BANK OF ENGLAND

Staff Working Paper No. 842

The empirics of granular origins: some challenges and solutions with an application to the UK

Nikola Dacic and Marko Melolinna

December 2019

BANK OF ENGLAND

Staff Working Paper No. 842

The empirics of granular origins: some challenges and solutions with an application to the UK

Nikola Dacic[1] and Marko Melolinna[2]

Abstract

We study the effects of firm-level microeconomic fluctuations on aggregate productivity in the United Kingdom. We show that a standard measure of residual productivity growth of the largest UK firms (the 'granular residual') produces results that are counter-intuitive and not statistically significant. To combat this, we introduce a unique production function approach to estimate firm-specific technology shocks, accounting for firm-level heterogeneity and common shocks. Using this measure, we find that firm-level shocks matter; the 'granular residual' explains around 30% of aggregate UK productivity dynamics. We also show that simplifications of our approach, which omit controlling for firm-level heterogeneity or do not account for common shocks, do not perform well, highlighting the importance of identifying firm-specific shocks correctly in order to properly test the 'granularity hypothesis'.

Key words: business cycle, aggregate volatility, granularity hypothesis, firm-level productivity.

JEL classification: E23, E24, E32.

(1) Goldman Sachs & Co. Email: n.dacic26@gmail.com
(2) Bank of England. Email: marko.melolinna@bankofengland.co.uk

We would like to thank Xavier Gabaix, John Moffat and Ricardo Reis for the useful discussions. We would also like to thank participants at the 2019 Scottish Economic Society Conference in Perth, UK, the 2019 EWIPA conference in London, UK, the 2019 North American Econometric Society Conference in Seattle, USA and the 2019 African Econometric Society Conference in Rabat, Morocco, for their useful comments. The views expressed in this paper are those of the authors, and not necessarily those of the Bank of England or its committees. This paper was written in Nikola Dacic's individual capacity and is not related to his role at Goldman Sachs. The analysis, content and conclusions set forth in this paper are those of the authors alone and not of Goldman Sachs & Co. or any of its affiliate companies. The authors alone are responsible for the content and all remaining errors. This work contains statistical data from ONS, which is Crown Copyright. The use of the ONS statistical data in this work does not imply the endorsement of the ONS in relation to the interpretation or analysis of the statistical data. This work uses research datasets, which may not exactly reproduce National Statistics aggregates. We remain solely responsible for the contents of this paper and any errors or omissions that could remain.

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 publications@bankofengland.co.uk

ISSN 1749-9135 (on-line)

1 Introduction

It has been widely noted that aggregate labour productivity has been low in the UK following the global financial crisis of 2007–2008 (see Figure 1). As of 2017, this shortfall amounts to productivity being about 20% below where it would have been had the average pre-crisis growth been maintained since 2007. The vast literature on the typical effects of financial crises on key macroeconomic variables (including productivity) suggests that the sharp fall in productivity during the peak years of the global financial crisis is in line with prior evidence from other countries that have experienced financial crises.[1] However, the slowdown in the *growth rate* of UK productivity post-2009 is puzzling, both in relation to historical evidence and the developments in other advanced economies during this period.

Various explanations of the weak productivity dynamics in the UK in the aftermath of the financial crisis have been proposed. Among others, those explanations include labour hoarding, misallocation of labour and capital, an elevated number of 'zombie firms', and increased firms' power in product and factor markets.[2] A unifying theme across all of these and other explanations is the question of whether the slowdown of productivity post-crisis has been cyclical or structural in nature. A wide range of approaches have been used, increasingly relying on firm- and industry-level data. Given the protracted slowdown, it is perhaps not surprising to see increasing evidence on the importance of structural factors.[3]

Instead of examining the nature of the slowdown in productivity (i.e. whether it is cyclical or structural), others have focused on *locating* the productivity (growth) puzzle. Tenreyro (2018) offers evidence pointing to the importance very few sectors in driving the growth puzzle, with finance and manufacturing being the largest contributors to the slowdown.[4] Using data on the universe of UK firms, Schneider (2018) argues that the same aggregate slowdown has been driven by shortfalls in the top quartile of the productivity distribution. However, whilst insightful per se, these conclusions do not point to the importance of *idiosyncratic* as opposed to *common shocks* in driving the slowdown.

Our focus in this paper is on establishing whether idiosyncratic shocks to the largest UK firms have had a significant effect on the dynamics of aggregate labour productivity in the UK. To our knowledge, there has not been a systematic treatment of the potential impact of firm-level idiosyncratic shocks and common shocks (e.g. technological or monetary policy shocks) may have had on aggregate productivity in the UK. This study thus aims to

[1] See, for example, Jorda, Schularick, and Taylor (2013), Oulton and Sebastia-Barriel (2017), and Cerra and Saxena (2008).

[2] On labour hoarding, see e.g. Barnett et al. (2014b) and Oulton and Sebastia-Barriel (2017). On 'zombie firms', see e.g. Carney (2017). On market power, see e.g. Haldane (2018).

[3] See, for example, Barnett et al. (2014a), Barnett et al. (2014b), Goodridge et al. (2014) and Riley et al. (2015).

[4] See also Mason et al. (2018), who reach a similar conclusion.

fill that gap and takes as its starting point the key insights from Gabaix (2011)—namely that aggregate fluctuations may arise due to shocks to the largest producers—as a way of examining the extent to which the dynamics of aggregate productivity in the UK can be explained by shocks to the largest firms.

The main premise in Gabaix (2011) can be summarised as follows: if the distribution of firm size is sufficiently skewed *and* firm-level volatility does not tend to vanish as firms scale up, then a small number of large firms may account for a disproportionally large fraction of aggregate fluctuations. The author then introduces the concept of a 'granular residual' (GR), which is a composite of idiosyncratic, firm-level shocks to a selected number of largest firms in the economy. Gabaix finds that the GR is a significant driver of aggregate GDP and productivity dynamics in the US. Other papers have since replicated the analysis for other economies, with generally similar findings.[5] We are aware of one which includes the UK (see Lin and Perez (2014)); they do not find the GR to be relevant for aggregate GDP fluctuations in the UK.

The identification of firm-level idiosyncratic shocks in the benchmark version of the GR from Gabaix (2011) is arguably rather crude; the firm-level shocks correspond to de-meaned firm-level productivity growth rates. In addition, the GR formula in Gabaix (2011) is a consequence of Hulten's theorem. Baqaee and Farhi (2019) show that Hulten's theorem—namely, that the aggregate impact of microeconomic TFP shocks is proportional to the Domar weight of the producer—is globally accurate only in an efficient Cobb-Douglas economy. The existence of distortive frictions in the economy and deviations from unitary elasticities of substitution (which characterise a Cobb-Douglas economy) may reduce the accuracy with which the GR approximates (even locally) the aggregate effect of microeconomic TFP shocks.

We therefore try to improve on the benchmark version of the GR by identifying firm-level idiosyncratic shocks using a modification of the approach in De Loecker and Warzynski (2012). More specifically, we allow for the possible existence of common shocks affecting multiple producers and we control for the unobserved, firm-specific differences in technology that do not vary over time. In addition, we consider a non-parametric second-order approximation of the impact of microeconomic TFP shocks, following Baqaee and Farhi (2019), in order to generalise the GR beyond a Cobb-Douglas efficient economy.

Our contributions to the literature are the following. Our main contribution is to introduce a unique way of estimating firm-level idiosyncratic TFP shocks taking into account firm heterogeneity and the existence of common shocks. In terms of our empirical

[5] Di Giovanni et al. (2014) and Friberg and Sanctuary (2016) find significant effects of firm-level shocks for France and Sweden, respectively, while Gnocato and Rondinelli (2018) find that idiosyncratic TFP shocks to large firms explain around 30% of aggregate TFP volatility in Italy. However, a recent paper by Gutierrez and Philippon (2019) finds that the contribution of shocks to the largest firms has become much smaller in the last decade compared to preceding decades.

results, first, we find that the distribution of firm size in the UK is of a power-law type, rendering the GR potentially relevant in explaining aggregate productivity dynamics. Second, we find that the benchmark GR (identical to that in Gabaix (2011)) has some explanatory power for aggregate productivity dynamics, when one excludes the financial crisis period, but over the whole sample, the results are insignificant and partly counter-intuitive. Third, and most importantly, when we introduce the control function approach of De Loecker and Warzynski (2012), allowing for time-invariant technological differences across firms as well as common shocks, we find that the GR explains around 30% of UK productivity dynamics over the past three decades. The approximate second-order effects of Baqaee and Farhi (2019) cause an improvement in explanatory power of this model.

Outline The rest of the paper is organised as follows. Section 2 motivates the 'granularity hypothesis' by considering the contribution of the largest firms to UK's labour productivity growth in the decomposition of Melitz and Polanec (2015). Section 3 sets up the theoretical framework both for the GR and the control function approaches. Section 4 analyses the importance of the GR in explaining the dynamics of UK aggregate productivity. Section 5 concludes.

2 Granularity in a Productivity Decomposition

In this section, we use a very popular decomposition of aggregate labour productivity proposed by Melitz and Polanec (2015) to investigate the extent to which the largest firms' contributions matter.[6] We view this exercise as motivating our analysis in the next section.

Melitz and Polanec (2015) define aggregate productivity at time t as:

$$\Phi_t = \sum_i s_{it}\phi_{it}, \tag{1}$$

where the employment shares $s_{it} \geq 0$ sum to 1 and ϕ_{it} denotes the log of firm i's productivity at time t. They derive a decomposition of aggregate productivity growth in terms of the contributions of three groups of firms: survivors, entrants, and exiters. In particular, they define aggregate productivity growth as:

$$\Delta\Phi_t = \underbrace{\left(\Phi_t^S - \Phi_{t-1}^S\right)}_{\text{survivors}} + \underbrace{s_t^E\left(\Phi_t^E - \Phi_t^S\right)}_{\text{entrants}} + \underbrace{s_{t-1}^X\left(\Phi_{t-1}^S - \Phi_{t-1}^X\right)}_{\text{exiters}}, \tag{2}$$

[6] Another popular decomposition is that of aggregate TFP proposed by Levinsohn and Petrin (2012). We could not implement this decomposition as we do not have the data required for this exercise readily available in the dataset we use in this paper.

where $s_t^G = \sum_{i \in G} s_{it}$ denotes the aggregate employment share of a group G of firms and $\Phi_t^G = \sum_{i \in G}(s_{it}/s_t^G)\phi_{it}$ is that group's aggregate (average) productivity. The first term corresponds to the contribution to aggregate productivity growth arising from the changing employment shares and/or productivity of the *surviving* firms. It can be rewritten as:

$$\underbrace{\Phi_t^S - \Phi_{t-1}^S}_{\text{survivors}} = \sum_{i \in S} \left(\frac{s_{it}}{s_t^S} \phi_{it} - \frac{s_{i,t-1}}{s_{t-1}^S} \phi_{i,t-1} \right). \tag{3}$$

Given that our focus is on the largest firms and that our available sample for this part of the analysis is relatively short (2004–2014), it seems reasonable to focus on the largest firms' contribution to this term only.[7] In each period t, we will consider the largest 100 surviving firms (where size is measured by turnover) and compute their total contribution to aggregate productivity growth, using equation (3). Figure 2 shows the contribution from the largest 100 surviving firms to aggregate labour productivity growth in the UK over 2004–2014. The contribution tracks the aggregate growth rate very closely and is sizeable. Interestingly, it suggests that the largest firms experienced a downturn during 2008 (due to their labour weight and/or their productivity falling), whereas aggregate productivity growth reached a trough a year later.

The Melitz-Polanec (2015) decomposition is a useful framework to account for the dynamics of aggregate productivity growth using individual firms' contributions. Importantly, those contributions may reflect various shocks that transmit through the economy (for instance, via the production network), which may both be common and/or idiosyncratic. A priori, it is therefore unclear whether the contributions of large firms in the Melitz-Polanec (2015) decomposition reflect their own idiosyncratic shocks, or idiosyncratic shocks to other firms and/or common shocks. The granularity hypothesis states that it is precisely the idiosyncratic shocks to the largest firms that will have non-trivial aggregate effects. Therefore, to empirically test this hypothesis, in the next section we aim to identify firm-specific shocks by estimating production functions, allowing for the potential existence of common shocks, assuming only that firms are cost minimizers and imposing no restrictions on the type of competition in factor and output markets.[8]

[7] Appendix A provides a description of the data used in the analysis throughout this paper.

[8] The Levinsohn-Petrin (2012) decomposition of aggregate TFP growth, for example, makes several simplifying assumptions in order to identify such shocks, namely that the production functions have constant returns to scale and are separable in intermediate inputs and value-added, and that output and input markets are perfectly competitive, in addition to the assumed cost-minimising behaviour on the part of firms. The approach we use does not rely on these assumptions.

3 Granular Residual

This section first lays out the theoretical foundations of the GR in the first subsection, as defined by Gabaix (2011). In the second subsection, we aim to generalise and improve the original version of the GR from Gabaix (2011) by, first, taking into account the second-order effects of shocks, and second, by identifying firm-specific TFP shocks using the control function approach, allowing for the existence of common shocks and firm fixed effects. Finally, we introduce the set of models that we take to the data in order to empirically test the granularity hypothesis for the UK.

3.1 Benchmark Models

3.1.1 First-Order Granular Residual

Gabaix (2011) derives the theoretical requirements for the idiosyncratic TFP shocks to the largest firms to be macroeconomically significant. For this to be possible, the distribution of firm size needs to follow a power law distribution (Zipf's distribution). In other words, there needs to be a long tail of large firms. In an economy with a fat-tailed distribution of firms, aggregate volatility decays at a much slower rate as the number of firms in the economy increases.

If these theoretical requirements are met, idiosyncratic shocks to the largest producers could have a signficant effect on aggregate productivity dynamics in an economy. To investigate the effects of such granular shocks on aggregate activity, one can use of Hulten's theorem, defining the GR as a Domar-weighted composite of shocks:

$$GR_t = \sum_{i=1}^{K} \frac{S_{i,t-1}}{Y_{t-1}} Shock_{it}, \tag{4}$$

where S_{it} denotes firm i's sales (turnover) in period t, Y_t is nominal GDP at time t, K denotes the number of largest firms selected and $Shock_{it}$ denotes an idiosyncratic technology shock to firm i in period t.

Gabaix's (2011) empirical implementation replaces $Shock_{it}$ with de-meaned productivity growth:

$$GR_t^B = \sum_{i=1}^{K} \frac{S_{i,t-1}}{Y_{t-1}} (g_{it} - \overline{g}_t^K), \tag{5}$$

where g_{it} denotes firm-level productivity growth from year $t-1$ to t (where productivity is defined as turnover divided by the number of employees) and $\overline{g}_t^K$ is the average productivity growth of the largest K firms. The term $(g_{it} - \overline{g}_t)$, denoting de-meaned firm-level productivity growth, is arguably the simplest measure of an idiosyncratic shock to firm i.[9]

As was pointed out earlier, the above version of the granular residual relies on Hulten's theorem. Up to a first-order approximation, Hulten's theorem (by in turn resorting to the envelope theorem) implies that the sufficient statistic for the impact of a technology shock to producer i is its size, as measured by its Domar weight (equal to its sales divided by GDP). This powerful result implies that—as long as the steady state is efficient—it does not matter whether or not factors or inputs are reallocated in response to a shock, or what structure the input-output network takes.[10] In the framework of Gabaix (2011), the economy is efficient so Hulten's theorem applies and the growth of aggregate TFP is given by the sales-share weighted average of idiosyncratic technology shocks.

Importantly, by identifying firm-specific shocks using de-meaned productivity growth, $g_{it} - \overline{g}_t$, we now demonstrate that it is to be expected that the resulting composite of 'shocks' will have significant predictive power for aggregate productivity growth. To see why, note that g_{it} equals $\Delta\phi_{it}$ in the Melitz-Polanec (2015) decomposition of aggregate productivity. We can thus rewrite equation (5) as:

$$GR_t^B = \sum_{i=1}^{K} \frac{S_{i,t-1}}{Y_{t-1}} \Delta\phi_{it} - \overline{\Delta\phi}_t^K \sum_{i=1}^{K} \frac{S_{i,t-1}}{Y_{t-1}}. \tag{6}$$

Taking differences on both sides of equation (1) yields:

$$\Delta\Phi_t = \left[\sum_{i\in K} s_{i,t-1}\Delta\phi_{it} + \sum_{i\in K} \Delta s_{it}\phi_{it}\right] + \underbrace{\left[\sum_{i\notin K} s_{i,t-1}\Delta\phi_{it} + \sum_{i\notin K} \Delta s_{it}\phi_{it}\right]}_{=\Delta\Phi_t^{i\notin K}}. \tag{7}$$

The first two terms on the right-hand side in equations (6) and (7) will be quantitatively similar if amongst the set of K largest firms $\frac{S_{i,t-1}}{Y_{t-1}} \approx s_{i,t-1}$, i.e. if their Domar weights are approximately equal to their shares of total employment. Figure 3 suggests that this is indeed the case in the UK. The second term within the first square bracket in equation (7) will be relatively small if the largest firms' employment shares do not change significantly year-on-year.

To empirically demonstrate this point more rigorously, we compute a version of the

[9] Gabaix (2011) also considers an alternative de-meaning procedure, where the average productivity is computed within the industry a given firm belongs to. This yields results that are very similar to those obtained under his benchmark approach, which we thus focus on here.

[10] See Baqaee and Farhi (2019) for a more detailed discussion.

GR with employment weights in place of Domar weights:

$$GR_t^E = \sum_{i=1}^{K} s_{i,t-1} \underbrace{(g_{it} - \overline{g}_t)}_{=\Delta\phi_{it} - \overline{\Delta\phi}_t}, \tag{8}$$

The only difference between GR_t^B and GR_t^E is in the use of Domar weights in the former and employment shares in the latter, which tend to be strongly positively correlated. Note that GR_t^E is, by definition, equal to $\Delta\Phi_t - \overline{\Delta\phi}_t \sum_{i=1}^{K} s_{i,t-1} - \sum_{i=1}^{K} \Delta s_{it}\phi_{it} - \Delta\Phi_t^{i\notin K}$, so GR_t^E is very likely to be correlated with $\Delta\Phi_t$. Therefore, by construction, testing the granularity hypothesis in a manner that involves identifying firm-level shocks crudely as de-meaned productivity growth rates, as in Gabaix (2011), is likely to mechanically provide evidence in favour of the granularity hypothesis.

Following the approach in Gabaix (2011), many papers have subsequently tested the granularity hypothesis by regressing GDP growth and, commonly, aggregate TFP growth, on the granular residual GR_t^B (typically including its first two lags as well). First, note that as we argued above, since Domar weights and employment shares tend to be strongly positively correlated, the benchmark granular residual (GR_t^B) will tend to be correlated with aggregate labour productivity growth derived bottom-up using the Melitz-Polanec decomposition ($\Delta\Phi_t$). Second, aggregate labour productivity growth based on the Melitz-Polanec decomposition will in turn be strongly correlated with aggregate labour productivity growth based on the official aggregate data. Third, aggregate labour productivity growth will tend to be strongly correlated with both aggregate TFP growth and GDP growth.[11] For instance, the correlation coefficient between aggregate labour productivity growth and TFP (GDP) growth in the UK over 1970-2016 was 0.96 (0.76).[12]

3.1.2 Second-Order Granular Residual

Baqaee and Farhi (2019) show that Hulten's theorem can in practice be very fragile due to significant non-linearities in how shocks are mapped to output in the presence of *general* input-output networks.[13] In other words, the full (non-linear) impact of idiosyncratic shocks can be quite different from the first-order approximation given by Hulten's theorem. For this reason, in addition to the benchmark ('first order') version of the GR models, we

[11] Assuming that the aggregation production function is Cobb-Douglas and that there is constant factor utilisation, real GDP growth is given by $\Delta y_t = \Delta TFP_t + \alpha\Delta k_t + (1-\alpha)\Delta l_t$ where α denotes the labour share of total income, and y_t, k_t, and l_t denote (the log of) real GDP, aggregate capital, and aggregate labour. Aggregate labour productivity growth is given by $\Delta y_t - \Delta l_t = \Delta TFP_t + \alpha(\Delta k_t - \Delta l_t)$. It is thus to be expected that real GDP growth, TFP growth, and aggregate labour productivity growth will be positively correlated.

[12] Focusing on the pre-crisis period only from 1970-2006 does not significantly alter the correlation coefficients; correlation between aggregate labour productivity growth and TFP (GDP) growth is 0.95 (0.69).

[13] See Section 7 in Baqaee and Farhi (2019).

will also consider a non-parametric approach proposed by Baqaee and Farhi (2018) which allows us to account for the second-order macroeconomic impact of microeconomic shocks. Denote by $GR_t^{2^{nd}}$ the corresponding (non-parametric) second-order granular residual, given by:

$$GR_t^{2^{nd}} = \sum_{i=1}^{K} \frac{1}{2} \left[\frac{S_{i,t-1}}{Y_{t-1}} + \frac{S_{i,t}}{Y_t} \right] Shock_{it}. \tag{9}$$

This second-order approximation aims to non-parametrically capture the idea that Domar weights themselves change in response to shocks. For example, a positive $Shock_{it}$ could be expected to result in a larger Domar weight for firm i.

3.2 Identifying Firm-Level Idiosyncratic Shocks

Fundamentally, regardless of the extent to which the conditions of Hulten's theorem are satisfied, empirically testing the granularity hypothesis requires estimating the firm-level idiosyncratic TFP shocks. In Gabaix (2011), the identification of these shocks is rather crude. Instead, we consider a modification of the well-established control function method for estimating production functions. In particular, we consider a setting in which there may exist input-output linkages between firms as well as common shocks, whilst also allowing for a time-invariant component characterising a given firm's technology.

Although Hulten's theorem does not require that $Shock_{it}$ contains no components that are common across producers (given the linearity in shocks), we will seek to identify firms' idiosyncratic shocks that have been purged from any common components. We do so for two reasons. First, the theoretical underpinnings of Gabaix's (2011) granularity hypothesis assume that firms' shocks are uncorrelated random variables. In the empirical testing of the granularity hypothesis, Gabaix's (2011) aims to provide evidence that the agruably idiosyncratic changes in the largest 100 firms' productivity growth explain an important fraction (one-third) of the movement of GDP growth, hence our focus on the idiosyncratic shocks themselves. Second, if the statistical processes describing firms' technology contain common shocks *and* if those common shocks are correlated over time—the latter being an assumption routinely made in the context of factor models—then the standard control-function approach to estimating production functions (and thus identifying shocks to firms' production) is inconsistent. Instead, we aim to identify firms' idiosyncratic shocks by allowing for common shocks which are potentially correlated over time.

To begin with and for simplicity, assume that the gross output production function of each producer i belonging to sector s takes the standard Cobb-Douglas form, which in its logarithmic form is given by:

$$y_{it} = \alpha^s v_{it} + \beta^s k_{it} + \omega_{it} + \epsilon_{it}, \tag{10}$$

where the parameters of the production function are specific to the sector s firm i belongs to, and where v_{it} denotes firm i's cost of goods sold and k_{it} denotes its net capital stock. In our data, we do not observe firms' purchases of intermediate inputs, which are usually part of their cost of goods sold. Since the cost of goods sold also includes payments to labour, we subsume the material and labour inputs—which we both treat as variable—into cost of goods sold.[14] We do not impose the assumption that the production function has constant returns to scale. Following De Loecker and Warzynski (2012), we assume in equation (10) that the overall level of factor-neutral technology is the sum of the producer's productivity, ω_{it}, and unanticipated shocks to production and i.i.d. shocks including measurement error, ϵ_{it}.

Conditional on the production function given by equation (10) being the true functional form, it has long been recognised that a simple OLS estimation of it will produce inconsistent estimates of (α^s, β^s) because of omitted variable bias: ω_{it} is unobserved, and optimal input choices are likely to be correlated with it. To overcome this source of endogeneity, the control function approach relies on the existence an invertible function $h(\cdot)$ such that $\omega_{it} = h(v_{it}, k_{it})$.[15].

A very popular reference in the literature on estimating production functions using the control function approach is De Loecker and Warzynski (2012), who assume that the process for TFP (ω_{it}) is of the following AR(1) form:

$$\omega_{it} = \rho^s \omega_{i,t-1} + u_{it}, \tag{11}$$

where ρ^s is a parameter to be estimated and the error term (u_{it}) is i.i.d..

As in equation (11), one implicit assumption that is typically made in the control-function approach is that there are no firm-specific, time-invariant elements of the technology process, ω_{it}. If this assumption is violated, then not controlling for these firm-specific fixed effects will tend to result in first-stage residuals (we describe the first stage of the control function approach below), which are potentially highly persistent and not zero on average. In addition, if these firm-specific effects are not explicitly controlled for in the second stage of the procedure, the resulting GMM conditions will be invalid and will thus yield inconsistent estimates.[16]

[14] This assumption has been commonly made in the recent literature on estimating production functions.

[15] See De Loecker and Warzynski (2012) or Ackerberg, Caves, and Frazer (2015).

[16] Lee, Stoyanov and Zubanov (2019) elaborate on this point more thoroughly. They find that controlling for firm-specific fixed effects outperforms the standard control function estimation approach in terms of capturing persistent unobserved heterogeneity in firm productivity.

Another implicit assumption that the control function approach makes is that there are no components of technology, ω_{it} that are common across producers, which is equivalent to assuming that ω_{it} are independent across i. if this assumption is satisfied and there are no firm-specific time-invariant components of ω_{it}, one can in principle consistently estimate the parameters in equation (10) using the generalised method of moments (GMM). However, in the presence of common shocks which are correlated over time, this approach will also be inconsistent.

We thus generalise the control function approach by relaxing the assumptions that there are no time-invariant firm-specific components of firms' technology or common shocks correlated over time. More specifically, suppose that technology ω_{it} follows an AR(1) process of the following form:

$$\omega_{it} = \rho^s \omega_{i,t-1} + \mu_t + c_i + u_{it}, \tag{12}$$

where the innovation consists of a common (μ_t) and an idiosyncratic i.i.d. component (u_{it}). We allow for the existence of non-zero firm-specific fixed effects, c_i. We assume that the common innovation (μ_t) is uncorrelated with the firm-specific innovation (u_{it}), and that it is potentially correlated over time.[17] These two assumptions are standard in the literature on factor models. As our aim is to elicit the firm-specific shocks, obtaining an estimate of the term u_{it} is what we are ultimately interested in.

Note that if $c_i \neq 0$, an OLS regression of equation (12) that does not include firm-specific dummies will yield inconsistent estimates, as $\omega_{i,t-1}$ will be correlated with the firm-specific fixed effect (c_i). In particular, it will tend to be biased upwards, thus attenuating our estimate of the firm-specific idiosyncratic shock. To allow for this possibility, we will allow for firm-specific intercepts in estimating equation (12).

We perform the *first stage* of the control function approach as proposed by De Loecker and Warzynski (2012), and first run a regression of y_{it} on v_{it}, k_{it} and their interactions up to a given order (in our case, second). Importantly, we also allow for firm-specific intercepts, thus obtaining estimates of c_i. This polynomial, which approximates the function $\omega_{it} = h(v_{it}, k_{it})$, yields an estimate of expected output (from the predicted value from this regression), $\hat{y}_{it}$, as well as an estimate of unanticipated shocks to production

[17]In principle , there is no reason why the loading in the μ_t needs to be unitary. For example, Gabaix and Koijen (2019) consider a more general specification in a factor model set-up. They assume that the loading on μ_{it} is a function of known firm-specific obervarbles, namely the firm's size. This approach is challenging to implement for our setup due to the relatively small sample size. We do, however, consider a case where we allow for a heterogeneous effect of the common shock across firms within a given sector depending on the firms' level of market capitalisation. More specifically, we set a threshold dummy to equal 1 if firm i had above-median market capitalisation within its sector at time $t-1$. Hence, the common shock may have an additional effect on firms that have relatively high market capitalisation. The method is rather *adhoc*, which is why we do not present the model here, but the results are similar, or even slightly better in terms of their significance than those for the best model we present below.

(and possibly also measurement error), $\hat{\epsilon}_{it}$, which is the residual from this regression.

In the *second stage*, we view technology ω_{it} as a function of the (yet) unknown parameters $\boldsymbol{\theta}^s$:

$$\omega_{it}(\boldsymbol{\theta}^s) = \hat{y}_{it} - f(v_{it}, k_{it}; \boldsymbol{\theta}^s), \tag{13}$$

where $\boldsymbol{\theta}^s \equiv (\alpha^s, \beta^s)$. Conditional on $\mu_t = 0$ and $c_i = 0$, i.e. there being no common shocks of firm-specific fixed effects, we can recover the firm-specific innovation to productivity given $\boldsymbol{\theta}^s$, $\xi_{it}(\boldsymbol{\theta}^s)$, by regressing $\omega_{it}(\boldsymbol{\theta}^s)$ on its lag, $\omega_{i,t-1}(\boldsymbol{\theta}^s)$. The standard control function approach would then amount to searching for those values of $\boldsymbol{\theta}^s$ that minimise the sample counterparts of the GMM conditions given by:

$$\mathbb{E}\left[\xi_{it}(\boldsymbol{\theta}^s)\begin{bmatrix} k_{it} \\ v_{i,t-1} \end{bmatrix}\right] = 0, \tag{14}$$

where k_{it} is assumed to have been chosen prior to period t and is thus a valid instrument, and labour and materials are lagged since they are more flexible and may respond contemporaneously to TFP shocks.

However, if the common shock μ_t is serially correlated, then the GMM moment conditions given in equation (14) are no longer valid: $\xi_{it}(\boldsymbol{\theta}^s)$ contains the serially correlated common shock, which makes it potentially correlated with all instruments. We thus need to purge $\xi_{it}(\boldsymbol{\theta}^s)$ from the common shock. Since we run the estimations on samples consisting of firm-year observations belonging to a given sector, we can control for the common shock (given our assumption of unitary loadings) by including a dummy variable for each year. Importantly, recall that we have purged ω_{it} from the firm-specific, time-invariant component in the first stage by including firm-specific intercepts. We then end up with the following set of GMM conditions, which are valid in the presence of common shocks, correlated over time, and firm fixed effects:

$$\mathbb{E}\left[\nu_{it}(\boldsymbol{\theta}^s)\begin{bmatrix} k_{it} \\ v_{i,t-1} \end{bmatrix}\right] = 0. \tag{15}$$

The second stage thus results in the estimated parameter values, $(\hat{\alpha}^s, \hat{\beta}^s)$, so we can obtain our measure of firm i's technology from:

$$\hat{\omega}_{it} = y_{it} - \hat{\epsilon}_{it} - \hat{\alpha}^s v_{it} - \hat{\beta}^s k_{it}, \tag{16}$$

where $\hat{\epsilon}_{it}$ comes from the first stage. For all firms in each sector s, we then regress $\hat{\omega}_{it}$ on its lag and a set of dummy variables for each year and each firm. Since the growth rate of

TFP of firm i belonging to sector s is given by:

$$\Delta\omega_{it} = (\rho^s - 1)\omega_{i,t-1} + \mu_t + u_{it}, \tag{17}$$

we finally end up with an estimate of the idiosyncratic shock to firm i's TFP, $\hat{u}_{it}$. Recall that the first-stage resulted in an estimate of ϵ_{it}, which denotes unanticipated shocks to production (and potential measurement error in firm i's output). Therefore, the combined firm-specific technology shock to production at time t is given by $\Delta u_{it} + \Delta\epsilon_{it}$, where we obtain the former term from the second stage and the latter term from the first stage of our procedure:

$$Shock_{it} = \Delta\hat{u}_{it} + \Delta\hat{\epsilon}_{it}. \tag{18}$$

We implement our approach by running the estimation separately on each of the S samples (where S is the number of sectors) consisting of all firm-year observations in a given sector.

3.3 GR models

We now summarise the above discussion by setting up five models that we use in the next section to estimate a version of the GR and test their significance for explaining aggregate productivity dynamics in the UK. Table 1 summarises our models. We use both a first-order and a second-order order versions (as per equation (9)) for all the models in the empirical investigation below.

The first model (Model 1) is the benchmark model, based on Gabaix's (2011) definition of firm-level shocks:

$$GR_t^B = \sum_{i=1}^{K} \frac{S_{i,t-1}}{Y_{t-1}}(g_{it} - \overline{g}_t^K), \tag{19}$$

where $K = 100$.

Model 2 is similar to Model 1, except that we use labour weights rather than GDP weights to investigate the extent to which the benchmark GR is likely to provide support for the granularity hypothesis *by construction*:

$$GR_t^E = \sum_{i=1}^{K} s_{i,t-1}(g_{it} - \overline{g}_t^K), \tag{20}$$

where $s_{i,t-1}$ are employment shares of the firms.

Models 3-5 are based on the control function (CF) approach. The GR is given by:

$$GR_t^{CF} = \sum_{i=1}^{K} \frac{S_{i,t-1}}{Y_{t-1}} Shock_{it}, \tag{21}$$

where the identification of $Shock_{it}$ is most restrictive in Model 3. For Model 3, we use the same definitions as in De Loecker and Warzynski (2012), with no firm fixed effects or common shocks. In Model 4, we allow for firm fixed effects bu no common shocks, and finally, in Model 5, we allow for firm fixed effects as well as common shocks (with unitary loading, as discussed above).

Table 1. Different variants of the granular residual

Name	Measure of Idiosyncratic Shocks
1. Benchmark (GR_t^B)	De-meaned firm productivity growth ($g_{it} - \overline{g}_t^K$)
2. Labour weights (GR_t^E)	De-meaned firm productivity growth ($g_{it} - \overline{g}_t^K$)
3. No firm fixed-effects ($GR_t^{\text{no firm-FE}}$)	$Shock_{it}$ given by eq. (18), $c_i = \mu_t = 0$ in eq. (12)
4. No common shocks ($GR_t^{\text{no common}}$)	$Shock_{it}$ given by eq. (18) and $\mu_t = 0$ in eq. (12)
5. Unitary loadings (GR_t^{unitary})	$Shock_{it}$ given by eq. (18) and allowing for c_i and μ_t in eq. (12)

Notes: The first-order versions of the four granular residuals involve the use of lagged Domar weights, whilst the second-order versions involve the use of an equally-weighted average of lagged and contemporaneous Domar weights.

4 Results for UK productivity

4.1 Main Results

Before estimating the GR with the five models detailed above, we first note that the population of UK firms does indeed follow a distribution very close to a power law.[18] Figure 4 shows the histogram of UK firms in the latest data (referring to financial year 2015-16) on a log-log scale. For the Zipf's distribution, the linear slope of the data plotted on a log-log scale should be close to -2. The slope in our case (-1.93) is very close to this and hence, it is reasonable to assume that the population of UK firms does follow the Zipf's distribution.[19]

[18] Based on the ONS Business Structure Database, which provides an annual snapshot of around 2-2.5 million UK firms.

[19] This finding is also corroborated by the so-called sales herfindahls, which measure the variation of firm-level turnover. We find very similar values (5-8%) for the UK over the past 20 years as Gabaix (2011) found for the US, which suggests that the distribution of UK firms is sufficiently skewed for the large firms to have enough volatility to be significant at the aggregate level.

As the UK firm population appears to have favourable features required for the firm-level shocks to potentially non-trivially affect aggregate productivity, we proceed by calculating the GR for the UK based on the models listed in Table 1, focusing on the 100 largest firms from 1988 to 2016.[20] Similarly to Gabaix (2011), we exclude the oil and finance sectors from our analysis.[21]

First, we note that the ratio of the largest 100 UK firms' sales to UK GDP is large (Figure 5). It is around 20-50% of UK GDP, depending on the particular year and whether we focus on firms' total or domestic sales. Gabaix (2011) obtained broadly similar findings for the US, although the slight downward trend in the share of the largest firms appears to be specific to the UK. Using total turnover gives by far the highest share. However, since we are interested in the GR contributions associated with domestic production, in our main models from Table 1 we exclude international sales[22]

Across our five models summarised in Table 1, we have two largely overlapping but slightly different samples, based on data coverage and availability. In the first sample, which we refer to as the 'GR sample', we include all firms for which data required to calculate labour productivity (i.e., turnover and employment) is available. In the second sample referred to as the 'CF Sample', we use firms for which the data on the variables used in our CF approach are available (i.e., turnover, cost of goods sold and capital). The coverage of the samples is nearly identical, although for the CF sample, the time span is shorter at the beginning due to the estimation requirements of the two-stage CF approach.[23]

Figure 6 shows the resulting GR for the UK associated with selected models from Table 1, namely the benchmark GR measure (model 1), as well as the control function (CF) approach with firm fixed effects and commons shocks (Model 5) as described above. Figure 7 shows the analogous series for the respective second-order versions of the GR. All the GR series are relatively volatile and there are large fluctuations around the financial crisis; for the CF model, these fluctuations appear to correlate more strongly with aggregate productivity growth dynamics than the benchmark GR model. The GR and the CF series in the charts look dissimilar, and the differences are due to the different measures

[20] We use the Thomson Reuters Worldscope dataset for this analysis. See Appendix A for details.

[21] More specifically, we exclude SIC07 section K ('Financial and insurance activities') and division 19 ('Manufacture of coke and refined petroleum products'). For section K, the exclusion is due to the fact that GR shocks are not a relevant concept for the financial account statement of financial firms, and for division 19, we want to avoid the fluctuations of global oil prices affecting our measures of GR shocks. In addition to this, Gabaix (2011) also excluded the energy sector more generally. However, we do not deem this to be necessary for the UK case, as the shocks to the largest firms in this sector are not positively correlated with global energy prices, and these firms generally have a strong domestic focus.

[22] This is in contrast to Gabaix (2011), who uses total turnover due to data availability issues. Gutierrez and Philippon (2019) conduct a GR exercise on US data using domestic US sales only.

[23] We have also carried out the analysis using the same sample for all the models, i.e., a somewhat smaller dataset for which all the required data is available. The results are not presented here, but they are very similar to the main results presented below.

of firm-specific shocks, $Shock_{it}$. The differences between the first- and second-order GR measures are generally small.

Next, in line with Gabaix (2011), we examine the significance of the GR for explaining aggregate productivity dynamics in a simple OLS regression. In Table A.1 (A.2), we report the results for selected regressions involving the first-order (second-order) GR.[24] We start with the benchmark GR model (Model 1). The results are not very encouraging; only the second lag is statistically significant, and the adjusted R-squared is very low at below 0.1. When we move to using labour weights (Model 2), the results look even worse, with a lower R-squared and a negative coefficient on the contemporaneous GR term.[25] So the traditional models commonly used to compute the GR do not perform very well in explaining aggregate productivity dynamics in the UK.

Next, we move to the CF models (models 3-5). First, in Model 3, we have the case where we do not allow for the firm-specific fixed effects or the common shocks (i.e., $\mu_t = 0$ and $c_i = 0$ in equation (12)). The results are poor; the model has no explanatory power for aggregate productivity. The explanatory power increases when we allow for firm-specific fixed effects in Model 4. The contemporaneous GR coefficient is now significant, and the R-squared is around 0.17. Model 5, which includes both firm fixed effects and a sector-specific common shock, performs the best, with statistically significant GR coefficients up to the first lag, and an adjusted R-squared of 0.23. It is now also notable that — unlike for the other models — the second order model performs clearly better than the first order one, as would be expected. This is our preferred model; it uses our theoretically justified method of identifying firm-specific TFP shocks, results in correctly signed and significant coefficients, and the second order version of the model explains more of the aggregate productivity dynamics than the first order one.

Overall, our analysis suggests that the commonly-used models to compute the GR may be sensitive to alterations in the specification of the sample of firms, and some of the results look counter-intuitive. We show that the CF model performs better, and it is important to allow for common shocks as well as firm-specific fixed effects in the model. Hence, by properly accounting for firm-level technology shocks, we find that the shocks to the largest 100 UK firms can explain around 30% of aggregate productivity dynamics in the UK. In other words, we find support for Gabaix's (2011) granularity hypothesis in a manner that overcomes some of the challenges that may render the existing empirical tests of it unreliable.

[24]The different versions in the tables correspond to the models in Table 1. The aim of the regressions is not to estimate the best possble fit for modelling UK productivity; clearly other explanatory variables would be important in such an exercise. Rather, the aim is to study the relevance of the GR terms alone, similarly to Gabaix (2011).

[25]Our results associated with Model 2 are subject to an important caveat: since our dataset does not contain any information on *domestic* employment, we proxy for it by assuming that the fraction of a firm's total employment that is UK-based is equal to the ratio of domestic to total turnover.

4.2 Robustness Checks

Next, we explore a number of robustness checks on the results reported above, especially regarding our preferred model. Selected results are presented in Table A.3.

First, the results in column (1) are from an estmation with the GR benchmark model, but with using total rather than domestic sales as weights in the GR calculation. This is exactly the same set-up as in Gabaix (2011), who, as mentioned above, used total sales of US firms. The results do not indicate significance, in other words, the original set-up does not produce good results for the UK.

Columns (2) and (3) check how sensitive the results of the CF model are to using different number of firms as the basis for the CF calculation. The results are very similar to the main one using $K = 100$, so they are robust to the choice of the number of largest firms.

The productivity dynamics during the financial crisis are clearly an important feature of that data. We explore the effects of the financial crisis period by dummying them out in Models (5) and (6), the former for the CF unity and the latter for the GR benchmark model. The signs are as expected, and the explanatory power in terms of the R-squared improves dramatically. This is especially true for the GR model, which has a higher adjusted R-squared than the CF model, and the GR coefficients are significant. For the CF model, the coefficients are of the expected sign, but they are no longer significant. Given the large variation in both the aggregate productivity as well as the GR and CF dynamics around the crisis period, it is not surprising that this is an important factor for the results. The CF model does seem to capture aggregate productivity dynamics better when there is a lot of variation in productivity, without having to exclude this variation with dummy variables.

The specification of our preferred model poses the question on how important it is to estimate the firm-specific TFP shocks with the De Loecker and Warzynski (2012) method, or would generic Cobb-Douglas type coefficients perform as well. To test this, we use economy-wide coefficients of 2/3 for labour inputs and 1/3 for capital inputs, without using the CF estimation procedure (Model 6 in Table A.3). The results are very poor; the model has no explanatory power for aggregate productivity. We thus conclude that our CF procedure is important for estimating the relevant TFP shocks.

5 Conclusion

This paper studies the granular effects of the largest 100 firms on aggregate productivity in the UK. We first estimate a basic version of the Granular Residual model, introduced by Gabaix (2011) and note that some of the empirical results are counter-intuitive and not statistically significant. We then introduce a control function approach, where we, uniquely, allow for idiosyncratic as well as common TFP shocks. Using this measure, we find that the Granular Residual can explain around 30% of aggregate UK productivity dynamics. Hence, firm-level shocks to large firms matter. The results are, however, sensitive to the financial crisis period; the benchmark GR performs better in the years outside it, while the control function based TFP shock measure captures the crisis dynamics relatively well.

Our results shed new light on the propagation and importance of firm-level TFP shocks on aggregate productivity. However, given the relatively small sample at our disposal for the analysis, the results are necessarily tentative. For example, boostrapping experiments would suggest that some of the explanatory power in our simple regressions may not be statistically significant. Therefore, it would be interesting to take our preferred modelling strategy to different and larger datasets of other countries and examine whether the results hold, or to study other methods for identifying firm-specific idiosyncratic shocks. We leave these questions for future research.

References

Ackerberg, DA., K. Caves and G. Frazer (2015). Identification properties of recent production function estimators. *Econometrica*, Vol. 83(6), pp. 2411–2451.

Baqaee, D. and Farhi, E. (2019). The Macroeconomic Impact of Microeconomic Shocks: Beyond Hulten's Theorem. *Econometrica* (forthcoming).

Barnett, A., Batten, S., Chiu, J., Franklin, J. and M. Sebastia-Barriel (2014a). The UK productivity puzzle. *Bank of England Quarterly Bulletin*, 2014Q2.

Barnett, A., Chiu, J., Franklin, J. and M. Sebastia-Barriel (2014b). The productivity puzzle: a firm-level investigation into employment behaviour and resource allocation over the crisis. *Bank of England Working Paper*, No. 495.

Carney, M. (2017). Policy Panel: Investment and growth in advanced economies. Speech at 2017 ECB Forum on Central Banking, Sintra, 28 June 2017.

Cerra, V. and Saxena, S. C. (2008). Growth Dynamics: The Myth of Economic Recovery. *American Economic Review*, 98 (1): 439-57.

De Loecker, J. and Warzynski, F. (2012). Markups and Firm-Level Export Status. *American Economic Review*, Vol. 102, No. 6, pp. 2437-71.

Di Giovanni, J., A. Levchenko and I. Mejean (2014). Firms, Destinations, and Aggregate Fluctuations. *Econometrica*, Vol. 82(4), pp. 1303-1340.

Friberg, R. and M. Sanctuary (2016). The contribution of firm-level shocks to aggregate fluctuations: The case of Sweden. *Economic Letters*, Vol. 147, pp. 8-11.

Gabaix, X. (2011). The Granular Origins of Aggregate Fluctuations. *Econometrica*, Vol. 79(3), pp. 73s3-772.

Gabaix, X. and Koijen, R. S. J. (2019). Granular Instrumental Variables. *Mimeo.*

Gnocato, N. and Rondinelli, C. (2018). Granular Sources of the Italian business cycle. *Banca D'Italia Working Papers*, No. 1190 (September 2018).

Goodridge, P., Haskel, J. and Wallis, G. (2018). Accounting for the UK Productivity Puzzle: A Decomposition and Predictions. *Economica*, 85: 581-605.

Gutierrez, G. and T. Philippn (2019). Fading Stars. *NBER Working Paper*, No. 25529, February 2019.

Haldane, A. (2018). Market Power and Monetary Policy. Speech at the Federal Reserve Bank of Kansas City Economic Policy Symposium, Jackson Hole, Wyoming, 24 August 2018.

Jorda, O., M. Schularik and A. Taylor. (2013). When Credit Bites Back. *Journal of Money, Credit and Banking*, Vol. 45(2), pp. 3-28.

Lee, Y. , Stoyanov, A. and Zubanov, N. (2019). Olley and Pakes-style Production Function Estimators with Firm Fixed Effects. *Oxford Bulletin of Economics and Statistics*, 81: 79-97.

Levinsohn, J. and Petrin, A. (2012). Measuring Aggregate Productivity Growth Using Plant-level Data. *The RAND Journal of Economics*, 43, no. 4, pp. 705-25.

Lin, S. and M.F. Perez (2014). Do Firm-Level Shocks Generate Aggregate Fluctuations?. *Mimeo.*

Mason, G., M. Mahoney, R. Riley (2018). What is Holding Back UK Productivity? Lessons from Decades of Measurement. *National Institute Economic Review*, Vol. 246(1).

Melitz, M. and Polanec, S. (2015). Dynamic Olley-Pakes productivity decomposition with entry and exit. *The RAND Journal of Economics*, Vol. 46, pp 362-375.

Oulton, N. and M Sebastia-Barriel (2017). Effects of financial crises on productivity, capital and output. *Review of Income and Wealth*, Vol. 63(1), February 2017.

Riley, R., C. Rosazza-Bondibene and G. Young (2015). The UK productivity puzzle 2008-13: evidence from British businesses. *Bank of England Working Paper*, No. 531.

Schneider, P. (2018). Decomposing differences in productivity distributions. *Bank of England Working Paper*, No. 740.

Tenreyro, S. (2018). The fall in productivity growth: causes and implications. Speech at Peston Lecture Theatre, Queen Mary University of London, 15 January 2018.

A. Data

Data used in the analysis comes from three data sources: Thomson Reuters Worldscope, Office for National Statistics (ONS) Business Structure Database (BSD) and OECD/ONS Labour Productivity data. This Annex describes the data we use from these datasets in detail.

Worldscope is a proprietary dataset that includes financial account information on large (mainly listed) UK firms since the 1980s. For our analysis, we use data on the following variables:

- **Turnover** for the largest firms for each year (minus international sales for those firms that have reported it).
- **Employment**, used as the denominator in the firm-level turnover productivity calculation (for those firms that report international sales, the same proportion of domestic employment is used as for domestic turnover).
- **Cost of goods sold**, used as a measure of variable costs in the GR control function approach.
- **Net property, plant and equipment** as a measure of firm-level capital.

We have noticed a number of mistakes in the Worldscope data. This is especially true for the employment data; as this is not a financial account item, there seems to be more errors in the recording of it in the Worldscope dataset. While correcting for all mistakes in the entire data would not be feasible, we have gone through the largest errors for those firms that contribute most to the dynamics of the GR measures and replaced the erroneous observations with the correct ones from the annual reports of the firms in question, when these have been available. In particular, the number of corrections we do are the following:

- Cost of good sold; corrections for 10 firms.
- International sales; corrections for 70 firms.
- Employment; corrections for 30 firms.

BSD covers the universe of all UK enterprises, based on an annual snapshot from the Inter-Departmental Business Register (IDBR). On average, there are over 2 million enterprises annually in the BSD. In the IDBR, turnover is updated via administrative sources (HMRC VAT and PAYE records) and ONS Business Surveys. We use enterprise level turnover for calculating the power law distributions, based on the BSD vintage of 2017.

To calculate the productivity contributions of the largest firms for the Melitz and Polanec (2015) methodology, we use the ONS **ARD** database. This database includes firm-level financial account data based on the Annual Business Inquiry (ABI), Annual Business Survey (ABS) and employment data based on the Business Register and Employment Survey (BRES). We use reporting unit level turnover and employment data to calculate estimates of turnover productivity. Note that the data is only available for 2002-2014, with methodological breaks during this period, so the results should only be seen as indicative and motivating the main analysis using the Worldscope data.

ONS Labour Productivity data is used for the GR regressions, where aggregate labour productivity (output per head) is the independent variable. In addition to this, Figure 1 uses OECD data on hourly labour productivity for selected economies. All the ONS and OECD aggregate data is publicly available.

B. Tables and Figures

B.1 Tables

Table A.1: Predictive Power of the First-Order Granular Residual for UK Productivity Growth

	(1) Benchmark	(2) Labour weights	(3) No fixed effects	(4) No common shocks	(5) Unitary loadings
GR_t	0.797	-4.161	1.037	2.356*	1.718**
	(1.155)	(3.208)	(0.723)	(1.156)	(0.738)
GR_{t-1}	0.609	1.509	0.276	2.087	1.696*
	(1.305)	(2.264)	(0.472)	(1.274)	(0.950)
GR_{t-2}	1.778**	3.415	-0.203	0.429	0.711
	(0.791)	(2.296)	(0.582)	(1.137)	(0.604)
(Intercept)	0.013***	0.013***	0.015***	0.017***	0.019***
	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)
N	28	28	25	25	25
R^2	0.173	0.142	0.139	0.269	0.326
Adjusted R^2	0.070	0.034	0.016	0.165	0.229

Notes: The columns correspond to the different versions of the granular residual. The dependent variable is annual aggregate productivity growth per head (GVA/employment). The sample is annual data from 1988 to 2016. All models exclude oil and finance sectors. HAC standard errors shown in parentheses. Statistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. See main text for further details on the specifications.

Table A.2: Predictive Power of the Second-Order Granular Residual for UK Productivity Growth

	(1) Benchmark	(2) Labour weights	(3) No fixed effects	(4) No common shocks	(5) Unitary loadings
GR_t	0.639	-4.339	1.127	2.211*	1.815**
	(1.250)	(2.867)	(0.753)	(1.119)	(0.670)
GR_{t-1}	0.462	0.032	0.265	2.119	1.867**
	(1.494)	(3.081)	(0.508)	(1.336)	(0.887)
GR_{t-2}	2.002**	3.508	-0.158	0.585	0.817
	(0.934)	(2.300)	(0.604)	(1.160)	(0.550)
(Intercept)	0.011***	0.013***	0.015***	0.017***	0.019***
	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)
N	28	28	25	25	25
R^2	0.172	0.147	0.142	0.244	0.406
Adjusted R^2	0.068	0.040	0.019	0.135	0.321

Notes: The columns correspond to the different versions of the granular residual. The dependent variable is annual aggregate productivity growth per head (GVA/employment). The sample is annual data from 1988 to 2016. All models exclude oil and finance sectors. HAC standard errors shown in parentheses. Statistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. See main text for further details on the specifications.

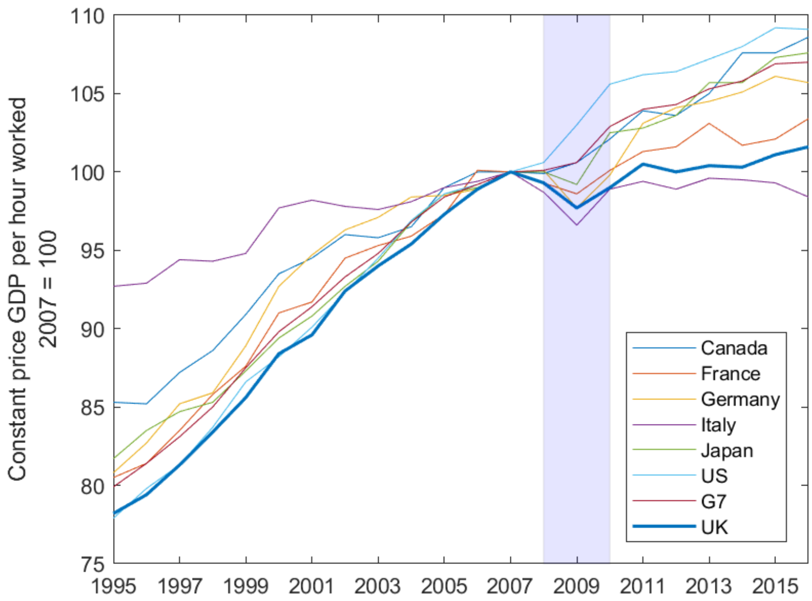
Table A.3: *Selected Robustness Checks on the Regression Results of the Granular Residual*

	(1) GR benchmark	(2)CF unity	(3) CF unity	(4) CF unity	(5) GR benchmark	(6) CF Cobb-Douglas
	1st order	2nd order	2nd order	2nd order	1st order	
	total sales	K=50	K=200	Crisis dummies	Crisis dummies	fixed C-D coeffs.
GR_t	-0.254	1.988**	1.676**	0.835	1.229*	0.0984
	(0.657)	(0.732)	(0.635)	(0.595)	(0.619)	(0.487)
GR_{t-1}	-0.255	2.254**	1.645*	0.838	1.341**	-0.414
	(1.086)	(0.991)	(0.828)	(0.669)	(0.484)	(0.691)
GR_{t-2}	1.468*	0.914	0.768	0.597	1.103	-0.0984
	(0.774)	(0.568)	(0.491)	(0.532)	(0.719)	(0.535)
Crisis dummy				-0.027***	-0.037***	
				(0.008)	(0.007)	
(Intercept)	0.014***	0.020***	0.020***	0.019***	0.016***	0.014***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.001)	(0.003)
N	28	25	25	25	28	27
R^2	0.158	0.417	0.412	0.612	0.701	0.026
Adjusted R^2	0.052	0.334	0.328	0.534	0.649	-0.101

Notes: The columns correspond to the different versions of the granular residual. The dependent variable is annual aggregate productivity growth per head (GVA/employment). The sample is annual data from 1988 to 2016. All models exclude oil and finance sectors.
HAC standard errors shown in parentheses. Statistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
See main text for further details on the specifications.

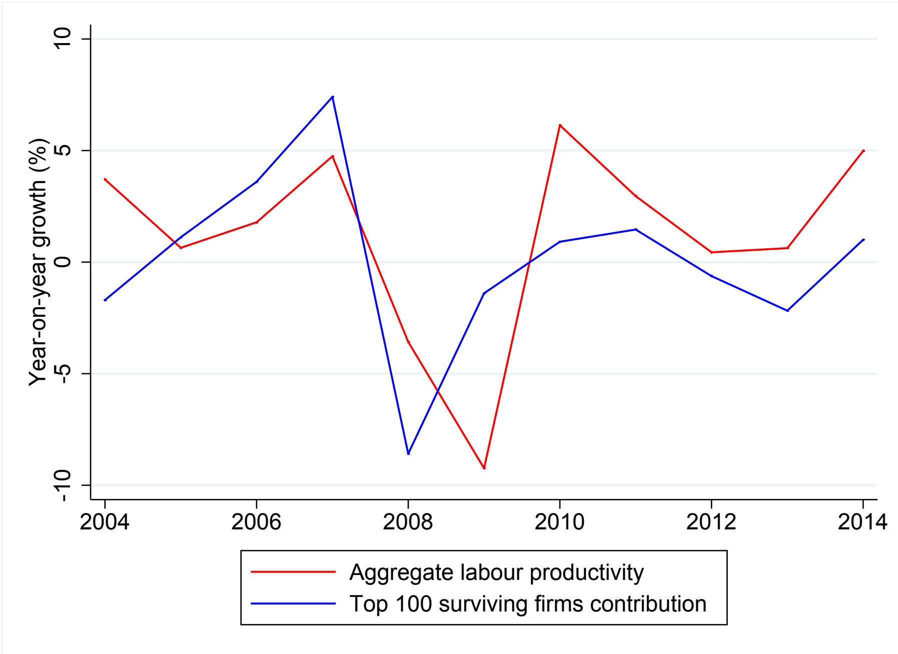
B.2 Figures

Figure 1: Labour productivity in G7 economies



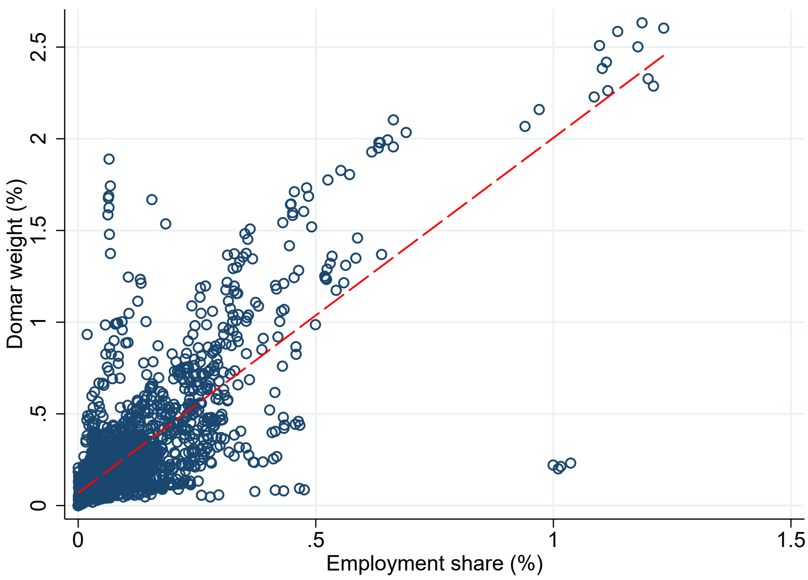
Sources: OECD, ONS.

Figure 2: *Contributions of top 100 firms to UK productivity growth (Melitz-Polanec method)*



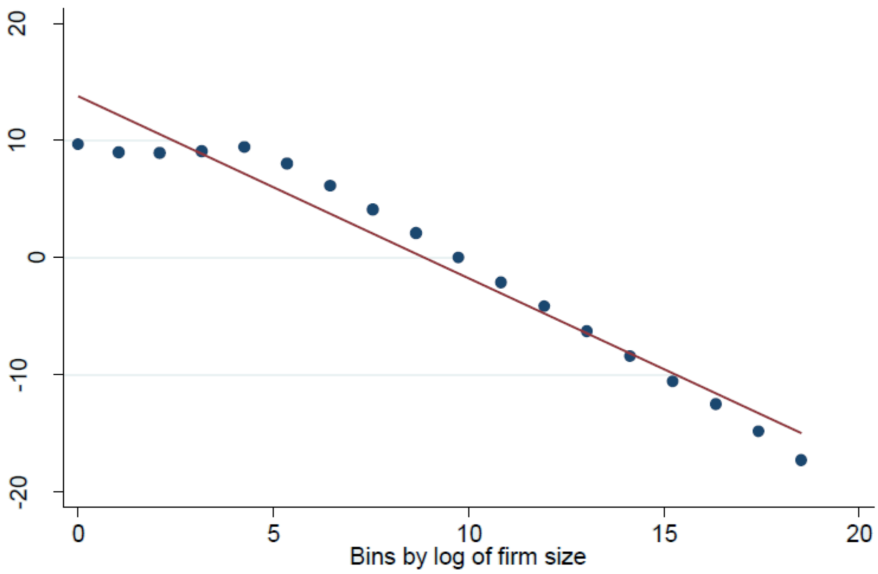
Sources: ONS, authors' calculations.

Figure 3: *Scatterplot of Domar weights and employment shares among the largest 100 firms*



Sources: Bank of England, Worldscope, authors' calculations.

Figure 4: *Histogram of firm size in the UK*



Source: ONS, authors' calculations.

Notes: The horizontal axis has 17 bins of equal size of turnover (in logs). The vertical axis has (log of) the number of firms in each bin, adjusted for the width of the bin.

Figure 5: *Share of top 100 firms of UK GDP*

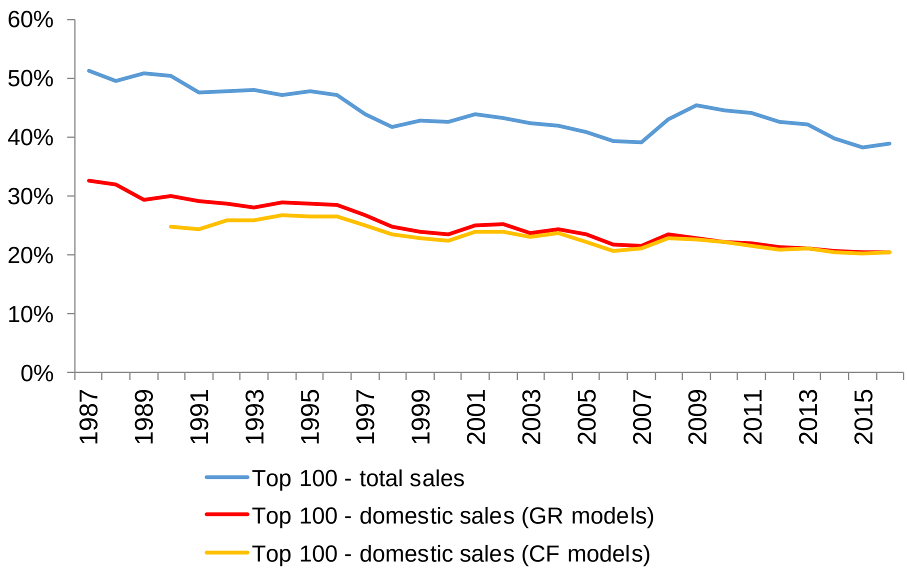


Figure 6: *1st Order Granular Residual based on the largest 100 firms listed in the UK*

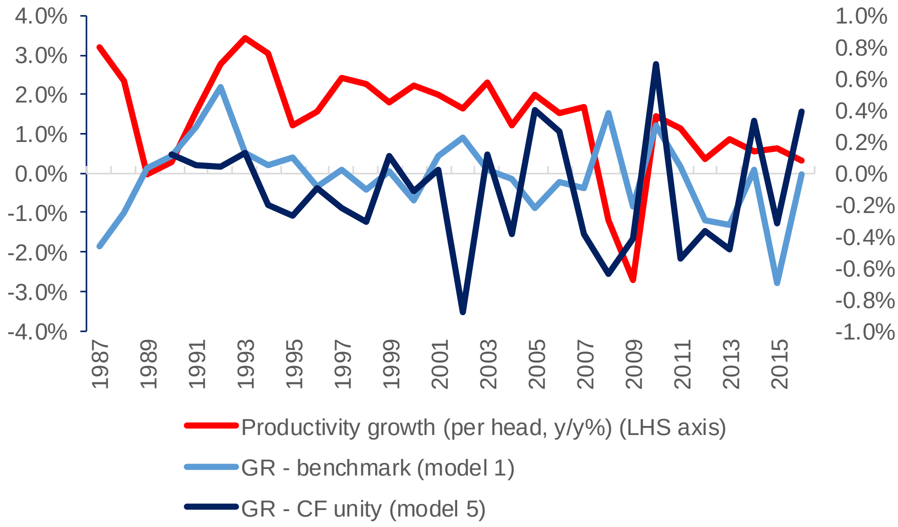


Figure 7: *2nd order Granular Residual based on the largest 100 firms listed in the UK*