Firm Dynamics at the Zero Lower Bound

(VERY) PRELIMINARY AND (VERY) INCOMPLETE.

Alex Clymo*

May 17, 2018

Abstract

Can the large decline in aggregate employment during the US's Great Recession be explained by the unusually large decline in employment of young firms? In this paper I develop a quantitative heterogeneous-firm model with both financial and nominal frictions and use it to analyse the recent recession through the lens of firm-level data. I first use BDS data for the US to show that 40% of the decline in aggregate employment during the recession can be accounted for by declining employment in young firms (age 0-5). While this supports the role of financial frictions in impeding employment, it still leaves 60% of the decline to be explained by old firms (age 6+). Decomposing this further into changes in average employment within firms, and the number of firms, I find that most of the total decline in employment can be accounted for by (1) a decline in the *number* of young firms by over 25%, and (2) a decline in employment *within* old firms of 7.5%. I then extend the heterogeneous firm model of Khan and Thomas (2013) to include nominally denominated firm debt, downwards nominal wage rigidity (DNWR), and the zero lower bound (ZLB). Absent nominal features, a financial shock is only able to match the response of young firms during the crisis, because declines in real factor prices reallocate resources from young to old firms. However, adding DNWR and the ZLB inhibits this fall in wages and interest rates, leading the financial crisis to spill over to old firms. This allows the model to better match the data across the whole firm age distribution. Additionally, rather than leading to the misallocation of resources across firms, the crisis instead manifests as a drastic decline in employment, as occurred in the US. Finally, I show how the power of fiscal policy at the ZLB depends on the firm distribution. The government spending multiplier is larger when more firms are financially constrained, and fiscal transfers to young firms are more powerful than transfers to older firms.

*Department of Economics, University of Essex. Email: a.clymo@essex.ac.uk. I thank Wouter den Haan for invaluable support. I also thank Francesco Caselli, Sergio de Ferra, Kyle Herkenhoff, Andrea Lanteri, Marcelo Pedroni, Morten Ravn, Kevin Sheedy, Christian Stoltenberg, Silvana Tenreyro, and John Van Reenen for helpful advice and comments, as well as seminar participants at Uppsala University, the University of Essex, Oxford University, and the University of Surrey. I thank the ACRM for financial support. All errors are my own.

1 Introduction

The financial crisis has spurred much work, both empirical and theoretical, on how shocks affect individual firms. This is very natural given the nature of the recession: For example, a shock-reduction in the availability of credit would hurt cash-poor firms more than it would hurt cash-rich firms. Identifying this channel empirically and modelling it theoretically necessitates the use of both firm-level data, and models which take firm heterogeneity seriously.

In this paper I first provide novel empirical evidence on the behaviour of firms during the crisis, and then develop a quantitative heterogeneous-firm model through which to interpret this data. In particular, I argue that existing models of the financial crisis are not able to capture the responses of firms across the whole firm-age distribution. I then show that adding nominal frictions to the model allows the model to do so, by bringing the model's implications for factor price adjustments closer to the data.

It is well known that the crisis was unusual in the US for how severe the effects on the labour market were. Unemployment rose to 10%, and measures of employment and hours worked fell by as much as 7% to 10% respectively. The crisis was also unusual in that total employment in young firms was particularly hard-hit, as one might expect given the financial nature of the recession (see, e.g., Dyrda, 2016). Given the importance of young firms for overall job creation (startup firms account for essentially all net job creation) one might expect that the large *aggregate* employment response can be accounted for by the problems at these younger firms. My first contribution is to bring new evidence on how the financial crisis affected employment at different firms across the firm-age distribution, and use this to provide a formal decomposition of the aggregate employment decline.

To this end, I use Business Dynamics Statistics (BDS) data, provided by the US Census Bureau. BDS data provides repeated cross sections of the universe of US firms, aggregated to narrow firm size and/or age bins.[1] I first show that, despite providing only 14% of total employment pre-crisis, 40% of the *decline* in aggregate employment during the recession can be accounted for by the *decline* in employment in young firms (age 0-5).[2] Indeed, total employment in young firms shrank by 25% from 2007 to 2010, while total employment in old firms shrank by a much more modest 5%. While this suggests the important role that young firms, and hence the shocks that particularly affect them, played during the crisis, it still leaves 60% of the aggregate employment decline to be explained by old firms (age 6+). Thus, contrary to the intuition that the severity of the aggregate employment response during the recession is due mostly to the issues with younger firms, understanding the declines in older firms is equally important for capturing aggregate employment dynamics.

This becomes even clearer after I further decompose these changes into changes in average employment within firms of a certain age, and changes in the number of firms of each age (driven by changes in entry and exit). The larger decline in total employment in young firms turns out to be driven almost entirely by an unprecedented drop in the number of young firms relative to old firms, and not by a larger decrease in average employment within young firms. In particular, I find that most of the aggregate decline in employment can be accounted for by (1) a decline in the *number* of young firms by 25%, and (2) a decline in employment *within* old firms of 7.5%. While average

[1] The BDS data is publicly available, and constructed by aggregating the confidential Longitudinal Business Database (LBD) data into narrow bins. Data is provided split by establishment or firm age. Since I focus on financial frictions, I use data binned by firm age.

[2] These results are robust to changing the definition of young versus old firms. Similar numbers have been reported by other authors, as I discuss in the literature review.

employment in young firms does decline more than in old firms (10% vs 7.5%) this difference is not large enough to be important for aggregate employment dynamics, compared to the fall in the number of young firms.

The fact that older firms shrank so drastically during the crisis presents an interesting challenge to the idea that the crisis was primarily a financial shock which impacted the economy by misallocating resources across firms. Indeed, if this was the case we would expect to see resources being reallocated from poor to rich firms. If we are willing to equate young with poor and old with rich, we do not see that at all: instead, firms of all ages shrink, and the recession is not one of mis- or even re-allocation, but just one of general decline. A key message of this paper is that reallocation of resources across firms only happens when incentivised by changes in factor prices. As I will discuss below, nominal frictions will enable my model to generate realistic factor price movements, which will be crucial in matching the firm-level reallocation patterns.

Having established a set of facts about the responses of employment across the firm-age distribution during the crisis, I then turn to building a quantitative heterogeneous firm model to explain them. I build on the seminal heterogeneous-firm model with financial shocks of Khan and Thomas (2013) and extend it in two directions.

Firstly, I extend the model's calibration in order for the model to replicate the joint distribution of employment and the number of firms by firm age and firm size pre-crisis. This is essential in order to be able to compare the model's responses by firm age to those from the BDS data described above. To do this, I introduce several realistic model features. Firms' productivities vary due to three distinct shocks: 1) I use permanent productivity differences across firms to match the very wide empirical distribution *across* firms. 2) Temporary productivity shocks are used to match the degree of *within-firm* volatility. 3) Firms' productivities grow over time, to replicate the evidence that firms *accumulate demand* as their customer base grows (Foster et al., 2015).[3] Along with financial frictions (non-contingent debt and collateral constraints) and the accumulation of net worth, capital adjustment costs, and firm exit, this creates a rich cross section of firms which is able to broadly match the key features of the empirical joint firm size/age distribution.

Secondly, I introduce two nominal frictions: downwards nominal wage rigidity (DNWR) and long-term nominally-denominated firm debt.[4] The central bank controls the nominal interest rate in order to meet its goals, and I explicitly model (in a fully non-linear setting) the zero lower bound (ZLB). While these frictions are not operative in the steady state (where inflation is constant) and hence do not affect the *steady-state* firm distribution, they will be important in altering how shocks are *propagated* through the firm distribution over time, especially when the ZLB binds.

My main experiments involve subjecting my model economy to a financial crisis, and investigating how the addition of nominal rigidities alters the propagation of the shock. Broadly speaking, this occurs through altering three channels: 1) *a real wage channel:* nominal wage rigidities hamper the ability of the real wage to fall during a crisis, 2) *a real interest rate channel:* the ZLB hampers the ability of the real interest rate to fall during a crisis, and 3) *a real debt channel:* deflation increases the real value of firms' (nominal) debts.

To illustrate these channels it is useful to first explain how a financial crisis impacts the economy

[3] I assume that firms have constant returns to scale production functions, but face downwards-sloping demand curves coming from a CES demand system. Under these assumptions, firms' physical productivities and demand shocks are interchangeable.

[4] Combined with the incomplete markets setting, nominally-denominated debt redistributes wealth across agents following unexpected changes in inflation. See, e.g., Gomes et al. (2016) for a recent application.

in a "real" model without any nominal features. These results thus mirror Khan and Thomas (2013), and correspond to my model if the central bank holds inflation constant, and neither DNWR nor the ZLB bind. I simulate a financial crisis by unexpectedly and permanently tightening firms' borrowing constraints, leading to a permanent deleveraging by firms, and compute the economy's transition towards the new steady state.

The tightened borrowing constraint affects financially constrained firms, who are predominantly young and middle-aged firms,[5] who are forced to shrink their investment, and hence desired employment. In response to this reduced demand for labour and investment, both the real wage and real interest rate must fall in order to restore market clearing to the labour and goods markets. Consider now an older firm, which has outgrown its financial frictions. This firm suffers no adverse effects from the financial shock itself, and hence in response to the decline in these two factor prices, the optimal strategy of the firm is to expand. This is in fact the source of misallocation following a financial shock in this class of models: financially unconstrained firms soaking up resources shed by financially constrained firms. However, this also highlights the fundamental difficulty these "real" models have in explaining the financial crisis data I presented earlier: market clearing wants old, financially unconstrained firms to expand when young, financially constrained firms shrink, whereas in the data we see that old firms actually shrunk, and by not much less than young firms.

The above makes clear that the behaviour of factor prices is very important for the response of firms, especially old, financially unconstrained firms, to a financial shock. However, do factor prices really behave in the data as they are predicted to in the real model? In my model with nominal rigidities, both factor prices will be unable to decline as much as they do in the real model: DNWR hampers the ability of the real wage to fall (the real wage channel), and the ZLB on nominal interest rates hampers the ability of the real interest rate to fall (the real interest rate channel). I present evidence that both real wages and interest rates were indeed elevated during the crisis.

What is the effect of a financial shock in the presence of binding nominal frictions? With the real wage and real interest rate unable to fall (as much, if at all) following the shock, old, financially unconstrained firms now face less incentive to expand and absorb the resources (labour and capital) shed by young firms. If the constraints are sufficiently binding, old firms will instead contract, following the young firms due to a *demand spillover*. This brings the nominal model in line with the firm-level data, which indeed showed a contraction in average employment in old firms.

I demonstrate these results across a number of experiments, in order to differentiate between the role of wages and interest rates. The main result is that nominal frictions are essential for the model to correctly match the response of firms across the whole distribution. One feature of the US recession was a decline in inflation, which helps DNWR and the ZLB to bind in my model. Although the main intuitions do not depend on it, the quantitative results suggest that the model performs better at replicating the financial crisis if the economy is hit by a simultaneous demand shock.

While I have stressed the role that nominal frictions play in helping the model match firm-level data better, they also fundamentally change the way that a financial shock affects the aggregate economy. Rather than leading to reallocation, as young firms shrink and old firms expand, now all firms shrink as the lack of demand coming from young firms spills over to the old. Financial shocks now no longer cause misallocation, but instead cause unemployment, as we saw in the recent recession. Thus, I show a pleasing link between aggregate price data and firm-level data: bringing

[5] As in Ottonello and Winberry (2017), the presence of demand growth at firms means that firms do not grow out of their financial constraints very quickly, because their profitability (and desired size) increases with age.

the model's implications for factor prices closer to the data improves it's ability to match firm-level data, and hence its implications for reallocation and aggregated responses. Additionally, long-term nominal debt amplifies these effects (the real debt channel), as deflation further tightens firms' borrowing constraints.

Having shown the importance of nominal frictions in explaining the behaviour of different firms during the recent financial crisis, I turn to analysing how fiscal policy can work at alleviating demand problems during a deleveraging episode in this framework. I perform two simple experiments to illustrate how the power of policy at the ZLB depends on the firm distribution. Firstly, I revisit the simple government spending multiplier. In the presence of financially constrained firms, the multiplier to a transitory increase in government spending at the ZLB can be larger than one due to an *investment multiplier*. The increase in demand coming from government spending increases firms' profits, which helps them overcome financial frictions and increase investment. The size of this multiplier depends on the mass of financially constrained firms, and is larger following negative shocks which increase this mass. Secondly, I look at a more nuanced policy targeted at the source of the shock: fiscal transfers to firms to help them overcome financial frictions. I find that transfers to younger firms are more effective than transfers to older firms, since these firms are more financially constrained and hence use the transfer to increase investment. However, interestingly there is a "hump shape" in the effectiveness of the transfers, with the most powerful effects coming from transfers to firms which are neither too old nor too young.

Related Literature. The paper closest to this one is Ottonello and Winberry (2017). They also build a heterogeneous firm model with both financial and nominal frictions, and both models could be described as "Heterogeneous-Firm New Keynesian" models. Differently to their paper, I focus on downwards nominal wage rigidity as the main source of nominal frictions, whereas they add a retail sector subject to Calvo pricing frictions. Their focus is on explaining firms' responses to monetary policy shocks, whereas I focus on financial crises. Gomes et al. (2016) study the interaction between nominal firm debt and financial frictions in a setting where the firm sector aggregates, which I extend to a full heterogeneous-firm setting.

Several papers have also started using Business Dynamics Statistics (BDS) data to investigate the financial crisis, and to calibrate heterogeneous firm models. Dyrda (2016) performs similar decompositions to mine during the financial crisis using the BDS data. However, he does not decompose the fall in employment within age groups into average employment within a firm versus the number of firms, which turns out to be important for interpreting the data. He uses the BDS data as targets for the steady state distribution of firms in his model, as do Jo and Senga (2017).

This paper is related to the recent literature investigating financial shocks through the lens of heterogeneous-firm models. Most closely, I build on the model of Khan and Thomas (2013), who construct 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. I show that changes to the behaviour of factor prices in their model (in my case changed by adding nominal frictions) have very important effects on how financial shocks propagate. Moving from a real model to one with nominal frictions can completely reverse the implications of a financial shock, shutting down its effect on misallocation and instead leading it to manifest as unemployment. 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. This misallocation happens through factor prices, and I show that again these channels are sensitive to how prices behave during a crisis. Other notable papers in this literature include Arellano, Bai & Kehoe (2012), Petrosky-Nadeau (2013), Buera, Fattal-Jaef & Shin (2014), Sepahsalari (2016), and Gilchrist et al. (2015),

Other papers have investigated how different price processes affect allocations over the business cycle in heterogeneous firm models. In a seminar series of papers, Thomas (2002), and Khan and Thomas (2003, 2008) demonstrate that the lumpiness of investment at the micro-level does not translate into lumpy aggregate dynamics once the interest rate is allowed to adjust in general equilibrium. Later, Winberry (2016) showed that introducing consumption habits into the Khan and Thomas (2008) model to bring the model's predictions about the cyclicality of the interest rate closer to the data, restores some of these effects. Lanteri (2017) shows how introducing a second hand market for capital, which leads to a procyclical price for used capital as opposed to the constant resale price in the standard model, reverses the predictions for reallocation over the cycle of the standard heterogeneous-firm model.

This paper draws on the vast New Keynesian literature, and applies its insights to heterogeneous firm models. In particular, I focus on the ZLB, which has long been understood as a constraint on central banks, as formalised in Eggertsson and Woodford (2003). I use downwards nominal wage rigidity as a source of nominal frictions. For a recent example of a paper using DNWR, see Schmitt-Grohé and Uribe (2015). My specification of DNWR follows closely Eggertsson et al. (2017).

In this class of models, several papers have conducted shock-decomposition exercises. These papers typically find that a combination of financial and nominal shocks are needed to replicate aggregates during the crisis. See, for example, Christiano et al. (2015) and Del Negro et al. (2017). I extend this literature by applying this same shock decomposition logic to a heterogeneous-firm model, and use firm-level, rather than aggregate, data to infer the sources of shocks.

While I focus on firm-level data split by firm age to infer the sources of shocks, other papers take a more direct approach and use data on firms' financial positions to test for the existence of financial shocks. Chodorow-Reich (2014) and Giroud and Mueller (2015) perform this exercise for the US.

I present evidence that the downwards adjustment of real wages and interest rates during the crisis has been impaired by DNWR and the ZLB. Other studies have investigated these issues separately. Regarding wages, in a cross section of US counties, Mian and Sufi (2014) show that counties which were harder hit by the collapse of the housing bubble had no larger wage adjustment than other counties, and larger unemployment increases, suggesting a role for sticky wages. Other papers study downwards nominal wage rigidity by looking for a spike at zero nominal wage changes, and an associated missing mass below zero. Daly and Hobijn (2014) document an increase in this spike during the Great Recession in the US. Regarding real interest rates, Del Negro et al. (2017) present evidence that the natural rate of interest has fallen below the level achievable due to the ZLB. They argue that this is due to an increased preference for safety and liquidity, and it implies that the interest rate is being held artificially high by the ZLB floor.

The rest of the paper is organised as follows. In section 2 I present my empirical evidence. In section 3 I set up my model, and I discuss the calibration in section 4. In section 5 I present results away from the zero lower bound, and in section 6 I present results at the ZLB, and discuss fiscal policy. In section 7 I conclude.

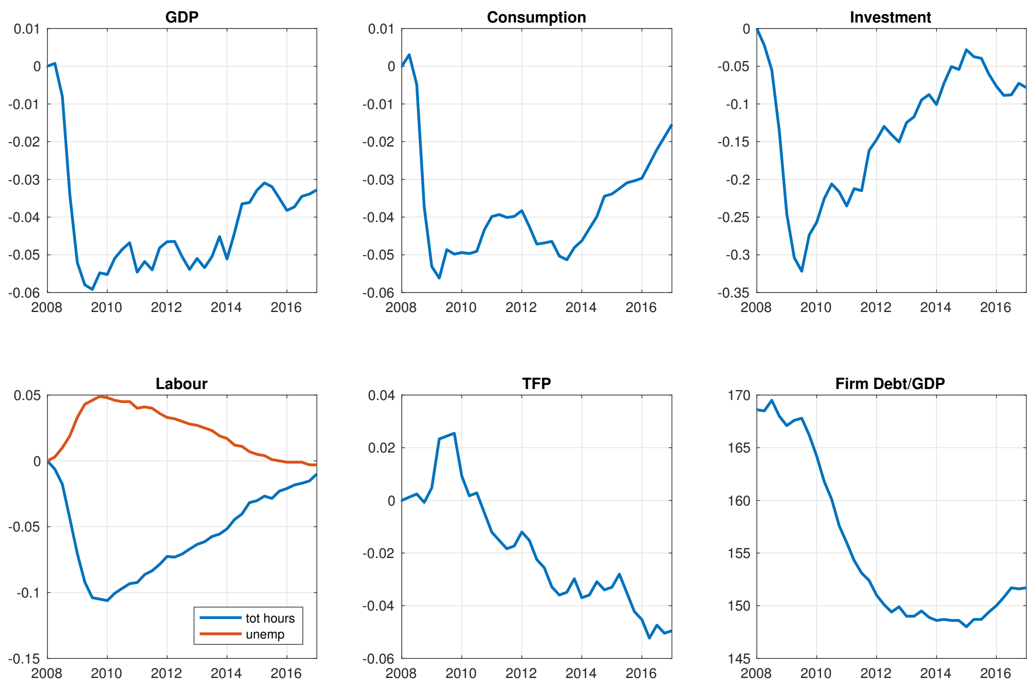
2 Empirical evidence

In this section I present evidence from various sources about how the Great Recession manifested. I begin with a review of common aggregates, and then perform a decomposition of aggregate employment using firm-level data. This decomposition will be the data that I use to test my models. I then discuss the behaviour of prices during the crisis, as a motivation for my focus on nominal rigidities.

2.1 Aggregates

The severity of the Great Recession is well known, and I document several common series in Figure 1. Output is shown in the top-left panel. Relative to the 2000-2007 trend, real GDP per capita falls 6% within two years, and is still 3% below trend at the end of the sample. One striking feature of the recession is how much it manifested in labour markets: the bottom-left panel shows that the unemployment rate rose by 5 percentage points, and total hours worked per capita by 10%. I will decompose the collapse of the labour market using firm level data in the next sections, but it is worth noting here that the huge disruption here is perhaps already a clue that the problems in the labour market were not constrained to a small subset of firms.

Figure 1: Great Recession: Aggregates



Quarterly national accounts data are from the BEA. Nominal variables are deflated with the GDP deflator. All but labour/debt are detrended with 2000-7 growth rates. TFP is util. adjusted and sourced from BFK. GDP, consumption, investment and hours are per capita. Debt/GDP data is from the Bank for International Settlements.

TFP is plotted in the bottom-centre panel, using Basu Fernald and Kimball's utilisation adjusted series. Relative to the pre-crisis trend, TFP does not fall on impact in the US. Rather, it actually increases by over 2%, before starting a gradual decline. This has two important implications about what kind of stories we can tell about the Great Recession. Firstly, and more obviously, a simple "TFP shock" story of the recession is hard to reconcile with the lack of a negative drop in TFP towards the beginning of the recession. This leaves us searching for other explanations for the crisis outside of the realm of standard RBC models.

Secondly, the lack of a TFP fall is also a problem for financial stories of the recession! As discussed in the literature review, models which use a tightening of financial frictions on firms to generate a recession typically lead to *endogenous* reductions in TFP. This is intuitive: tightened financial constraints misallocate resources across firms, which should reduce TFP. In this paper I will build a model which corrects this failing of existing models, and show how nominal rigidities will lead to a financial crisis manifesting in large declines in employment, rather than in misallocation.

In the bottom-right panel I plot the ratio of firm debt to GDP, provided by the Bank for International Settlements. The data shows a large and protracted deleveraging of firms, with debt falling by 20% from 170% of GDP pre-crisis to 150% of GDP post crisis. Debt has never returned to its pre-crisis levels, and I use this deleveraging as the key shock in my models.

2.2 Firm distribution pre-crisis

Before decomposing aggregate employment by firm age, it is helpful to illustrate what the firm distribution looked like before the crisis. This will give us a base to analyse the changes from, and will also illustrate the masses of the various groups of firms.

I use the Business Dynamics Statistics (BDS) data, provided by the US Census Bureau. BDS data provides repeated cross sections of the universe of US firms, aggregated to narrow firm size and/or age bins. Data is provided at the firm and establishment level, and I use on firm-level data. I will be primarily interested in financial frictions, and firm, rather than establishment, age is more likely to be related to firms' accumulated net worth. For ease of exposition, I aggregate these bins up to four age groups, and two size groups. Entrants firms are firms in their first year of operation. Young firms are 1 to 5 years old, Middle Aged firms are 6 to 15 years old, and Old firms are 16 years and older. Small firms have less than 50 employees, and Big firms have 50 or more.

Table 1 summarises the marginal distributions across firm age and size. The top row focuses on total employment within any bin. Starting with firm age, we see an interesting disconnect between total employment and the number of firms. Over 40% of firms are less than six years old (Entrant or Young firms) but they only account for 14% of total employment. Startups account for 10% of all firms, but only 3% of employment. Thus, young firms are small on average, which is intuitive. Old firms (16+ years) only account for 30% of firms, but account for nearly 70% of employment, and are hence large on average. Turning to firm size, we see a similar pattern: Small firms (1-49 employees) account for 96% of firms, but only 29% of total employment.

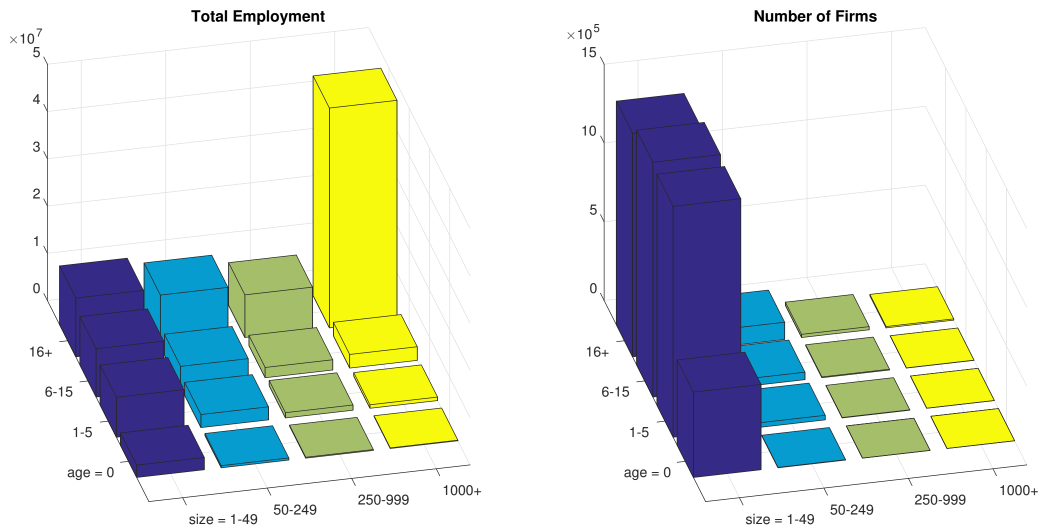
Rather than focusing on the marginal distributions of size and age, Figure 2 plots the joint distribution. In this plot I decompose firm size further into four bins. The most striking thing about the plots is the huge spike in total employment in the top-right bin: firms who are old (16+ years) and very large (1000+ employees) account for 40% of employment. This is despite accounting for only 0.16% of the number of firms (around 8200 firms)! The data also show that 99.5% of startup firms have less than 50 employees. This number is smoothly declining as firms get older, and falls to 91% for firms aged 16 and older.

Table 1: Firm distribution: summary statistics

	Small	Big	Entrant	Young	Middle Age	Old
Employment	0.293	0.707	0.029	0.113	0.171	0.686
No. of firms	0.956	0.044	0.106	0.291	0.300	0.302

Data are from Business Dynamics Statistics (BDS). Data is binned by firm age/size, and data is from 2005. Small: <50 employees. Large: 50+. Young: 1-5 years. Middle: 6-15. Old: 16+. Numbers are fraction of total employment and number of firms respectively.

Figure 2: The firm age / size distribution in 2005



Data are from Business Dynamics Statistics (BDS). Data is binned by firm age/size, and data is from 2005. Age = firm age. Size = no. of employees.

2.3 Responses by firm age

In this section I decompose the aggregate decline in employment by firm age. The BDS data is not panel data, but rather repeated cross sections. Hence I track how the employment in firms of different ages evolves during the crisis, and not the performance of fixed cohorts of firms.

In the left panel of Figure 3 I plot how employment evolved for three groups of firms. For ease of visualisation, here I combine Middle Aged and Old firms into a single bin containing all firms of age six and above. I calculate total employment in firms of each age bin each year, and plot it's deviation from the level of employment in 2007. The results are striking: total employment Startup and Young firms dropped by 20-25% from 2007, and does not recover. Total employment in Middle Aged and Old firms declines much less, falling by at most 5% before recovering.

Figure 3: Great Recession: total employment and number of firms by firm age

Data are from Business Dynamics Statistics (BDS). Data is binned by firm age and are repeated cross sections. "MA" refers to Middle Aged firms. Plotted as deviations from 2007 values.

This data thus emphasises the incredible disruption at Startup and Young firms during the crisis, and appears to place a large emphasis on them in explaining the aggregate decline in employment. However, there are two important things we must consider before making this judgement. Firstly, as shown in the last section, these firms make up a relatively small share of total employment (13%), so even a large decline in their employment may not translate into a large effect on aggregate employment. Secondly, the left panel measures total employment of firms of that age, and thus conflates the roles of the number of firms, and the average employment of each firm.

Indeed, moving to the right panel, I plot the change in the total number of firms in each age bin. We see that the number of Startup firms declined by 25% post-2007, which is surprisingly similar to the decline in total employment of Startups. Hence, simple intuition would suggest that the decline in employment in entrant firms is not driven by these firms becoming smaller on average, but rather a drastic decline in the rate of firm startups.

I make these points more formally in Table 2 and Table 8. In Table 2 I perform a shift-share decomposition of the total change in employment between 2007 and 2010. Denote total employment at time t as E_t, and total employment in group i as $E_{i,t}$. Denote the percentage change in employment by $\Delta E \equiv (E_{2010} - E_{2007})/E_{2007}$ and similarly for groups let $\Delta E_i \equiv (E_{i,2010} - E_{i,2007})/E_{i,2007}$. Denote by $S_i = E_{i,2007}/E_{2007}$ group is share of total employment in 2007. The shift share decomposition exactly decomposes ΔE as:

$$\Delta E = \sum_i \Delta E_i S_i \tag{1}$$

I define the contribution of group i as $\Delta E_i S_i/\Delta E$, so that the sum of contributions naturally adds up to 100%. Turning to the results in Table 2, we can now assess the importance of each age group of firms in explaining the total employment decline. Total employment declined by 7.4% from 2007 to 2010, and the table reveals that 40% of this decline is attributable to Startup and Young firms, while 60% is attributable to Middle Aged and Old firms.

Table 2: Decomposition of aggregate employment decline by firm age

	All	Entrant	Young	Middle Age	Old
ΔE_i	-7.4%	-20.8%	-20.8%	-8.6%	-4.4%
S_i	100%	2.5%	11.5%	16.5%	69.5%
Contribution	100%	7.2%	32.3%	19.2%	41.3%

See text for details of decomposition. Young = 1-5 years, Middle = 6-15 years, Old = 16+ years. "Contribution" refers to the percentage of the total employment decline ($\Delta E = -7.4\%$) explained by any one factor.

This is an interesting result, which highlights the importance of taking into account firm shares. The decline in total employment at younger firms was far more severe than at older firms (20.8% at younger firms vs less than 8.6% at Middle Aged and 4.4% at Old firms), but younger firms account for a smaller initial share of employment (13%). Hence in the decomposition, these two effects offset each other, leaving the contribution of younger and older firms much more balanced. How do we interpret these results? There are two main takeaways: 1) The collapse of employment at young firms has been enormous, and is interesting in its own right. 2) This large decline means that young firms account for more than their fair share of the decline in total employment, at around 40%. However, the smaller decline in employment at older firms, due to the sheer number of older firms, still accounts for 60% of the employment decline and cannot be ignored.

Table 3: Decomposition: average employment within firms vs number of firms

Change:	All	Entrant	Young	Middle Age	Old
$\Delta(E/N)$	-2.0%	7.9%	-10.1%	-7.5%	-7.7%
ΔN	-5.5%	-26.7%	-11.9%	-1.1%	3.6%
ΔE	-7.4%	-20.8%	-20.8%	-8.6%	-4.4%
Contribution:	All	Entrant	Young	Middle Age	Old
$\Delta(E/N)$	27.4%	-2.7%	15.7%	16.8%	72.5%
ΔN	74.1%	9.2%	18.6%	2.6%	-33.8%
ΔE	100%	7.2%	32.3%	19.2%	41.3%

$\Delta E_i = \Delta(E_i/N_i)\Delta N_i$. Young = 1-5 years, Middle = 6-15 years, Old = 16+ years. Average employment calculated at bin level. "Contribution" refers to the percentage of the total employment decline ($\Delta E = -7.4\%$) explained by any one factor.

I next turn to decomposing the results by incorporating firm entry and exit. We can exactly decompose the percentage change in total employment ($E_{i,t}$) in an age bin into a change in average employment ($E_{i,t}/N_{i,t}$) and a change in the number of firms ($N_{i,t}$). The exact decomposition is multiplicative:

$$1 + \Delta E_i = (1 + \Delta(E_i/N_i))(1 + \Delta N_i) \Rightarrow \Delta E_i = \Delta(E_i/N_i) + \Delta N_i + \Delta(E_i/N_i) \times \Delta N_i \tag{2}$$

Given that the multiple of the two growth rates will be very small (in the data it accounts for only 1% of the change in employment) we can safely ignore it and focus on the intuitive first order approximation:

$$\Delta E_i \simeq \Delta(E_i/N_i) + \Delta N_i \tag{3}$$

That is, the change in employment of firms of age i is roughly equal to the change in average employment plus the change in the number of firms. These changes are given in the top half of Table 8 (the numbers within a column do not add up exactly due to the missing multiplicative term). We learn several things from this decomposition. Firstly, the decline in employment at Startup and Young firms is overwhelmingly driven by a decline in the number of firms: the number of startups declined by 26.7%, and the number of Young firms by 11.9%. Startup firms on average actually got larger during the recession, by 7.9%, and this is likely due to a composition effect: only larger startups were able to successfully enter during the recession.

Turning to older firms, we see no such decline in the number of firms. The number of Middle Age firms does drop by 1.1%, but the number of Old firms increases by 3.6%. Thus, the decline in total employment in older firms is driven by a decline in average employment within these firms.

The key lesson of this decomposition is that, once entry and exit are controlled for, both younger and older firms all shrink during the crisis, and by a comparable amount: Young firms shrink on average by 10.1%, Middle Aged firms by 7.5% and old firms by 7.7%. While *average* employment at young firms does shrink by around 2.5 percentage points more, this difference is small compared to the difference in the behaviour in *total* employment.

We can compute the contribution of each of these elements to the total change in employment by combining this with the shift-share decomposition, $\Delta E = \sum_i \Delta E_i S_i$, giving us

$$\Delta E \simeq \sum_i (\Delta(E_i/N_i) + \Delta N_i)S_i = \sum_i (\Delta(E_i/N_i)S_i + \Delta N_i S_i) \tag{4}$$

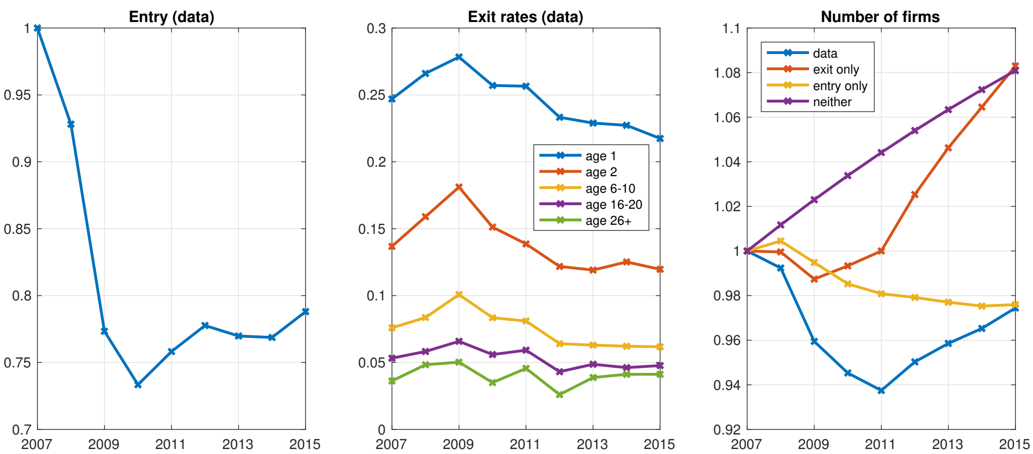
The results of this decomposition are given in the bottom half of Table 8. This table shows us that the single largest contribution to the decline in total employment was the decline in employment at Old firms, accounting for 72.5% of the decline. If we look at Entrant and Young firms, however, the larger contribution comes from the decline in the number of firms, with the decline across the two groups accounting for 27.8% of the decline in aggregate employment.

Interestingly, the decomposition also reveals the importance of digging into the aggregate numbers for understanding the contributions of average employment versus entry and exit. Looking at all firms in Table 8, the total employment decline of 7.4% is decomposed into a 2% decline in average employment and a 5.5% decline in the number of firms. But looking across (and down) to the contributions of different firm age bins, we see a more nuanced picture. The single biggest contributor to the decline in employment is the decline in average employment in old firms, accounting for fully 72.5% of the total decline. How do we resolve this apparent contradiction? Composition effects are key. The decline in entry during this period shifted the distribution of firms towards older firms, who tend to have higher employment. This mitigates the decline in average employment when looking at the sum of all firms. By performing a decomposition by firm age we correct for these effects, and can see the true importance of declines in average employment.

As one final exercise with this data, consider a counterfactual scenario whereby Young firms only shrank their average employment as much as old firms did. In the data they shrunk more, with average employment shrinking by 10.1% as opposed to the 7.5% of Middle Aged firms and 7.7% of old firms. This larger response of Young firms speaks to the role of financial frictions. However, if we compute a counterfactual where we impose that Young firms only shrink their average employment by 7.5% (the same as the Middle Aged) then total employment still falls by 7.1%. This is not much different from the 7.4% we see in the data. In other words, given the relatively small employment share of Young firms, the fact that they shrank average employment by 2.6 percentage points more than older firms has little hope of explaining much of the aggregate employment decline.

This exercise also speaks to the potential problems with the data coming from composition effects. It could be that the true decline in average employment at young firms is larger than the 10.1% reported in the binned data, if, for example, smaller firms exit more within any given bin. Even if this were true, the small share of young firms means this has only a small effect on the decompositions: allowing for a counterfactual 20% decline in employment in Young firms only increases the total employment decline to 8.4$. Supporting this, Siemer (2014) uses the confidential firm-level data behind the BDS and finds that the differences in employment growth between young and old firms are driven by entry and exit, and not differences in employment growth of continuing firms.

Figure 4: Entry and exit decomposition



Data are from Business Dynamics Statistics (BDS). Left panel shows total entry, relative to 2007 level. Middle panel shows the yearly exit rate of firms by selected firm ages. Right panel shows the number of firms in the data in blue, and three counterfactuals. "entry only" holds exit rates at 2007 levels and simulates employment allowing only entry to change. "exit only" holds entry at 2007 levels and allows exit rates to vary, and "neither" holds both entry and exit rates at their 2007 levels.

One concern with the above analysis is that it did not distinguish between entry and exit, and focused only on the number of firms. Thus I also perform an exercise where I decompose changes in the number of firms into the change coming from entry and exit. To do this, I calculate exit rates by firm age coming from the BDS data, and simulate counterfactual scenarios.[6] Figure 4 plots the results. The left panel shows the number of entrant firms by year relative to 2007, showing the massive decline in entry. The middle panel plots firm exit rates for selected firm ages. These increase in the early years of the crisis, especially at younger firms. The right panel plots three decompositions. The blue line plots the number of firms in the data. The yellow "entry only" line plots the simulated path holding all exit rates constant at their 2007 levels and only allowing the entry rate to fall. Vice versa, the red "exit only" line plots the simulated path only allowing exit rates to vary, and "neither" plots the simulation holding both entry and exit rates at their 2007 levels. The results show that, over the whole sample, it is the massive decline in entry which is more important for driving the number of firms. By 2015, the decline in entry alone can capture the whole decline, while the change in exit rates predicts continued growth in the number of firms. However, in the first few years of the crisis both entry and exit appear to be important, with heightened exit rates even being slightly more important for the first two years in explaining the decline in the number of firms.

The results of this section can be summarised in two findings. Firstly, the bulk of the aggregate decline in employment can be accounted for by two things: (1) a decline in the *number* of young firms by around 25%, and (2) a decline in employment *within* older firms of around 7.5%. This challenges narratives which rely on contractions in younger firms only to explain the crisis, with older firms accounting for about 60% of the employment fall. Secondly, both younger and older firms contracted significantly during the crisis. As we will see when I move on to analysing models of financial frictions, this is troubling for pure financial frictions stories which affect output via misallocation. These models want to predict a reallocation of resources towards financially constrained firms, which tend to be older firms.

In Appendix A I perform several additional exercises. Firstly, we might be worried that defining Old firms as firms aged 16+ might leave some relatively young or financially constrained firms defined as Old. I show that all of the results in this section are identical if we instead define Old firms as those of age 26 or above, which is the oldest age bin in the BDS data. The results are also robust to alternative start and end dates of the recession. Secondly, I provide a comparison with the previous two recessions where I show that the behaviour of average employment in Young versus Old firms is remarkably similar, and it is a combination of 1) the massive decline in entry and 2) larger average employment declines in all age bins, which distinguishes the Great Recession from previous recessions.

2.4 Real wage and interest rate data

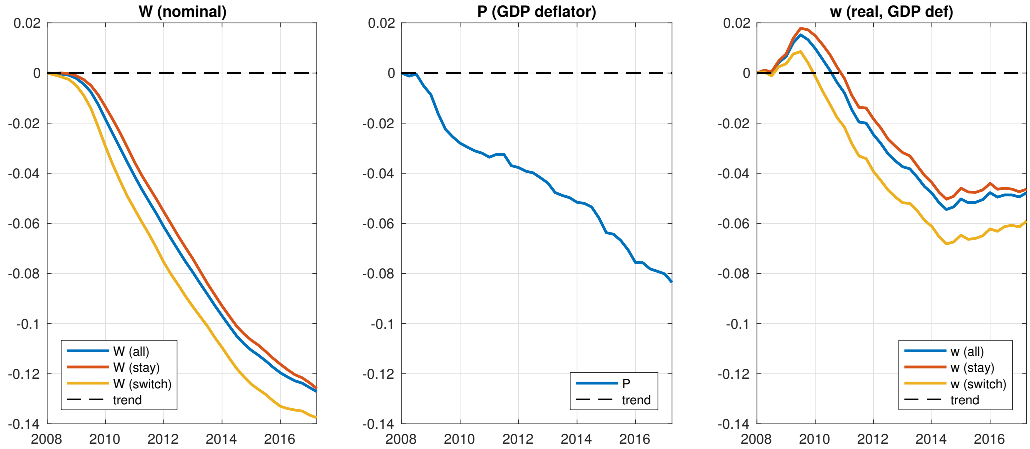
Finally, I turn to data on prices. This data is important for two reasons. Firstly, I use it to argue that the factor-price implications of the standard "real" financial frictions models are inconsistent with how factor prices actually evolved during the crisis. Secondly, I show how they motivate my focus on nominal rigidities as a factor in their actual evolution.

[6] I simulate employment according to $n_t^a = (1 - x_t^a)n_{t-1}^{a-1}$ where n_t^a is the number of firms of age a at time t, and x_t^a is the calculated exit rate. The procedure is complicated slightly by the binning of several ages into larger bins, and I assume the exit rates of all ages are the same within any bin.

2.4.1 Real wages

Figure 5 plots various measures of real and nominal wages during the financial crisis. Nominal wage data are from the Atlanta Fed's Wage Growth tracker, which uses microdata from the Current Population Survey (CPS). Nominal wage growth is the median percentage change in the hourly wage of individuals observed 12 months apart, and I use this to construct an index of nominal wages. In the left panel I plot three such indices. The blue line gives the index for the wage of all agents in the sample. This is decomposed in the red and yellow lines into wages of job movers, and job stayers. This is to alleviate concerns that wages of new hires are more flexible than wages in existing jobs, and that these wages are the relevant wages for hiring. All series are plotted as deviations from their 2000 to 2007 trend. We see declines in all measures of nominal wages, with the declines in the wages of new hires being the largest.

Figure 5: Measures of nominal and real wages during the crisis



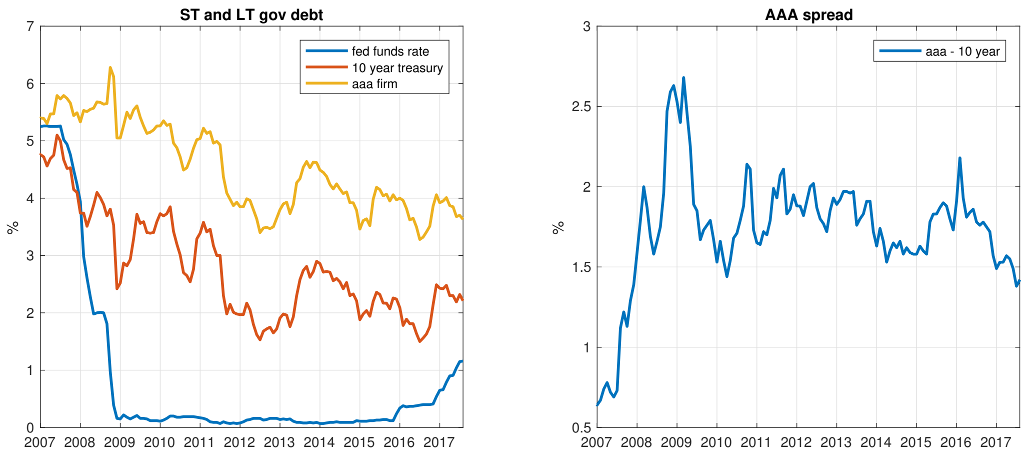
Data are from the Atlanta Fed's Wage Growth tracker, which uses microdata from the Current Population Survey (CPS). Growth is median percentage change in the hourly wage of individuals observed 12 months apart, and I use this to construct an index. Left panel shows nominal wages of all workers, and job stayers and switchers. Middle panel gives the GDP deflator, and right panel deflates nominal wages using the GDP deflator. All variables detrended with 2000-2007 growth rates.

However, these are only nominal wages. How do real wages evolve? The second panel plots the GDP deflator, which we see has fallen significantly behind trend during the crisis. That is, inflation (as measured by the GDP deflator) has been persistently below its pre-crisis trend. The right panel plots real wages by deflating the nominal wage series (left panel) with the price level (centre panel). Given the deflation, the real wage falls by much less than the nominal wage. In fact, due to the severe disinflation during the beginning of the recession, real wages actually rise by up to 2% during the beginning of the recession, a period where unemployment was hitting 10%. Even only focusing on job-switchers, real wages only fall by 2% in the first three years of the crisis, despite the large declines in employment. The lack of a significant decline in wages will turn out to be important in determining how a financial crisis is transmitted through the firm distribution. As seen by the difference between nominal and real wages, the decline of inflation during the crisis is very important (in a decompositional sense) for explaining the lack of a decline in real wages. I take this as strong suggestive evidence that nominal frictions, such as downwards nominal wage rigidity, combined with a declining price level contributed to keeping wages artificially high during the crisis.

2.4.2 Interest rates

Figure 6 plots measures of interest rates during the crisis. The left panel plots in blue the Federal Funds Rate, the key short term interest rate under the central bank's control. This rate dropped by 5% as the interest rate was lowered to the zero lower bound (ZLB) to combat the recession. However, there are two important questions to be asked. Firstly, how much further would this rate have fallen if it was not constrained by the ZLB? Secondly, much much did this fall in interest rates actually translate into cheaper funding for firms looking to borrow?

Figure 6: Measures of interest rates during the crisis



Left panel shows nominal interest rates: the federal funds rate and 10 year treasury rate. The right panel plots the AAA spread, which is the spread of safe (AAA) corporate debt over the 10 year treasury rate.

To answer the first question requires estimates of the natural rate of interest, which serves as a target for how the interest rate would have been adjusted without the ZLB constraint. Del Negro et al. (2017) provide estimates using both a pure empirical and a DSGE based specification. They estimate that the natural real interest rate could have been as low as -4.5% during the first years of the crisis. Due to the ZLB on nominal rates, the actual real interest rate was only able to fall to around -1.5%, giving a roughly 3% gap (see their Figure 15). This represents a sizeable deviation from the flexibility of the real interest rate assumed in the standard model of financial crises.

To answer the second question, I look at several other measures of interest rates. In the left panel of Figure 6 I also plot in red the 10 year treasury nominal interest rate and the super-safe (AAA) firms' borrowing rate, which is an average of long-term rates with average maturity. This shows that while the short-term policy rate fell drastically, the change in longer term rates was smaller. The rate on long-term government debt fell by around 2% in the first few years of the crisis, and eventually by 3% over the whole sample.

Since firms are typically making longer term investments, longer term borrowing rates are much more important for them in making investment decisions, and these came down less than short term rates.

Of course, firms borrow at different interest rates than the government,

These firms essentially never default, and so the spread between these rates represents the difference in liquidity of these assets. Interestingly, the spread increases dramatically during the crisis, peaking at a 2.5% spread, making a 2 percentage point rise from the pre-crisis level. This shows that even as safe *government* interest rates were falling, the rates that *firms* could borrow at were falling much less, because the increased demand for liquidity meant that people were not willing to lend to firms.

Thus, compared to the standard model, the data shows two key differences. Firstly, the ability of interest rates to fall was constrained by the ZLB. Secondly, due to an increased demand for liquidity, lending rates to firms came down even less than safe government rates.

2.5 Data summary

The story behind these various pieces of data can be summarised as follows. The financial crisis led firms to contract their size (average employment) across the firm age spectrum. While younger firms do contract more than older firms, even old firms contract employment by 7.5%. The total decline in employment can broadly be accounted for by a huge decline in firm entry, and the contraction in average employment in old firms, and both are required for a complete picture of the labour market. This speaks against stories of the crisis which focus just on trouble at small firms, and the misallocation of resources. We do not see a reallocation of resources towards old firms, and rather see declines in employment across the firm age spectrum. This is reflected in how the crisis manifested in aggregates: we see a large decline in aggregate employment, since all firms are shrinking, but we do not see any large decrease in TFP, which suggests little misallocation at the aggregate level.

Finally, I argue that must be related to the behaviour of factor prices, and show that, most likely due to nominal rigidities, real wages and interest rates paid by firms do not seem to fall as far as they "should" during the crisis. In the models below, I will show how this lack of movement in factor prices naturally generates both the firm level and aggregate level facts above, in contrast to the standard financial frictions model.

3 Model

In this section I build a dynamic, calibrated heterogeneous firm model (in the spirit of Hopenhayn and Rogerson (1993) and Khan and Thomas (2008)). The model builds most closely on Khan and Thomas (2013), who incorporate financial frictions and an aggregate financial shock. To this I add an explicit nominal structure, which I specify below. The experiments are conducted under perfect foresight.[7]

[7] 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.

3.1 Environment

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

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 CES bundle of intermediate goods. These are produced by a continuum of intermediate-goods producing firms, who are the main firms of interest. These firms will be hit by both permanent and idiosyncratic shocks, and these will be calibrated to match the firm distribution from the BDS data above. They produce via a constant returns to scale production function in capital and labour, have heterogeneous productivities and may be financially constrained. A fraction of firms will endogenously emerge as financially constrained depending on their productivity and net worth, and firms have well defined optimal sizes due to the assumption of monopolistic competition and downwards sloping demand functions. Finally, there is a central bank, who will conduct monetary policy by setting the nominal interest rate. In the final section I also consider a government, who may conduct fiscal policy and make fiscal transfers.

3.1.1 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:

$$Y_t = \left(\int_0^1 y_{i,t}^{\rho} \, \mathrm{d}i \right)^{\frac{1}{\rho}} \tag{5}$$

Where I define $\epsilon = 1/(1-\rho)$ as the elasticity of substitution between varieties. Final output, Y_t, is produced from a CES bundle of intermediates goods $y_{i,t}$, each produced by a different intermediate firm, i.[8] The final-goods firm is a price taker in both the final and intermediates markets. Profit is given by:

$$\pi_t^{fg} = \left(\int_0^1 y_{i,t}^{\rho} \, \mathrm{d}i \right)^{\frac{1}{\rho}} - \int_0^1 p_{i,t} y_{i,t} \, \mathrm{d}i \tag{6}$$

$p_{i,t}$ is the price (in terms of numeraire) of intermediate i. The first order condition for intermediate purchase from i gives:

$$y_{i,t} = p_{i,t}^{-\epsilon} Y_t$$

3.1.2 Intermediate good firm's static problem

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. Each firm has a CRS production function:

$$y_t = z_t k_{t-1}^{\alpha} l_t^{1-\alpha} \tag{7}$$

[8] $i \in [0, 1]$ 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.

We can 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. Real static profit is given by:

$$\pi_t^f(k_{t-1}, z_t) = \max_{l_t} z_t^\rho k_{t-1}^{\alpha\rho} l_t^{(1-\alpha)\rho} Y_t^{1-\rho} - w_t l_t \tag{8}$$

π^f denotes real firm profits, and $w_t = W_t/P_t$ is the real wage. Notice that despite the assumption of CRS production functions, the revenue function is decreasing returns to scale at the firm level due to the downwards sloping demand curve, as captured by ρ. Once maximised, optimised profit, labour demand, and output are given by

$$\pi_t^f(k_{t-1}, z_t) = \left(\nu^{\frac{\nu}{1-\nu}} - \nu^{\frac{1}{1-\nu}}\right) \left(Y_t^{1-\rho} w_t^{-\nu}\right)^{\frac{1}{1-\nu}} z_t^{\frac{\rho}{1-\nu}} k_{t-1}^{\frac{\alpha\rho}{1-\nu}} \tag{9}$$

$$l_t(k_{t-1}, z_t) = \left(\frac{\nu Y_t^{1-\rho}}{w_t}\right)^{\frac{1}{1-\nu}} z_t^{\frac{\rho}{1-\nu}} k_{t-1}^{\frac{\alpha\rho}{1-\nu}} \tag{10}$$

$$y_t(k_{t-1}, z_t) = \left(\frac{\nu Y_t^{1-\rho}}{w_t}\right)^{\frac{1-\alpha}{1-\nu}} z_t^{\frac{1}{1-\nu}} k_{t-1}^{\frac{\alpha}{1-\nu}} \tag{11}$$

One important thing to note about these functions is the inclusion of Y_t in profit and optimal choices. This is the "demand spillover" which arises in the CES model. Each firm's demand curve, $y_{i,t} = p_{i,t}^{-\epsilon} Y_t$, depends on aggregate output, Y_t. When aggregate output is low, each firm's demand as low and firms will optimally choose to respond to this by shrinking (unless offset by falling factor prices).

3.1.3 Intermediate good firm's dynamic problem

The rest of the firm's problem is dynamic, and concerns the decisions of how much to invest, how much to borrow, and how much to pay out as dividends. Firms maximise the discounted sum of dividends, and discount using the representative household's stochastic discount factor.

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.

Firms borrow using long term nominal debt. Following Gomes et al. (2016), these are decaying coupon bonds. One bond issued at time t promises to pay a constant nominal coupon c each period, until paying nominal face value 1 when it is retired. Bonds are retired each period with probability δ_c. Suppose that a firm starts the period with $\tilde{d}_{t-1}$ bonds outstanding. The cashflow requirement is to pay coupons for all bonds, $c\tilde{d}_{t-1}$ and to pay the principal on the δ_c which retire, $\delta_c\tilde{d}_{t-1}$. The firm can issue new bonds at nominal price Q_t^c. Suppose a firms issues f_t, then the new total amount of bonds starting next period is $\tilde{d}_t = f_t + (1-\delta_c)\tilde{d}_{t-1}$. A firm's nominal cashflow is given by

$$P_t(i_t + e_t) + (c + \delta_c)\tilde{d}_{t-1} = P_t\pi_t^f + Q_t^c f_t \tag{12}$$

Where $e_t \geq 0$ are real dividend payments, and i_t is real investment expenditures, or receipts from disinvestment. Plugging in the flow equation for $\tilde{d}_t$ to replace f_t and rearranging yields:

$$i_t + e_t + r_{t,t-1}d_{t-1} = \pi_t^f + d_t \tag{13}$$

where $d_t \equiv Q_t^c \tilde{d}_t / P_t$ is the real value of outstanding debt, and I define the "realised" interest rate as

$$r_{t-1,t} = \frac{c + \delta_c}{\pi_t Q_{t-1}^c} + \frac{Q_t^c}{Q_{t-1}^c \pi_t}(1 - \delta_c) \tag{14}$$

Inflation is defined as usual as $\pi_t \equiv P_t / P_{t-1}$. I discuss in later section how to determine the price of firms' debts, and hence $r_{t-1,t}$, using arbitrage with the one-period real interest rate, r_t.

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:

$$n_t = \pi_t + (1-\delta)k_{t-1} - r_{t-1,t}d_{t-1} \tag{15}$$

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

$$q_t k_t + e_t = \tilde{n}_t + d_t \tag{16}$$

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:

$$n_{t+1} = \left(\frac{\pi_{t+1}(k_t, z_{t+1})}{k_t} + (1-\delta)q_{t+1}^i - q_t r_{t,t+1} \right) k_t + r_{t,t+1}\tilde{n}_t - r_t e_t \tag{17}$$

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:

$$d_t \leq \lambda_t q^d k_t \tag{18}$$

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

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

$$\mathrm{E}_t\left[\beta\frac{u'(c_{t+1})}{u'(c_t)}\Big((1-\sigma)(n_{t+1}-(1-q^d)(1-\delta)k_t)+\sigma V_{t+1}(n_{t+1},k_t,z_{t+1})\Big)\right]\Bigg\} \tag{19}$$

where the maximisation is subject to the net worth transition, (17), with $q_t = 1$ and $\tilde{n}_t = n_t$. The maximisation incorporates the collateral constraint, which, combined with (16), 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

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

$$\mathrm{E}_t\left[\beta\frac{u'(c_{t+1})}{u'(c_t)}\Big((1-\sigma)(n_{t+1}-(1-q^d)(1-\delta)k_t)+\sigma V_{t+1}(n_{t+1},k_t,z_{t+1})\Big)\right]\Bigg\} \tag{20}$$

again subject to (17), 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:

$$V_t(n_t,k_{t-1},z_t) = \max\left\{V_t^i(n_t,k_{t-1},z_t), V_t^d(n_t,k_{t-1},z_t)\right\} \tag{21}$$

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 (17) and the borrowing policy function, $d_t = d_t(n_t, k_{t-1}, z_t)$, by (16).

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.

3.1.4 Entry and exit

As I discuss in the calibration section, I will use permanent differences across firms to match the wide firm size distribution. Entrant firms will draw a type, and also draw a value of idiosyncratic productivity from the ergodic distribution for their type. Exit is exogenous, and firms exit each period with i.i.d probability $(1-\sigma)$.

Firm entry is endogenous, and modelled in a reduced form way. Firms can be one of several firm types, indexed by $j = 1, ..., J$ (in the calibration there will be two types: "normal" and "growth" firms). If an entrepreneur sets up a firm of type j this period, they draw an initial value of productivity z_t (realised after entry) drawn from a type-specific productivity distribution $P_j(z_t)$.

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.

Hence the expected value of setting up a firm of type j is the expectation over the value function with $n_t = n_e$ and $k_{t-1} = 0$: $v_t^{e,j} = E_t^j V_t(n_e, 0, z_t)$ where E_t^j refers to the expectation with respect to the type-specific distribution $P_j(z_t)$.

Entry is assumed to depend positively on the expected value of setting up a firm. Let $\mu_t^{e,j}$ denote the mass of entrants of type j at time t. This depends on $v_t^{e,j}$ according to:

$$\mu_t^{e,j} = a_j \left(v_t^{e,j}\right)^{b_j} \tag{22}$$

where $a_j > 0$ controls the average level of entry and $b_j > 0$ controls the sensitivity of entry to expected firm value.

3.1.5 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 $K_{t-1} = \int_0^1 k_{t-1}\,\mathrm{d}i = \int k_{t-1}\,\mathrm{d}\mu_t$. While the model does not aggregate fully, and must be solved with heterogeneous firm methods, there is a useful partial aggregation result. The aggregate production function can be expressed as:

$$Y_t = Z_t K_{t-1}^{\alpha} L_t^{1-\alpha} \tag{23}$$

where TFP is endogenous and given by:

$$Z_t = \left(\int_0^1 z_t^{\frac{\rho}{1-\nu}} \tilde{k}_{t-1}^{\frac{\alpha\rho}{1-\nu}}\,\mathrm{d}i\right)^{\frac{1-\nu}{\rho}} = \left(\int z_t^{\frac{\rho}{1-\nu}} \tilde{k}_{t-1}^{\frac{\alpha\rho}{1-\nu}}\,\mathrm{d}\mu_t\right)^{\frac{1-\nu}{\rho}} \tag{24}$$

Additionally, we can derive an analytical formula for the aggregate labour demand, which takes the form of a constant labour share:

$$w_t L_t = (1-\alpha)\rho Y_t \tag{25}$$

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, (5). The labour share is thus constant, and equal to the usual Cobb Douglas formula, adjusted for market power, ρ. 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.

3.1.6 Household

There is a representative household, who supplies labour, consumes, owns firms, and provides funds to the firm sector. 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 in the case of rationing. The household maximises the discounted sum of utility

$$U = \sum_{t=0}^{\infty} \beta^t \left(u(C_t) - v(L_t)\right) \tag{26}$$

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

$$C_t + D_t = w_t L_t + r_{t-1,t} D_{t-1} + E_t \tag{27}$$

where D_t is lending to firms, and E_t is dividend income from firms, net of equity injections to entering firms. If we allow the household to trade in a risk-free one-period real asset, the standard Euler equation is

$$1 = \beta \frac{u'(C_{t+1})}{u'(C_t)} r_t \tag{28}$$

We can price a nominal riskless asset using the Fisher equation, $i_t = r_t/\pi_{t+1}$. I discuss how to price firm debt in the next section.

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:

$$w_t = \frac{v'(L_t^s)}{u'(C_t)} \tag{29}$$

This is the standard labour supply first order condition.

3.2 Pricing firm's long term debt

With perfect foresight, i.e. no aggregate uncertainty, we can price firm debt via arbitrage with the one-period real bond. The returns must be equal, giving:

$$r_{t,t+1} = \frac{c + \delta_c}{\pi_{t+1} Q_t^c} + \frac{Q_{t+1}^c}{Q_t^c \pi_{t+1}}(1 - \delta_c) = r_t \tag{30}$$

This allows us to back out the current bond price using a terminal condition on Q_t^c and backwards recursion on

$$Q_t^c = \frac{1}{r_t \pi_{t+1}} \left(c + \delta_c + (1 - \delta_c) Q_{t+1}^c \right) \tag{31}$$

The steady state bond price is given by

$$Q_{ss}^c = \frac{c + \delta_c}{\pi_{ss}(r_{ss} + \delta_c - 1)} \tag{32}$$

The only exception to this will be when I perform unanticipated shock experiments. Here, I start from the steady state and subject the economy to an unanticipated shock at time 1. Then, in the period that the unanticipated shock occurs, the ex-post interest rate is calculated as

$$r_{0,1} = \frac{c + \delta_c}{\pi_1 Q_{ss}^c} + \frac{Q_1^c}{Q_{ss}^c \pi_1}(1 - \delta_c) \tag{33}$$

and may differ from the ex-ante real interest rate, r_t, since the shock was not anticipated. We see from this equation how unanticipated inflation (π_1) increases the ex-post interest rate paid, increasing the value of firms' debt. With long term debt this is more severe, because the price of outstanding debt, Q_1^c is also affected.

3.3 Sticky wages

I specify a simple form of downwards nominal wage rigidity, DNWR, following Eggertsson et al. (2017):

$$W_t \geq \gamma \bar{\pi} W_{t-1} + (1-\gamma) P_t w_t^{mc} \tag{34}$$

where $\bar{\pi}$ is the indexation rate, γ controls the degree of rigidity, and w_t^{mc} is the market clearing wage, which would set labour supply equal to labour demand:

$$w_t^{mc} = \frac{v'(L_t)}{u'(C_t)} \tag{35}$$

This says that the nominal wage is free to fall, until it hits the barrier defined by DNWR. We can express this in real terms as

$$w_t \geq \frac{\gamma \bar{\pi} w_{t-1}}{\pi_t} + (1-\gamma) w_t^{mc} \tag{36}$$

To see how this works, consider that when $\gamma = 0$ the constraint reduces to $w_t \geq w_t^{mc}$, and DNWR are switched off. If $\gamma = 1$ then $w_t \geq \frac{\gamma \bar{\pi} w_{t-1}}{\pi_t}$, so the real wage can only fall by at most $\frac{\gamma \bar{\pi}}{\pi_t}$ each period. Since the constraint is defined in nominal terms, higher inflation helps to overcome DNWR, and it is more likely to bind under deflation. Hence, γ controls speed of adjustment to the market clearing wage, and the degree of indexation is also important in determining how likely DNWR is to bind.

3.3.1 Central bank

The central bank conducts policy by choosing values for the nominal interest rate, subject to the zero lower bound (ZLB):

$$i_t = \max\{i_t^*, 0\} \tag{37}$$

where i_t^* is the bank's unconstrained choice of the interest rate. I variously assume different rules for the central bank in the experiments below.

3.3.2 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

$$C_t + \sigma \left(\int_{i_t \geq 0} i_t \, \mathrm{d}\mu_t + \int_{i_t \leq 0} q^d i_t \, \mathrm{d}\mu_t \right) + (1-\sigma) \left(\int i_t^e d\Gamma_{e,z} - \int q^d (1-\delta) k_{t-1} \, \mathrm{d}\mu_t \right) = Y_t \tag{38}$$

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 difference between difference between labour demand and supply:

$$U_t \equiv L_t^s - L_t \tag{39}$$

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

3.3.3 Definition of equilibrium

Definition 1. *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, (19), (20), and (21).* $\{l_t, y_t, d_t, n_t\}_{t=0}^{\infty}$ *are defined by (10), (11), (16), and (17).*
2. $\{C_t, D_t\}_{t=0}^{\infty}$ *solve the household's problem, (26), subject to (27).*
3. *Output is given by the aggregator, (5). 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* $t = 0, 1, \ldots$*. Aggregate TFP is defined as the residual in (23).*
4. *The goods market clears for all* t*: (38).*
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 Calibration

The model is calibrated yearly. Firm level parameters are set to match moments of firm-level data in the model's stationary distribution, as discussed in detail in the next section. 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. The steady state wage, w_{ss}, is set to normalise labour supply to 1/3. The depreciation rate δ is set to 0.065. For comparability with Khan and Thomas' (2013) results, I specialise household utility to $u(c_t, l_t) = \log(c_t) - Dl_t$, although this choice is not important for the qualitative message of the paper.

4.1 Firm-level calibration

Firm-level parameters are chosen in two sets. The first set targets moments outside of the firm distribution, and the second set is chosen to target moments of the firm distribution from the BDS presented in this paper.

Starting with the first set, 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). 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 degree of irreversibility, the resale price of capital, from Khan and Thomas (2013) and set it to 0.95.

Table 4: 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	1.16	Normalise $Y_{ss} = 1$
w_{ss}	Mean wage	1.64	$L_{ss} = 1/3$
	Firm level:		
ρ	Good substitution	0.82	Cooper and Haltiwanger (2006)
σ_z	std($z_{i,t}^{tr}$)	0.06	std($ik_{i,t}$) $= 0.34$ *(large firms)*
ρ_z	autocorr. $z_{i,t}^{tr}$	0.65	Khan and Thomas (2013)
q^d	Capital resale price	0.95	Khan and Thomas (2013)
λ_{ss}	Collateral rate	0.51	$D_{ss}/A_{ss} = 0.37$
σ	Survival rate	0.9	Exit rate 10%
n_e	New firm equity	0.42	Entrant emp = 2.9%
$z^{G,0}/z^N$	Growth firm prod	1.35	Large firm emp = 71%
ρ^N	Frac. normal firms	0.1	Frac small = 2.9%
ρ^G	Prob z^G doubles	0.0667	Double in 15 years

Table 5: Firm distribution: summary statistics

Employment:	Small	Big	Entrant	Young	Middle Age	Old
Data	0.293	0.707	0.029	0.113	0.171	0.686
Model	0.294	0.706	0.029	0.170	0.410	0.391
No. of firms:	Small	Big	Entrant	Young	Middle Age	Old
Data	0.956	0.044	0.106	0.291	0.300	0.302
Model	0.955	0.045	0.100	0.369	0.346	0.185

Data are from Business Dynamics Statistics (BDS). Data is binned by firm age/size, and data is from 2005. Small: <50 employees. Large: 50+. Young: 1-5 years. Middle: 6-15. Old: 16+

Moving on to matching the firm distribution, I adopt a two-level structure for firms' productivity in order to simultaneously match the across- and within firm distributions. Firms' productivity is given by:

$$z_{i,t} = z_{i,t}^{p} z_{i,t}^{tr} \tag{40}$$

Total productivity, $z_{i,t}$ is split into a permanent, $z_{i,t}^{p}$, and a transitory, $z_{i,t}^{tr}$ component. The transitory component follows an AR(1) process, and is chosen to match within-firm volatility. The permanent component is chosen to match the across-firm distribution, and to capture demand growth facts.

In particular, there are two groups of firms: "normal" firms and "growth" firms. Normal firms have permanent productivity given by $z_{i,t}^{p} = z^N$, with z^N a low number. Thus, these firms are optimally small, relative to growth firms, and their demand does not grow over time. This form is chosen to match the large mass of small firms of all ages that we see in the firm-level data. In my calibration, normal firms will not be financially constrained in equilibrium.

Growth firms are born with the permanent component equal to $z^p_{i,t} = z^{G,0} > z^N$. This is higher than for normal firms, so growth firms will have a larger target size. Additionally, their productivity will grow over time: with probability ρ_G, their productivity doubles to $z^{G,1} = 2 \times z^{G,0}$. This is to match data from Foster et al. (2016) who show that demand doubles as firms age and expand their customer base. ρ_G is chosen so that it takes around 15 years on average for firms to receive the shock.

The relative masses of the two groups of firms in steady state are chosen to target the fraction of firms who are small in the data, which is 95.6%. Total entry is picked to normalise the total mass of firms to one. Letting $j = \{N, G\}$, this is achieved by picking the values of a_j in (22). The sensitivities of entry to firm value, the b_j, are chosen so that the entry of normal firms does not react to shocks ($b_N = 0$) and I set $b_G = 10$ for growth firms.

The relative productivity of growth firms, $z^{G,0}/z^N$ is chosen to match the share of employment in small firms, which is 29.3% in the data. The exogenous net worth injection, n_e, given to entering firms is chosen to match the total employment in startup firms, which is 2.9% of total employment in the BDS data. The exit rate, σ, is set so that firms exit every 10 years on average, in order to match the average 10% exit rate from the BDS data.

Turning to the transitory component, this is calibrated to within-firm data. 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. I take the autocorrelation of firm level shocks from Khan and Thomas (2013) as 0.65.

4.2 Performance at replicating firm distribution

The calibration matches the firm distribution quite well, as shown in Table 5. The fraction of small versus large firms, and employment in small versus large firms, is matched almost exactly. For the firm age distribution, the employment in and number of entrant firms is matched almost exactly. Moving up the firm age distribution, the model slightly over-predicts the mass of young and middle aged firms, and under-predicts the mass of old firms, because it applies a constant exit rate to all firms. In the data younger firms exit with a higher rate, which explains this discrepancy. The model matches the higher fraction of total employment in older firms relative to younger firms. It does not match the exact numbers yet, which is hard because in the data a small number of extremely large, older firms account for an inordinate share of employment. This is a challenging thing to match numerically.

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 B 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.

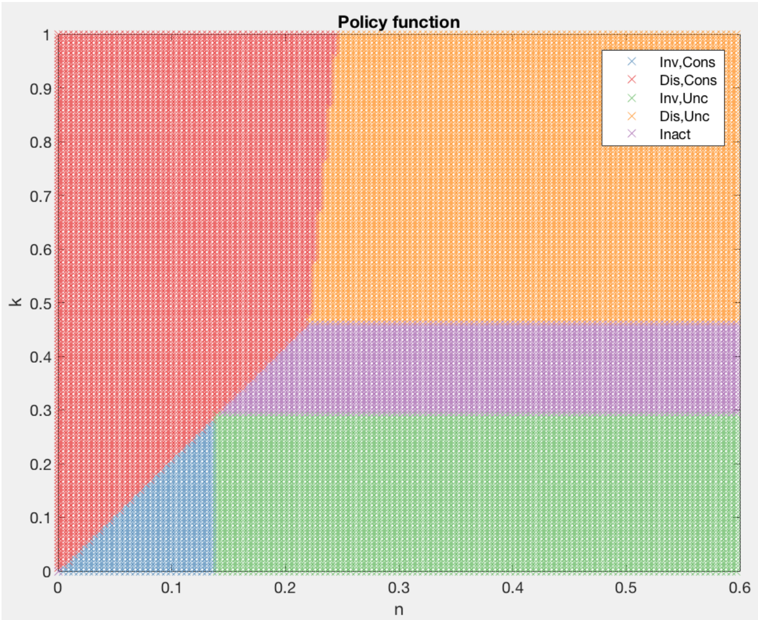
Simulation and calibration of the steady state requires calculating the distribution of (n_t, k_{t-1}, z_t) across firms. I do this using a non-stochastic simulation procedure, first proposed by Young (2009). These algorithms discretise the states (n_t, k_{t-1}, z_t), and and define the distribution 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 C.

4.4 Steady state policy functions and ergodic distribution

Figure 7 plots a top-down view of the capital function in the steady state. It illustrates the different regions, and shows how richer firms are financially unconstrained and can afford the unconstrained optimal choices of capital. They thus choose between investing, disinvesting, and inacting. Poorer firms are financially constrained, and are either investing or disinvesting.

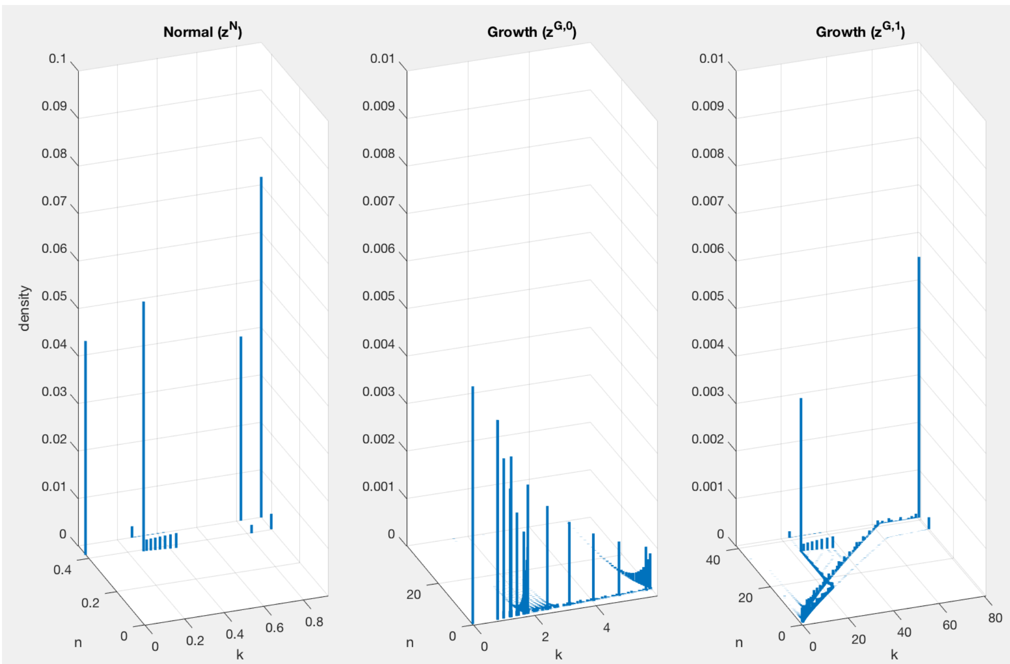
Figure 8 plots the ergodic distribution. The key moments are illustrated in Table 5 but this figure illustrates the challenges in solving this model. Normal firms and larger growth firms are nearly two orders of magnitude different in size as measured by capital and net worth. This necessitates the use of wide grids, and I choose grid points endogenously to make sure they are spread efficiently across this wide range.

Figure 7: Steady state policy function



Top-down view of steady state policy function regions. Plotted across net worth and capital for a particular level of productivity.

Figure 8: Ergodic distribution



Ergodic distribution across net worth and capital for different firm types.

5 Results: away from the ZLB

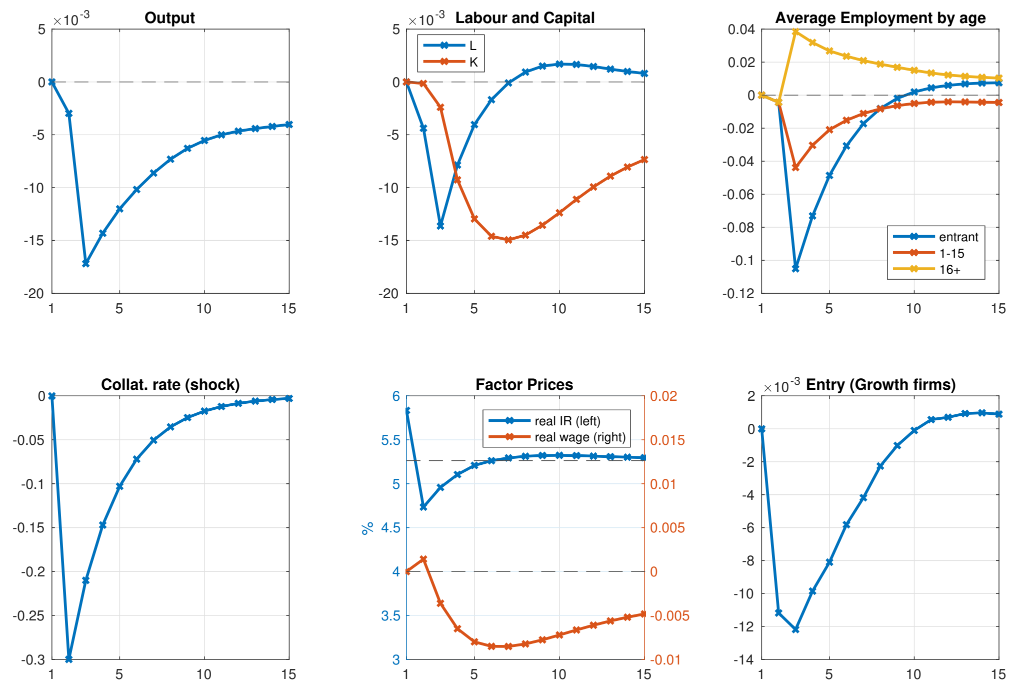
My first set of results concerns the behaviour of the model ignoring the ZLB constraint. This allows me to build intuitions in a simple way, before moving on to the ZLB in the next section. This will illustrate the *real wage channel* while leaving the real interest rate free to adjust.

5.1 Experiment 1: Financial shock, constant inflation

I model a financial crisis as an unexpected temporary deleveraging shock. In particular, at time $t = 1$ the economy is at steady state, with the collateral constraint equal to λ_{ss}. At time $t = 2$ the economy is unexpectedly hit by a shock which reduces λ_2 by 30% relative to steady state. λ_t then mean reverts back to steady state according to the AR(1) transition $\lambda_t = (1 - \rho_l)\lambda_{ss} + \rho_l \lambda_{t-1}$ with $\rho_l = 0.7$.

In order to focus first on how the model behaves in response to a pure financial crisis, I turn off all nominal features in the model. The ZLB will not bind for this shock, and I turn off downwards nominal wage rigidity by assuming that $\gamma = 0$ in the wage rule (34). Finally, I assume that the central bank is a strict inflation targeter who sets $\pi_t = \pi^*$ every period, where $\pi^* = 2\%$ is the inflation target. The central bank chooses the nominal interest rate each period to achieve this target.

Figure 9: Financial shock (with constant inflation: $\pi_t = \pi^*$)



Response of economy to unanticipated collateral constraint shock (bottom left panel). Average employment by firm age is calculated as average employment within firm-age bins, as in the BDS data.

Figure 9 plots key series from this experiment. With inflation always constant, and the DNWR not binding by construction, this model is in fact identical in spirit to the "real" model of Khan and Thomas (2013), just with my new calibration for the firm distribution.

The shock is plotted in the bottom left panel. In the top-left panel I plot output, which declines by 1.7% at its peak before gradually recovering. In the top middle panel I plot labour and capital. Labour initially declines by 1.5%, before recovering. Capital smoothly declines, and remains depressed even after labour has recovered.

Moving on to the firm distribution, in the top right panel I plot the responses of firms by firm age. Here I exactly replicate the empirical method I carried out on the BDS data in the empirical section. I group the firms by age and make repeated cross sections, I then plot the deviation of average employment within each bin from its steady state level across three age groups. The employment within startup firms is hardest hit, falling by 10% before recovering.

Young and Middle Aged firms (age 1-15 years) shrink by less, at around 4%. However, the big problem for this model is clearly the behaviour of old firms, here measured as 16+ years old. Rather than shrinking during the crisis, these firms actually expand by 4%. This stands in stark contrast to the BDS data, which showed that even old firms contracted during the financial crisis.

What is the source of this behaviour? The answer lies in factor prices. Old firms in this model are typically financially unconstrained, and hence unaffected by the financial shock directly. They do respond to factor prices, however. As can be seen from the bottom centre panel, both real wages and interest rates fall during the crisis (since inflation is constant, the nominal rate is equal to the real rate). This occurs because younger firms are unable to purchase capital and labour following the tightening of constraints, reducing their demand. To clear these markets, factor prices must fall. Older firms respond to this by increasing employment and purchasing more capital. The ex-post real interest in period 1 (between $t = 1$ and 2) rises due due the long-term debt structure, which slightly amplifies the crisis by redistributing resources away from constrained firms.

The expansion of old firms is a very robust feature of real financial shock models. Market clearing will always reduce factor prices following a financial shock, and unconstrained firms will always try to take advantage of this. This limits these models' ability to fully account for the Great Recession, where we see the opposite.

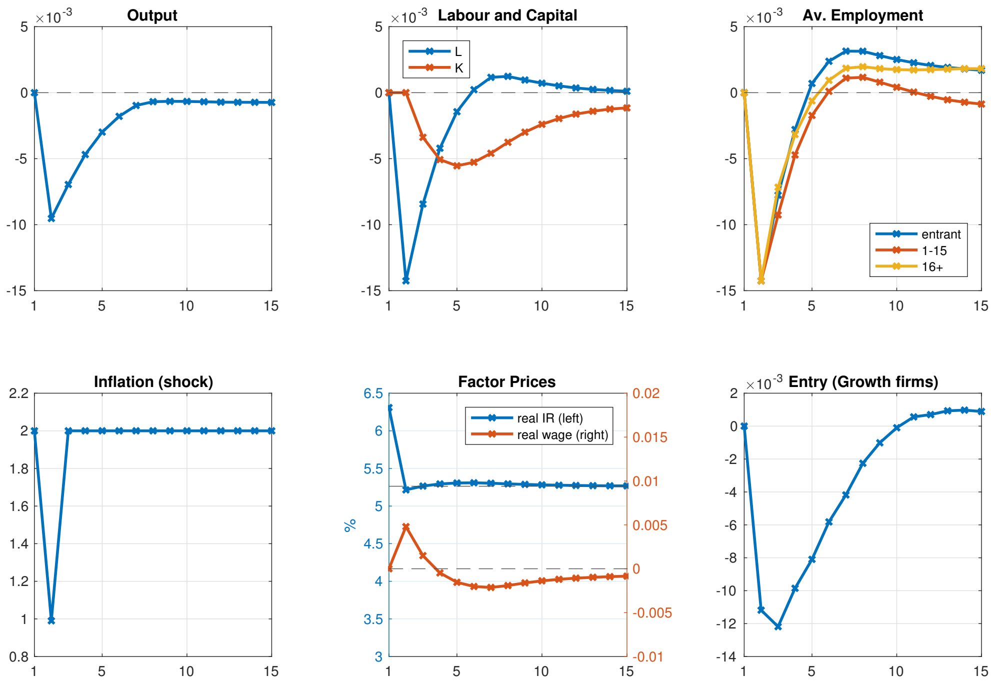
5.2 Experiment 2: Demand shock only

The next experiment is a pure monetary shock. In this experiment, I suppose that at $t = 2$ inflation temporarily drops from the inflation target to $\pi_2 = 1\%$ for one period only. I now impose downwards nominal wage rigidity by setting $\gamma = 0.6$ in the wage rule (34), implying that real wages move 40% of the way to their market clearing value within one year. The nominal wage norm is set to the inflation target, $\bar{\pi} = \pi^* = 2\%$ so that any decline in inflation below target will case DNWR to bind. I do not model here why inflation falls below target, but just study the effects of this happening. In the next section, inflation will drop because the central bank loses control of inflation at the ZLB.

Figure 10 shows the results. In the bottom left panel I plot the inflation series. The dip in inflation triggers DNWR to bind, pushing the real wage up for two periods. This triggers a recession, and in the top left panel we see that output falls by 1% on impact. Since the recession is now driven by a disruption in the labour market, the recession is now much more driven by labour than by capital, as seen in the top centre panel.

The responses across the firm age distribution are vastly different to the responses to the financial shock. In the top right panel we now see that all firms ages shrink their average employment regardless of their age. The results are identical on impact, and are slightly more persistent for Young and Middle Aged firms since the shock depletes their net worth which impedes their ability to invest. Thus we see that the demand shock is able to generate responses across the firm age distribution much more in line with the data, where we saw that even Old firms shrank their average employment. Given the failure of the financial shock alone to match this pattern, this suggests that incorporating demand features is one way to improve the model's performance.

Figure 10: Inflation shock (with constant collateral constraint: $\lambda_t = \lambda_{ss}$)



Response of economy to unanticipated inflation shock (bottom left panel). Average employment by firm age is calculated as average employment within firm-age bins, as in the BDS data.

The intuition for why the demand shock affects all firms, including Old firms, is very simple: This is an aggregate shock, rather than the financial shock which was targeted at younger firms. The rise in the real wage reduces the incentive to hire labour, causing all firms to shrink. In terms of entry, in the bottom right panel we see that the model also predicts a decline in entry in response to the demand shock, since the increased real wage reduces firm values.

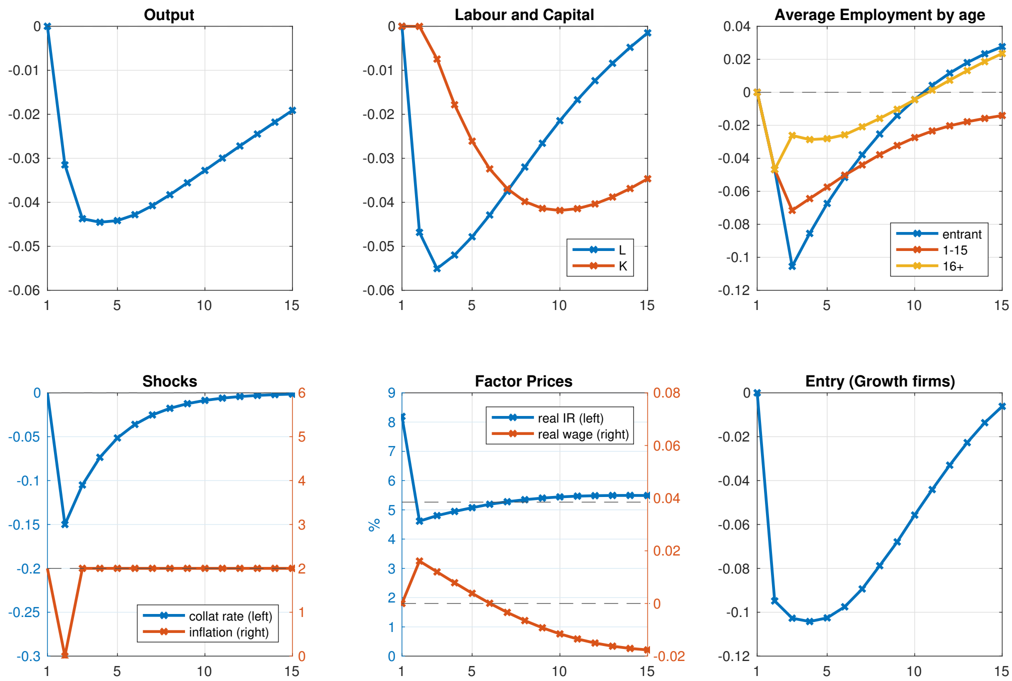
5.3 Experiment 3: Combined financial and demand shock

Finally, I investigate the economy's response to a combined financial and monetary shock. I subject the economy to a financial shock half the size of that in the first experiment (λ_2 drops 15% from steady state) and increase the drop in inflation to 2% while also increasing the degree of wage stickiness by setting $\gamma = 0.9$.

I plot the results in Figure 11. Compared to the results of the pure financial shock, the response of Old firms to the crisis is now improved. Rather than expanding, they now shrink, along with the younger firms. Why do older firms now contract? Without the decline in wages from the "real" model, there is less incentive for them to expand. Additionally, the recession itself reduces the demand these firms face (recall the demand externality in firms' demand curves). These effects combined encourage old firms to contract.

Finally, it is interesting to see how this difference in disaggregated responses adds up to to the aggregate responses. Since old firms are now also contracting, the effect of the financial crisis is more severe. Similarly, the response of labour is now much more pronounced. With DNWR active there is now equilibrium unemployment. Thus, we see that the additional of DNWR actually improves the model in two complementary ways. Firstly, it corrects the response of old firms to the crisis, and encourages them to shrink their employment too. Secondly, this adds up to a larger total response of employment to the crisis, allowing the model to better match the large reduction in aggregate labour.

Figure 11: Combined financial and inflation shocks



Response of economy to unanticipated financial and inflation shocks (bottom left panel). Average employment by firm age is calculated as average employment within firm-age bins, as in the BDS data.

6 Results: ZLB

Having presented the results away from the ZLB illustrating the real wage channel, I now move on to incorporating the ZLB and analysing the *real interest rate channel.* Solving the model at the ZLB presents several complications. I solve the model fully nonlinearly, and impose the ZLB explicitly as an occasionally binding constraint on the system.

6.1 Setup

One important feature of the recession that I demonstrated in the data was that the spread on safe AAA firm debt increased during the crisis relative to safe long term treasury debt. This turns out to be an important feature that needs to be incorporated in order for the ZLB to bind in this model. This spread reflects the increased preference for liquidity during the crisis, since both forms of debt are free from default risk. I incorporate this by introducing exogenous wedges in the household's Euler equations. Households now price firm debt according to:

$$1 = \beta \tau_t^f \frac{u_{c,t+1}}{u_{c,t}} i_t \pi_{t+1}^{-1} \tag{41}$$

Government bonds are priced according to:

$$1 = \beta \tau_t^c \frac{u_{c,t+1}}{u_{c,t}} i_t^c \pi_{t+1}^{-1} \tag{42}$$

where I use i_t^c to refer to a liquid government bond and i_t to refer to a firm bond. The wedges τ_t^f and τ_t^c capture non-modelled features which drive the household's preference for each type of bond away from the simple Euler equation. For example, if $\tau_t^c > \tau_t^f$ then households prefer (all else equal) to hold government bonds than firm bonds. This could be because we imagine them as being less liquid, and hence we can think of this wedges as liquidity preference shocks. Dividing the two we see that

$$i_t = \frac{\tau_t^c}{\tau_t^f} i_t^c \tag{43}$$

and so $\frac{\tau_t^c}{\tau_t^f}$ captures exactly the spread between these two bonds in the model. This will be important in driving the rate on government bonds to the ZLB even as firm rates do not fall as far, as we saw in the data.

I model the nominal features of the experiment after Rendahl (2016). In particular, I assume an extreme form of DNWR in (34) where $\gamma = 1$ and $\bar{\pi} = \pi^* = 2\%$. That is, the nominal wage is fully downwardly rigid, in the sense that it must grow at at least the inflation target regardless of the market clearing real wage:

$$w_t \geq \frac{\pi^*}{\pi_t} w_{t-1} \tag{44}$$

The central bank is assumed to choose interest rates inflation in order to target the labour market. Rather than being a pure inflation targetter, it chooses inflation to try and maintain the wage at the market clearing level, $w_t = w_t^{mc}$, which ensures zero unemployment in this model. Given the wage rule above, and assuming the central bank chooses the minimum value of inflation required to satisfy the inequality, the central bank will (whenever possible) set inflation to

$$w_t = w_t^{mc} \implies \pi_t = \pi^* i \frac{w_{t-1}}{w_t^{mc}} \tag{45}$$

This rule can be interpreted as the central bank aiming to achieve long-run inflation of π^*, but prioritising minimising unemployment fluctuations over inflation stability in the short run. However, if the ZLB binds the central bank might not be able to achieve this target, and inflation could fall below this values leading DNWR to bind.

I recalibrate the value of β in this version of the model to reflect the wedges in the Euler equation. The changes are chosen to not affect the steady state of the model. I assume that pre-crisis there is no wedge in the Euler equation for government bonds ($\tau_{ss}^c = 1$). I target a 5% nominal interest on government bonds pre-crisis, as in the data. Given an inflation target of 2% and $\tau_{ss}^c = 0$ this implies a value of β of $\beta = 1.02/1.05 = 0.9714$. To maintain the same steady state interest rate on firm debt as in the original calibration, I choose the steady state value of τ_{ss}^f to keep i_t at its original value. Letting $\beta_0 = 0.95$ denote the discount rate in the original calibration, this gives a value of $\tau_{ss}^f = 1/\beta \times \beta_0 = 0.95/0.9714 = 0.9780$.

A peculiar feature of the functional form assumptions made in this paper is that if DNWR is not binding, then the nominal interest rate will always be constant in equilibrium, regardless of the shocks hitting the model, and the behaviour of the real interest rate or inflation. To see this, note that I assumed log utility from consumption and linear utility from labour. Plugging these and optimal inflation from (45) into (42) gives

$$1 = \beta \tau_t^c \frac{c_t}{c_{t+1}} i_t^c \frac{w_{t+1}^{cm}}{\pi^* w_t^{mc}} = \beta \tau_t^c \frac{c_t}{c_{t+1}} i_t^c \frac{Dc_{t+1}}{\pi^* Dc_t} \implies i_t^c = \frac{1}{\beta \tau_t^c} \tag{46}$$

Thus, regardless of the shocks hitting the economy, the central bank (if it can) always moves inflation to exactly offset changes in the real interest rate, leading to a constant nominal interest rate. The exception is shocks to the liquidity preference for government bonds, τ_t^c, which we see above directly move the nominal interest rate. This result is a peculiar (and unexpected!) feature of the current assumptions on preferences and central bank behaviour, and will be relaxed in future versions of this draft.

However, one implication of this assumption is that we can very easily calculate the size of the demand (liquidity) shock required to take us to the ZLB directly from (46): any shock such that $\tau_t^c \geq 1/\beta$ would violate this condition. When this happens, the central bank is no longer able to achieve $w_t = w_t^{mc}$ and instead inflation this period falls to drive up the real wage and restore equilibrium by reducing current consumption so that $i_t^c = 1$ in equilibrium.

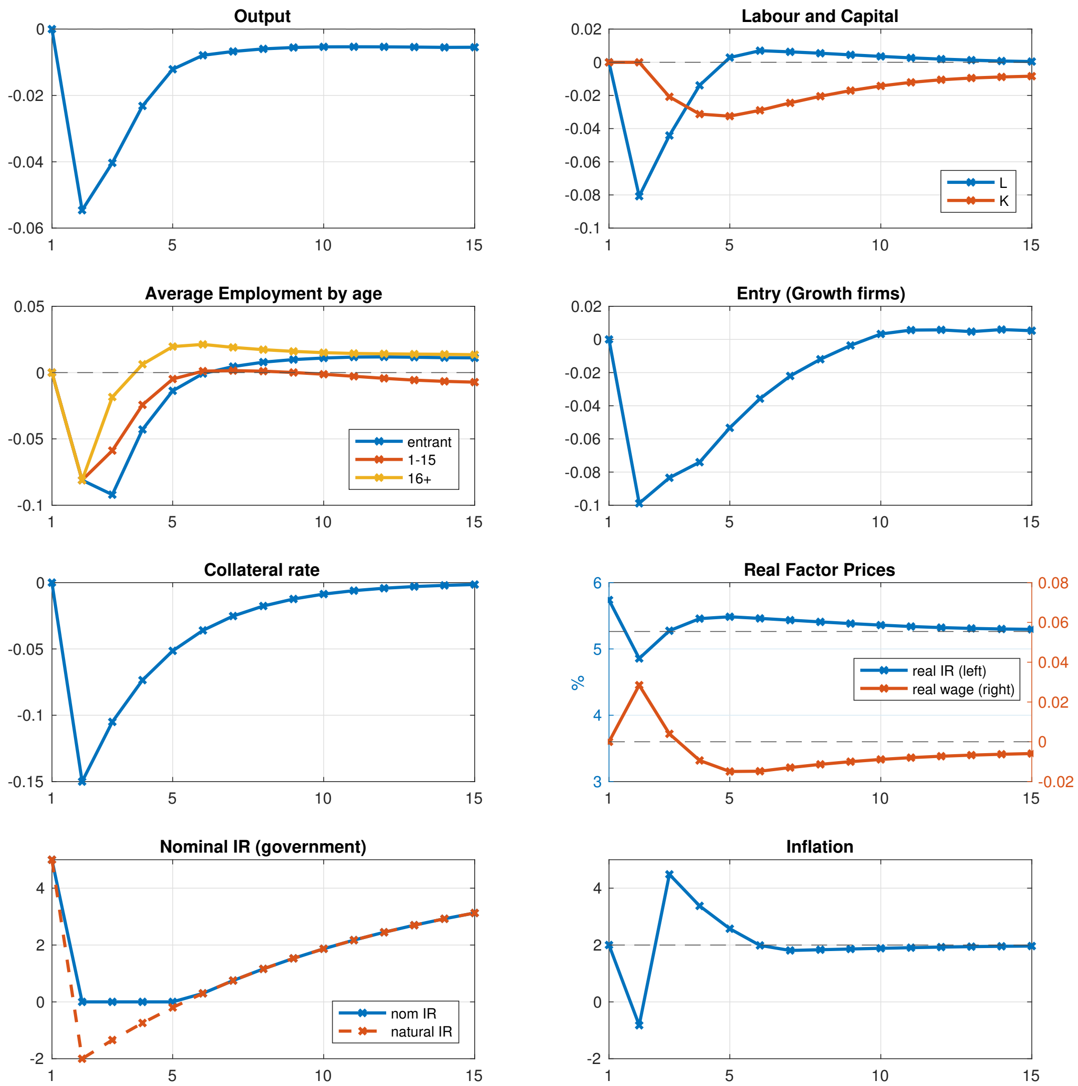
6.2 Experiment

I consider a baseline experiment which combines a financial shock with a demand shock which drives the economy to the ZLB. I then consider each of the two shocks in isolation to understand their contributions to the baseline dynamics. So far, the shocks are chosen for illustrative purposes only, to roughly capture behaviour during the crisis.

In this experiment I again consider a temporary deleveraging. In particular, I shock the economy by reducing the collateral rate to $\lambda_2 = 0.85 \times \lambda_{ss}$, and then allow it to recover according to $\lambda_t = (1 - \rho_l)\lambda_{ss} + \rho_l \lambda_{t-1}$ with $\rho_l = 0.7$.

I maintain the firm Euler wedge at its steady state value: $\tau_t^f = \tau_{ss}^f$. To generate the demand shock, I shock the liquidity preference in the government-bond Euler. In particular I set $\tau_2^c = 1.0714$ at then let it mean revert back to $\tau_{ss}^c = 1$ according to $\tau_t^c = (1-\rho_c)\tau_{ss}^c + \rho_c \tau_{t-1}^c$ with $\rho_c = 0.9$. Before the economy adjusts to restore equilibrium, this shock would imply a (natural) nominal interest rate equal to -2% on impact, and which stays below the ZLB for four years.

Figure 12: Combined financial and liquidity preference shocks (ZLB binds)



Response of economy to unanticipated financial shock and liquidity preference shocks. Average employment by firm age is calculated as average employment within firm-age bins, as in the BDS data.

Figure 12 plots the results. Counting from top to bottom and left to right, the shock to the collateral rate is shown in the fifth panel. The liquidity shock can be calculated from the natural nominal interest rate in the seventh figure using $\tau_t^c = 1/(\beta i_t^{c,nat})$. The two shocks combined lead to an output fall of around 5%. The economy is driven to the ZLB, as shown in the seventh panel, with the equilibrium nominal (government) rate above the natural rate for four periods.

Once the nominal interest rate hits the ZLB, the economy adjusts to restore equilibrium using the inflation rate, driving it away from the central bank's preferred value. The final panel plots the inflation rate, and we see inflation falling below the long run level in period 2. Due to DNWR this drives up the real wage, which rises by 4% on impact. From time 3 onwards we see that the model predicts inflation above the long run value, despite being at the ZLB. This inflation is needed to reduce the real wage, given the strong assumption of complete DNWR. However, while not predicting continued disinflation, the ZLB still implies that inflation is lower than the amount needed to clear the labour market, and the real wage remains artificially elevated during this period.

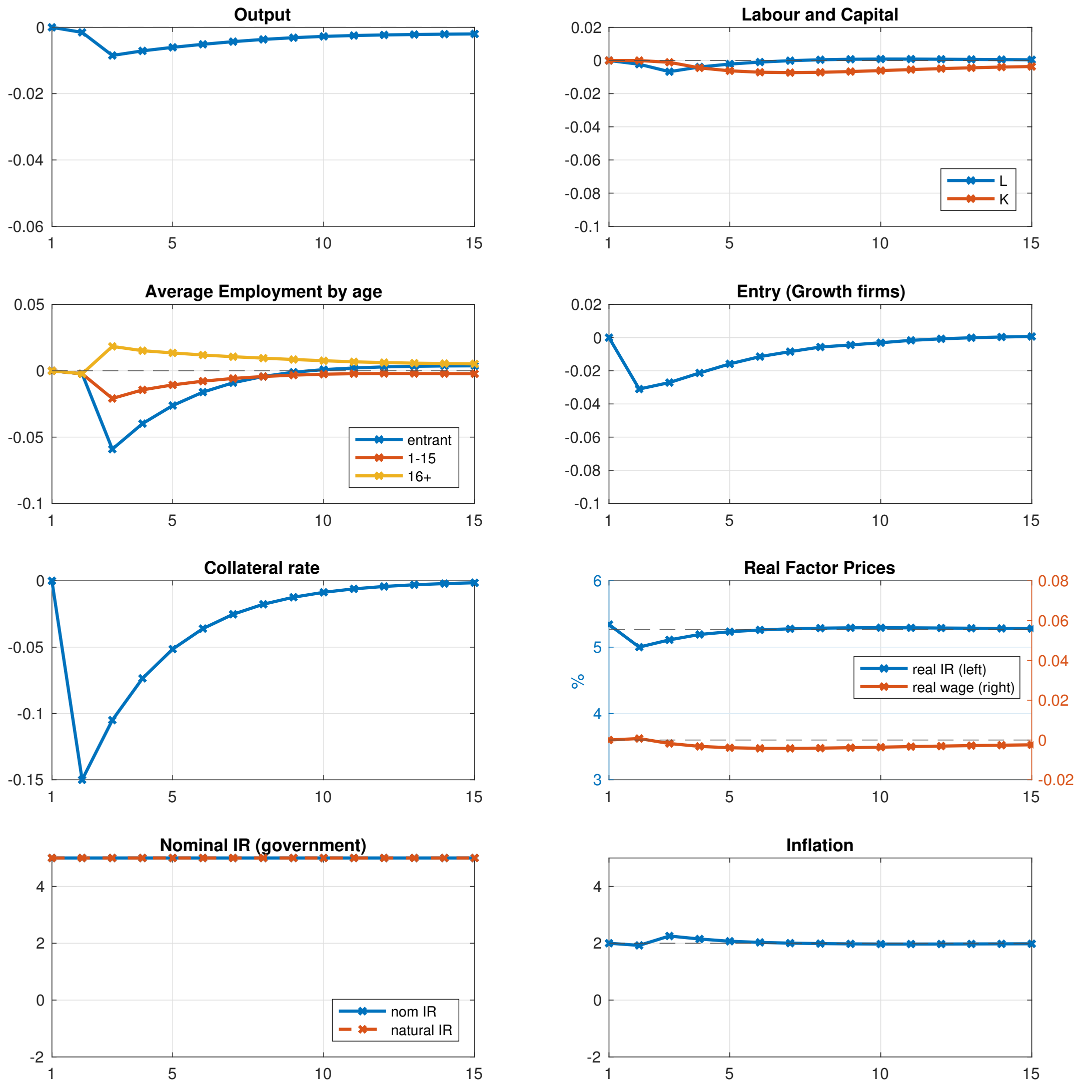
The (ex-post) real interest rate, shown in the fifth panel, spikes up in the first period reflecting the long-term debt assumption. In the period of the shock, the real interest rate falls, reflecting the models attempt to clear the capital market following the disinvestment of financially constrained firms caused by the financial shock. However, note that the real interest rate remains elevated above the steady state value during much of the sample. This reflects the real interest rate channel, and will be discussed further below.

In terms of behaviour at the firm level, we see that the combination of the two shocks allows the model to match the qualitative features of the data well. As in the simple experiments of the last section, this reflects the better properties of the real factor prices when demand shocks are included, which improve the behaviour of Old, financially unconstrained firms. In the third panel, we see that firms of all ages shrink their employment in the first two years of the crisis, although older firms do eventually expand once the ZLB episode passes. Young firms shrink their employment more than old firms, in line with the data, and entry of Growth firms falls by 10%.

Two failings of the model here should be noted. Firstly, the average employment of startup firms falls during the crisis, while in the data it rose. This presumably reflects the model's lack of a selection margin at entry. Secondly, the decline in entry is shown only for growth firms, who make up only 10% of entrants in my model. I assume that Normal firms have a constant entry rate, so the combined effect here is only a 1% drop in total entry, compared to the 25% in the data. The entry process assumed in this version of the model is stylised, and this failure relative to the data needs to be addressed in future versions.

In Figures 13 and 14 I plot the response of the economy to only the financial shock and only the liquidity preference shock respectively. The scales in the axis are held constant across the three figures for comparability. This comparison tells us several things. Firstly, it reveals the importance of the real interest rate channel. In Figure 13 we see as before that a financial shock alone incorrectly predicts expansion of average employment at Old firms. This is driven by factor prices: the aforementioned real wage channel, and additionally the fact that the real interest rate is free to fall to encourage Old firms to pick up capital shed by financially constrained firms. Comparing the real interest rate here vs the model with both shocks, we see that the real interest rate is more elevated in the model with both shocks. This shows how the ZLB, by preventing the nominal interest rate from falling, elevates the real interest rate and thus reduces the incentive of Old firms to pick up capital shed by the Young. Looking at Figure 14 we see that the elevated real interest rate is indeed inherited from the demand shock.

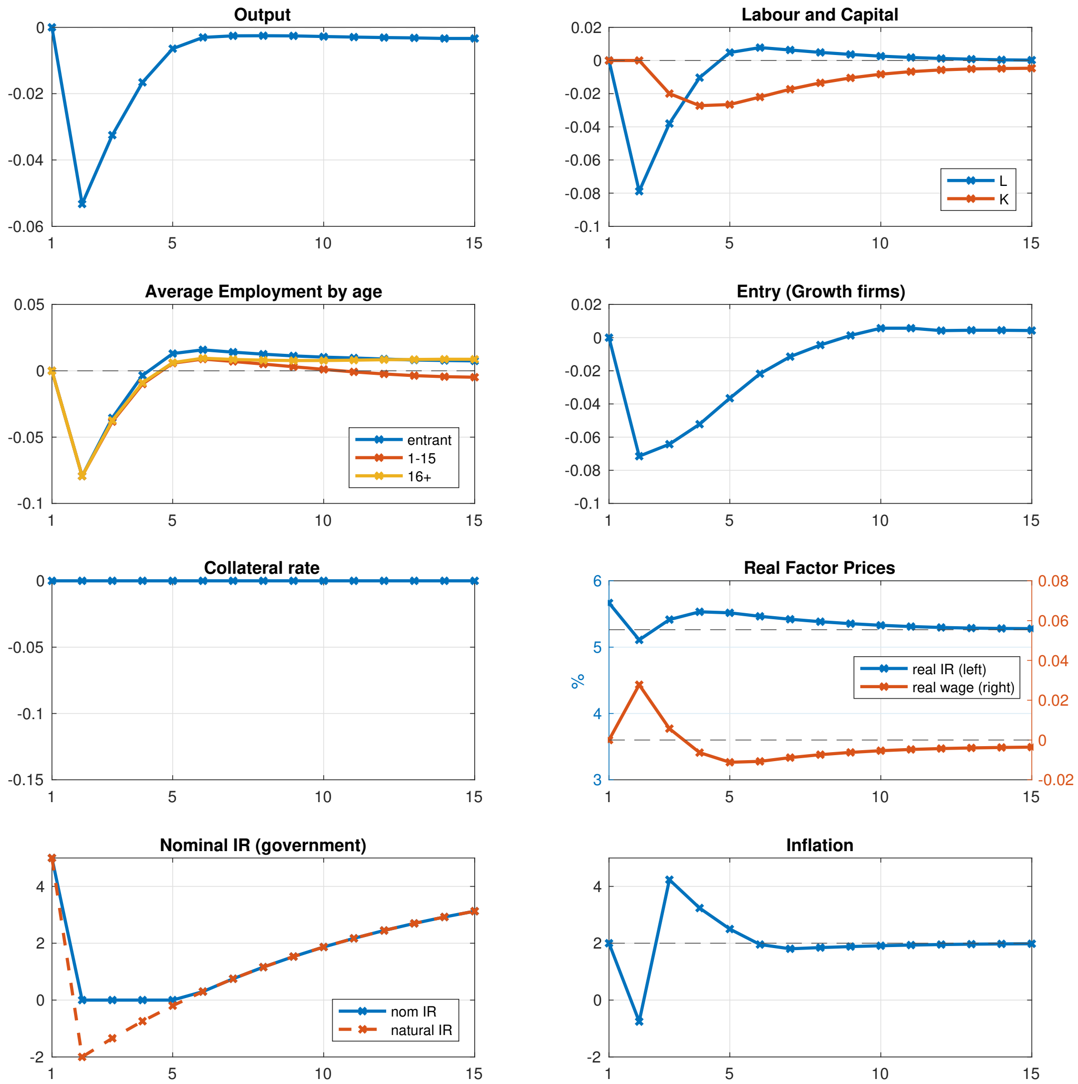
Figure 13: Financial shock only (ZLB doesn't bind)



Response of economy to unanticipated financial shock. Average employment by firm age is calculated as average employment within firm-age bins, as in the BDS data.

Secondly, the decomposition reveals the different roles that the two shocks are playing in matching the data. With the caveat that these results are only illustrative, they capture in an intuitive way how financial and demand shocks are both needed to capture the key features of the BDS data shown at the beginning of the paper. Firstly, the financial shock alone does a poor job of generating a large enough recession. This is despite the decline in employment at entrant and young firms, and is due to the offsetting effect of Old firms, who expand during the crisis, and the small initial employment

Figure 14: Liquidity preference shock only (ZLB binds)



Response of economy to unanticipated liquidity preference shock. Average employment by firm age is calculated as average employment within firm-age bins, as in the BDS data.

share of younger firms. The demand shock does a better job of capturing the large total employment fall, because firms of all ages decline. This maps nicely into the BDS data, which showed that the decline in average employment at old firms is important for matching the overall employment decline due to their large share.

However, the decomposition also shows that financial shocks play a role in matching the data by firm age. In the data, younger firms shrank their employment by more than older firms, which is something that is missing from the economy's response to a demand shock only, but that the financial shock generates.

The results of this section are subject to the caveat that the size of each shock was not chosen to be disciplined by some data. Hence the financial shock could be found to be more powerful at generating an overall employment decline if it was chosen to be larger, for example, or if entry fell more during the crisis. Hence, an important next step is to discipline the size of these shocks in a reasonable way to better assess the role of each shock in driving different parts of the data.

6.3 Fiscal policy at the ZLB (VERY PRELIMINARY)

Note: The results from this section are from an old calibration of the model featuring Growth firms only.

This model provides an interesting laboratory for thinking about fiscal policy at the ZLB. The presence of a rich firm distribution suggests that the effectiveness of fiscal policy will depend on the state of all firms, and how the policy is transmitted through the distribution. In this section I consider two simple experiments.

Firstly, I consider the simple government spending multiplier. To make things stark, I suppose a one-period increase in government spending during the period the ZLB binds. As is well known, the assumption of a one-period rise in government spending means that the government spending multiplier ($\Delta Y/\Delta G$) is typically exactly one: government spending does not crowd out investment or consumption, but does not encourage it either. It is possible to prove this is also true in my model, if firms face no financial frictions.

However, if firms do face financial frictions then the government multiplier can be greater than one even if spending is only increased for a single period. In the scenario above, I calculate the multiplier numerically to a wasteful fiscal stimulus of 0.5% of steady-state GDP. The multiplier is 1.28, significantly greater than one. This operates through financial frictions, and in particular through stimulating investment. The increase in government spending increases output, which increases firms' profits. This relaxing their financial frictions during the deleveraging crisis, and allows them to increase investment. In fact, almost the entire of the increase in the multiplier above one is driven by investment, rather than consumption. Additionally, the size of this multiplier is state dependent. Since it operates through financial frictions, the more these frictions bind the stronger is the multiplier. Comparing the multiplier at the depth of a financial recession to the multiplier in a recession driven by demand only, I find it to be 8% higher during a financial recession, when more firms are financially constrained.

Secondly, I consider a more targeted fiscal policy, which aims to address the source of the financial crisis. I consider fiscal transfers to firms by the government, funded by lump sum taxes on the household. I am interested in how the power of these transfers depends on the firms they are targeted towards, and hence consider a policy that directs the money only to firms of a certain age. In each experiment, the government gives a total transfer of 0.5% of steady state GDP to firms of a given age, split among the firms in proportion to their current net worth.

The results are given in Table 6, which gives the transfer multiplier across transfers to different firm ages. This is meant to approximate funding schemes aimed at helping young firms deal with the financial crisis. The results are very surprising. One might have expected the transfer to get more effective the younger the targeted firms are, since these firms are more likely to be financially constrained. However, there is actually a hump shaped response in the multiplier, such that targeting

Table 6: Fiscal transfers by firm age

Age:	-1	0	1	2	3	4	5	6+
$\Delta Y/\Delta T$:	0.08	0.3	0.48	0.62	0.72	1.34	1.4	0.9

the money to young *but not too young* firms is the most effective. Transfers to firms of age 5 give a multiplier of 1.4, while transfers to startups give a multiplier of 0.08 and transfers to old firms (6+ years) give a multiplier of 0.9.

Why is this? It turns out that the response of old firms to the transfer is crucial for the aggregate power of the transfer. Transferring money to financially constrained firms will naturally increase their investment, and hence boost their output, but the aggregate effect depends on what older firms do too. Since this is a firm lifecycle model, and firms are forward looking, it turns out that older firms are not always happy about these transfers being given to young firms. On the contrary, the transfers fund their future competitors. Hence old firms might choose to shrink their production in response to other firms being given transfers, as they anticipate being crowded out in the future and preemptively reduce their size. This is especially true when given transfers to startups, who can leverage this benefit over many periods to grow quickly to their optimal size. Giving the money to slightly older firms gives the same benefits of increasing output today, while scaring the older firms less, since these firms were nearly at their optimal size already and hence do not get as big a boost in future periods from the transfer.

These results highlight the benefits that using a heterogeneous firm model can bring to the question of fiscal policy at the ZLB. Not only can we study how fiscal policy is propagated through the firm distribution, but it also allows for discussion of targeted fiscal policies aimed at firms of different ages.

7 Conclusion

In this paper I developed a quantitative heterogeneous-firm model with both financial and nominal frictions and use it to analyse the recent recession through the lens of firm-level data. The financial crisis was notable for many reasons, two of them being 1) the remarkable collapse of firm entry, and hence total employment, at younger firms, and 2) the extent to which the crisis disrupted the labour market, reducing aggregate employment. I ask to what extent these phenomena are linked, by decomposing the decline in total employment along the firm age distribution.

I use Business Dynamics Statistics (BDS) data, and first show that, despite providing only 14% of total employment pre-crisis, 40% of the *decline* in aggregate employment during the recession can be accounted for by the *decline* in employment in young firms (age 0-5). While this suggests the important role that young firms, and hence the shocks that particularly affect them, played during the crisis, it still leaves 60% of the aggregate employment decline to be explained by old firms (age 6+). Decomposing the changes into changes in average employment within firms, and changes in the number of firms, the larger decline in total employment in young firms turns out to be driven almost entirely by the drop in the number of young firms relative to old firms, and not by a larger decrease in average employment within young firms. In particular, I find that most of the aggregate decline in employment can be accounted for by (1) a decline in the *number* of young firms by 25%, and (2) a decline in employment *within* old firms of 7.5%.

The fact that older firms shrank so drastically during the crisis presents an interesting challenge to the idea that the crisis was primarily a financial shock which impacted the economy by misallocating resources across firms. Indeed, if this was the case we would expect to see resources being reallocated from poor to rich firms. If we are willing to equate young with poor and old with rich, we do not see that at all: instead, firms of all ages shrink, and the recession is not one of mis- or even re-allocation, but just one of general decline.

I then turn to modelling the crisis, and use this data to ask what kinds of models and shocks are needed to replicate the crisis. I argue that models of the crisis which focus on financial shocks alone, such as Khan and Thomas (2013), while they can explain the behaviour of young firms, cannot explain the behaviour of old firms. Indeed, the key mechanism in these models is that old firms typically expand during a crisis, because factor prices fall and they are thus incentivised to grow and absorb the labour and capital shed by financially constrained firms.

Looking at data on factor prices, I argue that we do not see the decline predicted by these models, and propose adding nominal features to the models to arrest this decline. I do this by extending Khan and Thomas' (2013) model to incorporate downwards nominal wage rigidity (DNWR) and the zero lower bound (ZLB). I show how simply correcting the behaviour of factor prices in the model can simultaneously correct the model's implications for how a financial crisis affects the firm age distribution, and importantly change the aggregate implications of a financial shock.

References

[1] 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.

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

[3] 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.

[4] 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.

[5] Christiano, L. J., M. S. Eichenbaum, and M. Trabandt, 2015. *Understanding the Great Recession.* American Economic Journal: Macroeconomics, 7, 110?167.

[6] 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.

[7] Del Negro, Marco, Domenico Giannone, Marc P. Giannoni, and Andrea Tambalotti, 2017. *Safety, Liquidity, and the Natural Rate of Interest.* Federal Reserve Bank of New York Staff Reports 812, May 2017.

[8] Dyrda, Sebastian, 2016. *Fluctuations in Uncertainty, Efficient Borrowing Constraints and Firm Dynamics.* Mimeo.

[9] Eggertsson, Gauti, and Michael Woodford, 2003. *The Zero Bound on Interest Rates and Optimal Monetary Policy.* Brookings Papers on Economic Activity, Economic Studies Program, The Brookings Institution, vol. 34(1), pages 139-235.

[10] Eggertsson, Gauti B., Neil R. Mehrotra, and Jacob A. Robbins, 2017. *A Model of Secular Stagnation: Theory and Quantitative Evaluation.* NBER Working Papers 23093, National Bureau of Economic Research, Inc.

[11] Foster, L., Haltiwanger, J. and Syverson, C., 2016 *The Slow Growth of New Plants: Learning about Demand?* Economica, 83: 91?129.

Gilchrist, Simon, Raphael Schoenle, Jae Sim, and Egon Zakrajsek, 2017. *Inflation Dynamics during the Financial Crisis.* American Economics Review, Vol. 107, No. 3, March 2017, pp. 785-823.

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

[13] Gomes, João, Urban Jermann, and Lukas Schmid, 2016. *Sticky Leverage.* American Economic Review, Vol. 106, No. 12, December 2016.

[14] 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.

[15] Jo, In Hwan and Tatsuro Senga, 2017. *Aggregate Consequences of Credit Subsidy Policies: Firm Dynamics and Misallocation.* Mimeo.

[16] Khan, Aubhik, Julia K. Thomas, 2003. *Nonconvex Factor Adjustments in Equilibrium Business Cycle Models: Do Nonlinearities Matter?* Journal of Monetary Economics, March 2003.

[17] 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.

[18] 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.

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

[20] Lanteri, Andrea, 2017. *The Market for Used Capital: Endogenous Irreversibility and Reallocation over the Business Cycle.* Mimeo.

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

[22] Ottonello, Pablo and Thomas Winberry, 2017. *Financial Heterogeneity and the Investment Channel of Monetary Policy.* Mimeo.

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

[24] Rendahl, Pontus, 2016. *Fiscal Policy in an Unemployment Crisis.* Review of Economic Studies, vol 83 (3), 2016.

[25] Sepahsalari, Alireza, 2016. *Financial Market Imperfections andLabour Market Outcomes.* Mimeo.

[26] 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.

[27] Thomas, Julia, 2002. *Is Lumpy Investment Relevant for the Business Cycle?* Journal of Political Economy, June 2002.

[28] Winberry, Thomas, 2016. *Lumpy Investment, Business Cycles, and Stimulus Policy.* Mimeo.

Appendices

A Empirical appendix

A.1 Alternative definition of Old firms

In this section I redefine Old firms as being only firms aged 26 and above, and redefine Middle Aged firms as those aged 6 to 25. The idea is that we might worry that Old firms are not that old, and hence could plausible be financially constrained. Thus, by redefining Old firms as being 26 and above years old, we reduce the likelihood that they will be financially constrained.

The results are qualitatively similar to the results in the main text, and are even more supportive of the message of the paper. Defined in this way Young firms do not shrink their average employment any more than Middle Aged or Old firms. The Young (whose definition did not change) continue to shrink by 10.1%, while the Middle Aged and Old shrink by 10.3% and 11.4% respectively. Thus the results are robust to this alternative definition of firm age.

Table 7: Decomposition with Old firms defined as age 26+

Change:	All	Entrant	Young	Middle Age	Old
$\Delta(E/N)$	-2.0%	7.9%	-10.1%	-10.3%	-11.4%
ΔN	-5.5%	-26.7%	-11.9%	-2.1%	11.8%
ΔE	-7.4%	-20.8%	-20.8%	-12.1%	-1.0%
Contribution:	All	Entrant	Young	Middle Age	Old
$\Delta(E/N)$	27.4%	-2.7%	15.7%	45.1%	83.0%
ΔN	74.1%	9.2%	18.6%	9.0%	-85.5%
ΔE	100%	7.2%	32.3%	53.2%	7.3%

$\Delta E_i = \Delta(E_i/N_i)\Delta N_i$. Young = 1-5 years, Middle = 6-25 years, Old = 26+ years. Average employment calculated at bin level. "Contribution" refers to the percentage of the total employment decline ($\Delta E = -7.4\%$) explained by any one factor.

A.2 Firm responses by age in previous recessions

In this section I present data on employment by firm age during previous recessions covered in the BDS sample. The sample where we have enough data on firm ages to calculate the relevant bins also includes the 2001-2 recession, and the 1990-1 recession. Since all firms born before 1977 have no age classification and are included in the Left-Censored category, in order to be sure they are correctly placed in the oldest bin I use, in this section I have to combine the Middle Aged and Old bins into a single bin of firms aged six and above.

Comparing the 2007 recession and the previous two recessions, two interesting facts emerge. Firstly, the massive decline in entry is unique to the 2007 recession. The decline in the number of entrant firms is much larger in 2007, even controlling for the larger size of the recession. Secondly, the relative decline in average employment between Young and Old firms is quite stable across recessions. That is, Young firms always shrink more than Old firms during recessions, but this difference is not markedly pronounced during the 2007 recession. During the 1990-1 recession Young firms shrink

3.1pp more than Old firms, during the 2001-2 recession it was 4.3pp, and during the 2007 recession it was 3.7pp.

Thus, the reason that the employment decline during the 2007 recession was worse than the previous two recessions is because 1) entry fell more, and 2) all firms contracted their average employment more. Young firms did not shrink their average employment more than Old firms any more than in the previous recessions.

Table 8: Firm responses by age in previous recessions

2007-11:	All	Entrant	Young	MA + O
$\Delta(E/N)$	-2.0%	7.9%	-10.1%	-6.4%
ΔN	-5.5%	-26.7%	-11.9%	1.3%
ΔE	-7.4%	-20.8%	-20.8%	-5.2%
2001-2:	All	Entrant	Young	MA + O
$\Delta(E/N)$	-2.5%	5.6%	-6.5%	-2.2%
ΔN	0.6%	5.3%	-1.3%	0.7%
ΔE	-1.9%	11.2%	-7.6%	-1.4%
1990-1:	All	Entrant	Young	MA + O
$\Delta(E/N)$	-2.4%	-4.7%	-6.1%	-3.0%
ΔN	1.2%	-1.7%	-0.9%	3.2%
ΔE	-1.2%	-6.3%	-6.9%	0.1%

$\Delta E_i = \Delta(E_i/N_i)\Delta N_i$. Young = 1-5 years, Middle = 6-15 years, Old = 16+ years. Average employment calculated at bin level.

B Numerical model appendix

B.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 (19) 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 (20) ignoring the constraints on k_t.[9]

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

[9] 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 $k^{d,u}(n,k,z)$

computed the five policy functions I apply the following program to check which is optimal.[10]

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 $k^{i,u}$).

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.

B.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:

$$n_{t+1} = \pi(z_{t+1}, k_t^u(k_{t-1}, z_t)) + q^i(1-\delta)k^u(k_{t-1}, z_t) - rd_t \tag{47}$$

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

[10] 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.

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:

$$\underline{n}_t(k_{t-1}, z_t) = k_t^u(k_{t-1}, z_t) - \min \left\{ \bar{d}_t(k_{t-1}, z_t), \lambda q^d k_t^u(k_{t-1}, z_t) \right\} \tag{48}$$

If disinvesting:

$$\underline{n}_t(k_{t-1}, z_t) = q^d k_t^u(k_{t-1}, z_t) + (1 - q^d)(1 - \delta)k_{t-1} - \min \left\{ \bar{d}_t(k_{t-1}, z_t), \lambda q^d k_t^u(k_{t-1}, z_t) \right\} \tag{49}$$

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

B.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 $\mu_{t+1} = \Gamma_t(\mu_t)$.

B.4 Solving for inflation at the ZLB

How to update inflation when the ZLB is binding: The code simulates the economy forward, imposing market clearing. At time t, I solve for the values of Y_{t+1}, C_{t+1}, w_{t+1}, π_{t+1}, and i_t to clear the time $t + 1$ markets, and the bond market between t and $t + 1$.

If the ZLB is binding at time $t + 1$, we need to update the value of π_{t+1} to ensure that $i^c_{t+1} = 1$. This is done as follows. Government bonds are priced according to

$$1 = \beta \tau^c_{t+1} \frac{u_{c,t+2}}{u_{c,t+1}} i^c_{t+1} \pi^{-1}_{t+2} \tag{50}$$

And the real wage process gives

$$\pi_{t+2} = \bar{\pi} \frac{w_{t+1}}{w_{t+2}} \tag{51}$$

This holds regardless of whether the ZLB binds or not. Plugging this in and imposing $i^c_{t+1} = 1$ gives

$$1 = \beta \tau^c_{t+1} \frac{u_{c,t+2}}{u_{c,t+1}} \frac{w_{t+2}}{\bar{\pi} w_{t+1}} \tag{52}$$

Finally, use the wage equation one period forward to express w_{t+1} as

$$w_{t+1} = \frac{\bar{\pi}}{\pi_{t+1}} w_t \tag{53}$$

Plugging this in and solving for π_{t+1} gives

$$\pi_{t+1} = \frac{u_{c,t+1}\bar{\pi}^2 w_t}{\beta\tau^c_{t+1}u_{c,t+2}w_{t+2}} \tag{54}$$

In the simulation, all t and $t+2$ variables are held constant while solving for the $t+1$ variables, and so π_{t+1} can be solved for from the above. Before convergence, the Euler equation will not exactly hold since incorrect values of t and $t+2$ variables will be used, but once convergence is reached the Euler equation will hold exactly, by construction.

C 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.[11] 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

[11] 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.

this limit end up clustering on the same level of net worth, conditional on their (k, z).

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.