

Supplemental Appendix for:
“Are Complementary Policies Substitutes?
Evidence from R&D Subsidies in the UK”

Jacquelyn Pless

March 13, 2025

There are four appendices associated with this paper:

Appendix **A**: Data Preparation

Appendix **B**: Additional Robustness Checks

Appendix **C**: Additional Figures

Appendix **D**: Additional Tables

A Appendix: Data Preparation

This appendix details the process I followed for preparing the data sets and the notable results associated with the matching.

A.1 Data Preparation for Small Firm Analysis

Direct Grants for R&D.—To study the effects of R&D grants and tax credits on small firms, I combine two main data sets. I focus on firms in the United Kingdom that received grants through Innovate UK, the UK’s largest public funding body for private sector innovation. I start with the program’s public database of all grants allocated since its inception and gather information on grants provided from 2005 through 2017.¹ The database contains details such as the grant award date, recipient, total grant award amount, proposed project costs, competition title and number, etc. Importantly, it includes unique company registration numbers (CRNs) so that firms can be matched over time and to other firm-level data sets.

I carry out a few data management and cleaning steps to prepare the data before matching to other data sets. I omit observations for which projects were withdrawn and observations without firm identifiers – which primarily include those associated with academic institutions, public service organizations (PSOs), charities, etc. and thus not entities to which the Innovate UK funding thresholds apply. Next, I search for and drop any remaining organizations of these types that did have some information in the firm identifier field, so they were not dropped in the first step, as well as grants provided to projects associated with the Catapult Network Centres, as these organizations tend to be also supported through a network that formally provides expertise and facilities to accelerate the application and scaling of their research. It is thus difficult to disentangle whether changes in R&D are associated with the grant funding itself or other resources from the Catapult Centres. I also omit grants noted as supporting activities like procurement and partnership building as well as others for which the grant funding thresholds do not apply, such as “vouchers,” which are designed strictly for all smaller firms to seek expert advice.

The data set at this stage contains 16,250 observations across 7,492 organizations. I then examine the firm identifiers for obvious errors like non-standard formats and omit observations for which the CRN takes on the format of a non-UK country. I also drop observations that appear to include clear data entry errors in other grant-related variables, such as cases where the total amount of funding provided exceeds the project’s proposed

¹I use the “competition year” to determine the grant award date, and since these dates align with the application periods and span two calendar years, I use the second part of the competition time period, which is more likely to be associated with the year in which firms actually receive grants after the review process.

costs (as this would not align with the program’s funding rules) and cases where the start of the project is later than one year after the grant year. This process drops only a small number of observations and leads to the data set containing 16,077 observations across 7,399 organizations.

Finally, since some organizations receive more than one grant per year, I collapse the data to the firm-year level for matching purposes, considering the “grant year” to be the second part of the competition’s fiscal year. I sum quantitative variables such as grant amount and the number of grants received, and for non-quantitative variables, I use the value of the first observation for grants the firm receives that year, such as organization name and competition title.² The final clean Innovate UK data set that I match to other firm-level data contains 12,072 observations across 7,387 unique organizations for firms receiving grants between 2005 and 2017. On average, firms receive 1.63 grants over this time period.

Firm Balance Sheets and Profit and Loss Statement Data (Bureau van Dijk).—I collect firm-level financial data, such as R&D expenditures, employment, total assets, turnover, etc., from Bureau van Dijk’s Financial Analysis Made Easy (FAME) Database, which is a commercial data set containing detailed balance sheet and P&L data on companies and unincorporated businesses in the UK and Ireland. Its construction begins with official filings content from the UK’s Companies House and is enriched with additional efforts to ensure accuracy and fill in some gaps. Overall, the database covers over 11 million companies, including 2 million in a detailed format, 1.3 million companies that are active but have not yet filed accounts or are not required to file, and 6 million companies that are no longer active.

In the UK, all limited, PLC, LLP and LP companies are required to file accounts with Companies House, so the data capture at least basic information on the universe of firms in the UK, representing about 1 million companies as of 2015 in FAME (Kalemli-Ozcan and Yesiltas 2015). All companies in the UK must keep accounting records of all money received and expended, assets, and liabilities and report this information. However, as I discuss more below, small firms had the option to report less detailed accounts over my study period. The FAME data provides the latest account date, but some firms report quarterly. I follow Kalemli-Ozcan and Yesiltas (2015) and define the filing year based on the year of the latest filing date if the date is June 1 or later, and otherwise, I use the preceding year.

Matching Innovate UK Grant Data to Bureau van Dijk’s Financial Data.—I next match the prepared firm-year Innovate UK grant data with the balance sheet and P&L statement data from FAME. Of the 12,072 firm-year observations in the grant data, 11,750 observations

²This limits the usefulness of some of the non-quantitative grant-specific variables, so I primarily only use it when carrying out a few robustness checks, like when including competition-level fixed effects.

(97.3%) across 7,146 organizations match when merging on company registration numbers. Notably, the number of observations that did not match in years that I define as being the “low tax credit rate” (2005-2012) and “high tax credit rate” (2013-2017) periods when studying policy interactions for small firms is very similar.³

I apply a few standard data preparation rules to be sure that I omit potential data entry errors for key financial variables. For example, I drop observations for which variables like total assets are negative or when the observed year is earlier the firm’s founding year, and I convert monetary figures into real 2010 terms using the World Bank’s Consumer Price Index. Once this is complete, I again drop any observations associated with firms that did not receive a grant from 2005 to 2017 in this remaining “clean” sample. The data set at this stage is an unbalanced panel of 72,481 observations across 7,035 firms that received Innovate UK grants between 2005 and 2017, and these firms received an average of 1.72 grants over this time period.

Throughout the majority of the analyses when examining the effects of grant funding and policy interactions on small firms, I limit the sample to just the year in which firms receive a grant and the three years that follow. This provides a data set of 28,796 observations across 7,035 firms. Furthermore, for most of the estimations throughout the paper, I limit the sample to firms that also meet the total assets and turnover criteria for higher grant rate generosity as described below, which reduces the sample to 20,398 observations across 5,301 firms. The sample size then declines more substantially when restricting the data to firms with non-missing employment data in the year prior to receiving a grant, which is required for defining treatment status. I discuss this and the implications below as well as in the main text.

Advantages of Using FAME for Small Firms.—There are two key advantages of using FAME to study small firms relative to some other R&D data sets frequently used to study innovation in the UK that are important for my paper, such as the Business Enterprise Research and Development (BERD) survey. First, BERD has been shown to vastly under-cover small firms, and second, the sampling population conditions on already being an R&D performer. I discuss these challenges more below.

One of the key advantages of using FAME to study small firms over BERD is that BvD starts with a comprehensive list of firms from Companies House (and the data reported therein) and then enhances the officially-reported data with its own research. On the other hand, while BERD is well-suited for studying larger firms—which has been used in many other studies of UK innovation policies and which I indeed use to study larger firms as

³Only 189 observations did not match in the “low” period and 151 did not match in the “high” period.

well—and especially those that are already R&D performers, small firms are vastly under-covered. The sampling frame that ONS uses to identify firms for the BERD survey is a list of all known R&D performers in Great Britain, whereby R&D-performers are identified from responses on other business surveys (like the Annual Business Survey (ABS)). The top 400 businesses according to the size of previously reported R&D expenditures are selected automatically and then another 3,600 firms are randomly selected from the remaining firms on the pre-determined list of R&D performers (ONS 2022b).

Importantly, most of these “feeder surveys” from which firms are sampled carry out a census of all large businesses (i.e., those with more than 250 employees by their standard definition), so even though the random sample may consist of firms with lower R&D expenditures than the “biggest” R&D-performers, many small firms according to standard definitions are never included in these census surveys and thus are never covered by the BERD survey. Instead, data for non-sampled businesses are estimated and imputed using a ratio method based on employment.

The under-coverage of small firms in BERD is striking. The HMRC’s official R&D statistics historically indicated that R&D is higher than what BERD statistics suggest. For example, for the financial year 2020/21, R&D expenditures were 42% higher than the BERD estimates. To try to better-understand this divergence, the ONS and HMRC carried out a microdata sharing project and concluded that the overwhelming majority of it could be explained by the under-coverage of small firms (ONS 2022a). Using their newly-developed “uplift” method to assess the degree of under-representation, the report finds that the value of expenditures performed by UK firms would have been 43 billion GBP compared to the 26.9 billion GBP previously reported if the necessary methodological adjustments were made. Likewise, R&D expenditures would have been 15 and 15.6 billion higher in 2018 and 2019. While there are some other differences between HMRC’s and BERD’s data collection and R&D measurement methods that can explain small proportions of this difference, the report concludes that most of the difference is due to the small firm under-coverage issue.

Second, the sampling approach used to construct the BERD survey population conditions on firms already being R&D performers. Part of the objective of grant programs is to help firms overcome financial constraints, which are more likely to be firms with no previous R&D, and thus many firms of interest were likely never included in BERD. Part of my objective also is to study whether the combination of policies induced R&D entry—an indication of potentially overcoming financial constraints—which could not be done if all firms in the sample were already investing in R&D.

Final variable construction.—Lastly, I construct a few final variables required for the analyses. I convert monetary variables into real 2010 terms using the World Bank’s CPI for GBP,

and I convert total assets and turnover into euros using each year’s average exchange rate for the purposes of defining treatment status, as the Innovate UK generosity thresholds are defined using euros. For R&D expenditures, I consider all missing data to be zeros. I discuss the rationale and implications for this more below and in the main text and I also carry out a number of robustness checks to help ensure this assumption is not biasing the results. Essentially, though, the idea is that because I am already conditioning on firms that choose to report employment (since employment is the running variable), missing data for financial variables likely represent true zeros, as these pieces of information likely would have been reported as well if firms already choose to report employment.

To define whether the firm is treated and benefits from more generous grant rates, I use the values of employment, total assets, and turnover from the year prior to receiving a grant. To be eligible for the higher grant rates, firms must have fewer than 50 employees and either total assets or turnover must be lower than 10m euros. Since employment is the binding criteria, I restrict the estimation sample to include only firms that meet the total assets and/or turnover requirements in the year prior to receiving a grant and then define treatment using the 50-employee threshold. This allows me to take all three eligibility criteria into account and to rely on the binding criteria for treatment status. If both turnover and total assets data are missing in the year prior to receiving a grant, I keep them in the data set to preserve sample size but show that the results are robust to not imposing the turnover and assets restrictions as well.

I apply the same treatment status for the three years that follow grant receipt when constructing the treatment and running variables. That is, if a firm receives a grant in 2014, I use the 2013 employment, total assets, and turnover values to define treatment, and if it meets the eligibility requirements for the small firm grant generosity rate in 2013, I consider the firm treated in 2014, 2015, and 2016 as well (for both the treatment indicator and running variable construction). If the firm receives another grant during those years, I update the treatment status using the same procedure, replacing the original treatment variables with updated ones reflecting status associated with the most recent grant received.

Variable coverage.—A 97.2% match rate between the Innovate UK and FAME data suggests that BvD’s coverage of firms in FAME is quite comprehensive, even for relatively small firms. This is because BvD starts by drawing company information from Companies House, where all firms in the UK are required to file accounts with at least some basic information. At the same time, this matching rate does not reflect the coverage in my data set, since not all variables are fully populated, even with BvD’s additional research supplementing the Companies House data. This is because not all firms are required to report all information.

The first variable for which this is particularly important is employment, the running

variable used in the research designs described in Section ???. There is indeed a significant portion of observations for which employment data are missing. Of the 28,796 observations across 7,035 firms in the starting “cleaned” sample (which includes years in which firms receive grants and the three years that follow), current employment information is included for 15,111 observations (52.5%) across 4,756 (67.6%) of firms. Lagged employment is populated for 13,037 observations (45.3%) across 4,204 firms (59.8%). When considering lagged employment in the year prior to receiving a grant, which is what I use to define treatment status and then apply to all years associated with the grant and is thus most relevant for my estimation, 11,482 observations (40%) across 2,975 firms (42.3%) have populated data.

The most relevant sub-sample to consider in terms of coverage rates is the final starting data set containing 20,398 observations across 5,301 firms that I use throughout most of the paper (conditioning on also meeting the total assets and turnover eligibility requirements for more generous grant rates, as described above). Of this data, 4,703 observations (23%) across 1,614 firms (30.4%) have employment data for defining treatment status. Although these coverage rates might seem low, the sample size for matched firms with employment information is still much larger than when attempting to match Innovate UK data to BERD data that is not imputed.⁴ It is also likely a more representative sample of the types of firms I aim to study since it does not condition on already being an R&D performer.

The main potential concern that missing data introduces from an identification strategy perspective is that firms under the 50-employee threshold that choose to report certain information may differ systematically relative to those that do not. As such, the second variable for which it is important to consider missing data is R&D, the main outcome of interest. Throughout my analyses, I assume all missing R&D data are zeros. See the main text of the paper for a discussion of what this implies for my estimation. The main potential concern with doing this is that small firms in the UK can opt out of reporting all information on P&L statements. It is first worth noting that firms in my baseline estimation sample are more likely to report R&D (if they do invest) because I already inherently condition on reporting employment given the treatment variable definition. If firms report employment, missing information for R&D is more likely to represent a true zero. That said, if firms under the 50-employee threshold do choose to not report R&D even though they report employment, this could put either upward or downward pressure on the estimates if they systematically differ from firms that do not report R&D (yet do invest). I carry out tests that suggest it is likely not a threat in my setting in Section ??? and Section ???.

⁴I discuss the limitations of using BERD to study small firms in more detail above.

A.2 Data Preparation for Larger Firm Analysis

UK Data Services Secure Lab.—The regression analysis for large firms entails linking several microbusiness datasets that are legally protected and held by the UK’s Office of National Statistics (ONS). Accessing the data requires a special process, which begins with training and taking an exam regarding the use and protection of sensitive data to become a UK Accredited Researcher. A research proposal then must be submitted and approved, justifying the use of the data sets and providing the reasons that they must be accessed and linked in order to answer a question that is relevant for the UK’s public good. Once approved, all data use and analysis must be conducted in the UK Data Service Secure Lab.

Firm R&D Expenditures.—I use data from Business Enterprise Research and Development (BERD) survey to study R&D expenditures of large firms. The BERD survey is conducted by the ONS following the Frascati Manual methodology (OECD 2002), collecting information on R&D expenditures and other characteristics of firms identified as actively performing R&D. A stratified sampling approach is employed to select which enterprises will receive a BERD questionnaire. The ONS primarily uses the Annual Business Survey (ABS) to identify R&D-performing firms as well some other data sources such as the UK Community Innovation Survey and HMRC data on firms claiming R&D tax credits.

All questionnaires sent to those selected include a minimum set of questions on total R&D spending and R&D employment. The largest spenders on R&D receive “long form” questionnaires and the remainder receive a “short form.” The short form asks for basic information related to R&D, such as in-house and extramural expenditures and total headcount of R&D employees. The long form covers more detailed information, such as how R&D expenditures are spent based upon capital and non-capital expenditures. Enterprises not included in the stratified sampling, and responses to questions on the long form from firms that were just sent a short form, have imputed values. These are the mean values of the variable as a share of employment in the firm’s size band-sector group.

I collect BERD data from 2000 through 2014 and omit defense-related R&D investments. The full BERD datasets begin with about 30,000 observations per year. I take a number of steps to prepare the data for analysis. First, I do not use imputed values in order to avoid introducing measurement error. Omitting observations with imputed responses for the key outcome variable of interest (R&D expenditures) reduces the sample size to about 2,500 observations per year. Next, I omit observations where the Inter-Departmental Business Registrar (IDBR) reporting unit number seems as though it was recorded incorrectly due to taking on the wrong format. I also drop observations where the IDBR is duplicated, as there is no consistent way of understanding which entry is correct when the responses do

not align. In total, this process results in dropping <0.01 percent of the observations.

Finally, the BERD responses are observed at the IDBR reporting-unit level, but funding and tax credit eligibility rules are determined by firm characteristics at the “enterprise group” level, which is a larger statistical unit. The EU Regulation on Statistical Units defines enterprise groups as “an association of enterprises bound together by legal and/or financial links” (EEC 696/93). The reporting unit level is associated with a geographical unit, whereas enterprise groups capture all reporting units associated with an enterprise.

The BERD datasets for each year include all reporting unit-year observations that were identified by ONS as firms performing R&D in the UK, yet the assignment to treatment in this analysis depends on whether the enterprise group satisfies the eligibility criteria. I aggregate the BERD data to the enterprise group level so that it can be matched to the Business Structure Database (BSD), which provides data on the enterprise group’s total employment, and so that the R&D expenditure data captures the entire enterprise group’s R&D investment levels. Furthermore, the location where R&D funds are allocated to an enterprise might not be the same local-level reporting level that is observed in BERD. This aggregation process results in only a very small further reduction in the sample size (usually less than 100 observations per year). The final BERD data set consists of about 2,000 to 2,500 enterprise groups per year from 2000 through 2014.

Determining Funding Level Eligibility.—I use the UK’s Business Structure Database (BSD) to determine each enterprise group’s tax credit rate eligibility. The BSD is also held securely by the ONS and requires UK Data Services Secure Lab access. It includes information on a small set of variables for nearly all businesses in the UK, and since it allows for one to observe a reporting unit’s enterprise group, I use this to determine each enterprise group’s employment level and thus tax credit rate eligibility. The data are derived mostly from the IDBR, which is a live register of administrative data collected by HM Revenue and Customs including all businesses that are liable for VAT and/or has at least one member of staff registered for the Pay As You Earn (PAYE) tax collection system. The BSD only misses very small businesses, such as those that are self-employed, and covers almost 99 percent of the UK’s economic activity.

The BSD annual datasets include variables such as local unit-level and enterprise-level employment, turnover, company start-up date, postcodes, and the Standard Industrial Classification (SIC). I aggregate variables to the enterprise group level. If the observation is missing an enterprise number and does not belong to a larger enterprise group, I use the given observation’s values for each variable. There are about 3 million observations per year. The enterprise group numbers are anonymous but unique so that they can be linked to other data sets held by the ONS.

Final Data Sample Preparation.—A few final steps are taken to prepare the data. First, all expenditure and financial variables are converted into real 2010 terms using the World Bank’s Consumer Price Index. Observations associated with inactive firms are dropped from the sample, which results in dropping only 72 observations, and I omit the top 1% of the R&D distribution to address the highly skewed nature of R&D investments. The final data set includes about 2,000 to 2,500 firms/enterprise groups per year from 2000 through 2014.

B Appendix: Additional Robustness Tests

B.1 Other Policy Changes in 2013

As noted in Section ??, one potential concern is that the UK’s Patent Box policy (introduced in 2013) could confound the main policy interaction results if it enhanced the impact of grant funding. There are several reasons to believe this is not a significant threat. Experts have argued that it is poorly targeted for promoting research (Griffith, Miller and O’Connell 2010), though, consistent with evidence in the literature on patent box policies as well. While patent box regimes reduce patent transfers out of countries, they do not seem to significantly affect R&D (or patenting levels) (Gaessler, Hall and Harhoff 2021; Alstadsaeter, Barrios, Nicodeme, Skonieczna and Vezzani 2018).⁵ See Hall (2022) for a comprehensive review.

Second, data on patent box and R&D tax credit claims suggest that the latter are much more important for small firms. I gathered data on the number and value of claims for both policies from the HMRC’s official public statistics (see Appendix Table D.14). On average, between fiscal years 2013/14 and 2016/17, small firms filed only 250 patent box claims per year on average (totaling £6.3 million per year). To make the statistics comparable to tax credit claims, which are provided for SMEs as a whole, I also gather data on medium-sized firms, which filed 249 patent box claims per year (£21.7 million per year) for a sum of 499 patent box claims by SMEs per year. In contrast, SMEs made an average of 31,976 R&D tax relief claims per year (totaling £1,511 million per year) over this time period, which is about 64 times the number of annual patent box claims. One complication with this comparison is that SMEs under the tax credit policy include firms up to 500 employees whereas the cutoff is 250 for the patent box policy, but even if small firms account for only 10% of all R&D tax credit claims, this still amounts to 12.8 times more than patent box claims by small firms.

Results from two empirical tests also provide assurance that the patent box policy is probably not driving the results. First, omitting firms in the manufacturing and wholesale and trade industries—which make up more than 50% of patent claims each year—does not weaken the findings (see Appendix Table D.15).⁶ If the patent box policy was a major issue in my setting, one might expect a lower or non-existent policy interaction effect for firms in industries that do not make many claims. Second, one might expect firms with a longer track record of investing in R&D to benefit most from the patent box policy, as they are more likely to already own more patents or be closer to filing patents. The baseline regressions already

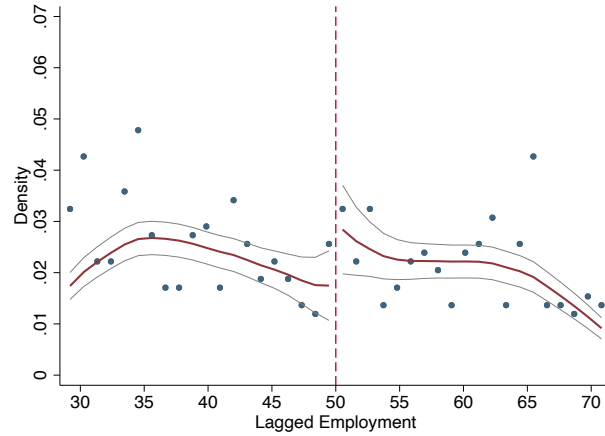
⁵One exception is Mohnen, Vankan and Verspagen (2017), who find that the patent box policy in the Netherlands is associated with more R&D person-hours. However, the Netherlands’ policy uniquely covers non-patentable R&D as well.

⁶Firms with SICs of 10 to 33 (manufacturing) and 45 to 47 (wholesale and trade) are excluded.

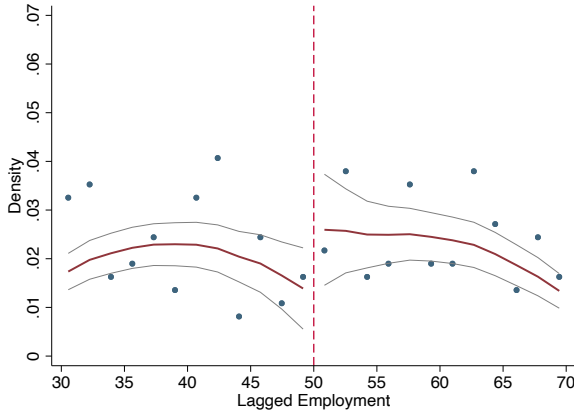
control for cumulative R&D, but as shown later in Section ??, the effects are particularly substantial for firms with *zero* cumulative R&D if anything (Column 2 of Table ??).

C Appendix: Additional Figures

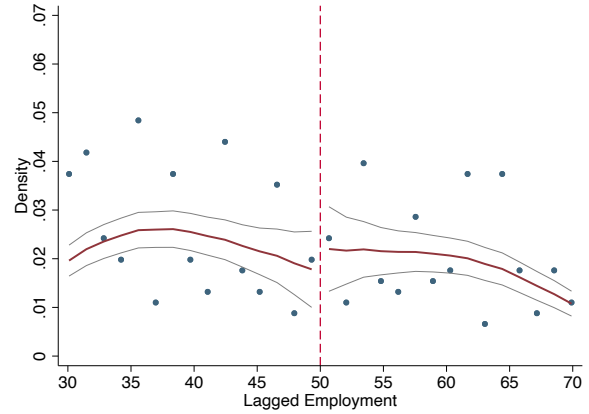
Figure C.1: McCrary Tests for No Manipulation of Firm Size at the Small Firm Grant Generosity Threshold



(a) Full Sample Period (2005-2017)



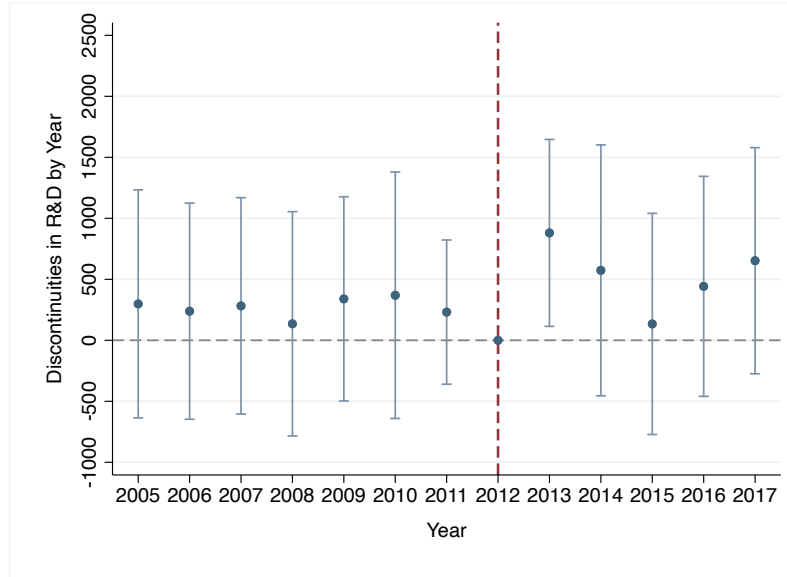
(b) Low Tax Credit Rate (2005-2012)



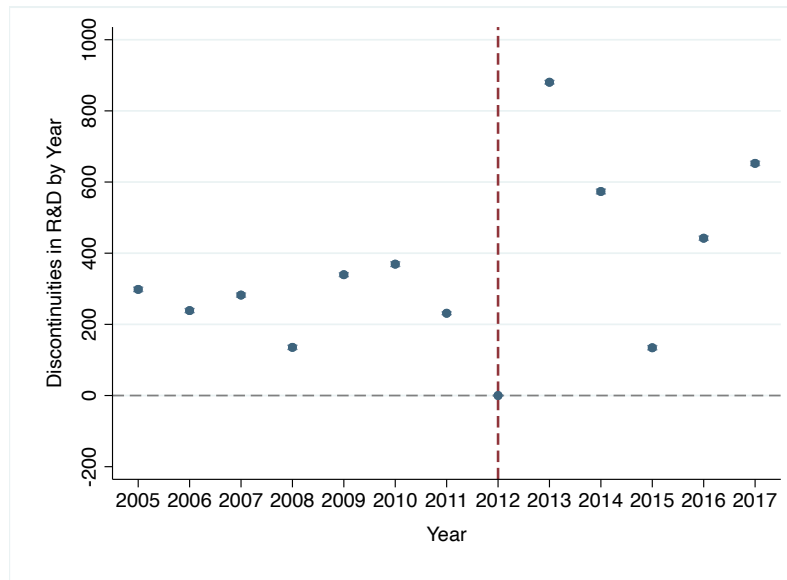
(c) High Tax Credit Rate (2013-2017)

Note: Figures provide McCrary tests for discontinuities in the distribution density of employment for the full sample period (Panel A) and then separately for the low tax credit rate period (Panel B) and high tax credit rate period (Panel C). Sample includes firms with 29 to 71 employees (lagged) that also meet the turnover and total assets grant generosity eligibility requirements. The discontinuity estimates (log differences in density height) at the small firm threshold (and standard errors) are 0.540 (0.283) in Panel A, 0.792 (0.462) in Panel B, and 0.267 (0.344) in Panel C. The discontinuities are not statistically different from zero and a t-test for the statistical difference between the low and high tax credit periods' discontinuities suggests that the diff-in-disc in the firm size distribution is also not statistically significant.

Figure C.2: Event Study Plot of Discontinuities in R&D by Year



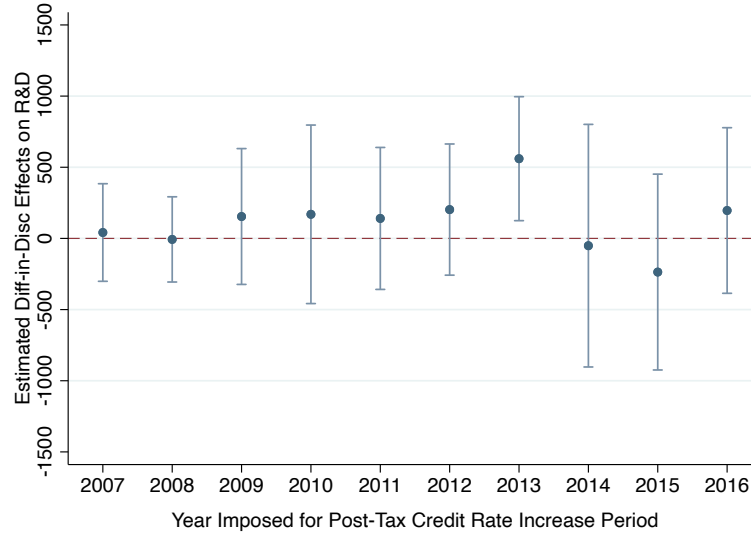
(a) **Estimates with 95% Confidence Intervals**



(b) **Estimates without Confidence Intervals**

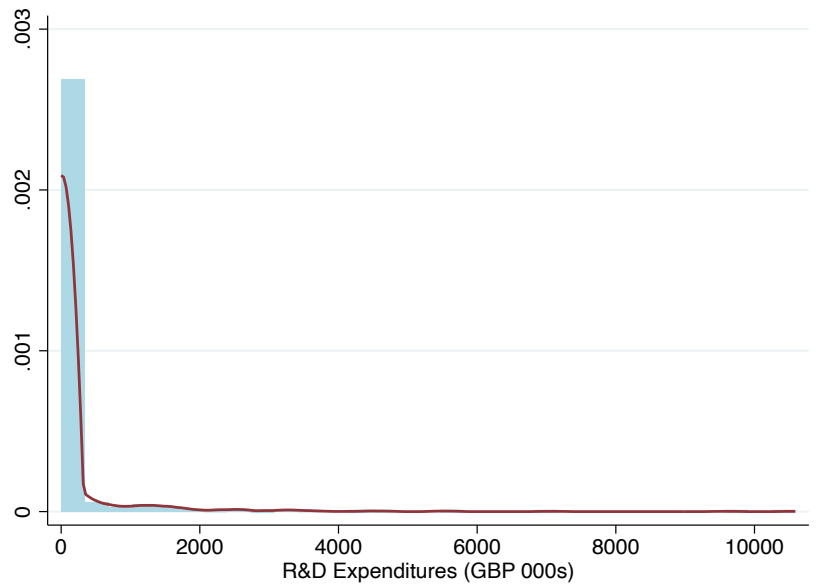
Note: Figure plots yearly discontinuities in R&D, which are found by estimating the diff-in-disc model of Equation ?? but interacting the treatment dummy and running variables with indicators equal to one for each year (using 2012 as the reference period). All fixed effects and controls are included in the regression. Panel A plots the results with 95% confidence intervals and Panel B omits the confidence intervals so the difference in coefficient estimate magnitudes is more transparent. Coefficient estimates and their standard errors are provided in Appendix Table D.10 for reference. The average of the point estimates in the low tax credit rate period is 270.7 and the average in the high tax credit rate period is 536.6.

Figure C.3: Placebo Tests for Tax Credit Rate Increase Timing



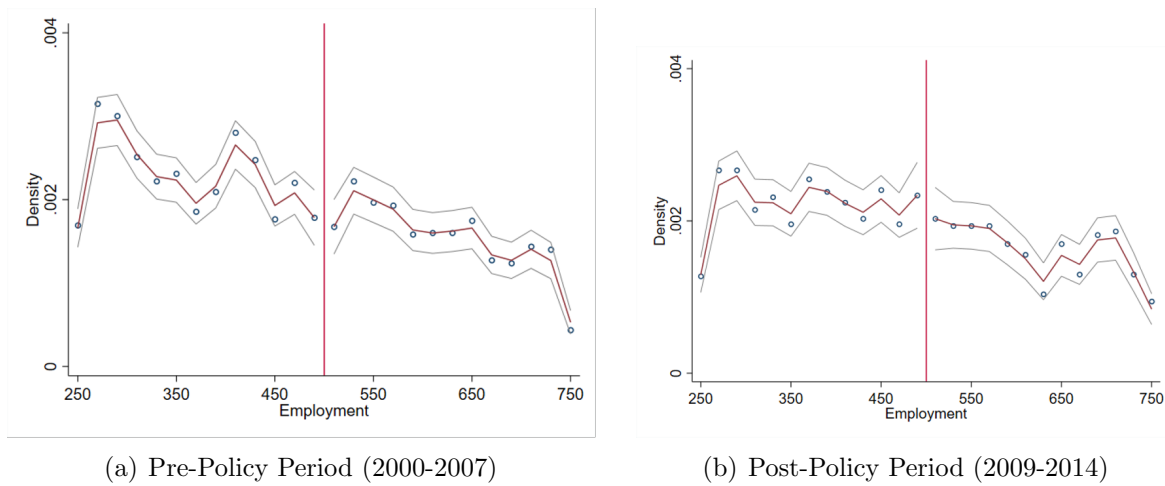
Note: Figure plots diff-in-disc estimates and their 95% confidence intervals from when estimating separate equations that impose different years as “pseudo” policy change years. For example, the 2007 coefficient is associated with estimating the diff-in-disc model of Equation ?? but assuming that tax credit rates enter a high tax credit rate period in 2007 rather than 2013. As expected, the only case in which there is a large, positive, and statistically significant diff-in-disc is for the actual treatment year (2013).

Figure C.4: Distribution of R&D Expenditures for Baseline Sample



Note: Figure plots R&D expenditure distribution and kernel density function (epanechnikov) for baseline estimation sample of firms with 29 to 71 employees including observations in the year in which firms receive grants and the three years that follow and after omitting the top 1% of the non-zero R&D distribution.

Figure C.5: McCrary Tests for No Manipulation of Firm Size at the “Larger” Firm Tax Credit Rate Generosity Threshold



Note: Figures provide McCrary tests for discontinuities in the distribution density of employment for firms with 250 to 750 employees in the pre-tax credit rate policy period (2000-2007) in Panel A and post-tax credit rate policy period (2009-2014) in Panel B. The discontinuity estimates (log differences in density height) at the 500-employee firm threshold (and standard errors) are -0.063 (0.284) in Panel A and -0.141 (0.289) in Panel B. The discontinuities and the difference in discontinuities are not statistically different from zero. Figure was created using data from the UK’s Business Structure Database and Business Enterprise Research and Development Database, Department for Business, Innovation, and Skills, Office for National Statistics.

D Appendix: Additional Tables

Table D.1: R&D Tax Credit Rates for Small and Medium-Sized Enterprises

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: R&D Tax Relief Scheme Enhancement Rates and Benefits								
Year	Enhancement Rate	Payable Credit	Low Corp. Tax	Main Corp. Tax	Profit-Making % Benefit		Loss-Making % Benefit	Average % Benefit
					Low Tax	Main Tax	No Tax	
2005	0.50	0.16	0.19	0.30	0.095	0.150	0.240	0.162
2006	0.50	0.16	0.19	0.30	0.095	0.150	0.240	0.162
2007	0.50	0.16	0.20	0.30	0.100	0.150	0.240	0.163
2008	0.75	0.14	0.21	0.28	0.158	0.210	0.245	0.204
2009	0.75	0.14	0.21	0.28	0.158	0.210	0.245	0.204
2010	0.75	0.14	0.21	0.28	0.158	0.210	0.245	0.204
2011	1.00	0.13	0.20	0.26	0.200	0.260	0.250	0.237
2012	1.25	0.11	0.20	0.24	0.250	0.300	0.248	0.266
2013	1.25	0.11	0.20	0.23	0.250	0.288	0.248	0.262
2014	1.25	0.15	0.20	0.21	0.250	0.263	0.326	0.280
2015	1.30	0.15	0.20	0.20	0.260	0.260	0.334	0.285
2016	1.30	0.15	0.20	0.20	0.260	0.260	0.334	0.285
2017	1.30	0.15	0.19	0.19	0.247	0.247	0.334	0.276

Panel B: Averages and Changes Between “Low” and “High” Rate Periods

	Low TC Period Avg. (2005-2012)	High TC Period Avg. (2013-2017)	Percent. Point Difference	% Change
Enhancement Rate	0.750	1.280	0.530	0.707
Tax Credit Benefit	0.200	0.277	0.077	0.385

Notes: Panel A provides each component that goes into determining the R&D tax credit benefit for firms qualifying as small- and medium-sized enterprises (SMEs) under the R&D tax relief scheme receive as well as the calculated benefit rates. The inputs into the equation include R&D enhancement rates (Column 1), the payable credit rate for loss-making firms (Column 2), the corporate tax rate for profit-making firms making less than 300k in profits (Column 3), and the main corporate tax rate for firms making more than 300k in profits (Column 4). Columns 5 and 6 provide the resulting benefit rate for profit-making firms based on the lower or main corporate tax rate, Column 7 provides the benefit rate for loss-making firms, and Column 8 provides the average of all three benefit rates. Panel B provides the average enhancement and benefit rates for the low and high tax credit rate periods, the percentage point difference, and the percentage change.

Table D.2: R&D Tax Credit Rates for “Large” Firms Over the Tax Credit Generosity Threshold of 500 Employees Relative to Under

	(1)	(2)	(3)	(4)	(5)
	<u>Difference from SMEs</u>				
Year	Enhancement Rate Rate	Corporate Tax	Tax Credit Benefit	Pct. Point Diff. in Enh. Rates	Pct. Point Diff. in Benefits
2008	0.3	0.28	0.084	-0.45	-0.120
2009	0.3	0.28	0.084	-0.45	-0.120
2010	0.3	0.28	0.084	-0.45	-0.120
2011	0.3	0.26	0.078	-0.70	-0.159
2012	0.3	0.24	0.072	-0.95	-0.194
2013	0.3	0.23	0.069	-0.95	-0.193
2014	0.3	0.21	0.063	-0.95	-0.217
Averages	0.3	0.254	0.076	-0.700	-0.160

Notes: Table provides R&D Tax Relief Enhancement Rates for firms with more than 500 employees (Column 1), the UK’s main corporate tax rate (Column 2), the R&D tax credit benefit based on the policy’s formula using enhancement rates and corporate tax rates (Column 3), and the percentage point differences for these “large” firms and SMEs (just under the threshold) in enhancement rates (Column 4) and tax credit benefit (Column 5).

Table D.3: Descriptive Statistics of Data Used for Larger Firm Analyses

<i>Firms in Sample:</i>	All Firms		Below Threshold		Above Threshold	
<i>Pre/Post Tax Credit Rate Policy</i>	Pre	Post	Pre	Post	Pre	Post
	(1)	(2)	(3)	(4)	(5)	(6)
Mean R&D expenditures (£000s)	1751.3 (3436.2)	2288.7 (4492.8)	1815.2 (3540.8)	2353.0 (4782.6)	1656.4 (3274.0)	2200.4 (4061.8)
Mean R&D direct subsidies (£000s)	22.49 (90.37)	61.98 (322.68)	22.94 (96.75)	69.62 (394.38)	21.82 (79.98)	51.48 (192.64)
Mean prop. of R&D subsidized (%)	0.030 (0.084)	0.046 (0.086)	0.031 (0.084)	0.046 (0.089)	0.029 (0.084)	0.045 (0.082)
Observations	2746	2118	1641	1226	1105	892
No. of Firms	1299	1011	900	702	573	445

Notes: Table provides descriptive statistics of data used to study larger firms for firms with 250 to 750 employees from the UK’s Business Enterprise Research and Development (BERD) survey and Business Structure Database (BSD). The full baseline sample is included in Columns 1-2, firms below the R&D tax credit generosity threshold of 500 employees are included in Columns 3-4, and firms above the threshold are included in Columns 5-6. Sample omits top 1% of the R&D expenditure distribution. Pre-policy period statistics (2000-2008) are provided in the odd-numbered columns and post-policy period statistics (2009-2014) are in even-numbered columns. Standard errors are in parentheses.

Table D.4: Covariate Balance Around Grant Rate Threshold Before Receiving Grants

<i>Dependent Variable:</i>	R&D	Cumulative R&D	Age	log(Total Assets)	log(Current Liabilities)
	(1)	(2)	(3)	(4)	(5)
Panel A: Discontinuities in Pre-Grant Years					
1[Empl. < 50]	85.621 (60.217)	232.945 (172.179)	2.517 (2.588)	0.001 (0.105)	0.201* (0.121)
Observations	1,281	1,281	1,281	1,281	1,279
Mean Dep. Var.	72.610	163.468	21.564	8.390	7.576
Panel B: Differences-in-Discontinuities in Pre-Grant Years					
1[Empl. < 50] * Post 2012	-0.549 (79.756)	-114.409 (407.362)	8.262 (6.645)	-0.226 (0.344)	-0.690 (0.492)
1[Empl. < 50]	86.509 (66.664)	254.292 (191.197)	1.654 (2.891)	0.025 (0.118)	0.278** (0.139)
Observations	1,281	1,281	1,281	1,281	1,279
Mean Dep. Var.	72.610	163.468	21.564	8.390	7.576

Notes: Table provides discontinuities (Panel A) and differences-in-discontinuities (Panel B) in covariates when using observations from only the years before firms receive their first Innovate UK grant. Coefficients are from estimating Equations ?? (Panel A) and ?? (Panel B) with different covariates as dependent variables conditional on first order polynomials of the running variable (included separately for each side of the threshold). The dependent variable is R&D in Column 1, lagged cumulative R&D in Column 2, firm age in Column 3, (log) total assets in Column 4, and (log) current liabilities in Column 5. Sample includes firms within the MSE-optimal window of 29 to 71 employees that also meet the turnover and total assets grant rate generosity criteria. Standard errors are clustered by firm. Asterisks denote * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.5: No Discontinuities in Likelihood of Receiving a Grant

<i>Dep. Var.:</i>	Received (1)	Received (2)	Received (3)	Received (4)	Received (5)	Received (6)
1[Empl. < 50]	-0.020 (0.021)	-0.018 (0.024)	-0.027 (0.026)	-0.021 (0.026)	-0.016 (0.029)	0.048 (0.038)
1[Empl. < 50] * Post 2012				-0.008 (0.043)	-0.019 (0.049)	-0.079 (0.070)
Observations	4,510	3,554	3,268	4,510	3,554	2,954
Mean Dep. Var.	0.149	0.154	0.157	0.149	0.154	0.159
Cond. on All Eligibility Criteria		x	x		x	x
Baseline Controls and FEs			x			x

Notes: Dependent variable is an indicator equal to one if the firm received a grant that year and zero otherwise. Data for all years (2005-2017) included for firms with (lagged) employment of 29 to 71 employees conditional on also meeting the total assets and turnover grant rate generosity criteria. In Columns 1 and 4, the sample is not yet limited to firms that also meet the turnover and total assets grant rate generosity criteria. In Columns 2, 3, 5, and 6, the other eligibility criteria are applied. Running variable controls are included in all regressions and the remaining controls and fixed effects of the baseline specification are included in Columns 3 and 6. Standard errors are clustered at the firm level. Asterisks denote * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.6: Sample of UK Policies Providing Benefits for Smaller Firms

Policy/Program	Description	Firms Affected
Small Business Rate Relief	Relief from property business rates charged on non-domestic properties.	Firms with rateable value less than £15k or business uses only one property.
Corporate Taxes	Single Corporation Tax rate of 20% for non-ring fence profits.	Determined by profits as opposed to turnover, employment, or total assets.
Employment Allowance	Discount on National Insurance bill.	Any business paying employers' Class 1 National Insurance
Venture Capital Schemes: Enterprise Investment Scheme, Seed Enterprise Investment Scheme, and Social Investment Tax Relief	Tax relief provided to investors of venture capital schemes. Relief provided against income tax or capital gains tax.	Tax relief is provided to investors as opposed to firms.
Enterprise Investment Scheme	A venture capital scheme that helps companies raise money.	Firms must not have gross assets worth more than £15m before shares are issued (and £16m afterwards), and must have fewer than 250 employees.
Seed Enterprise Investment Scheme	A venture capital scheme that helps companies raise money.	Firms must not have gross assets worth more than £200k at the time when shares are issued, and must have fewer than 25 employees.
Small Business: GREAT Ambition	Commitment to help small businesses grow.	No firm size definitions that align with the Innovate UK definitions.
British Business Bank	A business development bank committed to making finance markets work better for small businesses.	Support programs for start-ups and small businesses in general with no noticeable advantages to firms that align with the firm size definitions for grant generosity.
Employer NI Contributions	Employers pay secondary national insurance contributions to HMRC.	Rates are determined by profits as opposed to employment, turnover, or total assets.
Value Added Tax	VAT registration is required for firms of a certain size.	The threshold for VAT registration is £85k.
Pay As You Earn	Payment by employers as part of the payroll so that the HMRC can collect income tax and national insurance.	Income tax rates depend on how much of taxable income is above personal allowance, and rates are determined by earnings.
Export Credits Guarantee Scheme	Encourages exports by SMEs by ensuring successful implementation of scheme.	Applies to all SMEs, not just small firms.
Loan Guarantees for SMEs	Government agreement with large banks to extend loans to small businesses in the UK, increasing the availability of finance.	Applies to all SMEs, not just small firms.
Enterprise Capital Funds	Financial schemes to address the provision of equity finance to certain firms and to invest in high growth businesses.	Applies to all SMEs, not just small firms.
Business Angel Co-Investment Fund	A £100M investment fund for UK businesses.	Applies to all SMEs, not just small firms.

Notes: Table provides information on a sample of other policies in the UK that provide incentives for small businesses.

No policies that could confound the diff-in-disc estimates for small firms are found.

Table D.7: No Discontinuities in Likelihood of Reporting Profits or R&D Before Receiving First Grant

<i>Dep. Var. (indicator):</i>	Reported (1)	Reported (2)	Reported (3)	Reported (4)	Reported (5)	Reported (6)
1[Empl. < 50]	-0.020 (0.023)	-0.032 (0.028)	-0.045 (0.029)	0.047 (0.039)	0.036 (0.047)	-0.001 (0.042)
Observations	1,623	1,281	1,089	1,623	1,281	1,089
Mean Dep. Var.	0.957	0.947	0.953	0.093	0.092	0.100
DV = 1 if Reported Profits	x	x	x			
DV = 1 if Reported R&D				x	x	x
Cond. on All Eligibility Criteria		x	x		x	x
Baseline Controls and FEs			x			x

Notes: Dependent variable is an indicator equal to one if there is non-missing data for the (before tax) profits or R&D variables. Missing data could reflect either firms not reporting that information in the P&L statements that they file or BvD not gathering the additional information for that firm. Sample includes only years prior to when firms receive their first grant. In Columns 1 and 4, all firms in the baseline sample with (lagged) employment of 29 to 71 but without conditioning on the grant generosity threshold turnover and total assets criteria also being met. In Columns 2, 3, 5, and 6, the other eligibility criteria are applied. Running variable controls are included in all regressions and the remaining controls and fixed effects of the baseline specification are included in Columns 3 and 6. Standard errors are clustered by firm. Asterisks denote * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.8: Pseudo-Threshold Tests for RDD Grant Funding Rate Effects

<i>Dependent Variable:</i>	R&D (1)	R&D (2)	R&D (3)	R&D (4)	R&D (5)
Panel A: Triangular Weights and Even Bandwidth					
Pseudo-Thresh (5)	-14.94 (27.88)				
Pseudo-Thresh (10)		-60.66 (43.21)			
Pseudo-Thresh (15)			-19.89 (74.77)		
Pseudo-Thresh (85)				-51.69 (134.43)	
Pseudo-Thresh (90)					-145.64 (193.79)
Observations	954	1,666	2,136	597	495
Mean Dep. Var.	30.60	68.47	96.37	235.87	280.99
<i>Sample Empl. Range:</i>	1 to 9	1 to 19	1 to 29	64 to 106	69 to 111
Panel B: Uniform Weights and Baseline Bandwidth					
Pseudo-Thresh (5)	-0.32 (49.74)				
Pseudo-Thresh (10)		-99.32** (45.91)			
Pseudo-Thresh (15)			-43.64 (57.28)		
Pseudo-Thresh (85)				-47.81 (135.73)	
Pseudo-Thresh (90)					-130.78 (192.30)
Observations	2,017	2,213	2,406	597	495
Mean Dep. Var.	79.81	80.54	84.18	227.02	280.85
<i>Sample Empl. Range:</i>	1 to 26	1 to 31	1 to 36	64 to 106	69 to 111

Notes: Dependent variable is R&D expenditures. Results from estimating the RDD model using placebo thresholds at which no discontinuities should exist, and indeed, no discontinuities are detected. In Panel A, triangular weights are used and an even bandwidth for each threshold going up to the baseline of 21 when possible, and in Panel B, uniform weights are used and the baseline bandwidth when possible. All baseline running variable controls, fixed effects, and additional controls included in all regressions. Standard errors are clustered by firm. Asterisks denote * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.9: No Difference-in-Discontinuities in Years Prior to Winning First Grant

<i>Dep. Var.:</i>	R&D (1)	R&D (2)	R&D (3)	R&D (4)
1[Empl. < 50] * Post 2012	-0.55 (79.76)	43.35 (91.76)	25.05 (97.56)	-156.56 (175.57)
1[Empl. < 50]	86.51 (66.66)	24.48 (20.64)	42.26* (25.04)	21.49 (70.55)
Observations	1,281	1,279	1,136	951
Mean Dep. Var.	72.61	72.74	71.89	60.63
Baseline Controls		x	x	x
Firm FEs			x	x
Year x Industry FEs				x

Notes: Dependent variable is R&D expenditures (£000s). Table provides results from estimating the diff-in-disc model of Equation ?? for small firms when including only observations prior to when firms receive their first Innovate UK grant (i.e., when there should be no such discontinuities or differences in discontinuities). No statistically significant discontinuities or differences are detected. Baseline sample inclusion criteria are applied such that firms with 29 to 71 employees are included and only those that also meet the turnover and total assets grant generosity criteria. Running variable controls are included in all regressions and other baseline controls and fixed effects are included in Columns 2-4. Standard errors are clustered by firm. Asterisks denote * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.10: Event Study Point Estimates for Discontinuities in R&D by Year

Year (1)	Coefficient Estimate (2)	Standard Error (3)	P> t (4)
2005	298.33	474.89	0.530
2006	238.73	450.06	0.596
2007	282.29	450.64	0.532
2008	135.41	467.10	0.772
2009	339.62	425.27	0.425
2010	369.25	513.16	0.472
2011	231.25	300.32	0.442
2013	880.52	389.26	0.024
2014	573.44	522.47	0.273
2015	134.31	460.55	0.771
2016	442.17	458.25	0.335
2017	652.49	470.89	0.167

Average Discontinuities in Low vs. High Tax Credit Rate Periods

Low Tax Credit Rate Period (2005-12):	270.70
High Tax Credit Rate Period (2013-17):	536.59

Notes: Table provides point estimates, standard errors, and p-values for the coefficients from an event study version of Equation ?? estimating independent effect of more generous grant funding in R&D expenditures (£000s). The estimates are also plotted in Appendix Figure C.2. Coefficients capture the discontinuity in R&D at the grant generosity threshold each year conditional on the full baseline set of controls, running variables, and fixed effects. The model is estimated as one equation by interacting the grant generosity treatment indicator (and the running variables that differ at the threshold) with indicators for each year, omitting 2012 as the reference year. Standard errors are clustered by firm. Asterisks denote * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.11: Policy Interaction Effects for Small Firms When Widening Bandwidth

<i>Dependent Variable:</i>	R&D (1)	R&D (2)	R&D (3)	R&D (4)	R&D (5)	R&D (6)	R&D (7)
1[Empl. < 50] * Post 2012	654.95*** (224.97)	573.74*** (200.08)	530.95*** (189.32)	544.41*** (190.82)	569.80** (224.90)	529.16** (249.65)	470.42* (259.83)
1[Empl. < 50]	-111.40 (217.01)	-55.70 (203.77)	-80.70 (210.26)	-96.47 (233.04)	-336.04 (293.05)	-113.61 (326.05)	-141.29 (307.51)
Observations	1,460	1,737	2,105	2,428	2,862	2,944	2,970
Mean Dep. Var.	184.54	178.34	175.00	184.87	175.36	191.22	190.51
<i>Sample Empl Range:</i>	20 to 71	15 to 71	10 to 71	10 to 90	10 to 150	10 to 200	10 to 250
Baseline Controls	x	x	x	x	x	x	x
Firm FEs	x	x	x	x	x	x	x
Year x Industry FEs	x	x	x	x	x	x	x

Notes: Table provides diff-in-disc effects of the policy interactions on small firms' R&D expenditures (£000s) when widening the window around the grant rate threshold. Sample includes only firms that also meet the total assets and turnover grant generosity criteria. Standard errors are clustered by firm. Asterisks denote * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.12: No Evidence of Expenditure Relabelling

<i>Dependent Variable:</i>	Tangible Assets (1)	Investment (incl. depr.) (2)	Investment (no depr.) (3)
1[Empl. < 50] * Post 2012	173.14 (445.80)	-109.81 (215.89)	-175.12 (228.83)
1[Empl. < 50]	-511.52 (528.54)	155.26 (181.30)	219.87 (181.99)
Observations	982	958	982
Mean Dep. Var.	1182.26	226.54	19.00

Notes: Table provides estimates from estimating the difference-in-discontinuities in ordinary investment to test whether firms appear to relabel expenditures. Dependent variables are tangible assets (Column 1) and non-R&D ordinary investment (including depreciation in Column 2 and not including depreciation in Column 3). Baseline estimation sample is used and each regression includes all baseline running variables, controls, and fixed effects. Standard errors are clustered at the firm level. Asterisks denote * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.13: Pseudo-Threshold Tests for Diff-in-Disc Policy Interaction Effects

<i>Dependent Variable:</i>	R&D (1)	R&D (2)	R&D (3)	R&D (4)	R&D (5)
Panel A: Triangular Weights and Even Bandwidth					
Pseudo-Thresh (5) * Post 2012	-91.99 (84.11)				
Pseudo-Thresh (10) * Post 2012		-140.07 (179.70)			
Pseudo-Thresh (15) * Post 2012			171.78 (179.57)		
Pseudo-Thresh (85) * Post 2012				-232.70 (369.67)	
Pseudo-Thresh (90) * Post 2012					-174.66 (595.92)
Observations	641	1,259	1,694	562	460
Mean Dep. Var.	38.23	84.13	112.57	233.22	291.48
Empl. Range:	1 to 9	1 to 19	1 to 29	64 to 106	69 to 111
Panel B: Uniform Weights and Baseline Bandwidth					
Pseudo-Thresh (5) * Post 2012	-129.84 (197.94)				
Pseudo-Thresh (10) * Post 2012		18.87 (180.66)			
Pseudo-Thresh (15) * Post 2012			48.59 (277.40)		
Pseudo-Thresh (85) * Post 2012				-272.94 (375.58)	
Pseudo-Thresh (90) * Post 2012					-175.93 (622.53)
Observations	1,573	1,763	1,946	562	460
Mean Dep. Var.	97.69	96.62	98.74	223.61	291.78
Empl. Range:	1 to 26	1 to 31	1 to 36	64 to 106	69 to 111

Notes: Dependent variable is R&D expenditures. Results from estimating the diff-in-disc model using placebo thresholds at which no discontinuities should exist. No differences in discontinuities are detected, as expected. In Panel A, triangular weights are used and an even bandwidth for each threshold going up to the baseline of 21 when possible, and in Panel B, uniform weights are used and the baseline bandwidth when possible. All baseline running variable controls, fixed effects, and additional controls included in all regressions. Standard errors are clustered by firm. Asterisks denote * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.14: Patent Box and R&D Tax Credit Claims

Fiscal Year	<u>R&D Tax Credit Relief</u>		<u>Patent Box Relief</u>			
	SME Total	SME Total	Small	Medium	SME Total	SME Total
	Claims	Relief (£m)	Claims	Claims	Claims	Relief (£m)
	(1)	(2)	(3)	(4)	(5)	(6)
2004/05	5,310	190				
2005/06	4,960	185				
2006/07	5,270	200				
2007/08	5,990	245				
2008/09	6,670	265				
2009/10	7,470	320				
2010/11	8,280	355				
2011/12	10,030	435				
2012/13	13,140	615				
2013/14	15,585	705	170	175	345	14.8
2014/15	29,775	1,315	285	275	560	32.9
2015/16	37,105	1,760	280	285	565	31.2
2016/17	45,440	2,265	265	260	525	32.7
Averages (2013-17)	31,976	1,511	250	249	499	27.9

Notes: Table provides summary statistics of R&D Tax Credit Scheme claims for SMEs (Columns 1-2) and Patent Box claims (Columns 3-6). Data are compiled by author from annual Patent Box and R&D Tax Credit official statistics published on the HM Revenue and Customers (HMRC) website.

Table D.15: Robustness to Omitting Firms in Sectors with Many Patent Box Claims

<i>Dependent Variable:</i>	R&D (1)	R&D (2)	R&D (3)
1[Empl. < 50] * Post 2012	466.88* (242.49)	439.65** (187.13)	708.00** (332.79)
1[Empl. < 50]	-97.30 (68.15)	-81.84 (83.24)	-675.58* (371.17)
Observations	740	721	536
Mean Dep. Var.	197.29	199.41	237.08
Baseline Controls		x	x
Firm FEs			x
Year x Industry FEs			x

Notes: Dependent variable is R&D expenditures (£000s). Firms in sectors with majority of patent box claims (manufacturing and wholesale and trade sectors) are omitted and sample is otherwise the same as the baseline, including firms with 29 to 71 employees in the year before winning a grant conditional on also meeting the turnover and total assets grant rate generosity threshold. Running variable controls included in all regressions. Standard errors are clustered at the firm level. Asterisks denote $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

Table D.16: Robustness of Policy Interaction Effects to Different Winsorization Rules and Addressing R&D Reporting Concerns

<i>Dependent Variable:</i>	R&D (1)	R&D/Empl. (2)	R&D/SIC Total (3)	R&D (4)	R&D (5)
1[Empl. < 50] * Post 2012	507.69** (244.30)	10.57* (5.82)	0.11** (0.05)	627.78** (250.51)	687.27** (302.20)
1[Empl. < 50]	-192.84 (226.04)	-2.71 (6.12)	-0.05 (0.07)	-95.02 (282.20)	-328.79 (321.45)
Observations	1,043	974	1,047	877	717
Mean Dep. Var.	157.57	3.53	0.04	218.60	179.95
Firm FEs	x	x	x	x	x
Year x Industry FEs	x	x	x	x	x
Additional Controls	x	x	x	x	x
Winsorization	5%	1%	1%	1%	1%
Non-Missing Profit Data				x	
Non-Missing Cost of Sales Data					x

Notes: Dependent variable is R&D expenditures (£000s) in Columns 1, 4, and 5. In Column 2, the dependent variable is R&D per employee, and in Column 3, it is R&D as a proportion of the firm's 4-digit SIC total. In Column 1, I winsorize the top 5% of the distribution of non-zero R&D expenditures (for firms with fewer than 100 employees) rather than 1% as done in the baseline. In Columns 4 and 5, I limit the sample to only firms with non-missing data for profits and cost of sales variables, respectively. Sample includes firms with 29 to 71 employees in the year before winning a grant conditional on also meeting the turnover and total assets grant rate generosity eligibility criteria. All baseline running variables, fixed effects, and additional controls included in all regressions. Standard errors are clustered at the firm level. Asterisks denote * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.17: Robustness of Policy Interaction Effects to Different Decisions
Determining Sample Selection

<i>Dependent Variable:</i>	R&D (1)	R&D (2)	R&D (3)	R&D (4)	R&D (5)	R&D (6)	R&D (7)
1[Empl. < 50] * Post 2012	540.12** (260.53)	504.72** (227.61)	561.66** (222.74)	518** (236.28)	768.82*** (268.80)	781.12** (339.80)	923.88** (385.80)
1[Empl. < 50]	-392.07 (353.50)	76.00 (225.58)	-121.74 (259.50)	-123.60 (293.33)	-391.17 (361.58)	-394.33 (426.30)	-840.99* (502.46)
Observations	928	1,325	986	1,182	931	864	804
Mean Dep. Var.	199.18	211.09	195.49	181.34	210.29	220.10	226.38
<i>Years in Sample:</i>	2005-17	2005-17	2005-17	2005-17	2008-17	2009-17	2010-17
Firm FEs	x	x	x	x	x	x	x
Year x Industry FEs	x	x	x	x	x	x	x
Additional Controls	x	x	x	x	x	x	x

Notes: Dependent variable is R&D expenditures (£000s). In Column 1, the bandwidth is reduced to 19 such that the sample includes only firms with 31 to 69 employees in the year prior to receiving a grant (but still conditional on meeting the other eligibility criteria). In Column 2, I use the baseline MSE-optimal bandwidth but do not condition on firms also meeting the other grant rate generosity criteria. In Column 3, I omit observations associated with grants received after 2015 so there are at least 2 years of post-grant data. In Column 4, I include up to four years of data post-grant rather than just three. In Columns 5-7, I omit years leading up to the Great Recession. The estimation sample is otherwise the same as the baseline, including firms with 29 to 71 employees in the year before winning a grant conditional on also meeting the turnover and total assets grant rate generosity eligibility criteria. Running variable controls included in all regressions. Standard errors are clustered at the firm level. Asterisks denote * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.18: Additional Robustness Checks for Policy Interaction Effects on Small Firms' R&D Expenditures

<i>Dependent Variable:</i>	R&D Cluster by Industry (1)	R&D Uniform Weights (2)	R&D Quadratic Polynomials (3)	R&D Cubic Polynomials (4)
1[Empl. < 50] * Post 2012	560.28** (229.63)	489.98** (237.56)	711.88*** (233.45)	746.24*** (273.21)
1[Empl. < 50]	-122.49 (312.74)	94.59 (231.42)	-178.79 (385.23)	213.40 (638.88)
Observations	1,047	1,325	1,047	1,047
Mean Dep. Var.	186.70	212.00	186.70	186.70
<i>Polynomial Flexibility:</i>				
Linear (baseline)	x	x		
Quadratic			x	
Cubic				x
Firm FEs	x	x	x	x
Year x Industry FEs	x	x	x	x
Additional Controls	x	x	x	
<i>Clustering:</i>	Industry	Firm	Firm	Firm
<i>Kernel Weight:</i>	Tri.	Uni.	Tri.	Tri.

Notes: Dependent variable is R&D expenditures (£000s). Sample includes firms with 29 to 71 employees in the year prior to winning a grant conditional on also meeting the turnover or total assets eligibility criteria for more generous grant funding rates. Column 1 clusters standard errors at the 4-digit SIC level rather than firm level. In Column 2, I remove the requirement for the sample to only include firms that also meet the turnover and total assets grant rate generosity threshold. This sample selection criteria apply again though in Columns 3-5. In Column 3, I use uniform weights rather than triangular, and in Columns 4 and 5, I increase the polynomial flexibility of the running variable controls. Standard errors are clustered at the firm level in Columns 2-5. Asterisks denote * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.19: Details on Financial Constraint Proxy Variables

Proxy Variable	Description	Median in Baseline Sample (£000s)
Operating Profit	Gross profit minus operating expenses before accounting for net interest paid/received and taxes Constrained firms are more likely to have lower operating profits. Variable is constructed by author using FAME data.	£69
Available Funds	After-tax profits plus depreciation Constrained firms are more likely to have fewer available funds. Variable is taken directly from FAME database.	£174
Short-term Loans	Current liabilities such as group loans, bank overdrafts, short-term hire purchasing and leasing, etc. Constrained firms are more likely to have more short-term debt. Variable is taken directly from FAME database.	-£411

Notes: Table provides definitions of financial constraint proxies used when estimating the heterogeneous policy effects for constrained vs. unconstrained firms in Section ???. The proxies aim to capture the resources firms may have for self-financing R&D. Median values in the year prior to when firms in the baseline estimation sample receive a grant are also included, which I use when splitting the sample into constrained versus unconstrained firms. Results from the heterogeneous analyses are in Table ??.

Table D.20: Pre-Policy Covariate Balance for Larger Firms

<i>Dep. Var.:</i>	Grant Funds (£000s) (1)	Grant Funds per R&D Exp. (2)	Revenue (£m) (3)	Revenue per Empl. (4)	Age (years) (5)	Average Wages (6)	Number of Scientists (7)
1[Empl.<500]	2.67 (5.64)	0.01 (0.01)	-9.79 (9.83)	-31.57 (28.26)	-0.52 (0.86)	0.16 (0.33)	-0.49 (2.29)
Observations	2,746	2,746	2,746	2,746	2,746	2,746	2,746
No. of Firms	1299	1299	1299	1299	1299	1299	1299

Notes: Table provides evidence of balanced covariates around the tax credit policy threshold in the pre-policy period (2000-2007), suggesting that firms around the threshold are similar. The main regressor is a dummy variable equal to one if the firm has fewer than 500 employees within a regression discontinuity design. Firms with 250 to 750 employees are included in the sample. In all cases, triangular weights are used and first-order polynomials of the running variable (employment) are included separately for each side of the threshold. In Column 1, the dependent variable is direct subsidies for R&D (000s GBP). In Column 2, it is the proportion of R&D expenditures subsidized by government subsidies. In Column 3, it is revenue (millions GBP). In Column 4, it is labor productivity (turnover in millions over total number of employees). In Column 5, it is firm age. In Column 6, it is average wages (R&D worker salaries in total over total number of employees). In Column 7, it is the number of R&D scientists. The top 1% of pre-policy R&D expenditure distribution is dropped to account for outliers in all cases. Standard errors are clustered at the firm level. Asterisks denote $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

Table D.21: Independent Effect of Tax Credit Policy on R&D Expenditures for Larger Firms

<i>Dependent Variable:</i>	R&D (1)	R&D (2)	R&D (3)	R&D (4)
<i>Estimation Approach:</i>	Regression Discontinuity		Diff-in-Disc	
	Post-Policy	Pre-Policy		
1[Empl.<500]	1,000.01** (446.19)	191.07 (154.52)		
1[Empl.<500] * Post 2008			539.48** (238.93)	588.84** (241.67)
Observations	1,382	2,746	4,240	20,618
No. of Firms	584	1,299	1,061	5,493
Dep. Var. Mean	2,811	1,766	2,203	1,221
Sample Empl. Range	250-750	250-750	250-750	≤1000
Firm FEs			x	x
Year FEs			x	x

Notes: Table provides estimates for the effect of being eligible for more generous tax credit rates on larger firms' R&D expenditures (£000s). In Columns 1-2, I use an RDD to estimate the discontinuity in R&D in post-policy years (2009-2014) and pre-policy years (2000-2008). In Columns 3-4, I implement a difference-in-discontinuities design using data from all years. Sample includes firms with 250 to 750 employees in Column 3 and firms with 1,000 employees or less in Column 4. Samples in all regressions omit the top 1% of the R&D distribution to account for outliers. Triangular weights are used in Columns 1-3 and uniform weights are used in Column 4. All regressions include first-order polynomials of the running variable (employment) separately for each side of the threshold. The firm's average pre-policy R&D expenditures are included as a control in Columns 1, 3, and 4. Standard errors are clustered at the firm level. Asterisks denote * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.22: Robustness of Tax Credit Generosity Effects to Different Windows Around Threshold for Larger Firms

<i>Sample Empl. Range:</i>	300-700	275-725	225-775	200-800
<i>Dependent Variable:</i>	R&D	R&D	R&D	R&D
	(1)	(2)	(3)	(4)
Panel A: Discontinuities in R&D in Post-Policy Period (2009-2014)				
1[Empl.<500]	759.80* (452.19)	924.11** (441.02)	1,021.35** (441.99)	999.13** (434.71)
Observations	1,100	1,253	1,528	1,702
No. of Firms	466	538	635	704
Dep. Var. Mean	2920	2862	2828	2799
Panel B: Discontinuities in R&D in Pre-Policy Period (2000-2008)				
1[Empl.<500]	230.80 (163.25)	221.26 (159.94)	136.15 (149.83)	133.46 (145.15)
Observations	2,141	2,446	3,063	3,425
No. of Firms	1040	1155	1428	1603
Dep. Var. Mean	1809	1791	1754	1738

Notes: Table provides results from estimating the effects of being eligible for more generous tax credits on larger firms' R&D expenditures (£000s) using the RDD model but alternative firm size bandwidths around the 500-employee threshold. The windows of firm sizes are provided in the header. Dependent variable is R&D expenditures (£000s). Panel A provides the discontinuities in the post-policy period (2009-2014) and Panel B provides the discontinuities in the pre-policy period (2000-2008). In all regressions, triangular weights are used and first-order polynomials of the running variable (employment) are included separately for each side of the threshold. All regressions in Panel A also include a control for the firm's average pre-policy R&D expenditures and I winsorize by dropping the top 1% of the R&D expenditure distribution. Standard errors are clustered at the firm level. Asterisks denote $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

Table D.23: Replication of Tax Credit Generosity Effects for Larger Firms Using Alternative Data—Regression Discontinuity Design Estimates

<i>Sample Empl. Range:</i>	≤ 1000	100-900	200-800
<i>Dependent Variable:</i>	R&D	R&D	R&D
	(1)	(2)	(3)
Panel A: Discontinuities in R&D in Post-Policy Period (2009-2014)			
1[Empl.<500]	550.52** (234.85)	473.68* (243.22)	305.93 (273.19)
Observations	8,549	3,867	1,991
No. of Firms	1,849	878	494
Mean non-zero R&D	1,614	2,177	2,700
Panel B: Discontinuities in R&D in Pre-Policy Period (2000-2008)			
1[Empl.<500]	29.05* (16.34)	26.53 (19.28)	21.18 (19.52)
Observations	9,603	4,250	2,245
No. of Firms	1,695	790	490
Mean non-zero R&D	2,320	3,853	3,204

Notes: Table provides results from replicating the larger firm tax credit effect analysis when using the regression discontinuity design approach. Estimates reflect the effect of being eligible for more generous tax credit rates on larger firms' R&D expenditures (£000s) when using Bureau van Dijk data (rather than BERD). Panel A provides discontinuity estimates in post-policy years (2009-2014) and Panel B presents estimates for pre-policy years (2000-2008). Samples in all regressions omit the top 1% of the R&D distribution to account for outliers. The row with mean R&D is the mean of positive R&D values for the given estimation sample. Triangular weights are used in all cases and all regressions include first-order polynomials of the running variable (employment) separately for each side of the threshold. The firm's average pre-policy R&D expenditures as well as cumulative R&D interacted with firm age are included as controls (similar to the controls used for the small firm analysis using the BvD data). Standard errors are clustered at the firm level. Asterisks denote * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.24: Replication of Tax Credit Generosity Effects for Larger Firms Using Alternative Data—Diff-in-Disc Estimates

<i>Sample Empl. Range:</i>	≤ 1000	≤ 1000	100-900	100-900
<i>Dependent Variable:</i>	R&D	R&D	R&D	R&D
	(1)	(2)	(3)	(4)
1[Empl.<500] * Post 2008	469.21** (216.02)	659.07*** (218.04)	525.86** (251.87)	1261.55** (604.14)
Observations	17,946	17,946	8,015	3,132
No. of Firms	1,919	1,919	935	344
Mean non-zero R&D	1,659	1,659	2,255	2,255
Firm FEs	x	x	x	x
Year FEs	x	x	x	x
R&D Intensive Only				x
Kernel Weighting	Triangular	Uniform	Uniform	Uniform

Notes: Table provides results from replicating the larger firm tax credit effect analysis when using a difference-in-discontinuity approach. Estimates reflect the effect of being eligible for more generous tax credit rates on larger firms' R&D expenditures (£000s) in the post-policy period relative to the pre-policy period when using Bureau van Dijk data (rather than BERD). All years of data that overlap with the BERD data analysis are included here (2000-2014). Samples in all regressions omit the top 1% of the R&D distribution to account for outliers. Firms with 1,000 employees or less are included in Columns 1-2 and firms with 100 to 900 employees are included in Columns 3-4. In Column 4, I include only firms that have some positive reported value for R&D at some point through the sample period. The row with mean R&D is the mean of positive R&D values for the given estimation sample. Triangular weights are used in Column 1, and to mimic a diff-in-diff approach, uniform weights are used in Columns 2-4. All regressions include first-order polynomials of the running variable (employment) separately for each side of the threshold. The firm's average pre-policy R&D expenditures as well as cumulative R&D interacted with firm age are included as controls (similar to the controls used for the small firm analysis using the BvD data). Standard errors are clustered at the firm level. Asterisks denote * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.25: Falsification Tests for Policy Interaction Effects for Larger Firms

<i>Dependent Variable:</i>	R&D	R&D	R&D	R&D	Non-Cap R&D	Cap R&D
<i>Threshold Type::</i>	Pseudo	Pseudo	Pseudo	Pseudo	Real	Real
	(1)	(2)	(3)	(4)	(5)	(6)
Direct Subsidies * Threshold	0.46 (0.6)	0.33 (1.95)	1.94 (2.71)	4.02 (4.21)	-2.64** (1.34)	-0.19 (0.12)
Direct Subsidies	2.42*** (0.23)	2.63*** (0.34)	6.19*** (1.56)	6.46** (2.96)	3.62*** (1.18)	0.22** (0.11)
Observations	7,869	3,405	1,646	1,035	1,382	1,382
No. of Firms	5086	1768	719	450	584	584
Dep. Var. Mean	647.6	1998	2531	2897	2632	179.5
<i>Threshold (Empl.):</i>	100	350	650	900	500	500
<i>Sample Empl. Range:</i>	0-200	100-600	400-900	650-1150	250-750	250-750

Notes: Table provides results from falsification tests of larger firm policy interaction effects when estimating discontinuity-in-effects models. In Columns 1-4, the dependent variable is R&D expenditures (£000s) and I conduct placebo tests imposing fake pseudo-thresholds, finding no statistically significant discontinuities. In Columns 5 and 6, I return to using the actual tax credit generosity threshold and estimate the effects specifically on non-capital R&D (Column 5), where the substitution is most likely to occur, and capital R&D (Column 6), where there should be less or no substitution since these expenditures typically do not qualify for tax credits. In all regressions, triangular weights are used and first-order polynomials of the running variable (employment) are included separately for each side of the threshold. All regressions also include a control for the firm's average pre-policy R&D expenditures and I winsorize by dropping the top 1% of the R&D expenditure distribution in the post-policy period. Standard errors are clustered at the firm level. Asterisks denote * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.26: Using Lagged Employment to Define Tax Policy Treatment Status

<i>Dep. Var.:</i>	R&D	R&D
<i>Employment year(s) used to define tax credit treatment:</i>	Current + One Year Lag (1)	Current + Two Year Lags (2)
Direct Subsidies * 1[Empl.<500]	-2.28* (1.30)	-2.25* (1.29)
Direct Subsidies	3.38*** (1.11)	3.38*** (1.11)
Observations	1,382	1,380
No. of Firms	584	583

Notes: Dependent variable is R&D expenditures (£000s). Table provides results from placing more stringent requirements on how tax credit eligibility is defined for firms around the 500-employee threshold by using one year of lagged employment in addition to the current year in Column 1 and two years of lagged employment plus the current year in Column 2. In both regressions, firms with 250 to 750 employees are included, triangular weights are used, and first-order polynomials of the running variable (employment) separately for each side of the threshold are included as well as a control for the firm's average pre-policy R&D expenditures. I drop the top 1% of the R&D expenditure distribution to account for outliers. Standard errors are clustered at the firm level. Asterisks denote * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.27: Additional Robustness Checks of Policy Interaction Effects on Larger Firms' R&D Expenditures

<i>Dependent Variable:</i>	R&D (1)	R&D (2)	R&D (3)	R&D (4)
Panel A: Disc-in-Effects in Post-Policy Period (2009-2014)				
Direct Subsidies * 1[Empl.<500]	-2.86** (1.41)	-2.87** (1.42)	-3.36** (1.41)	-3.64*** (1.40)
Direct Subsidies	3.85*** (1.24)	3.86*** (1.25)	4.49*** (1.26)	5.04*** (1.22)
1[Empl.<500]	653.91 (509.54)	562.40 (544.69)	1,244.51*** (465.30)	643.17*** (211.11)
Observations	1,382	1,382	1,382	1,078
Dep. Var. Mean	2,811	2,811	2,766	1,453
Panel B: Disc-in-Effects in Pre-Policy Period (2000-2008)				
Direct Subsidies * 1[Empl.<500]	-0.42 (2.31)	-0.38 (2.31)	-0.22 (2.45)	1.69 (2.38)
Direct Subsidies	3.23** (1.29)	3.27** (1.30)	3.13*** (1.20)	5.83*** (1.56)
1[Empl.<500]	155.95 (475.24)	814.57 (626.37)	335.31 (315.36)	-12.29 (143.45)
Observations	2,746	2,746	2,746	2,269
Dep. Var. Mean	1766	1766	1751	970.9
<i>Polynomial Flexibility:</i>				
Linear			x	x
Quadratic	x			
Cubic		x		
Only if Subsidies>0				x
Kernel Weighting:	Triangular	Triangular	Uniform	Triangular

Notes: Table provides robustness checks of the policy interaction effects for larger firms when implementing the discontinuity-in-effects approach. Dependent variable is R&D expenditures (£000s). Panel A provides estimates for the post-policy period (2009-2014) when discontinuities are expected and Panel B provides estimates for the pre-policy period (2000-2008) where a positive correlation between direct subsidies and R&D is expected but no discontinuity. In Columns 1 and 2, I increase the flexibility of the running variable controls. In Column 3, I estimate the baseline model using first-order polynomials but apply uniform kernel weighting rather than triangular. In Column 4, I limit the sample to include only observations for which there is a positive value of direct subsidies. In all regressions, firms with 250 to 750 employees are included and the top 1% of the R&D expenditure distribution is omitted to account for outliers. Standard errors are clustered at the firm level. Asterisks denote * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.28: No Evidence that Effects are Driven by Changes in Type of Projects Funded, Competition, or Cumulative Funding

<i>Dependent Variable:</i>	Feasibility (1)	Concept (2)	Prototype (3)	Market (4)	R&D (5)	R&D (6)
1[Empl. < 50] * Post 2012	-0.13 (0.14)	-0.02 (0.06)	-0.13 (0.09)	0.02 (0.06)	828.66* (432.48)	964.38*** (280.29)
1[Empl. < 50]	-0.05 (0.12)	0.01 (0.03)	0.13 (0.09)	0.01 (0.04)	1371.07** (611.99)	-138.45 (282.21)
1[Empl. <50] * Post 2012 * Cumulative Grants						-129.55* (67.73)
1[Empl. <50] * Cumulative Grants						61.93 (71.04)
Cumulative Grants						-53.37 (52.96)
Observations	1,047	1,047	1,047	1,047	963	1,047
Mean Dep. Var.	0.13	0.03	0.09	0.02	171.67	186.70
Firm FEs	x	x	x	x	x	x
Year x Industry FEs	x	x	x	x	x	x
Baseline Controls	x	x	x	x	x	x
Grant Competition FEs					x	

Notes: Table provides results from investigating alternative explanations of positive interaction effects and are associated with discussion in Section ?? . In Columns 1-4, the dependent variables are indicators for whether projects are feasibility studies (Column 1), proofs of concept (Column 2), development of prototypes (Column 3), and proofs of concept (Column 4). In Column 5, I estimate the baseline model with grant competition-level fixed effects, and in Column 6, the main treatment variables are interacted with the firm's cumulative number of grants. Sample includes firms with 29 to 71 employees in the year prior to winning a grant conditional on also meeting the turnover and total assets grant rate generosity criteria. All baseline running variables, controls, and fixed effects are included in all regressions. Standard errors are clustered at the firm level. Asterisks denote * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Data References

Financial Analysis Made Easy (FAME), Bureau van Dijk. Available at:
<https://www.bvdinfo.com/en-gb/our-products/data/national/fame>

Department for Business, Energy & Industrial Strategy, Innovate UK, UK Research and Innovation. (2017). Transparency Data: Innovate UK funded projects since 2004. Available at: <https://www.gov.uk/government/publications/innovate-uk-funded-projects>

Office for National Statistics. (2017). Business Expenditure on Research and Development, 1994-2015: Secure Access. [data collection]. 6th Edition. UK Data Service. SN: 6690,
<http://doi.org/10.5255/UKDA-SN-6690-6>

Office for National Statistics. (2017). Business Structure Database, 1997-2017: Secure Access. [data collection]. 9th Edition. UK Data Service. SN: 6697,
<http://doi.org/10.5255/UKDA-SN-6697-9>

References

- Alstadsaeter, A., S. Barrios, G. Nicodeme, A.M. Skonieczna, and A. Vezani**, “Patent Boxes Design, Patents Location, and Local R&D,” *Economic Policy*, 2018, *33* (93), 131–77.
- Gaessler, F., B. Hall, and D. Harhoff**, “Should there be lower taxes on patent income?,” *Research Policy*, 2021, *50* (1).
- Griffith, R., H. Miller, and M. O’Connell**, “The UK will introduce a Patent Box, but to whose benefit? [Comment] IFS,” 2010.
- Hall, Bronwyn**, “Tax Policy for Innovation,” in Austan Goolsbee and Benjamin Jones, eds., *Innovation and Public Policy*, University of Chicago Press, 2022.
- Mohnen, P., A. Vankan, and B. Verspagen**, “Evaluating the Innovation Box Tax Policy Instrument in the Netherlands,” *Oxford Review of Economic Policy*, 2017, *33* (1), 141–56.
- ONS**, “Comparison of ONS business enterprise research and development statistics with HMRC research and development tax credit statistics,” 2022. Available at: <https://www.ons.gov.uk/economy/governmentpublicsectorandtaxes/researchanddevelopmentexpenditure/articles/comparisonofonsbusinessenterpriseresearchanddevelopmentstatisticswithhmrcresearchanddevelopmenttaxcreditstatistics/2022-09-291>, Released 29 September 2022, Accessed August 2023.
- , “Quality and Methodology Information Report, Business Enterprise Research and Development Survey QMI,” 2022. Available at: <https://www.ons.gov.uk/economy/governmentpublicsectorandtaxes/researchanddevelopmentexpenditure/methodologies/ukbusinessenterpriseresearchanddevelopmentsurveyqmi>, Accessed August 2023, ONS website last updated 17 November 2022.
- S., Sorenson B.-Villegas-Sanchez C.-Volosovych V. Kalemli-Ozcan and S. Yesiltas**, “How to Construct National Representative Firm Level Data from the ORBIS Global Dataset,” *Working paper*, 2015.