BANK OF ENGLAND

Staff Working Paper No. 924

Impacts of the Covid-19 crisis: evidence from 2 million UK SMEs

James Hurley, Sudipto Karmakar, Elena Markoska, Eryk Walczak and Danny Walker

June 2021

BANK OF ENGLAND

Staff Working Paper No. 924

Impacts of the Covid-19 crisis: evidence from 2 million UK SMEs

James Hurley,(1) Sudipto Karmakar,(2) Elena Markoska,(3) Eryk Walczak(4) and Danny Walker(5)

Abstract

We introduce a novel data set to analyze the impact of the Covid-19 crisis on SME cash flows. The crisis led to a sharp drop in economic activity in the UK, which hit SMEs harder than larger businesses. The data set comprises monthly information on all 2 million SMEs that have current accounts or debt with nine major banking groups, with roughly 5 billion data points in total. We document a few basic facts on UK SME cash flows during Covid-19. (1) The virus and the public health interventions coincided with a 30 percentage point reduction in turnover growth for the average SME. (2) There was significant heterogeneity in the turnover shock across SMEs, with the biggest reductions for younger SMEs in consumer-facing sectors in Scotland and London. (3) Cash flows were broadly flat on average and there was much less heterogeneity across SMEs. (4) SMEs with average turnover growth in 2020 were most likely to use the main government-guaranteed loan scheme for SMEs, as well as those in the hospitality sector in more affluent areas of the country. Our analysis provides a framework to monitor SMEs as the sector recovers from the pandemic.

Key words: Covid-19 pandemic, small and medium-sized enterprises, SMEs, government support schemes.

JEL classification: D22, E65, G30.

(1) Bank of England. Email: james.hurley@bankofengland.co.uk
(2) Bank of England. Email: sudipto.karmakar@bankofengland.co.uk
(3) Bank of England. Email: elena.markoska@bankofengland.co.uk
(4) Bank of England. Email: eryk.walczak@bankofengland.co.uk
(5) Bank of England. Email: daniel.walker@bankofengland.co.uk

The Bank's working paper series can be found at www.bankofengland.co.uk/working-paper/staff-working-papers

Bank of England, Threadneedle Street, London, EC2R 8AH
Email enquiries@bankofengland.co.uk

ISSN 1749-9135 (on-line)

1 Introduction

1.1 Context

The Covid-19 pandemic and the public health measures introduced to contain it have had a huge impact on the UK economy. According to the latest data, GDP was 10% lower in 2020 than in 2019. This could be the largest annual fall in around 300 years[1]. Lower demand for goods and services and disruption to production and supply chains have led the revenues of many businesses to decline sharply and persistently. This decline in revenues put pressure on cash flows and increased liquidity needs. Small and medium enterprises (SMEs)[2] were likely hit hardest, as set out in the Bank of England's August 2020 Financial Stability Report.

Beyond national accounts aggregates and relatively small-scale surveys, data on the performance of UK businesses through the Covid-19 crisis has been sparse. Only the largest listed companies publish timely accounting information and, unlike in the US, most of those only publish them twice a year. The larger private businesses have to file information on their profit and loss accounts at Companies House, which makes them available online, but this comes with lags of up to two years from when the activity took place. In this paper, we introduce a novel and data set that sheds light on the performance of 2 million UK SMEs. It contains detailed monthly information on all SME accounts held with nine major banking groups in the UK, covering current accounts[3] and bank debt of various forms. The data is provided confidentially to the Bank of England via Experian, a private sector information services company, with around a 6 week lag. This allows us to analyze the effects of the Covid-19 crisis on a very large sample of UK SMEs in detail for the first time. The latest data point that we analyze in this paper is for December 2020, which means our analysis does not cover the impacts since then, although future work could analyze it in detail.

There are around 3 million SMEs in the data set, 2 million of which are limited companies and 1 million of which take other legal forms. This compares to the 6 million businesses in

[1] See the GDP estimates produced by the UK's national statistics authority, compared to historical data compiled by the Bank of England.

[2] There is no commonly-accepted definition of SMEs. For the purposes of this paper we define them as businesses that have less than £25 million in turnover (another word for revenues). Other definitions reference numbers of employees.

[3] In the US current accounts are often referred to as 'checking accounts'. The data does not cover longer-term savings accounts.

total in the UK in 2020 according to aggregate statistics compiled by the UK government's business department, around 5.9 million of which were SMEs and 2 million of which were limited companies. We focus our analysis on SME limited companies, given this is where we have near-universal coverage of the population and where we can match the data set to other public data sources. This means for each of the 2 million SMEs we have recent information on their sector, address, age and some simple balance sheet characteristics, as well as current account flows and new loans on a monthly basis. At the start of 2020, UK SMEs employed around 17 million people on aggregate (61% of the private sector total) and generated turnover of £2.3 trillion (52% of the total). Around a third of UK banking sector exposures to UK businesses are to SMEs. A healthy SME sector is important for the sustained growth of the UK economy and the resilience of the financial sector. However, the prolonged nature of the pandemic, the severity of the shock and the uncertain outlook are likely to continue to threaten many SMEs in the months to come.

The UK government has implemented a range of fiscal measures to support SME liquidity through the Covid-19 crisis[4], many of which were still in place at the time of writing. The biggest in terms of its fiscal cost was the Coronavirus Job Retention Scheme (CJRS), which allowed businesses of any size to furlough employees and receive compensation from the government to help to cover their labor costs. There was also a range of direct cash grants, typically administered via local government authorities. In addition to these grants, the government introduced a number of loan guarantee schemes to ease the supply of credit, particularly to the smallest businesses. This included the Bounce Back Loan Scheme (BBLS), which came with a 100% credit risk guarantee and has so far seen around 1.5 million loans of up to £50,000 go to very small businesses. Around £70 billion of the £80 billion of new finance raised by UK businesses in 2020 came through the government loan schemes. [5]

This paper documents the following facts on UK SMEs:

1. The average UK SME saw a 30 percentage point fall in turnover growth between April 2020 and December 2020 relative to the period before the Covid-19 crisis. Turnover growth for the average SME was still well below its January 2020 level in December 2020.

[4] See the Office for Budget Responsibility's (latest report) for more details on these measures.

[5] See the Bank of England blog for more details on the loan schemes.

2. There was considerable heterogeneity in the turnover shock across UK SMEs. Younger SMEs, those that operate in the *Arts and Recreation* and *Accommodation and Food* sectors, in Scotland and London, saw larger turnover growth reductions than the average.

3. SMEs appear to have been able to reduce costs by as much as turnover on average, potentially supported by the government interventions. This means there was much less heterogeneity in the cash flow impact, which was relatively small for the average SME.

4. Usage of the BBLS was also quite heterogeneous. Firms with the highest and lowest turnover growth in 2020 were least likely to use the BBLS. Firms in *Accommodation and Food* were most likely to use BBLS loans. In terms of regional variation, firms in the north of England were most likely to have taken a BBLS loan and those in Northern Ireland were least likely. Firms in more affluent areas were more likely to use the BBLS.

1.2 Contribution to the literature

Our paper contributes to a rapidly emerging strand of literature on the impact of the Covid-19 pandemic on businesses. We are able to overcome two key challenges faced by prior studies: poor coverage of very small businesses; and small sample sizes. Our data comprises 2 million SMEs, which is almost the entire universe of limited companies in the UK. As well as studying average impacts in simple regressions, we document impacts along various dimensions of heterogeneity, including sector, region, firm age and firm size. Given the granularity of our data, we are also able to track the usage of government-guaranteed loan schemes across firms, showing the types of businesses that the schemes helped most. To the best of our knowledge, this is the first paper to provide such a detailed and representative view of the effect of the crisis on SMEs. Methodologically, we go beyond linear regressions and provide the reader with a glimpse of how more advanced machine learning algorithms could be used to analyze big data on businesses in the future.

1.3 Related literature

A large number of papers study the difficulties that small businesses have been facing through the Covid-19 pandemic and public health interventions. Bloom et al. [2021] use survey data on an opt-in panel of around 2,500 US small businesses to assess the impact of Covid-19. They document a significant negative sales impact that peaked in Quarter 2 of 2020, with an average loss of 29% in sales. The authors find that the smallest offline firms experienced sales drops of over 40% compared to less than 10% for the largest online firms. Similar results have been documented in Buffington et al. [2020], Bartik et al. [2020] and Chetty et al. [2020]. Chodorow-Reich et al. [2020], use loan-level data to document that SMEs obtain credit lines of shorter maturity, manage maturity less actively, post more collateral, pay higher interest rates and have higher utilization rates. The authors show that SMEs do not draw down in contrast to large firms despite SME demand, but that government support scheme loans helped alleviate the shortfall. Acharya and Steffen [2020] document a corporate "dash for cash" induced by the pandemic. The authors show that in the first phase of the pandemic, all firms drew down on bank credit lines to raise cash balances. This reversed in the second phase (following the adoption of stabilization policies) when only the highest rated firms switched to capital markets to raise cash. The adverse effect of the pandemic on corporate cash buffers is also documented in Banerjee et al. [2020] where they find that about 40% of a sample of firms from advanced economies would not be able to service their debt in 2020 in a scenario where 2020 revenues decline by 25% relative to 2019. Autor et al. [2020] study the efficacy of the Paycheck Protection Program (PPP) which was intended to help SMEs maintain employment and wages during the pandemic in the US. The authors estimate that the PPP boosted employment at eligible firms by 2% to 4.5% and imply that the PPP increased aggregate U.S. employment by 1.4 million to 3.2 million jobs through the first week of June 2020. Although the evidence is supportive of a causal effect of the PPP on aggregate employment, the authors are mindful of puzzles and view this work as preliminary.

Papanikolaou and Schmidt [2020] document that sectors in the US in which a higher fraction of the workforce is not able to work remotely experienced significantly greater declines in employment, significantly more reductions in expected revenue growth, worse stock market performance, and higher expected likelihood of default. Schivardi and Romano [2020] present a simple method to determine the number of firms that could become illiquid, and when, in Italy. The authors find that at the peak, around 200,000 companies (employing 3.3 million workers) could become illiquid due to a total liquidity shortfall of EUR72 billion. Carletti et al. [2020] go beyond the universe of small firms and forecast the drop in profits and the equity shortfall triggered by the Covid-19 lockdown, using a representative sample of around eighty thousand Italian firms. The authors predict that a three-month lockdown would entail an aggregate yearly drop in profits of about 10% of GDP and amount to financial distress for 17% of the firms included in the sample. Distress is forecasted to be more frequent for SMEs, for firms with high pre-Covid-19 leverage, and those belonging to the Manufacturing and Wholesale Trading sectors. At a cross-country level, Gourinchas et al. [2020] estimate the impact of the pandemic on SME failures, in seventeen countries. The authors estimate an increase in failure by nearly 9 percentage points, absent government support. Accommodation & Food Services, Arts, Entertainment & Recreation, Education, and Other Services are among the most affected sectors.

This paper also contributes to the growing literature that seeks to shed light on practical issues using machine learning techniques. In recent years, machine learning approaches have become more and more prominent as means to answer a range of research questions in economics. Athey and Imbens [2019] discusses the importance of utilizing machine learning models for economics problems. The authors highlight that in economics, methods with formal properties are typically favored to machine learning models. However, the focus on machine learning approaches allows for different types of formal results, such as error rates, model performance, predictive power, guarantees, etc. In this paper we present some initial results using machine learning methods applied to SME data.

The rest of the paper is organized as follows. Section 2 describes the data set and sets out some simple stylized facts. Section 3 shows how the Covid-19 crisis impacted UK SMEs over the course of 2020 and documents some of the considerable heterogeneity of those impacts across different types of SMEs operating in different parts of the UK. Section 4 presents analysis of which SMEs raised liquidity through the BBLS program. Section 5 includes some novel machine learning models that could be used in future for prediction. Section 6 sets out some conclusions and ideas for further work.

2 Data and stylized facts

This section describes the data set and presents some simple stylized facts, focusing on how UK SME turnover and cash flow evolved through 2020.

2.1 Data description

This paper introduces a novel data set containing information on the bank accounts of a very large number of UK SMEs. The Bank of England receives the data on a monthly basis via Experian, a private sector information services company. This arrangement is supported by the Small Business Enterprise & Employment Act, which was enacted in 2015 in the UK. It introduced the Credit Information Regulations, which require nine major banking groups[6] to send regular data on all of their SME customers to a list of designated credit reference agencies. The credit reference agencies have to share the data they collect with all finance providers, to promote competition in the SME lending market. The credit reference agencies also have to share the data with the Bank of England. The data is confidential and this paper does not present any information that identifies any individual SMEs or banks.

The legislation provides a specific definition of SMEs. It states that a UK business qualifies as an SME if it satisfies the following criteria:

- it has an address in the United Kingdom;
- it carries out commercial activities as its principal activity;
- it is not part of a group which as a whole has an annual turnover which is equal to or greater than £25 million.

On the face of it this definition is relatively precise and differs from some other sources, which define SMEs on the basis of their number of employees.[7] However, the data is collected from banks, which manage their SME portfolios and business lines in different ways. This means that in some cases the turnover cut-off they use in practice can be above or below

[6] Several other lenders share their data with the credit reference agencies on an optional basis.

[7] For example, the European Union defines SMEs as businesses with fewer than 250 employees, whilst the US authorities typically refer to businesses with fewer than 500 employees.

£25 million and the data therefore includes some businesses with higher turnover than this threshold. We retain these businesses in the sample we use for our analysis.[8]

There are around 3 million SMEs in the data set, 2 million of which are limited companies and 1 million of which take other legal forms. These non-limited businesses that take other legal forms are typically sole traders or partnerships. This compares to a total of around 5.9 million SMEs in the UK business population according to aggregate statistics compiled by the UK government's business department, 2 million of which are limited companies.[9] This implies that the data set has very high coverage of SMEs that take the form of limited companies and much lower coverage of other legal forms. It covers around half of SME limited company turnover and 60% of current account balances. Appendix C sets out more detail on the representativeness of the data. The data covers all sectors of the economy and all regions of the UK. The data contains monthly observations for all firms starting in 2017, soon after the introduction of the legislation, but data quality is highest from 2018 onward, so we focus on this period in our analysis. The SMEs in the data typically have one current account with one bank. Relatively few of them have separate debt accounts but most have overdraft limits associated with their current accounts. Debt accounts are classified into around 20 different categories, with credit cards the most common. There is also some information on missed debt repayments and defaults, which we do not analyze in this paper but intend to analyze in future research.

This data set contains information on very small businesses that few other studies in the literature have been able to analyze before, because they are typically under-represented in firm-level data and in credit registers. Many papers in the literature use information on large listed companies compiled via sources such as Compustat or Capital IQ. Recent papers covering the impact of Covid-19 on SMEs have used: survey data with small sample sizes [Buffington et al., 2020, Bartik et al., 2020, Bloom et al., 2021]; private sector data with large sample sizes but low representativeness [Chetty et al., 2020, Autor et al., 2020]; or banking sector data that explicitly excludes small businesses [Chodorow-Reich et al., 2020]. Even papers such as Bahaj et al. [2020] that draw on data from Companies House, where all UK

[8]Based on 2019 current account inflows, there appears to be around 3,000 firms with more than £25 million in turnover. Our results are robust to excluding them.

[9]Note that the aggregate data only allows us to identify SMEs on the basis of their number of employees, not their turnover as in the data set we use in the paper. For the comparisons here we focus on firms with less than 250 employees, which is a widely-used definition of SMEs.

limited companies have to publicly report their accounting information, only have detailed information on turnover and other flow variables for fewer than 50,000 firms, which tend to be relatively large.[10]

For all of the limited companies in the data set we have registration numbers that allow us to match them to Companies House data acquired via Bureau van Dijk. This gives us additional information on their firm description, age, the sector in which they operate (their SIC code), simple balance sheet variables (such as total assets and total liabilities) and in relatively few cases we can obtain information on their number of employees and profit and loss accounts. We match in the latest reported information for all companies, which typically refers to accounting periods that cover calendar years 2018 or 2019. We use these variables as lagged controls in our regression analysis.[11] We use the same unique IDs to merge in loan-level supervisory data obtained from 15 major UK banks collected in August 2020 on all of the companies that had obtained BBLS loans by that point.

2.2 Key variables

In this paper we focus much of our analysis on firm-level measures of turnover and cash flow, which we derive using the monthly current account transaction data. The turnover measure should be interpreted as the total monthly inflows into all current accounts held by each firm, expressed in nominal terms. Note the data does not allow us to split out the components of current account flows. We only have information on total inflows and outflows. More formally, we use the following formula to compute turnover in month t for firm i: $turnover_{i,t} = inflows_{i,t} - newloans_{i,t}$. We subtract any inflows that relate to new loans that we observe firms take out in a given month, which gives us a clean measure of the operating performance of the firm, although it differs from accounting measures of turnover in a few key ways.[12] We also analyze a measure of total costs: $costs_{i,t} = outflows_{i,t}$. Note

[10] The reporting criteria for Companies House depend on company size, with less onerous reporting requirements for smaller companies and micro-entities. Small companies (micro-entities) can opt to file only abridged (micro-entity) accounts if they fulfill at least two of the following criteria: a turnover of £10.2 million (£632,000) or less; £5.1 million (£316,000) or less in assets on balance sheet; 50 (10) employees or less.

[11] This means that we do not analyze non-limited companies in this paper. Further work could look at them in more detail.

[12] For example, it is a cash-based measure rather than accruals-based like accounting data. It may include government transfers and in some cases it could include financing flows from outside the nine major banking

that outflows will include total labor costs, non-labor costs and capital expenditure as well as any shareholder payouts and debt servicing costs. We are not able to separate out these different cost components out in the data.[13] Appendix B sets out more detail on the data cleaning process.

We compute year on year growth in turnover and costs, which helps to strip out seasonality. Focusing on year on year variables means that some firms drop out of our sample because they do not appear for at least one full year.[14] The firm-level growth measures we construct follow the approach outlined in Davis et al. [1996], which is often referred to in the literature as the "DHS growth" measure. This growth measure is produced by dividing the year on year change in turnover in a given month by the average turnover this year and in the same month the year before. Specifically we calculate the following for all firms, for both turnover and costs:

$$turnovergrowth_{i,t} = \frac{turnover_{i,t} - turnover_{i,t-12}}{\frac{1}{2}(turnover_{i,t} + turnover_{i,t-12})}$$

Where t denotes months and i is an individual firm. This measure helps to take account of intensive and extensive margins of growth and firms that face zero turnover or cash flow in a given month. For example, if a firm goes from zero turnover in May 2019 to positive turnover in May 2020, the turnover growth measure takes a value of $+2$. If a firm goes from positive turnover to zero turnover it takes a value of -2. Note it can only take values in the range -2 to $+2$. It is monotonically related to a conventional growth rate based on the formula $growth = \frac{2*DHS}{2-DHS}$. Indeed, for small growth rates it is very close to a conventional growth rate.

We exploit the data on loan accounts to identify borrowing that has taken place under the government-guaranteed loan schemes. Given the size of the firms in the sample, the BBLS is by far the most important of the schemes. We identify these loans based on the timelines of the scheme (it has been available since May 2020) as well as the size and type of

groups covered by the data set.

[13] Note that identifying the impact of government support is not straightforward. The CJRS would likely lead to cash outflows at the time that furloughed employees are paid their reduced wages, followed by cash inflows when the government compensates businesses for their labor costs at a later date.

[14] In practice this means our regression sample contains around 1.8 million SMEs compared to 1.98 million in the latest snapshot.

loans available.[15] We combine this approach with the loan-level data mentioned above. As of September 2020, our data shows around 700,000 BBLS loans totaling £24 billion with the top three sectors by absolute volume being *Construction*, *Wholesale and Retail* and *Professional and Scientific*. This suggests we have around two thirds coverage of the aggregate volumes under the scheme at that time.[16]

2.3 Stylized facts

Table 1 shows some simple summary statistics for the key variables in the data set. There are around 1.8 million limited companies. The median turnover in the data set is around £6,500 per month, or £80,000 per year, which implies that many firms in the data set are very small. The mean is above the median, which points to a positive skew in the firm size distribution. Assets data follows a similar pattern to turnover. The average firm is around 6 years old and most are in the 3 to 10 year range.

Table 1: Summary statistics, firm-level variables

	Turnover (£, monthly)	Age (years)	Assets (£, stock)
Mean	39620	8.36	1099280
Median	6551	6.12	36280
10th percentile	174	2.13	1450
25th percentile	2024	3.47	8880
75th percentile	19581	10.35	152450
90th percentile	63503	16.75	620250

Figure 1 shows how year on year turnover growth has evolved for UK SMEs since the start of 2018. Before 2020 the median firm typically faced zero growth in turnover year on year but this picture changed significantly in 2020. The median firm faced a reduction in turnover of around 30% year on year in May 2020, which was the peak of the crisis on this

[15]The BBLS allows firms to take out standardized amortizing term loans with fixed 2.5% interest rates in amounts up to the lower of 25% of pre-Covid-19 turnover or £50,000. We also identify some loans that could be CBILS loans, although given that the CBILS scheme is less standardized than BBLS we struggle to distinguish them from normal lending in the data. See the British Business Bank website for details.

[16]The missing component likely reflects a combination of slow reporting as a result of system changes at the major banks and the fact that our data only covers nine major banks.

measure. However, the median masks heterogeneity across firms. The 75th percentile of the distribution has remained relatively unchanged in 2020 compared to 2019. But the 25th percentile dropped by more than the median and had not yet recovered fully by December 2020. This suggests that the Covid-19 crisis led to a significant drop in turnover for many firms and a significant increase in dispersion across firms. Figure 9 in the appendix compares year on year SME turnover to national accounting aggregate data, showing that it broadly tracks an aggregate measure of private non-financial company earnings.

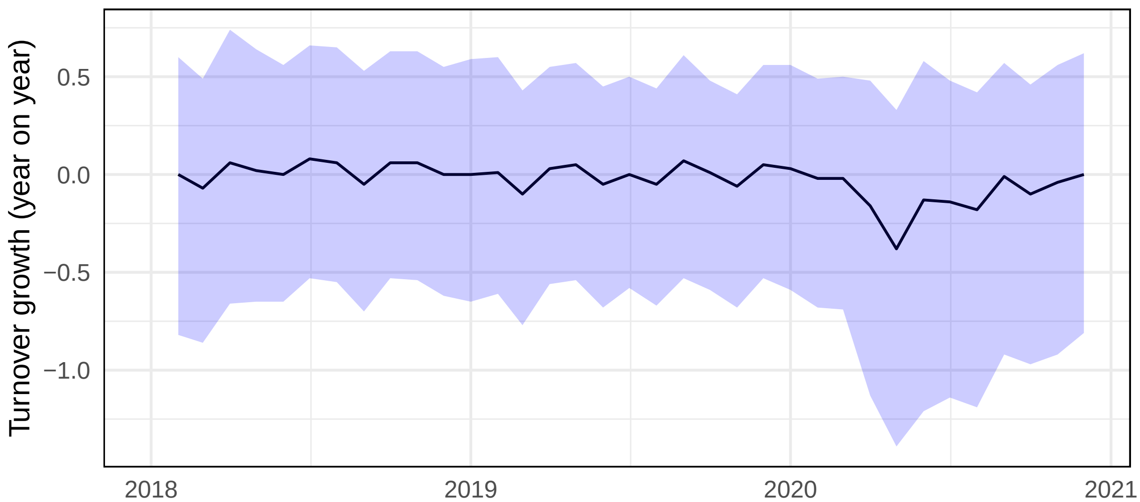

Chart shows the median (the line) and the interquartile range (the shaded area). DHS growth rates (see text for explanation). Values of −2 indicate firms going to zero turnover or leaving the sample.

Figure 1: Historical turnover growth by month

The appendix presents a number of other stylized facts. Figure 10 shows how turnover growth varied across firms in different sectors, revealing significant heterogeneity. The dispersion of median turnover growth across sectors increased markedly between January 2020 and May 2020, the peak of the Covid shock. The most impacted sectors in May 2020 were *Arts and Recreation* and *Accommodation and Food*. Figure 11 shows how median turnover growth varied across geographical areas, based on the 121 postcode areas in the UK. As before, it shows median turnover growth by area in January 2020 vs May 2020. Most areas saw lower median turnover growth in May, but the lowest growth rates were particularly concentrated in parts of Scotland and London. Figure 12 shows the proportion of firms in each area that took out a BBLS loan in our data set. There was significant regional heterogeneity in the probability of borrowing under the BBLS scheme.

3 Impact on turnover and costs

In the previous section, we showed that the first half of 2020 saw a large drop in turnover year on year for the average UK SME and that there appeared to be significant heterogeneity across firms. In this section we analyze the impact of the various public health measures (the 'lockdowns') using linear regression techniques. We also document the heterogeneity of the crisis across firms by analyzing the factors that correlate with the scale of the turnover change at firm level, including age, size, sector and local area.

3.1 Average impact of the lockdowns

The stylized facts presented in Section 2 suggest that many UK SMEs faced a significant reduction in turnover from April 2020 onward. This coincides with the first national lockdown, which began in late March 2020 and was not relaxed for many businesses until early July 2020.[17] We have constructed a series of dummy variables that record the businesses that were affected by each of the lockdown measures in each month of 2020, based on their registered business addresses. For example, a business that operates in Leicester was subject to the first national lockdown from March to July, then a local lockdown until November and then the second national lockdown in November. See Appendix D for more details on the assumptions underpinning this. We have used this measure as an explanatory variable in regressions to document the correlation between the public health measures and SME turnover and cash flows.

The baseline regression we use to analyze turnover growth takes the following form (we use the same specification for costs growth):

$$turnovergrowth_{i,t} = \beta lockdown_{i,t} + v_i + \epsilon_{i,t} \tag{1}$$

The dependent variable, $turnovergrowth$, is year on year growth in turnover for firm i in month t. Appendix A reports statistics on the distribution of this variable. $lockdown$ is a series of dummies denoting different levels of public health measures. In our simplest specification this comes in the form of a 'post-March' dummy, which takes a value of 1

[17] The lockdown saw the government introduce new legislation to contain the spread of Covid-19, which limited the number of people that could leave their houses and required many businesses to be completely closed.

for all months from April 2020 to December 2020. Appendix D has more detail on how these variables are constructed in practice. We include firm fixed effects, denoted v_i. The β coefficients should be interpreted as the correlation between a given set of lockdown measures and year on year growth in turnover for SMEs, relative to the period before, controlling for time-invariant characteristics such as their sector and region.

Table 2 presents regression results for the turnover growth variable. Column (1) shows the results of the simplest version of equation (1), a regression of year on year turnover growth on the 'post-March' dummy controlling for firm fixed effects.[18] The coefficient estimate suggests that for the average SME, turnover growth was around 30 percentage points lower in the period from April 2020 to December 2020 than it was before the pandemic occurred. Column (2) includes a richer set of lockdown dummies, which allow us to compare different sets of public health measures to one another. Note that the coefficient estimates implicitly represent effects that are measured relative to the period before April 2020. These estimates suggest that the first national lockdown, which came to an end in June 2020, was the most severe in terms of its impact on SME turnover growth. The other sets of public health measures were less severe and relatively similar to one another in terms of impact. Even in periods from April 2020 onward when there were no public health measures being applied, SME turnover and cost growth were both around 30 percentage points lower than in periods before the pandemic. Columns (3) and (4) run the same regressions but with a different dependent variable, year on year costs growth. Note these coefficient estimates should be interpreted as event study correlations and not *causal* effects of the public health measures per se.

[18]See Gaure [2013] for details on the R command we use.

Table 2: Baseline regression estimates for average lockdown impact

	Dependent variable:			
	Turnover growth		Costs growth	
	(1)	(2)	(3)	(4)
Post-March 2020	−0.332***		−0.325***	
	(0.0004)		(0.0003)	
National lockdown 1 (Q2 2020)		−0.358***		−0.356***
		(0.001)		(0.001)
National lockdown 2 (Q4 2020)		−0.291***		−0.260***
		(0.001)		(0.001)
Local lockdowns		−0.316***		−0.284***
		(0.003)		(0.002)
Tier 1		−0.310***		−0.318***
		(0.001)		(0.001)
Tier 2		−0.315***		−0.342***
		(0.001)		(0.001)
Tier 3		−0.292***		−0.335***
		(0.002)		(0.002)
Post-March 2020, no restrictions		−0.320***		−0.299***
		(0.001)		(0.001)
Firm fixed effects	*Yes*	*Yes*	*Yes*	*Yes*
Observations	43,442,410	37,714,382	46,222,428	40,123,098

*p<0.1; **p<0.05; ***p<0.01

Figure 2 explores the monthly impacts of the Covid-19 crisis in more detail. Specifically, it presents coefficient estimates and 99% confidence intervals from a regression that contains a full set of month dummies for all months in 2020, controlling for firm fixed effects. The figure plots results for different dependent variables, as in table 2. The figure shows that there was a large fall in turnover growth for the average SME from April 2020 onward, relative to the period before. The impact troughed in May 2020 at around a 40 percentage point drop and had not returned to the January 2020 level even by December 2020. Costs growth fell substantially, in line with turnover growth, supporting cash flows. When turnover growth dropped in April 2020, costs growth fell by even more. By June 2020 the impact on costs growth was smaller than the impact on turnover growth, implying some cash flow pressure for the average SME. Over the period as a whole, turnover growth and cost growth moved roughly in line with one another, implying that the average SME managed to offset the large reduction in turnover by reducing outgoings to maintain cash flows.

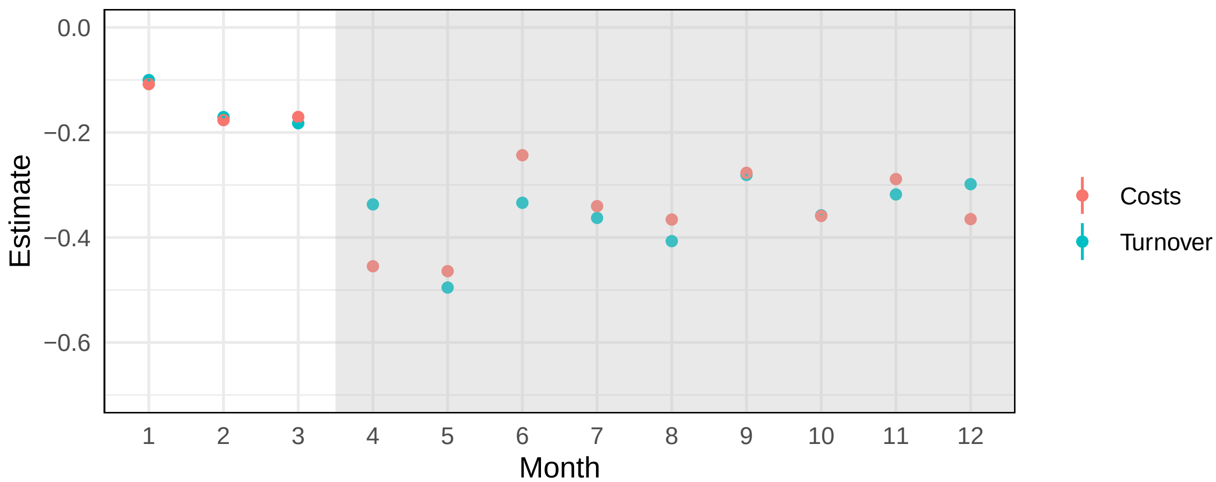

Chart shows the estimated coefficients on month dummies in 2020 in a regression as in equation (1), with different dependent variables. Lines show 99% confidence intervals (standard errors are typically very small). Shaded area shows the period after March 2020.

Figure 2: Estimated coefficients on month dummies for 2020

All of the analysis presented up to this point should be interpreted as referring to the experience of the average SME. We now explore firm heterogeneity by estimating how different types of SMEs performed. Specifically we interact the 'post-March' dummy with different sets of dummy variables, denoted Z, that reflect lagged firm characteristics, such as their sector or region:

$$turnovergrowth_{i,t} = \beta postMarch_{i,t} * \mathbf{Z_{i,t-1}} + v_i + \epsilon_{i,t} \tag{2}$$

In the next few sub-sections we present the results of this heterogeneity analysis. All of the results are based on the regression specification in equation (2). There is more analysis of firm-level heterogeneity in Appendix A.

3.2 Sectoral heterogeneity

The Covid-19 shock and the public health measures had very different impacts on firms in different sectors. We use information obtained from Companies House to identify the sector in which each SME operates. We use this information to create a variable that takes a value of 1 if a firm operates in a sector that is highly exposed to social distancing, which we refer to as a 'social sector'. This includes businesses in the non-food retail sector, passenger transport, hospitality, arts and leisure and personal care. We follow Joyce and Xu [2020] in determining the most exposed sectors. These are the businesses that we would expect to face the biggest direct impact of the lockdowns on turnover.

Figure 3 presents the results of the sectoral analysis. The coefficient estimates on the chart should be interpreted as estimates of the average year on year growth in turnover or costs for SMEs in each of the sectors, relative to the period before April 2020. As expected, there was significant sectoral heterogeneity in the turnover reduction. The average SME in the *Arts and Recreation* sector saw more than a 40 percentage point reduction in turnover growth year on year, compared to only around a 10 percentage point reduction for the average SME in the *Agriculture* sector. This pattern is borne out in the 'social sector' variable too. The costs effects imply that there was significantly less heterogeneity in terms of impacts on cash flow growth. Most sectors saw a large fall in costs growth as turnover growth fell. Some sectors, such as *Real Estate*, actually managed to cut costs growth by more than their fall in turnover growth on average.

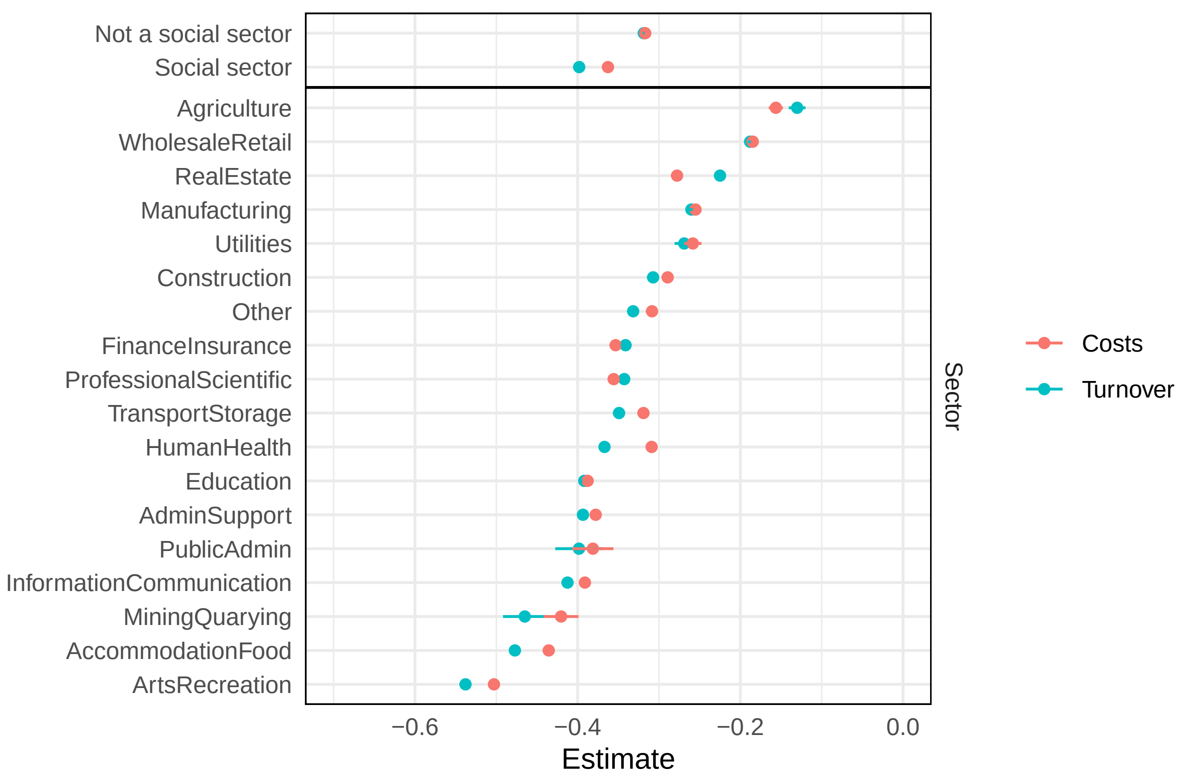

Chart shows the estimated coefficients on sector dummies interacted with 'post−March' dummy in a regression as in equation (2), with different dependent variables. Lines show 99% confidence intervals (standard errors are typically very small).

Figure 3: Estimated coefficients on sector interaction terms

3.3 Geographical heterogeneity

There is likely to have been substantial regional variation in how the Covid-19 shock played out. We have used information on the business address of SMEs to analyze this dimension.[19]. Postcodes allow us to match each SME to their NUTS 1 region of the UK e.g. London. We also analyzed regional heterogeneity based on the characteristics of the local area in which they operate. Analysis of SME performance in the US by Chetty et al. [2020] suggested that SME revenues fell by more in areas where higher income people tend to live, perhaps because they cut their spending on local goods and services by more than lower income people. To extend this analysis we used data on the deprivation of local areas in England, which measures living conditions in different areas of England.[20] Around half of the index weight is placed on income and employment, with the other half capturing factors like health and education. This allowed us to identify more and less affluent lower layer super output areas within regions (there are around 50,000 of these small areas in total).

Figure 4 presents the results of the regional analysis. This shows that there was some geographical heterogeneity in the impact on turnover growth, although less than for sectors. At one end of the spectrum, the average SME in Scotland or London faced around a 35 percentage point reduction in turnover growth over the April 2020 to December 2020 period compared to the period before the shock. At the other end, the average SME in Northern Ireland faced around a 25 percentage point reduction. There is also some evidence that the turnover shock was slightly more severe in the most affluent parts of the country, where incomes and employment levels are highest, although this effect appears to be small. The picture for costs growth is similar.

[19] In some instances we have filled in missing information using the address of the firm headquarters, which is recorded in Companies House

[20] Note this means that the analysis of deprivation levels leads to us losing information on businesses in Northern Ireland, Scotland and Wales.

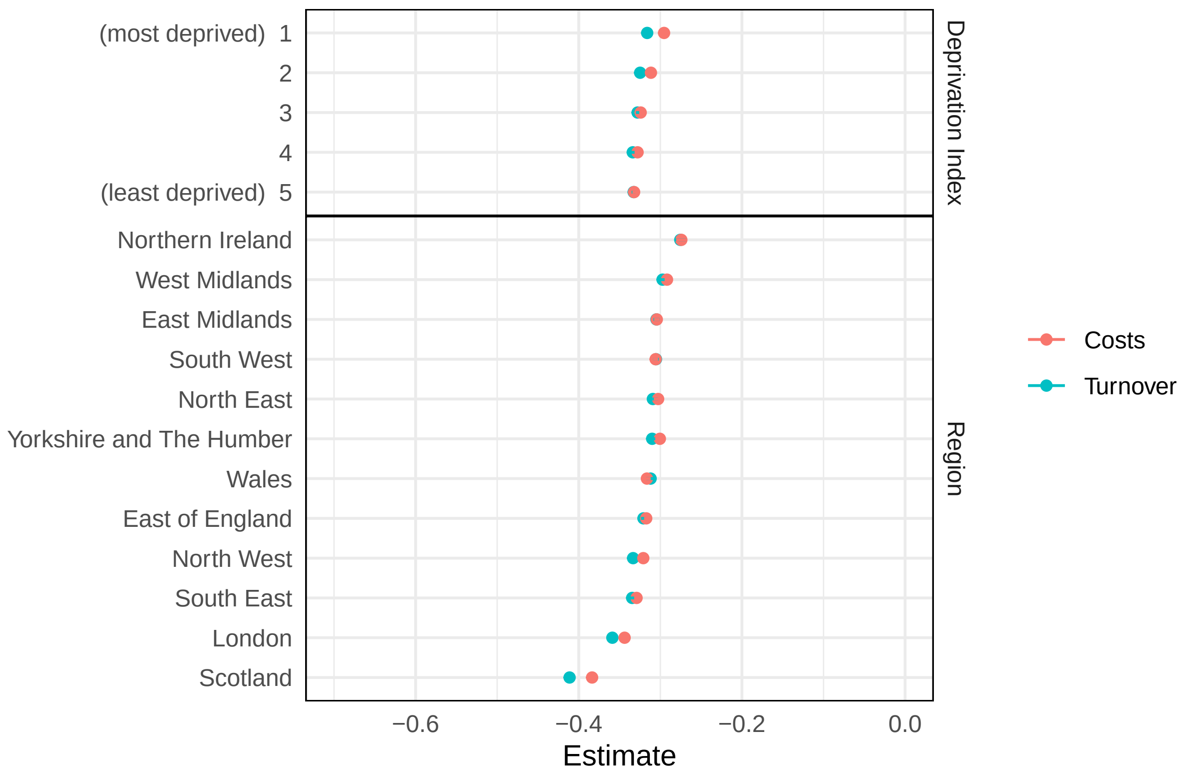

Chart shows the estimated coefficients on region dummies interacted with 'post−March' dummy in a regression as in equation (2), with different dependent variables. Lines show 99% confidence intervals (standard errors are typically very small).

Figure 4: Estimated coefficients on sector interaction terms

3.4 Firm age heterogeneity

Figure 5 presents analysis of the impact of the Covid-19 shock by firm age. We have used information from Companies House on the incorporation date of companies to compute their age in years. We split this into deciles by ranking firms from youngest to oldest. The chart shows a clear relationship between age and turnover growth during Covid-19. The youngest SMEs in the bottom decile of the age distribution faced around a 45 percentage point reduction in turnover growth relative to the period before the shock. The SMEs in the top decile of the distribution, which are more than 20 years old on average, faced only a 20 percentage point reduction in turnover growth. The gap between the turnover and costs dots on the chart shows that the youngest SMEs appear to have had the smallest cash flow hits in relative terms on average, perhaps because they were particularly well supported by government policy.

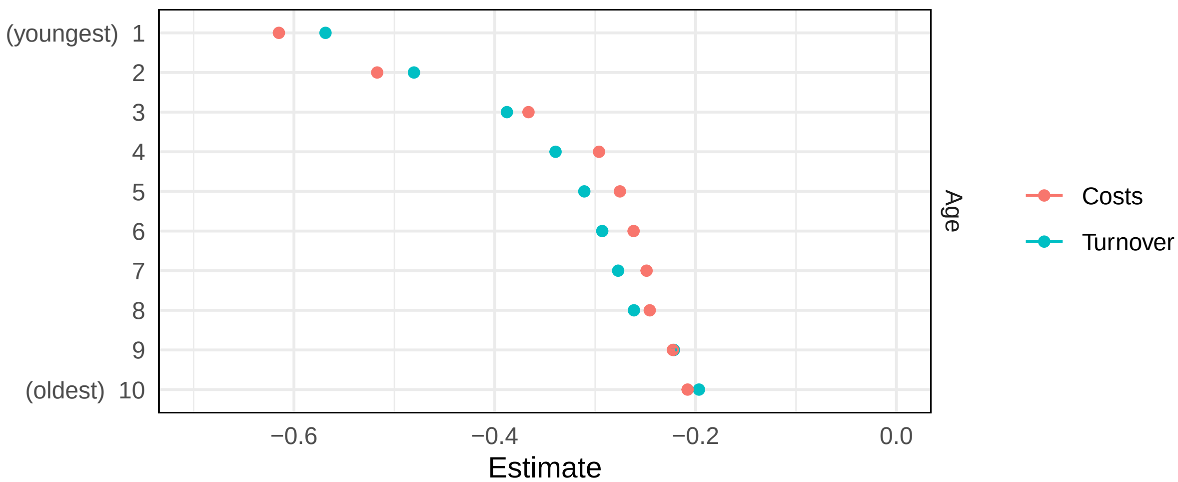

Chart shows the estimated coefficients on firm age dummies interacted with 'post−March' dummy in a regression as in equation (2), with different dependent variables. Lines show 99% confidence intervals (standard errors are typically very small).

Figure 5: Estimated coefficients on firm age interaction terms

3.5 Firm size heterogeneity

Figure 6 presents analysis of the impact of the Covid-19 shock by firm size. We have used historical data on average annual turnover for each firm to identify smaller and larger SMEs. The smallest SMEs appear to have had the smallest impact on turnover growth during the Covid shock. The largest SMEs had the most negative impact, although there are relatively few of these in our data set. The second most severely affected group of SMEs were those in the £100,000 to £1 million turnover bracket. There is some evidence that the largest SMEs managed to reduce their costs by more than smaller SMEs.

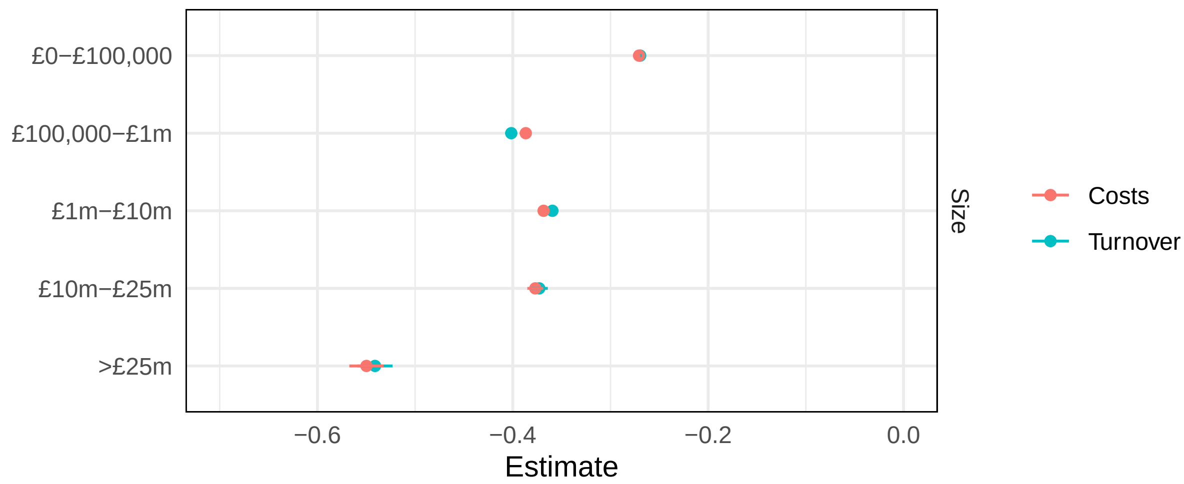

Chart shows the estimated coefficients on firm size dummies interacted with 'post−March' dummy in a regression as in equation (2), with different dependent variables. Lines show 99% confidence intervals (standard errors are typically very small).

Figure 6: Estimated coefficients on firm size interaction terms

4 Government-guaranteed loan schemes

In this section we analyze the probability of obtaining a BBLS loan, using a probit model. By the end of January 2020 the official government statistics suggested that around 1.5 million businesses had borrowed a total of around £45 billion under the BBLS scheme. We attempted to identify the borrowers in the credit reference agency data set. In the first instance we identified firms that had taken BBLS loans using a loan-level supervisory data set acquired from 15 major UK banks in August 2020, which covers 90% of loans issued under the scheme by that date. This gave us around 600,000 observations among the limited companies that appear in the credit reference agency data set. We added to this based on identifying all firms that borrowed any amount up to £50,000 (the BBLS scheme limit) after May 2020, when the scheme opened (this gave around another 100,000 firms).

We analyzed similar firm characteristics to those in the previous section (see Appendix A for the regression tables). Specifically, we fitted the following probit model:

$$P(TookOutBBLSloan) = \Phi[\alpha + \beta_1(\mathrm{turnovergrowthin2020_i}) + \beta_2(\mathrm{sizebracket_i}) + \beta_3(\mathrm{region_i}) + \beta_4(\mathrm{sector_i}) \ + \beta_5(\mathrm{agequintile_i}) + \epsilon_i] \tag{3}$$

Figure 7 shows that the probability of taking out a BBLS loans varies by firm size, region and sector. We found significant variation across sectors, with the effects on the chart shown relative to *Real Estate*, which had a relatively low probability of taking BBLS loans. The sectors which were most directly impacted by the lockdowns, namely *Accommodation and Food*, *Transport and Storage*, *Wholesale and Retail* were the most likely to have used BBLS. Looking at the impact of regions, relative to London, we found much less regional variation, controlling for sector. The smallest firms, as measured based on their 2019 turnover, were least likely to use the scheme, closely followed by very large SMEs. Firms in the North (North West and North East) were the most likely to have taken a BBLS loan, whereas those in Northern Ireland were the least likely. We also found a linear effect of age, with the oldest companies (higher quintiles) least likely to use the BBLS.

We also analyzed the probability of taking a BBLS loan as a function of turnover growth in 2020.[21] This reveals an inversed U-shape, where the firms in the bottom and top deciles of the growth distribution in 2020 were the least likely to take out BBLS loans.

[21] We have removed the entries where turnover growth was exactly zero, which are likely to be errors in the underlying data.

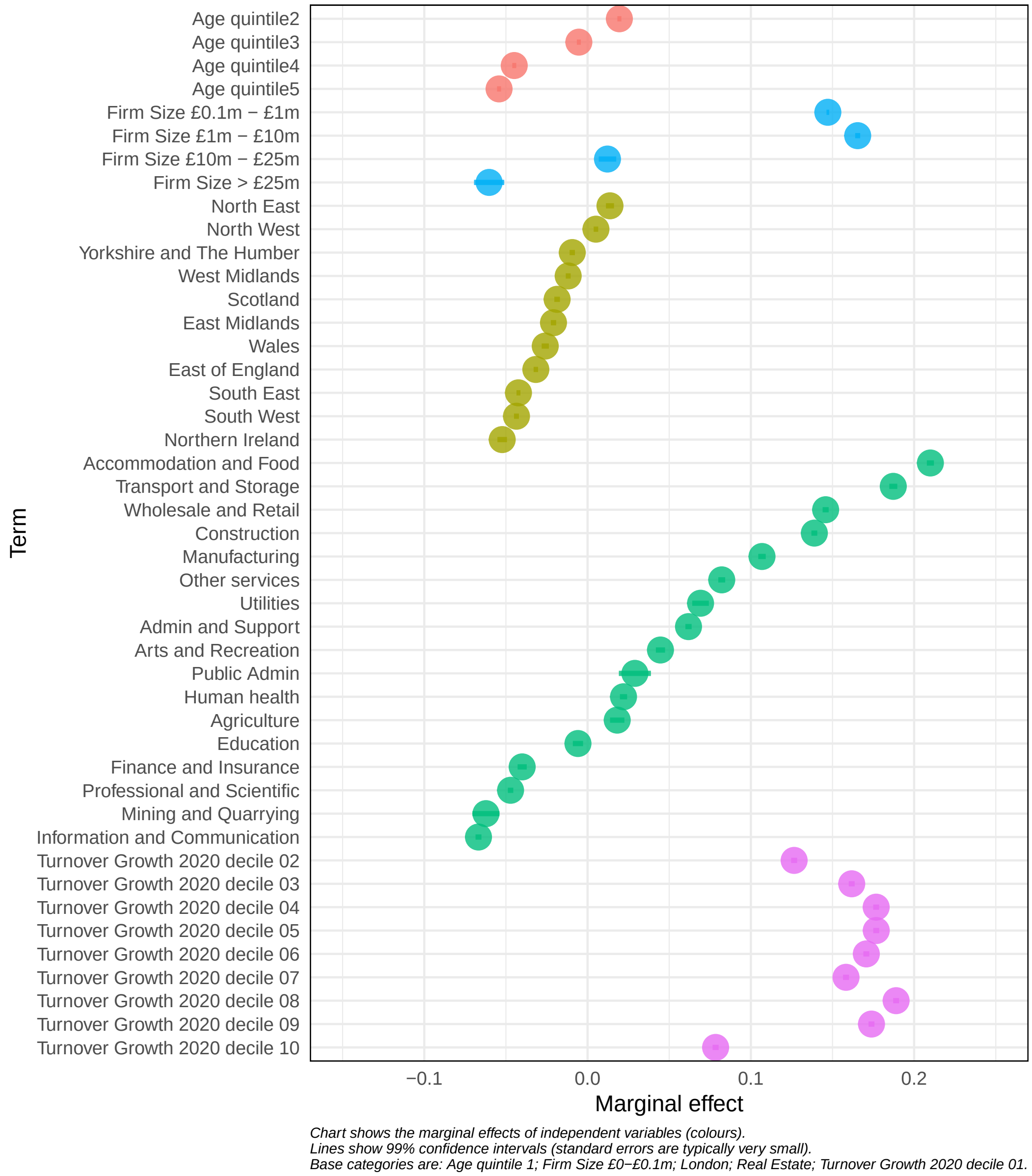

Figure 7: Estimated marginal effects from BBLS probit model

5 Machine learning models

The quality of our conclusions is reliant on the quality of the models we use. In prior sections we used traditional regression models to analyze firm-level correlations, fitting the entire data set to the model. While this approach is appropriate for our analysis, in this section we explore machine learning models, to evaluate whether they more accurately represent the data and its patterns. Furthermore, in linear regressions it is often very difficult to analyze individual companies, while machine learning models offer tools to do this.

Unlike regressions, a typical machine learning approach involves separating the data into a training and a testing subset. The training subset is used for aiding the algorithm to 'understand' the patterns in the data itself, configuring the machine learning algorithm in the process. Afterwards, the configured algorithm is evaluated on the testing subset of the data, to determine how well it had fit the data. The core benefit of a model that fits the data well and is robustly tested on out of sample data is the fact that conclusions derived from this model are reliable, whether related to retrospective understanding of the phenomena in the data, or future forecasting. We have experimented with several different machine learning algorithms, to understand which approach provides a better fit to the data. Each of the algorithms is trained on 80% of the data set, and tested on the remaining 20%. The split regarding which rows are to be used for training or testing is done randomly. A similar methodology was used in Bluwstein et al. [2020].

We have run experiments using the following machine learning algorithms: Random Forest, Support Vector Machines, Ada Boost [Pedregosa et al., 2011]. We have chosen these algorithms as they represent sufficiently varied approaches in a methodological sense. Random Forests are an expansion of the traditional decision tree algorithm, which splits the data according to different nodes in the tree, supporting decision making. An ensemble model of numerous decision trees makes a random forest. Decision trees have been widely known as a robust mechanism for a variety of classification and regression problems. The power of decision tress is only enhanced by their joined combination within ensemble models such as Random Forests, where each tree depends on values of randomly sampled subsets which maintain the same distribution for all trees in the forest. After the creation of the forest of decision trees, the final result stems from a majority vote among each of the specified trees. Random Forests have found numerous uses both in economics [Suss and Treitel, 2019] and computer science [Pal, 2005]. Support Vector Machines are similar to Linear Regression models in the sense that they attempt to produce a hyper-dimensional plane to split the data more meaningfully. Finally, Ada Boost is an example of boosting algorithms, which, like Random Forests, also focus on ensembles of smaller models, however in each iteration of generating a new model, the algorithm uses data points which are likely to have been misclassified in the previous iterations, thus boosting the algorithm's performance.

We use these algorithms to evaluate the fit to the data when classifying whether companies took out BBLS loans or not, and also evaluate the models for their performance regarding inference only, focusing on out of sample data. To perform this evaluation, we use different errors such as Accuracy, Precision, Recall, F1-micro, and F1-macro [Powers, 2020]. These are standard errors typically used to evaluate classification problems as they evaluate the differences between the true value and its predicted counterpart. The results from the classification of whether companies took out a BBLS loan are presented in table 3. All three models perform very similarly, with Random Forest offering the highest accuracy. This implies that Random Forest fits the data better than the other two algorithms. Machine learning approaches in general can also help us understand the interactions between different variables, which is not always very straight-forward when applying traditional regressions.

Table 3: Machine Learning algorithms applied for classification of whether companies took out a BBLS loan

	Random Forest	SVM	Ada Boost
Accuracy	0.670	0.663	0.682
F1-Micro	0.670	0.663	0.682
F1-Macro	0.767	0.748	0.722
Precision	0.935	0.895	0.800
Recall	0.670	0.663	0.682

This table presents the results of an evaluation of three machine learning algorithms using different performance metrics. Values are in the range 0 to 1, with 1 indicating the best performance.

Random Forest models allow us to rank all variables based on their contribution to explaining whether or not a firm has taken out loans. Bholat et al. [2017] describes how Random Forest models work in more detail. Figure 8 presents the results of this analysis, suggesting that the strongest predictor of whether or not a firm has taken out a loan is its size. Firms with turnover between £100,000 to £1 million are most likely to take out a loan. Firms in the *Food and Beverage serving* sector have also been highly affected by the pandemic and have thus been in greater need of loans.

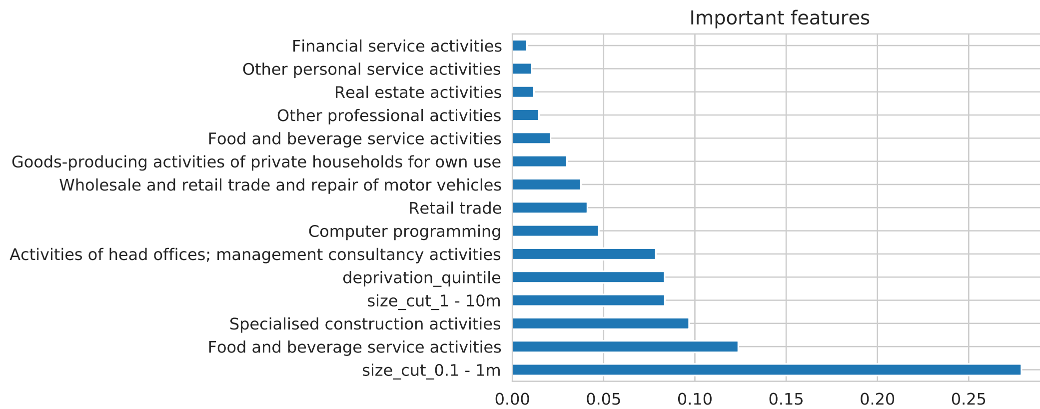

Figure 8: The importance of variables in the machine learning model

This figure plots estimates of the importance of each variable from a Random Forest model.

Shapley values [Joseph, 2020] are also a powerful tool for examining individual companies. Figure 18 in the appendix offers a visual representation of the use of Shapley values for determining whether examples of dummy firm has received a BBLS loan or not. The likelihood of the first firm receiving a BBLS loan is 0.31. The figure shows how different variables and their values contribute positively or negatively towards a higher or lower probability. For example, the fact that the firm's turnover is between £100,000 to £1 million is a strong predictor of the firm taking out a BBLS loan, however the stacked variables against a BBLS loan (being in the 5th age quintile, the Manufacturing industry, etc.) all contribute to lowering the probability.

6 Conclusions and next steps

Central banks around the world are beginning to make more use of novel big data sources to inform policy [Doerr et al., 2021]. This paper introduces a data set to track the performance of 2 million UK SMEs in the aftermath of the Covid-19 shock. The data comes from the SME portfolios of major UK banks and is delivered confidentially to the Bank of England via Experian, a private sector information services company. Unlike most papers on the corporate sector in the literature, the data contains information on very small firms, which were particularly exposed to the Covid-19 shock.

We apply this new data to document a few important facts on the performance of UK

SMEs through the Covid-19 crisis. There was a steep decline in turnover growth when the shock first took hold in April 2020, with a trough in May 2020. Even by December 2020, the latest data point that we analyze in this paper, the average SME had not yet recovered fully. There was significant heterogeneity across firms, with the youngest firms in consumer-facing sectors in Scotland and London facing the largest declines in turnover growth. Cash flow growth declined by much less than turnover growth for the average firm and showed less heterogeneity across firms, helped by cost cutting and substantial government support, including through the Coronavirus Job Retention Scheme (CJRS). The government-guaranteed loans were available to all SMEs but in practice were more likely to be extended to mid-sized SMEs (£100,000 to £1 million), those located in the north of England, younger firms, those with average turnover growth in 2020, and those in sectors particularly impacted by the lockdowns.

The data we have presented in this paper will be useful for future research and monitoring of UK SMEs as we emerge from the Covid-19 crisis. There are a number of potential angles for further analysis, including: the scarring effects of the crisis; zombification caused by the cheap credit that has been extended; relationships between banks and firms and how they affect credit provision; and any signs of distress that show up through missed payments and defaults on debt products. The granularity of the data should allow for broader and deeper insights into SME performance than much of the existing literature.

Appendices

A Additional charts and tables

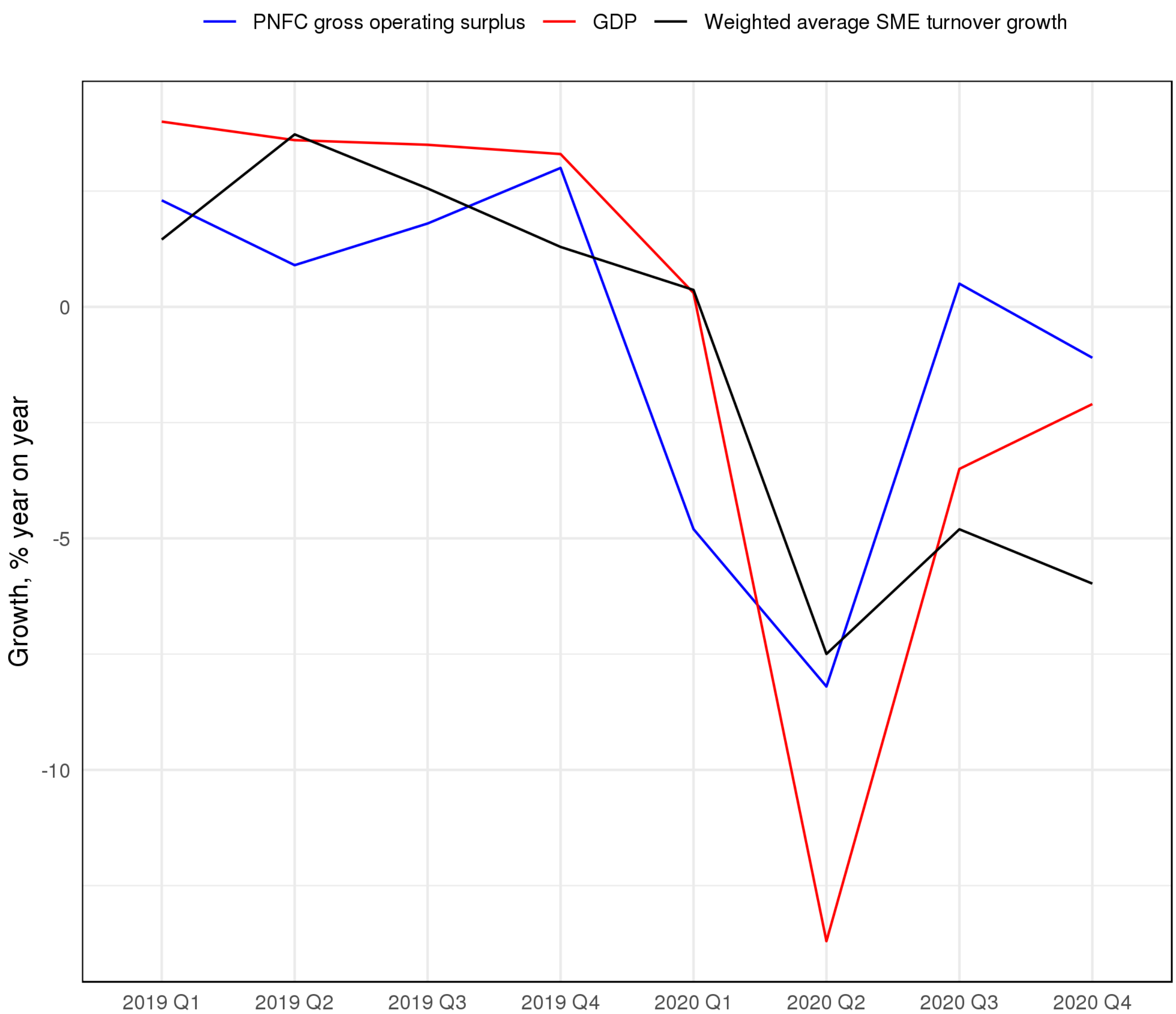

Chart shows growth in a given quarter compared to the same quarter a year earlier. All series are in nomi terms. The ONS data is not available on a monthly basis so we have run the comparison on a quarterly ba PNFC gross operating surplus is a measure of gross profits for private non-financial companies. The SME growth measure is weighted average DHS growth, converted to a standard growth rate. SME weights bas 2019 turnover.

Figure 9: Comparison of SME turnover growth with aggregate data

Table 4: Summary statistics, firm-level growth

	Turnover growth	
	Pre-March 2020	Post-March 2020
Mean	-0.04	-0.21
Median	0	-0.12
10th percentile	-1.98	-2
25th percentile	-0.62	-1.11
75th percentile	0.55	0.5
90th percentile	1.73	1.57
Observations (m)	30.7	12.75
Firms (m)	1.87	1.75

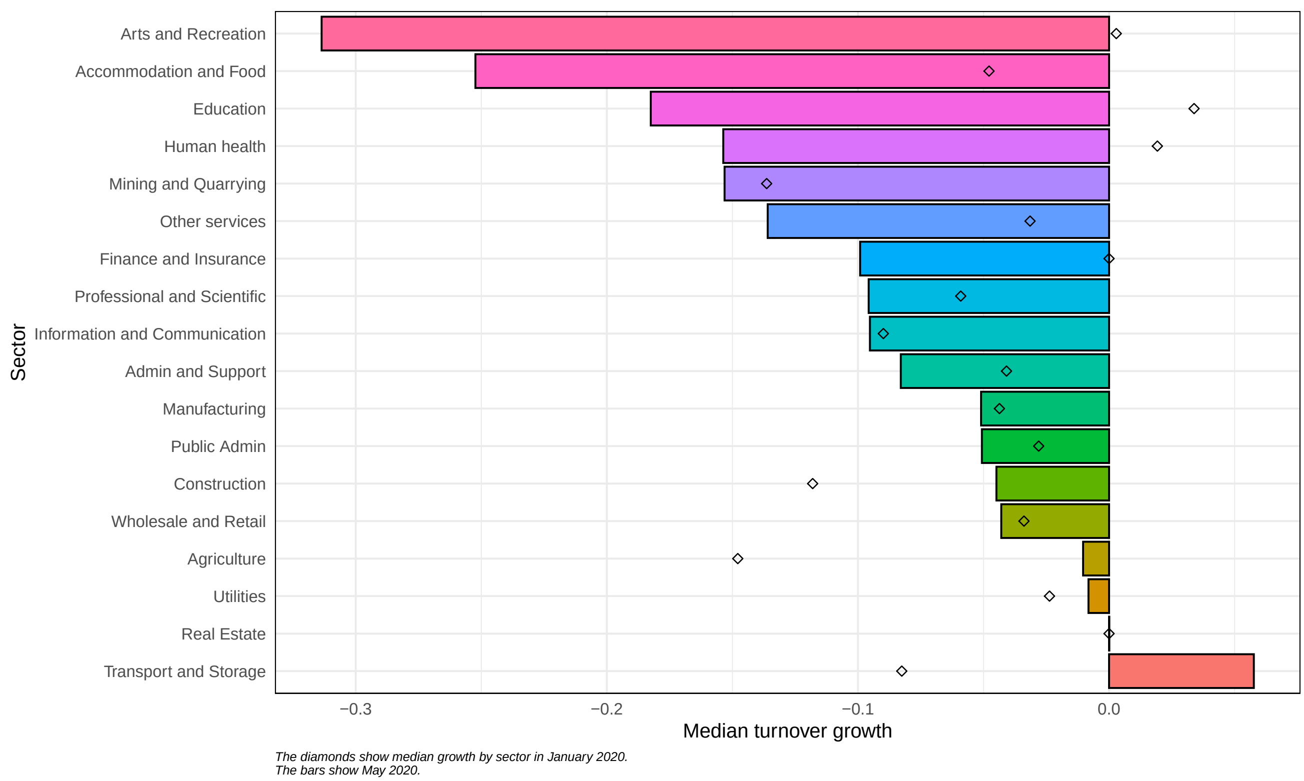

Figure 10: Median turnover growth by sector in January 2020 and May 2020

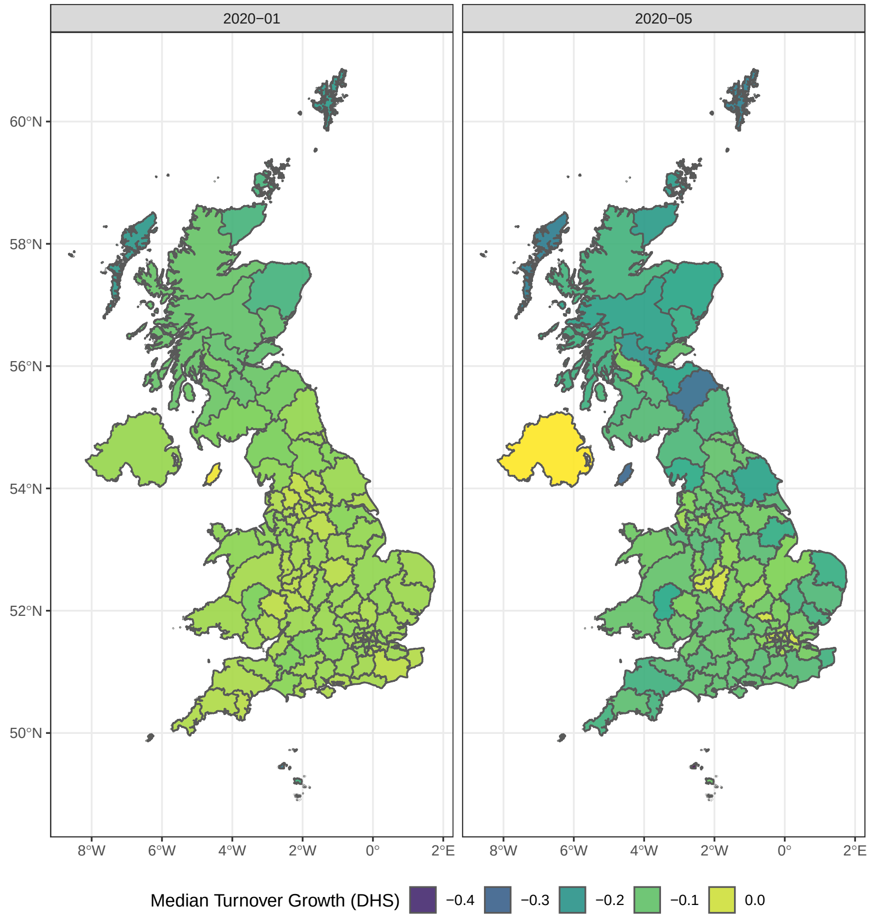

Figure 11: Median turnover growth by area in January 2020 and May 2020

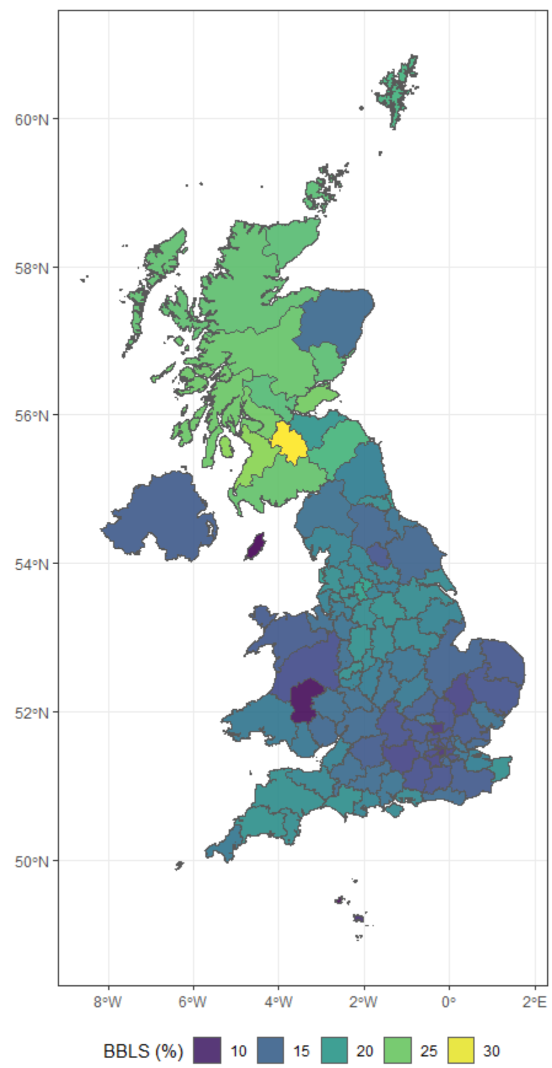

Figure 12: Proportion of firms by area that have taken out a BBLS loan

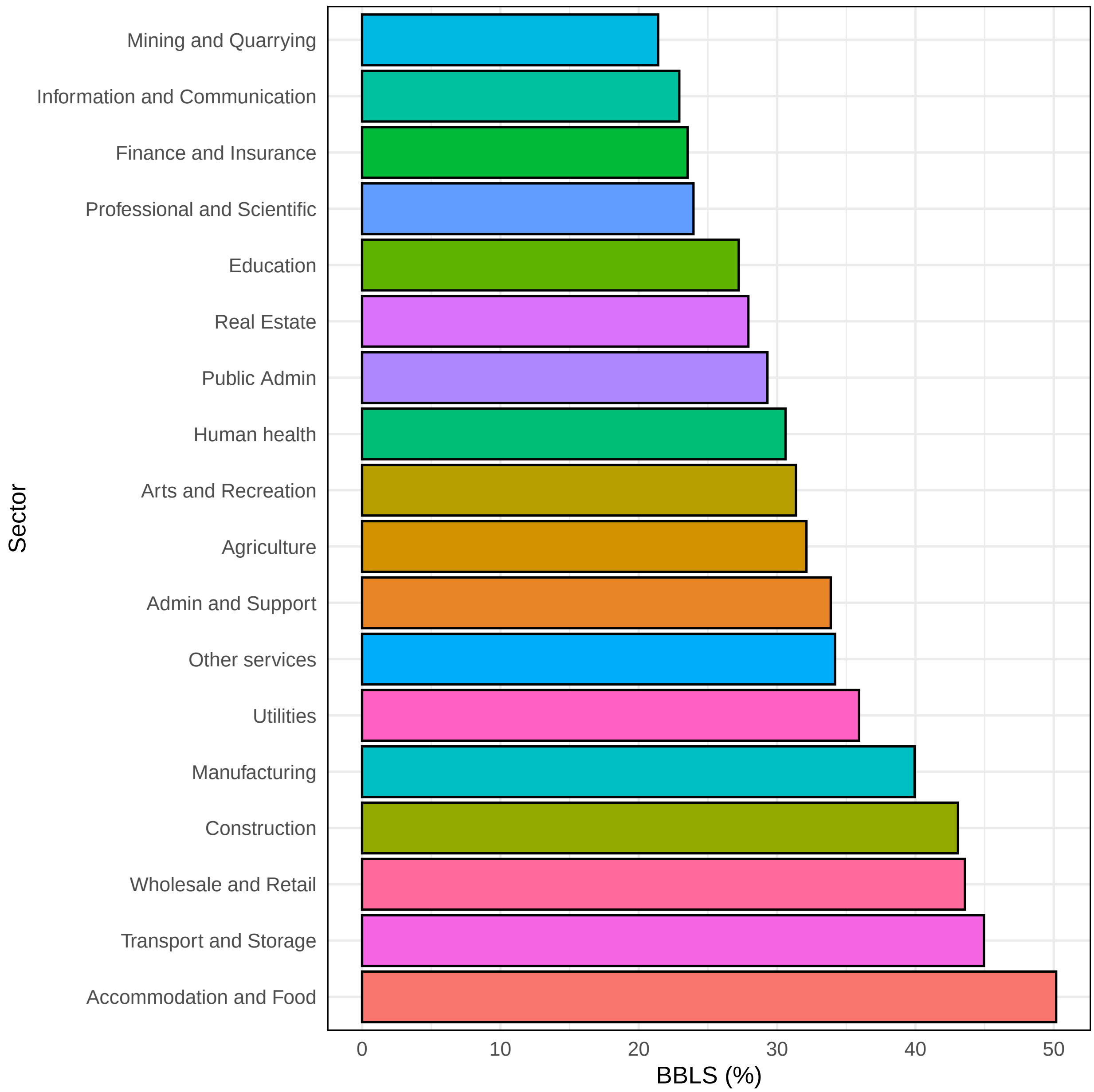

Figure 13: Proportion of firms by sector that have taken out a BBLS loan

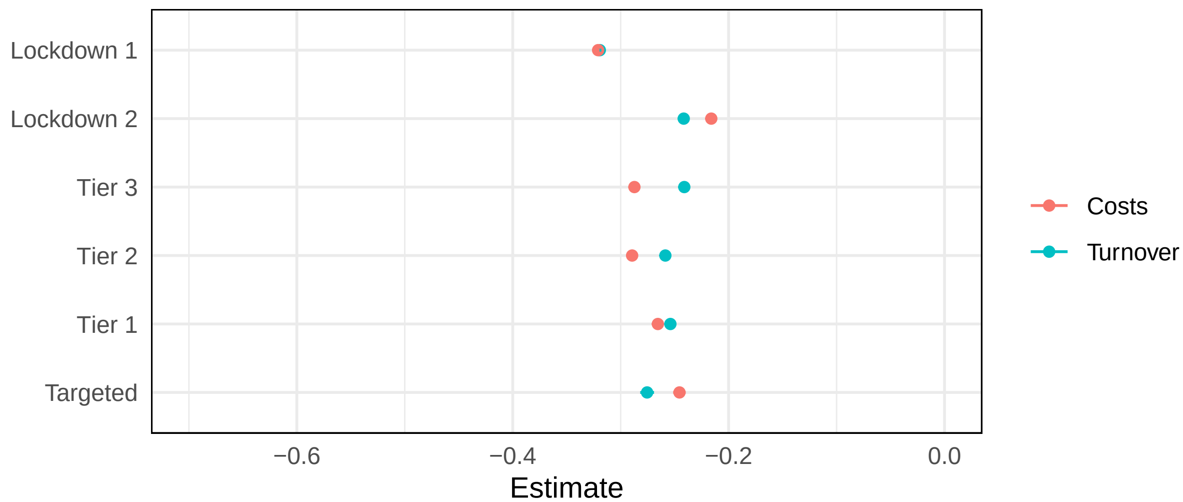

Chart shows the estimated coefficients on lockdown dummies in a regression as in equation (1), with different dependent variables. Lines show 99% confidence intervals (standard errors are typically very small).

Figure 14: Estimated coefficients on lockdown dummies for 2020

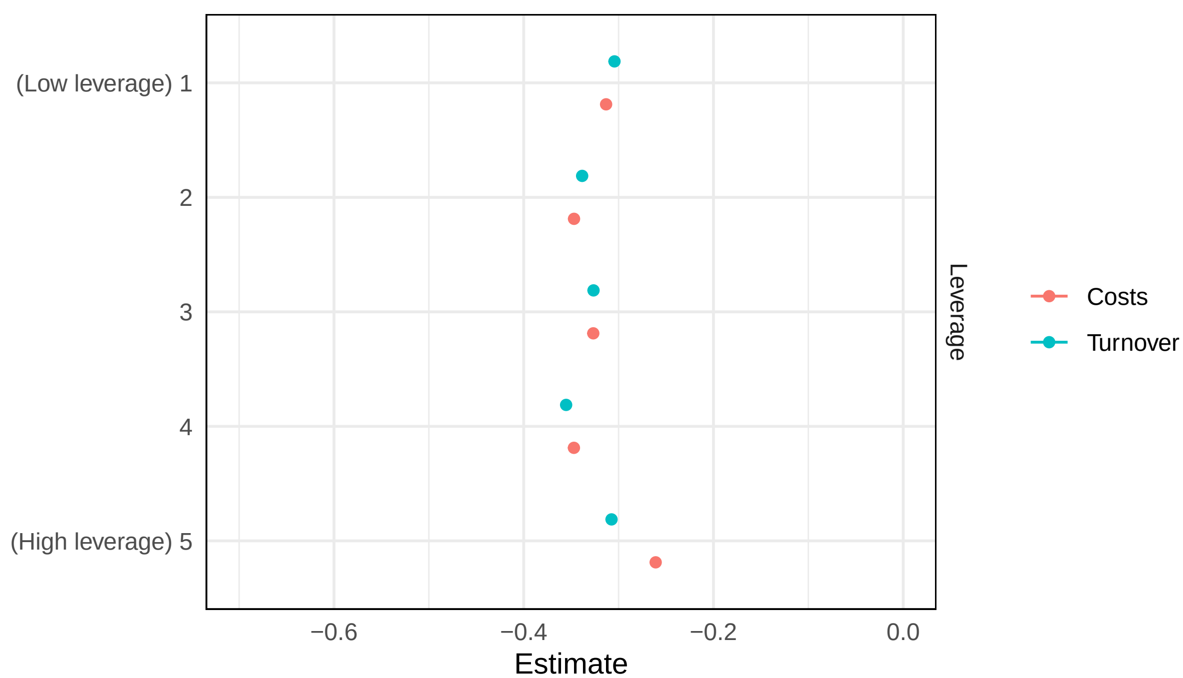

Chart shows the estimated coefficients on leverage quintile dummies interacted with 'post−March' dummy in a regression as in equation (2), with different dependent variables. Lines show 99% confidence intervals (standard errors are typically very small).

Figure 15: Estimated coefficients on leverage interaction terms

Table 5: Estimated coefficients from BBLS probit model

	Dependent variable:	
	TookOutBBLSloan	
	(1)	(2)
size_cut>25m	−0.166*** (0.026)	−0.117*** (0.028)
size_cut0.1 - 1m	0.388*** (0.002)	0.404*** (0.002)
size_cut1 - 10m	0.423*** (0.004)	0.480*** (0.004)
size_cut10 - 25m	0.032** (0.014)	0.110*** (0.015)
turnover_growth_2020_decile02	0.326*** (0.005)	0.323*** (0.005)
turnover_growth_2020_decile03	0.415*** (0.005)	0.428*** (0.005)
turnover_growth_2020_decile04	0.453*** (0.005)	0.473*** (0.005)
turnover_growth_2020_decile05	0.453*** (0.005)	0.472*** (0.005)
turnover_growth_2020_decile06	0.437*** (0.005)	0.464*** (0.005)
turnover_growth_2020_decile07	0.406*** (0.005)	0.437*** (0.005)
turnover_growth_2020_decile08	0.484*** (0.005)	0.519*** (0.005)
turnover_growth_2020_decile09	0.445*** (0.005)	0.463*** (0.005)
turnover_growth_2020_decile10	0.204*** (0.005)	0.219*** (0.005)
East Midlands	−0.056*** (0.004)	
East of England	−0.085*** (0.004)	
North East	0.036*** (0.006)	
North West	0.014*** (0.004)	
Northern Ireland	−0.143*** (0.008)	
Scotland	−0.050*** (0.005)	
South East	−0.114*** (0.003)	
South West	−0.118*** (0.004)	
Wales	−0.070*** (0.006)	
West Midlands	−0.032*** (0.004)	
Yorkshire and The Humber	−0.025*** (0.004)	
Accommodation and Food	0.535*** (0.006)	
Admin and Support	0.161*** (0.005)	
Agriculture	0.048*** (0.012)	
Arts and Recreation	0.117*** (0.007)	
Construction	0.358*** (0.005)	
Education	−0.016* (0.008)	
Finance and Insurance	−0.109*** (0.008)	
Human health	0.058*** (0.006)	
Information and Communication	−0.183*** (0.005)	
Manufacturing	0.275*** (0.006)	
Mining and Quarrying	−0.171*** (0.024)	
Other services	0.213*** (0.005)	
Professional and Scientific	−0.127*** (0.005)	
Public Admin	0.076*** (0.025)	
Transport and Storage	0.477*** (0.006)	
Utilities	0.179*** (0.013)	
Wholesale and Retail	0.375*** (0.005)	
age_quintile2	0.052*** (0.003)	0.032*** (0.003)
age_quintile3	−0.014*** (0.003)	−0.044*** (0.003)
age_quintile4	−0.121*** (0.003)	−0.164*** (0.003)
age_quintile5	−0.146*** (0.003)	−0.157*** (0.004)
deprivation_quintile2		−0.083*** (0.004)
deprivation_quintile3		−0.187*** (0.003)
deprivation_quintile4		−0.263*** (0.004)
deprivation_quintile5		−0.353*** (0.004)
social_distancingsocial_distancing		0.275*** (0.003)
Constant	−0.941*** (0.006)	−0.709*** (0.005)
Observations	1,760,814	1,586,604
Log Likelihood	-1,099,692.000	-1,002,390.000
Akaike Inf. Crit.	2,199,475.000	2,004,826.000
Note:		*p<0.1; **p<0.05; ***p<0.01

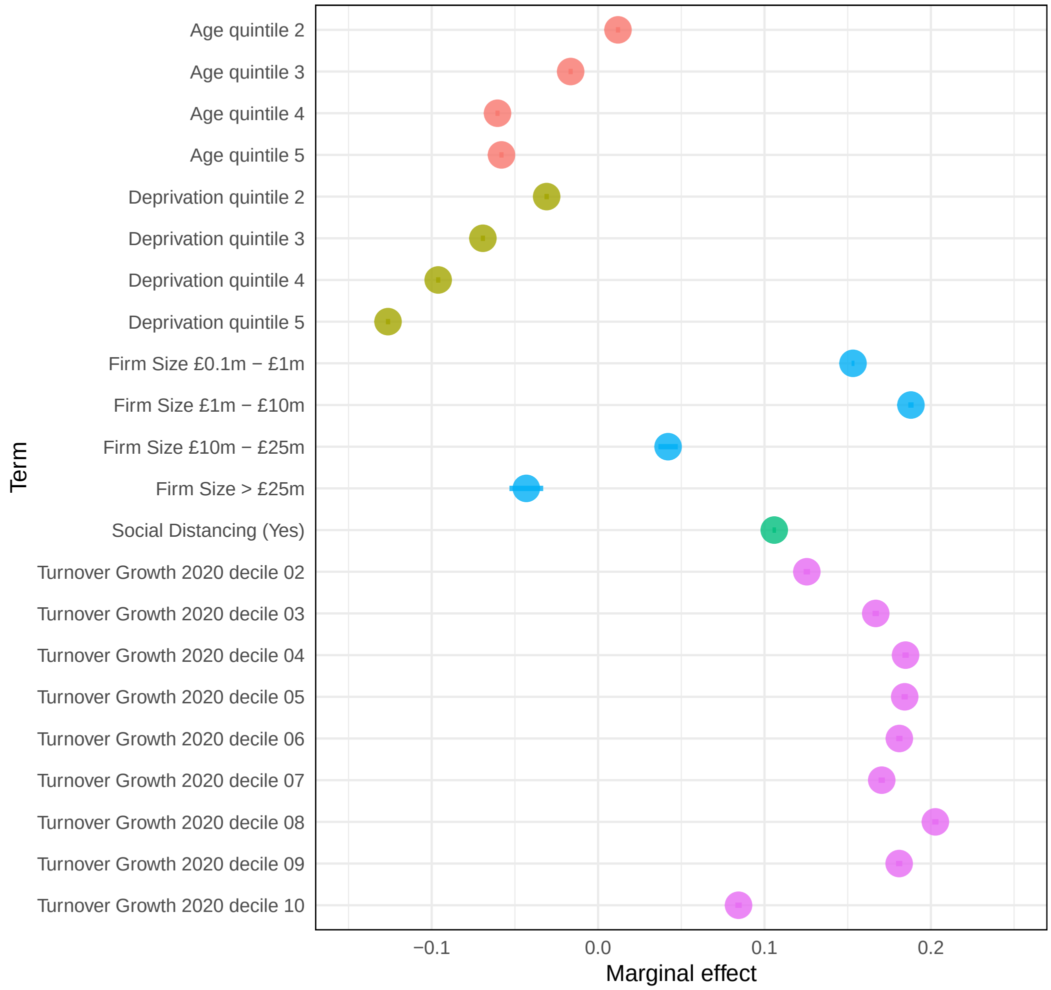

Chart shows the marginal effects of independent variables (colours).
Lines show 99% confidence intervals (standard errors are typically very small).
Base categories are: Age quintile 1; Deprivation quintile 1; Firm Size £0−£0.1m; Social Distancing (No); Turnover Growth 2020 decile 01.

Figure 16: Estimated marginal effects from BBLS probit model

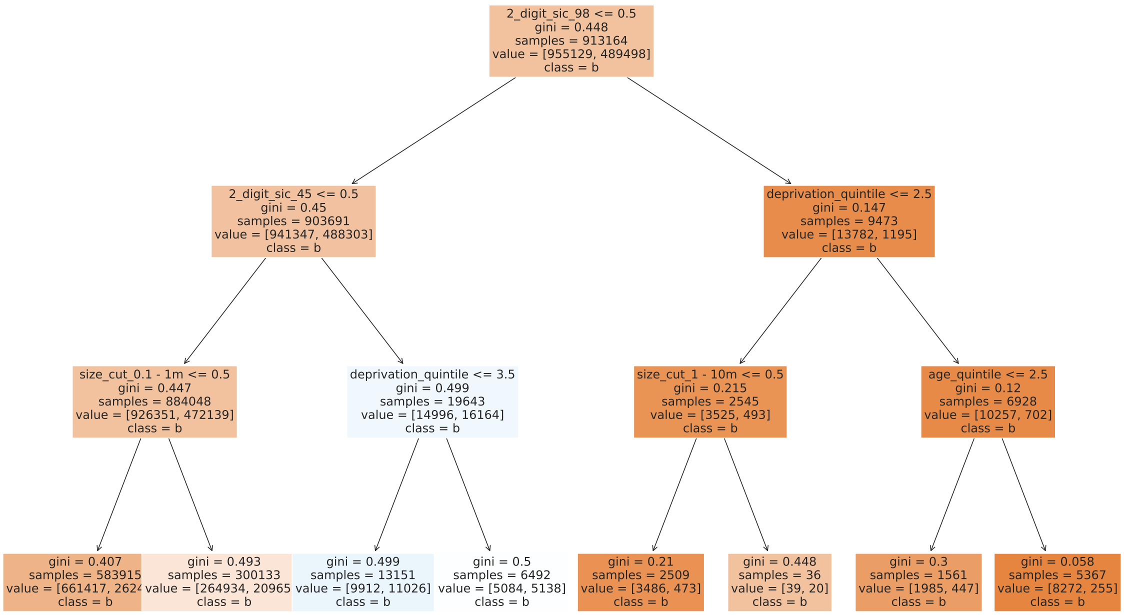

Figure 17: An example of a fitted decision tree

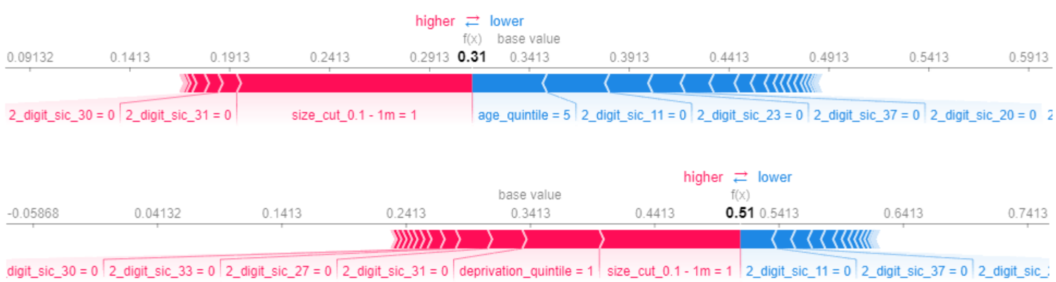

Figure 18: Shapley values for an individual dummy firm

This figure visualizes the importance of each variable using Shapley values. The total sum of the importance of all variables is 1.

B Data cleaning process

The initial data set we acquired from the credit reference agency consists of two files, covering transactions and accounts. The raw data set is stored in a very large CSV file, from which we produced tables to query subsets. Due to the size of the data set, we used Spark big data technology for all of our transformations of the data. The accounts file contains descriptive information on all accounts that companies have with all of their banks, while transactions contains monthly reports from all of the accounts within the accounts data set. The data in its raw form is relatively messy. Accounts are often reported on different days in the month, companies have multiple accounts with multiple banks and loan accounts and current accounts appear in the same data set.

We produced a standardized version of the data set in the form of an unbalanced firm-level panel, which required several transformations, including:

1. Identify all current accounts for limited companies.

2. Identify all debt accounts for limited companies.

3. Make a column signifying the total sum of loans taken out in each month. Aggregate by firm and month.

4. Aggregate current account flow and stocks information for the same firm where different banks have reported on exactly the same date.

5. Combine loan and current account flow information by month where not reported on exactly the same date, using the latest current account report after a firm takes out a loan.

6. Merge in additional information about the firm from Companies House data set.

7. Smooth flow variables by weighting according to the portion of the flow that came in each month. For example, a flow that covers 15 days in March and 15 days in April would be split in half and allocated to March and April.

8. Aggregate all variables by firm (so that firms with multiple accounts have one entry).

C Representativeness of data

We can use aggregate metrics to compare the coverage of the data to other sources of information on UK SMEs. Note that there is no commonly-shared definition of SMEs so these comparisons are all approximate. The Business Population Estimates (BPE) suggest that there are around 6.0 million SMEs with total turnover of £2.2 trillion. This is based on a definition that includes all businesses with fewer than 250 employees. The data set we use in the paper has high coverage of limited companies by number and slightly less than half by turnover as of end-2019. It has lower coverage of non-limited companies e.g. sole traders, which we exclude from our analysis in this paper.

Table 6: Representativeness, number of firms and turnover

	Aggregate data)			data set		
	Limited	Non-limited	Total	Limited	Non-limited	Total
Number (million)	2.0	4.0	6.0	1.98	0.960	2.94
Turnover (£ trillion)	1.7	0.5	2.2	0.78	0.084	0.864

UK Finance collect information from large banks on total SME deposits, defining SMEs based on a £25 million turnover threshold. It suggests that SME deposits totaled around £200 billion, with current account balances accounting for £117 billion of this. The data set we use in the paper covers around two thirds of total SME deposits in the end-2019 data.

Table 7: Representativeness, current account balances

	Aggregate data		data set		
	Current accounts	Total	Limited	Non-limited	Total
Cash (£ billion)	117	200	70	8	78

D Construction of lockdown variables

The national lockdown, tiers and local lockdown variables are based on the geographical location of firms and the point in time. The geographical location corresponds to the local authority district the policies were implemented in and the time dimension is monthly. We focused this part of the analysis on England because Scotland, Wales and Northern Ireland had different policies and the bulk of the firms in the data set are located in England.

- We constructed the national lockdown variable based on the dates of the first and second lockdowns according to government statements. We applied them to all local authorities in England for the months they were in force. They are shown in the red shaded areas in figure 19.

- The tier system was introduced for the months of October to December 2020 and different levels of restrictions applied at the local authority level. Figure 19 shows the proportion of the English population were under each tier at each point in time.

- We define 'targeted' local lockdowns as those that came before the tier system in October 2020. They include cities such as Leicester and Manchester, which had specific measures that were stricter than the rest of England at the time. Around 25 different local authorities faced 'targeted' local lockdowns. However, it is important to note that these local lockdowns differed in severity e.g. in Leicester non-essential shops were all closed but in Birmingham the measure only restricted the number of people that could meet.

Given the turnover data we used in our analysis was at monthly level, we used a simple approach to record lockdown measures by month. If more than half of the days in a given month came with lockdown restrictions, we coded that month to the lockdown variable for all firms in that area.

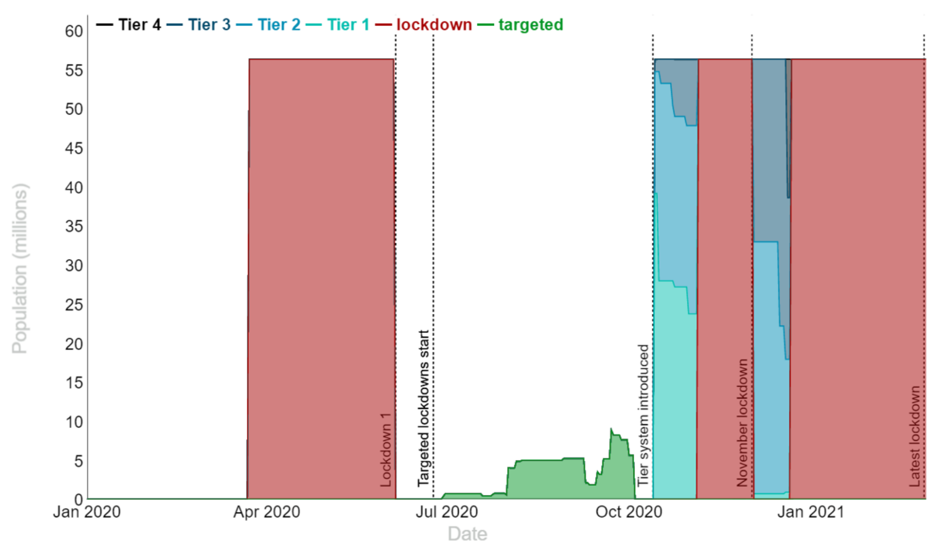

Figure 19: UK public health measures since January 2020

References

Viral V Acharya and Sascha Steffen. The risk of being a fallen angel and the corporate dash for cash in the midst of covid. (27601), July 2020. 4

Susan Athey and Guido W. Imbens. Machine learning methods that economists should know about. *Annual Review of Economics*, 11(1):685–725, 2019. 5

David Autor, David Cho, Leland D. Crane, Mita Goldar, Byron Lutz, Joshua Montes, William B. Peterman, David Ratner, Daniel Villar, and Ahu Yildirmaz. The risk of being a fallen angel and the corporate dash for cash in the midst of covid. July 2020. 4, 7

Saleem Bahaj, Angus Foulis, and Gabor Pinter. Home values and firm behavior. *American Economic Review*, 110(7):2225–70, July 2020. 7

Ryan Banerjee, Anamaria Illes, Enisse Kharroubi, and José María Serena Garralda. Covid-19 and corporate sector liquidity. (10), April 2020. 4

Alexander W. Bartik, Marianne Bertrand, Zoe Cullen, Edward L. Glaeser, Michael Luca, and Christopher Stanton. The impact of covid-19 on small business outcomes and expectations. *Proceedings of the National Academy of Sciences*, 117(30):17656–17666, 2020. ISSN 0027-8424. 4, 7

David Bholat, James Brookes, Chris Cai, Katy Grundy, and Jakob Lund. Sending firm messages: text mining letters from PRA supervisors to banks and building societies they regulate. (688), October 2017. 24

Nicholas Bloom, Robert S Fletcher, and Ethan Yeh. The Impact of COVID-19 on US Firms. (28314), January 2021. 4, 7

Kristina Bluwstein, Marcus Buckmann, Andreas Joseph, Miao Kang, Sujit Kapadia, and Ozgur Simsek. Credit growth, the yield curve and financial crisis prediction: evidence from a machine learning approach. (848), January 2020. 23

Catherine Buffington, Carrie Dennis, Emin Dinlersoz, Lucia Foster, and Shawn Klimek. Measuring the effect of covid-19 on u.s. small businesses: The small business pulse survey. 2020. 4, 7

Elena Carletti, Tommaso Oliviero, Marco Pagano, Loriana Pelizzon, and Marti G. Subrahmanyam. The COVID-19 Shock and Equity Shortfall: Firm-level Evidence from Italy. (566), May 2020. 5

Raj Chetty, John N Friedman, Nathaniel Hendren, Michael Stepner, and The Opportunity Insights Team. The economic impacts of covid-19: Evidence from a new public database built using private sector data. (27431), June 2020. 4, 7, 17

Gabriel Chodorow-Reich, Olivier Darmouni, Stephan Luck, and Matthew C Plosser. Bank liquidity provision across the firm size distribution. (27945), October 2020. 4, 7

S.J. Davis, J. Haltiwanger, and S. Schuh. Small business and job creation: Dissecting the myth and reassessing the facts. *Small Business Economics*, 8:297–315, 1996. 9

Sebastian Doerr, Leonardo Gambacorta, and José María Serena Garralda. Big data and machine learning in central banking. (930), March 2021. 25

Simen Gaure. lfe: Linear Group Fixed Effects. *The R Journal*, 5(2):104–116, 2013. 13

Pierre-Olivier Gourinchas, Sebnem Kalemli-Ozcan, Veronika Penciakova, and Nick Sander. Covid-19 and sme failures. (27877), September 2020. 5

Andreas Joseph. Parametric inference with universal function approximators. 2020. 25

Robert Joyce and Xiaowei Xu. Sector shutdowns during the coronavirus crisis: which workers are most exposed? (BN278), April 2020. 15

M. Pal. Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, 26(1):217–222, 2005. 23

Dimitris Papanikolaou and Lawrence D.W. Schmidt. Working remotely and the supply-side impact of covid-19. (27330), June 2020. 4

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12(85):2825–2830, 2011. 23

David MW Powers. Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation. *arXiv preprint arXiv:2010.16061*, 2020. 24

Fabiano Schivardi and Guido Romano. Liquidity crisis: Keeping firms afloat during covid-19. 2020. 4

Joel Suss and Henry Treitel. Predicting bank distress in the UK with machine learning. (831), October 2019. 23