New Keynesian Economics with Household and Firm Heterogeneity*

Thomas Winberry
Wharton and NBER
twinb@wharton.upenn.edu

Adrien Auclert
Stanford and NBER
aauclert@stanford.edu

Matthew Rognlie
Northwestern and NBER
matthew.rognlie@northwestern.edu

Ludwig Straub
Harvard and NBER
ludwigstraub@g.harvard.edu

December 2025

Abstract

The Heterogeneous-Agent New Keynesian literature has revisited the transmission of monetary and fiscal policy to consumption using models where heterogeneous households face idiosyncratic income risk and borrowing constraints. We show that the key lessons from this literature also apply to investment using a model where heterogeneous firms face idiosyncratic productivity risk and financial frictions: constrained firms' investment depends on their free cash flow, generating indirect effects of monetary policy and implying that transfer payments stimulate investment demand. Quantitatively, the strength of these new mechanisms is governed by firms' marginal propensities to invest (MPIs), similar to the role of marginal propensities to consume (MPCs) for households. But unlike MPCs, we currently lack quasi-experimental evidence about MPIs that we can use to directly discipline the new mechanisms.

*In preparation for the *Journal of Economic Literature*. We thank the editor David Romer, five anonymous referees, as well as Adrien Bilal, Mark Gertler, Calvin He, Ben Hebert, Patrick Kehoe, Chen Lian, Hanno Lustig, Yueran Ma, Matteo Maggiori, Emi Nakamura, Pablo Ottonello, Jón Steinsson, David Thesmar, Christian Wolf, and participants at the Stanford Macro lunch, the Penn-Wharton Macro lunch, the Nemmers Prize Conference in Honor of Michael Woodford, and UT Austin for helpful comments. Tomáš Tapák provided exceptional research assistance.

1 Introduction

Over the past few decades, the New Keynesian framework has emerged as a workhorse to study the aggregate effects of monetary and fiscal policy on the economy. Simple versions of the model build on the "three-equation" specification in which consumption is the single endogenous component of aggregate demand (Clarida, Gal and Gertler, 1999; Woodford, 2003, chapter 4, Gal, 2008), while quantitative versions of the model used for policy analysis add investment and other features (Woodford, 2003, chapter 5, Christiano, Eichenbaum and Evans, 2005; Smets and Wouters, 2007). Underlying all these models are the twin assumptions of a representative household and a representative firm. Yet, micro data reveals extensive heterogeneity across individual households and firms, raising a natural question: how does micro-level heterogeneity matter for the transmission of monetary and fiscal policy?

Over the last ten years, the heterogeneous-agent New Keynesian (HANK) literature has developed to answer this question, focusing on the household side (e.g., McKay, Nakamura and Steinsson, 2016; Kaplan, Moll and Violante, 2018; Auclert, Rognlie and Straub, 2024; see Kaplan and Violante, 2018, McKay and Wolf, 2023 and Auclert, Rognlie and Straub, 2025b for recent reviews). Comparatively less work has been done to incorporate heterogeneous firms, with the existing literature primarily focusing on the heterogeneous effects of policy across firms (e.g. Ottonello and Winberry, 2020). From an aggregate perspective, this imbalance is surprising because investment accounts for a large share of changes in output over the business cycle and in response to monetary shocks.

Our goal in this paper is to bring heterogeneous firms centrally into HANK models, exploiting analogies with heterogeneous households. To do so, we develop a canonical heterogeneous-firm model in which firms face idiosyncratic productivity shocks and make investment decisions subject to capital adjustment costs and financial constraints, as in Khan and Thomas (2013) and Ottonello and Winberry (2020). In our model, constrained firms have a positive marginal propensity to invest (MPI), which measures how much investment increases in response to a one-time, unexpected increase in the firm's current cash flow, holding all other determinants of investment fixed.[1]

[1]This definition is parallel to the standard definition of MPCs on the household side. However, it differs from other definitions of MPIs in the literature, such as when current cash flows signal higher demand in

The presence of high-MPI rms activates new mechanisms governing the transmission of monetary and scal policy to investment, just as the presence of high-MPC households activated new mechanisms for consumption. In particular, monetary policy a ects investment not just through traditional direct e ects (lower interest rates decrease rms' discount rates) but also new indirect e ects (equilibrium changes in cash ows a ecting the investment of constrained rms). In addition, de cit- nanced scal transfers to rms stimulate aggregate demand by increasing investment among constrained rms. Given the aggregate importance of investment, characterizing these new mechanisms on the rm side seems just as important as characterizing them on the household side.

Unlike MPCs, there are no quasi-experimental estimates of MPIs for a broad sample of rms that we can use to directly discipline these new mechanisms. Instead, we indirectly discipline the distribution of MPIs by matching other moments and relying on the structure of the model. Going forward, we believe that more empirical work directly estimating MPIs will spur progress in this space, similar to how direct estimates of MPCs have stimulated the development of heterogeneous-household models over the last decade.

We begin our analysis by developing a representative agent benchmark with capital investment. We extend the textbook three-equation New Keynesian model (Clarida et al., 1999, Woodford, 2003, Gal, 2008) to include capital investment subject to standard capital adjustment costs, and use sticky wages rather than sticky prices. The resulting model provides a launching point for incorporating micro-level heterogeneity throughout the paper.

Consumption in this representative agent benchmark follows the permanent income hypothesis (PIH). This property implies that monetary transmission is dominated by direct e ects, i.e. changes in interest rates generating intertemporal substitution. In contrast, indirect e ects resulting from changes in income are small because MPCs out of temporary income changes are low. Furthermore, Ricardian equivalence holds under the PIH, so transfer payments that are paid with future lump-sum taxes have no e ect on consumption.

Investment in the representative agent benchmark follows a version of the Q theory, which has similarly stark predictions for policy transmission. Monetary transmission is dominated

the future (Alati, Fischer, Froemel and Ozturk, 2024), or households' marginal propensity to save in speci c accounts or securities (e.g., Luetticke, 2021).

by direct effects, i.e. changes in interest rates affecting firms' discounting; indirect effects are small because the MPI out of changes in cash flows is zero. Furthermore, because the MPI is zero, definanced transfer payments have no effect on investment and are instead simply paid out as dividends.

Both the PIH model of consumption and Q theory of investment have been routinely rejected by the data. Early work testing the PIH typically found "excess sensitivity," i.e. that consumption is positively correlated with temporary changes in income (e.g. Flavin, 1981, Campbell and Mankiw, 1989; see Jappelli and Pistaferri, 2010 for a review). More recently, well-identified empirical studies have confirmed this finding using exogenous variation in income (e.g. tax rebates in Johnson, Parker and Souleles, 2006, or lottery winners in Fagereng, Holm and Natvik, 2021). On the firm side, early work testing the Q theory typically found that average Q was not a sufficient statistic for investment, as predicted by the theory, with investment highly correlated with cash flows instead (e.g. Fazzari, Hubbard and Petersen, 1988). The promise of heterogeneous agent models is that, by moving beyond the PIH and Q theory, we can develop richer, more realistic models of policy transmission.

In response to the failure of the PIH, the heterogeneous-household literature has focused on models that can match the distribution of MPCs estimated in micro data. We summarize the main takeaways of this literature using a canonical HANK model consistent with that data. As in Kaplan et al. (2018), monetary transmission propagates through indirect effects because high-MPC households are not very sensitive to interest rates but are quite sensitive to changes in income. As a result, the power of monetary policy depends on the initial distribution of MPCs, as well as on the response of fiscal policy and the effects on the labor market. Furthermore, as in Auclert et al. (2024) and Angeletos, Lian and Wolf (2024), high-MPC households break Ricardian equivalence, so definanced fiscal transfers increase consumption even when they must be paid back with future taxes. As a result, transfer payments are a potentially useful source of aggregate stimulus.

Our main contribution is to show that similar insights apply in a canonical heterogeneous-firm model of investment. In the spirit of the firm dynamics literature (Hopenhayn, 1992), heterogeneity in our model is driven by both life-cycle dynamics and idiosyncratic productivity shocks. In the spirit of the Q theory, investment is subject to quadratic capital

adjustment costs. But given the empirical relevance of cash flows to investment, we also assume that firms face financial frictions in the form of collateral constraints (as in the earlier representative-agent work of Kiyotaki and Moore, 1997 and Bernanke, Gertler and Gilchrist, 1999). As a result, constrained firms have a positive MPI because additional cash flows provide additional resources which can be spent on investment; in fact, since firms can use the new capital as collateral for additional borrowing, their MPI can be larger than one.

The new transmission mechanisms in our heterogeneous-firm model are governed by two features of the MPI distribution. To understand these features, consider a deficit-financed transfer to firms. If the transfer is uniformly distributed, it directly increases aggregate investment according to the *unweighted* MPI aggregated across all firms.[2] In our calibration, the unweighted MPI is large (0.50), so the firm transfer multiplier is large—even larger than the household multiplier. Given our assumption of sticky wages, the resulting increase in aggregate demand lowers the real wage, raising firms' cash flows and therefore further stimulating investment. Since the increase in cash flows is in proportion to the firm's initial labor demand, this indirect effect is governed by the *size-weighted* MPI. In our calibration, the size-weighted MPI is small (0.05) because high-MPI firms tend to be small. By the same logic, the indirect effects of monetary policy are also relatively small in our calibration, accounting for approximately 15% of the aggregate response of investment.

We emphasize that the modest size-weighted MPI is the result of our particular calibration, not a general property of these models. In particular, our model features a strong negative correlation between a firm's size and its MPI, akin to how the first generation of household models features a strong negative correlation between a household's income and its MPC. However, direct estimates of the distribution of MPCs then emerged which indicated a much weaker correlation between income and MPCs, and this spurred a next generation of models (e.g. Kaplan and Violante, 2014). On the firm side, it may be that new direct evidence on the distribution of MPIs similarly spurs a next generation of models with larger size-weighted MPIs and therefore stronger indirect effects of monetary policy.

[2] While real-world fiscal transfers to firms generally do not take the form of a simple uniform lump-sum transfer, analyzing this policy is useful for us to highlight the parallel with household transfers. Studying more realistic policies, such as debt forgiveness programs or guaranteed loans, is an interesting avenue for future research.

These firm-level transfers also highlight new mechanisms generated by combining heterogeneous firms with heterogeneous households. In particular, because the firm transfers stimulate aggregate demand, they also raise households' income; constrained households with high MPCs respond by increasing their consumption demand, which further stimulates aggregate demand, and so on. In this sense, our model captures the intuition that providing fiscal support to constrained firms also supports household consumption. Quantitatively, we find that this additional stimulus through combining heterogeneous households and firms accounts for about 20% of the aggregate effect of the fiscal transfer.

These new mechanisms unlocked by incorporating heterogeneous firms into the New Keynesian model are currently as underexplored as their cousins were on the household side ten years ago. To reduce barriers to entry, we have developed code for solving models with both heterogeneous firms and heterogeneous households using the sequence-space approach of Auclert, Bardóczy, Rognlie and Straub (2021). The resulting model, with both sources of heterogeneity and three state variables for firms, can be solved in a matter of minutes.

Outline of paper. The rest of our paper is organized as follows. Section 2 develops the representative agent benchmark and shows that policy transmission is dominated by direct effects. Section 3 incorporates household heterogeneity and reviews the main takeaways of the HANK literature for policy transmission. Section 4 develops a canonical heterogeneous-firm model and shows how borrowing constraints generate positive MPIs for constrained firms. Section 5 describes how we use observed data to discipline MPIs and calls for more empirical research to directly estimate them. Section 6 shows how high-MPI firms generate indirect effects of monetary policy and imply that de cit- nanced fiscal transfers stimulate aggregate investment. Section 7 studies the aggregate demand complementarities that arise from studying heterogeneous households and heterogeneous firms together. Finally, Section 8 summarizes the analysis by discussing various analogies between heterogeneous household and firm models and uses these analogies as a framework to suggest possible avenues for future research. Section 9 concludes.

2 Representative Agent Benchmark

We begin by studying policy transmission in a representative agent benchmark to serve as the foundation of our later models with heterogeneity. We show that this benchmark has stark predictions for policy transmission: monetary policy is driven by direct effects and fiscal transfers have no impact on aggregate spending. While these results are well-understood for consumption, we are not aware of a similar analysis for investment.

2.1 Model

Our benchmark extends the RBC model to incorporate sticky wages.[3] This benchmark is a useful middle ground between the simplest three-equation New Keynesian model, which lacks investment, and the quantitative versions, which add investment together with many other features, such as habit formation or adjustment costs to the flow of investment itself.

Time is discrete and infinite. Aggregate shocks are modeled as MIT shocks, i.e. small unexpected aggregate shocks at date 0 with perfect foresight after date 0. Due to aggregate certainty equivalence, the impulse responses to this model are the same as those in the equivalent model with aggregate uncertainty up to first order (see e.g. Auclert et al., 2021).[4]

Household. There is a representative household with preferences over sequences of consumption C_t and labor supply N_t described by the utility function $\sum_{t=0}^{\infty} \beta^t [u(C_t) - v(N_t)]$. Each period, the household receives the real wage $w_t = W_t/P_t$ per unit of labor N_t, may receive a lump-sum transfer T_t^h from the government, and can invest its wealth A_t in shares of a mutual fund. In turn, the mutual fund holds government bonds and shares of the firm.

The mutual fund provides a real realized return r_t^p per share at time t, so the household's budget constraint in period t is $C_t + A_t = (1 + r_t^p) A_{t-1} + w_t N_t + T_t^h$. Let $q_t \equiv \prod_{s=0}^{t-1} \frac{1}{1+r_{s+1}^p}$

[3] We use sticky wages, rather than sticky prices, for two reasons. First, wages change less than prices in the micro data: wages are reset on average every 12 months (Grigsby, Hurst and Yildirmaz, 2021), while prices are reset on average every 3 to 9 months (Bils and Klenow, 2004, Nakamura and Steinsson, 2008). Second, the opposite assumption of sticky prices but flexible wages generates highly (and counterfactually) countercyclical profits, and the distribution of these profits is well-understood to be key for aggregate outcomes once we add heterogeneity (Broer, Hansen, Krusell and Oberg, 2020, Bilbiie, 2025).

[4] The nonlinear impulse responses of this perfect-foresight model are also related to the generalized impulse responses of the model with aggregate uncertainty to higher order (see Auclert, Rigato, Rognlie and Straub, 2025c). How large these nonlinearities are is an open question, especially in the model with firm heterogeneity.

denote the discount factor associated with the mutual fund's expected returns. Iterating forward and using the transversality condition delivers the intertemporal budget constraint

$$\sum_{t=0}^{\infty} q_t C_t = (1 + r_0^p) A_{-1} + \sum_{t=0}^{\infty} q_t \left(w_t N_t + T_t^h \right) : \tag{1}$$

The nominal wage W_t is sticky and labor is demand-determined at each date t. The household supplies that labor demand N_t, but allows a labor union to negotiate for a new level of the nominal wage. As in Erceg, Henderson and Levin (2000), the wage is reset infrequently according to the mechanism in Calvo (1983), leading to a Phillips curve for nominal wage in ation $\pi_t^w \equiv \frac{W_t}{W_{t-1}} - 1$, which is given to rst order by[5]

$$\pi_t^w = \frac{v'(N_t)}{u'(C_t)} w_t + \beta \pi_{t+1}^w : \tag{2}$$

Mutual fund. The mutual fund invests household wealth in government bonds B_t, which pay a promised real return r_t between time t and $t+1$, and rm shares s_t, which have (real) price p_t at time t and pay a (real) dividend D_t per share. The mutual fund's balance A_t evolves as: $p_t s_t + B_t + (1 + r_t^p) A_{t-1} = (p_t + D_t) s_{t-1} + (1 + r_{t-1}) B_{t-1} + A_t$. Since portfolios are unrestricted, equilibrium imposes the no-arbitrage condition

$$1 + r_t^p = 1 + r_{t-1} = \frac{p_t + D_t}{p_{t-1}}; \quad \text{all } t \geq 1: \tag{3}$$

The discount factors q_t can therefore also be written as $q_t \equiv \prod_{s=0}^{t-1} \frac{1}{1+r_s}$.

Normalizing the total number of shares to $s_t = 1$, the end-of-period mutual fund balance equals the value of the underlying assets $A_t = B_t + p_t$ in each period t. At date 0, when an unexpected shock may occur, the return paid to mutual fund holders must then be given by

$$1 + r_0^p = \frac{p_0 + D_0}{p_{-1}} \frac{p_{-1}}{A_{-1}} + (1 + r_{-1}) \frac{B_{-1}}{A_{-1}}; \tag{4}$$

where $\frac{p_{-1}}{A_{-1}} = 1 - \frac{B_{-1}}{A_{-1}}$ is the fraction of the mutual fund's portfolio allocated to stocks prior to

[5] We assume, as in Erceg et al. (2000), that employment is subsidized to eliminate the monopolistic distortion associated with the wage-setting power of unions.

date 0. The value of a stock at date 0 is given by the present discounted value of dividends

$$p_0 + D_0 = \sum_{t=0}^{\infty} q_t D_t. \tag{5}$$

Firm. There is a representative firm with production function $Y_t = K_{t-1} N_t$, where $+ 1$. The firm hires workers in a frictionless labor market at real wage w_t. Given that our model features nominal rigidities, some form of capital adjustment cost is necessary to limit the movement of investment in response to shocks that affect the real interest rate.[6] Specifically, we assume that the resources required to go from K_{t-1} to K_t are the sum of investment $I_t \quad K_t \quad (1 \quad) K_{t-1}$ and adjustment costs $' \frac{K_t \quad K_{t-1}}{K_{t-1}} \quad K_{t-1}$, where $' ( )$ is a convex function satisfying $' (0) = 0$ so that no adjustment costs are needed to keep the capital stock constant. The firm can also receive lump-sum transfers T_t^f from the government.

The objective of the firm is to maximize firm value (5) by choosing K_t and N_t subject to the flow of funds constraint

$$D_t = Y_t \quad w_t N_t \quad I_t \quad ' \quad \frac{K_t \quad K_{t-1}}{K_{t-1}} \quad K_{t-1} + T_t^f; \tag{6}$$

which determines dividend payments D_t.

Monetary and fiscal policy. The fiscal authority issues government debt B_t and provides lump sum transfers to households T_t^h and firms T_t^f (with negative transfers representing lump-sum taxes). The government budget constraint is $B_t = (1 + r_{t-1}) B_{t-1} + T_t^h + T_t^f$. We assume that the government has no initial debt, $B_{-1} = 0$, and that any change in the path of fiscal policy $\left\{ T_t^h; T_t^f \right\}_{t=0}^{\infty}$ must respect the intertemporal budget constraint:

$$\sum_{t=0}^{\infty} q_t \left(T_t^h + T_t^f \right) = 0. \tag{7}$$

The monetary authority sets the nominal interest rate i_t to achieve a certain path for the

[6] Specifications of the New Keynesian model without adjustment costs suffer pathologies due to the extreme responsiveness of investment to changes in real interest rates; see, for instance, Dupor (2001) and Rupert and Sustek (2019).

ex-ante real interest rate r_t via

$$1 + i_t = (1 + r_t)(1 + \pi_{t+1}); \tag{8}$$

where π_t is consumer price inflation $1 + \pi_t = (1 + \pi^w_t)\frac{w_{t-1}}{w_t}$. This real rate rule, popularized by Woodford (2011), implies that the ex-ante real interest rate is $r_t = r_t$. As is well-known, a real rate rule generates equilibrium multiplicity; we select the unique equilibrium such that aggregates are back to steady state in the long-run.[7] We use this rule because it provides a simple description of monetary policy in our decompositions. It can be thought of as an intermediate case between very responsive monetary policy as under an active Taylor rule, and very passive monetary policy as at the zero lower bound (see e.g. Auclert et al., 2024).

Equilibrium. Given the monetary and fiscal policy $\{r_t; T^h_t; T^f_t\}$, equilibrium is a path of prices $\{w_t; \pi_t; \pi^w_t\}$ and quantities $\{C_t; I_t; K_t; Y_t; D_t; N_t\}$ such that all agents maximize and the goods market clears at all dates: $C_t + I_t + \left[\phi\left(\frac{K_t}{K_{t-1}} - 1\right) + 1\right] K_{t-1} = Y_t$.

2.2 Equilibrium analysis in the sequence space

We now analyze the equilibrium using the sequence-space approach. This approach is particularly useful for us because it allows us to cleanly decompose the different transmission channels of macro policies both in this benchmark and in the heterogeneous agent models.

Consumption function. The representative household maximizes utility subject to its intertemporal budget constraint (1). This yields a set of Euler equations in terms of the realized return r^p_t, which combined with the no-arbitrage equations (3) imply

$$u'(C_t) = \beta(1 + r_t)\, u'(C_{t+1}); \tag{9}$$

[7] To implement this rule in practice, we can let the mutual funds trade nominal government bonds in addition to real bonds, and assume that the nominal bonds are in zero net supply. No arbitrage then implies a Fisher equation $1 + r_t = \frac{1+i_t}{1+\pi_{t+1}}$, which combined with (8) implies $r_t = r_t$. The equilibrium we obtain with our selection rule is the limit of equilibria such that $1 + i_t = (1 + r_t)(1 + \pi_{t+1})$ as β approaches 1 from above. This selection is natural because it implies that monetary policy has no long-run effects.

In addition, combining the household's intertemporal budget constraint (1), the mutual fund revaluation equation (4), the asset pricing equation (5), and the government's intertemporal budget constraint (7), we obtain $\sum_{t=0}^{1} q_t C_t = \sum_{t=0}^{1} q_t (w_t N_t + D_t)$, where recall that $q_t \equiv \prod_{s=0}^{t} \frac{1}{1+r_s}$. Note that government transfers T_t^h drop out of the intertemporal budget constraint because they are paid back with future taxes|an expression of Ricardian equivalence. Hence, the solution to the household problem only depends on the sequence of interest rates $\{r_s\}_{s=0}^{1}$ and real incomes $\{w_s N_s + D_s\}_{s=0}^{1}$. Let $\omega_t = w_t N_t + D_t$ be real income at t.

We denote by $C_t(\{r_s\}, \{\omega_s\}_{s=0}^{1})$ the implied path of aggregate consumption given these objects. This consumption function is useful to understand the forces shaping households' consumption: first, movements in r_t cause an intertemporal substitution effect through the Euler equation (9), and second, movements in total income ω_t cause an income effect through the intertemporal budget constraint (1).

Since we know that consumption returns to steady state, we can also consider another consumption function $\tilde{C}_t(\{r_s\}_{s=0}^{1})$ obtained by solving only the consumption Euler equation (9) given the condition at infinity $C_1 = C_{-1}$. In equilibrium we must have at once that $C_t = C_t(\{r_s\}, \{\omega_s\}_{s=0}^{1}) = \tilde{C}_t(\{r_s\}_{s=0}^{1})$. The function $\tilde{C}_t(\cdot)$ is useful for solving for equilibrium, while $C_t(\cdot)$ is useful for understanding the underlying drivers of consumption.

Investment function. The representative firm maximizes the present discounted value of dividends (5), with dividends given by (6). This delivers a version of the Q theory. Since labor is chosen statically, we have that

$$w_t = MPN_t = (1-\alpha) K_{t-1}^{\alpha} N_t^{-\alpha}. \tag{10}$$

The first order condition for capital can be written as

$$1 + \phi'\left(\frac{K_t}{K_{t-1}} - 1\right) = \frac{1}{1+r_t}\left[MPK_{t+1} + 1 - \frac{\phi}{2}\left(\frac{K_{t+1}}{K_t} - 1\right)^2 + \phi'\left(\frac{K_{t+1}}{K_t} - 1\right)\frac{K_{t+1}}{K_t}\right] \tag{11}$$

where $MPK_{t+1} = \alpha K_t^{\alpha-1} N_{t+1}^{1-\alpha}$ is the expected marginal product of capital at date $t+1$, which depends on the wage at $t+1$ through the choice of labor N_{t+1}.[8] Note that government

[8] Specifically, we have $MPK_{t+1} = \alpha^{\frac{1}{1-\alpha}} K_t^{\frac{1-(\alpha+\nu)}{1-\alpha}} w_{t+1}^{\frac{-\alpha}{1-\alpha}}$.

transfers T_t^f do not enter either rst order condition because they are additively separable from these choices. Hence, the solution to the rm problem only depends on the sequence of interest rates $\{r_s\}_{s=0}^{1}$ and real wages $\{w_s\}_{s=0}^{1}$.

We denote the implied path of investment by $I_t(\{r_s; w_s\})$. Analogously to the consumption function, this investment function allows us to understand the forces shaping rms' investment. First, movements in r_t cause changes in discounting on the right-hand side of (11). Second, movements in the real wage w_t cause changes in the rm's expected marginal product of capital through the choice of labor N_{t+1}.

The rm's investment and hiring decisions determine the paths of aggregate capital and labor, and therefore aggregate output $Y_t = K_{t-1}N_t$. We denote this path as a function of prices by $Y_t(\{r_s; w_s\})$. Similarly, we denote the path of aggregate adjustment costs $\frac{K_t - K_{t-1}}{K_{t-1}} K_{t-1}$ by $_t(\{r_s; w_s\})$.

Equilibrium. Given a path of policies $\{r_t; T_t^h; T_t^f\}$ and recognizing that $r_t = r_t$, the key step in nding equilibrium is to obtain the sequence of real wages $\{w_s\}$ satisfying

$$C_t(\{r_s\}) + I_t(\{r_s; w_s\}) + _t(\{r_s; w_s\}) = Y_t(\{r_s; w_s\}): \tag{12}$$

Once the path of real wages and aggregates is obtained, one can nd wage in ation through (2) and price in ation via $1 + _t = (1 + _t^w)\frac{w_{t-1}}{w_t}$. However, given our real rate rule, the paths of in ation are irrelevant for equilibrium dynamics.[9]

To rst order, we can summarize the behavior of the model in terms of the derivatives of C, I, and Y functions with respect to the paths of prices, which are known as sequence-space Jacobians (Auclert et al., 2021). By the implicit function theorem, the path of real wages $dw = (dw_0; dw_1; :::)$ that results from a change in the path of real interest rates $dr = (dr_0; dr_1; :::)$ is given by

$$dw = (Y_w - I_w - _w)^{-1}(C_r + I_r + _r - Y_r)\,dr: \tag{13}$$

[9] The exible price version of the model is determined by a sequence of real interest rates and wages $\{r_s; w_s\}$ such that the goods market clears and households' labor supply is $u^0(N_t) = w_t u^0(C_t)$ at all t.

The path of any other aggregate, such as output Y_t, follows from $dY = Y_r dr + Y_w dw$. The heterogeneous-agent models that we develop below also satisfy a version of (12), but differ in their aggregate consumption and/or investment functions. Viewing the models in this way makes the problem "modular" in the sense that we can easily swap out either form of heterogeneity, or bring them both together.

2.3 Calibration

We now parameterize the model in order to study the impulse responses to monetary and fiscal policy shocks. For this quantitative work, we choose standard functional forms for utility and adjustment costs:

$$u(C) = \frac{C^{1-\frac{1}{\gamma}}}{1-\frac{1}{\gamma}}; \qquad v(N) = \frac{N^{1+\varphi}}{1+\varphi}; \qquad \Psi(x) = \frac{\psi}{2}x^2. \tag{14}$$

The labor disutility parameters φ and χ are irrelevant for the aggregate dynamics of real variables given sticky wages and our real interest rate rule.

Table 1 summarizes the parameters and targeted moments in our calibration. We set the first four parameters to ensure that steady-state moments match long-run averages in the data. In particular, we set the discount factor $\beta = 0.99$ to match an annualized real interest rate of $r = 4\%$. We set the depreciation rate $\delta = 2.5\%$ quarterly to match the investment-capital ratio, and the labor coefficient $\alpha = 0.6$ to match the labor share in the data. Finally, we calibrate the capital coefficient ν so that there are aggregate diminishing returns to scale, $\nu + \alpha < 1$. This proxies for a return to managerial skill and makes the optimal scale of firms well-defined in the version of the model with heterogeneous productivity (e.g. Lucas, 1978). In the aggregate, decreasing returns generate rents which get capitalized into the value of the firm. We set $\nu = 0.32$, which delivers an annual ratio of capitalized rents to GDP of 2, consistent with estimates of the markup and rent component of U.S. aggregate wealth (see, eg, Auclert, Malmberg, Rognlie and Straub, 2025a).[10]

The remaining two parameters|the elasticity of intertemporal substitution (EIS) $\frac{1}{\gamma}$

[10] In the steady state of our model, aggregate wealth is made of this rent-to-GDP ratio plus the capital-GDP ratio of $K/Y = 2.29$, so the total wealth-to-GDP ratio is equal to 4.29.

Table 1: Parameters and Targets for the Representative-Agent Model

Parameter	Description	Value	Target (all joint)	Value
	Discount rate	0.99	Real interest rate	4%
	Labor exponent	0.6	Labor share of output	60%
	Capital exponent	0.32	Rent/GDP ratio	200%
	Capital depreciation rate	0.025	Investment/Capital ratio	10%
$\frac{1}{}$	EIS	0.8	C IRF to monetary shock	0.2%
	Capital adjustment cost	9	I IRF to monetary shock	1%

Notes: The calibration is at quarterly frequency. The parameters in the left panel are not annualized; the targets on the right are annualized.

and the adjustment cost parameter control elasticities and therefore cannot be disciplined by steady-state moments alone[11]. Instead, we follow a long tradition in the literature of pinning down these parameters using the response of macro aggregates to identified monetary policy shocks; in particular, we match the peak impulse responses of consumption and investment from Christiano et al. (2005). While there exist many alternative ways of identifying monetary policy shocks in the data, these alternatives tend to deliver qualitatively similar aggregate responses.

In response to a typical monetary policy shock, Christiano et al. (2005) find that the real interest rate falls by 50 basis points, consumption rises by 0.2%, and investment rises by 1%[12]. We model their monetary shock as a decline in the real interest rate target $r_t = r - \frac{0.5}{400}(0.5)^t$ for $t \geq 0$, implying an impact effect at date 0 of 50 basis points (annualized) and an autocorrelation of 0.5. We then choose $^{1} = 0.8$ and $= 9$ to match the peak responses of consumption and investment.

The resulting paths for the real interest rate, output, consumption and investment are displayed in Figure 1. Early on, the decline in the real interest rate increases consumption and investment demand, resulting in excess demand in the goods market that must be resolved by a decline in the real wage. Later on, consumption and investment demand have

[11] In the heterogeneous agent models, individual consumption and investment decisions still vary in response to idiosyncratic shocks, even though aggregates are constant. In this sense, heterogeneous-agent models can also let one calibrate elasticities in steady state, though we do not pursue this route for ^{1} or .

[12] Since consumption represents 80% of private spending, this implies that investment makes up over 55% ($= \frac{0.2 \quad 1}{0.8 \quad 0.2+0 \quad .2 \quad 1}$) of the total response of private spending to monetary policy in the data. Hence, it is critical for quantitative models of monetary policy to feature investment as an endogenous component of aggregate demand.

returned to steady state but the capital stock is higher, resulting in excess supply in the goods market that must be resolved by an increase in the real wage.

In our model, as well as our later heterogeneous-agent models, the peak response to a monetary policy shock occurs on impact: the model does not feature ingredients to delay (and reduce) the consumption and investment responses, such as habit formation in consumption or adjustment costs on the change in investment. As a result, our calibrated adjustment costs are higher than typically estimated in the literature.[13]

Calibration approach across models. As we add micro-level heterogeneity in Sections 3 and 4, we will recalibrate the EIS ^{1} and adjustment costs to match the same aggregate responses to a monetary shock. We do so because aggregate data provide important discipline on the aggregate strength of various mechanisms operating at the micro level. In this sense, the exact numbers we target are less important than the fact that we maintain this target across all models we study.

Importantly, the fact that the aggregate response to monetary policy is the same across models does not mean that micro-level heterogeneity is irrelevant for those responses. First, the channels through which monetary policy a ects aggregates are di erent across models, implying that the monetary authority (and researchers) need to consider a wider variety of issues when formulating policy. Second, given these di erent transmission mechanisms, the response to other shocks|in our context, scal transfers|will be di erent as well.

2.4 Policy Transmission in the Representative Agent Benchmark

We now use the consumption and investment functions to decompose the transmission channels of monetary and scal policy.

Consumption and the Permanent Income Hypothesis. We rst consider the channels driving the response of aggregate consumption to a monetary shock. Using the consumption function, $C_t = C_t(\{r_s; l_s\})$, we can write the rst-order impulse response of consumption

[13] See Auclert, Rognlie and Straub (2020) for an example of how humps can be reconciled with household heterogeneity and MPCs.

Figure 1: Monetary Transmission in the Representative Agent Benchmark

Notes: impulse responses of aggregates (left), consumption (middle, blue line) and investment (right, orange line) to a monetary policy shock modeled as a 50bp annualized decline in the real interest rate with quarterly persistence of 0.5, similar to Christiano et al. (2005). All variables are in percent deviations from their steady state values, except the real interest rate which is in annualized percentage points. The middle panel also displays the decomposition of consumption into a direct and indirect effect following equation (15). The right panel also displays the decomposition of investment into a direct and indirect effect by considering the first order effects from equilibrium movements in r_t and MPK_{t+1} on I_t implied by equations (18)–(19).

$dC = (dC_0, dC_1, \ldots)$ as

$$dC = \underbrace{C_r dr}_{\text{direct effect}} + \underbrace{C_w dw}_{\text{indirect effect}} : \tag{15}$$

This consumption decomposition writes the change in consumption as the sum of the effects from the change in the real interest rate, which the literature calls the "direct effect," and the effect from equilibrium changes in wages and dividend, called the "indirect effect."

In the representative-agent model, the sequence-space Jacobians C_r and C_w have simple analytical expressions, which imply:

$$(C_r dr)_t = -\frac{C}{1-\beta}\left(\sum_{s=t}^{\infty} \frac{dr_s}{1+r} + (1-\beta)\sum_{s=0}^{t-1}\sum_{j=s}^{t-1}\frac{dr_j}{1+r}\right) \tag{16}$$

$$(C_w dw)_t = (1-\beta)\sum_{s=0}^{t-1} \beta^s dw_s. \tag{17}$$

The direct effect is the substitution effect: consumption is increased at t by any decline in real interest rates at t or beyond, net of a constant term that ensures that the present discounted value of the direct effect is zero. The scale of this effect is governed by the elasticity of

intertemporal substitution $\frac{1}{\bar{}}$. The indirect e ect is the income e ect: the present value of labor and dividend income rises, which increases the household's permanent income in their intertemporal budget constraint|Friedman (1957)'s Permanent Income Hypothesis (PIH). Each period, the household consumes the annuity value of that income, such that their MPC is equal to 1 .

The middle panel of Figure 1 shows that nearly all of the response of consumption to the monetary shock is driven by the direct e ect; indirect e ects are small because the MPC is 1 = 0:01. To delve even deeper into this result, recall that our real interest rate rule implies that we have the alternative consumption function $C_t(f r_s g)$, which implies $dC_t = \frac{C}{\bar{}} \sum_{s=t}^{P} \frac{1}{1+r} dr_s$. It follows that the indirect e ect in (17) is exactly o set by the second term in the direct e ect in equation (16).

Turning to scal policy, since transfers$f T_t^h; T_t^f g$ do not enter the consumption function, it immediately follows that households do not respond at all to de cit- nanced changes in transfers to either households and rms. Again, this is a re ection of Ricardian equivalence, whose logic is closely tied to the PIH: the present value of any change in transfers is zero, so transfers do not a ect the household's intertemporal budget constraint; hence, the e ective marginal propensity to consume out of current transfers is exactly zero.

We emphasize that the key mechanisms driving these results|the PIH and Ricardian equivalence|have been rejected by the empirical literature (see Jappelli and Pistaferri, 2010 for a survey). Instead, this literature has consistently found large marginal propensities to consume out of transitory windfall income shocks, as well as non-zero marginal propensities to consume out of de cit- nanced government transfers. The heterogeneous-household model in Section 3 resolves these shortcomings.

Investment and the Q theory. We now study the transmission channels of monetary policy to investment. To motivate our decomposition, it is useful to go back to the equation determining investment in (11). Using our functional form for adjustment costs (14), we get

$$\frac{I_t}{K_{t\ 1}} = \frac{1}{\bar{}} \left(\frac{1}{1+r_t} V_{t+1}^0 \right) 1 \ ; \tag{18}$$

where V^0_{t+1}, the marginal value of capital at $t+1$, is given by the recursion

$$V^0_{t+1} = MPK_{t+1}(w_{t+1}; K_t) \quad \frac{I_{t+1}}{K_t} \quad \frac{}{2} \quad \frac{I_{t+1}}{K_t} \quad ^2 + \frac{1}{1+r_{t+1}} \frac{K_{t+1}}{K_t} V^0_{t+2}: \qquad (19)$$

These equations encode the standard forces from the Q theory of investment: equation (18) says that net investment is positive as long as the discounted marginal value of capital (also known as marginal Q, Q_t) exceeds 1, while equation (19) governs the dynamic evolution of Q_t, capturing the present discounted value of the marginal product of capital in future periods.

Equations (18) and (19) clarify the direct and indirect effects of monetary policy in the representative-agent benchmark. The direct effect, dr, stems from the fact that lower real interest rates raise firm value, thereby increasing the incentive to invest.[14] This effect is highest on impact of the real interest rate decline, when the effect of the new path of interest rates on present discounted values is the highest. The indirect effect, dw, stems from the fact that higher real wages result in less hiring, N_{t+1}, and therefore a lower marginal product of capital in period $t+1$. Since this effect only runs through changes in MPK_{t+1}, we call this the indirect effect from the expected marginal product of capital. Note that changes in w_0 have no effect on investment at any date; they affect dividends but not investment.

The right panel of Figure 1 shows that the response of investment to monetary policy is primarily driven by the direct effects. In fact, the direct effects account for slightly more than 100% of the increase in investment, with the indirect effect from the change in the real wage being slightly negative over the first eight quarters of the shock. This occurs because firms incorporate future wages into (19) in a forward-looking way. While the initial period of low real wages would, on its own, raise investment, the following prolonged period of high real wages pushes investment down. However, this effect is quite small, since the path of wages does not significantly change the present discounted value of the marginal products.

Turning to fiscal policy, lump-sum transfers to firms at any date have no effect on investment because those transfers do not interact with the cost or benefit of investment—in other words, the MPI in this representative agent benchmark is 0. Since firms are already at

[14] For instance, consider a purely transitory change in the real interest rate dr. This essentially does not affect the present discounted value of capital V^0. Therefore, we have that $\frac{dI}{K} = \frac{1}{}dr$, so the semielasticity of investment to the real interest rate is approximately $\frac{1}{}$.

an optimum of their investment plan, they instead distribute the proceeds from the transfer back to their shareholders, per equation (6).[15]

Just like the PIH, Q theory was subject to extensive empirical testing starting in the 1980s. A key challenge for that literature is that marginal Q, $V'=(1+r)$, is not directly observable in the data. In the special case with constant returns to scale $+$ = 1, however, marginal Q is proportional to average Q, which we can compute as the rm's market value relative to its capital stock (Hayashi, 1982). In this case, equation (18) predicts that a regression of the investment rate on average Q should have an R^2 of 1 and a coe cient on average Q equal to $1=$. By and large, the literature has found that these types of regressions produce very low R^2 because observables beyond Q, especially cash ows, tend to predict investment (see, e.g., Fazzari et al., 1988).

The investment literature largely proceeded along two branches following the empirical failures of the Q theory. The rst branch interpreted the importance of cash ows as suggestive evidence that the MPI is positive and therefore studied various forms of nancial frictions (e.g. Gomes, 2001, Hennessy and Whited, 2007). We will follow this branch in Section 4 given our focus on indirect e ects due to changes in aggregate demand. The second branch of the literature questioned the assumption that the marginal cost of investment is smoothly increasing in the amount of investment, which is required to arrive at (18). This branch incorporated nonconvex costs of adjusting capital, motivated by the empirical lumpiness of investment at the micro level (e.g. Caballero and Engel, 1999, Cooper and Haltiwanger, 2006, and Khan and Thomas, 2008). While we will not review those models given the constraints of space, studying the role of xed costs in monetary transmission is a potentially fruitful avenue for future research.

3 Heterogeneous Households: A Review of HANK

We now replace the representative household with heterogeneous households facing idiosyncratic income risk and borrowing constraints (Zeldes, 1989, Deaton, 1991, Carroll, 1997;

[15] If we gave rms access to a borrowing and savings vehicle at rate $1+r_t$, the timing of the distribution of dividends would also not be pinned down, a re ection of the Modigliani-Miller theorem.

Bewley, 1977, Huggett, 1996, Aiyagari, 1994.). Given that the material here is relatively standard (see, for instance, Auclert et al., 2025b for a recent review of this literature), our presentation will be fairly succinct and focus on aspects of the model that parallel the findings with heterogeneous firms.

3.1 Model

A mass 1 of ex-ante identical, but ex-post heterogeneous households are subject to idiosyncratic risk to their productivity e and their rate of time discounting β.[16] The joint evolution of (e, β) is governed by a Markov transition matrix Π, and we assume that the cross-sectional distribution of (e, β) starts at the stationary distribution of Π, normalizing $E[e] = 1$. The environment is identical to that in Section 2, except that households' wage income in productivity state e is now $ew_t N_t$, and they face incomplete markets in the sense that the mutual fund payoffs cannot be made contingent on their realization of their idiosyncratic shocks (e, β). In addition, we assume that households face a borrowing constraint, which for simplicity we set to zero, so that the choice of assets a' must satisfy $a' \geq 0$.

The dynamic problem of a household with exogenous state (e, β) and incoming asset position a is described by the Bellman equation:

$$\begin{aligned} v_t(a; e, \beta) = \max_{c, a'} \; & u(c) + \beta E[v_{t+1}(a', e', \beta')] \\ \text{s.t.} \; & c + a' = (1 + r_t^p) a + ew_t N_t + T_t^h \\ & a' \geq 0. \end{aligned} \tag{20}$$

In the representative agent model, the solution was characterized by the Euler equation (9); here, the Euler equation only holds with the weak inequality

$$u'(c_t(a; e, \beta)) \geq (1 + r_{t+1}^p) \beta E[u'(c_{t+1}(a', e', \beta'))].$$

This inequality is strict whenever the borrowing constraint $a' \geq 0$ binds; in this case, con-

[16] Shocks to the discount factor help us jointly match the wealth distribution and average MPCs, as we describe below.

sumption is determined residually from the budget constraint so that $c_t = (1 + r_t^p)a + ew_t N_t + T_t^h$. For these households, consumption responds one-for-one to changes in current income, generating an MPC of 1.

We represent the aggregate consumption sequence from this model with the new consumption function $C_t\left(r_t^p; w_t N_t; T_t^h\right)$. Intuitively, this consumption function can be computed as follows. First, given the sequence of real interest rates $\{r_s^p\}$, labor income $w_t N_t$, and transfers T_t^h, compute households' decision rules by solving (20) as a function of individual states $(a; e;)$. Second, starting from the stationary distribution of households over $(a; e;)$, use the decision rules to iterate forward on the law of motion to compute the path of cross-sectional distributions over time. Finally, compute aggregate consumption by integrating individual consumption rules against the cross-sectional distribution for each t 0.

3.2 Calibration and the MPC vs. wealth tradeo

We now calibrate the model in order to assess its implications for monetary transmission. As discussed in Section 2, the MPC out of temporary changes in income is critical for determining the size of the indirect e ects of monetary policy and the e ects of scal transfers. The quasi-experimental empirical literature has settled on a consensus that the average MPC at a quarterly frequency is about 0.25 (see, e.g., the reviews in Kaplan and Violante, 2022 and Auclert, Bardóczy and Rognlie, 2023).

As discussed in Kaplan and Violante (2022), in a model with a single discount factor, there is a fundamental tension between matching both the average level of the MPC and the aggregate amount of wealth in the economy. In our model, aggregate assets equal the value of rms because steady state government debt is 0; the value of rms to GDP is $\frac{A}{Y} = 4:29$, the sum of the capital-to-GDP and the rent-to-GDP ratio; given the labor share is $\frac{wN}{Y} = 0:6$, aggregate assets are 7.15 times average labor income. At our annualized real interest rate of $r = 4\%$, a single- model calibrated to this amount of wealth features very patient agents that are away from the borrowing constraint and have on average low MPCs, while a model that hits the average MPC features very impatient agents that are close to the borrowing constraint and hold too little wealth on average.

Table 2: Parameters and Targets for the Heterogeneous Household Block

Parameter	Description	Value	Target (all joint)	Value
Markov Chain for $(e;)$ from Auclert et al. (2025b)				
r	Real interest rate	0.001	Real interest rate	4%
$_L$	High discount rate	0.994	Assets/labor income	715%
$_H$	Low discount rate	0.824	Average MPC	025
^{1}	EIS	0.61	C IRF to monetary shock	02%

Notes: parameters in the left panel chosen to match moments in the right panel. The calibration is at quarterly frequency. The parameters in the left panel are not annualized; the targets on the right are annualized. Production side parameters are as in the representative agent benchmark from Section 2. The asset-to-labor income$A=wN$ ratio and real interest rate are targeted to be the same as in that model, so that the aggregate steady-states are identical in both models.

We resolve this tradeo by assuming heterogeneity in, as in Krusell and Smith (1998).[17] We calibrate the model using the Markov chain for income and impatience described in Auclert et al. (2025b): the income process is from Kaplan et al. (2018), featuring a high degree of income risk, and the impatience process is a two-point process for $2 f _L; _H g$, describing the slow evolution of patience across generations as in Krusell and Smith (1998). We set $(_L; _H)$ to simultaneously match an (unweighted) average MPC of $E[MPC] = 0:25$ and an asset to wage income ratio$A=wN = 7:15$ at our calibrated annual$r = 4\%$.

Table 2 contains the parameter values for the heterogeneous household model. In addition to the parameters described above, we also re-calibrate the EIS^{1} so that we continue to match the impact response of aggregate consumption to a monetary shock. With heterogeneous households, this choice requires a signi cantly lower EIS than in the representative agent benchmark (0.61 vs 0.8). The reason is that, as we will show below, the presence of high-MPC households generates an ampli cation mechanism: lower interest rates stimulate investment and therefore income, which in turn increases consumption (see Auclert et al., 2020 and Bilbiie, Kanzig and Surico, 2022). O setting this e ect requires consumption to be less responsive to interest rates in the rst place, i.e., a lower EIS.[18]

[17] Other solutions in the literature certainly exist, with the leading approach being the introduction of a high-return illiquid account that households use for most of their savings, but that is not useful for smoothing period-to-period income uctuations (e.g. Kaplan and Violante, 2014).

[18] If adjustment costs were in nite, so that investment had a completely inelastic response to monetary policy, then the Werning (2015) result of equivalence between the aggregate consumption behavior of the heterogeneous-agent model and representative-agent model would apply and we would not need to recalibrate the EIS. In other words, it is the elasticity of the investment response that activates this ampli cation mechanism. See Auclert et al. (2020), Appendix A, for more discussion of this point.

Figure 2: Savings Decision Rules, MPCs, and Distribution with Heterogeneous Households

Notes: households' decision rules in steady state of heterogeneous-household model calibrated as in table 2. The left panel plots net savings $a'(e,a) - a$ for impatient households ($\beta = \beta_L$) with two different levels of idiosyncratic productivity e as a function of a. The right panel plots the marginal propensities to consume $\frac{\partial c(e,a)}{\partial (1+r)a}$ for these households. The grey bars show the location of the stationary household distribution.

Savings decisions and MPCs. The left panel in Figure 2 visualizes the steady state net savings policy $a'(a,e,\beta) - a$ as a function of households' incoming asset position a, for two types of households with relatively low income e_L and relatively high income e_H (both households in this illustration are impatient, β_L).

The high-income household in red has positive net savings for low initial assets a in order to reduce the probability of becoming borrowing constrained, which generates costly consumption fluctuations. On the other hand, the household has negative net savings for high initial assets a because $\beta_L(1+r) < 1$. Therefore, there is a unique target level of assets $a^*(e,\beta)$ to which the household will converge if their idiosyncratic state (e,β) were to remain constant. This target, or buffer stock, of assets balances their impatience with their desire for precautionary savings against future declines in income (see, eg, Carroll, 1997).

The low-income household in blue dissaves for any positive level of initial assets a. In other words, when a household receives a negative shock and enters this low-income state, they use their previously accumulated buffer stock to limit the drop in their consumption induced by the decline in income.

The right panel of Figure 2 plots the implied MPCs of these agents. We compute the