report not having confidence that the bank would say yes (2018 & 2019 only). “No confidence (2)” is the share of SMEs who plan to apply for finance in the next 12 months who report not having confidence that the bank would say yes (2018 & 2019 only). “Injected funds: no choice” is the share of SMEs who reported having injected person funds into the business in the last 12 months because they felt they had no choice but to do so.
Figure 25. Equity investment likelihood and deal size by region, 2011-17

Panel A: Likelihood of receiving equity investment

Panel B: Equity deal size

Source: Authors’ analysis of estimates from Wilson, Kacer, and Wright (2019). Note: Equity deals / deal value per SME or per high growth SME are reported directly in Wilson et al. (their Table 6). Coefficient estimates in Panel A (green squares) are odds ratios calculated by Wilson et al. from regressions of whether a firm received equity funding on firm characteristics and region dummies (their Table 18); coefficient estimates in Panel B (green squares) are regression coefficients on regressions of log deal size on firm characteristics and region dummies (their Table 21). We put asterisks by the green squares where the coefficient estimate was not significantly different from 1 (i.e. the difference in equity funding likelihood, or deal size, between the region in question and London was not statistically significant).
Figure 26. Private sector equity investments by region of investor

Source: Authors’ analysis of data provided by Wilson, Kacer, and Wright (2019).

Figure 27. Early-stage funding for London and non-London based firms

Source: British Business Bank (2022), using data from Beauhurst. Note: Charts show the cumulative number of funding deals received, and cumulative total investment amount received, for the cohort of all UK startups who received their first equity investment round in 2011, 2012, or 2013. “Year 0” refers to the first year they received funding, and “Year 7” refers to 7 years after that.
7. WHY HASN'T LABOUR MOBILITY REDUCED REGIONAL ECONOMIC INEQUALITY?

In Sections 3-6, we considered factors which might prevent the UK’s lagging regions from growing faster, and policy interventions which might tackle these and boost productivity. But within a single country, even if policy interventions do not succeed in boosting growth, labour mobility would under most models of spatial equilibrium be expected to work to reduce regional economic inequalities as people move from lower-productivity lower-income regions to higher-productivity higher-income regions (Blanchard and Katz 1992). Indeed, in some cases policy may have little ability to boost productivity, either because of insufficient information, because the forces generating path dependence and economic agglomerations are too strong to overcome, or because the degree of scarring from past economic shocks is too deep. From this perspective, the remedy for regional economic inequality is to enable migration to high-productivity areas.

Internal migration in the UK currently – and strikingly – exhibits the opposite pattern than would be needed if mobility were to reduce regional earnings inequalities. On net, people move from London, the region with the highest productivity and highest earnings, to other regions of the country – and this is true even of younger working-age individuals, with net outmigration from London for all age groups except people in their twenties (Appendix Figure 10). London’s population continues to grow not because of internal migration to the city, but because net immigration inflows from overseas outweigh net internal migration outflows to the rest of the UK (McCann 2016). Figure 28 illustrates that the UK’s internal mobility patterns are unusual in international perspective: across eight large industrialised countries, the UK and France stand out as having extremely high levels of net negative outmigration from their most productive capital city regions.

Since the wage premium in London relative to other UK regions is so high, why don’t more non-London residents move to London? The answer must be some combination of (1) UK residents have a low propensity to move (perhaps because of high economic costs of moving, or because of strong social ties or community attachment), or (2) the true net economic benefits of migration are lower than the nominal earnings differentials (either because of differences in regional costs of living, principally housing, or because of better amenities or quality of life in low-income than in high-income places).

We find no evidence to support the theory that UK residents have a particularly low propensity to move. According to OECD Regional Statistics data 2014-2019, 2.2% of residents of the UK had moved from a different UK region in the previous year – a gross interregional migration rate only a little smaller than the US at 2.3%, and substantially higher.

55 If regional labour demand is downward-sloping, and agglomeration effects and regional aggregate demand effects are not too large.
56 While some outmigration from London is to commuter belt parts of the East and South East of England, there is still net outmigration from London, the East, and South East combined, for all ages over 30.
57 Specifically, the figure plots each region’s net interregional mobility rate against their regional GVA per capita relative to the national average for 2014-18. Net interregional mobility is defined as the number of in-migrants from other regions within the country, minus the number of out-migrants to other regions within the country, as a share of the region’s population. Data is from OECD.
than in the other G7 countries. Moreover, unlike in the US – where there has been a sizeable fall in internal migration in recent decades, a factor often considered important in explaining persistent regional economic inequality (Molloy, Smith, and Wozniak 2011; Austin, Glaeser, and Summers 2018) – there is no evidence to suggest there has been a meaningful fall in interregional migration in the UK. Champion and Shuttleworth (2017), using address data from the decadal censuses in 1971, 1981, 1991, 2001, and 2011, find that the propensity to move distances of longer than 10km was constant throughout that forty year period, and Alvarex et al. (2021) find no evidence of any trend decline in UK year-to-year mobility between the nine large English regions and three nations over the more recent period 1996-2018.

On the other hand, the evidence strongly points towards high housing costs eroding the net economic benefits of moving to London for many people. While hourly wages are 29% higher in London than in the North East, for people of the same education level, age, and gender (Figure 5), housing costs are also substantially higher, with London house prices twice the UK average as of 2019. Using ONS data on incomes and rents for private rental sector tenants across English regions over 2017-19, in Figure 29 we show renters incomes’ at the 25th, 50th, and 75th percentiles, and estimated renters’ incomes net of housing costs at the 25th, 50th, and 75th percentiles. Household incomes are substantially higher in London than in other English regions at all three points in the income distribution. But this London earnings premium is entirely erased for the 50th percentile households when calculating income net of rent costs. For the 25th percentile households London is actually one of the lowest-income regions net of rent costs. Similarly, Agrawal and Phillips (2020) find that median household income in London is 14% higher than the UK average before housing costs, but only 1% higher than the UK average after housing costs. This data suggests that for lower- and middle-income individuals, there is little net economic incentive to move to London since rent is so high. Income convergence cannot occur through migration, because high house prices erode any net economic gains for most people from in-migration to London from other regions.

On the other hand, this data suggests that much of the productivity and wage increases generated by London’s economic rise have been capitalised into housing costs – making the net beneficiaries London’s pre-existing homeowners and landlords, even as it excludes individuals from other regions who may otherwise be induced by high wages to move into London.

Taken together, this evidence suggests that despite the UK’s relatively high level of interregional migration, there are still substantial barriers for most people to move from low-

---

58 See Appendix Figure 9. See also McCann (2016). In addition, the fraction of the UK population who change address from one year to the next is also relatively high in international context (Bell et al. 2018; Molloy, Smith, and Wozniak 2011). The propensity to change address is not as useful an indicator for our purposes, however, since most changes of address in most countries are within not across regions, and we are interested specifically in the propensity to move across regions. Not everyone is mobile: nearly half of UK residents without a degree and a third with a degree have always worked in the travel-to-work area where they were born (Bosquet and Overman 2019).

59 They do find that there has been a decline in very short-distance moves of less than 10km, mostly driven by a declining propensity of the elderly, widowed, or retired to move. Judge (2019) also finds evidence of a decline in rates of moving between local authorities. Short-distance moves are less relevant, however, for understanding regional economic convergence.

60 We calculate incomes net of housing costs at the ith percentile as income at the ith percentile minus rent at the ith percentile.

61 Cavallini et al (2021) show that, all else equal, interregional migration flows in the UK are highly responsive to both regional differentials in GDP per capita and house prices. Overman and Xu (2022) argue that the fact that high rents offset high earnings in richer places may explain the fact that spatial disparities in reported wellbeing are smaller than those in labour market outcomes.
income to high-income places – specifically, from many non-London regions to London. The London area’s limited housing supply and high house prices seem to be the key factors in explaining why there are net internal migration flows away from, rather than towards, the UK’s highest productivity regions.

Figure 28. Net interregional migration and GVA per worker, 2014-18


---

62 High levels of London-focused international migration explain why London’s population nevertheless continues to rise, which adds further upward pressure on housing costs and further dampens incentives for inter-regional migration. International migrants may be less sensitive to higher costs of living of London as compared to internal migrants, if a portion of their income is to be spent in their home country (either from savings or remittances), as documented for the US by Albert and Monras (2017).

63 Hilber and Vermeulen (2016) find that restrictive planning is an important factor in explaining high house prices and low housing affordability, particularly in London and the South East, where the average planning refusal rate (1979-2008) for major residential projects was substantially higher than in most of the rest of the country. Cavalleri et al (2021) in a study of internal labour mobility across 20 countries, finds very strong responsiveness of interregional migration flows to relative regional house prices, and that a higher elasticity of housing supply in a given region is associated with increased internal migration into that region. They also show that the UK had a much higher than average growth in interregional differences in house prices over 2005-2017.
8. CONCLUDING REMARKS

The UK’s large gaps in productivity, earnings, and living standards across regions are troubling, and the striking persistence – and indeed, over recent decades, expansion – of gaps between UK regions indicates that this problem is unlikely to fix itself. The analysis in this paper can help shed light on possible promising avenues for policy to tackle regional economic inequality.

Tackling regional economic inequality in the UK is difficult. Rising regional inequalities across most industrialised countries suggest that much of this phenomenon may be an inevitable result of long-term economic trends: scarring after deindustrialisation, and structural economic change favouring knowledge-intensive services and large urban agglomerations. One example of this is the strikingly similar trajectories of regional inequality in the UK and France, in spite of different policy regimes, as the London and Paris regions pulled away from the rest of their countries. And there is little evidence of highly successful regional economic development policies across countries in the face of these structural challenges (Ehrlich and Overman 2020b). Meanwhile, the UK’s structural challenge may be stronger than most: it was more exposed than peer countries to the global forces widening regional inequalities as London, with existing specialisation in knowledge-intensive business services, benefitted greatly from the global rise of these industries, while the Midlands, North of England, Wales, and Northern Ireland were among the most exposed regions in Europe to...
deindustrialisation. On the other hand, Germany’s experience since reunification – as the East German regions have grown much more rapidly than the West German regions (although remain much poorer), supported by policy and large-scale fiscal transfers – suggests that at least some narrowing of very large regional economic disparities is possible.

In this context, what should policy focus on? In this paper, we emphasise the large and persistent productivity differentials between London and its environs and the rest of the UK – driven largely by the underperformance of non-London cities – as the key force underlying the UK’s persistent regional economic inequality. This suggests a central focus for policy seeking to “level up” lagging regions: boosting productivity in regions outside London and the broader South of England, with a focus in particular on high productivity potential urban agglomerations.

How to do this? We analyse four (non-exhaustive) policy levers – education, infrastructure, innovation (via support for R&D), and access to finance – attempting to identify which specific areas may represent binding constraints which can be alleviated effectively through policy. We find little evidence consistent with the hypotheses (i) that low tertiary education shares remain the primary constraint on growth for the UK’s regions; (ii) that there is a generalised issue with access to finance for firms outside the South East; or (iii) that low or falling regional migration rates are to blame for the persistence of the UK’s regional economic inequalities. Instead, we find evidence consistent with (i) specific shortages of STEM degree-level skills; (ii) binding transport infrastructure constraints within major non-London conurbations; (iii) a failure of public innovation policy to support clusters beyond the South East, in particular through the regional distribution of public support for Research and Development (R&D); and (iv) missed opportunities for higher internal mobility due to London’s overheating housing market.

The relative importance of each of the factors we have identified depends on the underlying economic model. If regional economies best approximate the neoclassical model, alleviating constraints on STEM education and transport infrastructure will be paramount (alongside ensuring that general education and skills keep pace with firms’ demand). If knowledge spillovers are fundamentally important to innovative activity and the generation of high-productivity clusters in the UK’s non-London city-regions, doing more to increase public investment in R&D in areas outside London and the greater South East may be the most important. If the forces generating path dependency in formerly industrialised regions and agglomeration economies in the greater London area are too strong, it may be very difficult for policy to do much to stimulate productivity growth in the UK’s other city-regions. This might mean that the only policy tool remaining is improving housing affordability to allow people to move to opportunity – enabling equality of opportunity for individuals from each region, even if not equality of outcome across regions.

64 There are also likely to be complementarities between the different factors we evaluate. Increasing the STEM-educated workforce in regions outside London and the South East, for example, would likely boost the effectiveness of increased government support for R&D in these regions. See Bloom, Van Reenen, and Williams (2019) for an overview of the role of a STEM educated workforce in boosting innovation.
Our analysis does not allow us to distinguish which of these models best describes the UK economy. What it does illustrate, however, is that – outside the realm of education, where huge increases in attainment have occurred – the large-scale, systematic, and consistent policy action necessary under any of these three models has not in recent decades been undertaken in the UK. Transport investment has been consistently low outside London; public R&D spend has been heavily skewed towards London and the South East; and little has been done to ease London’s housing supply constraints. In a world of uncertainty about the true underlying economic model, policymakers seeking to tackle regional economic inequality would therefore do well to seek to alleviate constraints on each of these areas.

One aspect we have not been able to analyse quantitatively in this paper is the appropriate level at which government should perform these functions. The UK is one of the most centralised advanced democracies in the world (Hooghe et al. 2016), and there is some stylised and qualitative evidence for this acting as a constraint on growth (as in the recent UK2070 Commission). The effect of governance regimes on economic growth is difficult to analyse both because of difficulty quantifying different regimes and relatively few changes in governance within countries and regions over time, particularly in the UK where the move to devolution in Scotland, Wales, and Northern Ireland is relatively recent and only partial.

Nonetheless, theory and some evidence from other countries suggests that the UK’s policy centralisation may have impeded its ability to respond effectively to regional economic inequality: decentralisation may enable government to be more responsive to local needs (for example, in education or infrastructure) (Tiebout 1956; Oates 1972), and may help government utilise local, sometimes tacit knowledge, and realise coordination externalities through the delivery of simultaneous interventions in infrastructure, skills and other policy domains (Rodrik 2000, Rodrik and Sabel 2019). While evidence on the effects of governance structures on economic outcomes is difficult to come by, Bianchi, Giorcelli, and Martino (2021) provide quasi-experimental evidence for governance decentralisation driving improvements in public service provision in Italy, and earlier work by the IMF supports this finding more generally, provided sufficient thresholds are met in terms of actual expenditure and revenue-raising powers being decentralised (Razafimahefa and Sow 2015). And while there is no evidence of any correlation between the level of spending decentralisation in a country and its overall economic growth rate, spending decentralisation is associated with fewer geographic economic inequalities (Ezcurra and Rodríguez-Pose 2013), and tax decentralisation with improved regional productivity (Blöchliger and Égert 2013, Blöchliger and Akgun 2018).\(^{65}\) In addition to reducing responsiveness to local needs and knowledge, some have argued that the UK’s centralised spending rules may exacerbate regional inequalities by structurally underfunding poorer regions. Public capital investment, allocated with the aim of maximising national value for money, flows disproportionately to those areas most likely to see a higher land value uplift as a result. This creates a structural bias towards already thriving local economies (Coyle and Sensier 2020).

\(^{65}\) Decentralisation may of course come with a trade-off: greater responsiveness to local needs/preferences, but fewer efficiencies from economies of scale, and a risk of inequities or policy failure due to varying governance quality (Pike et al. 2016, Pike et al. 2020); returns to decentralisation are greater in regions with higher quality governments. (Muringani, Fitjar, and Rodríguez-Pose 2019).
While the UK has made various attempts to decentralise in recent decades – with notable successes in the emergence of the Devolved Administrations in Scotland and Wales and the re-establishment of local governance in Northern Ireland following the Good Friday Agreement – the general story is one of endemic policy churn and institutional instability in economic development (Zymek and Jones 2020). This failure to stabilise governance structures for development, alongside patchy attempts at policy devolution, could be considered an obstacle to progress in narrowing the UK’s regional inequalities – especially when we consider that London and Scotland, two of the country’s relative productivity success stories, have benefited from relative stability and decentralisation. The combination of instability in institutional arrangements, the challenges of direct measurement, and the small number of experiments in decentralisation in the UK preclude us from casting judgement here on whether or not the UK’s centralised government is a barrier to regional growth at this moment: further research assessing the UK-specific evidence for decentralisation and growth, and the role of local and national policy institutions in regional economic development, would be welcome. To this end, we have written a companion contemporary history paper summarising findings from over 80 practitioner interviews with top-level decision makers in UK regional economic development over the period 1979-2015 (Turner et al. forthcoming).

Finally, we emphasise that productivity is not the only outcome that matters. We focus on productivity because it is extremely important for living standards, both directly and indirectly. Higher productivity increases the size of the pie available for distribution, and tends to raise typical workers’ incomes and household living standards (Stansbury and Summers 2018; Oulton 2022). Higher productivity growth can also be linked to other, non-wage improvements: while real incomes after housing costs in London are not exceptionally high for typical workers, due to high housing costs in the capital, London’s lower- and middle-income residents have seen improvements in life expectancy and educational outcomes during the recent decades over which London’s productivity has risen dramatically (Office for National Statistics 2020; Department for Education 2020). Nonetheless, a focus on productivity should not come at the expense of attention to real incomes, poverty, costs of living, health, and other measures of wellbeing. London’s case illustrates that improving productivity can, but does not automatically, improve wellbeing: the share of households in poverty, after housing costs, is much higher in London than in any other UK region (Agrawal and Phillips 2020). Improvements in regional productivity are necessary to tackle regional economic inequality, but are not sufficient to generate welfare for the region’s residents unless policy action is also taken to ensure the fruits of this growth are broadly shared.
9. REFERENCES

Agrawal, Sarthak, and David Phillips. 2020. ‘ Catching up or Falling behind? Geographical Inequalities in the UK and How They Have Changed in Recent Years’. The Institute for Fiscal Studies, p2.


———. n.d. ‘Researching R&D’. tomforth.co.uk/researchingresearch.


Haldane, Andrew. 2016. ‘One Car, Two Car, Red Car, Blue Car’. *Speech given at Materials Processing Institute Redcar*.


Özgüzel, Cem. 2020a. ‘Agglomeration Economies in Great Britain’.


10. APPENDIX: FIGURES AND TABLES

Appendix Figure 1. Regional GVA per worker and Finance and Business Services employment share, 1980 and 2018

Source: ARDECO. Note: Portugal and Luxembourg are excluded from the 1980 graph.

Appendix Figure 2. GDP per capita and tertiary education share, 2018

Source: Gothenburg QOG data. Note: Lines of best fit weighted by population; bubble size represents population. All bubbles represent NUTS2 regions except for London, for which we use the NUTS1 London region rather than the 5 NUTS2 sub-regions.
Appendix Figure 3. Change in education shares of population, 2000-2018

Source: Gothenburg QOG data. NUTS2 regions. Note: Sorted by change in tertiary share. Missing data on Merseyside, Chester, West Central Scotland, Eastern Scotland, Southern Scotland, South Western Scotland, due to changes in regional definitions.

Appendix Figure 4. Industry productivity in Western Europe’s 16 most industrialised regions in 1980

Source: ARDECO
### Appendix Table 1. Biggest 10-year periods of deindustrialisation, EU regions, 1980-2019

<table>
<thead>
<tr>
<th>Region</th>
<th>Country</th>
<th>Period</th>
<th>Change in industry employment share</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sachsen-Anhalt</td>
<td>DE</td>
<td>1991-2001</td>
<td>-0.17</td>
</tr>
<tr>
<td>Sachsen</td>
<td>DE</td>
<td>1991-2001</td>
<td>-0.16</td>
</tr>
<tr>
<td>Thuringen</td>
<td>DE</td>
<td>1991-2001</td>
<td>-0.15</td>
</tr>
<tr>
<td>West Midlands</td>
<td>UK</td>
<td>1980-1990</td>
<td>-0.14</td>
</tr>
<tr>
<td>Yorkshire and The Humber</td>
<td>UK</td>
<td>1980-1990</td>
<td>-0.14</td>
</tr>
<tr>
<td>Macregiunea trei</td>
<td>RO</td>
<td>1990-2000</td>
<td>-0.14</td>
</tr>
<tr>
<td>Macregiunea patru</td>
<td>RO</td>
<td>1990-2000</td>
<td>-0.14</td>
</tr>
<tr>
<td>Brandenburg</td>
<td>DE</td>
<td>1991-2001</td>
<td>-0.13</td>
</tr>
<tr>
<td>East Midlands</td>
<td>UK</td>
<td>1980-1990</td>
<td>-0.13</td>
</tr>
<tr>
<td>East Midlands</td>
<td>UK</td>
<td>1982-1992</td>
<td>-0.13</td>
</tr>
<tr>
<td>East Midlands</td>
<td>UK</td>
<td>1983-1993</td>
<td>-0.13</td>
</tr>
<tr>
<td>East Midlands</td>
<td>UK</td>
<td>1981-1991</td>
<td>-0.13</td>
</tr>
<tr>
<td>Yorkshire and The Humber</td>
<td>UK</td>
<td>1981-1991</td>
<td>-0.13</td>
</tr>
<tr>
<td>West Midlands</td>
<td>UK</td>
<td>1981-1991</td>
<td>-0.12</td>
</tr>
<tr>
<td>Yorkshire and The Humber</td>
<td>UK</td>
<td>1982-1992</td>
<td>-0.12</td>
</tr>
<tr>
<td>Macregiunea trei</td>
<td>RO</td>
<td>1991-2001</td>
<td>-0.12</td>
</tr>
<tr>
<td>Macregiunea patru</td>
<td>RO</td>
<td>1991-2001</td>
<td>-0.12</td>
</tr>
<tr>
<td>West Midlands</td>
<td>UK</td>
<td>1982-1992</td>
<td>-0.12</td>
</tr>
<tr>
<td>Yorkshire and The Humber</td>
<td>UK</td>
<td>1983-1993</td>
<td>-0.11</td>
</tr>
<tr>
<td>West Midlands</td>
<td>UK</td>
<td>1983-1993</td>
<td>-0.11</td>
</tr>
<tr>
<td>Lietuva</td>
<td>LT</td>
<td>1990-2000</td>
<td>-0.11</td>
</tr>
<tr>
<td>West Midlands</td>
<td>UK</td>
<td>1997-2007</td>
<td>-0.11</td>
</tr>
<tr>
<td>Wales</td>
<td>UK</td>
<td>1980-1990</td>
<td>-0.11</td>
</tr>
<tr>
<td>West Midlands</td>
<td>UK</td>
<td>1998-2008</td>
<td>-0.11</td>
</tr>
<tr>
<td>East Midlands</td>
<td>UK</td>
<td>1984-1994</td>
<td>-0.11</td>
</tr>
<tr>
<td>West Midlands</td>
<td>UK</td>
<td>1996-2006</td>
<td>-0.11</td>
</tr>
<tr>
<td>Macregiunea unu</td>
<td>RO</td>
<td>1990-2000</td>
<td>-0.11</td>
</tr>
<tr>
<td>West Midlands</td>
<td>UK</td>
<td>1999-2009</td>
<td>-0.1</td>
</tr>
<tr>
<td>West Midlands</td>
<td>UK</td>
<td>1995-2005</td>
<td>-0.1</td>
</tr>
<tr>
<td>East Midlands</td>
<td>UK</td>
<td>1985-1995</td>
<td>-0.1</td>
</tr>
<tr>
<td>East Midlands</td>
<td>UK</td>
<td>1986-1996</td>
<td>-0.1</td>
</tr>
</tbody>
</table>

Source: ARDECO
Appendix Figure 5. Transport investment as a share of GDP across countries, 1995-2019

![Graph showing transport investment as a share of GDP for various countries from 1995 to 2019.]

Source: OECD

Appendix Figure 6. Transport spending per capita across English regions

![Graph showing capital and current spending per capita across English regions from 2000 to 2020.]

Appendix Figure 7. Hours lost to congestion per driver, US and Western European cities

Source: INRIX 2021 Global Traffic Scorecard data. 2019 hours lost to congestion is calculated as reported 2021 hours lost to congestion, adjusted by the reported % change from pre-COVID.

Appendix Figure 8. Train cancellations by operator, Great Britain 2014-19

Source: Office of Rail and Road passenger rail performance, Tables 3123 and 3133. Note: Quarterly data is averaged over 2014-2019 inclusive. We categorized operators primarily serving London and the South of England as Chiltern Railways (CR), Govia Thameslink (GT), Greater Anglia (GA), Heathrow Express (HE), London Overground (LO), South Western (SW), Southeastern (SE), TfL Rail (TIL), and c2c (c2c). The remaining operators are Hull Trains (HT), TransPennine Express (TPE), London North Eastern (LNE), Avanti West Coast (AWC), CrossCountry (CC), Grand Central (GC), Caledonian Sleeper (CS), Northern Trains (NT), Great Western (GW), East Midlands Railway (EM), West Midlands Trains (WM), TfW Rail (TfW), and ScotRail (SR).
Appendix Figure 9. Interregional mobility rates in the G7, 2014-19

Source: OECD Regional Statistics. Note: Inter-regional mobility rate is defined as the % of the total country’s population who were living in a different region in the same country the previous year (where regions are defined at the OECD TL2 level). Regions FRY and ITZZ are dropped. France averages 2014-18, the US averages 2015, 17, 19. All other countries averages 2014-19 inclusive.

Appendix Figure 10. Net internal migration into London, by age

Source: Office for National Statistics.