Heterogeneous Firms, Wages, and the Effects of Financial Crises

(click here for latest version)

Alex Clymo*

October 8, 2017

Abstract

The Great Recession manifested itself mostly as a decline in hours and employment in the US, but a decline in Total Factor Productivity in the UK. Why did these two economies experience the crisis differently? I argue that a third fact, that real wages fell further during the crisis in the UK, may provide an explanation. I show that how a financial crisis manifests in the real economy in standard heterogeneous firm models depends crucially on assumptions on wage adjustment, and use this result to explain the divergent experiences of the two countries during the recent financial crisis. Theoretically, while a decline in wages protects the labour market, I show that, in the presence of financial frictions, it also causes a fall in TFP by misallocating resources across firms. Financially unconstrained firms are able to take greater advantage of a decline in wages than constrained firms, leading to a relative reallocation of resources towards the unconstrained, which I show must always reduce TFP in my model. On the other hand, if wages are fully rigid, I show that a financial crisis will have no effect on TFP. Quantitatively, the model is able to explain the greater fall in hours during the crisis in the US, the lack of a significant fall in TFP in the US, and 1/3 of the UK's TFP decline, or "productivity puzzle".

JEL classification: E24, E32, E44, G01, L11.

Keywords: financial crises, misallocation, employment, productivity, heterogeneity.

*Department of Economics, University of Essex. Email: a.clymo@essex.ac.uk. I thank Wouter den Haan for invaluable support. I also thank Charlie Bean, Esther Bøler, Francesco Caselli, Sergio de Ferra, Kyle Herkenhoff, Reka Juhasz, Andrea Lanteri, Pascal Michaillat, Marcelo Pedroni, Jonathan Pinder, Morten Ravn, Kevin Sheedy, Robert Shimer, Christian Stoltenberg, Silvana Tenreyro, and John Van Reenen for helpful advice and comments, as well as seminar participants at the Econometric Society Winter Meetings 2015, RES Conference 2016, Computing in Economics and Finance Conference 2016, EEA Meetings 2016, Banque de France, Bank of England, Stockholm University, LSE, and University of Amsterdam. I thank the ESRC, Paul Woolley Centre, and ACRM for financial support. All errors are my own.

1 Introduction

The recent financial crisis caused a highly synchronised recession across much of the developed world. However, beneath the surface there are differences in how countries experienced this decline. In this paper, I focus in particular on two countries: the United States and the United Kingdom. Both of these countries have come under scrutiny in recent years because of the nature of their experiences of the Great Recession, with the former suffering an unusually large drop in employment, and the latter an unusually large drop in aggregate Total Factor Productivity (TFP). How can we rationalise this heterogeneous behaviour in response to a common global shock? I propose a novel mechanism to explain how a financial shock can be transmitted to the real economy in different ways, paying particular attention to the behaviour of real wages.

During the Great Recession, the US saw an unusually large fall in employment and hours. Between 2008 and 2011 total hours fell by roughly 10%. On the other hand, over that period aggregate TFP remained robust, and even increased slightly relative to trend. In the UK the labour market performed relatively better, with hours only falling by just over 5% over the same period, but the TFP performance has been dismal. Aggregate TFP has fallen by over 5% relative to trend, in what has been deemed the UK's "productivity puzzle". At the same time, the real wage behaviour of the two countries has been very different, with the US seeing wages grow in line with their trend over the period, and the UK seeing wages fall by at least 6% relative to trend. This difference in wage behaviour appears to be driven largely by differences in inflation, with the UK running consistently higher inflation than the US during the crisis.[1]

Using worker-flow data, I also show that the less severe fall in employment in the UK than in the US during the crisis is driven by more robust job creation in the UK, not differences in job destruction. This is evidence against a simple labour hoarding explanation of the difference between the two countries,[2] and also implies that the UK reallocated more labour during the crisis than the US, where people instead became unemployed. Additionally, a review of the available evidence suggests that the misallocation of resources across firms during the crisis has been worse in the UK than in the US, consistent with the idea that maintaining higher employment in the UK has come at the cost of lowering productivity via misallocation: around 1/3 of the UK's TFP decline can be attributed to misallocation (Barnett et al., 2014, Bordon et al., 2016).

Motivated by the above facts, in this paper I ask whether the countries' different wage paths during the crisis can explain why the US experienced the crisis mostly as a decline in hours, and the UK more as a decline in TFP. The key idea is that falling real wages protect employment by encouraging hiring, but since a financial shock disrupts which firms can afford to hire, this hiring will necessarily come from increasing the reallocation of labour across firms. Since this reallocation is driven by financial frictions, it will in fact be misallocation, and the protection against unemployment that falling wages brings will come at the expense of lowering aggregate TFP by reallocating resources to the wrong firms.

I present my model in two sections. I first present a static version of the model which delivers analytical results. I prove that a financial shock is transmitted more into aggregate TFP declines and less into hours declines the more the real wage falls following the shock, consistent with the

[1] In the appendix, I show that these patterns also holds in a cross section of OECD countries during the recent crisis, and hence that the results of this paper are applicable more broadly to developed economies.

[2] The US and UK score similarly in OECD employment protection indicators. The US has the least protected labour market in the OECD, while the UK has the third, behind only the US and Canada (Martin and Scarpetta, 2011).

behaviour of wages, hours, and TFP in the US and UK during the crisis.

The model builds on the work of Buera and Moll (2015) and Khan and Thomas (2013). It features firms that are heterogeneous in their productivities and net worths, and produce goods which are imperfect substitutes. Firms compete under monopolistic competition, giving well defined, interior optimal capital and labour choices. The model is parsimonious, and utilises a simple collateral constraint and constant markups (induced by a CES demand structure). I assume a flexible form of wage rigidity which nests both fully rigid and flexible real wages. While the underlying distribution of firms can be rich, these assumptions make firms' policy functions particularly tractable, allowing for novel analytical results.

I first prove the perhaps surprising result that if wages do not adjust, a financial crisis leads to no fall in aggregate TFP. Additionally, the lack of a wage decline makes the reduction in labour demand particularly severe, leading to a large reduction in hours. Thus the model is able to replicate the US experience of no TFP or wage decline and a large fall in total hours following the financial crisis.

Without a wage decline, a financial crisis will not affect TFP because the shock is transmitted even to firms that are not financially constrained via a demand spillover. Only an endogenous subset of firms – the poorest, or most productive – will be financially constrained in equilibrium. These firms are forced to shrink, reducing total output and hence demand for *all firms'* output. Without a decline in real wages, unconstrained firms optimally choose to downsize by the same proportion that constrained firms are forced to. This leads all firms to shrink by the same amount, meaning that there is no reallocation across firms and hence no fall in TFP.[3] Intuitively, with a sticky wage a financial shock also becomes a *demand* shock, leading even unconstrained firms to downsize.

I then show that it is only via induced declines in equilibrium wages that aggregate TFP will endogenously decline in this framework. Additionally, a wage decline boosts labour demand, reducing the fall in hours during a crisis. Thus, the model is also able to replicate the UK experience of both TFP and wage declines, and a smaller fall in total hours, following the crisis.

If wages are allowed to fall, this will protect hours by encouraging hiring. But how does this lead to a fall in TFP? A key question is which firms choose to hire. While, conditional on a choice of capital level, all firms will choose to hire more, firms that are financially unconstrained will also be encouraged to purchase more capital when wages fall because profits will rise. This leads to a reallocation of resources towards unconstrained firms, which I show must cause a fall in aggregate TFP in my framework. Intuitively, this is because financially constrained firms have been kept below their optimal size, and the efficient allocation would prefer more resources to be allocated to them, while a wage decline encourages the opposite. Using the terminology of Hsieh and Klenow (2009), I prove that financially constrained firms will always have higher idiosyncratic productivity (measured in terms of their revenue TFP, or TFPR) than unconstrained firms, and hence that reallocating resources towards unconstrained firms must lower aggregate TFP.

These results highlight the key tradeoff: during a financial crisis, wage declines will protect labour markets, but will do so at the expense of inducing misallocation, and hence lowering aggregate TFP. This allows the model to replicate both the UK's and US' experience of the crisis, once the fact that their wages behaved differently is taken into account.

I am able to obtain sharp analytical results by assuming a CES demand structure, no frictions in hiring at the firm level, and a simple borrowing constraint limiting capital purchases to a fixed

[3] As I discuss in the paper, in the case of perfect substitutability the same result goes through, not due to demand spillovers but due to lack of movement in the productivity threshold above which firms produce.

multiple of net worth. The results are robust to allowing several extensions to more complicated borrowing constraints. I additionally show that the results also apply on the firm entry/exit margin: For a fixed wage, a financial crisis will now *increase* TFP by reducing profits, causing the exit (or stopping the entry) of the least productive firms. A decline in the wage during a crisis will instead increase profits, reducing TFP by allowing less productive firms enter and survive. Finally, I construct business cycle wedges following Chari, Kehoe and McGratten's (2007) accounting procedure. This exercise shows that the US' recession can be explained mostly through the labour wedge, and the UK's recession mostly through the efficiency wedge. My model is also consistent with this evidence: I show that following a financial shock, the model generates only a change in the labour wedge if wages are fully rigid, and only a change in the efficiency wedge if wages are fully flexible.

In order to test the quantitative relevance of these mechanisms, I then extend the model to a dynamic setting and calibrate it to match firm-level data in each country. I introduce a life-cycle firm structure, where firms are born and eventually die, creating a rich distribution across firms. Firms enter poor and small, and accumulate net worth as they age, growing towards their optimal size. I also introduce capital adjustment costs at the firm level, in order to match firm-level investment patterns. These take the form of a resale price for capital below the purchase price, leading firms to follow (S, s) decision rules which feature inaction regions. The model is solved numerically using heterogeneous firm techniques (see, e.g., Khan and Thomas, 2008) which require tracking the distribution of firms over time.

I perform a series of partial equilibrium experiments to illustrate the effect of wage adjustment on the transmission of a financial crisis. In the first experiment, I simulate a financial shock big enough to explain the fall in output in each country, conditional on their observed wage paths. I then compute the time paths for total hours and TFP, and compare the models to the data. The model correctly predicts a larger fall in hours in the US, and a larger fall in TFP in the UK. The results are quantitatively relevant: the model generates around 1/3 of the TFP fall in the UK, the lack of a TFP fall in the US, and at its peak captures around 2/3 of the larger decline in hours in the US. Additionally, the size of the TFP decline generated for the UK is comparable to the component attributed to misallocation in the UK data.

Of course, a financial shock may not have been the only shock hitting the economy, particularly in the UK where there does also appear to be a within-firm drop in TFP alongside the increase in misallocation in the data. Thus, in a second experiment, I also allow for a common TFP shock, and use output and TFP data to identify the size of each shock. The model now endogenously accounts for 40% of the TFP difference between the two countries, and 83% of the hours difference. Overall, these two experiments reveal that wage adjustment has a quantitatively large effect on the manifestation of financial crises theoretically, and that this mechanism helps explain the divergent experiences of the US and UK during the recent crisis.

In order to generate the above results, I also make a contribution to numerical techniques for simulating heterogeneous firm models. The calibrated dynamic model is relatively complex, featuring both capital adjustment costs and financial frictions. I develop a new "non-stochastic simulation" procedure (Young, 2009) which overcomes the curse of dimensionality in existing procedures by choosing the nodes over which to approximate the firm distribution endogenously using firms' policy functions. This makes it feasible to apply non-stochastic simulation even though my model has multiple endogenous firm-level states, and the algorithm is applicable to many other existing models.

Having shown that the model is able to match empirical aggregates well, I then turn to evaluating its performance relative to more disaggregated data sources. As previously mentioned, the difference in employment between the two countries is primarily driven by differences in job creation, not destruction. I show that my model is consistent with this fact, in contrast to explanations which rely on greater labour hoarding in the UK. I also show that an extension of my model to include entry and exit is consistent with the fact that the number of firms declined more during the crisis in the US than in the UK. Looking at the composition of debt, the decline in lending to firms is more concentrated in bank lending (as opposed to, e.g., bonds) in the UK than in the US, which is consistent with the sizes of the financial shocks estimated in my model decompositions.

Finally, I look at data split by firm size, using firm size as a proxy for financial constraints. Small firms may be more likely to be financially constrained, and Crawford et al. (2013) show that investment declines were more severe at small, rather than large, firms in the UK. I replicate their methodology using US data and show that this is less true in the US. This is consistent with my model, which requires that unconstrained firms in the UK suffer less than in the US, benefiting from the extra wage declines.

The data on employment by firm size is less supportive, with the predictions not lining up with those of the model. While this could just be because firm size is an imperfect proxy for financial constraints, I argue that this failure of the model actually hints at an interesting mechanism that the model is missing. My model assumes a common wage across firms, and studies the effect of a fall in this common wage. In reality, firms offer heterogeneous wages, and wages in the UK fell more at small firms during the crisis than at large firms. This also appears to have happened less in the US, where wage declines were smaller and not as heterogeneous across firms. I argue that accounting for wage heterogeneity should reconcile my model with the data, and would even enhance the effects of a wage reduction on TFP.[4]

In summary, my paper makes two main contributions. Firstly, I document key stylised facts about how the UK and US experienced the financial crisis. I show that the crisis manifested more in hours and employment in the US, and TFP in the UK, and that wages fell further in the UK. Most importantly, I document that the greater employment fall in the US is driven by a greater fall in job creation, and not by a greater rise in job destruction. This rules out the simple explanation that the UK hoarded more labour during the crisis, and instead demonstrates that the UK maintained employment by reallocating more labour than the US during the crisis.

Secondly, I construct a theoretical model to explain the above differences between the two countries' behaviour during the financial crisis. The model takes the wage behaviour of both countries during the crisis as given, and can rationalise the other differences endogenously. The model performs well quantitatively, and is consistent with the dominant role of job creation in explaining the employment difference, and greater evidence of misallocation in the UK. Importantly, I thus show that a large part of the experiences of the two countries can be explained within a single model. This informs future modelling of financial crises by showing that a unified framework is able to explain different manifestations of crises, once the behaviour of, for example, wages is taken into account.

My results are also interesting because I show that there is a cost to increasing wage flexibility. While it will prevent large declines in hours during a crisis, this will come at the cost of reducing TFP, which will affect welfare by reducing resources available for consumption. This also applies to monetary policy during a crisis: loose monetary policy can protect labour markets, but will do so

[4]Specifically, TFP declines in my model disrupt the distribution of capital across firms, but capital-labour ratios at each firm are still correctly set since firms face a common wage. The extra wage dispersion in the UK induced by the crisis would additionally disrupt capital labour ratios, creating an additional source of misallocation.

by reducing real wages and thus reducing TFP. While I do not make any normative statements in this paper, it sets out these key tradeoffs which can be taken to a welfare context in future work. Finally, if the hypothesis entertained by my model is correct, the UK's productivity problems may not be entirely permanent, since they are driven by misallocation rather than technology. If wages recover in the coming years, either as the economy normalises, or in response to explicit policies, we may also see a recovery in productivity.

Related Literature. This paper is related to many theoretical and empirical papers that relate financial frictions to labour markets and productivity. Most relevant to my work are papers which study the effects of financial shocks on the real economy.[5] This is a growing literature, with different papers emphasising different mechanisms and results. For example, Khan and Thomas (2013) build a calibrated numerical model, with firms heterogeneous in productivity and net worth, subject to capital adjustment costs. They show that a negative financial shock leads to a fall in TFP by disrupting the allocation of resources across firms. In an analytical framework, Buera and Moll (2015) also show that if firms are heterogeneous in productivity, a financial shock will manifest as a fall in TFP, by reallocating resources towards less productive firms.

In contrast, Petrosky-Nadeau (2013) shows that a negative financial shock will *increase* TFP. How can we rationalise such different results? My work provides a link between them, and highlights the roles of factor prices in driving misallocation. In Petrosky-Nadeau's (2013) model, a financial crisis increases TFP because it reduces profits for all firms, which encourages the least productive firms to stop producing, increasing aggregate TFP. In Buera and Moll's (2015) model, a financial crisis reduces TFP because it reduces factor prices, such as the wage. Reducing the wage increases profits for all firms, which encourages unproductive firms to start producing, lowering aggregate TFP. Abstracting from interest rate movements, I show that if wages don't adjust there is no force increasing profits and hence no change in the productivity threshold, or aggregate TFP. In these two models, the effects of a financial shock on TFP can thus be understood as a tug-of-war between (a) increased borrowing costs reducing profits, and (b) reduced wages increasing profits, highlighting the key role of wages on the transmission of the shock.

I add to this literature by investigating how wage adjustment affects the manifestation of a financial shock. Thus, I clarify the mechanisms in these models and show that the result that financial shocks affect TFP is actually sensitive to assumptions about whether wages adjust. Other papers have investigated partial wage adjustment in heterogeneous firm models. Buera, Fattal-Jaef and Shin (2014) also generate a fall in TFP following a financial shock. In an extension to their model, they introduce sticky wages and show that unemployment increases further in their model when wages are sticky. TFP in their model appears to fall slightly less when wages are sticky, consistent with my results, although they do not discuss nor attempt to explain this. Arellano, Bai and Kehoe (2012) also consider a model with heterogeneous firms, and consider both flexible and sticky wages. Their focus is on matching labour market outcomes, and they do not compare endogenous productivity across their flexible and sticky-wage variants.

While I clarify the role of wage adjustment in transmitting productivity shocks, the existing literature has already emphasised the role of the interest rate in transmitting financial shocks. In particular, the idea is that a negative financial shock, by lowering the amount that firms can borrow,

[5]This literature builds on work studying the steady state or transitional effects of financial frictions on productivity, such as Moll (2014) and Midrigan and Xu (2014).

lowers borrowing demand and hence the interest rate, making it profitable for less productive firms to borrow and produce, lowering aggregate productivity. This mechanism is similar to my wage story, and is explored in Reis (2013) and Gopinath et al. (2016).

Hsieh and Klenow (2009) present a general theory of how misallocation affects aggregate TFP using the idea of firm-level "wedges". These capture deviations of firms' capital and labour choices from the optimum that would arise in the absence of any frictions. They emphasise that, in a certain class of models, dispersion in firm-level TFPR[6] is indicative of misallocation. I apply this insight to my model by showing that financial frictions induce dispersion in TFPR, raising the TFPR of financially constrained firms relative to unconstrained firms.

My work is informed by empirical work of several kinds. Firstly, firm level evidence of misallocation in the US and UK during the Great Recession. There is evidence of misallocation in both countries, with the balance appearing to show greater misallocation in the UK. For the US, see Foster et al. (2014) and Kurtzman and Zeke (2016), and for the UK see Barnett et al. (2014), Bordon et al. (2016), Riley et al. (2015), and Riley and Rosazza Bondibene (2016). I discuss these papers further in Section 2. My model also relies on the transmission of a financial crisis to the firm sector, which has been documented empirically in numerous studies, including Chodorow-Reich (2014) and Giroud and Mueller (2015) for the US and Franklin et al. (2015) for the UK.

Secondly, my paper is related to empirical work attempting to explain the UK's productivity puzzle. Pessoa and Van Reenen (2014) argue that the fall in output per worker in the UK is related to the UK's large fall in wages, which encouraged substitution away from capital towards labour. Goodridge et al. (2015) discuss the role of factor utilisation and certain key sectors in the puzzle, and I discuss their work further in Section 2.

Thirdly, my paper is related to papers which investigate why the financial crisis manifested differently across countries. Brinca et al. (2016) apply Chari et al.'s (2007) business cycle accounting procedure across countries during the Great Recession, as does Ohanian (2010). Barth et al. (2016) study employment across countries during the crisis, and Daly et al. (2014) study the breakdown of Okun's law across countries.

My paper is also related to other work which explores the role of factor prices in allocating resources and hence affecting aggregate productivity. The directed technical change literature emphasises the role of factor prices in determining where investment in factor-augmenting technical change is directed. Acemoglu (2002) lays the theoretical foundation for the modern literature. In the search-and-matching literature, Acemoglu and Shimer (1999, 2000) show that changes in the generosity of unemployment insurance have effects on measured aggregate productivity by changing the jobs that workers accept. More generous UI shifts the distribution towards higher productivity and wage jobs, increasing productivity while increasing unemployment.[7]

The rest of the paper is organised as follows. In Section 2, I present motivating evidence. Section 3 contains the analytical model and results, and Section 4 the quantitative results. In Section 5, I discuss how the firm- and worker-level implications of the model compare with the data, and in Section 6 I conclude. The appendices contain additional empirical work, including the international evidence and wedges decomposition, as well as a description of my new non-stochastic simulation algorithm.

[6]Revenue TFP, or TFPR, measures the efficiency at which a firm produces revenue for given capital and labour inputs, while quantity TFP, TFPQ, measures the productivity at which physical output is produced.

[7]Empirically, Acemoglu (1997) finds that increases in minimum wages and unemployment insurance are associated with higher wages and higher labour productivity, which is consistent with my theory.

2 Motivating evidence

In this section I present the stylised facts underlying this paper. I use a variety of data sources, and the results are summarised below.

1. The Great Recession manifested more as a fall in hours in the US, and a fall in aggregate TFP in the UK. This holds even after excluding the financial sector, using several labour market measures, and correcting TFP for factor utilisation.

2. Real wages fell more in the UK, measured both using a simple average wage measure, and controlling for workforce composition. The difference in the real wage response between the two countries is driven primarily by higher inflation in the UK, but institutional factors also played a role.

3. The larger employment fall in the US is driven by a larger fall in job creation, not more job destruction. This is evidence against greater labour hoarding in the UK as an explanation of the different employment responses.

4. The larger fall in job creation in the US also implies that the UK reallocated more labour during the crisis. Firm-level evidence suggests that the reallocation of resources across firms can explain around 1/3 of the greater aggregate TFP fall in the UK.

In Appendix A, I perform other exercises: a) I perform a business cycle accounting decomposition to show that the recession manifested more in the labour wedge in the US, and efficiency wedge in the UK, consistent with my model b) I give a discussion of why wages, and not real unit labour costs, are the appropriate measure for my exercise, and c) I take an international panel of countries during the crisis, and show that the same pattern described between the UK and US holds across more countries.

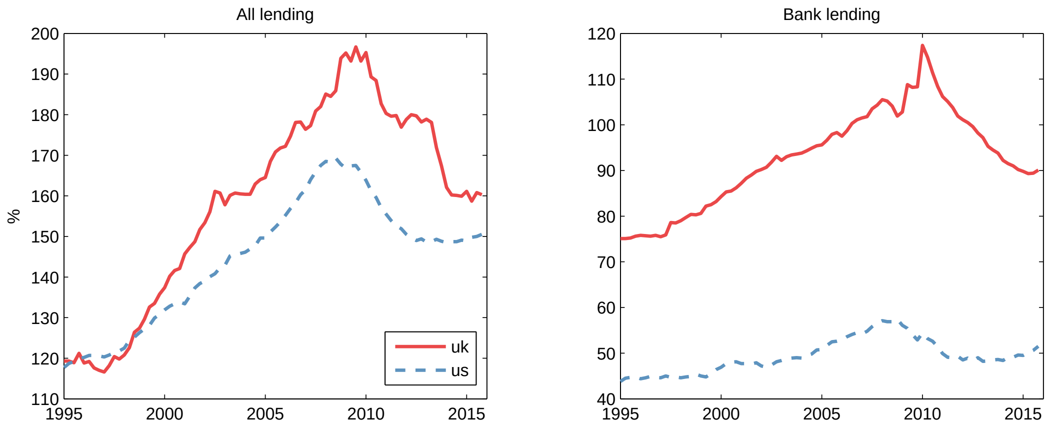
2.1 Lending

I will model the financial crisis as a reduction in lending to firms due to stresses in the financial sector. Before moving on to the main stylised facts, I thus first demonstrate that lending to firms fell in both countries during this period. In Figure 1, I plot the ratio of the stock of credit to the Private Non-Financial Sector to GDP for both countries. I include several years of pre-crisis data as a comparison. The left panel gives lending from all sources (which includes bank lending, bond issuances, and overseas lending), and the right panel lending only from domestic banks. Data is from the BIS, and credit is defined as loans, debt securities, and currency and deposits.

By either measure, firms use more debt in the UK, but the difference is not so large when measured in terms of all lenders, with the UK having a firm debt to GDP ratio of 185% in 2008, and the US 169%. The difference is more pronounced for bank lending, which the UK is much more reliant on. By both measures, we see a run up of debt before the crisis, which is reversed from 2008. The declines in total debt stocks are of a similar magnitude in both countries, falling by around 30pp in the UK and 20pp in the US, depending on the starting date used. Relative to the 2008 values, the declines are swifter for the US but recover faster, with the declines being relatively worse in the UK by the end of the sample.[8]

[8] I interpret these declines as signs of an exogenous reduction in lending to firms, which is the shock I use in my

Figure 1: Lending to Non-Financial Corporations during the Great Recession



The figures give outstanding credit to the Private Non-Financial Sector as a percentage of GDP. The left table gives total outstanding credit from all lending sources, and the right panel lending from domestic banks only. Data are from the BIS, and all data is quarterly.

2.2 Output, hours, and TFP

In Figure 2, I present the main stylised facts underlying the two countries' experiences of the Great Recession: I show that output fell in both countries, but that total hours fell more in the US, and TFP more in the UK. Data are taken from each country's national accounts, with details relegated to Appendix A. Output, Y_t, is taken as real per-capita GDP, deflated by the GDP deflator. Labour input, L_t, is measured by total hours worked, again per capita. TFP, Z_t, is measured simply as the residual in a standard Cobb-Douglas production function $Y_t = Z_t K_t^\alpha L_t^{1-\alpha}$, where K_t is capital (per capita, constructed via perpetual inventory) and $\alpha = 1/3$. Output and TFP are detrended by their 2000-2007 growth rates.[9]

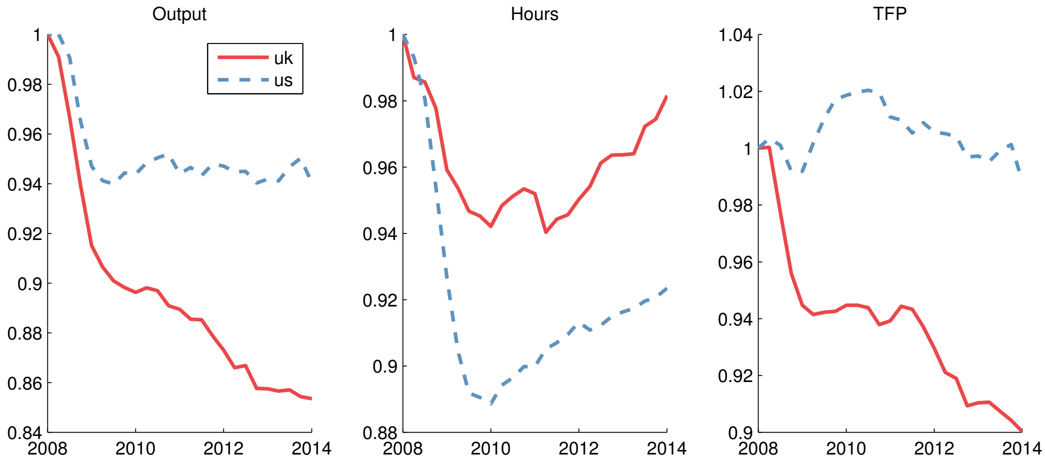
The left panel plots output for both countries. Both suffered sharp and persistent declines in output relative to trend: around 5% by 2011 in the US, and the UK faring worse at over 10%. Both countries continue to stagnate relative to trend even six years after the start of the recession. While the recession in the UK looks much worse, I show in the next section that once the financial sector is stripped out the recessions are of a similar magnitude.

However, both suffering a large recession is where the similarities between the two countries end. The central panel plots hours per capita, showing that the fall in hours in the US was nearly double that of the UK. The fall peaks at over 10% in the US, and is never more than 6% in the UK. Additionally, the recovery is much faster in the UK, while the US remains far below trend during the sample. On the other hand, the performance of TFP is much worse in the UK. TFP declines 6% from trend within the first three years in the UK, and around 10% by the end of the sample. In the

model. However, total debt is of course an equilibrium outcome, and the reductions could represent declines in demand rather than supply. Causal evidence from other sources suggests an important role for credit supply shocks originating from disruptions in the financial sector. See Chodorow-Reich (2014) for the US, and Franklin et al. (2015) for the UK.

[9]This is for consistency with previous work. The results are unchanged if the series are detrended by their long run growth rates. Throughout the paper I display all detrended quarterly data starting from 2008Q1, and all detrended yearly data starting in 2007.

Figure 2: Output, hours, and TFP during the Great Recession



Output and TFP are real and detrended by their 2000-2007 growth rates. Output and hours are per capita. All data is quarterly, and series are expressed as deviations from values in 2008Q1.

US, TFP actually rises slightly in the first couple of years, and remains elevated relative to trend throughout the sample. Thus, the main picture is of a deep recession in both countries, manifesting itself more in labour markets in the US, and more in TFP in the UK.

In the following sections, I refine my measures of output, hours, and TFP. In Appendix A.2 I show that the results for hours also hold for unemployment and Quality Adjusted Labour Input.

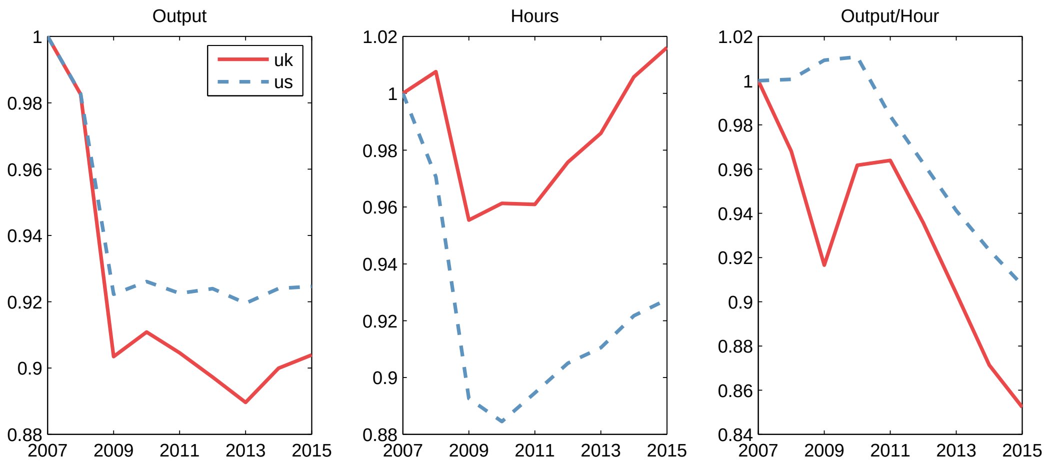
2.3 Excluding the financial sector

Since I am interested in how the reduction in lending from the financial sector affected the rest of the economy, I should strip out the financial sector from my aggregates. In this section I show that the stylised facts above also hold once the financial sector is removed.[10] The data is yearly, and uses employment data rather than hours. I construct hours data under the assumption that changes in hours per worker are identical across sectors, allowing me to apply the change in hours per worker at the aggregate level, for which data are available.

Excluding finance, the recessions in both countries are actually of a similar magnitude, as shown in the left panel of Figure 3. Output falls around 8% from trend in the US, and 10% in the UK. The falls are within the first two years, and output remains depressed for the rest of the sample. The fall in total hours is still worse in the US, as shown in the middle panel, with the magnitudes relatively unchanged. Since I do not have capital data at the sectoral level, I cannot compute TFP excluding the financial sector. I thus show a simpler productivity measure, output per hour, in the final panel. Here again, the declines are worse in the UK, with a peak difference of 8pp, and a final difference of around 6pp.

[10]Data details are in Appendix A.

Figure 3: Output, hours, and TFP excluding finance



Output and output per hour are real and detrended by their 2000-2007 growth rates. Output and hours are per capita. All data is yearly, and series are expressed as deviations from values in 20007.

2.4 Refining TFP measures

In this section, I show that the different TFP response of the two countries is robust to considering more sophisticated measures of TFP. TFP is essentially a residual productivity measure after controlling for whichever inputs are measured in the chosen production function. Since my model will offer one explanation of the TFP difference between the two countries (misallocation) it is important to control for any other potential causes when measuring TFP.

In this section I present the best available estimates for TFP for both countries. Data for the UK are from Goodridge, Haskell and Wallis (2015, henceforth GHW), who report the gap in TFP in 2011 relative to the 2000-2007 trend using yearly data. Labour is measured using QALI and capital broken into several subcategories. For the US data from Basu, Fernald and Kimball (2006) are constructed using a similar methodology.

In Table 1 I present the baseline TFP gaps in 2011, and the gaps after making a series of corrections. By this measure, TFP was 12.2% below trend in the UK in 2011, and only 3.84% in the US (a 8.36pp gap). After controlling for factor utilisation using BFK's method, the gap between the two countries remains large (6.86pp), as shown in the second column. After additionally removing the financial sector, the gap shrinks to 4.31pp.[11] Finally, GHW argue that increased capital scrapping post financial crisis could explain part of the TFP decline in the UK, and recalculate TFP assuming that capital depreciation rates increase by 25% from 2008 onwards. Repeating this exercise for the US[12] leads to a final gap in TFP of 4.31pp.

[11] For the UK, GHW compute TFP explicitly excluding the financial sector. For the US, I do not have access to sectoral capital data and instead use output per hour data: The 2011 gap for Y/H for all sectors is -1.88%, and the gap excluding finance is only 1.6%. Thus excluding finance increases Y/H by 14.45%, and I apply the same increase to the TFP number.

[12] Using the calibrated depreciation rate for the US, increasing the depreciation rate by 25% for three years reduces the capital stock by 5.07%, leading to an increase in measured TFP of $5.07\% \times \alpha = 1.81\%$.

Table 1: TFP gap in 2011 after adjustments

	TFP	+ utilisation adjust.	+ excl. finance	+ $\uparrow$ depreciation
UK	- 12.2%	-10.1%	-7.08%	-5.25%
US	- 3.84%	-3.24%	-2.77%	-0.96%

Private sector TFP gap in 2011 relative to 2000-2007 trend. UK numbers are from GHW. US TFP and utilisation adjusted numbers from BFK. The corrections are applied cumulatively from left to right.

2.5 Wages

I now turn to wage data, to establish that real wages in the UK fell further and faster than in the US. Given the issues in real wage measurement, such as composition and appropriate deflator, I discuss several measures of wages. I also decompose real wage movements into movements in nominal wages and inflation, and show that inflation dominates the differences in real wage movements, but that there is also a role for institutional factors. Data sources are discussed in Appendix A.

2.5.1 Real wages fall more in the UK

Figure 4 plots two measures of wages for each country. The left panel plots the average real wage calculated from the national accounts. This is constructed by dividing total labour income by total hours worked to construct an average wage. Consistent with the earlier data, it is deflated by the GDP deflator and expressed as a deviation from its 2000-2007 trend. By this measure real wages in the UK fell by 12% relative to trend in the six years after the crisis, while wages in the US fell by less than 8%. The differences are especially pronounced in the first three years of the crisis, where real wages in the UK fell by 7% and wages in the US by only 1%.

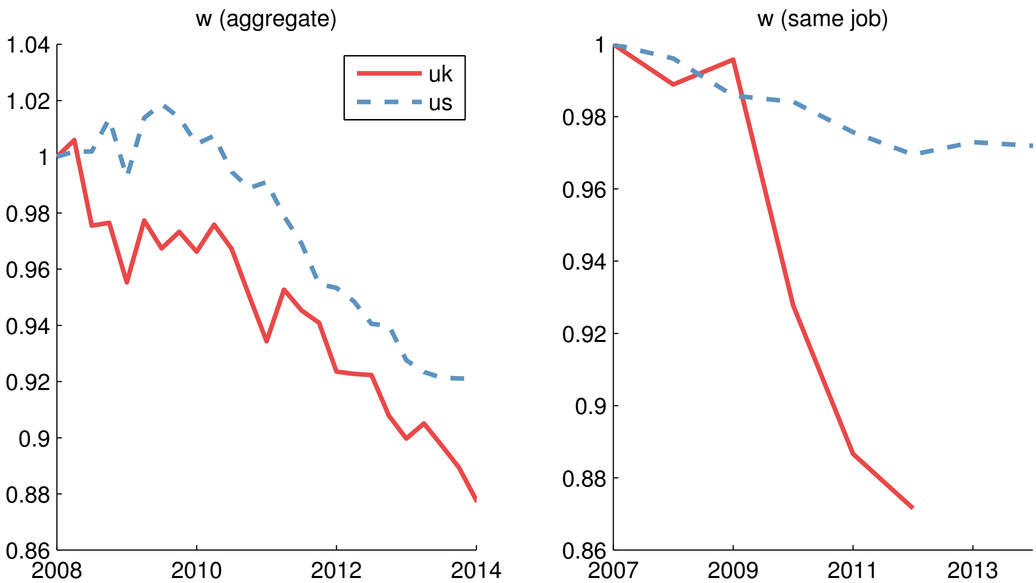
One issue with average wages constructed from aggregate data is that they ignore composition effects. Since job losses in recessions are typically concentrated in low wage workers, wages calculated from aggregate data are typically less cyclical than the wages of those who keep their jobs. This could exaggerate the difference between the US and the UK in aggregate wage data, since employment fell by more in the US. To address this concern, I turn to measures of wages adjusted for composition effects.

Daly and Hobijn (2016) use the Current Population Survey to construct a measure of wage growth adjusted for composition. They construct a measure of growth in median real wages within those who remain employed, deflated by the Personal Consumption Expenditures Price Index. I use this growth rate to construct a path for composition adjusted real wages, detrended by the pre-crisis growth rate, plotted in the right panel of Figure 4. Contrary to the aggregate wage series, this measure does show a reduction in real wages during the first three years of the crisis, but it is relatively muted at around 3%. Additionally, the series does not fall further during the last three years, as the aggregate series does, but instead plateaus.[13]

For the UK, Blundell, Crawford and Jin (2014) perform a similar exercise using the Labour

[13] Daly, Hobijn, and Pyle (2016) argue that the difference between the two series in the first three years stems from a composition effect as low wage workers disproportionately lose their jobs. In the second three years aggregate wages perform worse than composition adjusted because demographic changes start to kick in, with high-wage, older workers retiring.

Figure 4: Two wage measures during the Great Recession



Left panel shows quarterly real wage series calculated from the national accounts. This is an average wage measure, deflated by the GDP deflator. Right panel shows annual wages adjusted for workforce composition. US series is calculated by Daly and Hobijn (2016), and gives median wage of job stayers, and the UK series is calculated by UK, Blundell, Crawford and Jin (2014), and gives mean wages of job stayers. Both are deflated by a consumption price index. All are expressed as deviation from the 2000-2007 trend.

Force Survey. They construct a measure of growth in mean (*not* median) real wages within those who remain employed, deflated by the Retail Price Index. I again use this growth rate to construct a path for composition adjusted real wages, detrended by the pre-crisis growth rate. The greater magnitude of the fall in the UK is now even more pronounced. Although their calculations are only available up to 2012, by this time the real wage has fallen over 12% from trend.

One other issue in the measurement of real wages is how to deflate the nominal wage series. The result that real wages fell further in the UK is robust to the choice of deflator. The two series used above already used two different deflators: GDP and consumption. One can argue that producer prices could be a more appropriate choice, since the resulting real wage series than more closely resembles the real cost of labour to firms. If the aggregate wage series are deflated using each country's Producer Price Index the same result holds. Thus, overall, looking at several wage measures, I find that wages fell faster and further in the UK than in the US during the Great Recession.

2.5.2 Nominal wage stickiness, inflation, and institutional factors

I now decompose the real wage series into movements in nominal wages and inflation to investigate the underlying sources of the wage differences between these two countries. I show that the real wage differences are predominantly driven by differences in inflation, but that institutional factors may also have played a role. In this section I focus on my measure of aggregate wages, and the price index used is thus the GDP deflator.

Figure 5 plots nominal wages and the price index for each country. Neither series is detrended, and I include six years of pre-crisis data for comparison. Both series are normalised to one in 2008. The first thing to note from the left panel is the similarity in the nominal wage series over the whole sample, even during the Great Recession. Nominal wages were growing at the same rate in each country pre-crisis, and both slowed post crisis, without a persistent gap opening up between the series post-crisis.

On the other hand, in the right panel we see a wedge opening up between the price levels of the two countries almost immediately post-crisis, which has persisted throughout the sample. This wedge is caused by a slowdown in inflation in the US during the first few years of the crisis, while inflation in the UK continued at its pre-crisis rate, even increasing slightly. Combined, this means that differences in real wages must be explained mostly by inflation over this period, since nominal wages have behaved similarly in the two countries.

This is not the entire picture, however, because we do see a gap between nominal wages in the two countries appear during the first two years of the crisis. According to this measure, nominal wages in the UK actually fell, while they continued growing in the US. This is eventually corrected by faster growth in the UK, but during this initial period around 50% of the difference between real wages in the two countries can be explained by nominal wages.

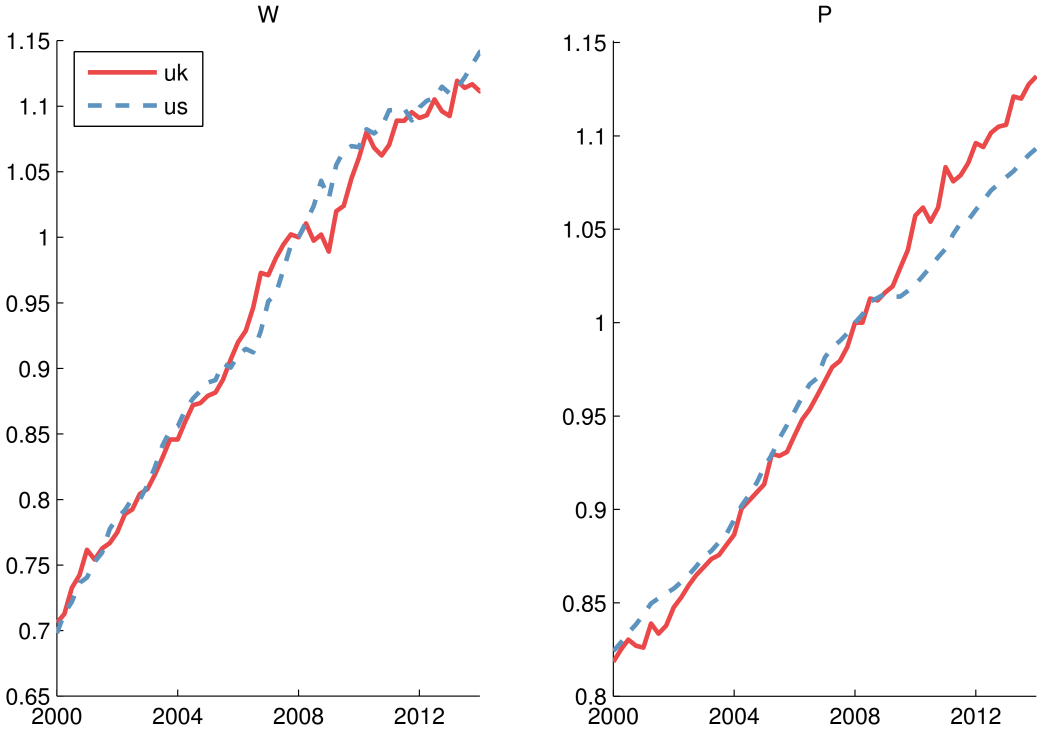
The dominant role of inflation differences clearly brings to mind a story of nominal wage stickiness coupled with differential inflation experiences. If one is willing to entertain that both countries suffer from similar degrees of downwards nominal wage rigidity then a compelling explanation of the differences in the real wage is that the UK was lucky, and ran higher inflation which eroded its real wages.

However, the fact that nominal wages also appeared to fall more in the UK during the early years suggests that nominal wage stickiness is not the entire story. An alternative is differences in *real* wage rigidity in the two countries during this period. In the UK, there has been a trend towards increasing labour market flexibility since the 1980s, which has been relatively untested since the last recession before the Great Recession was the 1990-91 recession.[14] In the US, on the other hand, some papers (Hagedorn, Karahan, Manovskii, and Mitman, 2015, Hagedorn, Manovskii and Mitman, 2015) have argued that the unprecedented extension of unemployment benefits during the Great Recession could have put upwards pressure on wages, by increasing the outside option of workers. These policies are discussed further in Appendix A.

Of course, the above policy changes only say that wages in the US could have become more rigid, while those in the UK have become less so. They do not say that wages in the US have become *more rigid than* those in the UK. However, combined with the differential inflation experiences of the two countries, they paint an interesting picture of the forces that could have pushed us in that direction.

[14]Blundell, Crawford and Jin (2014) summarise these changes, which include "the increasing number of welfare-to-work programmes available to jobseekers, the more stringent job search conditions attached to benefits claimed by the unemployed, those with disabilities and lone parents, and, more recently, the increase in the state pension age for women".

Figure 5: Nominal Wages and Prices during the Great Recession



Left panel gives average nominal wage calculated from the national accounts, and right panel the GDP deflator price index. Neither are detrended, and all series are normalised to one in 2008Q1.

2.6 Job flow measures

In this section I look at measures of worker flows during the Great Recession in both countries. I use data on flows between employment, unemployment, and inactivity. Data sources and methodology are detailed in Appendix A. I am interested in the evolution of the employment rate. I additionally construct two counterfactual series representing the contribution of job destruction rates and creation rates respectively. These are constructed by starting from the 2008 stocks, and simulating forward while allowing only certain flows to vary, while holding the others at their pre-2008 averages. Counterfactual employment allowing only job destruction to vary is calculated using the true rates for transitions *from* employment while holding the other flows constant. Counterfactual employment allowing only creation rates to vary is calculated using the true rates for all other transitions while holding the job destruction rates constant.[15]

[15]The inclusion of the inactivity state complicates the analysis from simple comparison of $e \rightarrow u$ and $u \rightarrow e$ flows, but is crucial to match the data. The result that destruction flows account for similar employment falls in both countries is unaffected by considering flows from unemployment or inactivity separately.

2.6.1 Results

The results are plotted in Figure 6, along with the actual data. While this data uses employment and not hours, the basic pattern that the fall in employment was larger and more persistent in the US remains. The solid black line gives the data, with a maximum fall of over 7% for the US, and less than 4% for the UK. Additionally, in both countries the increase in the job destruction rate alone can explain little of the employment fall, as shown by the blue dashed lines. Simulating employment allowing only the job destruction rates to vary leads to a decline in employment which never exceeds 2% in either country. The bulk of the employment fall is driven by the decline in job creation in both countries, and thus the difference between the two is driven by the fact that job creation fell by much more in the US. This is shown by the solid red line, which simulates allowing only creation rates to change, and tracks the true employment rate more closely.

2.6.2 Labour hoarding

The key role of job creation is evidence against the idea that the difference in employment changes between the two countries was driven by higher firing costs or labour hoarding in the UK. If that were the case, we would expect the difference between the two countries to be driven by differences in job destruction, while the opposite is true. Additionally, since the magnitude of the employment fall driven by increases in destruction is similar in both countries, and the employment fall was smaller in the UK, job destruction actually accounts for a higher fraction of the employment change in the UK.

One factor these graphs don't show is that labour flows are higher *on average* in the US. Typically, workers transition in and out of unemployment around 50% faster in the US than in the UK.[16] This means that, even if employment changes in both countries were driven entirely by changes in job creation, workers would remain in old jobs longer in the UK, since they are fired more slowly. One could think of this as there being higher labour hoarding in the UK "on average".

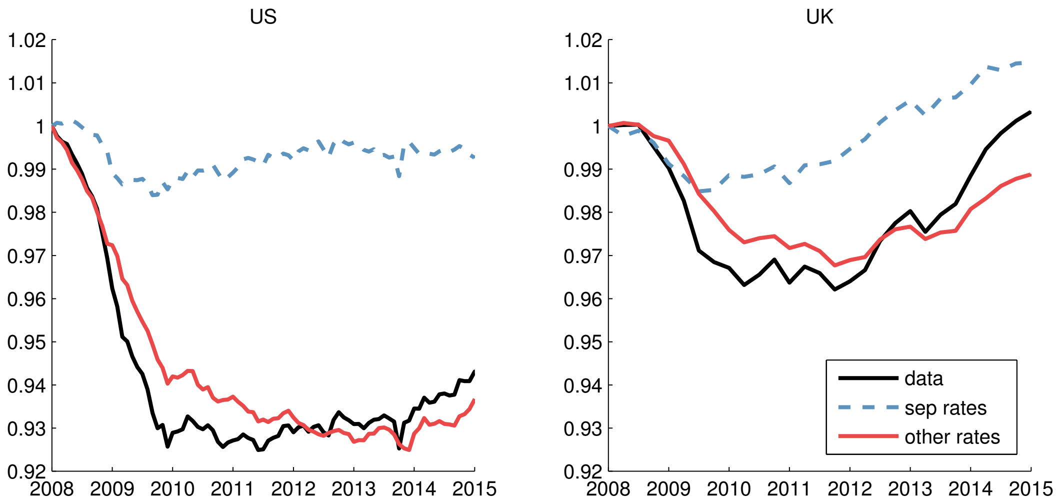
This could explain why TFP fell more in the UK, since workers were reallocated faster in the US, perhaps to more productive jobs. However, this story also does not hold up to the time series data. If slower flows were the cause of the muted employment response and TFP fall, relative to the US, we should expect the UK to "catch up" to the US during the crisis. Workers should eventually be fired, as they were in the US, reducing employment and reversing the fall in TFP. Instead, we see the opposite: a persistent fall in TFP, and a labour market which recovers faster than the US.

2.7 Evidence of misallocation

In this paper I argue that the differences in the aggregate TFP responses between the two countries are driven in part by differences in the allocation of resources across firms. In this section I investigate the extent to which misallocation was greater in the UK using worker- and firm-level data. I show that the UK reallocated more workers than the US during the crisis, and review evidence that misallocation of resources across firms increase more in the UK.

[16] For a detailed discussed including corrections for different data frequency between the two countries, see Gomes (2012).

Figure 6: Labour flows



Both panels plot the employment to population ratios as deviations from 2008 values. The (darker) black line plots the data, the solid (lighter) red line plots the simulated path holding employment separation rates constant and taking all other rates from the data. The dashed blue line plots the simulated path taking employment separation rates from the data and holding all other rates constant.

2.7.1 Job flows and reallocation

Davis and Haltiwanger (1992) argue that the sum of the job creation and destruction rates is a measure of reallocation. This is because a job which is destroyed and then created elsewhere represents a worker who has moved. By this measure, we see a difference in the worker reallocation patterns during the Great Recession between the two countries. While job destruction rates increases similarly in both, job creation fell by more in the US, leading to a relative increase in reallocation of labour in the UK during this period.

Intuitively, similar fractions of people were fired in both countries, but they remained unemployed in the US, while they were reallocated to new jobs in the UK. Of course, reallocation could be either productivity enhancing or misallocation inducing. In the next section I review evidence suggesting that this reallocation was indeed misallocation.

2.7.2 Firm-level evidence

Moving on to firm-level data, while differences in data sources and methodology across studies make an exact comparison challenging, the available work does suggest that misallocation worsened in the UK, while it did not in the US. Here I focus on papers which use firm-level data to decompose changes in aggregate productivity into changes in productivity within firm, and changes stemming from the reallocation of resources from low to high productivity firms.

For the UK, Barnett et al. (2014) combine the Annual Business Survey and Annual Business Inquiry data with the Inter-Departmental Business Register dataset to construct a population-level dataset for the UK economy. They find that one third of the slowdown in labour productivity growth since the Great Recession can be explained by "less – or less effective – reallocation of labour, and slower 'creative destruction'" (p22). They find that reallocation of labour led to productivity growth of 1.2% per year before the crisis, dropping to 0.5% post-crisis. Additionally, net entry of firms used to lead to productivity growth of 1.7% pre-crisis, falling to 0% post-crisis.[17] This totals to a drop of 2.4pp, around 1/3 of the 6pp slowdown in their dataset. There was little change in the capital-labour ratio during the crisis, so these results also hold for TFP.

Using a different dataset and methodology, Bordon et al. (2016) find similar numbers. They use the Orbis database of firms, and compute Hsieh and Klenow's (2009) measure of misallocation. They find that TFP was 3% lower post crisis in the UK due to increased misallocation, and that this can explain 1/3 of the slowdown in labour productivity.

To my knowledge, no comparable study using population-level data exists for the US. The closest work is that of Kurtzman and Zeke (2016), who use data on public firms from the Compustat database. They additionally use a different decomposition method, which they develop, from the standard one used by Barnett et al. (2014), and only study continuing firms, removing the ability to look at net entry. These issues aside, they find essentially no role for misallocation across firms during the Great Recession in explaining labour productivity or TFP. Reallocation contributes negatively to productivity growth both pre- and post-crisis, with the contribution becoming, if anything, more positive during the Great Recession.

Thus, the available evidence suggests a greater role for misallocation during the Great Recession in the UK than in the US. Future work should address the challenging task of constructing internationally comparable data to strengthen this conclusion.

2.8 Summary

Overall, the data paint a clear picture of a recession accompanied by steep declines in firm borrowing in both countries, with the recession manifesting itself more in labour markets in the US, and TFP in the UK. Additionally, wages fell further and faster in the UK, which is driven largely by higher inflation in the UK over this period. Labour flows data are not supportive of the hypothesis that the employment and TFP patterns are driven by higher labour hoarding in the UK. On the other hand, there is evidence of greater misallocation across firms in the UK. In the rest of the paper, I turn to explaining these facts.

3 Analytical Results

In the remainder of the paper I build a model to explain the above stylised facts. In this section I set up a simplified version of the model in which I am able to prove analytical results. I then move on to a richer version which I use to quantitatively test the model.

3.1 Environment

The model in this section is static, and has several key features. Firstly, firms have heterogeneous net worths and productivity. Secondly, they face a simple financial friction which limits their ability to purchase capital. A fraction of firms will endogenously emerge as financially constrained depending on their productivity and net worth. Thirdly, wages are partially sticky. Finally, I introduce

[17] Specifically, during the crisis, entrant firms became less productive, and exiting firms became more productive.

imperfect substitutability across goods via a CES aggregator and monopolistic competition at the firm level.

The model adapts features of Buera and Moll (2015) and Khan and Thomas (2013). The focus on analytical results, wedges, and simple borrowing limits comes from Buera and Moll, while the addition of monopolistic competition brings decreasing returns to scale (in revenue) at the firm level, leading to well defined firm size distributions. Despite the relatively general framework (for example, I can assume an arbitrary joint distribution over productivity and net worth) I am able to prove analytical results.

There is a single time period, but alternatively, this can be thought of as a two period model without discounting. There is a representative household who consumes a final good and supplies labour. A representative final-goods producer produces the final good, which is used for consumption and production of capital, from a bundle of intermediate goods. These are produced by a continuum of intermediate-goods producing firms, who are the main firms of interest. They produce via a constant returns to scale production function in capital and labour, have heterogeneous productivities and are financially constrained.

The exercises presented in this section are comparative statics exercises, studying the change in the equilibrium in response to changes in parameter values.

3.2 Final good producer

There is a numeraire final good, which can be converted one for one into consumption and investment. This is produced by a representative firm with production function:

$$Y = \left(\int_0^1 y_i^\rho \,\mathrm{d}i\right)^{\frac{1}{\rho}} \tag{1}$$

Where I define $\epsilon = 1/(1-\rho)$ as the elasticity of substitution between varieties. Final output, Y, is produced from a CES bundle of intermediates goods y_i, each produced by a different intermediate firm. The final-goods firm is a price taker in both the final and intermediates markets. Profit is given by:

$$\pi = \left(\int_0^1 y_i^\rho \,\mathrm{d}i\right)^{\frac{1}{\rho}} - \int_0^1 p_i y_i \,\mathrm{d}i \tag{2}$$

p_i is the price (in terms of numeraire) of intermediate i. The first order condition for intermediate purchase from i gives:

$$y_i = p_i^{-\epsilon} Y$$

3.3 Intermediate good firm

Intermediate firms produce their good via a CRS production function $y_i = z_i k_i^\alpha l_i^{1-\alpha}$, where z_i is firm-level productivity, k_i is capital and l_i labour. They take the wage as given, and can pay for labour using their production. Productivity is heterogeneous across firms, and known at the time inputs are chosen.

In this one-period set up, I also allow firms to purchase capital, which becomes immediately productive.[18] However, I assume that capital must be paid for in advance of the receipt of revenues

[18] Alternatively, this model is equivalent to a two-period model with no discounting and one-period time to build in

from the final good, and hence must be purchased using the firm's initial wealth, n_i, and within-period borrowing from the household, d_i. Firms have heterogeneous initial wealth and are assumed to be able to borrow at most an exogenous multiple $\lambda - 1$ of this from the household: $d_i \leq (\lambda - 1)n_i$. Since the borrowing is within period, and there is no option of default, the firm pays no interest on this borrowing. The balance sheet constraint gives $k_i = d_i + n_i$, and total spending on capital thus cannot exceed $(\lambda - 1)n_i + n_i$, giving $k_i \leq \lambda n_i$.[19] Capital is produced one-for-one from the final good, giving a fixed price of one. Since this is a one period model, capital becomes useless after production an has resale price of zero.[20]

Firms are thus heterogeneous over both productivity, z_i, and initial wealth, n_i, and the distribution of these variables over firms is represented by the density function $f(n, z)$. In this one-period set up the source of initial net worth is unmodelled, while in a dynamic framework this would come from retained earnings, which could introduce correlation between net worth and productivity. My results hold for any arbitrary joint distribution over productivity and net worth, so are consistent with any mechanism by which net worth evolves.

Firms are monopolistically competitive, and understand that the amount they produce will affect their price according to the final good firm's first order conditions, which imply $p_i = \left(\frac{Y}{y_i}\right)^{1-\rho}$. Revenue, in terms of the numeraire, is thus $p_i y_i = y_i^{\rho} Y^{1-\rho}$. The firm pays out all earnings at the end of the period as a dividend, e_i, and is assumed to maximise dividends. End of period earnings are revenues less the wage bill and the repayment of debt: $e_i = p_i y_i - w l_i - d_i$. Combining this with the balance sheet condition, the firm's optimisation can be represented as a profit maximisation problem:

$$\pi_i = \max_{l_i \geq 0, 0 \leq k_i \leq \lambda n_i} z_i^{\rho} k_i^{\alpha\rho} l_i^{(1-\alpha)\rho} Y^{1-\rho} - w l_i - k_i \tag{3}$$

Where $e_i = \pi_i + n_i$, and maximising profits is equivalent to maximising earnings. The solution to this optimisation differs depending on whether perfect ($\rho = 1$) or imperfect ($\rho < 1$) substitutability is assumed. I focus on the case $\rho < 1$, and discuss $\rho = 1$ separately in Section B.2.2. The solution can be split into two regions, depending on whether the borrowing constraint binds. If the borrowing constraint does bind then the optimal capital choice is given by:

$$k_i^* = (\alpha\rho)^{\frac{1}{1-\rho}} \left(\frac{1-\alpha}{\alpha}\right)^{\frac{\nu}{1-\rho}} Y z_i^{\frac{\rho}{1-\rho}} w^{\frac{-\nu}{1-\rho}} \tag{4}$$

Where k_i^* is used to denote the unconstrained optimal choice, and $\nu \equiv (1 - \alpha)\rho$. This is not feasible if it violates the borrowing constraint, leading to a capital policy function:

$$k_i = \begin{cases} k_i^* & \text{if } k_i^* \leq \lambda n_i \\ \lambda n_i & \text{otherwise} \end{cases} \tag{5}$$

capital.

[19] This simple borrowing limit arises from a limited commitment problem where firms cannot commit to repay borrowing, which can only be collateralised against a fraction of capital. In the robustness section I discuss the robustness of my results to more general borrowing constraints.

[20] The static model and lack of time to build implies that part of the capital stock used today will be produced using today's output. This peculiarity is easily resolved in the two-period interpretation of the model.

Given this value of capital, output and labour are given by:

$$y_i = \left(\frac{\nu Y^{1-\rho}}{w}\right)^{\frac{1-\alpha}{1-\nu}} z_i^{\frac{1}{1-\nu}} k_i^{\frac{\alpha}{1-\nu}} \tag{6}$$

$$l_i = \left(\frac{\nu Y^{1-\rho}}{w}\right)^{\frac{1}{1-\nu}} z_i^{\frac{\rho}{1-\nu}} k_i^{\frac{\alpha\rho}{1-\nu}} \tag{7}$$

3.4 Aggregation

Intermediate outputs, y_i, are aggregated into the final good according to the CES technology (1). Aggregate capital and labour are found by integrating over firms:

$$K \equiv \int_0^1 k_i \,\mathrm{d}i = \int k \,\mathrm{d}f(n,z), \qquad L \equiv \int_0^1 l_i \,\mathrm{d}i = \int l \,\mathrm{d}f(n,z) \tag{8}$$

The model doesn't aggregate in the sense that we can use a representative firm. However, we can do some useful partial aggregation to get expressions for output, TFP, and labour in terms of the distribution of productivity and capital. Derivations are given in Appendix B. In particular, taking firms' capital levels as given and using the optimality condition for labour demand, (7), we can aggregate the firms with three equations. A standard production function:

$$Y = ZK^{\alpha}L^{1-\alpha} \tag{9}$$

Here Z is defined as the residual in a Cobb Douglas production function, the standard measure of aggregate TFP comparable to the data presented in the last section. Labour demand aggregates to:

$$wL = (1-\alpha)\rho Y \tag{10}$$

That is, despite the heterogeneity, and DRS in revenue, labour share is constant at $(1-\alpha)\rho$ in this model, giving an equation similar to the standard $wL = (1-\alpha)Y$ from a representative firm model. Finally, using the firm level optimal output and labour choices, (6) and (7), the definition of L, and the aggregator (1), we can show that aggregate TFP is given by

$$Z = \left(\int_0^1 z_i^{\frac{\rho}{1-\nu}} \tilde{k}_i^{\frac{\alpha\rho}{1-\nu}} \,\mathrm{d}i\right)^{\frac{1-\nu}{\rho}} = \mathrm{E}_{n,z}\left[z^{\frac{\rho}{1-\nu}} \tilde{k}^{\frac{\alpha\rho}{1-\nu}}\right]^{\frac{1-\nu}{\rho}} \tag{11}$$

Where for each firm $\tilde{k} = k/K$ is the ratio of the capital stock operated by that firm to the total capital stock. $\mathrm{E}_{n,z}$ refers to the expectation taken over (n,z), and it is understood that $\tilde{k}$ is a function of (n,z) according to the optimal choices derived above.

This expression will be central to the results, since it concerns how the distribution of resources across firms of varying productivities affects measured aggregate productivity. Note that aggregate productivity takes the form of an average of firm productivities, weighted by a power of their relative size.

3.5 Household

Since the focus of the model is on the firm side, the household structure is left stylised. There is a representative household with utility function $U(C, L^s) = u(C) - v(L^s)$ where C is consumption and L^s hours worked. Consumption utility satisfies $u'(C) > 0$ and $u''(C) \leq 0$, and labour disutility satisfies $v'(L^s) > 0$ and $v''(L^s) \geq 0$, as is standard. The household receives labour income and net earnings paid out by firms. The household's budget constraint is:

$$C = wL^s + E + N_h \tag{12}$$

Where $E = \int_0^1 e_i \, \mathrm{d}i$ the net income received from firms, and N_h is the initial net worth of the household. Total lending to firms, D, is within period with no interest payments, and hence does not affect the household's budget. The household takes wages as given, and chooses how much labour to supply. Since I assume a degree of wage stickiness, and given the simple labour market structure, we might end up with labour market rationing in some states of the world. The household understands this and takes it into account in its optimisation. If the household is not rationed in its labour supply in equilibrium, then its labour supply first order condition must hold:

$$\frac{v'(L^s)}{u'(C)} = w \tag{13}$$

I restrict to preferences such that equilibrium aggregate labour supply is upwards sloping in the wage. However, if in equilibrium the market doesn't clear the household understands that it will only be able to work $L^s \leq L$ hours, where L is aggregate labour demand from firms (equal to observed labour), and the labour supply condition will not hold.

3.6 Market clearing

The initial net worth of firms, n_i, and the household, N_h, consists of any (unmodeled) initial claims between them, and claims to the pre-existing undepreciated capital stock, $(1 - \delta)K_0$. Since these initial claims must net out, we have that $N_h + \int_0^1 n_i \, \mathrm{d}i = (1 - \delta)K_0$. Combining this with the household's budget constraint and firms' balance sheets and earnings constraints yields the goods market clearing condition:

$$C + K = Y + (1 - \delta)K_0 \tag{14}$$

Since I assume sticky wages, the labour market may not clear, and I replace the labour market clearing condition with a flexible parameterisation of sticky wages. Define the market clearing wage as the one which satisfies the household's labour supply FOC:

$$w_{mc} = \frac{v'(L)}{u'(C)} \tag{15}$$

Then the equilibrium wage is assumed to satisfy:

$$w = w_0^{\gamma} w_{mc}^{1-\gamma} \tag{16}$$

Where w_0 is the reference wage level around wage wages are sticky. $\gamma = 1$ corresponds to fully rigid wages, which are unresponsive to equilibrium in comparative statics exercises, and $\gamma = 0$ corresponds to fully flexible wages. If the labour market is rationed, unemployment is defined as $U \equiv L^s - L$.

3.7 Wedges

I will be interested in the model's implications for the Chari et al. (2007) wedges, which I define here. The efficiency wedge, τ^e, is just endogenous TFP, Z. The labour wedge, τ^l, is defined, as in the empirical exercises, as the wedge between the marginal rate of substitution between labour and consumption and the marginal product of labour:

$$\frac{v'(L)}{u'(C)} = (1-\tau^l)(1-\alpha)\frac{Y}{L} \tag{17}$$

3.8 Definition of equilibrium

I focus on equilibria where the labour market does not clear, labour supply being rationed, leading to unemployment. All other markets clear, and I define the notion of a labour market rationing equilibrium below:

Definition 1. *A labour market rationing equilibrium is an allocation $(Y, C, L, D, K, U, L^s, Z)$, price, w, and functions, (k, d, y, l), such that, for given initial conditions, (f, N_h, K_0, w_0), and a given borrowing constraint, λ:*

1. *The policy functions, (k, d, y, l), all functions of (n, z), solve the firm's problem and are defined by (5), (6), (7), and $d = k - n$.*
2. *Output is given by the aggregator, (1). Aggregate labour and capital are defined by (8). Aggregate TFP is defined as the residual in (9).*
3. *The goods market clears: (14).*
4. *The bond market clears: $D = \int d df$.*
5. *The equilibrium wage satisfies (16).*
6. *L^s is given by $v'(L^s)/u'(C) = w$, and labour market disequilibrium by $U = L^s - L$.*

3.9 Firm-level productivity measures

There are two measures of productivity at the firm level: Quantity Total Factor Productivity (TFPQ) and Revenue Total Factor Productivity (TFPR). TFPQ measures the productivity at which output is produced, while TFPR measures the productivity at which revenue is produced:

$$TFPQ_i \equiv \frac{y_i}{k_i^\alpha l_i^{1-\alpha}} = z_i \tag{18}$$

$$TFPR_i \equiv \frac{p_i y_i}{k_i^\alpha l_i^{1-\alpha}} \propto z_i^{\frac{\alpha\rho}{1-\nu}} k_i^{\frac{-\alpha(1-\rho)}{1-\nu}} \tag{19}$$

Where the proportionality for $TFPR_i$ refers to proportionality across firms holding an aggregate scaling factor constant. The proportionality result holds for a given level of capital at each firm, once labour is chosen optimally. Note that TFPQ is exogenous, while TFPR is endogenous, responding to firm size.

3.10 Efficient allocation

Before moving on to the results, it is instructive to understand the efficient allocation. A necessary feature of the efficient allocation is that it maximises output, Y, for a given amount of inputs of total capital, K, and labour, L. Maximising (1) for given total inputs gives the following requirements, which I break down into two steps. Firstly, *given an amount of capital allocated to each firm*, labour should be allocated according to:

$$l_i \propto z_i^{\frac{\rho}{1-\nu}} k_i^{\frac{\alpha\rho}{1-\nu}} \tag{20}$$

Firms with higher TFP and more capital should be given more labour. Secondly, capital should be allocated across firms according to:

$$k_i \propto z_i^{\frac{\rho}{1-\rho}} \tag{21}$$

Firm level capital should be proportional to a power of firm-level TFP.[21] If both capital and labour are allocated optimally then one can show that aggregate TFP is independent of total K or L, and is given by:

$$Z^* = \mathrm{E}_z \left[z^{\frac{\rho}{1-\rho}} \right]^{\frac{1-\rho}{\rho}} \tag{22}$$

3.11 Equalisation of TFPR across firms

Hsieh and Klenow (2009) show that in the absence of any distortions, TFPR should be equalised across firms. I show that this is the case in my model, and that financial frictions distort TFPR in a particular way:

Lemma 1. *In the efficient allocation, and model without financial frictions ($\lambda \to \infty$), $TFPR_i$ is equalised across firms. In the model with financial frictions ($\lambda < \infty$), $TFPR_i$ is equalised for all financially unconstrained firms, and higher for financially constrained firms.*

Proof. For the efficient allocation and model without financial frictions, rearranging the optimal capital choice ((21) and (4) respectively) yields that $z^{\frac{\rho}{1-\nu}} k_i^{-\frac{1-\rho}{1-\nu}} = c$ for some c. Comparing this to (19) reveals that $TFPR_i$ must be equal across firms. In the model with financial frictions, doing the same exercise with (4) reveals that $TFPR_i$ should be equalised among unconstrained firms. For constrained firms, note that, by nature of being constrained, their capital must be below the optimal choice of an unconstrained firm with the same level of z_i. Since $-\frac{1-\rho}{1-\nu} < 0$ this implies that $z^{\frac{\rho}{1-\nu}} k_i^{-\frac{1-\rho}{1-\nu}} > c$. Using (19) this implies that $TFPR_i$ is higher than for unconstrained firms. □

Intuitively, in the efficient allocation TFPR is equalised across firms. Financial frictions force any constrained firms to be too small, which pushes up their revenue by restricting supply of their good and increasing its price. This artificially increases TFPR of any constrained firms relative to unconstrained firms.

3.12 TFP in the presence of financial frictions

How do these requirements for the distribution of resources in the efficient allocation compare to the conditions which emerge in the equilibrium of my model? Comparing firms' labour choices,

[21] Combining the two one show that if capital is optimally allocated, we also require that $l_i \propto z_i^{\frac{\rho}{1-\rho}}$.

(7), with (20) reveals that *conditional on the distribution of capital*, labour is allocated optimally in equilibrium: as in the efficient allocation, labour is chosen to be proportional to $z_i^{\frac{\rho}{1-\nu}} k_i^{\frac{\alpha\rho}{1-\nu}}$. Deviations from the maximised TFP, Z^*, will occur therefore because capital is not correctly allocated.

It is helpful to ask first what would happen in the absence of financial frictions (the limiting case where $\lambda \to \infty$). In this case, all firms are unconstrained in their capital choice and choose capital according to (4), and thus choose capital to be proportional to $z_i^{\frac{\rho}{1-\rho}}$. Comparing this to (21), we see that this is the condition for capital to be distributed efficiently. Hence, in the absence of financial frictions resources will be efficiently allocated in my model, and TFP will always be equal to the maximised value Z^*.

In the presence of financial frictions, TFP might differ from Z^* since capital will no longer be allocated efficiently. The capital policy function is now given by (5), which is no longer proportional to $z_i^{\frac{\rho}{1-\rho}}$ due to the presence of the financial friction, which limits capital to be equal to λn_i for some firms.

How do financial frictions affect TFP? To understand this, it is useful to revisit the expression for TFP for a given distribution of capital, (11). Differentiating this expression with respect to the capital share of any one firm gives:

$$Z = \mathrm{E}_{n,z}\left[z^{\frac{\rho}{1-\nu}}\tilde{k}^{\frac{\alpha\rho}{1-\nu}}\right]^{\frac{1-\nu}{\rho}} \Rightarrow \frac{\partial Z}{\partial \tilde{k}_i} \propto TFPR_i^{\frac{1}{\alpha}}$$

This measures the effect on TFP of marginally increasing any firm's share of the capital stock. Intuitively, this is proportional a firm's TFPR: all else equal, giving more capital to a firm with higher TFPR will increase aggregate TFP.

Note that the fact that TFPR is equalised across firms in the efficient allocation implies that, on the margin, aggregate TFP is unaffected by moving capital across firms. This is precisely the efficiency condition for maximising aggregate TFP, otherwise aggregate TFP could be increased by reallocating resources.

In the model with financial frictions, this is no longer the case. As was previously shown, financially constrained firms (a) have higher TFPR than financially unconstrained firms, and (b) have been kept too small, by definition. This means that aggregate TFP could be increased by reallocating resources towards constrained firms.

3.13 Results

In this section I present the main results. Note that the proofs presented are for the case of less-than-perfect substitution between varieties ($\rho < 1$), but identical results hold for the case of perfect substitution ($\rho = 1$).[22] All omitted proofs are relegated to Appendix B.1.

A financial crisis is modelled as a tightening of firms' borrowing constraints, i.e. a reduction in λ. I first present two partial equilibrium results, which are interesting in their own right and are later combined in general equilibrium. The first result concerns the response of productivity and hours to a financial crisis holding the wage constant. The result respects all market clearing and equilibrium conditions except the wage determination equation, (16), and thus ignores the labour supply curve.

[22] The intuitions in this case are slightly different, and are discussed along with the proofs in Appendix B.2.

Proposition 1. *Allow for an arbitrary* (n, z) *distribution. In partial equilibrium with a fixed wage, ignoring (16),* $\partial Z/\partial\lambda = 0$ *and* $\partial L/\partial\lambda > 0$. *Thus, following a tightening of borrowing constraints, aggregate productivity remains constant and hours fall.*

I relegate the proof of this proposition to Appendix B.1, and here provide intuition and a graphical exposition. All of the proofs are given for infinitesimal changes, but one can show that they also hold for discrete changes. The proof that a financial shock leads to no change in aggregate TFP if wages are constant ($\partial Z/\partial\lambda = 0$) relies on showing that a financial crisis leads to no change in the relative capital, $\tilde{k}_i$, held by any firm. This then trivially implies no change in Z according to (11). Intuitively, if all firms shrink by the same proportion (regardless of whether they are financial constrained) then there is no change in the distribution of resources across firms, only the average size of all firms.

This result is illustrated for a discrete change in λ in Figure 7. The left panel plots the optimal capital choices of firms across their net worth levels, n. I do this for a given level of productivity, alternatively one can think of this as a simplification of the model with just a single productivity level. The line k_0 plots the capital policy function for an initial borrowing constraint, λ_0. For low n firms are financially constrained, and capital increases linearly with their net worth. Once net worth reaches $\tilde{n}_0$ they have enough net worth to purchase the unconstrained optimal level of capital, k_0^*. The right panel plots $\tilde{k}$ across net worth levels, which is just the capital policy from the left panel scaled by (endogenous) aggregate capital, K_0.

I now consider a financial shock, tightening the borrowing constraint to $\lambda_1 < \lambda_0$. The line k_1 gives the hypothetical new capital policy function *ignoring that unconstrained firms might want to adjust their optimal capital.* In this case, constrained firms must reduce their capital to $\lambda_1 n < \lambda_0 n$. This implies that the point at which you become unconstrained will shift right, to $\tilde{n}_1$. Total capital will shrink to $K_1 < K_0$, meaning that unconstrained firms (who didn't adjust their capital yet) would end up with a higher share of the capital stock: $\tilde{k}_1^* = k_0^*/K_1 > k_0^*/K_0 = \tilde{k}_0^*$. Similarly, constrained firms end up with a lower share of the capital stock, represented by the shift in the line $\tilde{k}_1$.

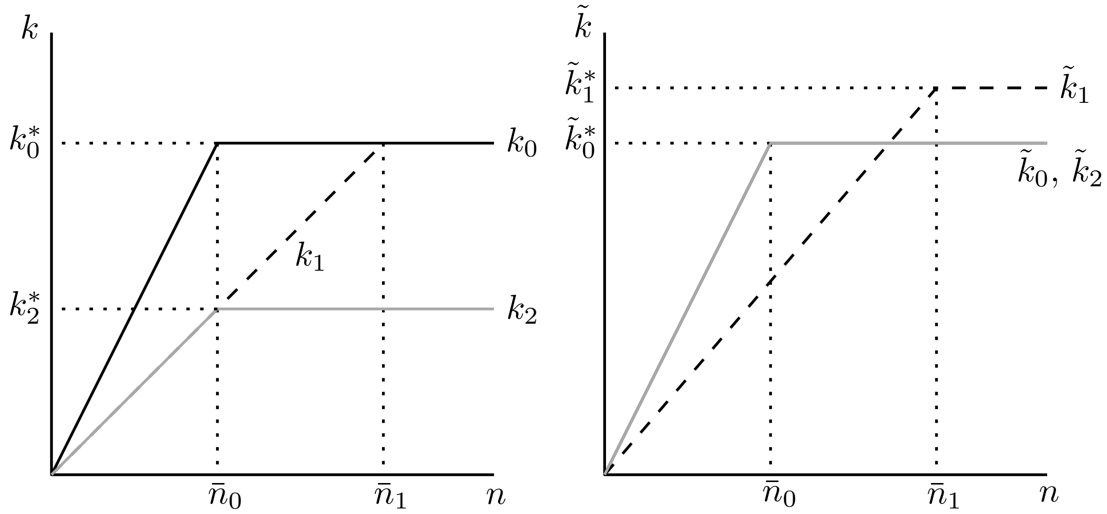
However, once we consider that unconstrained firms will choose to adjust their capital, all of these effects disappear. Without any change in the wage, there is a *100% spillover of the financial shock to unconstrained firms.* Even though unconstrained firms are not directly affected by the change in the financial friction, they are indirectly because it lowers aggregate output, lowering the demand for their product via imperfect substitutability: Holding their price fixed, their demand falls linearly with GDP according to $y_i = p_i^{-\epsilon} Y$. This leads unconstrained firms to optimally lower their size (in terms of output, capital, and labour) by exactly the same proportion as constrained firms.[23] This leads the true policy function after the financial shock to be k_2. Since all firms shrink by the same proportion, they all retain the same share of the capital stock, and the shares return to their original values, meaning that the lines $\tilde{k}_2$ and $\tilde{k}_0$ are identical. With no change in the relative distribution of resources across firms, there can be no change in aggregate TFP.

To prove the second part of the proposition (that total labour falls following a negative financial shock) simply note that total output falls since all firms shrink, and that with a fixed wage this implies that total labour falls according to (10).

The, perhaps surprising, implication of this proposition is that a financial crisis does not by itself

[23] This result holds for any degree of substitutability apart from perfect substitutability, in which case the same results go through via a different mechanism. The sharpness of the result comes from the assumption of a constant elasticity of substitution, not the specific elasticity.

Figure 7: Response to a financial shock with fixed wage



Left panel plots capital policy function $k(n, z)$ by net worth for a given z, and right panel plots the same, scaled by aggregate capital. Lines labelled "0" denote pre-shock functions (i.e. with borrowing constraint λ_0). Lines labelled "1" denote post shock (λ_1) functions, before allowing unconstrained firms to adjust their target capital values, and lines denoted "2" give the actual post-shock functions.

lead to any misallocation which directly lowers aggregate TFP, and only leads to declines in labour. This is despite the relative richness of the firm distribution: some firms are constrained while others are not, and there is a well defined distribution over firm sizes induced by decreasing returns to scale in revenue and heterogeneous productivities and net worth.

How then could a financial crisis lead to declines in TFP in this framework? The only channel by which this is possible in this model is indirectly, by affecting other endogenous variables. In the next result I show that changes in the wage lead to changes in TFP. This is a partial equilibrium result, allowing for equilibrium in all other markets but ignoring the wage determination equation, (16). This partial equilibrium result is then used to establish that a financial crisis can affect TFP, but only via wage movements.

Proposition 2. *Allow for an arbitrary (n, z) distribution. In partial equilibrium, ignoring (16), $\partial Z/\partial w \geq 0$, with $\partial Z/\partial w > 0$ as long as there is a positive mass of both constrained and unconstrained firms. That is, decreasing the wage decreases aggregate productivity. Additionally, $\partial L/\partial w < 0$, so hours increase in response to a decrease in wages.*

Again, I relegate the proof of this proposition to Appendix B.1. That lowering wages increases hours is intuitive, since it increases labour demand at both constrained and unconstrained firms. To understand the intuition for why $\partial Z/\partial w \geq 0$, recall that a change in the wage can only affect aggregate productivity by changing the distribution capital across firms, via $\tilde{k}$. For constrained firms we simply have $k = \lambda n$, and hence there can be no change in capital in response to a marginal wage change. Thus, if there are to be changes in $\tilde{k}$ across firms it must be because of the behaviour of unconstrained firms.

Financially unconstrained firms choose their capital according to (4). From here we see that reducing the wage will lead unconstrained firms to increase their capital, because cheaper labour increases labour use and thus the marginal product of capital, and hence also encourages a higher capital choice. This is illustrated in Figure 8, which shows the response to a discrete reduction in the wage from w_0 to $w_1 < w_0$. For the purposes of exposition, the figure is drawn starting from the final state of Figure 7, after the borrowing constraint has been tightened to λ_1. In the top left panel I show the response of capital to the wage reduction. Constrained firms are not able to expand, but unconstrained firms do, increasing target capital to k_3^*. The increase in capital at unconstrained firms increases the total capital stock, leading to a reduction in the relative capital at constrained firms and a relative redistribution of capital towards unconstrained firms. This is represented by the shift in the top right panel from $\tilde{k}_2$ to $\tilde{k}_3$.

In Section 3.12 I showed that financially unconstrained firms always have lower TFPR than financially constrained firms. I also showed that redistributing capital to firms with low TFPR will reduce aggregate TFP. Hence the reduction in the wage, precisely by encouraging the relative redistribution of resources to financially unconstrained firms, must lower aggregate TFP. This is shown in the bottom right panel, where TFPR is constant for all initially unconstrained firms, and increasing as firms become more constrained.

Thus I have established that (1) a financial crisis without adjustment in wages will lead to no TFP fall, but a fall in hours, and (2) a decrease in wages will lead to a fall in TFP, and an increase in hours. I can combine these partial equilibrium results to compute the general equilibrium effect of a financial crisis, which I do in the following proposition:

Proposition 3. *Suppose that $w_0 = w_{mc}$, so the reference sticky wage level is the initial market clearing wage. In response to a negative financial shock (i.e. marginal decrease in λ):*

1. *The more flexible are wages (smaller is γ), the larger the fall in Z and smaller the fall in L*
2. *If wages are perfectly flexible ($\gamma = 0$) there is no labour wedge and the efficiency wedge falls*
3. *If wages are perfectly sticky ($\gamma = 1$) there is no change in the efficiency wedge and the labour wedge rises*

Proof. Point 1: TFP can be written as a function of the borrowing constraint and wages, $Z(\lambda, w)$. Additionally, equilibrium wages can be written as a function of the borrowing constraint, $w(\lambda)$. This gives the total effect of a marginal change in λ on Z as:

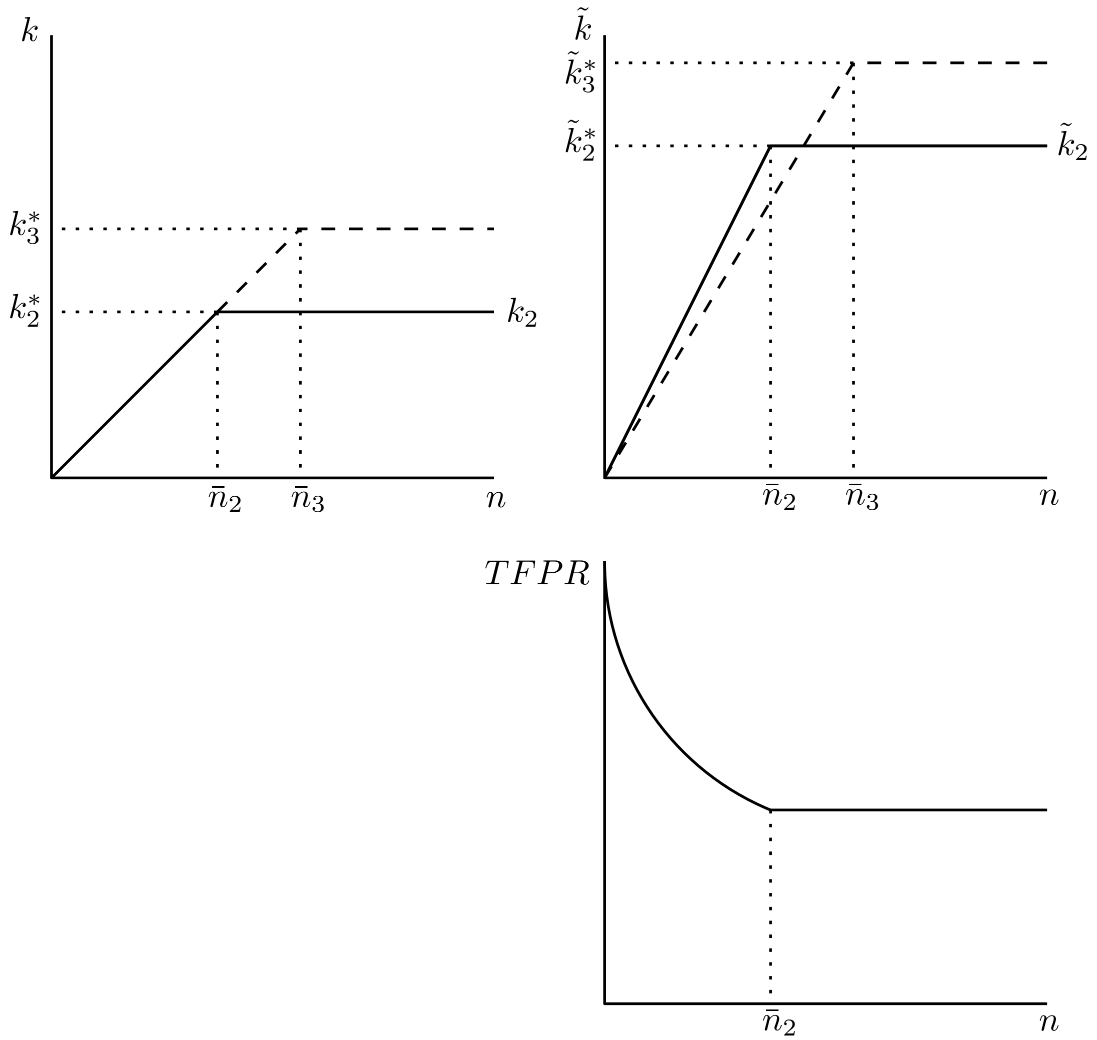
$$Z' = \frac{\partial Z}{\partial \lambda} + \frac{\partial w}{\partial \lambda}\frac{\partial Z}{\partial w} \tag{23}$$

Where I established in the previous two propositions that $\frac{\partial Z}{\partial \lambda} = 0$ and $\frac{\partial Z}{\partial w} > 0$. Differentiating (16) with respect to λ gives:

$$\frac{\partial w}{\partial \lambda} = (1-\gamma) w_0 \frac{\frac{\partial L}{\partial \lambda} v''(L) u'(C) - \frac{\partial C}{\partial \lambda} u''(C) v'(L)}{u'(C)^2} > 0 \tag{24}$$

Notice that wages (weakly) fall in response to a financial tightening, because both $\frac{\partial L}{\partial \lambda} > 0$ and $\frac{\partial C}{\partial \lambda} > 0$, and fall more the more flexible wages are (i.e. lower is γ). Thus the more flexible wages are, the larger is the weight on $\frac{\partial Z}{\partial w}$ and the larger the fall in productivity. A similar exercise for L confirms that the fall in hours is smaller when wages are more flexible.

Figure 8: Response to a reduction in wages



Top left panel plots capital policy function $k(n, z)$ by net worth for a given z, and top right panel plots the same, scaled by aggregate capital. Lines labelled "2" are pre-shock values (i.e. with initial wage w_0). Lines labelled "3" are post-shock (w_1) functions. Bottom right panel plots the marginal contribution to TFP, evaluated at the initial relative capital level.

Point 2: If wages are perfectly flexible, note that combining (16) and (10) leads to a labour wage which is independent of λ. In proving (1) we proved that TFP falls, which is a fall in the efficiency wedge.

Point 3: If wages are fully rigid, then I have already proved that TFP, i.e. the efficiency wedge, does not fall in response to a fall in λ. Combining the definition of the labour wedge with (10) yields:

$$\tau^l = 1 - \frac{v'(L)}{u'(C)} \frac{\rho}{w_0} \tag{25}$$

Following a fall in λ, both L and C fall, leading to an increase in τ^l since $v''(L) > 0$ and $u''(C) < 0$. $\square$

This completes the main results of this section. Greater wage flexibility leads to a financial crisis manifesting more as a fall in TFP, and less as a fall in hours. In terms of wedges, this translates into a financial crisis showing up more in either the efficiency or labour wedge. The results of this section are thus qualitatively consistent with the behaviour of the US and UK during the financial crisis.

3.14 Robustness and extensions

In Appendix B.2 I consider robustness and extensions to the above results. The first thing to note is that the results are not confined to the simple sticky wage structure presented above. Indeed, other shocks which increase the real wage, such as contractionary monetary shocks in a sticky nominal wage model, or negative labour supply shocks, also increase TFP in the above model. If these shocks occur coincidentally with a negative financial shock this has the same effects as exogenously sticky wages.

Much of the sharpness of the results above derives from the assumption of a simple collateral constraint, $k \leq \lambda n$, and I also investigate how generalising this constraint affects the results. The result that TFP is unaffected by a financial shock is robust to generalising the constraint to the form $k \leq \lambda g(n, z)$ for any function g. This form allows the constraint to depend on firm specific characteristics, but requires that the financial shock affects all firms proportionately. The result does not hold exactly for more general constraints of the form $k \leq g(n, z; \lambda)$. However, for the natural case of less productive firms having their constraints tightened more during a crisis, the result that TFP does not fall is still preserved, since TFP will now instead increase. The result that wage declines lead to TFP falls is robust to this more general constraint. However, a constraint which links borrowing to wages would weaken the result.

Finally, I show that the results are robust to, and even enhanced by, considering firm entry and exit. Including this extensive margin, a financial crisis without a decline in wages will now lead to an increase in aggregate TFP. This is because the crisis reduces profits, leading the least productive firms to exit, and only productive firms to enter, increasing TFP via a composition effect. Declines in wages increase profits, and hence allow less productive firms to survive, and encourage less productive firms to enter, reducing TFP as in the baseline model.

4 Quantitative model

While the model in the last section was useful in giving analytical results, it does not reveal if these effects are quantitatively relevant. In this section I build a dynamic, calibrated heterogeneous firm model (in the spirit of Hopenhayn and Rogerson (1993) and Khan and Thomas (2008)) to assess the size of these effects.

I conduct a series of partial equilibrium experiments, where I study the effect on TFP, hours, and other aggregates of a financial shock. The experiments are conducted under perfect foresight,[24] and chosen to replicate features of the Great Recession from both countries.

[24] I focus on perfect foresight to help build understanding of the role of the firm distribution. The mass of firms at the constraints, and how this varies conditioning on firm productivity, is important for the effects on TFP of the shock.

4.1 Environment

The model is very similar to the simpler model above, except that I now consider dynamics and enrich the firm structure. In particular, I maintain the basic structure of a representative household, final goods producer, and continuum of intermediate goods producers. Intermediate goods producers will now additionally enter and accumulate net worth over their lifetimes, and face capital adjustment costs, which are important for matching cross-sectional investment patterns.

Time is discrete, and the horizon infinite. I consider a notion of equilibrium where all prices are flexible except for the wage level, which is exogenously given. I will impose market clearing in all markets except for the labour market, where there will be rationing. I restrict myself to studying cases where labour supply exceeds labour demand, leading to unemployment.

4.1.1 Final good producer

I maintain the production function from the previous section, adding explicit dependence on time:

$$Y_t = \left(\int_0^1 y_{i,t}^{\rho} \, \mathrm{d}i \right)^{\frac{1}{\rho}} \tag{26}$$

Y_t is output at t, and $y_{i,t}$ is the input of intermediate i. Note there remains a unit mass of varieties of intermediates at any one time.[25] Letting $p_{i,t}$ be the price (in terms of the numeraire) of intermediate i, the first order condition for intermediate purchases from i gives the demand equation $y_{i,t} = p_{i,t}^{-\epsilon} Y_t$.

4.1.2 Intermediate good firm's static problem

As in the static model, intermediate goods producers are monopolistically competitive, and understand that the price of their goods depends on the amount they produce, according to $y_{i,t} = p_{i,t}^{-\epsilon} Y_t$. From now on I drop the firm level subscript i for convenience. In this dynamic version of the model, I introduce time to build and adjustment costs in capital.

However, we can still simplify the firm's problem by solving for the optimal labour choice, l_t, conditional on installed capital, k_{t-1}, idiosyncratic productivity, z_t, and the state of the economy, represented by indexing policy functions by t. The firm has the same CRS production function as the static model: $y_t = z_t k_{t-1}^{\alpha} l_t^{1-\alpha}$. Static profit is given by:

$$\pi_t(k_{t-1}, z_t) = \max_{l_t} z_t^{\rho} k_{t-1}^{\alpha\rho} l_t^{(1-\alpha)\rho} Y_t^{1-\rho} - w_t l_t \tag{27}$$

Once maximised, optimised profit, labour demand, and output are given by

$$\pi_t(k_{t-1}, z_t) = \left(\nu^{\frac{\nu}{1-\nu}} - \nu^{\frac{1}{1-\nu}} \right) \left(Y_t^{1-\rho} w_t^{-\nu} \right)^{\frac{1}{1-\nu}} z_t^{\frac{\rho}{1-\nu}} k_{t-1}^{\frac{\alpha\rho}{1-\nu}} \tag{28}$$

$$l_t(k_{t-1}, z_t) = \left(\frac{\nu Y_t^{1-\rho}}{w_t} \right)^{\frac{1}{1-\nu}} z_t^{\frac{\rho}{1-\nu}} k_{t-1}^{\frac{\alpha\rho}{1-\nu}} \tag{29}$$

[25] $i \in [0, 1]$ indexes firms who produce at time t. As detailed below, a fraction of firms will exit after production and be replaced by new entrant firms who will then produce next period. Each entrant inherits an index, i, from one exiting firm.

$$y_t(k_{t-1}, z_t) = \left(\frac{\nu Y_t^{1-\rho}}{w_t} \right)^{\frac{1-\alpha}{1-\nu}} z_t^{\frac{1}{1-\nu}} k_{t-1}^{\frac{\alpha}{1-\nu}} \tag{30}$$

4.1.3 Intermediate good firm's dynamic problem

A firm in operation in period t might shut down with exogenous probability $(1 - \sigma)$ every period. If they shut down at t then they produce in that period using their installed capital, and then sell their capital, repay their debts, and pay any remaining money out as dividends.

The mass of entering firms is exogenous at σ, chosen to keep the total mass of firms constant at one. Firms do not produce the period they enter, but are endowed with a fixed amount of net worth, n_e, from the household, which they use to purchase capital, and start producing the next period. Idiosyncratic productivity (TFPQ) evolves according to a Markov process: $z_{t+1} \sim \Gamma_z(z_t)$, and entrant firms draw productivity from the ergodic distribution of Γ_z, which I denote $\Gamma_{e,z}$. Each firm has a fixed unit mass of shares, owned by the representative household. A firm's cash flow any period must satisfy:

$$i_t + e_t = \pi_t - r_t d_{t-1} + d_t \tag{31}$$

Where $e_t \geq 0$ are dividend payments, d_t is borrowing in one-period, non-defaultable debt, and r_t is the interest rate. i_t is investment expenditures, or receipts from disinvestment. Capital adjustment costs are modelled as a wedge between the purchase and resale prices of capital: capital can be purchased at price 1, but must be resold at price $q^d < 1$, representing costs from reallocating capital which is partially specific to each firm. Thus, $i_t = k_t - (1-\delta)k_{t-1}$ if the firm invests ($k_t > (1-\delta)k_{t-1}$), and $i_t = q^d(k_t - (1-\delta)k_{t-1})$ if the firm disinvests ($k_t < (1-\delta)k_{t-1}$). If the firm decides to simply let its capital depreciate ($k_t = (1-\delta)k_{t-1}$) I label this inaction, and in this case $i_t = 0$.

This allows me to recover a balance sheet equation split into two regions. First define net worth, n_t:

$$n_t = \pi_t + (1-\delta)k_{t-1} - r_{t-1}d_{t-1} \tag{32}$$

We can combine this definition with the cashflow equation and the definition of investment to form a generic balance sheet:

$$q_t k_t + e_t = \tilde{n}_t + d_t \tag{33}$$

Where we have $q_t = 1$ and $\tilde{n}_t = n_t$ for investment or inaction, and $q_t = q^d$ and $\tilde{n}_t = n_t - (1 - q^d)(1-\delta)k_{t-1}$ for disinvestment. Taking the definition of n_t forward one period, we can construct a transition for net worth:

$$n_{t+1} = \left(\frac{\pi_{t+1}(k_t, z_{t+1})}{k_t} + (1-\delta)q^i_{t+1} - q_t r_t \right) k_t + r_t \tilde{n}_t - r_t e_t \tag{34}$$

Implicit in the above definitions is the restriction that the firm cannot raise equity: dividends must be positive ($e_t \geq 0$) and the firm cannot issue new shares. I additionally introduce a collateral constraint:

$$d_t \leq \lambda_t q^d k_t \tag{35}$$

This requires that borrowing cannot exceed a certain fraction of the resale value of your capital tomorrow, and can be motivated by a simple limited commitment problem. Firms maximise the present value of dividends, discounted using the household's stochastic discount factor. I denote the firm's maximised value by $V_t(n_t, k_{t-1}, z_t)$, where the t subscript allows for the possibility that value changes along transitions due to changes in the aggregate state.

Due to the non-convexity introduced by the capital adjustment costs, the value function must further be split into two regions. Denote by $V_t^i(n_t, k_{t-1}, z_t)$ the maximised value, conditional on investing. This can be expressed recursively as

$$V_t^i(n_t, k_{t-1}, z_t) = \max_{e_t \geq 0, (1-\delta)k_{t-1} \leq k_t \leq \frac{n_t}{q_t^i - \lambda_t q_{t+1}^d}} \Bigg\{ e_t +$$

$$\mathrm{E}_t \left[\beta \frac{u'(c_{t+1})}{u'(c_t)} \Big((1-\sigma)(n_{t+1} - (1-q^d)(1-\delta)k_t) + \sigma V_{t+1}(n_{t+1}, k_t, z_{t+1}) \Big) \right] \Bigg\} \tag{36}$$

where the maximisation is subject to the net worth transition, (34), with $q_t = 1$ and $\tilde{n}_t = n_t$. The maximisation incorporates the collateral constraint, which, combined with (33), places an upper bound on the feasible capital purchase. With probability $(1-\sigma)$ the firm exits tomorrow and pays out all remaining net worth as dividends, after adjusting for the fact that capital must be resold at the lower resale price. With probability σ the firm continues in operation. Similarly for the value of disinvesting, $V_t^d(n_t, k_{t-1}, z_t)$, we have

$$V_t^d(n_t, k_{t-1}, z_t) = \max_{e_t \geq 0, k_t \leq \min\{\frac{n_t - (q_t^i - q_t^d)(1-\delta)k_{t-1}}{q_t^d - \lambda_t q_{t+1}^d}, (1-\delta)k_{t-1}\}} \Bigg\{ e_t +$$

$$\mathrm{E}_t \left[\beta \frac{u'(c_{t+1})}{u'(c_t)} \Big((1-\sigma)(n_{t+1} - (1-q^d)(1-\delta)k_t) + \sigma V_{t+1}(n_{t+1}, k_t, z_{t+1}) \Big) \right] \Bigg\} \tag{37}$$

again subject to (34), now with $q_t = q^d$ and $\tilde{n}_t = n_t - (1-q^d)(1-\delta)k_{t-1}$. The maximised value is then the maximum over these two value functions, as the firm decides whether it is more profitable to invest or disinvest:

$$V_t(n_t, k_{t-1}, z_t) = \max \Big\{ V_t^i(n_t, k_{t-1}, z_t), V_t^d(n_t, k_{t-1}, z_t) \Big\} \tag{38}$$

If inaction is the optimal choice, then both of the conditional value functions will be maximised at $i_t = 0$, in which case they yield identical values. The solution to this maximisation defines the firm's capital policy function, $k_t = k_t(n_t, k_{t-1}, z_t)$, and dividend policy function, $e_t = e_t(n_t, k_{t-1}, z_t)$. The optimal evolution of net worth, $n_{t+1} = n_t(n_t, k_{t-1}, z_t, z_{t+1})$ is then implicitly defined by (34) and the borrowing policy function, $d_t = d_t(n_t, k_{t-1}, z_t)$, by (33).

Firms may accumulate enough net worth that they can permanently escape their financial constraints. In this case, they behave as Modigliani-Miller firms, and are indifferent about whether to retain earnings or pay them out as dividends. I follow Khan and Thomas (2013), and assume that these firms follow a "minimum savings policy", paying out the maximum dividends they can without running the risk of becoming constrained again.

4.1.4 Aggregation

Let $\mu_t(n_t, k_{t-1}, z_t)$ denote the marginal joint distribution of (n_t, k_{t-1}, z_t) across firms who produce at time t. Aggregate capital in use at t is given by $K_{t-1} = \int_0^1 k_{t-1}\,\mathrm{d}i = \int k_{t-1}\,\mathrm{d}\mu_t$. The model aggregates in the same way as the static model. I repeat the equations here with added time subscripts:

$$Y_t = Z_t K_{t-1}^{\alpha} L_t^{1-\alpha} \tag{39}$$

$$Z_t = \left(\int_0^1 z_t^{\frac{\rho}{1-\nu}} \tilde{k}_{t-1}^{\frac{\alpha\rho}{1-\nu}}\,\mathrm{d}i\right)^{\frac{1-\nu}{\rho}} = \left(\int z_t^{\frac{\rho}{1-\nu}} \tilde{k}_{t-1}^{\frac{\alpha\rho}{1-\nu}}\,\mathrm{d}\mu_t\right)^{\frac{1-\nu}{\rho}} \tag{40}$$

$$w_t L_t = (1-\alpha)\rho Y_t \tag{41}$$

Aggregate labour is defined as the integral of labour across firms, $L_t = \int_0^1 l_t\,\mathrm{d}i = \int l_t\,\mathrm{d}\mu_t$, and aggregate output is the given by the CES aggregator, (26). However, note that the above system of equations is sufficient to calculate Y_t and L_t from the underlying distribution of firms, whose impact is sufficiently summarised by Z_t and K_{t-1}. The evolution of μ_t must be tracked to simulate the economy. This is represented by a functional equation, which maps μ_t into μ_{t+1} and holds for all (n_t, k_{t-1}, z_t). This is denoted by $\mu_{t+1} = \Gamma_t(\mu_t)$, which will depend on firms' policy functions in equilibrium, and takes into account exit and entry.

4.1.5 Household

I maintain the representative household assumption from the previous model. Period utility is given by $U(C_t, L_t) = u(C_t) - v(L_t)$, where L_t is labour actually supplied in equilibrium, to be distinguished from *desired* labour supply, L_t^s. The household maximises the discounted sum of utility

$$U = \sum_{t=0}^{\infty} \beta\left(u(C_t) - v(L_t)\right) \tag{42}$$

where $\beta < 1$ is the discount factor. The household's budget constraint is

$$C_t + D_t = w_t L_t + r_{t-1} D_{t-1} + E_t \tag{43}$$

where D_t is lending to firms, and E_t is dividend income from firms, net of equity injections to entering firms. The solution leads to a standard Euler equation, $1 = \beta r_t u'(C_{t+1})/u'(C_t)$. Since I focus on labour disequilibrium where the household is off its labour supply curve, I do not allow the household to optimise labour supply, since it understands that this will be equal to the level of labour demand, L_t. However, I do specify the desired labour supply were the household allowed to choose labour, which is useful for defining a notion of unemployment. I define desired labour supply, L_t^s, as the solution to the standard labour optimality condition, $v'(L_t^s)/u'(C_t) = w_t$.

4.1.6 Market clearing

Goods market clearing now must take into account the partial irreversibility in capital. Consumption and capital are produced one-for-one from output. Disinvesting firms convert their capital back to output, although a fraction $(1-q^d)$ is lost. Goods market clearing gives

$$C_t + \sigma\left(\int_{i_t\geq 0} i_t \,\mathrm{d}\mu_t + \int_{i_t\leq 0} q^d i_t \,\mathrm{d}\mu_t\right) + (1-\sigma)\left(\int i_t^e d\Gamma_{e,z} - \int q^d(1-\delta)k_{t-1}\,\mathrm{d}\mu_t\right) = Y_t \qquad (44)$$

where $i_t(n_t, k_{t-1}, z_t) \equiv k_t(n_t, k_{t-1}, z_t) - (1-\delta)k_{t-1}$ and $i_t^e(z_t) \equiv k_t(n_e, 0, z_t)$. The terms multiplied by σ are investment and disinvestment by the fraction σ of firms who do not exit after production. The terms proceeded by $1-\sigma$ are the investment of entrant firms, and full disinvestment of exiting firms.

The labour market may not clear due to the partial equilibrium structure. In this case, disequilibrium unemployment is defined as the different between difference between labour demand and supply:

$$U_t \equiv L_t^s - L_t \qquad (45)$$

Market clearing in the intermediate goods markets is imposed by the monopolistic competition structure, and equilibrium in the bonds market follows from Walras' law.

4.1.7 Definition of equilibrium

Definition 2. *A labour market rationing equilibrium is a a sequence of allocations* $\{Y_t, C_t, L_t, D_t, K_t, U_t, L_t^s, Z_t\}_{t=0}^{\infty}$, *prices,* $\{r_t\}_{t=0}^{\infty}$, *functions,* $\{V_t, V_t^i, V_t^d, k_t, e_t, d_t, y_t, l_t, n_t\}_{t=0}^{\infty}$, *and distributions,* $\{\mu_t\}_{t=0}^{\infty}$, *such that, for given initial conditions,* (μ_0, r_{-1}, D_{-1}), *and a given sequence of borrowing constraints,* $\{\lambda_t\}_{t=0}^{\infty}$, *and wages,* $\{w_t\}_{t=0}^{\infty}$:

1. *The value functions,* $\{V_t, V_t^i, V_t^d\}_{t=0}^{\infty}$, *and policy functions,* $\{k_t, e_t\}_{t=0}^{\infty}$, *solve the firm's recursions, (36), (37), and (38).* $\{l_t, y_t, d_t, n_t\}_{t=0}^{\infty}$ *are defined by (29), (30), (33), and (34).*
2. $\{C_t, D_t\}_{t=0}^{\infty}$ *solve the household's problem, (42), subject to (43).*
3. *Output is given by the aggregator, (26). Aggregate labour and capital are defined as* $L_t = \int l_t \,\mathrm{d}\mu_t$ *and* $K_{t-1} = \int k_{t-1}\,\mathrm{d}\mu_t$ *for* $t = 0, 1, ...$ *Aggregate TFP is defined as the residual in (39).*
4. *The goods market clears for all* t*: (44).*
5. *The bond market clears:* $D_t = \int d_t \,\mathrm{d}\mu_t$ *for all* t.
6. $\{L_t^s\}_{t=0}^{\infty}$ *is given by* $v'(L_t^s)/u'(C_t) = w_t$, *and labour market disequilibrium by* $U_t = L_t^s - L_t$ *for all* t.
7. *The distributions* $\{\mu_t\}_{t=0}^{\infty}$ *evolve according to the policy functions* $\{k_t, n_t\}_{t=0}^{\infty}$ *and the Markov transition function,* Γ_z.

4.2 Calibration

The model is calibrated yearly. Firm level parameters are set to match moments of firm-level data in the model's stationary distribution. The US calibration follows closely Khan and Thomas (2013), and I replicate this calibration for the UK using UK data.

The calibration of aggregates is relatively standard. I set $\alpha = 1/3$, and set average firm productivity, $\mathrm{E}(z)$ to normalise steady state output to one in each country. The steady state wage, w_{ss}, is set to normalise labour supply to 1/3. I specialise to a risk neutral household ($u(c) = c$) allowing me to hold the real interest rate constant at $r_t = 1/\beta$ and focus on the effects of wage movements. The depreciation rate δ is set to 0.065.

Table 2: Calibration

	Interpretation	Value	Source
	Aggregates:		
β	Discount factor	0.95	–
α	$y = zk^{\alpha}l^{1-\alpha}$	0.33	–
δ	Depreciation rate	0.065	$I_{ss}/K_{ss} = 0.065$
$\mathrm{E}[z]$	Mean firm prod	1.70	Normalise $Y_{ss} = 1$
w_{ss}	Mean wage	1.64	$L_{ss} = 1/3$
	Firm level: (common)		
ρ	Good substitution	0.82	Cooper and Haltiwanger (2006)
ρ_z	autocorr. $z_{i,t}$	0.65	Khan and Thomas (2013)
q^d	Capital resale price	0.95	Khan and Thomas (2013)
	Firm level: (US)		
σ_z	$\mathrm{std}(z_{i,t})$	0.06	$\mathrm{std}(ik_{i,t}) = 0.34$ *(large firms)*
λ_{ss}	Collateral rate	0.49	$D_{ss}/A_{ss} = 0.37$
σ	Survival rate	0.9	Exit rate 10%
n_e	New firm equity	0.04	$L_e/L = 0.1$
	Firm level: (UK)		
σ_z	$\mathrm{std}(z_{i,t})$	0.09	$\mathrm{std}(ik_{i,t}) = 0.54$ *(all firms)*
λ_{ss}	Collateral rate	0.53	$D_{ss}/Y_{ss} = 1.15 \times (D_{ss}/Y_{ss})_{US}$
σ	Survival rate	0.89	Exit rate 11%
n_e	New firm equity	0.085	$L_e/L = 0.15$

For the US firm level calibration, the steady state borrowing limit, λ is set to match the average firm debt-to-asset ratio in the data as reported by Khan and Thomas (2013). The exit rate, σ, is set so that firms exit every 10 years on average, following the data used by Khan and Thomas (2013). The standard deviation of firm shocks is set to match the standard deviation of investment rates from Cooper and Haltiwanger (2006) of 0.33. Their sample is of large firms, so I take that sample as the subsample of my firms who are rich enough to become unconstrained. The exogenous net worth injection, n_e, given to entering firms is chosen to match the size of entering firms vs. the average firm as reported by Khan et al. (2014). In particular, it is chosen so that the average labour of entrant firms is 10% of the labour of the average firm.

For the UK, I choose the steady state borrowing limit so that the aggregate firm debt to output ratio is 15% higher than in the US calibration, in line with the BIS data I reported in Section 2. The exit rate is set to 11%, which is the average of the entry rate (12%) and exit rate (10%) in recent UK data, as reported in Barnett et al. (2014). The standard deviation of firm shocks is set to match the standard deviation of investment rates reported by Bayer (2005), this time for all firms in the sample since the data are for a representative group of firms. Finally, the net worth injection given to small firms is chosen so that the average size of entrant firms is 15% of the average firm size, following data from Brandt (2004).[26]

[26] Specifically, Brandt (2004) reports that, averaging across services and manufacturing, total employment at entrant firms is just over 1.5% of total employment. Since entrant firms make up approximately 10% of all firms in the model,

The final three firm level parameters are calibrated using US data, but applied to both countries. ρ, the CES parameter, is chosen to match the curvature in the profit function with respect to capital as measured by Cooper and Haltiwanger (2006). They report a curvature of 0.6, which corresponds to $\alpha\rho/(1-\nu)$ in my model once labour has been optimised. This leads to $\rho = 0.82$. I take the autocorrelation of firm level shocks from Khan and Thomas (2013) as 0.65, and the resale price of capital as 0.95.

4.3 Solution method

The firm's problem is solved via value function iteration. The solution is complicated by the occasionally binding borrowing constraint, and separate investment and disinvestment regions. In Appendix C I describe a procedure which simplifies the selection of the correct region. Additionally, the solution for firms who become rich enough to become forever unconstrained is simplified since it is no longer required to track their net worth as a state. I use Khan and Thomas' (2013) procedure to calculate the point at which firms become permanently unconstrained.

It is simple to show that all other firms find it optimal to pay zero dividends, removing one choice variable from the maximisation. The only other choice variable is capital, which implicitly defines a borrowing level. This is chosen by maximising over a grid, with allowances to ensure that exact inaction is feasible. Finally, while the state for any firm is three dimensional, (n_t, k_{t-1}, z_t), I show in the appendix that the maximisation stage can be performed on a two dimensional grid: $(\tilde{n}_t, z_t)$, and the results interpolated onto the (n_t, k_{t-1}, z_t) state. Given the expense of the maximisation step, this speeds up computation substantially.

Simulation and calibration of the steady state requires calculating the distribution of (n_t, k_{t-1}, z_t) across firms. I do this using a non-stochastic simulation procedure, first proposed by Young (2009). These algorithms discretise the states (n_t, k_{t-1}, z_t), and and define the distribution μ_t over this approximated state space. However, having two endogenous states makes existing algorithms inefficient due to the curse of dimensionality, because the number of nodes required grows exponentially with the number of variables. To overcome this problem I develop a new procedure which endogenously generates the nodes, and only uses nodes which are actually visited along any simulation. This algorithm has many useful applications in other models, and I discuss it further in Appendix E.

4.4 Steady state and ergodic distribution

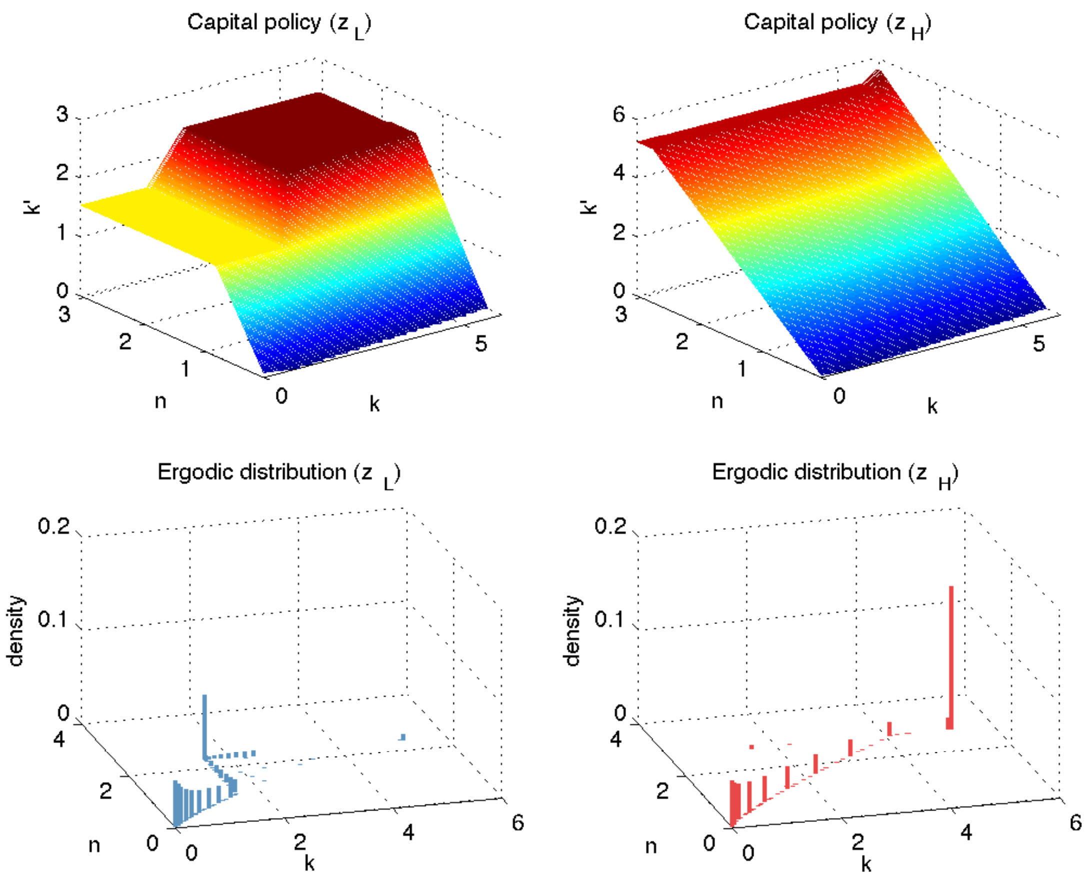
I first present the solution to the steady state of model without aggregate shocks. All aggregates remain at steady state values, but individual firms still experience idiosyncratic shocks. In Figure 9 I plot key elements of this steady state for the US calibration of the model.

In the top panels I plot the capital policy function, $k(n_t, k_t, z_t)$, with the left panel doing so for the low productivity state (z_l) and the right for high productivity (z_h). In both cases, for low enough net worth the firm is constrained and hence purchases capital up to the borrowing constraint. This leads to the linear increase in k_{t+1} as we increase n_t.

For high enough net worth, the firm becomes unconstrained, and the features introduced by irreversibility become more apparent. For low productivity, in the left panel, if capital is sufficiently high the firm will want to disinvest. Once the choice to disinvest has been made, the disinvestment target k_{t+1} is independent of k_t, leading to the flat region in the top right of the figure.

this leads to an average size of entrant firms of 15% of average employment.

Figure 9: Steady state: policy functions and ergodic distribution



The top panels give the steady state capital policy function $k(n, k, z)$, with the left panel for low productivity and the right for high. The bottom panels plot the densities in the ergodic distribution, with the left panel for low productivity and the right for high.

If the firm has sufficiently low capital, it will invest up to the target level of capital when investing, which is again independent of k_t. This leads to the flat regions to the top left of each figure. The regions where k_{t+1} is sloping up in k_t (in the middle of the z_l plot and to the top right of the z_h plot) are where current capital is sufficiently close to the target levels, and the firm chooses not to act, and simply let capital depreciate.

The bottom two panels plot the ergodic distribution, for low and high productivity respectively, with the height of the bars denoting the density at a given node. Roughly 50% of firms are unconstrained in the steady state, and most of these firms are represented by the largest two spikes in the density, which give the target capital levels for z_l and z_h. The figures illustrate the benefit of my simulation algorithm, since most of the points in the (k, n) space have zero mass in the ergodic distribution. Standard non-stochastic simulation algorithms would use an equi-spaced grid over the whole region, while my algorithm operates only on those points with positive mass.

The typical lifecycle for a firm can be summarised as follows. They are born young, with low net worth and no capital. They then purchase capital, and begin accumulating net worth and capital. They are financially constrained until they accumulate enough net worth. Firms with high productivity accumulate net worth faster, but also remain constrained longer since they have higher target capital stocks. Productivity shocks lead firms to invest and disinvest, with unconstrained firms also undergoing periods of inaction. Financially unconstrained firms with high productivity invest up to the target capital level. If productivity switches to the low state, they disinvest a portion of their capital, and let the remainder depreciate until they hit the target capital level for low productivity. Firms eventually shut down.

4.5 Comparative statics

To test the model's implications for aggregate outcomes, I subject it to several financial crisis experiments, meant to capture the rough features of the UK and US during the crisis. The first exercise is a simple comparative statics exercise designed to give an indication of the rough magnitudes one can expect from a financial shock, and varying wage decline, in this model. This experiment is carried out in the US calibration of the model, and I work with both countries' calibrations in the next sections.

The comparative statics exercise is to permanently tighten the borrowing constraint, λ, and recompute the steady state of the model. I then recompute aggregate output, Y, labour, L, and TFP, Z, and compute the percentage change from the initial steady state. I do this in partial equilibrium, allowing for different changes in the wage between the initial and new steady states.

In each case, I choose a financial tightening to generate exactly a 10% fall in output between the two steady states. I do this for wage adjustments between no adjustment, and a 5% fall. Notice that the size of the financial shock required to generate a 10% fall in output is larger when the wage is allowed to fall further, because the wage decline gives a boost to the economy. Thus, the interpretation of this exercise is that if one was to observe two economies, both with a 10% fall in output driven by a financial shock but with different observed wage adjustment, what would the model predict for TFP and hours in each economy?

Table 3: Permanent financial shock with different amount of wage adjustment

Δw	0%	-1%	-2%	-3%	-4%	-5%
ΔZ	0.1%	-0.3%	-0.7%	-1.0%	-1.4%	-1.7%
ΔL	-10%	-9%	-8%	-7%	-6%	-5%

The table shows the endogenous change in TFP and hours in response to a financial shock large enough to generate a 10% fall in output, for different degrees of wage adjustment. This is a partial equilibrium exercise, comparing the original steady state to the new steady state with a permanently tighter financial constraint.

The results of this exercise are given in Table 3, with different columns corresponding to different wage decline experiments. At the first extreme, a shock large enough to generate a 10% fall in output with no change in wages leads to a 10% fall in hours, and no decline in aggregate TFP (if anything, a 0.1% increase). At the other extreme, if wages are allowed to fall by 5% then the model instead generates only a 5% fall in hours, but a 1.7% fall in TFP.

Thus, we see exactly the pattern predicted by the analytical model replicated here: the larger the decline in wages post crisis, the smaller the decline in hours and the larger the decline in TFP.

Additionally, we see that the effects are quantitatively relevant: a difference of five percentage points in observed wage adjustment leads to a five percentage point difference in hours, and a 1.8 percentage points difference in TFP. Given that this wage difference was chosen to be similar to that observed between the US and UK (if anything slightly smaller), this is the first indication that the model might be able to explain a reasonable fraction of the roughly four percentage point gap in TFP and eight percentage point gap in hours observed during the crisis.

4.6 Decomposition 1: 2007-11

The next exercise I perform takes the US and UK data more seriously, and performs a simulation exercise using time series data from the Great Recession. As in the last exercise, I target output changes and choose the size of the financial shock required to generate them. The exercise runs from 2007-11 and thus asks: *conditional on matching output and wage paths*, how well does the model match data on hours and TFP in the two countries?[27]

I assume that in 2007 both countries are at their calibrated steady states, and agents believe that there will never be any aggregate shocks. At the beginning of 2007 all agents in the economy learn that an aggregate shock has occurred. The shock is modelled as a sequence of borrowing constraints, $\{\lambda_t\}_{t=2007}^{\infty}$, and wages, $\{w_t\}_{t=2008}^{\infty}$. The agents learn about the whole new sequence in 2007.[28]

Rather than pick a sequence for the borrowing shock, I instead target a transition path for output, and then back out the required sequence of borrowing constraints to create this path in each country. The data I use are annualised versions of the data presented in Figure 2 and aggregate wage series in Figure 4. I pick the sequence of borrowing constraints to match the output paths exactly, and assume that from 2011 onwards output and the wage remain at their 2011 values.[29]

Figure 10 gives the main results, with the paths for the UK given in solid blue, and the US in dashed red. The top left panel shows the path for output, and the path for wages is given in the bottom left panel. Both of these paths are exactly as they are in the data by construction.

The right panels give the model responses for hours and TFP. The model correctly predicts that hours should fall more in the US, where they experience a peak fall of over 8% compared to only 5% in the UK. It also correctly predicts that aggregate TFP should fall more in the UK, where it falls over 2.5%, while it remains elevated in the US most of the sample.

How do these responses compare to the data for each country? In Figure 11 I repeat the predicted hours and TFP paths from the model and compare them to the data. The top row gives the results for the UK. The model replicates the hours path very well. For TFP, the model is unable to match all of the 7% decline in TFP given by this measure, however with a maximum fall of around 2.5% the model can generate around 1/3 of the fall endogenously.

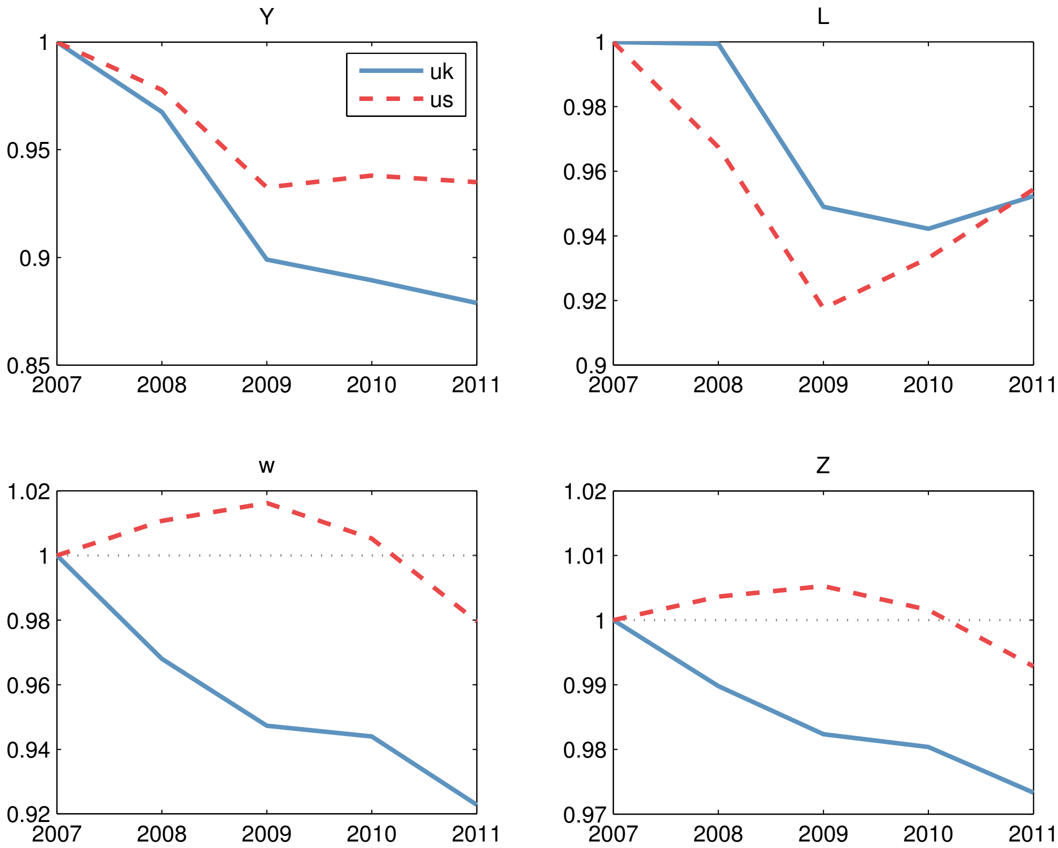
For the US, in the bottom row, the model predicts a peak fall in hours of 8%, while the peak fall is just below 12% in the data, so the model is able to capture around 2/3 of the fall. For TFP, the model correctly predicts essentially no movement, as in the data.

[27] I only run the experiment to 2011 for two reasons. Firstly, the corrected TFP data I use in the next section only runs to 2011. Secondly, by focusing on a shorter window I am able to focus on changes in TFP that are more likely related to the financial crisis, and not changes in longer term trends.

[28] This is a one-off, unanticipated shock sequence. One issue with unanticipated shocks is that they do not allow for precautionary behaviour by firms, who might want to hold more savings if they know a crisis could occur. To deal with this issue, I assume that in the initial steady state firms ascribe a positive, but vanishingly small, probability to a crisis occurring. This then only affects their dividend payout functions, not capital choices, encouraging them to retain more earnings.

[29] The results are robust to allowing for different final steady states.

Figure 10: Transition to financial shock



Response to the unanticipated financial shock. Y, Z, L, and w are aggregate output, TFP, labour and wages respectively. Blue solid lines are for the UK, and red dashed for the US. Y and w paths are targeted to match the data, L and Z paths are not.

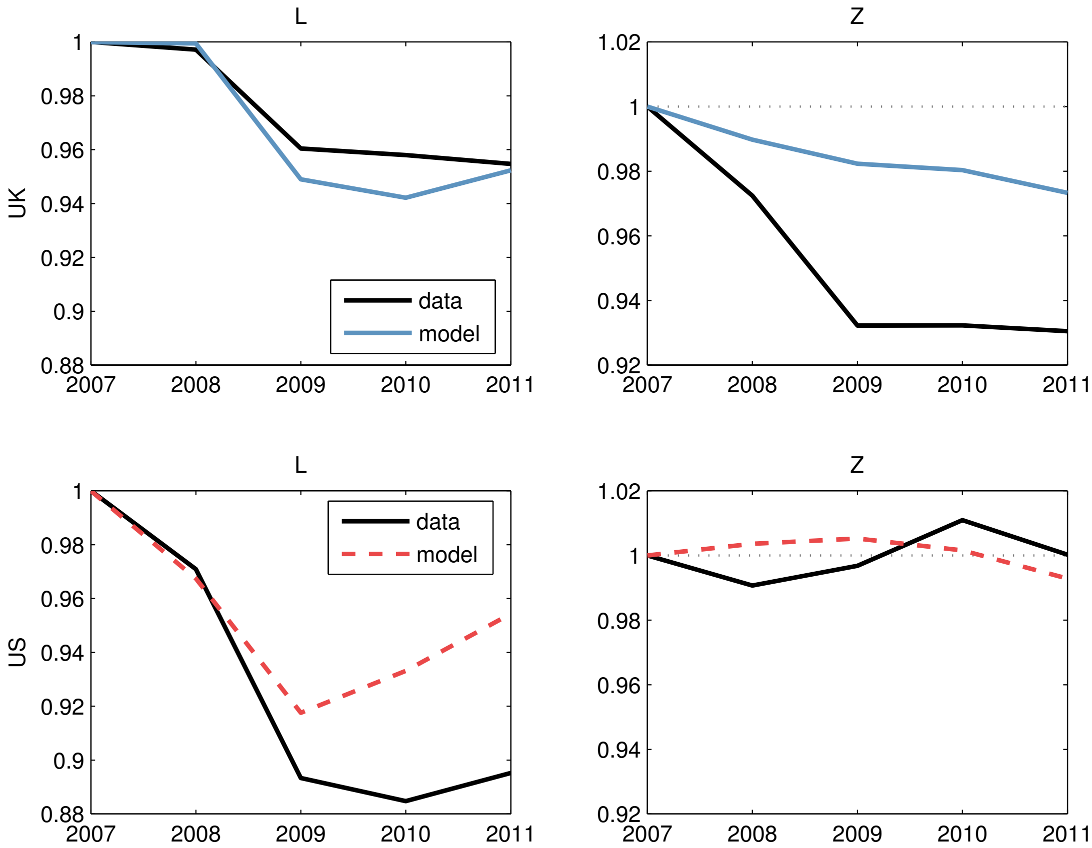
Overall, the model does a relatively good job at matching the paths for TFP and hours in both countries. In particular, the model generates the stylised fact that the recession manifested itself more in hours in the US, and more in productivity in the UK. The model matches the lack of a TFP fall in the US, approximately 1/3 of the TFP fall in the UK, and 2/3 of the hours fall in the US.

4.7 Decomposition 2: allowing for more shocks

The next exercise overcomes two shortcomings of the previous exercise. Firstly, in order to be able to create a rich time series with yearly data, the previous exercise used unadjusted data as its source, such as TFP data without the factor utilisation correction. This exercise uses the corrected TFP numbers, but this data is only available for the change in TFP between 2007 and 2011. Secondly, the last exercise attributed the whole fall in output in each country to a financial shock, while in reality there could have been other shocks at play. In this exercise I also allow for a common productivity shock and show that the results are robust to this.

For this exercise, I use output and hours data excluding the financial sector (Figure 3), and the TFP data after excluding finance, controlling for utilisation, and increasing capital scrapping

Figure 11: Transition to financial shock: model and data



Response of TFP (Z) and hours (L) to the unanticipated financial shock in the model and data. Top row gives UK and bottom US. Data are in solid (darker) black, and the (lighter) blue solid lines are model responses for the UK, and red dashed for the US.

(Table 1). Finally, I use the composition-adjusted wage data (right panel of Figure 4). Since the TFP data are only available as a change between 2007-11, I lose the full time series detail. Thus, instead of a period-by-period decomposition, as I did in the last section, I now perform a comparative statics exercise.

I augment the model by adding a common exogenous TFP shock, Z_t^c, to firms' production functions, which are now given by $y_{i,t} = z_{i,t} Z_t^c k_{i,t-1}^{\alpha} l_{i,t}^{1-\alpha}$. I perform a comparative statics exercise where I adjust the wage, borrowing constraint, and exogenous TFP shock and compare the new steady state to the original steady state.

As in the previous exercises, I choose the size of the financial shock to target the output fall the model generates. For each country, I choose the exogenous TFP shock based on data on aggregate TFP between 2007-11 combined with the evidence on misallocation. The misallocation data for the UK during the crisis report a 3% decline in TFP due to misallocation, leaving $5.25\% - 3\% = 2.25\%$ of the total TFP fall in the data unexplained. I thus give the UK a 2.25% negative TFP shock. The US data showed no increase in misallocation, so I attribute the US a negative TFP shock equal to the entire 0.96% fall in TFP in the data.

Table 4: Decomposition 2: results

	UK:		US:		Gap:	
	Data	Model	Data	Model	Data	Model
Y	-9.54%	-9.54%	-7.74%	-7.74%	-1.80pp	-1.80pp
w	-11.3%	-9.54%*	-2.43%	-2.43%	-8.87pp	-7.11pp
Z	-5.25%	-4.26%	-0.96%	-1.27%	4.29pp	-2.99pp
L	-3.91%	0%	-10.5%	-5.44%	6.59pp	5.44pp
Z_c	–	-2.25%	–	-0.96%	–	-1.31pp
$Z - Z_c$	–	-2.01%	–	-0.31%	–	-1.70pp

* *UK wage fall capped at size of output fall. Numbers are percentage changes between 2007-11 in the data and model for each country. "Gap" refers to the percentage point difference between the US and UK.*

In Table 4 I give the data used in the first column for each country respectively, and also compute the gap between the two. By this data, TFP fell by over 5% in the UK and less than 1% in the US, while hours fell over 10% in the US and less than 4% in the UK. Wages fell over 11% in the UK and less than 3% in the US. As in the previous exercise, I input the wage fall seen in the data to the model for each country. For the UK, the wage fall is rather severe, and I conservatively cap the wage fall at the size of the output decline seen in the data.

The second column for each country gives the model results. Z refers to the change in aggregate TFP, and L total hours. In the final two rows I give the exogenous TFP shock, Z_c, and the portion of the change in aggregate TFP which is endogenous, $Z_c - Z$. This is simply the total change in TFP less the exogenous change.

For the UK, the model generates an endogenous 2.01% fall in aggregate TFP, which is 38% of the total fall in the data. For the US, the model generates a 0.31% endogenous fall in TFP, leading to a 1.7pp difference in the endogenous TFP responses of the two countries. Given the total difference in TFP response of 4.29pp, the model thus generates 40% of the TFP gap between the two countries endogenously.

For labour, the model generates a 5.44% fall in hours in the US, which is 52% of the fall in the data. The model performs worse for the UK, where it actually predicts no fall in hours. The model generates a 5.44pp difference between the two countries, which is 83% of the gap seen in the data.

Overall, with the improved data and even allowing for alternative shocks, the model does a good job at replicating the data from the two countries. The difference in wage response contributes to the model being able to explain 40% of the TFP difference between the two countries endogenously.

5 Micro implications

In the last section, I showed that the model is able to do a relatively good job at matching the aggregate series in the US and UK during the Great Recession. I now turn to more disaggregated data to test the model's underlying mechanisms.

5.1 Effects by firm size

As discussed in the analytical results, the reason that wages affect aggregate TFP in my model is that they affect the distribution of resources across firms. Thus, it is important to check whether the underlying movements in these distributions match patterns observed in the data during the Great Recession. I do this now, for investment and employment.

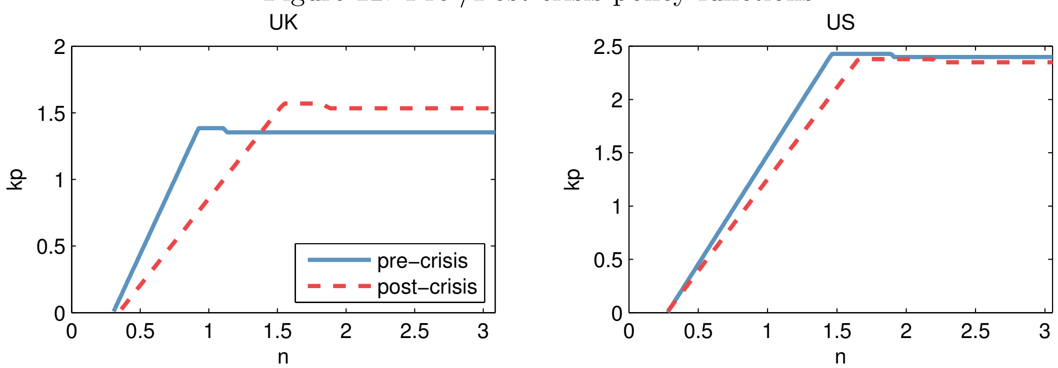
5.1.1 Investment

The larger decline in TFP in the UK in my model is driven by more reallocation of resources to financially unconstrained firms than in the US. This is illustrated in Figure 12, where I plot the capital policy function pre- and post-crisis in from my second decomposition experiment in Section 4.7. The policy function, $k_t = k(n_t, k_{t-1}, z_t)$ is plotted across net worth levels, for a given value of productivity and previous capital.

The left panel plots the UK, with the pre-crisis policy function in solid blue and post in dashed red. We see that poor firms, with low net worth, are forced to purchase less capital as the (binding) borrowing constraint tightens. Rich firms, on the other hand, increase their capital purchases, since the low wage makes the unconstrained optimal investment level higher.

The US is plotted in the right panel. Poor firms are again forced to reduce their investment, but by less than in the UK. Additionally, large firms do not increase their investment by as much as in the UK, and if anything decrease it slightly.

Figure 12: Pre-/Post-crisis policy functions



Both panels plot the steady state capital policy function $k_t = k(n_t, k_{t-1}, z_t)$ *evaluated at the low productivity state and highest capital state. The solid blue line is the original steady state, and the dashed red lune is the new steady state after the financial shock.*

For a given level of productivity, note that more financially constrained firms (with lower net worth) will be smaller. Hence it is common to use size as a proxy for financial constraints. Given the above, we should thus expect that the gap in investment between small and large firms should increase more in the UK during the crisis than it does in the US.

To test this requires firm-level data for the UK and US. For the UK, Crawford et al. (2013) run suitable regressions using Annual Respondents Database (ARD), which gives almost a census of large firms, and a random sample of smaller firms. I then use Compustat data for the US to run comparable regressions and compare the two countries.

Table 5: UK Investment decline by firm size

Size:	All	<50	50-249	≥250
Post08	-0.099***	-0.107***	-0.157***	-0.042*

Table gives the coefficient on Post08 in regression (46). Columns give regressions run only using firms with average employees in the given range.

Crawford et al. (2013) run the following regression on firm level investment data:

$$\log(Inv_{i,t}) = \alpha_i + \beta Post08_t + \gamma Year_t + \delta Post06_t + u_{i,t} \tag{46}$$

The specification estimates firm level investment as a simple function of a time trend, firm-level fixed effects, and a dummy equal to one during the financial crisis ($Post08_t$ is equal to one for 2008-9). Given the use of firm fixed effects, the coefficient on $Post08_t$ measures the average percentage within-firm fall in investment during the first two years of the crisis.[30] They run the regression for different firm groups, splitting the sample into small (<50 employees) medium (50 to 249 employees) and large ($\geq$ 250 employees) firms. The data are yearly and run from 1997-2009.

I give their coefficient on the post-2008 dummy in Table 5. The first column shows that, across all investing firms, there was a 10% drop in the average size of investment during the crisis. However, later columns show that this was particularly concentrated among smaller firms. Large firms saw an average 4.2% fall in employment, while the fall for medium firms was nearly four times larger at 15.7%, and small firms saw a fall of 10.7%.

Thus, we do see the small-large investment gap widening in the UK, consistent with my model. However, to be fully consistent it should be that the small-large gap increases more in the UK than in the US. To this end, I replicate Crawford et al.'s methodology on US data. I use the Compustat dataset, which gives a sample of approximately 30,000 listed firms.[31]

I first run exactly the same specification as they run, again splitting by firm size and using 1997-2009 data. However, the presence of the Dot-Com bubble and crash make the linear time trend specification problematic. I report the results from their specification in the appendix, which supports the results I give below but gives parameter values which are less easy to interpret. I instead report here the results from a more general specification:

$$\log(Inv_{i,t}) = \alpha_i + \beta_t + u_{i,t} \tag{47}$$

This specification replaces the time trend and dummy approach with a more general approach allowed a time dummy for each year. I only report the regressions for medium and large firms, since the Compustat sample of small firms is too limited, and the time dummies are imprecisely estimated. After keeping only firms with at least two observations, I have 5,495 large firms (27,440 firm-year observations) and 1,932 medium firms (7,320 firm-year observations). I plot the time dummies from

[30]Given the use of a log specification, the authors exclude firm-years with zero investment. The additionally add a dummy for post-2006 to control for a data revision.

[31]Following Crawford et al., I drop the top and bottom 1% of firms by employment and sales, and deflate investment using the industry specific GDP deflator.

this regression in Figure 13, with the 2007 dummies normalised to 0 for each group.

Figure 13: US investment: medium vs large firms

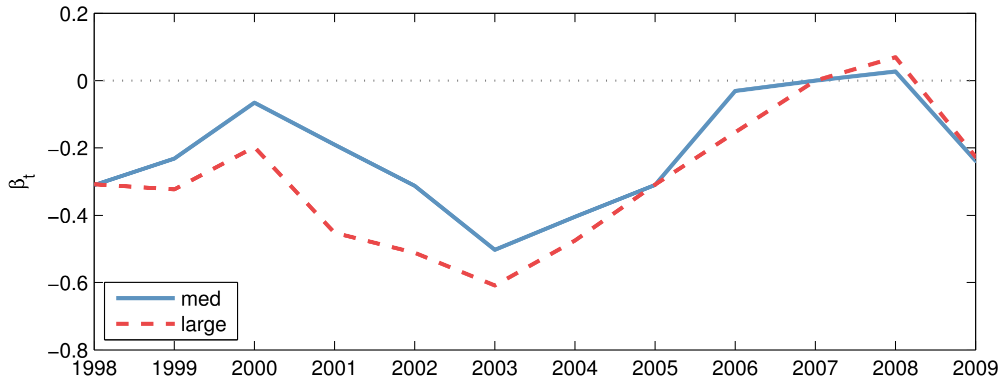


Figure plots the time fixed effect in the regression $log(Inv_{i,t}) = \alpha_i + \beta_t + u_{i,t}$ *using US Compustat data. Blue solid line is the regression only for medium sized firms (measured by employment) and the dashed red only for large firms.*

The results reveal that the decline in investment during the first two years of the crisis was similar for both medium and large firms. The pre-trends from 2003 onwards are similar, with both groups then seeing a similar decline between 2007 and 2009. Thus, in contrast to the UK, there is no evidence of a widening of the small-large investment gap in the US during the first few years of the crisis. This is consistent with the predictions of my model.

5.1.2 Wages and employment

Given the spot labour market with a common wage across firms, labour at the firm level moves in line with capital. Thus, my model predicts that the small-large employment gap should increase in the UK, relative to the US during the crisis. Is this what happens in the data? It turns out that the opposite is true. Sahin et al. (2011) report that employment at small firms was hit harder than large firms in the US. On the other hand, Crawford et al. (2013) report that employment at large firms was hit harder in the UK. Thus the small-large gap falls in the UK, relative to the US, contrary to the model's predictions.

These findings are supported by causal evidence using bank-firm relationships. In the US, Chodorow-Reich (2014) finds that firms with relationships with worse performing banks had to reduce employment more during the crisis, and this only holds for small firms. In the UK, Franklin et al. (2015) do not report a significant relationship between bank distress and firm employment, while they do find effects on capital intensity.

This is a key failure for the model, since this behaviour is central to the model being able to generate the aggregate facts I want to explain: It is precisely *because* the small-large gap goes up more in the UK that the model predicts that TFP falls in the UK and not the US. Thus, this evidence calls in to question the model's ability to explain the data. Is it possible to reconcile the model with this data?

The data on employment and wages at the firm level actually points to an interesting mechanism the model is missing, which can reconcile the model with the data. While they find that employment in the UK was hit harder in large firms, Crawford et al. (2013) also report an interesting reason why this might be the case: small firms reduced their real wages more than large firms during the crisis, and similarly for labour productivity. These findings are causally linked to financial distress by Franklin et al. (2015), who find that firms linked to distressed banks reduced their wages more than other firms. They also find that these firms reduced their capital intensity, and thus experienced falls in their labour productivity.

Together, these two studies paint an interesting picture. In the UK, it appears that financially constrained firms had to reduce their investment, but were able to substitute their capital for labour, which explains why labour did not fall so much. This was optimal at the firm level because financially constrained firms were able to reduce their wages, encouraging the substitution towards labour, which became cheaper as capital became harder to purchase. These effects did not operate at large firms, where the lack of a decline in labour productivity suggests no reduction in capital intensity, and since they did not lower their wages much, the reduction in their output was naturally met with a large reduction in employment.

Combined, this suggests that the financial crisis led to a situation where financially constrained firms shed capital and (relatively) replaced it with labour, while large firms shed labour and (relatively) replaced it with capital. This is *even worse* for productivity in the UK than the situation outlined in my model, because capital ends up being used by one group of firms, and labour by another, while ideally they should both be using a mix of the two.[32]

If the ability of small firms to lower their wages was not as pronounced in the US, perhaps because with lower inflation there was less ability to offer real wage cuts, these effects would not have operated in the US. This appears to be the case: while Wang (2015) does report that small firms cut wages relative to large firms during the recession, this gap is small and only emerges between 2010 and 2011.

This data suggests that incorporating *wage heterogeneity* is crucial for my model to fully match the data on employment flows. So far the model focused on the fact that wages fell more in the UK than the US, but it appears that it is also important that wage heterogeneity increased more in the UK than in the US. As discussed above, incorporating that wages fell more in small firms in the UK, and did not in the US, can help reconcile the model with employment flows data, and should make the effect of wages on aggregate TFP even stronger.

5.2 Job flows

In Section 2.6.2 I showed that the difference in employment experience between the two countries is driven by differences in hiring, and not firing. In this section I show that my model is consistent with this finding.

While my labour markets are frictionless, apart from a sticky wage, I can still create labour market flows by assuming that, instead of firms randomly rehiring labour on the spot market each period, people remain with the same firm until the firm needs to terminate a relationship, or they quit. For this exercise, I thus fix workers' intensive margins, and interpret movements in total hours only as movements in employment.

I assume an exogenous quit rate, ρ_x, of 20% per year.[33] Firing also happens endogenously when

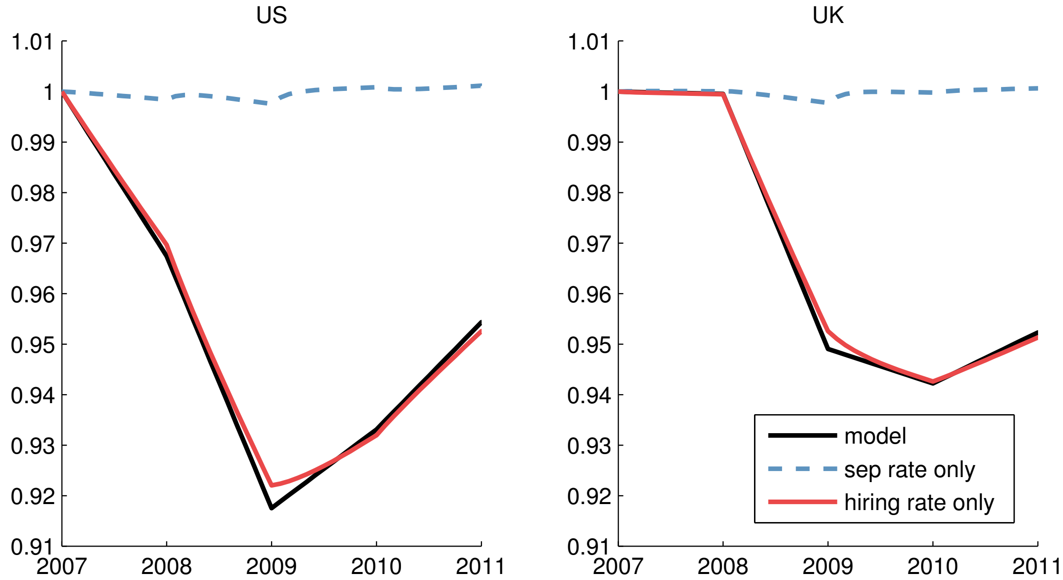
[32] In terms of firm level wedges, note that my model only disrupts the distribution of capital across firms, while the amount of labour at each firm is optimal given their capital stock. In the data it appears that not only capital but also labour decisions are distorted, giving an extra kick to misallocaiton.

[33] This number brings the total hiring and firing flow in steady state close to the US data. The results are robust to using different numbers.

firms want to downsize by more than this, or when a firm exogenously shuts down. For a given firm, if $l_{t+1} < (1-\rho_x)l_t$ then total fires are $f_t = l_t - l_{t+1}$ and total hires are $h_t = 0$. If $l_{t+1} \geq (1-\rho_x)l_t$ then total fires are $f_t = \rho_x l_t$, and total hires are $h_t = l_{t+1} - (1-\rho_x)l_t$.

Note that the assumption of a constant quit rate is natural, given the limited role for endogenous separations in the data. However, the model still does allow a role for endogenous separations through firing. To test the role of hiring and firing, I calculate total hires and fires in the time series data generated by the model in the first decomposition from Section 4.6, and perform the same decomposition exercise as I perform on the data.

Figure 14: Model labour flows



Both panels plot model-predicted employment as deviations from 2007 values. The (darker) black line plots the model-generated data, the solid (lighter) red line plots the simulated path holding the employment separation rate constant and taking the hiring rate from the model. The dashed blue line plots the simulated path taking the employment separation rate from the model and holding the hiring rate constant.

Figure 14 gives the results of the decomposition. In both model variants, employment is driven almost entirely by hiring, as in the data. This is driven by two effects. Firstly, the financial shock directly hits young firms. These firms are growing every period, and hence always tend to be hiring, and any shocks affecting them thus affect the hiring rate. Secondly, the high quit rate (needed to match the data) means that older firms, who we know behave differently between the two model variants, can adjust labour to their target levels almost entirely by adjusting hiring with no need to fire.

5.3 Composition of debt

The model has predictions for the composition of debt use across firms during the crisis which qualitatively match patterns observed in the two countries. In the decomposition of Section 4.7, the model estimates that the borrowing constraint tightened by 20% in the US, and 60% in the UK. Thus the supply of credit appears to be more disrupted in the UK.

While this difference is large, it should not be interpreted as a large difference in the response of total debt across the two economies. To see this, note that the difference in the fall in total capital between the two model economies is much more muted: capital falls by 16.4% in the UK model, and 9.6% in the US model. Thus, the large difference in the λ falls reflects more a change in the *composition* of debt: the UK model predicts that debt at financially constrained firms falls particularly hard, while debt use at richer firms is less affected. In the US model, debt use does fall more at constrained firms, but the difference is less pronounced.

This pattern holds in the debt data from the BIS presented in Figure 1. This data split firm debt into total debt (left panel) and debt only issued by domestic banks (right panel). This allows us to decompose the fall in total debt in each country into the fall in debt from bank lending, and from other sources. Doing so reveals that around 60% of the fall in total lending in the UK is accounted for by the fall in bank lending, while this figure is only 30% for the US. To the extent that young, financially constrained firms will be more reliant on bank lending than other sources (such as bonds) the fall in credit was particularly targeted at financially constrained firms in the UK, as predicted by my model.

5.4 Firm entry and exit

Entry and exit behaved differently in the two countries during the recession, and it is interesting to see if the model can match these patterns. Using BDS data, Siemer (2014) reports that, in the US, entry dropped and exit increased significantly during the recession. Specifically, this leads to a 5% decline in the number of firms from 2007-10, with the establishment entry rate falling by 25% and the exit rate increasing by 15%. Hathaway and Litan (2014) report similar numbers, with the entry and exit rates remaining dislocated through 2011.

In the UK, on the other hand, these effects are much more muted. Whether measured by establishments or firms, there is a smaller decline in the number of firms during the recession.[34] Looking at entry and exit rates, Riley et al. (2015) do report changes in both rates, but, unlike the US, they have recovered by 2011. The increase in exit rates is particularly small, increasing by 20% but only for a single year.

Can my model explain this difference between the two countries? In Section B.2.2 I analysed a version of my model with entry and exit. In this model, holding wages constant, a financial crisis reduces entry and increases exit, by reducing firm profits. This is thus consistent with the experience of the US. Additionally, a decline in wages promotes entry and discourages exit by increasing profits, and hence the model can explain why the rates moved by less in the UK.

6 Conclusion

The US and the UK experienced the financial crisis differently, with the former suffering a large decline in hours worked, and the latter a large decline in TFP. While one explanation for such a difference could be that the UK experienced more labour hoarding than the US, I show that this explanation does not square with the data. The differences in employment between the two countries are driven by larger declines in job creation in the US, and not differences in job destruction as would

[34] Data from the Inter-Department business register shows a 1% drop in the number of establishments from 2007-10 if we include firms with zero employees, or a 2% increase including such firms. Rhodes (2015) reports a 2.3% decline in the number of firms with employees from 2007-10 using BIS Business Population Estimates.

be predicted by a labour hoarding explanation. Instead, the smaller decline in job creation in the UK implies that the UK reallocated more labour than the US during the crisis, where labour instead moved out of the labour force. At the same time, there has been a larger increase in misallocation across firms during the crisis in the UK than the US, which can explain around 1/3 of the UK's TFP decline during the beginning of the crisis.

I thus investigate an alternative explanation for the different responses to the crisis of the two countries, motivated by the above facts. A recent theoretical literature (e.g. Khan and Thomas, 2013, Buera and Moll, 2015) shows how financial shocks can reduce aggregate TFP by inducing misallocation. In this paper I argue that these standard models of heterogeneous firms can explain the experience of both countries, if one accounts for the fact that their real wages behaved differently during the crisis.

Wages in the UK declined further and faster than in the US, and I show that this has implications for the manifestation of the financial crisis. I build a heterogeneous firm model with financial frictions and derive both analytical and quantitative results. I first show that a financial crisis, modelled as a tightening of borrowing constraints, has no effect on TFP if real wages do not adjust. This is despite the fact that the model admits a rich firm structure, with both constrained and unconstrained firms endogenously emerging in equilibrium.

In the model firms produce goods which are imperfect substitutes using a constant returns to scale technology, and compete via monopolistic competition. A financial crisis only directly affects financially constrained firms, but imperfect substitutability induces a demand spillover across firms, so that even unconstrained firms choose to downsize in response to the shock if wages do not adjust. In the analytical model, I show that this leads to no reallocation of resources across firms, since all firms shrink by the same amount, and hence no change in TFP. Intuitively, with sticky wages a financial shock also becomes a demand shock, leading even unconstrained firms to shrink.

If wages do adjust downwards during the crisis, I show that this endogenously lowers TFP. Low wages encourage firms to hire, but financially unconstrained firms are in a better position to do so than constrained firms, which reallocates resources to unconstrained firms. Constrained firms have been kept too small by definition, and a social planner would increase TFP by reallocating resources towards these firms. I show that constrained firms will always have higher revenue TFP (TFPR) than unconstrained firms, and hence that reallocating resources to unconstrained firms will always lower TFP.

While wage declines thus cause TFP declines during a crisis, they naturally protect labour markets by encouraging firms to hire. Thus the model is able to replicate both the US experience (of a large decline in hours and no TFP decline) and the UK experience (of a large TFP decline and smaller hours decline) conditional on each country's wage path. The model is also able to generate the behaviour of each country's labour and efficiency wedge (Chari et al., 2007).

I then construct a richer, dynamic version of the model, and calibrate it to firm-level data from each country. The model features a life cycle of firms who accumulate net worth endogenously, and invest subject to partial irreversibilities in capital. I perform a series of partial equilibrium experiments, asking whether *conditional on matching output and wage paths during the crisis* the model is able to correctly predict the behaviour of hours and TFP in both countries. In one experiment I allow only for a financial shock, and in another also for a common productivity shock, and find across both exercises that the model generates around 1/3 of the difference in TFP between the two countries. Depending on the exercise, the model can also explain 2/3 to all of the difference in hours, and can also explain around 1/3 of the UK's TFP decline, or "productivity puzzle."

Solving this model requires simulating a continuum of firms with multiple endogenous states. I do this via a new non-stochastic simulation algorithm which extends Young's (2009) original algorithm to help overcome the curse of dimensionality. The algorithm uses firms' policy functions to endogenously create the grid over which the population distribution is approximated. This allows for many fewer gridpoints than existing algorithms, which use equispaced grids for each endogenous state, making the procedure feasible in higher dimensions.

Finally, I confront the model with disaggregated data to test the underlying mechanisms. I find that the model is consistent with the dominant role of job creation (as opposed to destruction) in explaining the difference in employment between the two countries during the crisis, and an extension to include firm entry and exit also matches the greater decline in the number of firms in the US. The model also predicts a greater tightening of debt at constrained (as opposed to richer) firms in the UK, consistent with the behaviour of bank vs. total lending in both countries.

I then look at effects by firm size, taking size as a proxy for financial constraints. The model matches the behaviour of investment by firm size during the crisis, but fails to match the behaviour of employment by firm size. This is the key challenge for future work, and I discuss the likely resolution of this failure. My model assumed common wages across firms, and studied the effects of adjustment of this common wage. In reality, firms charge different wages, and evidence suggests that *wage dispersion* increased more in the UK during the crisis. Small firms cut wages more than large firms, presumably in response to their declining profitability as their access to investment was cut off, leading small firms to substitute towards labour. This appears to have happened less in the US, leading to another key difference between the two countries that could be important for explaining their differing productivity performance. Future versions of the model should take wage dispersion more seriously, in order to bring the model closer to the data.

Ideally, such work should be carried out with detailed firm-level data for each country. This would also allow me to construct more detailed measures of misallocation (such the misallocation of capital vs. misallocation of labour) to further discipline the model. More detailed data would also allow better tests of the underlying mechanisms, more detailed calibrations, as well as investigations into alternative explanations.

On the normative side, the results of this paper suggest a new tradeoff for monetary policy following a crisis: Attempts to stimulate employment by increasing inflation (and hence reducing the real wage) may come at the expense of lowering TFP. Monetary policy is a blunt tool, and the central bank cannot decide *where* resources will flow following a stimulus. More targeted financial policies may be more effective at increasing employment without detrimental productivity effects, since they can be targeted at financially constrained firms. My model provides a framework for analysing the effects of such competing policies, and designing optimal interventions in response to future shocks.

References

[1] Acemoglu, Daron, 1997. *Good Jobs vs Bad Jobs: Thoery and Some Evidence.* Working paper. Cambridge: MIT.

[2] Acemoglu, Daron, 2002. *Directed Technical Change.* The Review of Economic Studies, Vol. 69, No. 4 (Oct., 2002), pp. 781-809.

[3] Acemoglu, Daron, and Shimer, Robert, 1999. *Efficient Unemployment Insurance.* The Journal of Political Economy, Vol. 107, No. 5 (Oct., 1999), pp. 893-928.

[4] Acemoglu, Daron, and Shimer, Robert, 2000. *Productivity gains from unemployment insurance.* European Economic Review, Vol. 44, pp. 1195-1224.

[5] Arellano, Cristina, Bai, Yan, and Kehoe, Patrick J, 2012. *Financial Frictions and Fluctuations in Volatility.* Federal Reserve Bank of Minneapolis Research Department Staff Report 466. July 2012.

[6] Baldwin, J.R., R. Jarmin and J. Tang, 2002. *The Trend to Smaller Producers in Manufacturing: A Canada/U.S. Comparison.* Economic Analysis Research Paper Series, No. 3, Statistics Canada.

[7] Barnett, Alina , Adrian Chiu, Jeremy Franklin and María Sebastiá-Barriel, 2014. *The productivity puzzle: a firm-level investigation into employment behaviour and resource allocation over the crisis.* Bank of England Staff Working Papers, no. 495. 17 April 2014.

[8] Barth, Erling, James Davis, Richard B. Freeman, Sari Pekkala Kerr, 2016. *Weathering the Great Recession: Variation in Employment Responses by Establishments and Countries.* NBER Working Paper No. 22432.

[9] Bayer, Christian, 2005. *Investment dynamics with fixed capital adjustment cost and capital market imperfections.* Mimeo.

[10] Blundell, Richard, Crawford, Claire and Jin, Wenchao, 2014. *What Can Wages and Employment Tell Us about the UK's Productivity Puzzle?* The Economic Journal, May 2014. Volume 124, Issue 576, Pages 377-407.

[11] Bordon, Anna, Mico Mrkaic, and Kazuko Shirono, 2016. *A Firm-Level Analysis of Labor Productivity in the United Kingdom.* IMF Country Report No. 16/58, February 2016.

[12] Brandt, Nicola, 2004. *Business Dynamics and Policies.* OECD Economic Studies No. 38, 2004/1.

[13] Brinca, Pedro, V. V. Chari, Patrick J. Kehoe, and Ellen McGrattan, 2016. *Accounting for Business Cycles.* Minneapolis Fed Staff Report 531, June 2016.

[14] Buera, Francisco, and Moll, Benjamin, 2015. *Aggregate Implications of a Credit Crunch.* American Economic Journal: Macro, vol. 7, 2015.

[15] Buera, Francisco, Fattal-Jaef, Roberto, and Shin, Yongseok, 2015. *Anatomy of a Credit Crunch: From Capital to Labor Markets.* Review of Economic Dynamics, Elsevier for the Society for Economic Dynamics, vol. 18(1), pages 101-117, January.

[16] Chari, V. V., Kehoe, Patrick J., and McGrattan, Ellen R., 2007. *Business Cycle Accounting.* Econometrica, Econometric Society, vol. 75(3), pages 781-836, 05.

[17] Chodorow-Reich, Gabriel. 2014. *The Employment Effects of Credit Market Disruptions: Firm-level Evidence from the 2008-09 Financial Crisis.* Quarterly Journal of Economics 129 (1): 1-59.

[18] Cooper, Russell W. and John C. Haltiwanger, 2006. *On the Nature of Capital Adjustment Costs.* The Review of Economic Studies, Vol. 73, No. 3 (Jul., 2006), pp. 611-633

[19] Crawford, Claire, Wenchao Jin, and Helen Simpson, 2013. *Productivity, Investment and Profits during the Great Recession: Evidence from UK Firms and Workers.* Fiscal Studies, vol. 34, no. 2, pp. 153-177.

[20] Daly, Mary C. and Hobijn, Bart, 2014. *Downward Nominal Wage Rigidities Bend the Phillips Curve.* Federal Reserve Bank of San Francisco Working Paper 2013-08, January 2014.

[21] Daly, Mary C., Hobijn, Bart and Wiles, Theodore S., 2012. *Dissecting Aggregate Real Wage Fluctuations: Individual Wage Growth and the Composition Effect.* Federal Reserve Bank of San Francisco Working Paper Series. Working Paper 2011-23, May 2012.

[22] Daly, Mary C., John G. Fernald, Òscar Jordà, Fernanda Nechio, 2014. *Labor Markets in the Global Financial Crisis: The Good, the Bad and the Ugly.* National Institute Economic Review 228: R58-R64, May 2014.

[23] Druant, M., Fabiani, S., Kezdi, Gabor, Lamo, Ana, Martins, Fernando, and Sabbatini, R., 2009. *How are Firms' Wages and Prices Linked: Survey Evidence in Europe.* Working Papers w200918, Banco de Portugal, Economics and Research Department.

[24] Foster, Lucia, Cheryl Grim, John Haltiwanger, 2014. *Reallocation in the Great Recession: Cleansing or Not?* NBER Working Paper No. 20427.

[25] Franklin, Jeremy, May Rostom and Gregory Thwaites, 2015. *The banks that said no: banking relationships, credit supply and productivity in the United Kingdom.* Bank of England Staff Working Paper No. 557.

[26] Giroud, Xavier, Mueller, Holger M, 2015. *Firm Leverage and Unemployment during the Great Recession.* NBER Working Paper No. 21076. Issued in April 2015.

[27] Gomes, Pedro, 2012. *Labour market flows: Facts from the United Kingdom.* Labour Economics, Volume 19, Issue 2, April 2012, Pages 165-175.

[28] Goodridge Peter, Jonathan Haskel, and Gavin Wallis, 2015. *Accounting for the UK Productivity Puzzle: A Decomposition and Predictions.* Mimeo.

[29] Gopinath, Gita, Sebnem Kalemli-Ozcan, Loukas Karabarbounis, and Carolina Villegas-Sanchez, 2016. *Capital Allocation and Productivity in South Europe.* Harvard University working paper.

[30] Gregg, Paul, Machin, Stephen and Fernández-Salgado, Mariña, 2014. *Real Wages and Unemployment in the Big Squeeze.* The Economic Journal, May 2014. Volume 124, Issue 576, Pages 408-432.

[31] Hagedorn, Marcus, Karahan, Fatih, Manovskii, Iourii, and Mitman, Kurt, 2015. *Unemployment Benefits and Unemployment in the Great Recession: The Role of Macro Effects.* NBER Working Paper 19499, January 2015.

[32] Hagedorn, Marcus, and Manovskii, Iourii, 2015. *The Impact of Unemployment Benefit Extensions on Employment: The 2014 Employment Miracle?* NBER Working Paper 20884, January 2015.

[33] Hanlon, W. Walker, 2015. *Necessity Is the Mother of Invention: Input Supplies and Directed Technical Change.* Econometrica, Volume 83, Issue 1, pages 67-100, January 2015.

[34] Hathaway, Ian, and Robert Litan, 2014. *Declining Business Dynamism in the United States: A Look at States and Metros.* Economic Studies at Brookings, May 2014.

[35] Hopenhayn, Hugo and Rogerson, Richard, 1993. *Job Turnover and Policy Evaluation: A General Equilibrium Analysis.* Journal of Political Economy, University of Chicago Press, vol. 101(5), pages 915-38, October.

[36] Hsieh, Chang-Tai and Peter J. Klenow, 2009. *Misallocation and Manufacturing TFP in China and India.* The Quarterly Journal of Economics (2009) 124 (4): 1403-1448.

[37] Karabarbounis, Loukas, 2014. *The Labor Wedge: MRS vs. MPN.* Review of Economic Dynamics, Elsevier for the Society for Economic Dynamics, vol. 17(2), pages 206-223, April.

[38] Kaur, Supreet, 2014. *Nominal Wage Rigidity in Village Labor Markets.* Mimeo.

[39] Khan, Aubhik, Julia K. Thomas, 2008. *Idiosyncratic Shocks and the Role of Nonconvexities in Plant and Aggregate Investment Dynamics.* Econometrica, Volume 76, Issue 2, pages 395-436, March 2008.

[40] Khan, Aubhik, Julia K. Thomas, 2013. *Credit Shocks and Aggregate Fluctuations in an Economy with Production Heterogeneity.* Journal of Political Economy, University of Chicago Press, vol. 121(6), pages 1055 - 1107.

[41] Khan, Aubhik, Tatsuro Tenga and Julia K. Thomas, 2014. *Default Risk and Aggregate Fluctuations in an Economy with Production Heterogeneity.* Mimeo.

[42] Kurtzman, Robert J. and Zeke, David, 2016. *Accounting for Productivity Dispersion over the Business Cycle.* Finance and Economics Discussion Series 2016-045, Board of Governors of the Federal Reserve System.

[43] Leung, Danny, Césaire Meh, and Yaz Terajima, 2008. *Firm Size and Productivity.* Bank of Canada Working Paper 2008-45.

[44] Martin, John, and Scarpetta, Stefano, 2011. *Setting It Right: Employment Protection, Labour Reallocation and Productivity.* IZA Policy Paper No. 27, 2011.

[45] Mian, Atif, and Sufi, Amir, 2014. *What Explains the 2007-2009 Drop in Employment?* Econometrica, Volume 82, Issue 6, pages 2197-2223, November 2014.

[46] Midrigan, Virgiliu and Daniel Yi Xu, 2014. *Finance and Misallocation: Evidence from Plant-Level Data.* American Economic Review, Vol. 104 No. 2 February 2014.

[47] Moll, Benjamin, 2014. *Productivity Losses from Financial Frictions: Can Self-Financing Undo Capital Misallocation?* American Economic Review 2014, 104(10): 3186-3221.

[48] Mulligan, Casey, 2015. *Fiscal Policies and the Prices of Labor: A Comparison of the U.K. and U.S.* NBER Working Paper No. 21358, July 2015.

[49] Ohanian, Lee E. 2010. *The Economic Crisis from a Neoclassical Perspective.* Journal of Economic Perspectives, 24(4): 45-66.

[50] Olivei, Giovanni and Tenreyro, Silvana, 2007. *The Timing of Monetary Policy Shocks.* American Economic Review, American Economic Association, vol. 97(3), pages 636-663, June.

[51] Olivei, Giovanni and Tenreyro, Silvana, 2010. *Wage-setting patterns and monetary policy: International evidence.* Journal of Monetary Economics, Elsevier, vol. 57(7), pages 785-802, October.

[52] Pessoa, João Paulo and Van Reenen, John, 2014. *The UK Productivity and Jobs Puzzle: Does the Answer Lie in Wage Flexibility?* The Economic Journal, May 2014. Volume 124, Issue 576, Pages 289-699.

[53] Petrosky-Nadeau, Nicolas, 2013. *TFP during a credit crunch.* Journal of Economic Theory, Elsevier, vol. 148(3), pages 1150-1178.

[54] Rajan, Raghuram, and Luigi Zingales, 1995. *What Do We Know about Capital Structure? Some Evidence from International Data.* The Journal of Finance, Vol L, No. 5, December 1995.

[55] Reinhart, Carmen, and Rogoff, Kenneth, 2009. *This Time is Different. Eight Centuries of Financial Folly.* Princeton University Press.

[56] Reis, Ricardo, 2013. *The Portuguese Slump and Crash and the Euro Crisis.* Brookings Papers on Economic Activity, March 21, 2013.

[57] Rhodes, Chris, 2015. *Business Statistics.* Briefing paper, Number 06152, 7 December 2015, House of Commons Library.

[58] Riley, Rebecca, Chiara Rosazza-Bondibene and Garry Young, 2015. *The UK productivity puzzle 200813: evidence from British businesses.* Bank of England Staff Working Paper No. 531.

[59] Sahin, Aysegul, Sagiri Kitao, Anna Cororaton, and Sergiu Laiu, 2011. *Why Small Businesses Were Hit Harder by the Recent Recession.* Current Issues in Economics and Finance, Volume 17, Number 42011, Federal Reserve Bank of New York.

[60] Siemer, Michael, 2014. *Firm Entry and Employment Dynamics in the Great Recession.* Finance and Economics Discussion Series 2014-56, Board of Governors of the Federal Reserve System.

[61] Van Ark, B., and E. Monnikhof, 1996. *Size Distribution of Output and Employment: A Data Set For Manufacturing Industries in Five OECD Countries, 1960s-1990.* OECD Economics Department Working Paper No. 166.

[62] Wang, Yicheng, 2015. *Can Wage Dynamics in Long-term Employment Relationships Help Mitigate Financial Shocks?* Mimeo.

[63] Young, Eric, 2009. *Solving the Incomplete Markets Model with Aggregate Uncertainty Using the Krusell-Smith Algorithm and Non-Stochastic Simulations.* Mimeo.

Appendices

A Empirical appendix

A.1 Data sources

A.1.1 Aggregate data

US. Notes: All data are seasonally adjusted. The quarterly national accounts data are presented in yearly rates, and are thus divided by 4 to get quarterly values. The labour series is normalised to have average 1/3 over the sample. Consumption and investment are deflated by the GDP deflator and not their individual deflators. This is standard in the RBC model, since the model does not allow for movements in the relative prices of output, consumption or investment.

- GDP (Y_t): Chained value taken from line 1 of NIPA table 1.1.6. Deflator taken from line 1 of NIPA table 1.1.4.
- Consumption (C_t): non-durables plus services. Nominal, then deflated by gdp deflator. Taken from line 5 and 6 of NIPA table 1.1.6.
- Investment (X_t): Gross domestic private investment (NIPA 1.1.6 line 7) + durable consumption (NIPA 1.1.6 line 4). Both nominal, deflated by GDP deflator. Note that I treat investment from date t in the table as not operational until $t+1$.
- Initial capital stock (K_0): Year end capital stock constructed from BEA Fixed Asset table 1.1 (yearly data). To match investment data, use Private Fixed Assets (line 3) + Consumer Durables (line 15). These are current cost measures, and I deflate by the yearly GDP deflator. My initial capital stock is the year end value the year before my investment data begins.
- Depreciation (δ): Current cost depreciation data from Fixed Asset table 1.3. Computed for private fixed assets (line 3) + consumer durables (line 15) and again deflated by the GDP deflator. Yearly depreciation rates are computed as $\delta_{y,t} = Dep_t/K_{t-1}$. Implied quarterly depreciation rates are computed as the solution to $\delta_{q,t} + \delta_{q,t}(1-\delta_{q,t}) + \delta_{q,t}(1-\delta_{q,t})^2 + \delta_{q,t}(1-\delta_{q,t})^3 = \delta_{y,t}$. I compute the depreciation rate as the average of these quarterly depreciation rates.
- Capital stock (K_t): Constructed using the perpetual inventory method starting with the initial capital stock and using $K_t = X_t + (1-\delta)K_{t-1}$.
- Hours worked (L_t): Hours worked is the series "Nonfarm Business Sector: Hours of All Persons, Index 2009=100, Quarterly, Seasonally Adjusted", series HOANBS downloaded from FRED.
- Population (N_t): "Civilian Noninstitutional Population, Thousands of Persons, Monthly, Not Seasonally Adjusted" available from the Federal Reserve Economic Data (FRED). I take every third datapoint to construct quarterly data, and then take a one year moving average to deseasonalise.
- Labour share (LS_t): The data for the labour share come from Gross Domestic Income data, since income breakdowns are not available for the GDP data. Since there are small discrepancies between the GDI and GDP data, instead of taking the wage directly from the GDI data,

I simply compute the labour share from this data. I can then use the labour share (under the assumption that it is equal in the GDI and GDP data) to back out the implied wage consistent with the GDP data. The nominal GDI data is from NIPA table 1.10 and are constructed along the lines of Karabarbounis (2014). I first construct the unadjusted labour share, LS_t^u as the share of unambiguous labour income to GDI. This is compensation of employees paid (line 2) divided by GDI (line 1). I then attribute the fraction LS_t^u of ambiguous income to labour. Ambiguous income is "Proprietors' income with inventory valuation and capital consumption adjustments" (line 13) + "Taxes on production and imports" (line 7) - "Less: Subsidies" (line 8). The final labour share is then computed as unambiguous labour income plus the fraction LS_t^u of ambiguous income all divided by GDI.

- Real Wages (w_t): The real wage is calculated as $w_t = LS_t y_t / l_t$

UK. The available UK data which is comparable to the US data is limited, and starts from 1997Q1. In particular, a longer time series for capital is available for the UK, but it does not distinguish between government capital and private capital. In the wedges exercise for the US, following Chari, Kehoe and McGratten (2007), capital is defined as only private capital. For this reason I am restricted to data for the UK from 1997, when a breakdown of the capital stock was first released. Most data are from the UK quarterly national accounts, unless otherwise stated.

- GDP (Y_t): The GDP deflator is series "Implied Deflators: Gross domestic product at market prices" (series YBHA) from table "A1: National Accounts Aggregates". Nominal GDP is the series "Current prices: Gross domestic product at market prices" (series YBGB) from table "A2: National Accounts Aggregates". Real GDP is calculated as nominal deflated by the price deflator.

- Consumption (C_t): non-durables, semi-durables and services. Nominal, then deflated by GDP deflator. Taken from series UTIR, UTIJ and UTIN of table "E2: Household final consumption expenditure (goods and serives) at current prices".

- Investment (X_t): Business investment + Private Sector: Dwellings + Private Sector: Costs of Ownership + Durable Consumption. Investment data from table "G9: Gross Fixed Capital Formation by sector and type of asset". Series: NPEK, L62T and L62U. Durable consumption data from table E2, series UTIB. Nominal, then deflated by GDP deflator.

- Initial capital stock (K_0): Unfortunately, data on the stock of the specific sub-types of investment which I use for my capital series are not available for 1996Q4. I thus experiment with various initial stocks, and the results for the Great Recession episode are not sensitive to the initial choice. The initial choice I settle on is the one that leads to a path for the Solow residual which looks closest to being one with stationary fluctuations around a constant growth rate.

- Depreciation (δ): Again, individual depreciation rates are not available for the sub-types of capital. Since the sub types were chosen to be similar to those for the US, I take the same depreciation rate as the US.

- Capital stock (K_t): Constructed using the perpetual inventory method starting with the initial capital stock and using $K_t = X_t + (1-\delta)K_{t-1}$.

- Hours worked (L_t): Hours worked is the series "HOUR01 Actual weekly hours of work", seasonally adjusted, available from the ONS.
- Population (N_t): Annual "Mid-year population estimates and annual change for the UK mid-1964 onwards" data are interpolated to quarterly data by assuming a constant within-year quarterly growth rate. Data available at `http://www.ons.gov.uk/ons/rel/pop-estimate/population-estimates-for-uk--england-and-wales--scotland-and-northern-ireland/2013/sty-population-changes.html` at time of writing.
- Labour share (LS_t): Data are constructed as for the US, using the table "D: Gross Domestic Product: by category of income". Unambiguous labour income is taken as "UK Total compensation of employees" (DTWM). Ambiguous income is taken as "Other income" (CGBX). Total income is "Income based GVA at factor cost" (CGCB).
- Real Wages (w_t): The real wage is calculated as $w_t = LS_t y_t / l_t$

A.1.2 Sectoral level data

Sectoral level data is used to construct the GDP and employment data excluding the financial sector. For the UK, the employment and value added data are accessed at the two digit SIC2007 level. Value added data are from the ONS dataset "UK GDP (O) Low Level Aggregate", and employment data from "JOBS03: Employee Jobs by Industry". The totals excluding finance exclude sectors 64-66: "Financial service activities, except insurance and pension funding", "Insurance, reinsurance and pension funding", and "Activities auxilliary to financial services and insurance activities".

For the US, value added data are from the BEA, provided at the two digit BEA code level. To exclude finance, I exclude "Finance and Insurance", which is line 55 in the BEA tables. By BEA code I exclude "521CI: Federal Reserve banks, credit intermediation, and related activities", "523: Securities, commodity contracts, and investments", "524: Insurance carriers and related activities", and "525: Funds, trusts, and other financial vehicles". The employment data are from the Current Employment Statistics survey. These are provided by 2007 NAICS code, which I match to the BEA codes.

A.1.3 Labour market transitions

Data on labour market flows for the UK are published under the series "X02: Labour market flows". These are available from the ONS, and calculated from the Labour Force Survey. They can be accessed here: `http://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeet` `articles/labourmarketflows/2015-05-13`.Data for the US are from the Current Population Survey, distributed by the BLS, available here: `http://www.bls.gov/webapps/legacy/cpsflowstab.htm`. Data for the US are monthly, and data for the UK are quarterly.

There are three states: employed, e, unemployed, u, and not in the labour force, n. For any state, $x = \{e, u, n\}$, the stock of workers at the beginning of period t satisfies

$$x_t = f_t^{e,x} + f_t^{u,x} + f_t^{n,x} + f_t^{m,x} \tag{48}$$

Where $f_t^{y,x}$ is the flow of workers to x from y between $t-1$ and t. The state m represents flows in and out of the sample. I convert this to flow rates:

$$x_t = \rho_t^{e,x} e_{t-1} + \rho_t^{u,x} u_{t-1} + \rho_t^{n,x} n_{t-1} + f_t^{m,x} \tag{49}$$

Where $\rho_t^{y,x} \equiv f_t^{y,x}/y_{t-1}$. With knowledge of an initial state (e_0, u_0, n_0) and all flow rates $f_t^{y,x}$ these equations can be simulated forward to exactly recover the data. Counterfactual simulations are calculated by replacing certain true flow rates with averages from the pre-crisis data, as described in the text. Averages are computed for the year 2007. There are nine rates tracking the flows between the three states, which I can construct period by period. Starting from initial stocks at some time t and simulating forward with the flows data will trivially recover the resulting series.

I am interested in the evolution of the employment rate, $er_t \equiv e_t/p_t$, where p_t is the population. I construct two counterfactual series representing the contribution of job destruction rates and creation rates respectively. These are constructed by starting from the 2008 stocks, and simulating forward while allowing only these flows to vary, while holding the others at their pre-2008 averages. Counterfactual employment allowing only job destruction to vary is calculated using the true rates for transitions *from* employment, $(\rho_{ee,t}, \rho_{eu,t}, \rho_{en,t})$, while holding the other flows constant. Counterfactual employment allowing only creation rates to vary is calculated using the true rates for all other transitions, $(\rho_{ue,t}, \rho_{uu,t}, \rho_{un,t}, \rho_{ne,t}, \rho_{nu,t}, \rho_{nn,t})$, while holding the destruction rates constant. These simulations create counterfactual series for e_t, u_t, and n_t, which I use to construct the counterfactual employment rate series, er_t^{jd} and er_t^{jc}. These are plotted in Figure 6, along with the actual data.

A.2 Quality adjusted labour input and unemployment

The fact that the labour market performed worse in the US during the crisis is also apparent in other measures of labour market performance. In Figure 15 I plot Quality Adjusted Labour Input (QALI, per capita) and the unemployment rate for both countries.[35]

Again, the performance is worse in the US. QALI behaviour is quantitatively similar to the headline hours data, with peak falls of 10% in the US and 5% in the UK. The US fared worse with unemployment too, with unemployment starting at 4% and peaking at 10%, while it started at 5% and peaked at 8% in the UK. The longer run differences are, however, less pronounced when looking at the unemployment rate, with the rate falling to 6% in both countries by 2014 while their employment rates (and hours) still diverged. The difference can be explained by a larger decrease in participation in the US.[36]

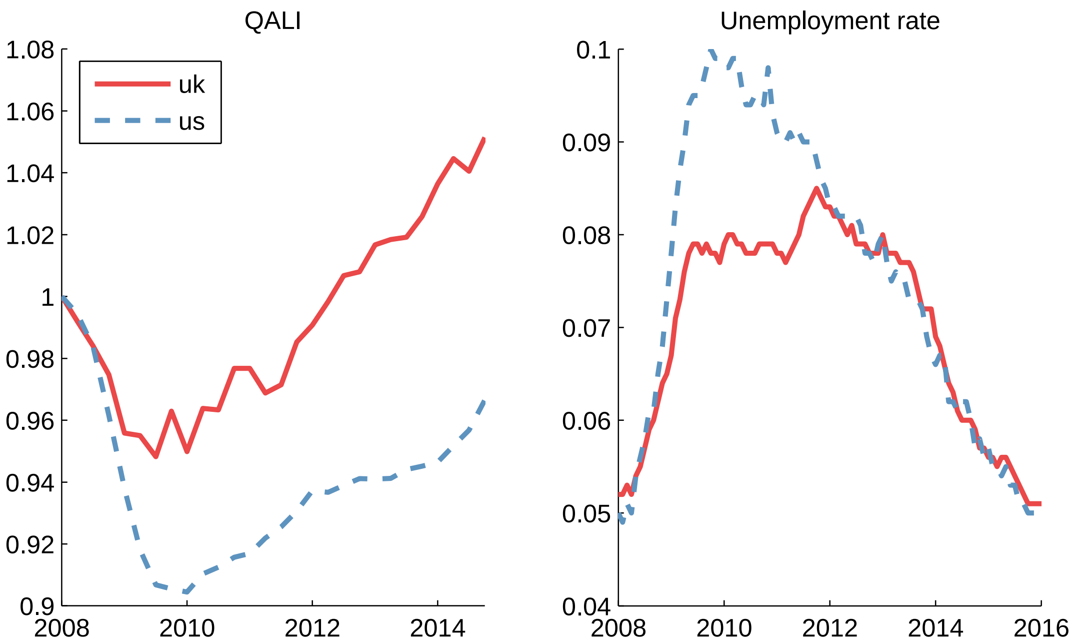
A.3 Labour market policies during the Great Recession

Gregg, Machin, and Fernandez-Salgado (2014) show that unemployment has become more of a moderating force on real wages in the UK, and even identify a structural break in the unemployment-wage relationship around 2003. They show that union wages are less responsive to unemployment than non-union wages, and hence argue that declining union membership can partly explain why overall wages have become more flexible.

[35]QALI measures the total hours worked in the economy adjusted for the quality of different types of labour. Higher productivity (measured by wage) labour is given a higher weight. QALI data for the UK are from the ONS release "Quality Adjusted Labour Input, Estimates to 2014", and for the US from Basu, Fernald, and Kimball (2006, henceforth BFK). Unemployment data are from the ONS and BLS.

[36]This decrease in participation is broad based across age groups, and thus unlikely to be driven purely by demographics.

Figure 15: Quality Adjusted Labour Input and Unemployment



QALI data are from the ONS for the UK, and Basu, Fernald, and Kimball (2006) for the US, and are per capita, and expressed as percentage deviations from 2008Q1. The unemployment rate data is from the ONS and BLS using the ILO definition of unemployment.

In the US, on the other hand, some papers (Hagedorn, Karahan, Manovskii, and Mitman, 2015, Hagedorn, Manovskii and Mitman, 2015) have argued that the extension of unemployment benefits during the Great Recession could have put upwards pressure on wages.[37] They claim that this is due to a "macro effect" whereby more generous unemployment benefits increase the outside option of workers, increasing their bargaining power and hence pushing up wages. Consistent with this, they find a statistically significant, positive relationship between wages and benefits using their identification strategy.

In a recent paper, Mulligan (2015) analyses the effects of taxes and fiscal policies on the incentives of firms to hire and workers to provide labour in the US and UK. He finds that during the first three years of the crisis, changes in taxes and subsidies in the US reduced employees' reward to work, whereas changes in the UK actually increased employees' reward to work. These policies, by relatively improving the outside option of not working in the US and worsening it in the UK, could explain why real wages behaved so differently in the first three years of the crisis.

[37]Hagedorn, Karahan, Manovskii, and Mitman (2015) use an identification strategy that exploits a policy discontinuity at state borders to estimate that unemployment in the US could have been as much as 2.5pp higher in 2011 due to the extension of benefits. Using the same methodology, Hagedorn, Manovskii and Mitman (2015) argue that 61% of the increase in employment in 2014 can be attributed to the expiration of the benefit extension.

A.4 Business Cycle Accounting

In this section I perform Chari, Kehoe, and McGratten's (2007) business cycle accounting procedure for both countries during the financial crisis. Technical details of how I do this can be found in the PhD version of this paper (Clymo, 2016), and I present a brief discussion here.

The prototype economy is standard. There is a representative household with growing population N_t. It has log utility over per-capita consumption, c_t, and convex disutility over hours worked per capita, l_t:

$$U(c_t, l_t) = \log(c_t) - v(l_t) \tag{50}$$

The household discounts the future with the discount factor β. The household's budget constraint is:

$$c_t + (1 + g_{N,t})k_t - (1 - \delta)k_{t-1} = (1 - \tau_t^l)w_t l_t + (1 - \tau_{t-1}^x)r_t k_{t-1} \tag{51}$$

Where $k_{t-1} \equiv K_{t-1}/N_t$ is per-capita capital which is productive at time t. w_t is the hourly wage, and capital depreciates at rate δ. $g_{N,t}$ is the population growth rate between t and $t+1$. τ_t^l and τ_t^x are percentage taxes on labour and capital income respectively. I assume that the tax on capital income is known at the time the relevant investment decision is made, namely one period in advance. There is a representative firm with Cobb-Douglas production function, $Y_t = Z_t K_{t-1}^{\alpha} L_t^{1-\alpha}$, whose static optimality conditions equate prices with marginal products. The equilibrium of the economy can be summarised by the following four equations:

$$\tilde{y}_t = e^{\tau_t^e} \tilde{k}_{t-1}^{\alpha} l_t^{1-\alpha} \tag{52}$$

$$v'(l_t)\tilde{c}_t = (1 - \tau_t^l)(1 - \alpha)\frac{\tilde{y}_t}{l_t} \tag{53}$$

$$1 = \frac{\beta}{1 + g_z} \mathrm{E}_t \left[\frac{\tilde{c}_t}{\tilde{c}_{t+1}} \left((1 - \tau_t^x)\alpha \frac{\tilde{y}_{t+1}}{\tilde{k}_t} + 1 - \delta \right) \right] \tag{54}$$

$$\tilde{c}_t + (1 + g_{N,t})(1 + g_z)\tilde{k}_t - (1 - \delta)\tilde{k}_{t-1} + \tau_t^g \tilde{y}_t = \tilde{y}_t \tag{55}$$

Lowercase variables with a tilde refer to per-capital variables which have been detrended by the average growth rate of TFP, g_z.[38] Each of these equations corresponds to a different wedge, which we can measure using data on output, capital, hours and consumption. The first equation, the aggregate production function, identifies the "efficiency wedge", τ_t^e. This is the log deviation of measured TFP from trend. The second equation, the labour market optimality condition, identifies the "labour wedge", τ_t^l. This is our labour income tax from the model, and the data measures it as the required labour tax in order to rationalise the observed level of hours. The third equation, the Euler equation, identifies the "investment wedge". This is our capital income tax, and the data measures it as the tax required to rationalise the level of investment. The final equation, the resource constraint, identifies the "government wedge", τ_t^g. This is measured from the data as the residual in the national accounts after subtracting consumption and investment from output, and hence actually contains both government spending and net exports. It is expressed for convenience as a fraction of output. Define the set of time t wedges as $\tau_t = \{\tau_t^e, \tau_t^l, \tau_t^x, \tau_t^g\}$.

[38] Z_t is measured TFP, and $e^{\tau_t^e}$ is thus the deviation of measured TFP from trend. For any variable x_t, $\tilde{x}_t \equiv x_t/(1 + g_z)^t$ except for capital, for which $\tilde{k}_{t-1} \equiv k_{t-1}/(1 + g_z)^t$.

A.4.1 Results

As is commonly found in these exercises,[39] the investment and government wedges account for very little movement in output, and hence I focus on the efficiency and labour wedges. Figure 16 summarises the results, with the top row showing graphs for the US, and the bottom the UK. The first column summarises the exercise from the beginning of the financial crisis up to the most recently available data. My data is quarterly, so I start in the first quarter of 2008.[40] The top left panel shows the dramatic fall in US output over this period relative to trend: a rapid fall of nearly 7% over the first year, followed by a further gradual decline. We see that the labour wedge is able to account very well for the fall in output, even slightly over-predicting it initially, and continuing to provide a reasonable account over the whole period. The efficiency wedge, on the other hand, is not able to account for the dynamics of output at all, and predicts that output should be slightly above trend over the whole period.

The UK shows the opposite pattern, with the efficiency wedge doing a better job at explaining output than the labour wedge. The bottom left panel documents the dramatic fall in output relative to trend, which is around 9% within the first year, and consistently two percentage points worse than the US over the period. The efficiency wedge explains the initial fall well, and continues to explain the bulk of the fall in output over the period. The labour wedge, while getting the sign of the output movements right, cannot match the magnitude of the fall, and additionally predicts that output should have returned to trend by 2014.

The remaining two columns provide additional information. The middle column plots the data on hours worked. A reflection of the more important labour wedge in the US is the worse performance of hours during the recession: it fell by nearly 11% in the US, whereas it fell by just over 5% in the UK. As is to be expected, the labour wedge is important for explaining the movements in hours in both the US and the UK: it is well known that the benchmark RBC model is unable to generate the required movements in hours worked, which explains why the labour wedge still plays a role.

The final column plots the wedges themselves. For the efficiency wedge, I plot the exponential of the wedge, which gives the deviation of TFP from trend. For the US, we see a peak increase of over 2% in detrended TFP during the crisis, which is eventually reversed, but still leaves detrended TFP less than 1% below trend by the end of the sample period. The UK, on the other hand, has TFP fall by 5% from trend within the first year, and continue falling to around 7% below trend by the end of the sample period. The labour wedge increases in both countries, but by roughly twice as much in the US. Additionally, the US labour wedge remains severely elevated at the end of the sample period, whereas the UK labour wedge actually ends the sample less severe than at the beginning.

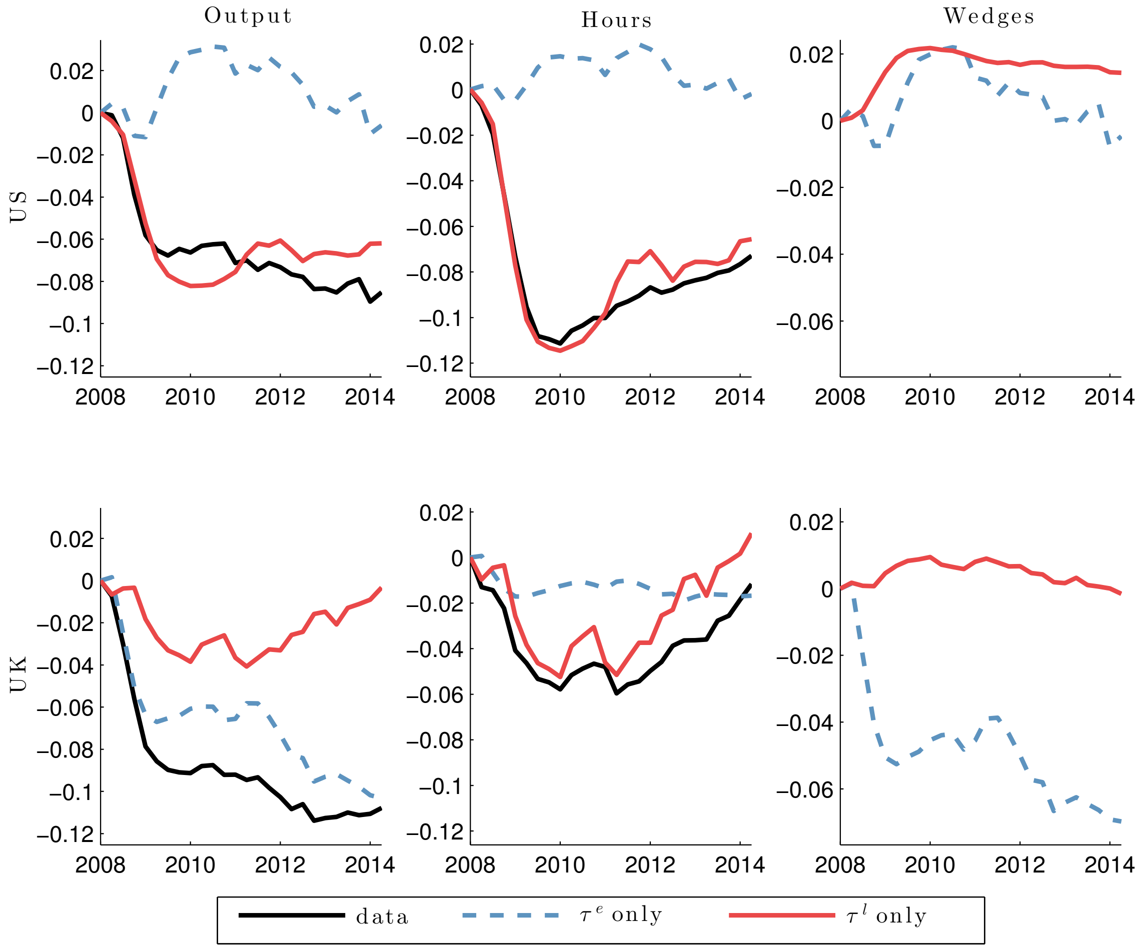
A.4.2 Decomposing the labour wedge

In this section I perform a decomposition of the labour wedge along the lines of Karabarbounis (2014), which is used to ask whether the labour wedge is caused mostly by distortions on the firm or consumer side. The idea is as follows. Suppose that instead of just considering a prototype economy where the consumer pays a labour income tax, we consider an economy where the consumer pays a labour

[39] See, for example, the original Chari, Kehoe, and McGratten (2007) paper.

[40] The NBER dates the US recession as starting in December 2007, and the results are robust to moving the start point by a few quarters in either direction. Interestingly, the data actually suggests that, relative to the long term trend, the US (but not UK) economy started slowing down in early 2006, and that this slowdown is associated with the efficiency wedge. Since I want to focus on financial crises , and this period precedes the events of 2008, I will not focus on it in this paper.

Figure 16: Business cycle accounting results



All variables are expressed as a fractional deviation from the value in the initial period. The first two columns plot output and hours, and their simulated paths subject only to one wedge. The final column plots the exponential of the efficiency wedge (giving the deviation of TFP from trend), and the value of the labour wedge. Output is expressed as the deviation from the estimated trend.

income tax, τ_t^{lc}, and the firm also pays a payroll tax, τ_t^{lf}. These are the decomposed labour wedges, which I'll refer to as the "consumer side" and "firm side" labour wedges, respectively. Then if we have data on real wages, the consumer and firm's labour optimality conditions can be rearranged to solve for these two wedges:

$$\tau_t^{lc} = 1 - \frac{v'(l_t)}{w_t u'(c_t)} \tag{56}$$

$$\tau_t^{lf} = (1-\alpha)\frac{y_t}{w_t l_t} - 1 \tag{57}$$

And it follows trivially from the definition of these wedges that they decompose the overall labour wedge in the following sense: $(1-\tau_t^l) = (1-\tau_t^{lc})/(1+\tau_t^{lf})$. This can be rearranged to construct the overall labour wedge from our two sub wedges:

$$\tau_t^l = 1 - \frac{1 - \tau_t^{lc}}{1 + \tau_t^{lf}} \tag{58}$$

The consumer-side labour wedge measures the wedge between the wage and the consumer's marginal rate of substitution between labour and consumption. It hence measures whether the representative consumer's labour supply is being distorted away from its optimal value. The firm-side labour wedge measures the wedge between the wage and the marginal product of labour. It hence measures whether the representative firm's labour demand is being distorted.

This decomposition thus allows us to see whether the labour wedge we see, especially the wedge in the US during the crisis, arises more from distortions on the firm or consumer side. To do this, I reconstruct a hypothetical labour wedge for the crisis for each country and each sub-wedge using (58), and allowing one sub-wedge to vary at a time while holding the other at its initial value. The results of this exercise are presented in the first column of Figure 17. For both the US and the UK we see that the cyclical movements in the labour wedge over the crisis are driven almost entirely by the consumer side labour wedge. This is also true of the simulated paths for output and labour: if I use these hypothetical wedges in the simulation exercise it is the consumer-side wedges which deliver movements closest to the data.

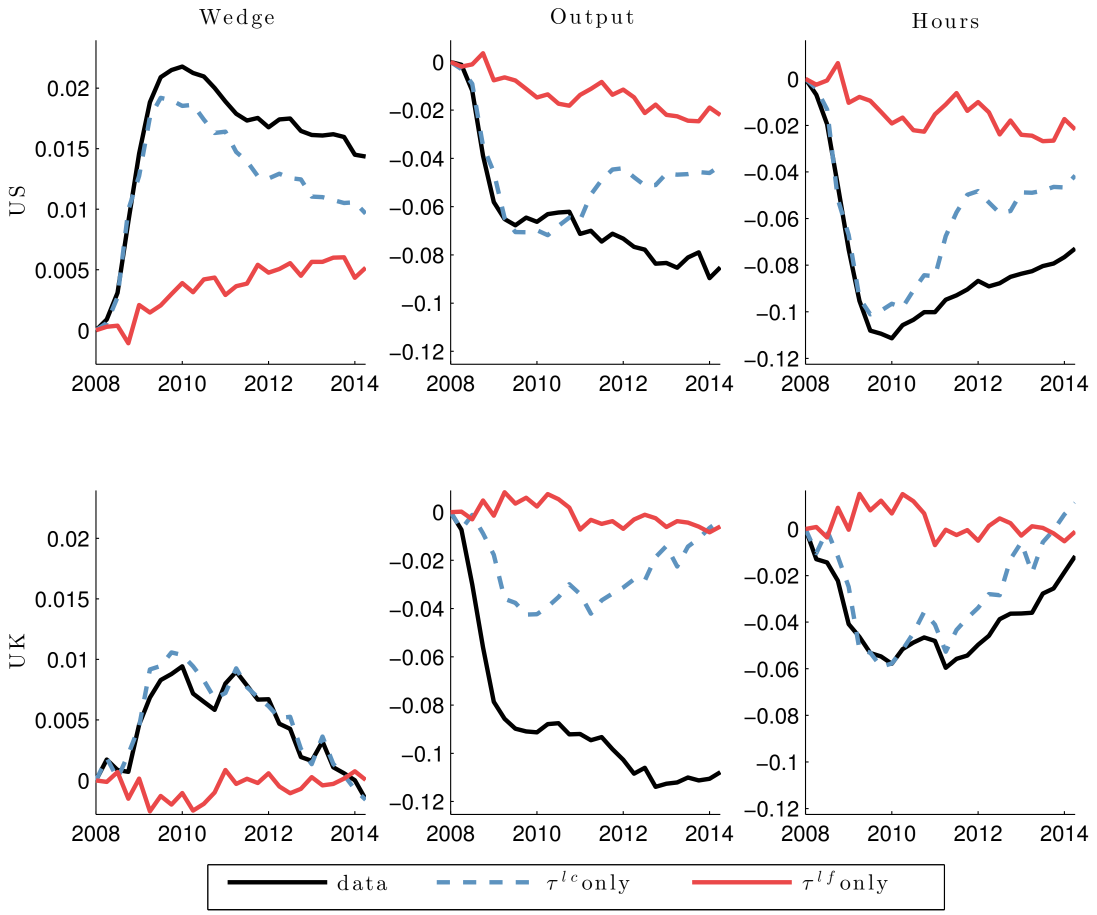
In other words, we can think of firms as being roughly on their labour demand curves, and it is consumers whose labour supply decisions are being distorted. Given the sign of the labour wedge, this means that during the crisis we can think of consumers as wanting to work *more* given the current wage, and firms being happy with their employment levels given the current wage. This evidence is consistent with a model of rationing unemployment, where the wage is stuck above the market clearing level. In this scenario, firms are on their demand curves, and hire little. Consumers are rationed and work less than the desired amount, taking them off their supply curves and causing the large consumer-side labour wedge. In my model I assume sticky wages, leading to rationing unemployment of the type discussed above. This evidence is supportive of this assumption, and further implies that my model is able to match not only movements in the labour wedge, but also movements in the two sub wedges.

A.5 Real unit labour costs

One might be concerned that wages are not the right measure of labour costs for my exercise. In particular, it is also common to look at "real unit labour cost" (RULC), which is defined as the wage bill divided by output: wL/Y. The argument for this measure is that it measures the labour cost of producing one unit of output, and hence measures labour costs better than the wage. In this section I argue that this is not the right measure for my exercise, and that wages are appropriate.

Firstly, in a simplified setting without composition effects from the labour side, wages are clearly the correct measure of labour costs for my model. Theoretically, the composition of firms is determined by wages, not RULC, as shown in my analytical model. Additionally, RULC is a very endogenous object, even if wages are fully exogenous. To see this, note that in my analytical model, RULC is always equal to $(1-\alpha)\rho$ by firm optimisation. This is because, regardless of the wage, firms adjust hours until $MPL = w$. Similarly, in the basic RBC model without monopolistic competition RULC are constant at $1-\alpha$. It is hard to think about how to interpret movements in a variable which should be constant at $1-\alpha$ in the baseline macroeconomic production function.

Figure 17: Decomposing the labour wedge



All variables are expressed as a fractional deviation from the value in the initial period. In the first column, "data" refers to the measured value of the overall labour wedge, τ_t^l, and the other two lines compute the value of the labour wedge would take using (58) if we allow one of the sub-wedges to vary, and hold the other at its first period value. In the other two columns, the counterfactuals for output and labour are computed using these counterfactual labour wedges. Output is detrended by the estimated trend growth rate for TFP.

Of course, RULC does move in the data. However, recent work has shown that these movements are relatively small compared to the size of fluctuations in the labour market. This is the message of the labour wedge decomposition of Karabarbounis (2014). The firm-side labour wedge actually measures movements in RULC, and its movements can explain essentially none of the movements in output or hours at business cycle frequencies. I confirm this result for the Great Recession in the US and UK in Appendix A.4.2.

Secondly, even if we add labour-side composition effects it is not clear that RULC controls for these. This is the supposed benefit of RULC, since by measuring the total cost of labour per unit of output it controls for workforce composition. In the appendix of Clymo (2016) I show that this claim may not be true, since the result that RULC should be constant also holds in a model where worker composition varies, if workers are perfect substitutes for each other. In the main text of this paper, I examine data which deals with composition effects more directly by using worker-level data to strip out composition effects.

A.6 International evidence

In this section I show that the my stylised facts hold across a broader range of countries using data from the OECD.

A.6.1 Data sources

Most of the cross-country data are from the OECD dataset "Growth in GDP per capital, productivity and ULC", which I denote OECD1, and is available at `https://stats.oecd.org/Index.aspx?DataSetCode=PDB_GR` at the time of writing. Selected variables for certain countries have been replaced with supplementary series in order to fill in missing data, and I document any changes below. All data are yearly.

- TFP: "Multifactor productivity" series from OECD1.
- Hours: Hours per capita is calculated as "Total hours worked" from OECD1 divided by "Total population; persons; thousands" from the OECD dataset "Level of GDP per capita and productivity", which is available at `https://stats.oecd.org/Index.aspx?DataSetCode=PDB_LV` at the time of writing.
- Wages: The nominal wage is taken as "Labour compensation per hour worked" from OECD1.
  - The wage series for New Zealand is replaced by the wage calculated from preliminary quarterly estimates to extend the series. Specifically, I take Labour Compensation per Employed Person from the OECD dataset "Unit Labout Costs and labour productivity (employment based), total economy" (available at `http://stats.oecd.org/Index.aspx?DataSetCode=ULC_EEQ` at the time of writing) which is then converted to an hourly wage using data on employment and hours from OECD1.
  - The wage series for Australia is replaced using Australian national accounts data. Specifically, I take Compensation of Employees from the national accounts (available at `http://www.abs.gov.au/` at the time of writing) and divide by hours from OECD1 to create a wage series.
- Prices: The price level is taken as the GDP deflator, which is calculated from OECD1 as real GDP ("GDP, constant prices") over nominal GDP ("Gross Domestic Product (GDP); millions").
- Credit intermediation ratio: Taken from the OECD dataset "Financial Indicators – Stocks", which is available at `https://stats.oecd.org/Index.aspx?DataSetCode=FIN_IND_FBS` at the time of writing.

From the total of 20 countries in the database I drop Switzerland due to missing wage data, and Korea due to irregularities in the wage data. Ireland is initially left in the sample, but is dropped as an outlier.[41] This leaves a total of 17 countries in the baseline comparison. The OECD data is annual, and runs up to 2011, starting at different dates for different countries, the earliest being 1970. I take the Great Recession to be the period from 2008 to 2011, and study correlations between various log-changes over this period. I detrend TFP using a constant growth rate estimated for each country using all available pre-crisis data. I also detrend the real wage rate using the same deterministic time trend. Hours worked is expressed in per capita terms, and not detrended. The price level is expressed in levels and is not detrended.

A.6.2 Simple correlations

I perform two exercises with this data. The first is a simple cross-country comparison. For each country, I construct the log change in variable x over the Great Recession as $\tilde{x}_i = \log(x_{i,2011}/x_{i,2007})$. I plot selected relationships between my four variables (detrended TFP, z, hours per capita, l, detrended real wages, w, and the price level, P) in Figure 18.

The top left panel gives the key correlation: countries which experienced larger falls in hours over the recession tended to experience less severe declines in TFP. The top right and bottom left panels give the correlations between wages and TFP and hours respectively. Countries which experienced higher wage growth experienced higher TFP growth. On the other hand, countries which experienced higher wage growth experienced larger falls in hours. Notice that these three panels all speak against a simple TFP shock interpretation of the data: in that case we would expect a positive correlation between hours, wages and TFP. Indeed, the downward slope of the relationship between hours and wages suggests that movements along the labour demand curve dominate the evolution of labour markets over the period. Finally, the bottom right panel plots the correlation between price changes and real wage changes over the period. Countries with higher inflation experienced lower real wage growth. This suggests a role for sticky nominal wages combined with differential inflation outcomes in determining real wages.

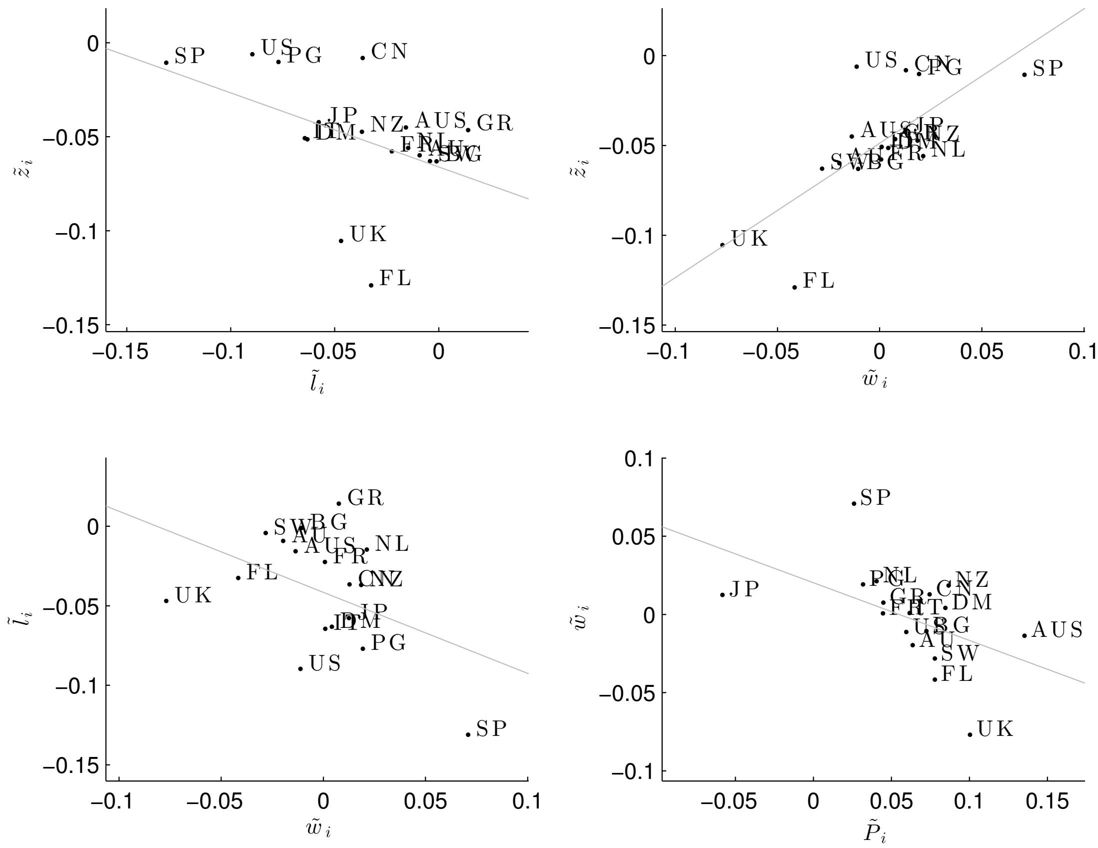
A.6.3 Partial correlations

While I am not looking to uncover causal relationships, there are obviously problems with taking such a simplistic cut of the data. For example, the correlations above could be driven not by differential wage behaviour during the recession (as my story claims), but simply by preexisting differences across countries. I would thus like to construct evidence that the above correlations hold within a hypothetical country over the recession. I do this by creating a short panel structure from my data. Instead of just looking across countries, I now look across both countries and time. Specifically, for each variable x I construct the log change from 2007 to year t in country i: $\tilde{x}_{i,t} = \log(x_{i,t}/x_{i,2007})$. I do this for $t = \{2008, 2009, 2010, 2011\}$ giving me four years of data across 17 countries, and a total of 68 data points. Intuitively, this lets me look at the relationship between variables both across and within countries.

Partial correlations capture the relationships between variables after controlling for their relationships with other variables. For example, the relationship between TFP and hours after controlling

[41] Ireland has been dropped as an outlier because its experience has been extreme relative to the rest of the sample. Ireland experienced the worst fall in hours (over 20%) in the sample, and third worst fall in TFP (over 10%). While the negative TFP-hours correlation holds robustly across the rest of the sample, Ireland is clearly a counterexample of a country which experienced both very bad TFP and very bad labour market performance and is worthy of further independent study.

Figure 18: Relationships between selected variables across countries.



Lines are OLS lines of best fit between the two variables. The country names refer to: AUS = Australia, AU = Austria, BG = Belgium, CN = Canada, DM = Denmark, FL = Finland, FR = France, GR = Germany, IT = Italy, JP = Japan, NL = Netherlands, NZ = New Zealand, PG = Portugal, SP = Spain, SW = Sweden, UK = United Kingdom, US = United States.

for the fact that some of their correlation derives from the fact that they both depend on country characteristics. I control for country, f_i, and time, q_t, fixed effects, allowing me to focus on variation in the variables unrelated to country characteristics, and the year of the recession. I also control for the size of the financial shock in a given country-year using data on the "credit intermediation ratio" from the OECD National Accounts. This is the ratio of loans from the financial sector to the non-financial sector to the total liabilities of the non-financial sector. In other words, it is a measure of the ability of non-financial firms to raise funds from the financial sector.[42] The first step is to

[42] This ratio has strong predictive power for output: in the sample of countries used the correlation between the growth rates of output and the credit ratio is 0.4685, significant above the 0.1% level. The item can be found under Financial Dashboard, Financial Indicators - Stocks, Private Sector Debt. The data is not available for New Zealand, so I drop it for this exercise. Additionally, there are concerns with the data for Finland. Specifically, a casual plot of output and the credit ratio during the crisis reveals a strong, positive relationship for all countries, except for Finland who experienced a severe recession while the credit ratio increased making it a severe outlier. Finland is thus dropped, but the results are unaffected by including it.

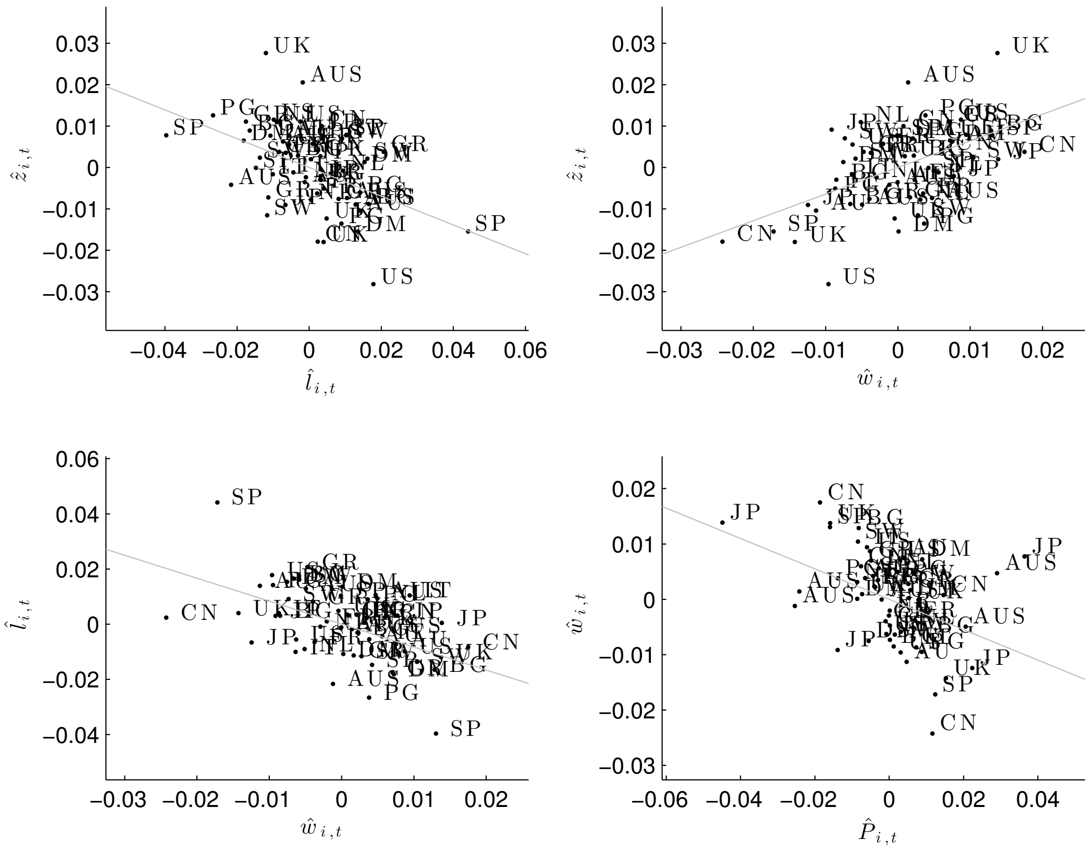
regress each of the variables of interest on the (common) control variables and calculate the residual:

$$\hat{x}_{i,t} = \tilde{x}_{i,t} - \beta_x' X_{it}, \tag{59}$$

$\hat{x}_{i,t}$ is the residual for variable $\tilde{x}$. $X_{i,t}$ is the set of controls. β_x' is the OLS estimator of $\tilde{x}$ on X. The residual is thus the component of $\tilde{x}$ not explained by the control variables. The partial correlation between variables $\tilde{x}^1$ and $\tilde{x}^2$ is then the correlation between their residuals, $\hat{x}^1$ and $\hat{x}^2$, which is the correlation between the components of $\tilde{x}^1$ and $\tilde{x}^2$ not explained by X.

In Figure 19 I repeat Figure 18, replacing the values of the variables with the residuals used in the partial correlations (each country now has four datapoints in each plot, one per year). The relationship between the variables are still apparent in this exercise, highlighting that the correlations appear to be driven by events *during* the Great Recession, and not driven by preexisting country characteristics.

Figure 19: Relationships between selected variables: Panel structure



Lines are OLS lines of best fit between the two variables. The country names refer to: AUS = Australia, AU = Austria, BG = Belgium, CN = Canada, DM = Denmark, FR = France, GR = Germany, IT = Italy, JP = Japan, NL = Netherlands, NZ = New Zealand, PG = Portugal, SP = Spain, SW = Sweden, UK = United Kingdom, US = United States.

The panel structure, by giving me more power, also allows me to more precisely measure the sizes of the partial correlations, and test their precision. Table 6 gives the estimated correlations behind the four plots. All of the correlations are significant at at least the 5% level.[43] These results are qualitatively robust to dropping outliers. For example, both Spain and the UK may appear to drive some of the results, but the significance and magnitudes are not greatly affected by their exclusion.

Table 6: Partial correlations

	corr(z,l)	corr(z,w)	corr(l,w)	corr(w,P)
	-0.4574***	0.5286***	-0.5192***	-0.4286***

, **, and * denote significance at the 10%, 5%, and 1% levels respectively.*

B Analytical model appendix

B.1 Proofs

Proof of Proposition 1. $\partial Z/\partial\lambda = 0$: The proof relies on showing that, if wages are fully rigid, a financial crisis leads to no change in the relative capital, $\tilde{k}$, held by any firm. This then trivially implies no change in Z according to (11). Since $\tilde{k} \equiv k/K$, this implies that all firms shrink by the same amount, regardless of whether they are financially constrained or not. Within this proof, I define $x' = \partial x/\partial\lambda$ as shorthand for the partial derivative of any variable x with respect to λ. I use $\dot{x} = x'/x$ to refer to the partial scaled by the value, which gives the (marginal) percentage change.

Starting with constrained firms, their capital is simply the maximum they can afford to purchase: $k = \lambda n$. Since n is exogenous, the partial derivative with respect to λ implies that $\dot{k}_{cons} = \frac{1}{\lambda}$. Where I use $\dot{k}_{cons}$ to refer to the common percentage change across all constrained firms. I now guess and verify that all unconstrained firms also optimally adjust their capital choice by the same ratio, that is $\dot{k}_{unc} = \frac{1}{\lambda}$ too. Differentiating $K \equiv \int_0^1 k_i \,\mathrm{d}i = \int k \,\mathrm{d}f(n,z)$ then yields[44]

$$\dot{K} = \int_n \int_z \dot{k}\tilde{k} f(n,z) dndz \tag{60}$$

$$= \frac{1}{\lambda} \int_n \int_z \tilde{k} f(n,z) dndz = \frac{1}{\lambda} \tag{61}$$

By the definition of $\tilde{k}$, this then implies that all firms' capital stocks remain in the same proportion, which was required to deliver $Z' = 0$:

$$\tilde{k} \equiv \frac{k}{K} \Rightarrow \tilde{k}' = \tilde{k}\left(\dot{k} - \dot{K}\right) = 0 \tag{62}$$

To complete the proof I must verify the conjecture that $\dot{k}_{unc} = \frac{1}{\lambda}$. Differentiating the expression for their optimal capital choice, (4), with fully rigid wages yields $\dot{k}_{unc} = \dot{Y}$ So for their optimal capital choice to fall by ratio $1/\lambda$ it must be that $\dot{Y} = 1/\lambda$. Using (1), the definition of Y, we see that this is true if $\dot{y} = 1/\lambda$ for all firms. That is, aggregate output changes by ratio $1/\lambda$ if all intermediates

[43] The panel structure could introduce serial correlation into the errors, which could be be a problem for the significance tests. However, given the relatively small sample size, clustering the standard errors is problematic.

[44] Throughout I use that, for $\rho < 1$, the capital choice of the marginally unconstrained firm is equal to that of the marginally constrained firm to ignore the integral-limit terms in Leibniz rule, since they cancel out.

change by that same ratio. For constrained firms, output is given by:

$$y_i = \left(\frac{\nu Y^{1-\rho}}{w}\right)^{\frac{1-\alpha}{1-\nu}} z_i^{\frac{1}{1-\nu}} k_i^{\frac{\alpha}{1-\nu}} \tag{63}$$

And for unconstrained it is given by:

$$y_i = \alpha^{\frac{\alpha}{1-\rho}} (1-\alpha)^{\frac{1-\alpha}{1-\rho}} \rho^{\frac{1}{1-\rho}} Y z_i^{\frac{1}{1-\rho}} w^{-\frac{1-\alpha}{1-\rho}} \tag{64}$$

For unconstrained firms differentiating (64) yields this immediately: $\dot{y}_{unc} = \dot{Y} = 1/\lambda$. For constrained firms, differentiating (63) yields the same result:

$$\dot{y}_{cons} = \left(1 - \frac{\alpha}{1-\nu}\right)\dot{Y} + \frac{\alpha}{1-\nu}\dot{k}_{cons} = \left(1 - \frac{\alpha}{1-\nu} + \frac{\alpha}{1-\nu}\right)\frac{1}{\lambda} = \frac{1}{\lambda} \tag{65}$$

I have thus verified my conjecture, completing the proof.

$\partial L/\partial\lambda > 0$: I proved that $\dot{Y} = 1/\lambda > 0$. The aggregated labour demand equation, (10) with a fixed wage implies that $\dot{L} = \dot{Y} = 1/\lambda > 0$. □

Proof of Proposition 2. $\partial Z/\partial w \geq 0$: Key to the proof is understanding how constrained and unconstrained firms' capital choices respond to a change in the wage. For constrained firms we simply have $k = \lambda n$, and hence no change in capital in response to a marginal wage change: $\dot{k}_{cons} = 0$. For unconstrained firms differentiating (4) yields:

$$\dot{k}_{unc} = -\left(\frac{\nu}{1-\rho}\frac{1}{w} - \dot{Y}\right) < 0 \tag{66}$$

This expression is negative because it is possible to prove that $\dot{Y} < 0$, that is that output falls in partial equilibrium as the wage is increased because firms hire and invest less. This allows us to compute the change in the total capital stock and capital shares:

$$\dot{K} = \dot{k}_{unc}\tilde{K}_{unc} < 0 \qquad \dot{\tilde{k}}_{cons} = -\dot{k}_{unc}\tilde{K}_{unc} > 0 \qquad \dot{\tilde{k}}_{unc} = \dot{k}_{unc}\tilde{K}_{cons} < 0 \tag{67}$$

Where $\tilde{K}_{unc} \equiv \int\int_{unc} kf(n,z)dndz/K$ is the fraction of the capital stock managed by unconstrained firms, and similarly for $\tilde{K}_{cons}$. $\int\int_{unc}$ refers to the double integral over the regions of (n,z) where firms are unconstrained, and similarly for constrained firms. Crucially, following an increase in the wage constrained firms relatively expand and unconstrained relatively shrink, because unconstrained firms optimally reduce their scale in response to an increase in the wage. To understand how these changes in relative size affect aggregate productivity, differentiate the expression for TFP, (11), and use the above expressions to yield:

$$\frac{Z'}{\alpha Z^{\frac{1-\nu-\rho}{1-\nu}}} = \int\int_{unc} z^{\frac{\rho}{1-\nu}} \tilde{k}^{\frac{\alpha\rho}{1-\nu}} \dot{\tilde{k}} f(n,z)dndz + \int\int_{cons} z^{\frac{\rho}{1-\nu}} \tilde{k}^{\frac{\alpha\rho}{1-\nu}} \dot{\tilde{k}} f(n,z)dndz \tag{68}$$

$$= \left(-\dot{k}_{unc}\right)\tilde{K}_{unc}\tilde{K}_{cons}\left[\frac{\mathrm{E}_{cons}\left[z^{\frac{\rho}{1-\nu}}\tilde{k}^{\frac{\alpha\rho}{1-\nu}}\right]}{\bar{K}_{cons}} - \frac{\mathrm{E}_{unc}\left[z^{\frac{\rho}{1-\nu}}\tilde{k}^{\frac{\alpha\rho}{1-\nu}}\right]}{\bar{K}_{unc}}\right] \tag{69}$$

Where $\bar{K}_{unc} \equiv \tilde{K}_{unc}/N_{unc}$ gives the average capital share among unconstrained firms (N_{unc} being the mass of unconstrained firms) and similarly for constrained firms. This expression immediately delivers that there is no productivity effect if either all or no firms are constrained, since in each case either $\tilde{K}_{unc} = 0$ or $\tilde{K}_{cons} = 0$. Since $\left(-\dot{k}_{unc}\right) > 0$, the result that $Z' > 0$ follows as long as the term in square brackets is positive.

To see that this is the case, (4) implies that for any unconstrained firm we can write $\tilde{k} = cz^{\frac{\rho}{1-\rho}}$ for some constant c. Rearranging implies that $z^{\frac{\rho}{1-\nu}}\tilde{k}^{\frac{\alpha\rho}{1-\nu}} = c^{\frac{\alpha\rho}{1-\nu}-1}\tilde{k}$, and hence

$$\frac{\mathrm{E}_{unc}\left[z^{\frac{\rho}{1-\nu}}\tilde{k}^{\frac{\alpha\rho}{1-\nu}}\right]}{\tilde{K}_{unc}} = c^{\frac{\alpha\rho}{1-\nu}-1} \tag{70}$$

For any constrained firm, the fact that capital is constrained to be too low implies that $z^{\frac{\rho}{1-\nu}}\tilde{k}^{\frac{\alpha\rho}{1-\nu}-1} \geq c^{\frac{\alpha\rho}{1-\nu}-1}$, with strict inequality for any strictly constrained firms. Letting $f_c(n, z)$ denote the density conditional on being constrained, and multiplying by $\tilde{k}f_c(n, z)$ yields $z^{\frac{\rho}{1-\nu}}\tilde{k}^{\frac{\alpha\rho}{1-\nu}}f_c(n, z) \geq c^{\frac{\alpha\rho}{1-\nu}-1}\tilde{k}f_c(n, z)$. Integrating over all constrained firms gives:

$$\frac{\mathrm{E}_{cons}\left[z^{\frac{\rho}{1-\nu}}\tilde{k}^{\frac{\alpha\rho}{1-\nu}}\right]}{\tilde{K}_{cons}} \geq c^{\frac{\alpha\rho}{1-\nu}-1} = \frac{\mathrm{E}_{unc}\left[z^{\frac{\rho}{1-\nu}}\tilde{k}^{\frac{\alpha\rho}{1-\nu}}\right]}{\tilde{K}_{unc}} \tag{71}$$

Since this holds strictly for any strictly constrained firms, we have that the term in square brackets must be strictly positive, and hence $Z' \geq 0$.

$\partial L/\partial w < 0$: This follows because $\partial l/\partial w < 0$ for both constrained and unconstrained firms, according to (7). □

B.2 Analytical model robustness

B.2.1 Generalising the borrowing constraint

Much of the sharpness of the results above derives from the assumption of a simple collateral constraint, $k \leq \lambda n$, and it is worth investigating how generalising this constraint affects the results. For example, we might want to allow more productive firms to borrow more, or net worth to enter in a non-linear fashion.

The result that TFP is unaffected by a financial shock is robust to generalising the constraint to the form $k \leq \lambda g(n, z)$ for any function g, where a financial shock is still defined as a marginal change in λ. The proof relied on all constrained firms adjusting their capital by the same proportion in response to a financial shock, which this constraint preserves.

Thus we can break the exactness of the result with a more general constraint, $k \leq g(n, z; \lambda)$. If $g_\lambda(n, z; \lambda)$ is not equal across firms, a change in λ will disrupt the (relative) distribution of capital and hence affect TFP. However, if we consider the natural case where the least productive firms are affected more by a financial shock ($g_\lambda(n, z_0; \lambda) > g_\lambda(n, z_1; \lambda)$ for $z_0 < z_1$) a financial shock should actually redistribute resources to more productive firms and increase productivity. Hence the result that a financial shock should not, on its own, reduce productivity is preserved.

The result that reducing the wage reduces TFP is robust to the introduction of the more general $k \leq g(n, z; \lambda)$ constraint, since it is only that constrained firms cannot adjust their capital in response to a wage change which matters, not the specifics of the constraint. However, the result would be dampened by a constraint that linked borrowing ability to expected profits, since reducing the wage would increase expected profits and raise the borrowing ability of constrained firms. This would reduce the size of the relative redistribution of resources towards unconstrained firms, protecting TFP. However, this feature is implicitly present in my numerical model, since I model the evolution of net worth over time. Lowering the wage increases the profits and net worth of financially constrained firms, allowing them to borrow more, and the same results go through quantitatively.

B.2.2 Entry and exit

How do entry and exit affect my main results? It turns out that they enhance them, since similar channels operate on the extensive margin. In particular, in the presence of entry and exit a financial crisis will still not reduce aggregate TFP, and will actually increase it. Additionally, wage declines will endogenously lower TFP.

Analytical results are harder to obtain when adding entry and exit, so I turn to a special case of the more geneal model above. Consider the version where firms' goods are perfect substitutes ($\rho = 1$). I first outline the solution without entry or exit. In this case the solution to the firm's problem simplifies, because the problem becomes linear and solutions become "bang bang". In this case there is a threshold for productivity above which firms will ideally want to purchase infinite capital, and hence all such firms are at their borrowing constraints. Below this threshold firms choose to invest nothing:

$$k_i = \begin{cases} \lambda n_i & : aw^{1-\frac{1}{\alpha}} z_i^{\frac{1}{\alpha}} \geq 1 \\ 0 & : aw^{1-\frac{1}{\alpha}} z_i^{\frac{1}{\alpha}} < 1 \end{cases} \tag{72}$$

Where $a \equiv \left((1-\alpha)^{\frac{1}{\alpha}-1} - (1-\alpha)^{\frac{1}{\alpha}}\right)$. Suppose that productivity and initial wealth are uncorrelated. Without entry and exit, we can show that in this case productivity can be expressed as $Z = \mathrm{E}\left[z^{\frac{1}{\alpha}} | aw^{1-\frac{1}{\alpha}} z^{\frac{1}{\alpha}} \geq 1\right]^{\alpha}$. Firms with productivity above a cut-off threshold invest in as much capital as they can afford, and those below invest nothing and produce nothing. TFP is thus the average productivity amongst producing firms. If w is fixed then a crisis (tightening of λ) again has no effect on TFP, and a fall in w reduces TFP by reducing the productivity threshold for investing, reducing the average productivity of producing firms. Thus the results of the previous section are preserved in the case of $rho = 1$.

I first extend the model by adding entry, but abstracting from exit. I suppose that at the beginning of the period there are no incumbent firms, and a unit mass of potential entrants. All potential entrants are endowed with identical net worth n_e, and know their productivity were they to enter. Entrants must pay a fixed entry cost c.

A firm will enter if the profit from doing so exceeds the entry cost. If a firm borrows up to the borrowing constraint they will make profit $\pi_i = \left(az^{\frac{1}{\alpha}} w^{-\frac{1-\alpha}{\alpha}} - 1\right) \lambda n_e$. This implies that firms will only enter if $az^{\frac{1}{\alpha}} w^{-\frac{1-\alpha}{\alpha}} \geq 1 + c/\lambda n_e$. This immediately tells us that (i) tightening the borrowing constraint increases the productivity threshold because it reduces profits, and (ii) lowering the wage reduces the productivity threshold by increasing profits.

This is reflected in the expression for endogenous TFP: $Z = \mathrm{E}\left[z^{\frac{1}{\alpha}} | aw^{1-\frac{1}{\alpha}} z^{\frac{1}{\alpha}} \geq 1 + \frac{c}{\lambda n_e}\right]^{\alpha}$. Now, unlike the model without entry, a financial crisis without a change in wages will actually *increase* TFP, because only more productive firms will enter. As in the baseline model, reducing the wage will reduce TFP, with this effect now also enhanced because reducing the wage reduces the minimum productivity of firms who choose to enter.

The above example can be easily reinterpreted to show that the effects of firm exit are identical.

Suppose instead that we start with a unit mass of existing firms, and remove entry. Firms must pay a fixed cost c to produce, else they must exit, and all firms have common net worth n. In this case, the expression for TFP is identical. A financial crisis increases TFP because it encourages the least productive firms to exit, and reducing the wage reduces TFP because it increases profits and allows less productive firms to survive.[45]

Thus, overall we see that the results are robust to, and indeed complemented by, introducing firm entry and exit. A financial crisis in this model by itself would increase TFP by increasing the minimum productivity required to enter, and causing the least productive firms to exit. Reductions in the wage increase profits, undoing these effects.

C Numerical model appendix

C.1 Details of solution to firm problem

The solution contains many regions that must be computed separately: does the firm hit the borrowing constraint? Does she invest or disinvest or just let capital depreciate? In this section I provide an algorithm to solve the optimisation.

The core of the algorithm is the calculation of five policy functions (conditional on the next period value function, $V_{t+1}(n_{t+1}, k_t, z_{t+1})$). The first two require maximisation: 1) $k^{i,u}(n, k, z)$ is the investment policy computed ignoring both the financial and irreversibility constraints. I.e. the solution to (36) ignoring the constraints on k_t. 2) Similarly, $k^{d,u}(n, k, z)$ is the disinvestment policy computed ignoring both the financial and irreversibility constraints. This is the solution to (37) ignoring the constraints on k_t.[46]

The remaining policies can be computed without maximisation: 3) $k^{i,c}(n, k, z) = n/(1 - \lambda_t q^d)$ is the investment policy when constrained 4) $k^{d,c}(n, k, z) = (n - (1 - q^d)(1 - \delta)k)/(q^d - \lambda_t q^d)$ is the disinvestment policy when constrained. 5) $k^{in}(n, k, z) = (1 - \delta)k$ is the inaction policy. Having computed the five policy functions I apply the following program to check which is optimal.[47]

1) $k^{i,u}(n, k, z)$ valid $\Rightarrow (1 - \delta)k < k^{i,u}(n, k, z) < k^{i,c}(n, k, z)$. Neither $k^{i,c}$ nor k^{in} can be optimal since they were both feasible choices in the $k^{i,c}$ maximisation and are hence dominated. $k^{d,u}$ can't be feasible since $k^{d,u} > k^{i,u} > (1 - \delta)$. $k^{d,c}$ also can't be feasible because $(1 - \delta)k < k^{i,u} < k^{i,c} \Rightarrow (q_t^i - \lambda_t q_{t+1}^d)(1 - \delta)k < n \Rightarrow (n - (q_t^i - q_t^d)(1 - \delta)k)/(q_{t+1}^d - \lambda_t q_{t+1}^d) > (1 - \delta)k \Rightarrow k^{d,c} > (1 - \delta)k$.

2) $k^{d,u}(n, k, z)$ valid $\Rightarrow k^{d,u} < (1 - \delta)k$ and $k^{d,u} < k^{d,c}$. Again neither $k^{d,c}$ or k^{in} can be optimal since feasible and dominated. $k^{i,u} < k^{d,u} < (1 - \delta)k$ and hence $k^{i,u}$ isn't feasible. For $k^{i,c}$ consider two cases. If $k^{i,c} \leq k^{i,u}$ then $k^{i,c} \leq k^{i,u} < k^{d,u} < (1 - \delta)k$ and not feasible. If $k^{i,c} > k^{i,u}$ then $k^{i,c}$ isn't optimal even if it is possible (because it is dominated by $k^{i,u}$).

[45] The sharpness of this result relies on the assumption that z and n are uncorrelated, since n is identical across firms. Note that firms with more net worth can afford more capital and make more profit, and hence are more likely to survive. If productivity and net worth are negatively correlated, then this effect could overturn the above results. For example, tightening the borrowing constraint would force low net worth firms to exit, and if these firms are high productivity then TFP could fall. However, negatively correlated productivity and net worth seems unlikely, and hence the result is likely to survive in a realistic calibration.

[46] The solution can be further simplified because the dependence of these functions on the state k_{t-1} is actually very limited. Note that $k^{i,u}(n, k, z)$ does not actually depend on k_{t-1} at all. $k^{d,u}(n, k, z)$ does, but actually n_t and k_{t-1} only affect it via their effect on the pseudo-state $\tilde{n}_t$. Thus the maximisation can actually be done to solve for a function $k^{d,u}(\tilde{n}, z)$, which is then interpolated over (n, k, z) to give $k^{d,u}(n, k, z)$

[47] This procedure relies on assuming that $k^{d,u} > k^{i,u}$, which is intuitive since $q^d < 1$ increasing the optimal capital purchase. This is easy to prove analytically in the case without financial frictions, with financial frictions I check that the condition holds numerically. A single-peakedness assumption must also be verified.

3) $k^{i,c}(n,k,z)$ valid $\Rightarrow (1-\delta)k < k^{i,c}(n,k,z) < k^{i,u}(n,k,z)$. Now $k^{i,u}$ is not feasible because it violates borrowing constraint. $k^{d,u} > k^{i,u} > (1-\delta)k$ and hence not valid. A single peaked argument says that k^{in} can't be optimal. $k^{in} < k^{i,c} < k^{i,u}$ means that we are to the left of $k^{i,u}$ so $v(k^{in}) < v(k^{i,c}) < v(k^{i,u})$. As in case (1), $k^{d,c}$ also can't be valid because $(1-\delta)k < k^{i,u} < k^{i,c} \Rightarrow (q_t^i - \lambda_t q_{t+1}^d)(1-\delta)k < n \Rightarrow (n - (q_t^i - q_t^d)(1-\delta)k)/(q_{t+1}^d - \lambda_t q_{t+1}^d) > (1-\delta)k \Rightarrow k^{d,c} > (1-\delta)k$

4) $k^{d,c}(n,k,z)$ valid $\Rightarrow k^{d,c} < (1-\delta)k$ and $k^{d,u} > k^{d,c}$. Now $k^{d,u}$ is not feasible because it violates borrowing constraint. $k^{i,c}$ is not valid because $k^{d,c} = (n - (q_t^i - q_t^d)(1-\delta)k)/(q_{t+1}^d - \lambda_t q_{t+1}^d) < (1-\delta)k \Rightarrow n/(q_t^i - \lambda_t q_{t+1}^d) < (1-\delta)k \Rightarrow k^{i,c} < (1-\delta)k$ and hence violates irreversibility. To see that $k^{i,u}$ is not feasible consider two cases. First if $k^{i,u} \leq k^{i,c}$ then the previous result gives that $k^{i,u} < (1-\delta)k$ and it also violates irreversibility. If $k^{i,u} > k^{i,c}$ then it violates the borrowing constraint and is not feasible anyway. Finally, k^{in} violates the borrowing constraint. To see this, we already showed that $n/(q_t^i - \lambda_t q_{t+1}^d) < (1-\delta)k$. But to afford inaction we require net worth at least $n/(q_t^i - \lambda_t q_{t+1}^d) \geq (1-\delta)k$ which is a contradiction.

5) $k^{in}(n,k,z)$: We have shown that whenever any other policy function is possible, then k^{in} is always dominated. Then the only time k^{in} can be chosen must be when all other options are infeasible.

C.2 Point where firms become unconstrained

At some point firms become so rich they won't ever be constrained again. How do we solve for this point? Once this happens, we know that these firms will play the unconstrained capital policy $k_t = k_t^u(k_{t-1}, z_t)$. We want to solve for a minimum level of worth such that the firm can afford to play this policy forever from now on without violating borrowing or positive dividend constraints. This minimum is state contingent and denoted $\underline{n}_t(k_{t-1}, z_t)$. This is the minimum net worth required at the beginning of t in order to play $k_t^u(k_{t-1}, z_t)$ from now on. It is time dependent along an aggregate transition path, and can be computed as a fixed point, $\underline{n}(k_{t-1}, z_t)$, in steady state.

We can characterise this level recursively. If you choose the unconstrained capital policy function and borrow d_t, next period's net worth is:

$$n_{t+1} = \pi(z_{t+1}, k_t^u(k_{t-1}, z_t)) + q^i(1-\delta)k^u(k_{t-1}, z_t) - rd_t \tag{73}$$

You cannot choose such high d_t that this leaves $n_{t+1} < \underline{n}_{t+1}(k_t^u(k_{t-1}, z_t), z_{t+1})$ for any z_{t+1}. This gives a maximum level of borrowing $d_t = \bar{d}_t(k_{t-1}, z_t)$, which depends on $\underline{n}_t$+1. The borrowing constraint places another constraint on how much you can borrow: $d_t \leq \lambda_t q^d k_t^u(k_{t-1}, z_t)$. Thus the maximum you can borrow without violating today's borrowing constraint or leaving too little net worth tomorrow is $\min\left\{\bar{d}_t(k_{t-1}, z_t), \lambda q^d k_t^u(k_{t-1}, z_t)\right\}$. Using the balance sheet, this gives us the required minimum net worth to fund the purchase of $k_t^u(k_{t-1}, z_t)$ given the maximum amount of borrowing. If investing:

$$\underline{n}_t(k_{t-1}, z_t) = k_t^u(k_{t-1}, z_t) - \min\left\{\bar{d}_t(k_{t-1}, z_t), \lambda q^d k_t^u(k_{t-1}, z_t)\right\} \tag{74}$$

If disinvesting:

$$\underline{n}_t(k_{t-1}, z_t) = q^d k_t^u(k_{t-1}, z_t) + (1-q^d)(1-\delta)k_{t-1} - \min\left\{\bar{d}_t(k_{t-1}, z_t), \lambda q^d k_t^u(k_{t-1}, z_t)\right\} \tag{75}$$

This establishes a recursive procedure for solving for $\underline{n}_t(k_{t-1}, z_t)$.

C.3 Densities, aggregation and transition

This section provides clarification of the correct densities to use for the computation of various aggregates. Since firms exit each period, care must be taken to ensure the correct densities are used. At the beginning of the period there is a mass of existing firms, $\mu_t(n,k,z)$ who all produce. Of these, a fraction $1-\sigma$ exit, and a mass of $1-\sigma$ firms enters to replace them. The survivors and entrants then invest at t, and all produce at $t+1$.

The density at the end of t, after exit and entry, is denoted $\mu_t'(n,k,z) \equiv (1-\sigma)\mu_t(n,k,z) + \mu^e(n,k,z)$, where $\mu^e(n,k,z) \equiv 1(n,k)\Gamma_{e,z}$ and $1(n,k)$ is an indicator function only equal to one for $(n,k) = (n_e, 0)$. The density of producers next period is then found by applying firms' policy functions to μ_t', to calculate the implied choices and realisations of (n', k', z'). This is summarised by the functional equation $\mu_{t+1} = \Gamma_t(\mu_t)$.

D Investment regressions

Note that these results support my finding that there is no difference between small and large firms in the US. However, the time trend is badly estimated due to the Dot-Com bubble, and the regression thus reports investment as being above trend for all size groups.

Full results of Crawford et al.'s (2013) regression for UK data:

Size:	All	<50	50-249	>250
Post08	-0.099***	-0.107***	-0.157***	-0.042*
Post06	0.131***	0.115***	0.139***	0.135***
Year	-0.046***	-0.044***	-0.048***	-0.045***
No. obs	208,169	77,145	79,752	51,272
No. RUs	75,772	40,253	24,349	11,170

Replication of Crawford et al.'s (2013) specification on US Compustat sample

Size:	All	<50	50-249	>250
Post08	0.110***	0.233**	0.111	0.098***
Post06	0.382***	0.709***	0.318***	0.361***
Year	-0.006	-0.051***	0.003	-0.003
No. obs	38,285	3,525	7,320	27,440
No. RUs	8,425	998	1,932	5,495

E Non-stochastic simulation with endogenous nodes

In this section I describe my algorithm for performing non-stochastic simulations on an endogenously created grid. To simulate my model using exact population moments, I must track the density $\mu_t(n, k, z)$ over time. Given that the states n and k are continuous, a computer cannot store the exact value of this function, and can only store an approximation using discrete points.

The standard non-stochastic simulation algorithm (Young, 2009) does this by discretising all continuous variables, and defining an equispaced grid over all variables. This can become expensive with many states: with two continuous states each requiring a few thousand nodes for accuracy, millions of nodes may be required.

My algorithm uses the following insight. If the shocks hitting firms (here the z values) are discrete, there are conditions under which the resulting distributions over (n, k, z) can be described as a finite number of mass points.[48] Thus, while the potential values the continuous states (n, k) can take are infinite, the values actually observed will be finite. An equispaced grid will necessarily create many extra nodes which have zero mass in the true distribution. My algorithm uses firms' policy functions to endogenously construct a grid only over values of (n, k, z) with positive mass in the distribution. This can be applied to both the ergodic distribution, and to distributions along stochastic simulations or deterministic transitions.

When will the number of mass points in, for example, the ergodic distribution be finite? We can understand this using a simple example. Suppose we start with a unit mass of firms, and ignore entry and exit. Suppose that the productivity shock is i.i.d. and can only take two values, z_l and z_h with equal probability. Suppose all firms start at time 0 with the bad productivity shock and the same values of the endogenous states, which I label (n_0, k_0, z_l). The distribution at time 0 is thus described by a single mass point at (n_0, k_0, z_l).

At time 0 all firms make the same choice of k_1, since they have the same state. However, half of the firms will have the bad productivity shock tomorrow, leading to state $(n_{1,l}, k_1, z_l)$, and the other half will get the good shock, leading to state $(n_{1,h}, k_1, z_h)$. The number of mass points thus doubles between periods 0 and 1. Similarly between periods 1 and 2 the number of nodes will again double to four. Thus, without any other forces, the ergodic distribution, calculated by simulating forward infinitely, must have a countably infinite number of nodes since the number doubles every period.

However, many models actually have forces which stop this growth, leading to a finite number of nodes. For example, in my model firms follow a minimum savings rule, meaning that they pay out net worth as dividends if it exceeds a threshold $\bar{n}_t(k, z)$. Thus, all firms whose net worth exceeds this limit end up clustering on the same level of net worth, conditional on their (k, z).

[48] Even if these conditions fail and the number of mass points becomes countably infinite, this algorithm can still be applied. It will then just be an approximation, rather than an exact description of the distribution.

Similarly, once my firms become unconstrained forever they follow the policy function of a firm facing only partial irreversibility. Unless inaction is chosen, all of these firms choose the same level of k' if they invest or disinvest. This again places limits on the number of nodes. If these forces are strong enough, the number of nodes in the ergodic distribution will be finite. For example, in my model with two values of productivity, there are only roughly 10,000 nodes in the ergodic distribution, around 5,000 per value of the productivity shock. Additionally, as can be seen from the plots of the density, much of the mass is actually clustered in a few key nodes.

The following algorithm constructs the nodes featured in the ergodic distribution. A similar version can be used along simulations. The version given simply constructs the nodes, and does not calculate the densities. The densities can then be calculated afterwards using standard iterations of the policy functions over the constructed grid. However, it is trivial, and more efficient, to calculate the density simultaneously with the grids. Finally, if the number of nodes in the distribution is too large (either too large but finite, or countably infinite) it is simple to add a projection step which reduces the number of nodes if they become to large.

In the following, I assume (for simplicity of exposition) that idiosyncratic productivity can take two values, z_l and z_h, indexed by $i_z = \{1, 2\}$.

1. Start with an initial number N_0 of nodes. For each node, $i \in N_0$, assign a state (n_i, k_i, z_i). E.g. start with just one node per productivity state with no capital and net worth n_e, to mimic starting only with entrant firms.

2. Follow the following iterative procedure. At each step, m, start with N_m nodes with states (n_i, k_i, z_i) assigned to each node.

    (a) For every i compute the capital choice $kp_i = k(n_i, k_i, z_i)$. Each node will become two nodes next period, one for each potential productivity realisation. Using the policy functions, compute the net worth at each potential next-period productivity: $np_{i,i_z} = np(n_i, k_i, z_i, i_z)$.

    (b) Create a new collection of $N' = N_m \times 2$ nodes, indexed by i and i_z. Index each node by $i' \in N'$, and use $i_z = g(i')$ and $i = h(i')$ to track which i and i_z is associated with each i'.

    (c) Assign states to each i' node using $n_{i'} = np_{h(i'),g(i')}$, $k_{i'} = kp_{h(i')}$ and $z_{i'} = zgrid(g(i'))$.

    (d) If any nodes are duplicates, delete duplicates to create a unique list of $N_{m+1} \subset N'$ nodes, each with associated (n_i, k_i, z_i).

3. Repeat until the number of nodes converges.

Note since this procedure builds the new set of nodes from the choices made each iteration, it also automatically throws out any nodes which are not used in the ergodic distribution.