Lumpy Investment, Business Cycles, and Stimulus Policy*

Thomas Winberry
Chicago Booth and NBER
Thomas.Winberry@chicagobooth.edu

February 14, 2020

Abstract

I study the aggregate implications of micro-level lumpy investment in a model consistent with the empirical dynamics of the real interest rate. I find that the elasticity of aggregate investment with respect to shocks is procyclical because more firms are likely to make an extensive margin investment in expansions than in recessions. Matching the dynamics of the real interest rate is key to generating this result; otherwise, counterfactual behavior of the model would eliminate most of the procyclical responsiveness. Therefore, data on interest rates places important discipline on the role of general equilibrium in aggregating micro-level investment behavior.

*I am extremely grateful to Richard Rogerson, Pablo Ottonello, Greg Kaplan, Mike Golosov, and Esteban Rossi-Hansberg for many conversations that shaped this paper. I also thank Mark Aguiar, Diane Alexander, Adrien Auclert, Gabe Chodorow-Reich, Jesus Fernandez-Villaverde, Oleg Itskhoki, Nobu Kiyotaki, Yann Koby, Ben Moll, Ezra Oberfield, Chris Sims, Christian vom Lehn, Christian Wolf, Eric Zwick, and four anonymous referees. Financial support from Chicago Booth is gratefully acknowledged.

1 Introduction

Aggregate investment is one of the most volatile components of GDP over the business cycle, accounting for 38% of the decline in GDP during recessions.[1] These large swings in aggregate investment are primarily driven by changes in the number of firms undertaking an investment project (the extensive margin) rather than changes in the size of investment projects (the intensive margin).[2] However, most DSGE macro models make the simplifying assumption that all changes in aggregate investment are the result of a representative firm operating along the intensive margin. Therefore, a key question for business cycle modeling is: does this abstraction matter for understanding aggregate investment dynamics? In a benchmark real business cycle (RBC) model, the answer to this question is no; general equilibrium changes in the real interest rate bring the aggregate investment series in line with the preferences of the representative household, regardless of whether that investment occurs along the extensive or intensive margin at the micro-level.[3]

In this paper, I argue that accounting for the extensive margin at the micro level is important for understanding aggregate investment dynamics. My argument has two main components. First, I show that the dynamics of the real interest rate which drive the irrelevance results in the RBC environment are at odds with the data. Second, I build a heterogeneous firm model consistent with both the importance of the extensive margin of investment and the observed dynamics of the real interest rate, and find that the behavior of aggregate investment is substantially different than in the representative firm model. In particular, the elasticity of aggregate investment with respect to shocks is procyclical; in expansions, more firms are close to making an extensive margin investment, so an additional shock generates more total investment than it would otherwise. These results illustrate how data on interest rates place discipline on the role of general equilibrium in determining the aggregate implications of micro-level investment behavior.

In the benchmark RBC environment, changes in the real interest rate eliminate the pro-

[1] Computed as the average contribution to the percentage change in GDP from BEA Table 1.1.2 during NBER recession dates, 1953-2018.

[2] See, for example, the evidence in Doms and Dunne (1998) or Gourio and Kashyap (2007).

[3] This irrelevance result was established in an important series of papers by Thomas (2002), Khan and Thomas (2003), Khan and Thomas (2008), and further elaborated upon by House (2014).

cyclical responsiveness to shocks because investment is extremely interest-sensitive; therefore, small but procyclical movements in the real interest rate are able to restrain the large movements in the extensive margin of investment. However, I show that these movements in the interest rate are inconsistent with two key features of the data. First, the interest rate is negatively correlated with aggregate output and productivity, suggesting that it does not dampen cyclical movements in investment demand. Second, the interest rate is an order of magnitude more volatile in the data than in the RBC model, suggesting that investment is not as interest-sensitive as it is in the model.

Motivated by this evidence, I extend a simple heterogeneous firm model with extensive margin investment decisions to capture the empirical dynamics of the real interest rate. In the model, there is a fixed mass of firms who make investment decisions subject to both fixed and convex capital adjustment costs. The presence of fixed costs generates the extensive margin choice of whether to invest or not. There is a representative household whose preferences feature habit formation and aggregate dynamics are driven by shocks to aggregate productivity. The equilibrium dynamics of the real interest rate are determined by the strength of habit formation – which controls the sensitivity of investment supply to aggregate shocks – and the strength of adjustment costs – which control the sensitivity of investment demand to shocks. I calibrate these features of the model to match both the dynamics of the real interest rate and the behavior of micro-level investment.

Quantitatively, my calibrated model predicts that aggregate investment is nearly 30% more responsive to an aggregate shock starting from a brisk expansion than it is starting from a deep recession. As described above, this procyclical responsiveness reflects the fact that more firms are close to making an extensive margin investment in expansions, so an additional shock induces more firms to invest. However, as Khan and Thomas (2008) show, embedding this mechanism into an otherwise standard RBC model (without habit formation or convex adjustment costs) generates little variation in the responsiveness to shocks. In my version of their calibration, aggregate investment is only 10% more responsive to shocks in the expansion compared to recession.

My model generates substantial variation in responsiveness for two related reasons. First, matching the negative comovement between the real interest rate and aggregate productivity

implies that the interest rate does not directly dampen the effect of a productivity shock. Second, matching both the dynamics of the interest rate and the behavior of investment indirectly disciplines the interest-sensitivity of investment, which Koby and Wolf (2019) show is key to determining aggregation in this class of models. In my version of Khan and Thomas (2008)'s calibration, the semi-elasticity of aggregate investment with respect to the real interest rate is approximately -1075, so small changes in the interest rate have a strong influence on aggregate investment. In my model, the corresponding semi-elasticity is only -6.73, giving general equilibrium a much smaller influence over aggregate dynamics.

I illustrate two implications of the model for investment stimulus policies, such as investment tax credits or the bonus depreciation allowance. First, the aggregate effect of investment stimulus is also state dependent and falls in recessions; therefore, predictions based on linear models overstate the effectiveness of these policies in recessions. Second, I develop a simple size-dependent stimulus policy that increases cost effectiveness by 30% compared to existing size-independent policies. The main insight of this alternative policy is to avoid subsidizing investment that would have been done even without the policy. Because investment primarily occurs along the extensive margin, most of this inframarginal waste is accounted for by subsidizing firms that would have made an extensive margin investment without the policy. In my model, as in the data, small firms grow faster than average and are therefore more likely to be inframarginal to the policy.

A key challenge throughout the analysis is efficiently computing the equilibrium of the model, which involves approximating the entire cross-sectional distribution of firms. I use the method developed concurrently in Winberry (2018), which approximates the distribution with a flexible but finite-dimensional parametric family. I find that this approach captures how changes in the shape of the distribution affect the dynamics of aggregate variables more accurately than simply using the mean capital stock.

Related Literature This paper contributes to three main strands of literature. First, it addresses the long-standing question of how the extensive margin of investment impacts aggregate dynamics. Early work, analyzing firms' decision rules with fixed prices, finds that the

extensive margin generates procyclical responsiveness to shocks (as in my model).[4] However, Thomas (2002), Khan and Thomas (2003), and Khan and Thomas (2008) show that most of this time-varying elasticity disappears when prices are endogenized in an otherwise standard RBC framework, rendering the extensive margin irrelevant for aggregate dynamics. House (2014) suggests that these irrelevance results are driven by the extreme sensitivity of investment to changes in the relative price of investment goods in a stylized partial equilibrium model. In recent work, Koby and Wolf (2019) analytically show that the elasticity of investment with respect its user cost is a sufficient statistic for the aggregation properties of a wide class of models with firm heterogeneity. They argue against an extreme user cost elasticity by matching the empirical response of investment to tax changes measured in Zwick and Mahon (2017). I show that matching the dynamics of the real interest rate also requires breaking this extreme sensitivity and, therefore, a key source of the irrelevance results.[5]

To match the dynamics of the real interest rate, I follow Beaudry and Guay (1996) in using habit formation and capital adjustment costs. Boldrin, Christiano and Fisher (2001) also use this approach to match interest rate dynamics and the level of the equity premium. These papers work in representative agent environments; my results show that many of their lessons carry over to a heterogeneous firm environment in which adjustment costs are disciplined with micro-level investment data.

Finally, this paper contributes to a large literature which studies investment stimulus policy. Many papers estimate the effect of stimulus policy through linear regression models, most recently House and Shapiro (2008) and Zwick and Mahon (2017). Edge and Rudd (2011) introduce the Bonus Depreciation Allowance into a linearized New Keynesian model, which rules out state dependence by construction. I focus on the effect of stimulus policy

[4]See, for example, Caballero et al. (1995) or Caballero and Engel (1999).

[5]Other papers challenge the irrelevance results on other grounds. Gourio and Kashyap (2007) and Miao and Wang (2014) show that the results are sensitive to the distribution of fixed adjustment costs and the degree of returns to scale. Bachmann, Caballero and Engel (2013) argue that Khan and Thomas (2008)'s calibrated fixed costs are implausibly small and that increasing them to empirically reasonable levels breaks the irrelevance results. Bachmann and Ma (2016) and Bayer and Tjaden (2016) argue that the irrelevance results are sensitive to the precise form of general equilibrium; Bachmann and Ma (2016) allow for two savings vehicles (physical capital and inventories) and Bayer and Tjaden (2016) allow for multiple countries. Finally, and most closely related to this paper, Cooper and Willis (2014) parameterize an interest rate process from the data and solve firms' decision problems given this process. My paper produces such an interest rate process endogenously in general equilibrium.

over the business cycle and how micro-level targeting can increase its cost effectiveness.[6]

Road Map The rest of this paper is organized as follows. Section 2 describes the role of the real interest rate in driving the existing irrelevance results in the literature and argues that the implied interest rate behavior is counterfactual. Section 3 develops my quantitative heterogeneous firm model, which Section 4 parameterizes to jointly match micro-level investment behavior and macro-level interest rate dynamics. Section 5 shows that the existence of the extensive margin implies that aggregate investment is more responsive to shocks in expansions than in recessions, and argues that matching the dynamics of the interest rate is key to generating this result in general equilibrium. Section 6 introduces stimulus policy into the model, shows that the effectiveness of these policies falls in recessions, and develops the alternative size-dependent stimulus to increase cost effectiveness. Finally, Section 7 concludes.

2 Role of Real Interest Rate Dynamics

This section motivates the features of real interest rate dynamics on which I will focus for the rest of the paper. Section 2.1 uses a simplified RBC model to illustrate the role of the interest rate in rendering the extensive margin of investment irrelevant for aggregate dynamics. Section 2.2 shows that the key features of the interest rate which drive this result are inconsistent with the data.

2.1 Irrelevance of Extensive Margin in a Simple RBC Model

I use a stylized model to simplify the exposition. In this model, I take returns to scale to be nearly constant, which implies that investment is nearly infinitely sensitive to changes in the real interest rate. Therefore, small but procyclical movements in the real interest rate are able to bring aggregate investment in line with the representative household's desired path

[6] Berger and Vavra (2015) analyze a related class of consumer durable stimulus policies in a model of lumpy durable investment. They find that stimulus policies are less effective in recessions for similar reasons as here; however, they focus on detailed features of the micro data while I focus on the role of real interest rate dynamics in aggregation and on designing more cost effective policies.

of consumption, regardless of the existence of fixed costs.

This analysis builds heavily on related work in the literature. Miao and Wang (2014) also show that fixed costs are irrelevant for aggregate dynamics under constant returns to scale, though they do not relate that mechanism to interest-sensitivity or allow for idiosyncratic productivity shocks. House (2014) shows that low depreciation rates can also generate extreme price-sensitivity of investment and render fixed costs irrelevant for aggregates. Finally, Koby and Wolf (2019) show that the user-cost elasticity of investment is a sufficient statistic for characterizing the aggregation properties of a general class heterogeneous firm models. The main value added of my analysis here is to summarize the mechanism in a transparent way and to study its empirical implications for interest rate dynamics.

Simple RBC Model with Fixed Costs Consider a discrete time environment with heterogeneous firms indexed by $j \in [0, 1]$. Firm j produces output y_{jt} using the production function

$$y_{jt} = z_t \varepsilon_{jt} k_{jt}^{\alpha}, \tag{1}$$

where z_t is aggregate productivity, ε_{jt} is idiosyncratic productivity, k_{jt} is the firm's capital stock, and the parameter $\alpha \leq 1$ controls the returns to scale. Idiosyncratic productivity ε_{jt} follows a first-order Markov process with finite support $\varepsilon \in \{\varepsilon_1, ..., \varepsilon_{n_\varepsilon}\}$. Firms have perfect foresight over the path of aggregate productivity z_t; since firms are owned by the representative household (as described below), the absence of aggregate uncertainty implies that firms use the risk-free rate r_t to discount future profits. The capital stock k_{jt} is predetermined at time t and each period the firm chooses next period's capital $k_{jt+1} = (1 - \delta)k_{jt} + i_{jt}$, where i_{jt} is gross investment and δ is the depreciation rate. Nonzero gross investment incurs a fixed resource adjustment cost $\overline{\xi}$ which is rebated lump-sum to the representative household. The initial distribution of idiosyncratic productivity and capital across firms is stationary absent changes in aggregate TFP z_t.

There is a representative household with preferences over consumption C_t represented by the utility function

$$\sum_{t=0}^{\infty} \beta^t \frac{C_t^{1-\sigma} - 1}{1 - \sigma},$$

where β is the discount factor and $1/\sigma$ is the elasticity of intertemporal substitution. The household owns all firms in the economy. Total output can be used for consumption or investment, which implies the aggregate resource constraint

$$Y_t = C_t + K_{t+1} - (1-\delta)K_t,$$

where $Y_t = \int y_{jt}\mathrm{d}j$ and $K_t = \int k_{jt}\mathrm{d}j$.

***Proposition* 1.** *Let the returns to scale parameter $\alpha < 1$. Let $k_t^*(\alpha)$ be the optimal capital accumulation policy at date t for a firm with idiosyncratic productivity $\widetilde{\varepsilon} = \max_i \mathbb{E}[\varepsilon'|\varepsilon_i]$, conditional on paying the fixed cost. Let $\pi_t^*(\alpha)$ denote the flow profits associated with that choice. If $\overline{\xi} \leq \pi_t^*(\alpha)$, then*

$$r_t + \delta \to z_{t+1}\widetilde{\varepsilon} \quad \text{as } \alpha \to 1. \tag{2}$$

Furthermore, aggregate output Y_t, investment I_t, and consumption C_t approach the outcomes of a representative firm model with aggregate productivity $\widetilde{Z}_t = Z_t\widetilde{\varepsilon}$ and initial capital stock K_0 equal to the aggregate capital stock in the invariant distribution.

Proof. See Appendix A. ■

In the limit of Proposition 1, the fixed cost $\overline{\xi}$ is irrelevant in the sense that the aggregate outcomes can be derived from a representative firm without the fixed cost. As the returns to scale approaches one, firms become infinitely elastic with respect to changes in the real interest rate because their profit function becomes linear with respect to capital. Since general equilibrium requires consumption to be positive and finite at all dates t – and therefore investment be finite at all dates t – the real interest rate adjusts to ensure that firms with the highest value of expected future productivity earn zero profits. At this real interest rate, only firms with $\mathbb{E}[\varepsilon'|\varepsilon] = \widetilde{\varepsilon}$ accumulate capital and all other firms do not.

The irrelevance of fixed costs is ensured by the real interest rate adjusting to (2) in order to generate an equilibrium which satisfies the household's preference for positive consumption. Although quantitative models do not exactly satisfy the conditions in Proposition 1, they are often close to that limit, giving the representative household's preferences for smooth consumption a powerful influence over aggregate investment. This occurs in quantitative

models for two reasons. First, allowing for modest decreasing returns $\alpha < 1$ does not break the extreme sensitivity of investment with respect to the real interest rates. The semi-elasticity of investment with respect to the real interest rate for adjusting firms is

$$\frac{\partial i_{jt}/i_{jt}}{\partial r_t} = -\frac{1}{\delta}\frac{1}{1-\alpha}\left(\frac{1+r_t}{r_t+\delta}\right). \tag{3}$$

As $\alpha \to 1$, this semi-elasticity (3) becomes infinite. However, even under a more typical calibration – for example, with $\alpha = 0.7$, $\delta = 0.025$, and $r_t = 0.01$ – the semi-elasticity is approximately -3,847.[7] Second, the requirement of small fixed costs $\overline{\xi}$ is often guaranteed by assuming that fixed costs are a random draw from a $U[0, \overline{\xi}]$ distribution, which ensures that there is always a positive mass of firms with arbitrarily small fixed costs.[8]

2.2 Comparing Interest Rate Dynamics to the Data

The simple framework described in Section 2.1 implies that the real interest rate must move one-for-one with productivity to ensure that the zero variable profit condition (2) holds, generating exact aggregation to a representative firm. In more general models, the real interest rate adjusts to ensure that the zero profit approximately holds, generating approximate aggregation. In this subsection, I show that these real interest rate dynamics are counterfactual.

Measurement I study the joint dynamics of the real interest rate, aggregate total factor productivity (TFP), and aggregate output in the U.S. data 1954q1 - 2016q4. I measure the real interest rate r_t as the nominal return on 90-day Treasury bills adjusted for realized CPI inflation. I measure aggregate productivity Z_t as the aggregate Solow residual. Finally, I measure output Y_t as real GDP. Details of the data construction are contained in Appendix B.1. I compare the data to the benchmark RBC model, which is quantitatively close to models

[7] Gourio and Kashyap (2007) calibrate a strong degree of decreasing returns and show that helps break this irrelevance result.

[8] Even in models in which fixed costs are non-random and strictly positive, aggregate investment may be extremely interest-sensitive. For example, House (2014) shows that when $\delta \to 0$, the elasticity of the timing of investment episodes with respect to the relative price of capital is infinite, playing a similar role to the infinite interest elasticity here.

Table 1
Cyclical Dynamics of Risk-Free Rate

	$\sigma(r_t)$	$\rho(r_t, y_t)$	$\rho(r_t, z_t)$
Whole sample	1.73%	-0.11^{*}	-0.20^{***}
(p-value)		(0.09)	(0.00)
No Volcker	1.61%	-0.03	-0.18^{***}
		(0.66)	(0.01)
Pre-1983	1.57%	-0.39^{***}	-0.18^{*}
		(0.00)	(0.06)
Post-1983	1.86%	0.21^{**}	-0.24^{***}
		(0.01)	(0.00)
RBC	0.16%	0.96	0.97

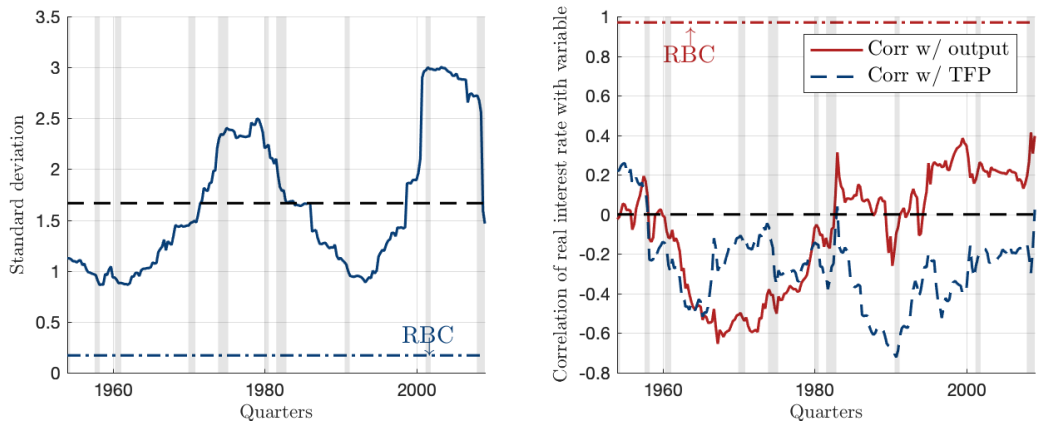
Notes: real interest rate measured as the return on 90-day Treasury bills adjusted for realized CPI inflation, expressed in annual percentage points. Output measured as real GDP. TFP measured as the aggregate Solow residual. All variables have been HP-filtered and expressed as percentage deviation from an HP trend. "Whole sample" refers to the 1954q1 - 2016q4 time series. "No Volcker" excludes 1979q1 - 1983q4. "Pre-1983" refers to the 1954q1-1982q4 sample. "Post-1983" refers to the 1983q1-2016q4 sample. "RBC" refers to the benchmark RBC model described in Appendix C.

in which fixed costs are irrelevant for aggregate dynamics (as discussed in the introduction). The detailed model specification and calibration, which follows standard practice in the business cycle literature, is contained in Appendix C.

Descriptive Results Table 1 shows that the RBC model is counterfactual along two key dimensions. First, the real interest rate is negatively correlated with aggregate TFP (-0.20), while in the model the two are extremely positively correlated (0.97). Second, the standard deviation of the interest rate is an order of magnitude higher in the data (1.73%) than in the model (0.16%). Table 1 also shows that these two findings hold in three different sub-samples of the data: pre-1983, post-1983, and excluding the years near the Volcker recession.

Figure 1 further illustrates the stability of these two statistics by plotting eight-year rolling windows of each statistic over the postwar sample. The correlation of the real interest rate with aggregate productivity is negative for nearly the entire sample. The correlation of the real interest rate with GDP is negative before 1983 but positive after 1983, consistent with the fact that productivity is less procyclical in the later period. Nevertheless, both correlations are consistently below the level implied by the RBC model. The standard deviation

FIGURE 1: Stability of Cyclical Dynamics of Risk-Free Rate



Notes: statistics computed in forward-looking 8 year rolling windows over the sample. The left panel plots the standard deviation of the real interest rate. The right panel plots the correlation of the real interest rate with output and TFP. Dashed lines correspond to the population statistics in the benchmark RBC model described in Appendix C. Real interest rate measured as the return on 90-day Treasury bills adjusted for realized CPI inflation, expressed in annual percentage points. Output measured as real GDP. TFP measured as the aggregate Solow residual. All variables have been logged and HP-filtered.

of the real interest rate is consistently above the prediction of the RBC model over the entire sample.

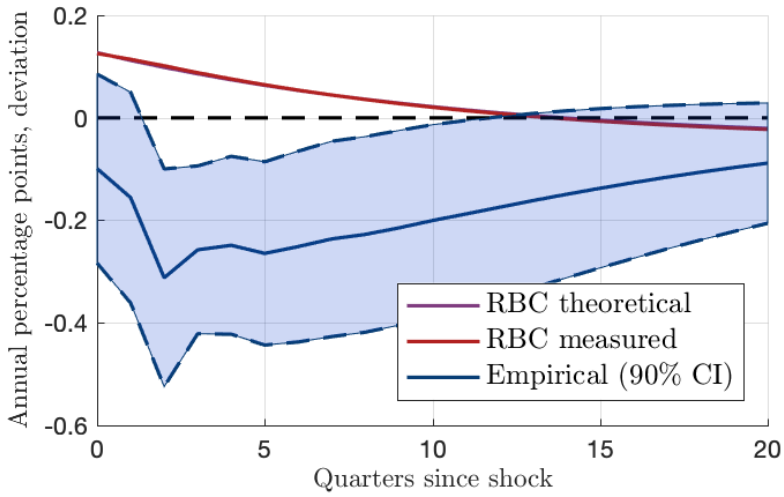
Impulse Response to TFP Shock In order to make a consistent comparison between the data and the RBC model, I estimate the impulse response of the real interest rate to a TFP shock using a simple bivariate VAR:

$$X_t = \sum_{j=1}^{p} \Gamma_j X_{t-j} + e_t, \tag{4}$$

where $X_t = (Z_t, r_t)^{\mathrm{T}}$, p is the lag length, Γ_j are coefficient matrices, and e_t are residuals. I choose the lag length $p = 3$ following the AIC. I identify TFP shocks by assuming that shocks to the interest rate equation do not affect TFP upon impact.[9] I use the adjusted TFP measure from Fernald (2014) in order to isolate changes in productivity that are not driven by endogenous changes in factor utilization.

[9]Note that this identifying assumption is weaker than the assumption that TFP is exogenous with respect to the real interest rate at all dates.

FIGURE 2: Impulse Response of the Real Interest Rate to TFP Shock



Notes: impulse response of the real interest rate to a TFP shock identified from a bivariate VAR with TFP ordered first. TFP is adjusted for changes in utilization following Fernald (2014). Lag length of 3 chosen by the AIC criteria. "RBC theoretical" refers to the theoretical impulse response from the benchmark RBC model described in Appendix C. "RBC measured" refers to the impulse response identified using the VAR estimation on simulated data from the model. "Empirical (90% CI)" refers to the empirical impulse response and 90% error bands.

Figure 2 shows that the RBC model fails along the same two dimensions highlighted in the descriptive analysis above. First, the empirical response is negative while the model's is positive, consistent with the differences in correlations presented in Table 1. Second, the magnitude of the empirical response is larger than in the model, consistent with the differences in volatilities in Table 1. The figure also shows that the model's theoretical impulse response is nearly identical to the response estimated using the VAR (4) on model-simulated data, validating the specification within that model.

Robustness The key mechanism driving the irrelevance result in Section 2.1 is that the user cost of capital $r_t + \delta$ moves one-for-one with aggregate productivity. While the empirical results above indicate that is not the case, one may be concerned that a richer specification of the user cost does move one-for-one with productivity, re-establishing the zero profit condition. For instance, in a model in which consumption goods cannot be transformed into investment goods one-for-one, the user cost is $q_t(1 + r_t) - (1 - \delta)q_{t+1}$ where q_t is the relative price of capital. Table 2 shows that this generalized user cost also does not move one-for-one

Table 2
Cyclical Dynamics of User Cost u_t

	$\sigma(u_t)$	$\rho(u_t, y_t)$	$\rho(u_t, z_t)$
Whole sample	1.98%	−0.06	−0.11*
(p-value)		(0.33)	(0.07)
No Volcker	1.93%	0.05	−0.07
		(0.43)	(0.26)
Pre-1983	2.39%	−0.13	0.10
		(0.18)	(0.31)
Post-1983	1.56%	0.07	−0.51***
		(0.40)	(0.00)
RBC	0.16%	0.96	0.97

Notes: user cost defined as $u_t = q_t(1 + r_t) - (1 - \delta)q_{t+1}$, where r_t is the real interest rate (measured as in Table 1) and q_t is the price of investment goods relative to consumption goods (constructed by Riccardo DiCecio), log-linearly detrended. User cost is expressed in annualized percentage points. Output measured as real GDP. TFP measured as the aggregate Solow residual. All variables have been HP-filtered and expressed as percentage deviation from an HP trend. "Whole sample" refers to the 1954q1 - 2016q4 time series. "No Volcker" excludes 1979q1 - 1983q4. "Pre-1983" refers to the 1954q1-1982q4 sample. "Post-1983" refers to the 1983q1-2016q4 sample. "RBC" refers to the benchmark RBC model described in Appendix C.

with productivity; in fact, its time-series behavior is similar to that of the real interest rate r_t.[10] Therefore, I focus on the simpler case in which $q_t = 1$ for the rest of the paper.[11]

Appendix B.2 contains a number of additional robustness checks of these empirical results. The appendix shows that the results continue to hold if one uses a VAR to construct expected inflation, are robust to using core CPI rather than headline CPI to deflate the nominal interest rate, and are robust to different choices of business cycle filters. It also displays the impulse response of the ex-ante real interest rate to a TFP shock, which is targeted in the model calibration in Section 4.

[10] One could further extend the user cost to include capital taxes, but the cyclical variation in taxes is limited.

[11] Most empirical estimates of the user cost elasticity of investment are low and therefore also provide evidence against the extreme interest-sensitivity driving the irrelevance results. In fact, Koby and Wolf (2019) show that the user cost elasticity from Zwick and Mahon (2017)'s estimates implies that fixed costs are quantitatively important for aggregate dynamics.

3 Model

I now develop a quantitative heterogeneous firm model to be consistent with both the behavior of investment at the micro level and the dynamics of the real interest interest rate at the macro level.

3.1 Environment

The model is set in discrete time.

Firms The firm side of the model builds heavily on Khan and Thomas (2008), extended to include convex adjustment costs and the corporate tax code.[12] There is a fixed mass of firms $j \in [0, 1]$ that produce output y_{jt} using the production function

$$y_{jt} = z_t \varepsilon_{jt} k_{jt}^{\theta} n_{jt}^{\nu},$$

where z_t is an aggregate productivity shock, ε_{jt} is an idiosyncratic productivity shock, k_{jt} is the firm's capital stock, n_{jt} is its labor input, and θ and ν are parameters satisfying $\theta + \nu < 1$. The aggregate shock z_t is common to all firms and follows the AR(1) process

$$\log z_{t+1} = \rho_z \log z_t + \omega_{t+1}^z, \text{ where } \omega_{t+1}^z \sim N(0, \sigma_z^2).$$

The idiosyncratic shock ε_{jt} is independent across firms but within firm follows the AR(1) process

$$\log \varepsilon_{jt+1} = \rho_\varepsilon \log \varepsilon_{jt} + \omega_{t+1}^\varepsilon, \text{ where } \omega_{t+1}^\varepsilon \sim N(0, \sigma_\varepsilon^2).$$

Each period, a firm j observes these two shocks, uses its pre-existing capital stock, hires labor from a competitive labor market at wage w_t, and produces output y_{jt}.

After production, the firm decides how much capital in which to invest for the next period. Gross investment i_{jt} yields $k_{jt+1} = (1 - \delta)k_{jt} + i_{jt}$ units of capital in period $t + 1$. This investment is subject to two capital adjustment costs. First, if $i_{jt} \neq 0$, then the firm must

[12] I include the tax code in order to study investment stimulus policy in Section 6.

pay ξ_{jt} units of labor.[13] The fixed cost ξ_{jt} is a uniform random variable with support $[0, \overline{\xi}]$, distributed independently across firms and time. Second, any nonzero amount of investment incurs the quadratic adjustment cost $-\frac{\varphi}{2}\left(\frac{i_{jt}}{k_{jt}}\right)^2 k_{jt}$ units of output.

After production and investment, the firm pays a linear tax rate τ on its revenue y_{jt} net of two deductions. First, the firm deducts its labor costs $w_t n_{jt}$. Second, it deducts capital depreciation costs according to the following geometric schedule. The firm enters the period with a stock of depreciation allowances d_{jt}, of which it writes off $\widehat{\delta} d_{jt}$ from its tax bill. The firm also writes off the same fraction $\widehat{\delta}$ of new investment i_{jt} from its tax bill. The remaining portion is then carried into the next period, so that $d_{jt+1} = (1-\widehat{\delta})(d_{jt} + i_{jt})$.[14] In total, the tax bill in period t is

$$\tau\left(y_{jt} - w_t n_{jt} - \widehat{\delta}(d_{jt} + i_{jt})\right).$$

Households There is a representative household with preferences represented by the expected utility function

$$\mathbb{E}\sum_{t=0}^{\infty}\beta^t \log\left(C_t - \chi\frac{N_t^{1+\eta}}{1+\eta} - X_t\right),$$

where C_t is consumption, N_t is labor supplied to the market, and X_t is habit stock. The habit stock X_t is

$$X_t = \lambda \widehat{C}_{t-1}, \tag{5}$$

where $\widehat{C}_t = C_t - \chi\frac{N_t^{1+\eta}}{1+\eta}$ is consumption net of the disutility of work and λ controls the sensitivity of habit with respect to $\widehat{C}_t$.[15] I assume that the household does not take into account the fact that their current choices impact the future habit stock X_{t+1}. The total

[13] Khan and Thomas (2008) allow for any investment in $[-ak_{jt}, ak_{jt}]$ to be free of the fixed costs. I set $a = 0$ for simplicity, but results are robust to allowing for empirically reasonable values of a.

[14] In reality, the U.S. tax code follows an annual straight line depreciation schedule with a half-year purchase convention rather than this simple quarterly geometric schedule. Section 3.2 shows that only the present value of this schedule per unit of investment affects firm's decisions in my model. See Xu and Zwick (2018) for an analysis of the implications of the details of the tax code using a model of firm-level tax management.

[15] I assume this particular form of preferences for two reasons. First, following Greenwood, Hercowitz and Huffman (1988), they eliminate the wealth effect on labor supply and allow the model to generate procyclical hours worked with a countercyclical real interest rate. With standard KPR preferences, the fact that the real interest rate falls in expansions potentially induces households to intertemporally substitute future leisure for current leisure in expansions, leading to a fall in hours worked. Second, assuming habit formation over the consumption bundle $\widehat{C}_t$ simplifies the analysis of the stochastic discount factor. The main results of the paper also hold if habit is defined over actual consumption C_t only.

time endowment per period is 1, so that $N_t \in [0,1]$. The household owns all firms in the economy and markets are complete.

Government The government collects the corporate tax and transfers the proceeds lump sum to the household. In period t, this transfer is

$$T_t = \tau \left(Y_t - w_t N_t - \widehat{\delta} \left(D_t + I_t \right) \right), \tag{6}$$

where Y_t is aggregate output, N_t the aggregate labor input, D_t the aggregate stock of depreciation allowances, and I_t is aggregate investment.

3.2 Firm Optimization

I characterize the firm's optimization problem recursively. The firm's individual state variables are ε_{jt}, its current draw of the idiosyncratic productivity shock; k_{jt}, its pre-existing stock of capital; d_{jt}, its pre-existing stock of depreciation allowances; and ξ_{jt}, its current draw of the fixed cost. I denote the aggregate state vector $\mathbf{s}_t$ and postpone discussion of its elements until I define the recursive competitive equilibrium in Section 3.4.

The firm's value function $v(\varepsilon, k, d, \xi; \mathbf{s})$ solves the Bellman equation

$$\begin{aligned} v(\varepsilon, k, d, \xi; \mathbf{s}) = \tau \widehat{\delta} d + \max_n \left\{ (1-\tau) \left(z \varepsilon k^\theta n^\nu - w(\mathbf{s}) n \right) \right\} \\ + \max \left\{ v^a(\varepsilon, k, d; \mathbf{s}) - \xi w(\mathbf{s}), v^n(\varepsilon, k, d; \mathbf{s}) \right\}. \end{aligned} \tag{7}$$

The first max operator represents the optimal choice of labor and the second max operator represents the extensive margin choice of investment. These two choices are independent because the choice of labor is a purely static problem.

If the firm chooses to pay the fixed cost, then it achieves the choice-specific value function $v^a(\varepsilon, k, d; \mathbf{s})$ defined by the Bellman equation:

$$\begin{aligned} v^a(\varepsilon, k, d, \mathbf{s}) = \max_i - \left(1 - \tau \widehat{\delta}\right) i - \frac{\varphi}{2} \left(\frac{i}{k} \right)^2 k + \mathbb{E}[\Lambda(z'; \mathbf{s}) v(\varepsilon', k', d', \xi'; \mathbf{s}') | \varepsilon, k, d] \\ \text{s.t. } k' = (1-\delta)k + i \text{ and } d' = \left(1 - \widehat{\delta}\right)(d + i), \end{aligned} \tag{8}$$

where $\Lambda(z';\mathbf{s})$ is the stochastic discount factor. I denote the implied "target" capital stock $k^a(\varepsilon,k,d;\mathbf{s}) = (1-\delta)k + i^a(\varepsilon,k,d;\mathbf{s})$.

If the firm chooses not to pay its fixed cost then it achieves the choice-specific value function $v^n(\varepsilon,k,d;\mathbf{s})$ defined by the Bellman equation:

$$v^n(\varepsilon,k,d;\mathbf{s}) = \mathbb{E}[\Lambda(z';\mathbf{s})v(\varepsilon',k',d',\xi';\mathbf{s}')|\varepsilon,k,d] \tag{9}$$
$$\text{s.t. } k' = (1-\delta)k \text{ and } d' = \left(1-\widehat{\delta}\right)d.$$

The only difference from the unconstrained Bellman equation (8) is that investment is constrained to be $i=0$. I call the implied "constrained" capital stock $k^n(\varepsilon,k,d;\mathbf{s}) = (1-\delta)k$.

The firm will choose to pay the fixed cost if and only if the value from doing so is higher than the value of not paying the fixed cost, i.e., if and only if $v^a(\varepsilon,k,d;\mathbf{s}) - \xi w(\mathbf{s}) \geq v^n(\varepsilon,k,d;\mathbf{s})$. For each tuple $(\varepsilon,k,d;\mathbf{s})$, there is a unique threshold $\widehat{\xi}(\varepsilon,k,d;\mathbf{s})$ which makes the firm indifferent between these two options:

$$\widehat{\xi}(\varepsilon,k,d;\mathbf{s}) = \frac{v^a(\varepsilon,k,d;\mathbf{s}) - v^n(\varepsilon,k,d;\mathbf{s})}{w(\mathbf{s})}. \tag{10}$$

For draws of the fixed cost ξ below $\widehat{\xi}(\varepsilon,k,d;\mathbf{s})$, the firm pays the fixed cost; for draws of the fixed cost above $\widehat{\xi}(\varepsilon,k,d;\mathbf{s})$, it does not. This threshold is increasing in the "capital imbalance" $|k^a(\varepsilon,k,d;\mathbf{s}) - k^n(\varepsilon,k,d;\mathbf{s})|$ since the value from adjusting is higher when the target capital stock is further away from the constrained capital stock. The firm will only find it optimal to pay the fixed cost infrequently, generating the lumpy investment patterns observed in the micro data.

It is possible to simplify the problem by eliminating the tax depreciation allowances d from the firm's state vector. In this model, firms only care about the present value of the depreciation allowances generated by their investment because these allowances enter the firm's flow profits separately from the other terms. Therefore, the tax depreciation schedule simply decreases the implicit price of new investment by the present value of depreciation allowances which that investment generates. I formalize this logic in Proposition 2:

***Proposition* 2.** *The firm's value function is of the form* $v(\varepsilon,k,d,\xi;\mathbf{s}) = v^1(\varepsilon,k,\xi;\mathbf{s}) +$

$\tau PV(\mathbf{s})d$ where $PV(\mathbf{s})$ is defined by the recursion $PV(\mathbf{s}) = \widehat{\delta} + \left(1 - \widehat{\delta}\right) \mathbb{E}\left[\Lambda(z';\mathbf{s})PV(\mathbf{s}')\right]$. Furthermore, $v^1(\varepsilon, k, \xi; \mathbf{s})$ is defined by the Bellman equation

$$v^1(\varepsilon, k, \xi; \mathbf{s}) = \pi\left(\varepsilon, k; \mathbf{s}\right) + \max_i \left\{ \begin{array}{c} -q(\mathbf{s})i - \frac{\varphi}{2}\left(\frac{i}{k}\right)^2 k - \xi w(\mathbf{s})\mathbb{1}\left\{i \neq 0\right\} \\ +\mathbb{E}[\Lambda(z';\mathbf{s})v^1\left(\varepsilon', (1-\delta)\,k + i, \xi'; \mathbf{s}'\right)] \end{array} \right\}, \tag{11}$$

where $q(\mathbf{s}) = (1 - \tau PV(\mathbf{s}))$ is the tax-adjusted relative price of investment.

Proof. See Appendix D. ■

In Appendix D, I show that this result can be leveraged to simplify the model's equilibrium following the strategy developed in Khan and Thomas (2003).

3.3 Household Optimization

Since investment is chosen by firms, there are no dynamic linkages in the household's choices and their decision problem is equivalent to the following static problem state by state:

$$\max_{C,N} \ \log\left(C - \chi\frac{N^{1+\eta}}{1+\eta} - X(\mathbf{s})\right) \text{ subject to } C \leq w(\mathbf{s})N + \Pi(\mathbf{s}) + T(\mathbf{s}), \tag{12}$$

where $\Pi(\mathbf{s})$ are profits from the firms and $T(\mathbf{s})$ is government transfers. Markets are complete with respect to aggregate risk, so the stochastic discount factor used by firms is equal to the household's intertemporal marginal rate of substitution state by state:

$$\Lambda(z';\mathbf{s}) = \beta\frac{\widehat{C}(\mathbf{s}) - X(\mathbf{s})}{\widehat{C}(\mathbf{s}') - X(\mathbf{s}')}. \tag{13}$$

3.4 Definition of Equilibrium

The aggregate state vector is $\mathbf{s} = (z, X, \mu)$, where z is the aggregate productivity shock, X is the household's habit stock, and μ is the distribution of firms over their individual state vector (ε, k, ξ, d).

***Definition* 1.** *A **Recursive Competitive Equilibrium** for this economy is a list of functions $v(\varepsilon, k, d, \xi; \mathbf{s})$, $n\left(\varepsilon, k; \mathbf{s}\right)$, $i^a\left(\varepsilon, k; \mathbf{s}\right)$, $\widehat{\xi}\left(\varepsilon, k; \mathbf{s}\right)$, $C(\mathbf{s})$, $N(\mathbf{s})$, $T(\mathbf{s})$, $w(\mathbf{s})$, $\Pi(\mathbf{s})$, $\Lambda(z';\mathbf{s})$, $X'(\mathbf{s})$, and $\mu'(\mathbf{s})$ such that*

(i) (Household Optimization) Taking $w(\mathbf{s})$, $\Pi(\mathbf{s})$, and $T(\mathbf{s})$ as given, $C(\mathbf{s})$ and $N(\mathbf{s})$ solve the utility maximization problem (12).

(ii) (Firm Optimization) Taking $w(\mathbf{s})$, $\Lambda(z';\mathbf{s})$, $X'(\mathbf{s})$, and $\mu'(\mathbf{s})$ as given, $v(\varepsilon,k,d,\xi;\mathbf{s})$, $n(\varepsilon,k;\mathbf{s})$, $i^a(\varepsilon,k;\mathbf{s})$, and $\widehat{\xi}(\varepsilon,k;\mathbf{s})$ solve the firm's maximization problem (7) - (10).

(iii) (Government) For all $\mathbf{s}$, $T(\mathbf{s})$ is given by (6).

(iv) (Consistency) For all $\mathbf{s}$,

(a) $\Pi(\mathbf{s}) = \int [\ (1-\tau)\left(z\varepsilon k^{\theta} n(\varepsilon,k;\mathbf{s})^{\nu} - w(\mathbf{s}) n(\varepsilon,k;\mathbf{s})\right) + \tau\widehat{\delta} d - \left(1-\tau\widehat{\delta}\right) i(\varepsilon,k,\xi;\mathbf{s}) - \frac{\varphi}{2}\left(\frac{i(\varepsilon,k,\xi;\mathbf{s})}{k}\right)^2 k - \xi w(\mathbf{s}) \mathbb{1}\left\{\frac{i(\varepsilon,k,\xi;\mathbf{s})}{k} \neq 0\right\}]\ \mu(d\varepsilon,dk,dd,d\xi)$, where $i(\varepsilon,k,\xi;\mathbf{s}) = i^a(\varepsilon,k,\xi;\mathbf{s})$ if $\xi \leq \widehat{\xi}(\varepsilon,k;\mathbf{s})$ and $i(\varepsilon,k,\xi;\mathbf{s}) = 0$ otherwise.

(b) $\Lambda(z';\mathbf{s})$ is given by (13).

(c) $X'(\mathbf{s})$ follows (5).

(d) For all measurable sets $\Delta_\varepsilon \times \Delta_k \times \Delta_d \times \Delta_\xi$, $\mu'(\Delta_\varepsilon \times \Delta_k \times \Delta_d \times \Delta_\xi) = \int p(\varepsilon' \in \Delta_\varepsilon|\varepsilon)d\varepsilon' \times \mathbb{1}\{\ i(\varepsilon,k,\xi;\mathbf{s}) + (1-\delta)k \in \Delta_k\ \} \times \mathbb{1}\{\ (1-\widehat{\delta})\ (i(\varepsilon,k,\xi;\mathbf{s}) + d) \in \Delta_d\ \} \times G(\Delta_\xi) \times \mu(d\varepsilon,dk,dd,d\xi)$, where $G(\xi)$ is the CDF of ξ and $p(\varepsilon'|\varepsilon)$ is the p.d.f. of idiosyncratic shocks ε.

(v) (Market Clearing) For all $\mathbf{s}$, $N(\mathbf{s}) = \int \left(n(\varepsilon,k,\xi;\mathbf{s}) + \frac{\widehat{\xi}(\varepsilon,k,d;\mathbf{s})^2}{2\overline{\xi}}\right)\mu(d\varepsilon,dk,dd,d\xi)$.

3.5 Solution Method

The key challenge to solving the model is that the aggregate state vector $\mathbf{s}$ contains the cross-sectional distribution of firms, which is an infinite-dimensional object. I overcome this challenge using the computational method concurrently developed in Winberry (2018) (which itself builds on Campbell (1998), Reiter (2009) and Algan, Allais and Haan (2008)). The method approximates the distribution at any point in time using a flexible but finite-dimensional parametric family; the parameters of that family are then endogenous aggregate state variables of the approximated model. A good approximation of the distribution requires 5-10 endogenous parameters, leaving globally accurate approximation methods infeasible due

to the curse of dimensionality. Therefore, I solve for the aggregate dynamics of the model using a second-order perturbation. See Appendix E for details of the implementation.

I have found that Winberry (2018)'s method has two advantages over the usual approach of approximating the distribution with moments, as in Krusell and Smith (1998). First, forecasts of key variables based only on the aggregate capital stock are inaccurate (which I show in Appendix E). This fact indicates that higher-order features of the distribution are relevant in determining aggregate dynamics. Second, the method is computationally efficient due to the use of perturbation methods with respect to the aggregate state vector. A local approximation with respect to the aggregate state is appropriate because aggregate shocks are small relative to idiosyncratic shocks. The main advantage of using Krusell and Smith (1998)'s method would be that the solution is globally accurate with respect to aggregate shocks.

4 Model Parameterization and Validation

I parameterize the model to jointly match the behavior of investment at the micro level and the dynamics of the real interest rate at the macro level.

4.1 Parameterization

I parameterize the model in two steps. First, I fix a set of parameters to match standard macroeconomic targets in steady state. Second, given the values of those parameters, I choose the remaining parameters to match targets in the data. A model period corresponds to one quarter.

Fixed Parameters Table 3 lists the parameters that I fix. I set the discount factor $\beta = 0.99$. I set the Frisch elasticity of labor supply to 2, within the range of macro elasticities identified by Chetty et al. (2011). I set the labor share $\theta = 0.64$ and choose the capital share so that the total returns to scale is 85%. The returns to scale lies within the range considered in the literature, from 60% in Gourio and Kashyap (2007) to 92% in Khan and Thomas (2008). I set $\delta = 0.025$ so that the steady state aggregate investment rate is 10%, in

Table 3
Fixed Parameter Values

Parameter	Description	Value
β	Discount factor	0.99
η	Inverse Frisch	0.5
θ	Labor share	0.64
ν	Capital share	0.21
δ	Capital depreciation	0.025
ρ_z	Aggregate TFP AR(1)	0.97
σ_z	Aggregate TFP AR(1)	0.0078
τ	Tax rate	0.35
$\widehat{\delta}$	Tax depreciation	0.119

Notes: parameters fixed in calibration.

line with the average in the postwar data. I set the stochastic process for TFP to $\rho_z = 0.97$ and $\sigma_z = 0.0078$, the values of the empirical process (which was used to estimate the impulse response of the real interest rate to TFP shocks in Section 2.2).

I set the tax rate $\tau = 0.35$ to match the top marginal income tax rate in the U.S. federal corporate income tax code over most of the sample period. Given this value of the tax rate, I choose the slope of the tax depreciation schedule $\widehat{\delta}$ to match the average present value of tax depreciation allowances per unit of investment from Zwick and Mahon (2017). Proposition 2 shows that this present value summarizes how the tax depreciation schedule affects firms' decisions in my model.

Fitted Parameters I choose the remaining parameters, listed in Table 5, in order to match the empirical targets in Table 4.[16] The micro-level investment targets are computed from annual IRS corporate income tax returns, reported in Zwick and Mahon (2017).[17] The IRS

[16] I exogenously fix the persistence of the productivity shocks $\rho_\varepsilon = 0.9$. Clementi and Palazzo (2015) show the persistence parameter is only weakly identified when using investment data alone.

[17] Much of the literature with firm heterogeneity and investment calibrates models to match investment behavior from the Census of manufacturing firms, reported by Doms and Dunne (1998) or Cooper and Haltiwanger (2006). Zwick and Mahon (2017)'s data has three important advantages over Census data in the context of this paper. First, it covers all sectors of the economy rather than just manufacturing, and therefore allows for a more representative sample of the economy than previous studies. Second, it covers a more recent sample period (1998-2010) than the Census data (1972-1988). Third, the IRS data is at the firm level, which is the appropriate unit of analysis for studying tax policy in Section 6.

However, Zwick and Mahon (2017)'s data also has two disadvantages relative to the Census data. First, the

Table 4
Empirical Targets

Micro Investment (annual)		
Target	**Data**	**Model**
Average investment rate (%)	10.4%	10.8%
Standard deviation of investment rates	0.16	0.14
Spike rate (%)	14.4%	18.9%
Positive investment rates (%)	85.6%	81.1%
Interest Rate Dynamics (quarterly)		
Target	**Data**	**Model**
Cumulative impulse response	−0.41	−0.29
$\sigma(I_t)/\sigma(Y_t)$	2.87	2.73

Notes: micro investment moments from annual firm-level IRS data, 1998 - 2010, as reported in Zwick and Mahon (2017) Appendix Table B.1. Statistics drawn from distribution of investment rates pooled over firms and time. Spike rate is fraction of observations with investment rate greater than 20%. Positive investment is fraction of observations less than 20%. "Cumulative impulse response" refers to the cumulated response of the ex-ante real interest rate to a TFP shock over the first year, identified from the empirical VAR (4). $\sigma(I_t)/\sigma(Y_t)$ is the standard deviation of HP-filtered aggregate investment relative to the standard deviation of HP-filtered aggregate output.

sample features significant micro-level lumpiness, in line with previous findings in Census data (see, for example, Cooper and Haltiwanger (2006)). About one fourth of firm-years in the sample feature essentially zero investment while simultaneously one sixth of firm-years have investment rate spikes greater than 20%.

I target two statistics related to the dynamics of the real interest rate. First, I target the one-year cumulative response of the expected real interest rate to a TFP shock identified using the VAR (4).[18] Second, I target the volatility of aggregate investment relative to the volatility of aggregate output. As I discuss in Section 4.2 below, increasing the strength of either habit formation or adjustment costs has similar implications for the dynamics of the real interest rate. However, they have opposite implications for the volatility of investment; adjustment costs make investment less volatile while habit formation makes it more volatile.

tax data only record investment expenditures, while the Census data also records retirement and sales of capital. Second, measured investment in the Zwick and Mahon (2017) sample mainly includes equipment goods while the measured capital includes both equipment and structures. On net, I prefer to work with the Zwick and Mahon (2017) data because of its greater sample coverage across sectors and time. I conjecture that my results are robust to calibrating the model to the Census data since both datasets indicate substantial lumpiness of investment in terms of the frequency of inaction and the frequency of spikes.

[18] The impulse response of the expected real interest rate is plotted in Figure 9 in Appendix B.2.

Targeting the volatility of investment therefore places discipline on the relative strength of these two forces.[19]

Although the model is nonlinear and overidentified, with six moments determining four parameters, it nonetheless fits the targets in Table 4 fairly well.[20] The model captures the frequency of spikes relative to the frequency of non-spike observations, which is informative about the strength of fixed costs. The model also captures the dispersion of investment rates across firms, which is informative about the size of idiosyncratic shocks and strength of the convex adjustment costs.[21] While the model matches the negative response of the real interest rate to a TFP shock, it only captures around two thirds of the overall decline. This failure is primarily due to the hump-shaped nature of the empirical response, which the simple habit formation process (5) does not generate.

Table 5 shows that the calibrated parameter values are broadly comparable to previous findings in the literature. The upper bound on the fixed cost $\overline{\xi}$ is within the (admittedly wide) range of 0.0083 in Khan and Thomas (2008) and 4.4 in Bachmann, Caballero and Engel (2013). The calibrated value implies that the average fixed cost paid conditional on adjusting is 9.3% of firms' average quarterly output. The dispersion of idiosyncratic TFP shocks is in line with direct measures surveyed in Syverson (2011); for example, the 90-10 ratio of log productivity is 0.31 in my model vs. 0.65 in the data. The average size of the habit stock is 75% of the households consumption bundle $\hat{C}_t$, close to the 65% in Christiano, Eichenbaum and Evans (2005).

[19] Habit formation has also been shown to improve the empirical performance of DSGE models along two other dimensions. First, habit improves the internal propagation of shocks onto aggregate consumption; for example, Christiano, Eichenbaum and Evans (2005) show that habit formation allows their model to match the hump-shaped response of consumption to a monetary policy shock. Second, habit helps match various features of asset prices (see, for example, Boldrin, Christiano and Fisher (2001)).

[20] I time-aggregate the model's quarterly observations to the annual frequency by summing over flow investment done within a year and using the end-of-year capital stock.

[21] A natural target which I omit is the occurrence of investment inaction, often defined as the frequency of observations with investment rates less than 1% annually. I do not target inaction for two reasons. First, as discussed by Cooper and Haltiwanger (2006), the precise definition of inaction in the data is rather arbitrary given heterogeneity in depreciation rates or in the types of investment episodes (e.g., maintenance vs. large new projects). Second, since the IRS data only reports investment expenditure, observed inaction may reflect firms which do not purchase capital but which nonetheless sell or retire capital. It is straightforward to allow for some degree of inaction in the model by allowing $a > 0$, i.e. a small amount of investment that is not subject to the fixed cost.

Table 5
Fitted Parameter Values

Micro Heterogeneity		
Parameter	**Description**	**Value**
$\overline{\xi}$	Upper bound on fixed costs	0.703
φ	Convex adjustment cost	2.950
ρ_ε	Idiosyncratic productivity AR(1) (*fixed*)	0.900
σ_ε	Idiosyncratic productivity AR(1)	0.053
Habit Formation		
Parameter	**Description**	**Value**
λ	Sensitivity of habit w.r.t. consumption bundle	0.756

Notes: Parameters chosen to match moments in Table 4. I exogenously fix the persistence of idiosyncratic productivity following the discussion in footnote 16.

4.2 Identification

The identification of these parameters can be understood in two broad steps. First, the dynamics of the real interest rate pin down the overall strength of habit formation and adjustment costs. This occurs because the model's equilibrium real interest rate is determined by the interaction of investment supply, which is influenced by the degree of habit formation, and investment demand, which is influenced by the degree of adjustment costs. Second, given the overall strength of adjustment costs, the micro-level investment data pins down the split between fixed and convex adjustment costs and the dispersion of idiosyncratic shocks.

In order to more formally understand the first step of this process, Figure 3 plots the impulse response of key model variables under four different parameter configurations: no habit formation or adjustment costs; only habit formation; only adjustment costs; and the calibrated model with both habit formation and adjustment costs.

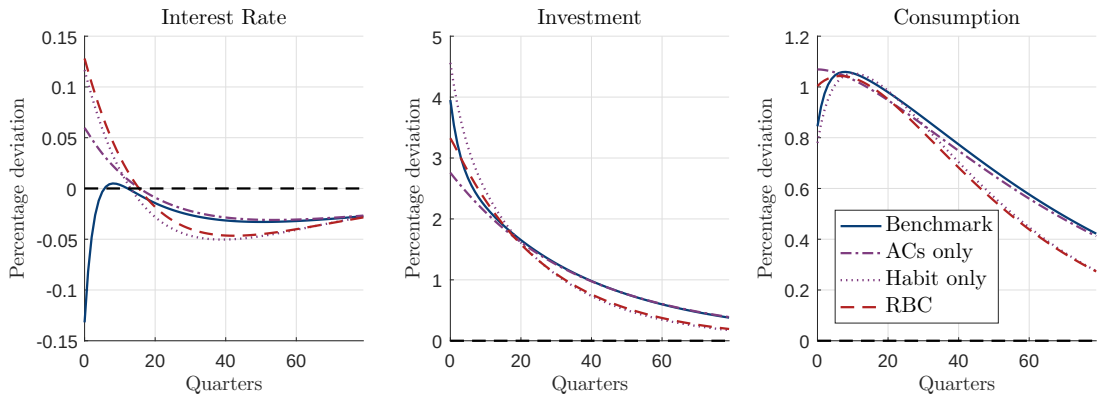
First consider the cases without habit formation. The household's Euler equation is

$$1 + r_t = \frac{1}{\beta}\mathbb{E}_t\left[\frac{\hat{C}_t}{\hat{C}_{t+1}}\right]^{-1},$$

which relates the real interest rate to expected consumption growth.[22] While a positive

[22] Due to Greenwood, Hercowitz and Huffman (1988) preferences, the Euler equation is technically in terms of the consumption bundle $\hat{C}_t = C_t - \chi\frac{N_t^{1+\eta}}{1+\eta}$ rather than consumption C_t itself. Quantitatively, the dynamics

FIGURE 3: Identification of Habit Formation and Adjustment Costs



Notes: impulse responses to a TFP shock in different parameterizations of the model. "Benchmark" refers to calibrated model. "ACs only" refers to model in which habit formation has been turned off (by setting $\lambda = 0$). "Habit only" refers to model in which adjustment costs have been turned off (by setting $\overline{\xi}$ and $\varphi = 0$ to low values). "No habit or ACs" refers to model in which both habit formation and adjustment costs have been turned off. In this case, the firm side of the model approximately aggregates to a representative firm.

TFP shock unambiguously increases both consumption C_t and investment I_t, its effect on consumption *growth* C_{t+1}/C_t – and therefore the real interest rate – depends on the responsiveness of investment because that determines the size of the capital stock in period $t+1$. Without adjustment costs, the capital stock K_{t+1} rises by enough to allow C_{t+1} to rise relative to C_t, therefore causing the real interest rate to rise. Adjustment costs dampen the response of investment, which dampens the rise of consumption growth and therefore dampens the rise of the real interest rate. However, adjustment costs alone are not quantitatively strong enough to fully account for the response of the real interest rate in the data.

Habit formation further brings the model in line with the data by breaking the tight link between consumption growth and the real interest rate. With habit formation, the household's Euler equation is

$$1 + r_t = \frac{1}{\beta}\mathbb{E}_t\left[\frac{\hat{C}_t - X_t}{\hat{C}_{t+1} - X_{t+1}}\right]^{-1},$$

where X_t is the stock of habit at time t. This equation implies that the real interest rate will fall following a positive TFP shock if the growth in the habit-adjusted consumption

of this consumption bundle are similar to consumption itself.

bundle $(\hat{C}_t - X_t)$ falls. And indeed, holding the path of the consumption bundle $\hat{C}_t$ fixed, stronger habit formation decreases the response of habit-adjusted consumption growth; with strong enough habit formation, $\widehat{C}_t - X_t$ increases by more than $\widehat{C}_{t+1} - X_{t+1}$ because X_t is predetermined during the period of the shock while X_{t+1} is not. However, stronger habit formation also increases the incentive to smooth the consumption bundle $\hat{C}_t$, which may undo some of its effect on the real interest rate. In fact, Figure 3 shows that, without adjustment costs, this consumption smoothing undoes nearly all of the effect of habit formation on the real interest rate. The increased desire to smooth consumption implies that consumption is less responsive to the shock and therefore investment is more responsive. Since there are no adjustment costs, only a small change in the path of the real interest rate is required to induce firms to increase their investment.

It is only when habit formation is combined with adjustment costs, which impede the household's ability to smooth consumption over time, that the real interest rate falls in response to a TFP shock. In this case, a large decline in the real interest rate path is required to induce firms to increase their investment and therefore partially accommodate the household's desire for smooth consumption.[23]

4.3 Relationship Between the Real Interest Rate and the SDF

The dynamics of the risk-free rate that I target in my calibration are tightly related to the stochastic discount factor $\Lambda(z'; \mathbf{s})$ which firms use to value the benefits of investment. Note that the expected present value of the firm can be decomposed as

$$\begin{aligned}\mathbb{E}[\Lambda(z'; \mathbf{s}) v(\varepsilon', k', \xi'; \mathbf{s}')|\varepsilon, k, \mathbf{s}] = & \frac{1}{1 + r(\mathbf{s})} \mathbb{E}[v(\varepsilon', k', \xi'; \mathbf{s}')|\varepsilon, k, \mathbf{s}] \\ & + \mathbb{C}ov(\Lambda(z'; \mathbf{s}), v(\varepsilon', k', \xi'; \mathbf{s}')|\varepsilon, k, \mathbf{s}) \end{aligned} \tag{14}$$

[23] This discussion makes clear that the response of the real interest rate to a productivity shock is informative about the degree of capital adjustment frictions and, therefore, the interest-sensitivity of investment. However, matching this impulse response does not imply realistic unconditional dynamics of the real interest rate; its volatility in the model is 0.15% (compared to 1.98% in the data) and its correlation with TFP is −0.66 (compared to −0.20 in the data). Adding more shocks would increase the volatility of the real interest rate and, therefore, require an even lower interest-sensitivity of investment to rationalize the observed volatility of investment in the data.

where $r(\mathbf{s}) = \frac{1}{\mathbb{E}[\Lambda(z';\mathbf{s})|\mathbf{s}]} - 1$ is the risk-free rate. The first term in the decomposition (14) captures the risk-free discounting of the value function, i.e. the implications of the stochastic discount factor (SDF) for intertemporal comparisons. The second term captures the covariance between the SDF and the value function, i.e. the implications of the SDF for risk. Hence, my calibration strategy directly targets the intertemporal component of the SDF and places no direct discipline on the risk component.[24]

Capturing the dynamics of the risk component is outside the scope of this paper for two reasons. First, in order to capture movements in the expected risk premium – which are informative about the risk component – the macro asset-pricing literature often appeals to stochastic volatility or a time-varying market price of risk, both of which are not in the model. Second, in the data, the covariance between the excess return to equity and aggregate TFP is small and does not significantly fluctuate over time. To the extent that returns to equity reflect the returns to capital, these facts suggest that the scope for time-variation in risk component in response to TFP shocks is small.[25]

4.4 Model Validation

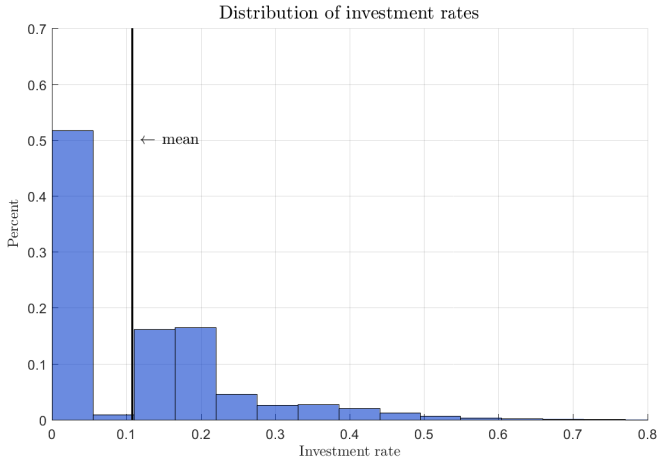
Before presenting the main results of the paper, I show that the model performs well along dimensions that were not targeted in the calibration.

Micro Investment Behavior The stationary distribution of realized investment rates across firms is broadly comparable to the empirical distribution reported in Zwick and Mahon (2017). Figure 4 plots the histogram of investment rates in the model's stationary distribution. Due to the fixed cost, there is a large mass of observations with zero investment together with a large mass of observations with large positive investment spikes. Overall, the distribution is highly non-normal; it features both positive skewness (1.55, compared to 3.60 in the data) and excess kurtosis (6.77, compared to 17.6 in the data). Investment rates are

[24]Note that the empirical analysis in Section 2.2 ignores changes in the inflation risk premium, which are likely to be small on a quarter-to-quarter basis.

[25]My model also does not match the average level of the equity risk premium in the data. Matching the level of the risk premium would increase the average level of the covariance term in (14), but not directly affect how the term responds to business cycle shocks.

Figure 4: Distribution of Annualized Investment Rates in Steady State



Notes: histogram of investment rates in the model's steady state. Investment rates are time-aggregated to the annual level in order to compare to the data. The distribution features positive skewness (1.55 in the model compared to 3.60 in the data) and excess kurtosis (6.77 in the model compared to 17.60 in the data).

Table 6
Unconditional Business Cycle Statistics

Volatility			*Cyclicality*		
Statistic	**Data**	**Model**	**Statistic**	**Data**	**Model**
$\sigma(Y_t)$	1.62%	1.77%			
$\sigma(C_t)/\sigma(Y_t)$	0.50	0.71	$\rho(C_t, Y_t)$	0.79	0.99
$\sigma(I_t)/\sigma(Y_t)$	2.87	2.73	$\rho(I_t, Y_t)$	0.77	0.97
$\sigma(N_t)/\sigma(Y_t)$	1.17	0.71	$\rho(N_t, Y_t)$	0.87	1.00

Notes: All series have been logged and HP-filtered with smoothing parameter 1600. "Data" refers to the data described in Appendix B.1. "Model" refers to calibrated model.

slightly negatively autocorrelated in the model (-0.021) because large investment spikes are often followed by zero investment.

Unconditional Business Cycle Statistics Table 6 shows that the model matches standard business cycle statistics roughly as well as the benchmark RBC model, despite the fact that the model has much stronger habit formation and adjustment costs. Investment is more volatile than output and consumption is less volatile than output in both the model and the data. The volatility of hours is lower in the model than in the data, which is a well-known de-

fect of the benchmark RBC model as well. Finally, all macroeconomic aggregates are highly correlated with each other due to the fact that there is a single aggregate shock.

5 Procyclical Responsiveness to Shocks

I now use the calibrated model to quantitatively analyze the implications of fixed costs for the dynamics of aggregate investment. Section 5.1 shows that fixed costs imply that aggregate investment is more responsive to shocks in expansions than in recessions. Section 5.2 shows that quantitatively matching the dynamics of the real interest rate is key to generating this procyclical responsiveness in general equilibrium.

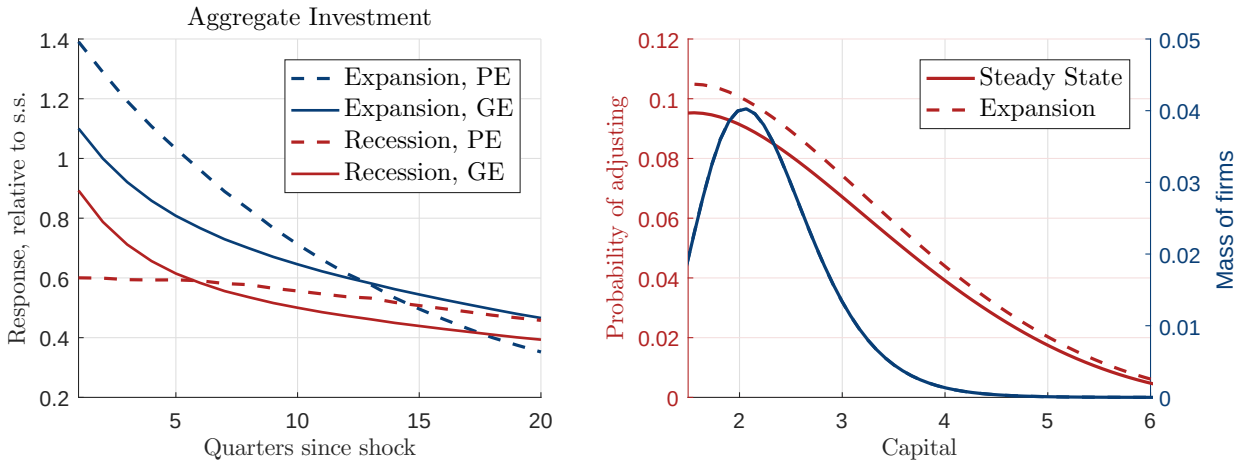
5.1 Role of Fixed Costs

I begin by describing how fixed adjustment costs generate the procyclical responsiveness of aggregate investment to an TFP shock. In order to isolate the role of firms' behavior in driving aggregate dynamics, I perform this analysis in "partial equilibrium," i.e. by aggregating firms' decision rules with prices held fixed at their steady state values. This analysis provides a natural benchmark against which to compare the general equilibrium results in Section 5.2.

Procyclical Impulse Responses Figure 5 compares the response of aggregate investment to a TFP shock starting from two different points in the business cycle. The first starting point is an example recession generated by a history of -2.5 standard deviation shocks over the previous year, which generates 25% decline in annual investment in the full general equilibrium model and is comparable to the observed decline during the 2008 recession. The second point is an expansion generated by a history of +2.5 standard deviation positive shocks over the previous year, which generates a similarly-sized increase in aggregate investment and is comparable to, e.g., the mid-1980s boom. I compute the response of aggregate investment to a +1 standard deviation shock starting from these two points. Since the model is nonlinear, I replicate this procedure over many simulations of the model and then take the average of these responses.

The left panel of Figure 5 shows the same shock generates more than double the amount

FIGURE 5: Procyclical Impulse Responses of Aggregate Investment



Notes: left panel plots the impulse responses to a one standard deviation positive shock to aggregate TFP computing following Koop, Pesaran and Potter (1996). "Expansion" refers to a history of four 2.5 standard deviation shocks and "recession" refers to history of four -2.5 standard deviation negative shocks. "Partial equilibrium" refers to aggregating firms' decisions holding prices fixed at their steady state values and "general equilibrium" refers to the full general equilibrium model. Since the model is nonlinear, I compute the impulse responses by (i) drawing a random series of aggregate shocks, (ii) adding the history of shocks to generate an expansion and recession, (iii) computing the difference between the simulation in which there is the additional shock and the original simulation, and (iv) repeating this procedure 200 times and taking the average of all the differences produced in step (iii).

Right panel plots how the adjustment probability for firms (conditional on the average realization of idiosyncratic productivity ε) responds to a positive aggregate shock starting from steady state. The blue line (measured against the right axis) plots the steady state distribution of firms over capital k. The red lines (measured against the left axis) plot the probability paying the fixed cost and adjusting capital. The solid red line is in steady state and the dashed line is following a one standard deviation positive TFP shock (with prices held fixed at their steady state values).

of investment starting from the expansion than starting from the expansion. The cumulative differences over time are smaller than the impact difference due to intertemporal substitution; starting from the expansion, some firms pull forward investment they would have done in the future.

Role of Fixed Costs The state dependence in these impulse responses is induced by more firms making an extensive margin investment starting from the expansion. In order to understand this result, first note that in steady state the average firm holds less capital than its target stock $k^a(\varepsilon, k; \mathbf{s})$ because of capital depreciation and convex adjustment costs (see footnote 27 below). Now consider a history of negative shocks which generates a recession.

Since the negative shocks decrease the marginal product of capital, they decrease the target capital stock and therefore bring the average firm closer to its target. In this case, the probability of a firm paying its fixed cost – which is proportional to the adjustment threshold $\widehat{\xi}(\varepsilon, k)$ – falls.[26] Furthermore, additional shocks will have a relatively small effect on the adjustment probability $\widehat{\xi}(\varepsilon, k)$.

On the other hand, a history of positive shocks will move the average firm even further below its target, i.e. induce $k^n(\varepsilon, k; \mathbf{s}) << k^a(\varepsilon, k; \mathbf{s})$, and increase the adjustment probability $\widehat{\xi}(\varepsilon, k; \mathbf{s})$. In this region of the state space, it turns out that further changes in $k^a(\varepsilon, k; \mathbf{s})$ have larger effects on the adjustment probability $\widehat{\xi}(\varepsilon, k; \mathbf{s})$. Hence, the fact that the adjustment probability $\widehat{\xi}(\varepsilon, k; \mathbf{s})$ is increasing in the distance from target $|k^a(\varepsilon, k; \mathbf{s}) - k^n(\varepsilon, k; \mathbf{s})|$ is the key source of procyclical responses to shocks (see Caballero and Engel (2007)).

The right panel of Figure 5 plots how the adjustment probabilities of firms respond to a positive productivity shock starting from steady state (conditional on the average realization of idiosyncratic productivity ε). Consistent with this discussion, the adjustment probability is increasing in $|k^a(\varepsilon, k; \mathbf{s}) - k^n(\varepsilon, k; \mathbf{s})|$.[27] The positive shock increases the target capital stock for all firms and therefore shifts the adjustment probability function up and to the right.

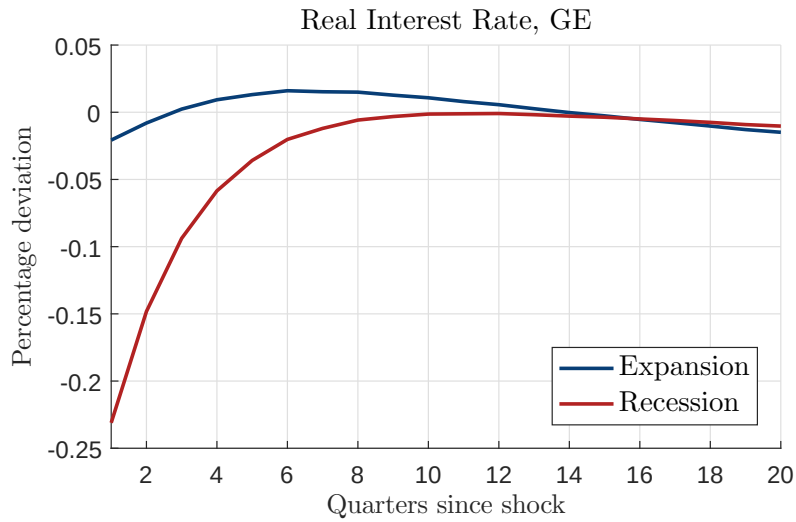
5.2 Role of Prices in General Equilibrium

The left panel of Figure 5 shows that the procyclical responses described above survive in general equilibrium. The shock generates 23% more investment upon impact starting from the expansion than starting from the expansion than starting from the recession. The procyclicality of the impact response is lower in general equilibrium for two reasons. First, the real wage w_t is procyclical, which decreases the marginal revenue product of capital in response to the shock. Second, as Figure 6 shows, the real interest rate r_t falls by more in the recession than in the expansion, reflecting the fact that marginal utility growth falls by more

[26] Since firms' draws of the fixed cost ξ are i.i.d., for each value of productivity and capital (ε, k) a fraction $\widehat{\xi}(\varepsilon, k; \mathbf{s})$ of firms will adjust while the remaining fraction will not.

[27] The adjustment probability function is positive throughout the distribution of firms due to the existence of convex adjustment costs φ. Convex adjustment costs imply that, for high enough levels of capital k, the target capital stock is a decreasing function of current capital since capital decreases the marginal adjustment cost.

FIGURE 6: Countercyclical Responsiveness of the Real Interest Rate



Notes: impulse response of the real interest rate to a one standard deviation positive shock to aggregate TFP computed following Koop, Pesaran and Potter (1996). "Expansion" refers to a history of four +2.5 standard deviation shocks and "recession" refers to history of four -2.5 standard deviation negative shocks. Since the model is nonlinear, I compute the impulse responses by (i) drawing a random series of aggregate shocks, (ii) adding the history of shocks to generate an expansion and recession, (iii) computing the difference between the simulation in which there is the additional shock and the original simulation, and (iv) repeating this procedure 200 times and taking the average of all the differences produced in step (iii).

starting from the recession. However, the procyclicality of the cumulative response is actually higher in general equilibrium than in partial equilibrium. This amplification is due to the fact that persistently low real interest rates following the shock weaken the intertemporal substitution motive for firms to pull forward their investment into the period of the shock.

In order to quantify the amount of time-variation in the impulse response function over a long simulation of the model, I follow Caballero and Engel (1999) and Bachmann, Caballero and Engel (2013) and compute the "impact responsiveness index" RI_t

$$RI_t = 100 \times \log \left(\frac{I(z_t + \sigma_z, X_t, \mu_t) - I(z_t, X_t, \mu_t)}{I(\sigma_z, X^*, \mu^*) - I(0, X^*, \mu^*)} \right), \tag{15}$$

where $I(z, X, \mu)$ is aggregate investment given the aggregate state $\mathbf{s} = (z, X, \mu)$.[28] This responsiveness index measures the impact effect of a TFP shock at a given point in time,

[28] Bachmann, Caballero and Engel (2013) use a more general measure that accounts for asymmetries in the response to a positive and negative shock. These asymmetries are small in my model, so I ignore them here for the sake of simplicity.

Table 7
Fluctuations In Responsiveness Index Over Time

	Impact RI_t				Cumulative $\widehat{RI}_t$	
	95-5 ratio	90-10 ratio	$\rho(RI_t, Y_t)$	$\rho(RI_t, adj_t)$	95-5 ratio	90-10 ratio
Benchmark Calibration (PE interest elasticity $d\log I_t/dr_t = -6.73$)						
Partial eq.	58.5%	44.9%	0.72	0.97	26.1%	19.6%
General eq.	26.3%	20.3%	0.99	0.65	16.2%	12.7%
GE dampening	55.0%	54.8%			37.9%	35.2%
Khan and Thomas (2008) Calibration (PE interest elasticity $d\log I_t/dr_t = -1075.83$)						
Partial eq.	1032%	419%	0.29	0.62	168%	106%
General eq.	10.0%	7.9%	0.99	0.72	2.4%	2.0%
GE dampening	99.0%	98.1%			98.6%	98.1%

Notes: impact responsiveness index RI_t defined in (15) and cumulative responsiveness index $\widehat{RI}_t$ defined in (16) of the main text. adj_t computes the fraction of firms which pay their fixed cost. "Partial equilibrium" refers to aggregating firms' decision rules holding prices fixed at their steady state values. "General equilibrium" refers to the full general equilibrium model. "Benchmark calibration" refers to the calibrated model. "Khan and Thomas (2008) Calibration" refers to eliminating the convex adjustment costs ($\varphi \to 0$), reducing the fixed costs ($\overline{\xi} = 0.0083/4$), changing the idiosyncratic shock process ($\rho_\varepsilon = 0.859^{1/4}$ and $\sigma_\varepsilon = 0.022/4$), increasing the returns to scale ($\theta = 0.256$), eliminating the tax code ($\tau = 0$), eliminating habit formation ($\lambda = 0$), and using separable preferences between consumption and labor supply ($\log C_t - \chi \frac{N_t^{1+\eta}}{1+\eta}$). "PE interest elasticity" is the effect of a one-time unexpected change in the real interest rate starting from steady state.

relative to the effect starting from steady steady state. I also compute the "cumulative responsiveness index" $\widehat{RI}_t$

$$\widehat{RI}_t = 100 \times \log\left(\frac{\widehat{I}(z_t + \sigma_z, X_t, \mu_t) - \widehat{I}(z_t, X_t, \mu_t)}{\widehat{I}(\sigma_z, X^*, \mu^*) - \widehat{I}(0, X^*, \mu^*)}\right), \tag{16}$$

where $\widehat{I}(z_t, X_t, \mu_t)$ is the amount of investment generated over the ten years following the shock. This long-run effect is closely related to the cumulative change in the capital stock and, therefore, output and consumption.

Table 7 shows that the model generates a significant amount of procyclical responsiveness over a long simulation. The impact responsiveness index RI_t is 26% higher than the $5th$ percentile; this variation is positively correlated with aggregate output and the amount of firms making an extensive margin investment, consistent with the mechanism described in Section 5.1. Similarly, the $95th$ percentile of the cumulative responsiveness index $\widehat{RI}_t$ is 16%

higher than the $5th$ percentile.

Relationship to Proposition 1 These quantitative results indicate that the model is far from the limiting case of Proposition 1, in which fixed costs are irrelevant for aggregate dynamics. In that case, small but extremely procyclical movements in the real interest rate are able to bring the dynamics of investment in line with the desires of the representative household because investment is extremely sensitive to changes in interest rates. In my calibrated model, the semi-elasticity of aggregate investment with respect to the real interest rate is only -6.73, so general equilibrium has a smaller influence of aggregate investment dynamics.[29]

Table 7 roughly replicates Khan and Thomas (2008)'s calibration at the quarterly frequency and compares it to the predictions of my model.[30] The partial equilibrium version of this calibration generates enormous variation in both responsiveness indices; the $95th$ percentile of the impact responsiveness index RI_t is ten times as large as the $5th$ percentile, while the $95th$ percentile of the cumulative responsiveness index $\widehat{RI}_t$ is four times as large as the $5th$ percentile. The partial equilibrium variation is so high because, without convex adjustment costs, firms are extremely sensitive to changes in the incentive to invest (as discussed in Section 2.1).[31] However, general equilibrium price movements eliminates roughly 99% of this variation in the responsiveness indices RI_t and $\widehat{RI}_t$. Consistent with the logic of Proposition 1, general equilibrium is powerful in this calibration because investment is extremely interest sensitive; the elasticity of aggregate investment with respect to the real interest rate is approximately -1075. Furthermore, the remaining degree of variation in the responsiveness index RI_t is also present in the version of the model without fixed costs, which

[29] I compute the interest elasticity as the effect of an unexpected increase in the real interest rate starting from the nonstochastic steady state. I assume that the increase in the real interest rate is for one period and is not driven by any aggregate shock in the model (i.e., it is an exogenous increase in the interest rate).

[30] I roughly replicate the Khan and Thomas (2008) calibration by eliminating the convex adjustment costs ($\varphi \to 0$), reducing the fixed costs ($\bar{\xi} = 0.0083/4$), changing the idiosyncratic shock process ($\rho_\varepsilon = 0.859^{1/4}$ and $\sigma_\varepsilon = 0.022/4$), increasing the returns to scale ($\theta = 0.256$), eliminating the tax code ($\tau = 0$), eliminating habit formation ($\lambda = 0$), and using separable preferences between consumption and labor supply ($\log C_t - \chi \frac{N_t^{1+\eta}}{1+\eta}$).

[31] The fact that investment is so responsive to shocks in partial equilibrium also implies that the local approximation may become inaccurate. Therefore, the precise numbers for the partial equilibrium Khan and Thomas (2008) calibration should be interpreted with caution. I set the convex costs $\varphi = 0.005$ to generate some curvature and help stabilize the partial equilibrium dynamics.

aggregates to a representative firm. Hence, in the Khan and Thomas (2008) model fixed costs are essentially irrelevant for aggregate dynamics relative to the representative firm model.[32]

Koby and Wolf (2019) provide a thorough analysis of the role of interest-sensitivity in driving the irrelevance results of the previous literature and argue that quantitative models should target the interest-sensitivity of investment. My calibration disciplines the interest-sensitivity of investment by separately targeting the volatility of investment and the dynamics of the interest rate in the data. The presence of convex adjustment costs φ is key to dampening the interest-sensitivity because it leads to an upward-sloping marginal cost curve for investment. Koby and Wolf (2019)'s analysis applies to a much broader class of models with a richer set of adjustment frictions (e.g. financial frictions) and in which the relative price of investment is not tied as directly to the real interest rate as it is in my model. They discipline the price-sensitivity of investment by targeting the response of investment to the Bonus Depreciation Allowance estimated in Zwick and Mahon (2017). While my strategy is less direct than Koby and Wolf (2019), it does not require the additional structure necessary to accommodate Zwick and Mahon (2017)'s difference-in-differences empirical specification. In addition, the precise dynamics of the real interest rate are quantitatively relevant in shaping the aggregate implications of micro-level frictions in models that feature a realistic price-sensitivity of investment. For example, as discussed above, the fact that the real interest rate is persistently low following a TFP shock amplifies the variation in the cumulative responsiveness index.

Role of Model Ingredients Table 8 decomposes the role of three key model ingredients in driving the procyclical responsiveness to shocks in my model. First, decreasing the size of the fixed costs decreases the variation in both responsiveness indices, which is natural given that the fixed costs are a key source of the state dependence (as described in Section 5.1).

Second, decreasing the size of the convex adjustment cost also decreases the variation in the responsiveness indices. This comparative static balances two opposing forces. On the one hand, lower convex costs make the extensive margin more responsive to shocks and, therefore, increases the degree of state dependence in partial equilibrium (see the partial

[32] In fact, Khan and Thomas (2008) show that the entire distribution of aggregate investment rates is nearly identical in the two models.

Table 8
Role of Key Model Ingredients

		Impact RI_t		Cumulative $\widehat{RI}_t$	
	PE $d\log I_t/dr_t$	95-5 ratio	90-10 ratio	95-5 ratio	90-10 ratio
Full model	-6.73	26.3%	20.3%	16.2%	12.7%
Smaller fixed costs	-7.70	21.8%	16.9%	6.2%	5.0%
Smaller convex costs	-8.38	18.7%	14.5%	10.5%	8.3%
Smaller habit	-6.73	25.9%	19.8%	15.8%	12.2%

Notes: impact responsiveness index RI_t defined in (15) and cumulative responsiveness index $\widehat{RI}_t$ defined in (16) of the main text. "Smaller fixed costs" keeps all parameters the same as the full model except decreases the upper bound on the distribution of fixed costs $\bar{\xi}$ by 50%. "Smaller convex costs" keeps all parameters the same as the full model except decreases the convex adjustment cost φ by 50%. "Smaller habit" keeps all parameters the same as in the full model except decreases the habit formation parameter λ by 50%. "PE $d\log I_t/dr_t$" is the effect of a one-time unexpected change in the real interest rate in steady state.

equilibrium Khan and Thomas (2008) calibration in Table 7). On the other hand, lower convex costs also increase the interest-sensitivity of investment and, therefore, decrease the degree of state dependence in general equilibrium. This latter force quantitatively dominates the former force in the calibrated model.

Third, decreasing the strength of habit formation slightly reduces the variation in the responsiveness indices, but the effect is minor. This comparative static also balances two forces. First, weakening habit formation reduces the countercyclicality of the real interest rate, which in turn decreases the degree of state dependence. Second, weakening habit formation also decreases the procyclical responsiveness of the real interest rate to shocks shown in Figure 6. To understand this result, first note that the fact that investment is more responsive to shocks in expansions implies that consumption growth is more responsive in expansions as well. In turn, the fact that consumption growth is more responsive to shocks in expansions implies that marginal utility growth – and therefore the real interest rate – is also more responsive in expansions, which dampens the procyclical responsiveness of investment. Quantitatively, the strength of this mechanism is increasing in the strength of habit formation. While it is presumably possible to reverse engineer a functional form for habit formation that would eliminate these nonlinearities (see, for example, Campbell and Cochrane (1999)), I do not pursue that approach here given that the ultimate impact on

aggregate investment dynamics is small.

All that said, it important to emphasize that habit formation plays a crucial role in generating realistic interest rate dynamics and micro-level investment behavior. As Figure 3 in Section 4.2 shows, eliminating habit formation in my calibration would generate counterfactually procyclical interest rate dynamics. Matching the dynamics of the real interest rate with adjustment costs alone would require significantly larger adjustment costs and, therefore, generate counterfactual implications for investment. Hence, habit formation gives the model enough degrees of freedom to jointly match interest rate dynamics and investment behavior.[33]

Appendix F contains three additional robustness checks on the results in this section. First, it shows that the results also hold with separable preferences over consumption and labor supply. Second, it shows that the results hold when the tax code is eliminated. Third, it shows that the results are robust to changing the returns to scale. Gourio and Kashyap (2007) argue that lowering the returns to scale has a strong effect on the degree of state dependence; these results suggest that, conditional on a given interest-sensitivity of investment, the returns to scale are relatively unimportant in generating state dependence.

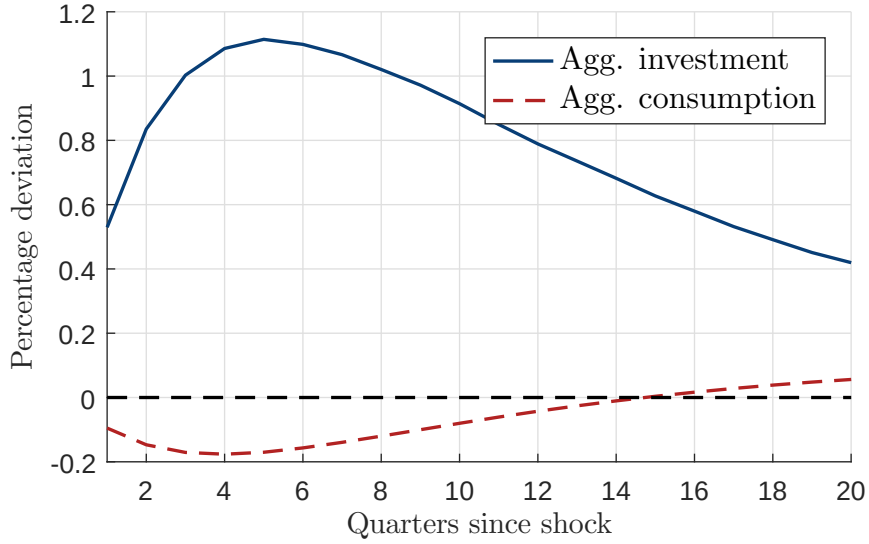
6 Implications for Stimulus Policy

In this section, I briefly study two implications of the model for investment stimulus policies. First, as with productivity shocks, the effectiveness of investment stimulus is state dependent and falls in recessions.[34] Second, the importance of the extensive margin at the micro level implies that targeting firms by their size can increase cost effectiveness up to 30% compared to existing size-independent policies. The goal of this second exercise is to illustrate how taking the micro-level behavior of investment seriously can affect policy design; the exercise abstracts from other considerations, such as financial frictions, which may affect the correlation of firms' size and their probability of adjustment.

[33] Koby and Wolf (2019)'s analysis avoids this issue by directly targeting the price-sensitivity of investment using variation in taxes rather than variation in the real interest rate. Indeed, they present a model without habit formation which nonetheless generates state-dependent impulse responses.

[34] State dependence in the effect of policy does not necessary follow from the results in Section 5 because policy shocks have different general equilibrium implications than productivity shocks.

FIGURE 7: Average Impulse Response to Investment Stimulus Shock



Notes: impulse response of aggregate investment and aggregate consumption to a one-standard deviation positive investment stimulus shock ω.

I model investment stimulus as an exogenous shock to the tax-adjusted price of capital, $q(\mathbf{s})$, derived in Proposition 2. In this section only, I assume that the relative price is $q(\mathbf{s}) = 1 - \tau(PV(\mathbf{s}) + \omega)$, where ω is the investment stimulus shock. Appendix G shows that the two most common investment stimulus policies in the U.S., the investment tax credit and the bonus depreciation allowance, map into different values for the shock ω. I assume that the shock ω follows an AR(1) process:

$$\omega' = \rho_\omega \omega + \varepsilon'_\omega,$$

where $\varepsilon_\omega \sim N(0, \sigma_\omega^2)$. I choose the standard deviation of the shock $\sigma_\omega = 0.035$ to roughly match the size of a 50% Bonus Depreciation Allowance and the quarterly autocorrelation $\rho_\omega = 0.91$ to match a half-life of two years.

Figure 7 plots the impulse responses of aggregate investment and consumption to a one standard deviation positive stimulus shock starting from steady state. The shock immediately decreases the relative price of investment $q(\mathbf{s})$, which then increases investment. Since output is fixed upon impact, this higher investment must be met with lower consumption. Note that the stimulus shock is isomorphic to an investment-specific technological shock, which also in-

Table 9
Responsiveness Index for Investment Stimulus Shock

	95-5 ratio	90-10 ratio	75-25 ratio	$\rho(RI_t, \log Y_t)$
Impact RI_t^ω	31.1%	24.6%	10.5%	0.89
Cumulative $\widehat{RI}_t^\omega$	16.7%	12.9%	6.1%	0.89

Notes: time-variation in the responsiveness index for the stimulus policy shock. Impact responsiveness index RI_t^ω defined in (15) and cumulative responsiveness index $\widehat{RI}_t^\omega$ defined in (16) of the main text.

duces a negative comovement between consumption and investment. Over time, persistently higher investment increases the capital stock, output, and therefore consumption.

State Dependent Effect of Investment Stimulus In order to quantify the degree of state dependence in response to the stimulus shock, I construct two policy responsiveness indices in the spirit of Section 5. The first measures the impact effect of the shock:

$$RI_t^\omega = 100 \times \log\left(\frac{I(z_t, \omega_t + \sigma_\omega, X_t, \mu_t) - I(z_t, \omega_t, X_t, \mu_t)}{I(0, \sigma_\omega, X^*, \mu^*) - I(0, 0, X^*, \mu^*)}\right),$$

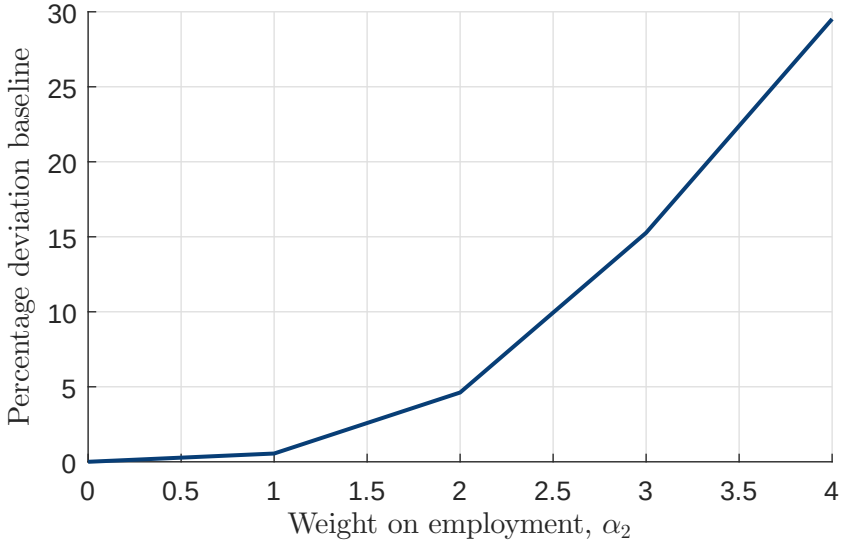
where $I(z_t, \omega_t, X_t, \mu_t)$ is aggregate investment given the expanded aggregate state vector $\mathbf{s}_t = (z_t, \omega_t, X_t, \mu_t)$. The second index measures the cumulative effect of the shock:

$$\widehat{RI}_t^\omega = 100 \times \log\left(\frac{\widehat{I}(z_t, \omega_t + \sigma_\omega, X_t, \mu_t) - \widehat{I}(z_t, \omega_t, X_t, \mu_t)}{\widehat{I}(0, \sigma_\omega, X^*, \mu^*) - \widehat{I}(0, 0, X^*, \mu^*)}\right),$$

where $\widehat{I}(z_t, \omega_t, X_t, \mu_t)$ is the cumulative amount of investment starting from aggregate state $\mathbf{s}_t$ and reverting back to study state.

Table 9 shows that both the impact and cumulative effect vary significantly over time; the $95th$ percentile of the impact effect is 31% higher than the $5th$ percentile, and the $95th$ percentile of the cumulative effect is nearly 17% higher than the $5th$ percentile. Both indices are positively correlated with output, implying that the effectiveness of policy falls in recessions. A linear forecasting model, such as a VAR, would abstract from this state dependence and therefore be biased up in recessions.

FIGURE 8: Cost Effectiveness of Size-Dependent Policy



Notes: aggregate investment generated by a size-dependent investment stimulus shock $\omega \times \alpha_1 n(\varepsilon, k; \mathbf{s})^{\alpha_2}$. For each value of α_2, I solve for the value of α_1 that makes the aggregate cost of the subsidy equal to the aggregate cost with $\alpha_1 = 1, \alpha_2 = 0$.

Increasing Cost Effectiveness with Micro-Targeting A general issue with investment stimulus policies is that much of their cost is due to subsidizing investment that would have been done even without the policy. Therefore, an important challenge to policymakers is identifying and disregarding this inframarginal investment. A key insight of my model is that, because investment primarily occurs along the extensive margin, most of the wasted subsidy to inframarginal investment is accounted for by subsidizing firms that would have made an extensive margin investment even without the policy. This insight simplifies the problem to identifying these inframarginal firms.

Figure 8 shows that a simple size-dependent implementation of this idea is quantitatively powerful. I now assume that the change in a given firm's relative price of investment $q(\mathbf{s})$ is

$$\omega \times \alpha_1 n(\varepsilon, k; \mathbf{s})^{\alpha_2},$$

where $n(\varepsilon, k; \mathbf{s})$ is the firm's employment and α_2 captures the weight of the policy on large firms. For each value of α_2, I solve for the value of α_1 that makes the aggregate cost of the policy equal to the aggregate cost under the size-independent policy, i.e. with $\alpha_1 = 1$ and

$\alpha_2 = 0$. Figure 8 shows that increasing the weight on large firms α_2 increases the amount of investment generated by the policy up to 30%. This occurs because, as in the data, small firms grow faster than the average firm (due to mean reversion in idiosyncratic productivity ε). In order to grow, these firms are more likely to invest, making them more likely to be inframarginal to the policy.[35]

7 Conclusion

In this paper, I have argued that accounting for the importance of the extensive margin in micro-level investment decisions matters for our understanding of aggregate investment dynamics because it implies that aggregate investment is more responsive to shocks in expansions than in recessions. Matching the dynamics of the real interest rate is key to generating this result; in an otherwise standard RBC model, counterfactual movements in the real interest rate eliminate most of this procyclical responsiveness. More generally, these results show that data on interest rates place strong discipline on the role of general equilibrium in determining the aggregate implications of firm-level investment behavior.

[35] Of course, the quantitative effect of this size-dependent policy relies on this particular model of firm growth. Clementi and Palazzo (2016) have used a similar model to study firms' lifecycles and argue that it provides a good fit to the data. However, other models may have different implications for the correlation between size and responsiveness to investment stimulus. For example, a model with financial frictions may imply that small firms are more likely to be financially constrained and therefore are more responsive to the policy. The goal of my exercise here is simply to illustrate the magnitude of the cost savings associated with micro-targeting firms along the extensive margin rather than strongly advocate for this particular size-dependent implementation.

References

Algan, Y., O. Allais, and W. D. Haan (2008): "Solving Heterogeneous-Agent Models with Parameterized Cross-Sectional Distributions," *Journal of Economic Dynamics and Control*, 32, 875–908.

Bachmann, R., R. J. Caballero, and E. M. Engel (2013): "Aggregate implications of lumpy investment: new evidence and a DSGE model," *American Economic Journal: Macroeconomics*, 5(4), 29–67.

Bachmann, R., and L. Ma (2016): "Lumpy investment, lumpy inventories," *Journal of Money, Credit and Banking*, 48(5), 821–855.

Bayer, C., and V. Tjaden (2016): "Large open economies and fixed costs of capital adjustment," *Review of Economic Dynamics*, 21, 125 – 146.

Beaudry, P., and A. Guay (1996): "What do interest rates reveal about the functioning of real business cycle models?," *Journal of Economic Dynamics and Control*, 20(9-10), 1661–1682.

Berger, D., and J. Vavra (2015): "Consumption dynamics during recessions," *Econometrica*, 83(1), 101–154.

Boldrin, M., L. J. Christiano, and J. D. Fisher (2001): "Habit persistence, asset returns, and the business cycle," *American Economic Review*, 91(1), 149–166.

Caballero, R. J., and E. M. Engel (1999): "Explaining investment dynamics in US manufacturing: a generalized (S, s) approach," *Econometrica*, 67(4), 783–826.

——— (2007): "Price stickiness in Ss models: New interpretations of old results," *Journal of monetary economics*, 54, 100–121.

Caballero, R. J., E. M. Engel, J. C. Haltiwanger, M. Woodford, and R. E. Hall (1995): "Plant-level adjustment and aggregate investment dynamics," *Brookings papers on economic activity*, pp. 1–54.

Campbell, J. (1998): "Entry, Exit, Embodied Technology, and Business Cycles," *Review of Economic Dynamics*, 1, 371–408.

Campbell, J. Y., and J. H. Cochrane (1999): "By force of habit: A consumption-based explanation of aggregate stock market behavior," *Journal of Political Economy*, 107(2), 205–251.

Chetty, R., A. Guren, D. Manoli, and A. Weber (2011): "Are micro and macro labor supply elasticities consistent? A review of evidence on the intensive and extensive margins," *American Economic Review*, 101(3), 471–75.

Christiano, L. J., M. Eichenbaum, and C. L. Evans (2005): "Nominal rigidities and the dynamic effects of a shock to monetary policy," *Journal of Political Economy*, 113(1), 1–45.

Clementi, G. L., and B. Palazzo (2015): "On The Calibration of Competitive Industry Dynamics Models," .

Clementi, G. L., and B. Palazzo (2016): "Entry, Exit, Firm Dynamics, and Aggregate Fluctuations," *American Economic Journals: Macroeconomics*, 8, 1–41.

Cooper, R., and J. Willis (2014): "Discounting: Investment Sensitivity and Aggregate Implications," *Unpublished manuscript*.

Cooper, R. W., and J. C. Haltiwanger (2006): "On the nature of capital adjustment costs," *The Review of Economic Studies*, 73(3), 611–633.

Den Haan, W. (2010): "Assessing the Accuracy of the Aggregate Law of Motion in Models with Heterogeneous Agents," *Journal of Economic Dynamics and Control*, 1, 79.

Doms, M., and T. Dunne (1998): "Capital adjustment patterns in manufacturing plants," *Review of Economic Dynamics*, 1(2), 409–429.

Edge, R. M., and J. B. Rudd (2011): "General-equilibrium effects of investment tax incentives,"

Journal of Monetary Economics, 58(6-8), 564–577.

Fernald, J. (2014): "A quarterly, utilization-adjusted series on total factor productivity," Federal Reserve Bank of San Francisco.

Gourio, F., and A. K. Kashyap (2007): "Investment spikes: New facts and a general equilibrium exploration," *Journal of Monetary Economics*, 54, 1–22.

Greenwood, J., Z. Hercowitz, and G. W. Huffman (1988): "Investment, Capacity Utilization, and the Real Business Cycle," *American Economic Review*, 78(3), 402–17.

House, C. L. (2014): "Fixed costs and long-lived investments," *Journal of Monetary Economics*, 68, 86–100.

House, C. L., and M. D. Shapiro (2008): "Temporary investment tax incentives: Theory with evidence from bonus depreciation," *American Economic Review*, 98(3), 737–68.

Khan, A., and J. K. Thomas (2003): "Nonconvex factor adjustments in equilibrium business cycle models: do nonlinearities matter?," *Journal of monetary economics*, 50(2), 331–360.

——— (2008): "Idiosyncratic shocks and the role of nonconvexities in plant and aggregate investment dynamics," *Econometrica*, 76(2), 395–436.

Koby, Y., and C. Wolf (2019): "Investment Stimulus: From Micro Elasticities to Macro Counterfactuals," *Unpublished manuscript*.

Koop, G., M. H. Pesaran, and S. Potter (1996): "Impulse Response Analysis in Nonlinear Multivariate Models," *Journal of Econometrics*, 74, 119–147.

Krusell, P., and A. A. Smith, Jr (1998): "Income and wealth heterogeneity in the macroeconomy," *Journal of Political Economy*, 106(5), 867–896.

Miao, J., and P. Wang (2014): "A Q-theory model with lumpy investment," *Economic theory*, 57(1), 133–159.

Reiter, M. (2009): "Solving heterogeneous-agent models by projection and perturbation," *Journal of Economic Dynamics and Control*, 33(3), 649–665.

Syverson, C. (2011): "What determines productivity?," *Journal of Economic Literature*, 49(2), 326–65.

Thomas, J. K. (2002): "Is lumpy investment relevant for the business cycle?," *Journal of Political Economy*, 110(3), 508–534.

Winberry, T. (2018): "A method for solving and estimating heterogeneous agent macro models," *Quantitative Economics*, 9(3), 1123–1151.

Xu, Q., and E. Zwick (2018): "Kinky Tax Policy and Abnormal Investment Behavior," .

Young, E. (2010): "Solving the Incomplete Markets Model with Aggregate Using the Krusell-Smith Algorithm and Non-Stochastic Simulations," *Journal of Economic Dynamics and Control*, 34, 36–41.

Zwick, E., and J. Mahon (2017): "Tax policy and heterogeneous investment behavior," *American Economic Review*, 107(1), 217–48.

Appendix (For Online Publication Only)

A Proof of Proposition 1

Consider the optimization problem of a firm j choosing capital accumulation k_{jt+1} in period t. Conditional on paying the fixed cost $\overline{\xi}$, the choice k_{jt+1} affects the discounted value of the firm's profits through the terms

$$-k_{jt+1} + \frac{1}{1+r_t}\left(z_{t+1}\mathbb{E}[\varepsilon_{jt+1}|\varepsilon_{jt}]k_{jt+1}^{\alpha} + (1-\delta)k_{jt+1}\right).$$

First consider the limiting case $\alpha = 1$ and $\overline{\xi} = 0$; I will discuss convergence to this limit below. In this limiting case, the expression becomes

$$\left[\frac{1}{1+r_t}\left(z_{t+1}\mathbb{E}[\varepsilon_{jt+1}|\varepsilon_{jt}] + (1-\delta)\right) - 1\right]k_{jt+1}. \tag{17}$$

Since the expression (17) is linear in capital accumulation k_{jt+1}, the optimal policy of the firm is to set $k_{jt+1} = 0$ if $\frac{1}{1+r_t}\left(z_{t+1}\mathbb{E}[\varepsilon_{jt+1}|\varepsilon_{jt}] + (1-\delta)\right) - 1 < 0$, set $k_{jt+1} \to \infty$ if $\frac{1}{1+r_t}\left(z_{t+1}\mathbb{E}[\varepsilon_{jt+1}|\varepsilon_{jt}] + (1-\delta)\right) - 1 > 0$, and can be any $k_{jt+1} \in [0, \infty)$ otherwise.

General equilibrium requires that the firm with the highest expected future productivity earns zero variable profits, i.e.

$$1 + r_t = (1-\delta) + z_{t+1}\tilde{\varepsilon}. \tag{18}$$

If $1 + r_t < (1-\delta) + z_{t+1}\tilde{\varepsilon}$, then the firms with $\mathbb{E}[\varepsilon_{jt+1}|\varepsilon_{jt}] = \widetilde{\varepsilon}$ would strictly prefer to let $k_{jt+1} \to \infty$, violating the finite resource constraint.[36] If $1 + r_t > (1-\delta) + z_{t+1}\tilde{\varepsilon}$, then no firms would find it profitable to invest, which would imply $C_{t+1} = 0$ and violate the consumer's Inada condition.

Condition (18) implies that only firms for which $\mathbb{E}[\varepsilon_{jt+1}|\varepsilon_{jt}] = \widetilde{\varepsilon}$ will accumulate capital for the next period; all other firms have strictly lower expected productivity and therefore set $k_{jt+1} = 0$. The choice $k_{jt+1} = \frac{K_{t+1}}{\mu}$ is optimal for the active firms, where μ is the mass

[36] Note that there is a positive mass of such firms because ε_{jt} has finite support.

of firms with $\mathbb{E}[\varepsilon_{jt+1}|\varepsilon_{jt}] = \widetilde{\varepsilon}$ and K_{t+1} is the aggregate capital accumulation implied by the household's Euler equation

$$C_t^{-\sigma} = \beta \left(z_{t+1}\tilde{\varepsilon} + 1 - \delta \right) C_{t+1}^{-\sigma}.$$

Aggregate output in period $t+1$ is therefore $Y_{t+1} = z_{t+1}\tilde{\varepsilon}K_{t+1}$. Hence, aggregate outcomes are identical to a representative firm with production function $Y_{t+1} = z_{t+1}\tilde{\varepsilon}K_{t+1}$. Note that average productivity among active firms in period $t+1$ is $\tilde{\varepsilon}$ by the law of large numbers.

Now consider the general firm's problem with $\alpha < 1$ and $\overline{\xi} > 0$. Following the statement of the proposition, let $k_{jt}^*(\alpha)$ be the optimal policy of a firm with productivity $\widetilde{\varepsilon}$, conditional on paying the fixed cost. Further denote the mass of these firms by μ_t, which may be time-varying depending on how many pay the fixed costs. Finally, let

$$\pi_t^*(\alpha) = -k_t^*(\alpha) + \frac{1}{1+r_t}\left(z_{t+1}\widetilde{\varepsilon}k_t^*(\alpha)^\alpha + (1-\delta)k_t^*(\alpha) \right)$$

be the contribution of the capital choice to the value of the firm's discounted profits, net of the fixed cost $\overline{\xi}(\alpha)$.

Now consider the limit as $\alpha \to 1$. The optimal policy $k_{jt}^*(\alpha)$ will converge to the optimal policy with $\alpha = 1$ if the fixed cost does not outweigh flow profits, i.e., $\overline{\xi} \leq \pi_t^*(\alpha)$. Since $\pi_t^*(\alpha) \to 0$ as $\alpha \to 1$, this requires $\overline{\xi} \to 0$. Hence, by the same logic as above, in the limit it must be that $r_t + \delta \to z_{t+1}\widetilde{\varepsilon}$ to ensure that the finite resource constraint of the economy is respected. Since active firms will be indifferent, the choice $\frac{K_{t+1}}{\mu_t}$ will be optimal, and aggregate output will be given by $Y_t = \mu_t \times z_{t+1}\widetilde{\varepsilon}\frac{K_{t+1}}{\mu_t} = z_{t+1}\widetilde{\varepsilon}K_{t+1}$.

B Data

B.1 Data Sources and Variable Definitions

I construct the variables used in the empirical analysis as follows.

- Real interest rate r_t: $400\left(\frac{1+r_t^{\text{nom}}}{1+\pi_{t+1}} - 1\right)$, where r_t^{nom} is the average yield on 90-day Treasury bills (FRED series DTB3) and π_{t+1} is realized CPI inflation (FRED series CPI-

AUSCL).

- Relative price of investment goods q_t: constructed by Riccardo DeCicio (FRED series PIRIC).
- Real GDP Y_t: nominal GDP, quarterly (NIPA Table 1.1.5) divided by GDP deflator (NIPA Table 1.1.9).
- Real consumption C_t: nominal expenditures on consumption goods (NIPA Table 1.1.5) divided by implicit price deflator (NIPA Table 1.1.9) plus nominal expenditures on services (NIPA Table 1.1.5) divided by implicit price deflator (NIPA Table 1.1.9). Converted each to 2009q1 dollars so units are comparable.
- Real investment I_t: nominal expenditures on nonresidential fixed investment (NIPA Table 1.1.5) divided by implicit price deflator (NIPA Table 1.1.9).
- Hours worked N_t: hours of all persons in nonfarm business sector (FRED series HOANBS).
- Total factor productivity z_t: downloaded from FRBSF database https://www.frbsf.org/economic-research/indicators-data/total-factor-productivity-tfp/

B.2 Robustness of Empirical Results

I perform four robustness checks on the empirical results in Table 1. First, I show in Table 10 that the main empirical results hold when I use core CPI rather than headline CPI to correct for inflation. Second, I show in Table 11 that the results hold when inflation expectations are computed from a VAR (rather than realized inflation as in the main text).[37] Third, I show in Figure 9 that the impulse response of this ex-ante real interest rate to a TFP shock is similar to the response of the ex-post real interest rate presented in the main text. Fourth, I show in Table 12 that the results hold when I detrend the data using a linear trend, a bandpass filter, or first differences.

[37] The VAR contains four lags of inflation, output growth, consumption growth, investment growth, and unemployment.

Table 10
Cyclical Behavior of Risk-Free Rate, Core CPI Inflation

	$\sigma(r_t)$	$\rho(r_t, y_t)$	$\rho(r_t, z_t)$
Whole sample	1.21%	0.09	−0.05
(p-value)		(0.17)	(0.41)
No Volcker	1.01%	0.26**	0.01
		(0.00)	(0.87)
Pre-1983	1.50%	−0.09	0.05
		(0.34)	(0.64)
Post-1983	0.92%	0.46***	−0.23***
		(0.00)	(0.01)

Notes: real interest rate measured as the return on 90-day Treasury bills adjusted for core CPI inflation, expressed in annual percentage points. Output measured as real GDP. TFP measured as the aggregate Solow residual. All variables have been HP-filtered and expressed as percentage deviation from an HP trend. "Whole sample" refers to the 1954q1 - 2016q4 time series. "No Volcker" excludes 1979q1 - 1983q4. "Pre-1983" refers to the 1954q1-1982q4 sample. "Post-1983" refers to the 1983q1-2016q4 sample.

Table 11
Cyclical Behavior of Risk-Free Rate, VAR Inflation Expectations

	$\sigma(r_t)$	$\rho(r_t, y_t)$	$\rho(r_t, z_t)$
Whole sample	1.19%	0.01	−0.13**
(p-value)		(0.92)	(0.05)
No Volcker	1.08%	0.12*	−0.11
		(0.07)	(0.11)
Pre-1983	1.16%	−0.06	0.06
		(0.54)	(0.52)
Post-1983	1.21%	0.09	−0.34***
		(0.30)	(0.00)

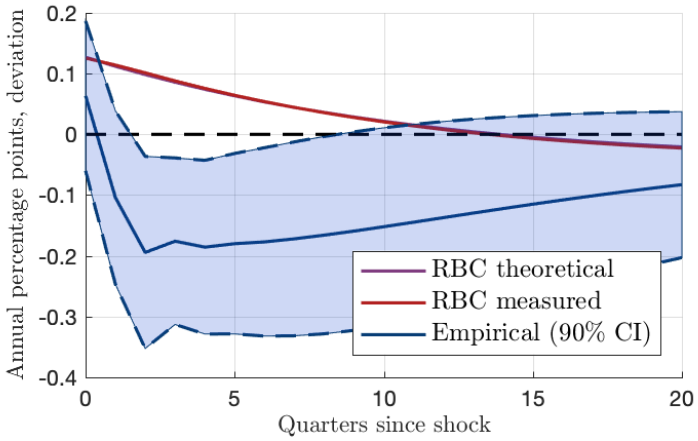
Notes: real interest rate measured as the return on 90-day Treasury bills adjusted for inflation expectations from a VAR, expressed in annual percentage points. The VAR contains four lags of inflation, output growth, consumption growth, investment growth, and unemployment. Output measured as real GDP. TFP measured as the aggregate Solow residual. All variables have been HP-filtered and expressed as percentage deviation from an HP trend. "Whole sample" refers to the 1954q1 - 2016q4 time series. "No Volcker" excludes 1979q1 - 1983q4. "Pre-1983" refers to the 1954q1-1982q4 sample. "Post-1983" refers to the 1983q1-2016q4 sample.

Table 12
Cyclical Behavior of Risk-Free Rate, Different Filters

	$\sigma(r_t)$	$\rho(r_t, y_t)$	$\rho(r_t, z_t)$
HP filter	1.73%	-0.11^{*}	-0.20^{***}
(p-value)		(0.09)	(0.00)
Linear trend	2.58%	0.15^{**}	-0.13^{**}
		(0.02)	(0.03)
Bandpass	1.22%	-0.18^{***}	-0.33^{***}
		(0.01)	(0.00)
First differences	2.58%	0.05	0.05
		(0.39)	(0.41)

Notes: real interest rate measured as the return on 90-day Treasury bills adjusted for realized CPI inflation, expressed in annual percentage points. Output measured as real GDP. TFP measured as the aggregate Solow residual. All variables have been HP-filtered and expressed as percentage deviation from an HP trend. "HP filter" refers to detrending all variables using an HP filter. "Linear trend" refers to removing a linear trend from output and TFP. "Bandpass" refers to removing a bandpass filter from all variables with minimum periodicity of 6 quarters and maximum periodicity of 32 quarters. "First differences" refers to expressing output and TFP in log-differences. All statistics are computed over the 1954q1-2016q4 sample.

Figure 9: Impulse Response of the Ex-Ante Real Interest Rate to TFP Shock



Notes: impulse response of the real interest rate to a TFP shock identified from a bivariate VAR with TFP ordered first. Lag length of 3 chosen by the AIC. Real interest rate measured as the nominal return on 90-day treasury bills adjusted for expected inflation. Expected inflation computed using a VAR with four lags of inflation, output growth, consumption growth, investment growth, and unemployment. "RBC theoretical" refers to the theoretical impulse response. "RBC measured" refers to the impulse response identified using the VAR estimation on simulated data from the model. "Empirical (90% CI)" refers to the empirical impulse response and 90% error bands.

C Benchmark Real Business Cycle Model

There is a representative firm with production function $Y_t = Z_t K_t^{\alpha} N_t^{1-\alpha}$, where Z_t is aggregate productivity, K_t is the aggregate capital stock, and N_t is labor supply. Aggregate productivity Z_t follows the log-AR(1) process $\log Z_t = \rho \log Z_{t-1} + \omega_t$, where $\omega_t \sim N(0, \sigma_z^2)$. There is a representative household which has separable preferences over consumption C_t and labor supply N_t represented by the expected utility function $\mathbb{E} \sum_{t=0}^{\infty} \beta^t \left(\frac{C_t^{1-\sigma}}{1-\sigma} - \chi \frac{N_t^{1+\eta}}{1+\eta} \right)$, where χ controls the disutility of labor supply and the $1/\eta$ is the Frisch elasticity.

A model period is one quarter, so I set the discount factor $\beta = 0.99$. I set the elasticity of intertemporal substitution $1/\sigma = 1$ and the Frisch elasticity of labor supply $1/\eta = 2$. I choose the disutility of labor supply χ to ensure that steady state hours worked is 1/3 of available time. I set the labor share $1 - \alpha = 0.64$ and the depreciation rate of capital $\delta = 0.025$. I set the process for aggregate TFP to the standard values $\rho = 0.95$ and $\sigma_z = 0.007$.

I solve the RBC model using a second-order perturbation implemented in `Dynare`. As I describe in Appendix E, I also solve for the aggregate dynamics of the heterogeneous firm model using a second-order perturbation in `Dynare`.

D Characterizing Equilibrium

In this Appendix I characterize the recursive competitive equilibrium defined in Section 3.4. I use this characterization to numerically compute the equilibrium in Appendix E. For the sake of generality, I allow firms that do not pay the fixed cost to choose any investment $i \in [-ak, ak]$. The main text sets $a = 0$.

Firm's Decision Problem I begin by simplifying the firm's decision problem in a series of three propositions. These propositions eliminate two individual state variables, which greatly simplifies the numerical approximation.

For ease of notation, define after-tax revenue net of tax writeoffs:

$$\pi(\varepsilon, k; \mathbf{s}) = \max_n \left\{ (1-\tau) \left(z\varepsilon k^{\theta} n^{\nu} - w(\mathbf{s}) n \right) \right\}$$

By construction, this object does not depend on current depreciation allowances d or the fixed adjustment cost ξ.

I begin by proving Proposition 2 in the main text. This proposition shows that the firm's value function $v(\varepsilon, k, d, \xi; \mathbf{s})$ is linear in the pre-existing stock of depreciation allowances d. I exploit this property in the other propositions to simplify the decision rules. For ease of reading, I restate the proposition below:

***Proposition* 3.** *The firm's value function is of the form* $v(\varepsilon, k, d, \xi; \mathbf{s}) = v^1(\varepsilon, k, \xi; \mathbf{s}) + \tau PV(\mathbf{s})d$ *where* $PV(\mathbf{s})$ *is defined by the recursion* $PV(\mathbf{s}) = \widehat{\delta} + \left(1 - \widehat{\delta}\right) \mathbb{E}\left[\Lambda(z'; \mathbf{s}) PV(\mathbf{s}')\right]$. *Furthermore,* $v^1(\varepsilon, k, \xi; \mathbf{s})$ *is defined by the Bellman equation*

$$v^1(\varepsilon, k, \xi; \mathbf{s}) = \pi(\varepsilon, k; \mathbf{s}) + \max_i \left\{ \begin{array}{c} -(1 - \tau PV(\mathbf{s}))\, i - \frac{\varphi}{2}\left(\frac{i}{k}\right)^2 k - \xi w(\mathbf{s}) 1\{i \notin [-ak, ak]\} \\ +\mathbb{E}[\Lambda(z'; \mathbf{s}) v^1(\varepsilon', (1-\delta)\, k + i, \xi'; \mathbf{s}')] \end{array} \right\} \tag{19}$$

Proof. First, I show that the value function is of the form $v(\varepsilon, k, d, \xi; \mathbf{s}) = v^1(\varepsilon, k, \xi; \mathbf{s}) + \tau PV(\mathbf{s})d$ for some function $v^1(\varepsilon, k, \xi; \mathbf{s})$. I begin by showing that the operator T defined by the right hand side of the Bellman equation maps functions of the form $f(\varepsilon, k, \xi; \mathbf{s}) + \tau PV(\mathbf{s})d$ into functions of the form $g(\varepsilon, k, \xi; \mathbf{s}) + \tau PV(\mathbf{s})d$. Applying T to f, we get:

$$\begin{aligned} T(f)(\varepsilon, k, \xi; \mathbf{s}) &= \pi(\varepsilon, k; \mathbf{s}) + \tau\widehat{\delta} d \\ &+ \max_i \left\{ \begin{array}{c} -\left(1 - \tau\widehat{\delta}\right) i - \frac{\varphi}{2}\left(\frac{i}{k}\right)^2 k - \xi w(\mathbf{s}) 1\{i \notin [-ak, ak]\} \\ +\mathbb{E}[\Lambda(z'; \mathbf{s})(f(\varepsilon', (1-\delta)k + i, \xi'; \mathbf{s}') + \tau PV(\mathbf{s})(1 - \widehat{\delta})(d + i))] \end{array} \right\} \end{aligned}$$

Collecting terms,

$$\begin{aligned} T(f)(\varepsilon, k, \xi; \mathbf{s}) &= \pi(\varepsilon, k; \mathbf{s}) + \tau\left(\widehat{\delta} + (1 - \widehat{\delta})\mathbb{E}[\Lambda(z'; \mathbf{s}) PV(\mathbf{s}')]\right) d \\ &+ \max_i \left\{ \begin{array}{c} -\left(1 - \tau\widehat{\delta} - \tau(1 - \widehat{\delta})\mathbb{E}[\Lambda(z'; \mathbf{s}) PV(\mathbf{s}')]\right) i - \frac{\varphi}{2}\left(\frac{i}{k}\right)^2 k \\ -\xi w(\mathbf{s}) 1\{i \notin [-ak, ak]\} + \mathbb{E}[\Lambda(z'; \mathbf{s}) f(\varepsilon', (1-\delta)k + i, \xi'; \mathbf{s}')] \end{array} \right\} \end{aligned} \tag{20}$$

By the definition of $PV(\mathbf{s})$, we have that

$$\begin{aligned} \tau\left(\widehat{\delta} + (1 - \widehat{\delta})\mathbb{E}[\Lambda(z'; \mathbf{s}) PV(\mathbf{s}')]\right) d &= \tau PV(\mathbf{s}) \\ -\left(1 - \tau\widehat{\delta} - \tau(1 - \widehat{\delta})\mathbb{E}[\Lambda(z'; \mathbf{s}) PV(\mathbf{s}')]\right) i &= -(1 - \tau PV(\mathbf{s}))\, i \end{aligned}$$

Plugging this back into (20) and rearranging gives

$$T(f)(\varepsilon,k,\xi;\mathbf{s}) = \tau PV(\mathbf{s})d+ \underbrace{\pi(\varepsilon,k;\mathbf{s}) + \max_i \left\{ \begin{array}{c} -\left(1-\tau PV(\mathbf{s})\right) i - \frac{\varphi}{2}\left(\frac{i}{k}\right)^2 k - \xi w(\mathbf{s}) 1\left\{i \notin [-ak, ak]\right\} \\ +\mathbb{E}[\Lambda(z';\mathbf{s}) f\left(\varepsilon', (1-\delta)k + i, \xi';\mathbf{s}'\right)] \end{array} \right\}}_{g(\varepsilon,k,\xi;\mathbf{s})}$$

which is of the form $\tau PV(\mathbf{s})d + g(\varepsilon,k,\xi;\mathbf{s})$. Hence, T maps functions of the form $\tau PV(\mathbf{s})d + f(\varepsilon,k,\xi;\mathbf{s})$ into functions of the form $\tau PV(\mathbf{s})d + g(\varepsilon,k,\xi;\mathbf{s})$. This is a closed set of functions, so by the contraction mapping theorem, the fixed point of T must lie in this set as well. Since the fixed point of T is the value function, this establishes that $v(\varepsilon,k,d,\xi;\mathbf{s}) = v^1(\varepsilon,k,\xi;\mathbf{s}) + \tau PV(\mathbf{s})d$.

To derive the form of $v^1(\varepsilon,k,\xi;\mathbf{s})$, plug $v(\varepsilon,k,d,\xi;\mathbf{s}) = v^1(\varepsilon,k,\xi;\mathbf{s}) + \tau PV(\mathbf{s})d$ into both sides of the Bellman equation to get

$$v^1(\varepsilon,k,\xi;\mathbf{s}) + \tau PV(\mathbf{s})d = \pi(\varepsilon,k;\mathbf{s}) + \tau\widehat{\delta}d+ \max_i \left\{ \begin{array}{c} -\left(1-\tau\widehat{\delta}\right) i - \frac{\varphi}{2}\left(\frac{i}{k}\right)^2 k - \xi w(\mathbf{s}) 1\left\{i \notin [-ak, ak]\right\} \\ +\mathbb{E}[\Lambda(z';\mathbf{s})(v^1(\varepsilon', (1-\delta)k + i, \xi';\mathbf{s}') + \tau PV(\mathbf{s})(1-\widehat{\delta})(d+i))] \end{array} \right\}$$

Rearranging terms as before shows that

$$v^1(\varepsilon,k,\xi;\mathbf{s}) + \tau PV(\mathbf{s})d = \pi\left(\varepsilon,k;\mathbf{s}\right) + \tau PV(\mathbf{s})d+ \max_i \left\{ \begin{array}{c} -\left(1-\tau PV(\mathbf{s})\right) i - \frac{\varphi}{2}\left(\frac{i}{k}\right)^2 k - \xi w(\mathbf{s}) 1\left\{i \notin [-ak, ak]\right\} \\ +\mathbb{E}[\Lambda(z';\mathbf{s}) v^1\left(\varepsilon', (1-\delta)k + i, \xi';\mathbf{s}'\right)] \end{array} \right\}$$

Subtracting $\tau PV(\mathbf{s})d$ from both sides establishes (19). ■

The above proposition shows that the depreciation allowances d do not interact with the other state variables of the firm. The next proposition shows that this implies that investment decisions do not depend on d. To ease notation, I first define the ex ante value function:

$$v^0\left(\varepsilon,k;\mathbf{s}\right) = \int_0^{\overline{\xi}} v^1(\varepsilon,k,\xi;\mathbf{s}) \frac{1}{\overline{\xi}} d\xi.$$

***Proposition* 4.** *The investment decision rule is independent of d and given by*

$$i(\varepsilon, k, \xi; \mathbf{s}) = \left\{ \begin{array}{l} i^a(\varepsilon, k; \mathbf{s}) \;\; \textit{if } \xi \leq \widehat{\xi}(\varepsilon, k; \mathbf{s}) \\ i^n(\varepsilon, k; \mathbf{s}) \;\; \textit{if } \xi > \widehat{\xi}(\varepsilon, k; \mathbf{s}) \end{array} \right\}$$

where

$$i^a(\varepsilon, k; \mathbf{s}) = \arg\max_i -(1 - \tau PV(\mathbf{s}))\, i - \frac{\varphi}{2}\left(\frac{i}{k}\right)^2 k + \mathbb{E}[\Lambda(z'; s) v^0(\varepsilon', (1-\delta)\, k + i; \mathbf{s}')$$

$$i^n(\varepsilon, k; \mathbf{s}) = \left\{ \begin{array}{c} ak \;\; \textit{if } i^a(\varepsilon, k; \mathbf{s}) > ak \\ i^a(\varepsilon, k; \mathbf{s}) \;\; \textit{if } i^a(\varepsilon, k; \mathbf{s}) \in [-ak, ak] \\ -ak \;\; \textit{if } i^a(\varepsilon, k; \mathbf{s}) < -ak \end{array} \right\}$$

$$\widehat{\xi}(\varepsilon, k; \mathbf{s}) = \frac{1}{w(\mathbf{s})} \times \left\{ \begin{array}{c} -(1 - \tau PV(\mathbf{s}))(i^a(\varepsilon, k; \mathbf{s}) - i^n(\varepsilon, k; \mathbf{s})) \\ -\frac{\varphi}{2}\left(\left(\frac{i^a(\varepsilon,k;\mathbf{s})}{k}\right)^2 - \left(\frac{i^n(\varepsilon,k;\mathbf{s})}{k}\right)^2\right) k \\ +\mathbb{E}[\Lambda(z'; \mathbf{s})(v^0(\varepsilon', (1-\delta)\, k + i^a(\varepsilon, k; \mathbf{s}); \mathbf{s}') \\ -v^0(\varepsilon', (1-\delta)\, k + i^n(\varepsilon, k; \mathbf{s}); \mathbf{s}'))] \end{array} \right\}$$

Proof. The form of $i^a(\varepsilon, k; \mathbf{s})$ follows directly from the Bellman equation, using the law of iterated expectations and the fact that ξ' is i.i.d. The form of $i^n(\varepsilon, k; \mathbf{s})$ also follows from the Bellman equation, which shows that the objective function in the no-adjust problem is the same as the adjust problem and the choice set is restricted. The form of $i(\varepsilon, k, \xi; \mathbf{s})$ comes from the following argument. At $\xi = 0$, the objective function of adjusting must be weakly greater than the no-adjust problem, again because the no-adjust problem has the same objective function as the adjust problem but has a restricted choice set. Further, the payoff of adjusting is strictly decreasing in ξ. Therefore, there must be a cutoff rule. Setting the adjust and no adjust payoffs equal gives the form of the threshold $\widehat{\xi}(\varepsilon, k; \mathbf{s})$. ■

The above proposition shows that knowing $v^0(\varepsilon, k; \mathbf{s})$ is enough to derive the decision rules. The next and final proposition defines the Bellman equation which determines $v^0(\varepsilon, k; \mathbf{s})$.

***Proposition* 5.** *$v^0(\varepsilon, k; \mathbf{s})$ solves the Bellman equation*

$$
\begin{aligned}
v\left(\varepsilon, k; \mathbf{s}\right) = \pi(\varepsilon, k; \mathbf{s}) \\
&+ \frac{\widehat{\xi}\left(\varepsilon, k; \mathbf{s}\right)}{\overline{\xi}} \left\{ \begin{array}{c} -\left(1 - \tau PV(\mathbf{s})\right) i^a\left(\varepsilon, k; \mathbf{s}\right) - \frac{\varphi}{2}\left(\frac{i^a(\varepsilon,k;\mathbf{s})}{k}\right)^2 k \\ -\frac{\widehat{\xi}(\varepsilon,k;\mathbf{s})}{2} w(\mathbf{s}) + \mathbb{E}[\Lambda(z';\mathbf{s}) v^0(\varepsilon', (1-\delta)k + i^a\left(\varepsilon, k; \mathbf{s}\right); \mathbf{s}')] \end{array} \right\} \\
&+ \left(1 - \frac{\widehat{\xi}\left(\varepsilon, k; \mathbf{s}\right)}{\overline{\xi}}\right) \left\{ \begin{array}{c} -\left(1 - \tau PV(\mathbf{s})\right) i^a\left(\varepsilon, k; \mathbf{s}\right) - \frac{\varphi}{2}\left(\frac{i^n(\varepsilon,k;\mathbf{s})}{k}\right)^2 k \\ +\mathbb{E}[\Lambda(z';\mathbf{s}) v^0(\varepsilon', (1-\delta)k + i^n\left(\varepsilon, k; \mathbf{s}\right); \mathbf{s}')] \end{array} \right\}
\end{aligned}
$$

Proof. This follows from integrating $v^0\left(\varepsilon, k; \mathbf{s}\right) = \int v^1\left(\varepsilon, k, \xi; \mathbf{s}\right) \frac{1}{\xi} d\xi$, using the expression for $v^1\left(\varepsilon, k, \xi; \mathbf{s}\right)$ from Proposition 3 and the form of the policy function from Proposition 4. ■

A Characterization of the Equilibrium The series of propositions above show that firms' decision rules are determined by the alternative value function $v^0\left(\varepsilon, k; \mathbf{s}\right)$. I now embed this alternative value function into a simplified characterization of the recursive competitive equilibrium. In addition to simplifying firms' decisions, this characterization eliminates household optimization by directly imposing the implications of optimization on firm behavior through prices, as in Khan and Thomas (2008). To do so, define the marginal utility of consumption in state $\mathbf{s}$ as $p(\mathbf{s})$. Abusing notation, I normalize the value function through

$$v(\varepsilon, k; \mathbf{s}) = p(\mathbf{s}) v^0(\varepsilon, k; \mathbf{s})$$

This normalization leaves the decision rules unchanged and I continue to denote them $i^a(\varepsilon, k; \mathbf{s})$, etc. In a final abuse of notation, I denote the distribution of firms over measurable sets $\Delta_\varepsilon \times \Delta_k$ as μ.

***Proposition* 6.** *The recursive competitive equilibrium from Definition 1 is characterized by a list of functions $v\left(\varepsilon, k; \mathbf{s}\right)$, $w(\mathbf{s})$, $p(\mathbf{s})$, $X'(\mathbf{s})$, and $\mu'(\mathbf{s})$ such that*

(i) (Firm optimization) $v(\varepsilon, k; \mathbf{s})$ solves the Bellman equation

$$
\begin{aligned}
v(\varepsilon, k; \mathbf{s}) &= p(\mathbf{s})\pi(\varepsilon, k; \mathbf{s}) \\
&+ \frac{\widehat{\xi}(\varepsilon, k; \mathbf{s})}{\overline{\xi}} \left\{ \begin{array}{c} -p(\mathbf{s})\left(1 - \tau PV(\mathbf{s})\right) i^a(\varepsilon, k; \mathbf{s}) - p(\mathbf{s})\frac{\varphi}{2}\left(\frac{i^a(\varepsilon,k;\mathbf{s})}{k}\right)^2 k \\ -p(\mathbf{s})\frac{\widehat{\xi}(\varepsilon,k;\mathbf{s})}{2} w(\mathbf{s}) + \beta\mathbb{E}[v(\varepsilon', (1-\delta)k + i^a(\varepsilon, k; \mathbf{s}); \mathbf{s}')] \end{array} \right\} \\
&+ \left(1 - \frac{\widehat{\xi}(\varepsilon, k; \mathbf{s})}{\overline{\xi}}\right) \left\{ \begin{array}{c} -p(\mathbf{s})\left(1 - \tau PV(\mathbf{s})\right) i^n(\varepsilon, k; \mathbf{s}) - p(\mathbf{s})\frac{\varphi}{2}\left(\frac{i^n(\varepsilon,k;\mathbf{s})}{k}\right)^2 k \\ +\beta\mathbb{E}[v(\varepsilon', (1-\delta)k + i^n(\varepsilon, k; \mathbf{s}); \mathbf{s}')] \end{array} \right\}
\end{aligned}
$$

where $i^a(\varepsilon, k; \mathbf{s})$, $i^n(\varepsilon, k; \mathbf{s})$, and $\widehat{\xi}(\varepsilon, k; \mathbf{s})$ are derived from $v(\varepsilon, k; \mathbf{s})$ using

$$
i^a(\varepsilon, k; \mathbf{s}) = \arg\max_i -p(\mathbf{s})\left(1 - \tau PV(\mathbf{s})\right) i - p(\mathbf{s})\frac{\varphi}{2}\left(\frac{i}{k}\right)^2 k + \beta\mathbb{E}[v(\varepsilon', (1-\delta)k + i; \mathbf{s}')
$$

$$
i^n(\varepsilon, k; \mathbf{s}) = \left\{ \begin{array}{c} ak \ \textit{if } i^a(\varepsilon, k; \mathbf{s}) > ak \\ i^a(\varepsilon, k; \mathbf{s}) \ \ \textit{if } i^a(\varepsilon, k; \mathbf{s}) \in [-ak, ak] \\ -ak \ \textit{if } i^a(\varepsilon, k; \mathbf{s}) < -ak \end{array} \right\}
$$

$$
\widehat{\xi}(\varepsilon, k; \mathbf{s}) = \frac{1}{p(\mathbf{s})w(\mathbf{s})} \times \left\{ \begin{array}{c} -p(\mathbf{s})\left(1 - \tau PV(\mathbf{s})\right)(i^a(\varepsilon, k; \mathbf{s}) - i^n(\varepsilon, k; \mathbf{s})) \\ -p(\mathbf{s})\frac{\varphi}{2}\left(\left(\frac{i^a(\varepsilon,k;\mathbf{s})}{k}\right)^2 - \left(\frac{i^n(\varepsilon,k;\mathbf{s})}{k}\right)^2\right) k \\ +\beta\mathbb{E}[(v(\varepsilon', (1-\delta)k + i^a(\varepsilon, k; \mathbf{s}); \mathbf{s}') \\ -v(\varepsilon', (1-\delta)k + i^n(\varepsilon, k; \mathbf{s}); \mathbf{s}'))] \end{array} \right\}
$$

and $PV(\mathbf{s})$ is defined by the recursion

$$
p(\mathbf{s})PV(\mathbf{s}) = p(\mathbf{s})\widehat{\delta} + \left(1 - \widehat{\delta}\right)\beta\mathbb{E}\left[p(\mathbf{s}')PV(\mathbf{s}')|\mathbf{s}\right].
$$

(ii) (Labor market clearing)

$$
\left(\frac{w(\mathbf{s})}{\chi}\right)^{\frac{1}{\eta}} = \int \left(n(\varepsilon, k; \mathbf{s}) + \frac{\widehat{\xi}(\varepsilon, k; \mathbf{s})^2}{2\overline{\xi}}\right) \mu(d\varepsilon, dk)
$$

where $n(\varepsilon, k; \mathbf{s}) = \left(\frac{z\varepsilon k^\theta \nu}{w(\mathbf{s})}\right)^{\frac{1}{1-\nu}}$.

(iii) (Consistency)

$$p(\mathbf{s}) = \left(C(\mathbf{s}) - X(\mathbf{s}) - \chi \frac{\left(\left(\frac{w(\mathbf{s})}{\chi} \right)^{\frac{1}{\eta}} \right)^{1+\eta}}{1+\eta} \right)^{-\sigma}$$

where $C(\mathbf{s})$ *is derived from the decision rules by* $C(\mathbf{s}) = \int (z\varepsilon k^{\theta} n(\varepsilon, k; \mathbf{s})^{\nu} - i(\varepsilon, k; \mathbf{s}) - AC(\varepsilon, k; \mathbf{s}))\mu(d\varepsilon, dk)$ *using* $i(\varepsilon, k; \mathbf{s}) = \frac{\widehat{\xi}(\varepsilon,k;\mathbf{s})}{\overline{\xi}} i^{a}(\varepsilon, k; \mathbf{s}) + \left(1 - \frac{\widehat{\xi}(\varepsilon,k;\mathbf{s})}{\overline{\xi}}\right) i^{n}(\varepsilon, k; \mathbf{s})$ *and*

$$AC(\varepsilon, k; \mathbf{s}) = \frac{\widehat{\xi}(\varepsilon, k; \mathbf{s})}{\overline{\xi}} \left(\frac{\varphi}{2} \left(\frac{i^{a}(\varepsilon, k; \mathbf{s})}{k} \right)^{2} k \right) + \left(1 - \frac{\widehat{\xi}(\varepsilon, k; \mathbf{s})}{\overline{\xi}} \right) \left(\frac{\varphi}{2} \left(\frac{i^{n}(\varepsilon, k; \mathbf{s})}{k} \right)^{2} k \right).$$

(iv) (Law of motion for habit stock)

$$X'(\mathbf{s}) = \lambda \left(C(\mathbf{s}) - \chi \frac{\left(\left(\frac{w(\mathbf{s})}{\chi} \right)^{\frac{1}{\eta}} \right)^{1+\eta}}{1+\eta} \right)$$

(v) (Law of motion for measure) For all measurable sets $\Delta_{\varepsilon} \times \Delta_{k}$,

$$\mu'(\mathbf{s})(\Delta_{\varepsilon} \times \Delta_{k}) = \int p(\varepsilon' \in \Delta_{\varepsilon} | \varepsilon) \Big(\frac{\widehat{\xi}(\varepsilon, k; \mathbf{s})}{\overline{\xi}} 1\{(1-\delta)k + i^{a}(\varepsilon, k; \mathbf{s}) \in \Delta_{k}\} + \left(1 - \frac{\widehat{\xi}(\varepsilon, k; \mathbf{s})}{\overline{\xi}}\right) 1\{(1-\delta)k + i^{n}(\varepsilon, k; \mathbf{s}) \in \Delta_{k}\} \Big) d\varepsilon' \mu(d\varepsilon, dk)$$

Proof. Condition (i) follows from Propositions 3 - 5, using the definition $v(\varepsilon, k; \mathbf{s}) = p(\mathbf{s}) v^{0}(\varepsilon, k; \mathbf{s})$ and noting that $\Lambda(z'; \mathbf{s}) = \frac{\beta p(\mathbf{s}')}{p(\mathbf{s})}$. Condition (ii) follows from the household's FOC, the firms' FOC, and labor market clearing. Condition (iii) follows from output market clearing and the definition of $p(\mathbf{s})$. Condition (iv) directly reproduces conditions iv(c) and iv(d) from Section 2.4 in the main text. Condition (v) follows from the original law of motion in condition iv(e) in the main text, eliminating d as an individual state variable and integrating out ξ. ■

E Solution Method

I solve the model using the method concurrently developed in Winberry (2018). I provide a brief overview of the method in this appendix and refer to the interested reader to Winberry (2018) for details. Broadly, the method involves three key steps. First, for each period t I approximate the equilibrium objects – including the cross-sectional distribution of firms – using a finite-dimensional parametric approximation. Second, I solve for the steady state of this discretized model in which there are no aggregate shocks (but there are still idiosyncratic shocks). Third, I solve for the dynamics of the discretized model by perturbing it around this steady state.

The main challenge in applying the method is approximating the value function $v_t(\varepsilon, k)$ and distribution $\mu_t(\varepsilon, k)$ in the first step. I approximate the value function using a weighted sum of Chebyshev polynomials, indexed by the vector of weights θ_t.[38] I approximate the density function of the distribution, denoted $g(\varepsilon, \log(k))$, using the parametric family

$$g\left(\varepsilon, \log(k)\right) \cong g_0 \exp\{g_1^1\left(\varepsilon - m_1^1\right) + g_1^2\left(\log(k) - m_1^2\right) + \tag{21}$$

$$\sum_{i=2}^{n_g}\sum_{j=0}^{i} g_i^j\left[\left(\varepsilon - m_1^1\right)^{i-j}\left(\log(k) - m_1^2\right)^j - m_i^j\right]\}, \tag{22}$$

where n_g indexes the degree of approximation, $\left\{g_i^j\right\}_{i,j=(1,0)}^{(n_g,i)}$ are parameters, and $\left\{m_i^j\right\}_{i,j=(1,0)}^{(n_g,i)}$ are centralized moments of the distribution. The fact that the parameters and moments must be consistent with each other implies that the parameters $\mathbf{g}_t$ are pinned down by the moments $\mathbf{m}_t$. I then approximate the law of motion of the distribution using the law of motion for these moments. With all of these approximations, the discretized equilibrium of the model is characterized by a sequence of state vectors $\mathbf{x}_t = (\mathbf{m}_t, X_t, z_t)$ and control vectors $\mathbf{y}_t = (\theta_\mathbf{t}, \mathbf{g}_t, p_t, w_t)$ which satisfy

$$\mathbb{E}_t[f(\mathbf{x}_t, \mathbf{x}_{t+1}, \mathbf{y}_t, \mathbf{y}_{t+1})] = \mathbf{0},$$

where f is a function returning equilibrium condition residuals. This is a standard canonical

[38] The notation in this discussion follows the exposition of Winberry (2018), which provides further details.

Table 13
Forecast Accuracy Based on Aggregate Capital

	Maximum DH	Mean DH	R^2
Capital accumulation K_{t+1}	1.2%	0.3%	0.999
Marginal utility $(\hat{C}_t - X_t)^{-1}$	3.5%	0.3%	0.996
Real wage w_t	13.0%	1.0%	0.989

Notes: results from forecasting using the system of equations (23) - (25). "Maximum DH" refers to the maximum absolute difference between realized series and series forecasting by iterating on (25) for $10,000$ periods (as suggested by Den Haan (2010)). "Mean DH" refers to mean absolute difference between these two series. "R^2" refers to simple R^2 of the regressions.

form in the perturbation literature and Winberry (2018) shows how it can be solved using `Dynare`. For the analysis of state-dependence in Sections 5 and 6, I solve the model using a second-order perturbation in order to capture the nonlinear aggregate dynamics. However, for the remaining analysis, I solve the model using a first-order perturbation. A first-order perturbation features the same average behavior and is computationally feasible enough to perform the calibration. I have verified that the features of the model targeted in the calibration are nearly indistinguishable in a first- vs. second-order calibration.

Table 13 shows that "approximate aggregation" does not hold in this model. The table reports results from the forecasting equations

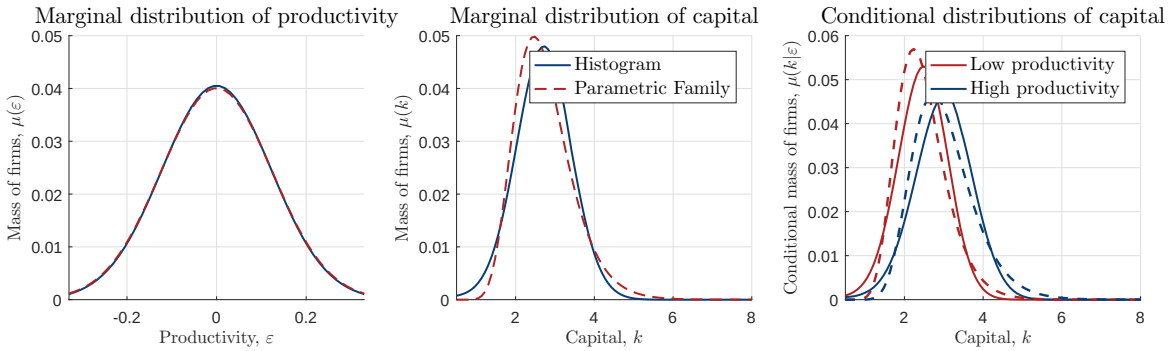
$$\log(\hat{C}_t - X_t)^{-1} = \alpha_0^C + \alpha_1^C \log z_t + \alpha_2^C K_t \tag{23}$$

$$\log w_t = \alpha_0^C + \alpha_1^C \log z_t + \alpha_2^C K_t \tag{24}$$

$$\log K_{t+1} = \alpha_0^C + \alpha_1^C \log z_t + \alpha_2^C K_t. \tag{25}$$

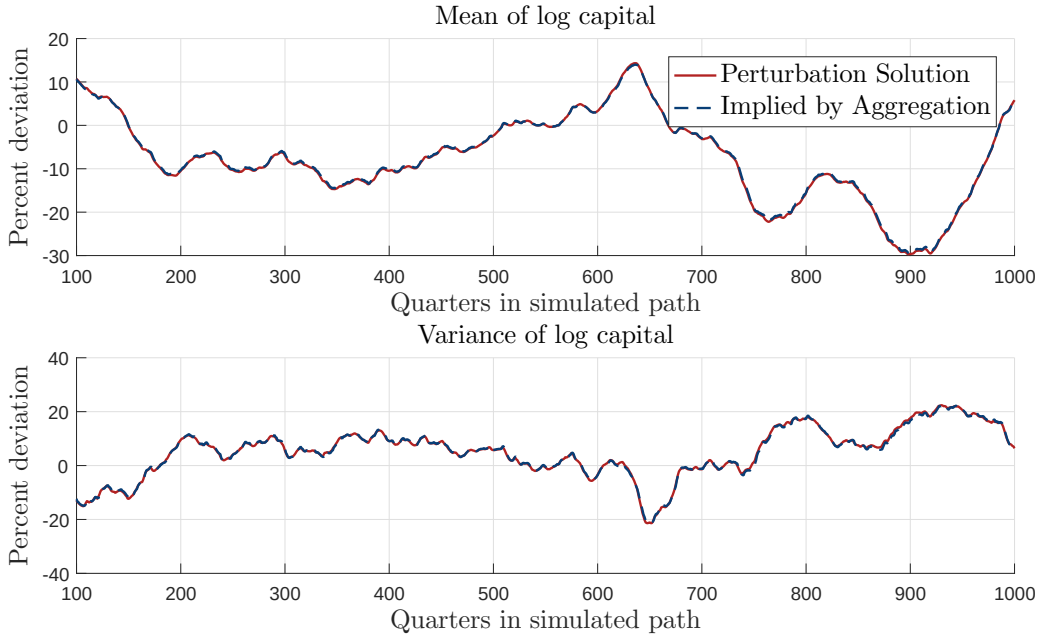
If approximate aggregation holds, then forecasts of the path of prices (marginal utility and the real wage) based on equations (23) - (25) would be extremely accurate. Following Den Haan (2010), I assess the forecasting power of this system by iterating (25) forward $T = 10,000$ periods to compute a path of capital and then using equations (23) and (24) to compute an implied path of prices. Table 13 shows that the implied forecasts are at times substantially different than the actual prices which occur in equilibrium, which suggests that applying Krusell and Smith (1998)'s methodology would require adding additional moments

FIGURE 10: Steady State Distribution, Histogram vs. Parametric Family



Notes: steady state distribution of firms. "Histogram" is the steady state distribution computed using a fine histogram instead of the parametric family. "High productivity" and "low productivity" correspond to approximately +/− two standard deviations of the distributions of idiosyncratic productivity shocks.

FIGURE 11: Accuracy of Distribution Dynamics



Notes: results from $T = 10,000$ quarter simulation. In each quarter, I use the state variable implied by the approximated solution and compute two objects: (i) the next quarter's moments $\mathbf{m}_{t+1}$ implied by the approximation ("perturbation solution") and (ii) the actual moments in the next quarter computed by aggregating the decision rules exactly ("implied by aggregation"). Lines are percentage deviation from steady state values.

to accurately summarize the distribution. This approach would be computationally costly and render the simulation-based calibration in Section 4.1 infeasible.

My method remains accurate even in the absence of approximate aggregation because it

Table 14
Alternative Model Specifications

	Impact RI_t			Cumulative $\widehat{RI}_t$	
	95-5 ratio	90-10 ratio	$\rho(RI_t, Y_t)$	95-5 ratio	90-10 ratio
Full model	26.3%	20.3%	0.99	16.2%	12.7%
Separable preferences	15.2%	12.0%	0.95	8.3%	6.4%
No taxes	18.2%	13.9%	0.99	10.4%	8.6%
Lower returns to scale	19.7%	15.3%	0.99	7.1%	5.8%

Notes: impact responsiveness index RI_t defined in (15) and cumulative responsiveness index $\widehat{RI}_t$ defined in (16) of the main text. adj_t computes the fraction of firms which pay their fixed cost. "Full model" refers to the calibrated model. "Separable preferences" refers to the preference specification $\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left(\log(C_t - X_t) - \chi \frac{N_t^{1+\eta}}{1+\eta} \right)$. "No taxes" refers to setting $\tau = 0$. "Lower returns to scale" refers to setting $\theta = 0.16$. All variables have been HP-filtered with smoothing parameter $\lambda = 1600$.

approximates the entire distribution of firms. Of course, the key restriction in this approximation is that the distribution is contained within the parametric family (22). Figure 10 shows that the parametric distribution in steady state is a tight fit to the true stationary distribution (computed using a fully nonparametric histogram, which is feasible in steady state).[39] In order to assess the accuracy of the dynamics of the distribution, I simulate the model for $T = 10,000$ quarters and, for each quarter t in the simulation, compute two objects: (i) the next quarter's moments $\mathbf{m}_{t+1}$ implied by the approximation and (ii) the actual moments in the next quarter computed by aggregating the decision rules exactly. If the method is not accurate, then the actual moments in step (ii) may fall outside the parametric family. Figure 11 shows that this is not the case; the series (i) and (ii) are nearly indistinguishable, indicating that the true distribution stays within the parametric family (22) in response to aggregate shocks. The correlation between the two series is over 0.997 in all cases.

F Business Cycle Analysis Appendix

Table 14 shows that the key results from Table 7 in the main text hold true in three alternative specifications of the model. First, the results hold if households have separable preferences

[39] I compute the histogram over a fine grid following the approach of Young (2010).

over consumption and labor supply represented by[40]

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left(\log(C_t - X_t) - \chi \frac{N_t^{1+\eta}}{1+\eta} \right).$$

Second, the results hold if there are no taxes, which is more directly comparable to Khan and Thomas (2008). Third, the results are also similar when using lower returns to scale $\theta = 0.16$ than in the paper $\theta = 0.21$. My reading of this result is that, conditional on generating a similar interest-sensitivity of investment, the exact degree of returns to scale is relatively unimportant.

G Policy Analysis Appendix

In this appendix, I show how common investment stimulus policies can be mapped into the investment stimulus shock ω defined in the main text.

Institutional Details I begin with a brief description of the U.S. corporate tax code. Firms pay taxes on their revenues net of business expenses. Most of those expenses are for nondurable inputs such as labor, energy, or materials. These nondurable inputs are fully deducted from the firm's tax bill because they are completely used in the fiscal year. However, since capital is a durable good, investment expenses are deducted over time. The schedule for these deductions is given by the IRS's Modified Accelerated Cost Recovery System, or MACRS.

Historically, there have been two main implementations of investment stimulus policies in the US: the investment tax credit, which was often used before the 1986 tax reform, and the bonus depreciation allowance, which has been used as countercyclical stimulus in the last two recessions. In order to understand how these policies work, consider the example of a firm which purchases \$1000 in computer equipment.[41] Table 15 reproduces the depreciation schedule for this \$1000 purchase under three regimes: the standard MACRS schedule, a 50%

[40] I found that using the value of the habit formation parameter $\lambda = 0.75$ with these preferences implies rather unstable aggregate dynamics. Therefore, I set $\lambda = 0.5$ for this exercise, which implies stable dynamics and a roughly similar response of the real interest rate to a TFP shock as in the main parameterization.

[41] This example draws heavily from Table 1 in Zwick and Mahon (2017).

Table 15
Tax Depreciation Schedule

Standard MACRS Schedule (No policy)								
Year	0	1	2	3	4	5	Total	PV, 7%
Deductions	200	320	192	115	115	58	1000	890
50% Bonus Depreciation								
Year	0	1	2	3	4	5	Total	PV, 7%
Deductions	500+100	160	96	57.5	57.5	29	1000	945
5% Investment Tax Credit								
Year	0	1	2	3	4	5	Total	PV, 7%
Deductions	$\frac{50}{35\%}$+190	304	182.4	109.3	109.3	55	1093	1093

Notes: tax depreciation schedule for purchase of $1000 computer. Top panel: standard schedule absent stimulus policy. Middle panel: 50% bonus depreciation allowance. Bottom panel: 5% investment tax credit. Present value computed using 7% discount rate. Example drawn heavily from Table 1 in Zwick and Mahon (2017).

bonus depreciation allowance, and a 5% investment tax credit.

First consider the standard MACRS schedule. The schedule specifies that the recovery period for a computer is five years and also specifies the fraction of the purchase that can be written off each each of those years. This fraction declines over time to reflect the economic depreciation of the computer. At the end of five years, the firm will have written off the full $1000 purchase.[42]

Now consider how the schedule changes under the two investment stimulus policies. The 50% bonus depreciation allowance allows the firm to immediately deduct 50% of the $1000, or $500. The firm then applies the standard MACRS schedule to the remaining $500. Hence, the bonus does not change the total amount of tax writeoffs, just their timing. Since more writeoffs are taken in the present, the bonus increases the present value of tax deductions, making investment more attractive to the firm. The 5% investment tax credit reduces the firm's tax bill by 5% of the $1000, or $50; expressed in terms of tax writeoffs, this is $50/35%, where 35% is the example tax rate. The firm then applies the standard schedule to the remaining $950. The investment tax credit thus increases both the total amount of tax deductions and the present value of these deductions, making investment more attractive.

[42]This discussion abstracts from the fact that firms do not pay taxes if they make a loss in that fiscal year; I abstract from loss carryforwards/carrybacks for computational simplicity.

Introducing Stimulus Policy into the Model In these two examples, the present value of tax deductions is a useful summary of how various schedules affect the incentive to invest. Proposition 2 shows that, in my model, the present value *completely* characterizes how the tax code affects the incentive to invest through the tax-adjusted price $q(\mathbf{s}) = 1 - PV(\mathbf{s})$. Any changes in tax depreciation allowances can therefore be mapped into changes in this tax-adjusted price, which I defined as ω in the main text. The 50% bonus depreciation allowance is captured by $\omega = 0.5 \times (1 - PV(\mathbf{s}))$; it is as if the firm receives the baseline depreciation schedule on all investment, plus gets an extra subsidy on 50% of its investment. The extra subsidy ω is equal to how much the firm values output in the current period, 1, relative to a stream of output through the depreciation schedule, $PV(\mathbf{s})$. Similarly, the 5% investment tax credit is captured by $\omega = 0.05 \times \left(\frac{1}{\tau} - PV(\mathbf{s})\right)$; the implicit subsidy ω equals how much the firm values the tax writeoff $\frac{1}{\tau}$ relative to the baseline schedule $PV(\mathbf{s})$.