
DOI: 10.1111/jmcb.12761

FELIPE ALVES
GREG KAPLAN
BENJAMIN MOLL
GIOVANNI L. VIOLANTE

A Further Look at the Propagation of Monetary Policy Shocks in HANK

We provide quantitative guidance on whether and to what extent different elements of Heterogeneous Agent New Keynesian (HANK) models amplify or dampen the response of aggregate consumption to a monetary policy shock. We emphasize four findings. First, the introduction of capital adjustment costs does not affect the aggregate response, but does change the transmission mechanism so that a larger share of indirect effects originates from equity prices rather than from labor income. Second, incorporating estimated unequal incidence functions for aggregate labor income fluctuations leads to either amplification or dampening, depending on the data and estimation methods. Third, distribution rules for monopoly profits that allocate a larger share to liquid assets lead to greater amplification. Fourth, assumptions about the fiscal reaction to a monetary policy shock have a stronger effect on the aggregate consumption response than any of the other three elements.

JEL codes: D14, D31, E21, E52

Keywords: monetary policy, heterogeneous agents, New, Keynesian, unequal incidence, investment adjustment cost, profit distribution, fiscal accommodation, Taylor rule

A RECENT LITERATURE THAT INCORPORATES microheterogeneity into New Keynesian models of the macroeconomy has advanced our

We are grateful to our discussant Gian Luca Benigno for his insightful comments on an early draft, to Fatih Guvenen and Serdar Ozkan for kindly sharing their data, and to Hugo Lhuillier for his excellent research assistance. We also thank many seminar participants for their comments.

FELIPE ALVES is with Bank of Canada (E-mail: FAlves@bank-banque-canada.ca). GREG KAPLAN is a University of Chicago and NBER (E-mail: gkaplan@uchicago.edu). BENJAMIN MOLL is with London School of Economics, CEPR, and NBER (E-mail: b.moll@lse.ac.uk). GIOVANNI L. VIOLANTE is a Princeton University, CEPR, IFS, IZA, and NBER (E-mail: violante@princeton.edu).

Received July 16, 2019; and accepted in revised form August 25, 2020.

understanding of the transmission mechanism of monetary policy.¹ In these Heterogeneous Agent New Keynesian (HANK) models, the general equilibrium effects of an interest rate cut, which operate through an increase in household incomes from higher labor demand, outweigh the direct effects that primarily operate through intertemporal substitution. This pattern of transmission stands in stark contrast to the Representative Agent New Keynesian (RANK) models that served as a point of departure for this literature, in which monetary policy affects aggregate consumption almost exclusively through intertemporal substitution and in which the indirect channel is negligible.

In this new framework, however, the effect of model assumptions and parameterizations on the consumption response to an interest rate cut is less understood. This is because the HANK framework incorporates several (realistic) elements that are either inconsequential or not even well defined in representative agent versions. Examples include the unequal incidence of aggregate fluctuations across households, the distribution of profits and capital gains, the cyclical risk and borrowing capacity, and the fiscal reaction to a monetary expansion. To paraphrase Sims (1980), once we depart from the representative household, we enter the “wilderness of heterogeneous agent macro.”

In an attempt to tame this wilderness, a growing literature starting from Werning (2015) has used stylized versions of HANK models that can be solved analytically to provide theoretical guidance on the model features that determine the extent of propagation (see, for example, Acharya and Dogra 2018, Bilbiie 2017, Debortoli and Gali 2018, Auclert 2019, Bernstein 2019, Bilbiie, Känzig, and Surico 2019). This literature clarifies the channels through which HANK model elements contribute to amplification and dampening. However, little is currently known about which elements are *quantitatively* important departures from RANK models, or whether the insights from these simple analytical models carry through to empirically relevant versions of HANK models.

In this paper, we address this gap by providing some quantitative guidance on the relative importance of different candidate propagation mechanisms for the case of a monetary policy expansion. Our starting point is the two-asset HANK model studied by Kaplan, Moll, and Violante (2018). We first incorporate two additional ingredients that are common in quantitative RANK models, but that were missing in the first generation of HANK models: aggregate capital adjustment costs and a Taylor rule with some degree of smoothing. Next, we use this setup to explore the quantitative impact of three potential amplification channels that determine how the change in aggregate labor, capital, and government (taxes and transfers) income induced by the monetary policy shock is distributed across households. Our main results are as follows.

1. See Guerrieri and Lorenzoni (2011), Oh and Reis (2012), Gornemann, Kuester, and Nakajima (2014), Den Haan, Rendahl, and Riegler (2015), Luetticke (2015), Werning (2015), McKay and Reis (2016), Auclert (2019), McKay, Nakamura, and Steinsson (2016), Ravn and Sterk (2017), Bilbiie (2017), Patterson (2018), Auclert and Rognlie (2018), Kaplan, Moll, and Violante (2018), Bayer, Luetticke, Pham-Dao, and Tjaden (2019), Kekre (2019), Lenel and Kekre (2019), Cui and Sterk (2019), and Berger, Bocola, and Dovis (2019), among others.

Adding capital adjustment costs has a negligible impact on the aggregate consumption response, but changes the dynamics of investment and asset prices dramatically, with interesting effects on the transmission mechanism of a policy rate cut. While most of the consumption response still comes from indirect general equilibrium effects as opposed to direct effects of the real rate change—in line with Kaplan, Moll, and Violante (2018)—adjustment costs alter the relative contribution of labor versus financial income. In particular, they curtail the investment response by increasing the price of capital. Less investment translates into more moderate movements in output and hence in households’ labor income. But, at the same time, the rise in the price of capital boosts financial wealth and hence more of the gains from the monetary expansion accrue to wealthy shareholders.

The partial-adjustment Taylor rule has almost no effect on the aggregate consumption and investment responses. Moreover, it does not seem to matter for the decomposition between different channels and for the distribution of gains from the monetary expansion across households.

Next, we study the different candidate amplification mechanisms based on the insight that in the presence of marginal propensity to consume (MPC) heterogeneity, redistributing resources across households has real effects. Part of our contribution here is to provide simple parameterized functional forms for each channel that are amenable to quantitative analysis, and to discipline these empirically in some cases.

We pay particular attention to the parameterization and estimation of various “incidence functions”—a concept that has also been used by Werning (2015), Auclert and Rognlie (2018), Bilbiie (2017), and Patterson (2018). An incidence function describes a rule for how a time-varying aggregate quantity is allocated across the distribution of households in the economy. It answers questions such as: when aggregate income rises by one percent, how is this additional income distributed across the population? We are interested in short-run incidence functions, for aggregate income fluctuations occurring at the business cycle frequency.

We propose a convenient parameterization for a general class of incidence functions and separately estimate incidence functions for labor income and government transfer income, using various sources of micro data for the United States: the Annual Social and Economic (ASEC) supplement of the Current Population Survey (CPS), the Survey of Consumer Finances (SCF), and tabulated statistics from the Master Earnings File of the Social Security Administration (SSA).

Depending on the data source used to estimate the incidence functions, the unequal distribution of income over the cycle can either dampen or amplify the consumption response to a monetary shock. For example, estimates using ASEC data suggest that households with low permanent income and higher MPCs are the most heavily exposed to fluctuations in aggregate labor income. This leads to an amplification of the aggregate consumption response of 2% to 20% relative to a model with equal incidence.² In contrast, estimates using SSA data *dampen* the effect of a rate cut, relative

2. We also explain that these findings are nonetheless consistent with Patterson (2018), who reports that her estimated incidence function results in amplification of up to 40%.

to an equal incidence benchmark. This is because these data suggest a U-shaped elasticity, that is, exposure is high not only at the bottom of the distribution (where MPCs are high) but also at the top (where MPCs are low), and this second effect dominates thereby resulting in dampening. Interestingly, the effects of unequal incidence are muted in the presence of capital adjustment costs because the smaller reaction of labor income, due to investment responding weakly to the shock, reduces the overall importance of this channel.

Next, we explore the effects of dividends and capital gain distribution. As in the basic New Keynesian model, profits are countercyclical under a monetary policy shock in our model, making them a countervailing force for consumption expansion. In our two-asset model, not only does it matter how this reduction in profits is distributed across households, but also whether profits are paid into households' liquid accounts or retained in their illiquid accounts. We find that if dividends are reinvested into the illiquid account, this significantly dampens the consumption response to the monetary policy shock. After a monetary expansion, intermediate producers' profits fall. Lower profits reduce investment. A weaker investment response, in turn, weakens the general equilibrium effect on household income, which dampens the consumption response.³ These findings further underscore the importance of how monopoly profits are distributed in HANK models. Broer et al. (2016) have emphasized that, even in standard New Keynesian models with the worker–capitalist dichotomy, the income effect of countercyclical profits on labor supply is crucial for the transmission of an interest rate cut. In HANK models, particularly those with both liquid and illiquid assets, the distribution of profits plays an even more critical role.

Kaplan, Moll, and Violante (2018) showed how the consumption response to the shock depends on the reaction of fiscal policy. We further investigate this important dimension of monetary transmission. We model fiscal policy in terms of a rule for the government primary surplus and explore both the role of the timing of the fiscal adjustment (e.g., raising transfers today or in the future) and the role of the choice of fiscal instrument (e.g., transfers, taxes, or government expenditures) to achieve a particular path for the surplus. Different assumptions on the fiscal side lead to the largest changes in aggregate consumption among all the channels we explore, with some scenarios amplifying the consumption response by a factor of two. Moreover, this amplification effect is robust in the sense that it survives the presence of capital adjustment costs and a partial adjustment Taylor rule.

The rest of the paper proceeds as follows. Section 1 describes how we estimate incidence functions. Section 2 outlines the model and calibration strategy. Section 3 collects the results from our experiments. Section 4 concludes the paper.

3. If profits go to the liquid account instead, they also dampen the aggregate consumption response by directly reducing household disposable income. What our quantitative experiments shows, however, is that the negative impact of reducing investments on disposable income is stronger than the direct effect.

1. INCIDENCE FUNCTIONS IN THEORY AND IN THE DATA

An incidence function describes an allocation rule of an aggregate quantity across the distribution of households in the economy. In this section, we first explain why unequal incidence can affect the propagation mechanism of aggregate shocks, and then estimate incidence functions for labor earnings and government transfers from microdata. An important caveat is that we estimate unconditional incidence functions that combine all sources of aggregate fluctuations, rather than incidence functions conditional on monetary shocks. We return to this limitation in the conclusion.

1.1 Unequal Incidence in Theory

To illustrate in more detail the mechanism by which unequal incidence of aggregate fluctuations may lead to amplification, we adopt the reduced-form approach of Bilbiie (2017) and Patterson (2018) and consider the effect of changes in aggregate income on aggregate consumption in a simple static framework (essentially a heterogeneous-agent version of the classical Keynesian cross).

There is a unit continuum of individuals indexed by i . Each individual's consumption c_i depends on her income y_i in a potentially nonlinear fashion $c_i = g_i(y_i, \theta_i)$, where θ_i are other demand shifters. Aggregate consumption is $C = \mathbb{E}_i[c_i]$ and aggregate income is $Y = \mathbb{E}_i[y_i]$ where the expectation operator \mathbb{E}_i computes the cross-sectional average. Consider an aggregate shock, such as a monetary policy disturbance, that induces a change in aggregate income dY distributed across individuals in an unequal fashion dy_i . Notice that consistency requires $\mathbb{E}_i[dy_i] = dY$. Denoting $MPC_i := \partial g_i / \partial y_i$, we can write the indirect general equilibrium effect of the shock on C , that is, how the shock impacts aggregate consumption through the change in aggregate income as

$$dC = \mathbb{E}_i[MPC_i \cdot dy_i].$$

Let $\gamma_i = \frac{dy_i}{dY} \frac{Y}{y_i} \simeq d \log y_i / d \log Y$ measure the individual income elasticity to aggregate income, and note that the income-weighted mean of γ_i equals one. Using this expression for γ_i in the equation above, and letting income-weighted operators be defined by the $\tilde{\cdot}$ symbol, we obtain

$$dC = \tilde{\mathbb{E}}_i[MPC_i \cdot \gamma_i] dY.$$

Then, the expression above can be rewritten as

$$dC = \tilde{\mathbb{E}}_i[MPC_i] dY + \mathbb{C}\tilde{\mathbb{O}}\tilde{\mathbb{V}}_i(MPC_i, \gamma_i) dY,$$

where $\mathbb{C}\tilde{\mathbb{O}}\tilde{\mathbb{V}}_i$ is the income-weighted cross-sectional covariance. The first term is the income-share weighted average MPC in the population times the change in total income dY . It shows that the size of the aggregate MPC of the economy affects the magnitude of the general equilibrium feedback of a shock. The second term is

directly related to unequal incidence: it involves the income-weighted covariance between individual MPCs and the elasticity of individual income to aggregate income γ_i .

The covariance term is the component highlighted by Bilbiie (2017) and Patterson (2018). If there is equal incidence, $\gamma_i = 1$ for all i , then the covariance term is zero. If individuals who are more exposed to fluctuations in aggregate income (high γ_i) are also those with high MPCs, then this term is positive. In this case, unequal incidence is an amplification mechanism. If, instead, the correlation between MPCs and individual exposures is negative, unequal incidence is a dampening mechanism for shocks.

This simple exposition suggests that an analysis of the quantitative importance of unequal incidence as an amplification mechanism of monetary shocks requires two key ingredients. First, an empirically disciplined parameterization of the elasticities γ_i , which is what we discuss next. Second, a parameterization of how these elasticities co-vary with individuals' MPCs. Here, our approach differs from Patterson (2018). We do not make an attempt to directly estimate individual MPCs from microdata and correlate them with the degree of individual exposure to shocks, as she does. Rather, we rely on the endogenous distribution of MPCs generated by our model. Section 2.2 articulates this point further.

1.2 Unequal Incidence in the Microdata

We now describe our functional form for the incidence function and then proceed to the estimation of the incidence function's parameters. Since it is not feasible to estimate the degree of exposure individual by individual, we group individuals based on some fixed characteristic, which we summarize in the variable z .⁴

Functional form. Let Y_t be the aggregate variable of interest—earnings or transfers in our case—at date t . We assume that the allocation of this variable to an individual of type z (the incidence function for Y) is

$$\Gamma_y(z, Y_t) = \frac{\bar{v}_y(z)(Y_t/\bar{Y})^{\gamma_y(z)}}{\mathbb{E}_i[\bar{v}_y(z_i)(Y_t/\bar{Y})^{\gamma_y(z_i)}]} Y_t, \quad (1)$$

where \bar{Y} is a long-run average, which corresponds to a model's steady state. Note that our incidence function satisfies the consistency condition that $\mathbb{E}_i[\Gamma_y(z_i, Y_t)] = Y_t$.

The incidence is parameterized by two sets of coefficients: (i) $\bar{v}_y(z)$, which denotes the long-run (or steady-state) income share that accrues to an individual of type z , that is, $\bar{v}_y(z) = \bar{y}(z)/\bar{Y}$, and (ii) $\gamma_y(z)$, which captures the elasticity of the type z income to Y_t , if we impose the normalization $\mathbb{E}_i[\bar{v}_y(z_i)\gamma_y(z_i)] = 1$. To see this, log-differentiate (1) and evaluate at $Y_t = \bar{Y}$

$$\frac{\partial \log \Gamma_y(z, Y_t)}{\partial \log Y_t} = \gamma_y(z), \quad (2)$$

4. This is also useful when we map these estimates to our model, because in the model individuals are indexed by a finite number of state variables.

where we have used the normalization $\mathbb{E}_i[\bar{v}_y(z_i)\gamma(z_i)] = 1$.⁵ Hence, $\gamma(z)$ is the elasticity for type z of the variable at the individual level to its aggregate counterpart. In order to make (1) operational we need to estimate two sets of coefficients: income shares $\bar{v}_y(z)$ and elasticities $\gamma_y(z)$ for each group type z .

Estimation. Our first choice is how to proxy the grouping characteristic z . It is known that individual traits such as gender, age, education, occupation are all determinants of the exposure to business cycles. For example, earnings and hours worked are more cyclical for women, younger workers, less skilled workers, and for certain occupations and industries such as manufacturing. For ease of computation, we summarize all these characteristics into one variable only, the permanent component of labor income, which we denote by z .

Our first data source is the ASEC supplement of the CPS, which is conducted every March. This supplement to the CPS has the longest and largest sample as well as the most comprehensive collection of data on labor force status, work experience, and different types of income. We use data from 1967 to 2017 for all individuals between the ages of 26 and 55. The total average annual sample size is around 66,000 observations per year.

Labor income is defined as total pretax wage and salary income—that is, money received as an employee—over the calendar year.⁶ Government transfers are defined as income received from Social Security, from all welfare programs (e.g., TANF, SNAPS, Housing Assistance), from other major government programs other than Social Security and welfare (e.g., unemployment compensation, disability insurance), and from the Earned Income Tax Credit.

We proceed in two steps. First, for each individual in the data, we measure its position in the distribution of permanent income and bin individuals into quantiles. Next, we estimate the shares and the elasticities for each quantile.

To construct our measure of permanent income z , we first run a Mincer-style regression. We regress log labor income on dummies for gender, race, marital status, education, age, and occupation, as well as interactions between education and age, and between gender and age, to capture some heterogeneity in life-cycle earnings profiles. The adjusted R^2 of these regressions varies between 0.32 and 0.47, with higher values in the earlier years. We bin individuals into 50 quantiles based on their predicted level of permanent labor income.

5. To see this first log-differentiate (1):

$$\begin{aligned} \frac{\partial \log \Gamma_y(z, Y_t)}{\partial \log Y_t} &= \gamma_y(z) - \frac{\partial \log \mathbb{E}_i[\bar{v}_y(z_i)(Y_t/\bar{Y})^{\gamma_y(z_i)}]}{\partial \log Y_t} + 1 \\ &= \gamma_y(z) - \frac{\mathbb{E}_i[\bar{v}_y(z_i)\gamma_y(z_i)(Y_t/\bar{Y})^{\gamma_y(z_i)}]}{\mathbb{E}_i[\bar{v}_y(z_i)(Y_t/\bar{Y})^{\gamma_y(z_i)}]} + 1. \end{aligned}$$

Next, evaluate at $Y_t = \bar{Y}$ and use the restrictions $\mathbb{E}_i[\bar{v}_y(z_i)] = 1$ and $\mathbb{E}_i[\bar{v}_y(z_i)\gamma(z_i)] = 1$ to obtain (2).

6. We also used a broader definition of labor income with an imputation of 2/3 of self-employment income and results are very similar. For both definitions, we dropped top-coded observations.

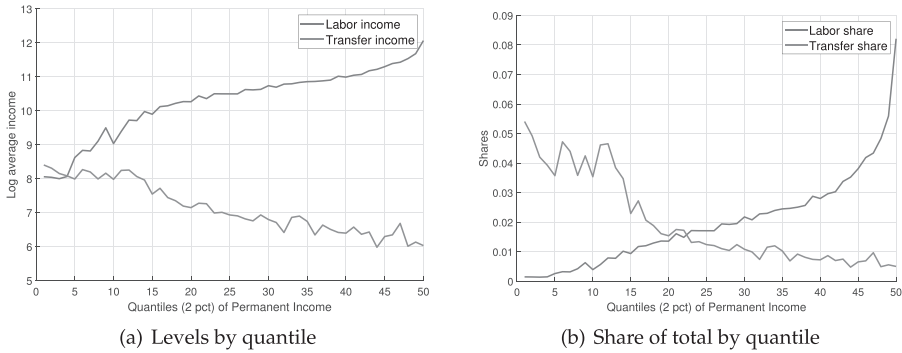


Fig 1. Log-Level and Shares of Labor Income and Government Transfers by Quantiles of Permanent Income in 2015.

Figure 1 reports the log of average labor and transfer income (left panel) and the share of total income (right panel) by quantile of permanent income z for the year of 2015.⁷ Average labor income is increasing in permanent income, especially at the top. Transfers are, instead, decreasing in permanent income with the low quantiles receiving five to seven times the transfers of the highest earners. The information on income shares is displayed the right panel, and directly maps to coefficients $\bar{v}(z)$ in (1).⁸

Next, we move to the elasticities $\gamma(z)$. As an exploratory step, we take the average earnings for different quantile bins of the permanent income distribution and compute their deviations from a linear trend.⁹ We then analyze how these quantile fluctuations relate to aggregate income fluctuations over time. Figure 2 plots the results: it has aggregate income log-deviations on the x-axis and log-deviations for three points of the permanent income distribution (corresponding to the 25th, 50th, and the 75th percentiles) on the y-axis. It is clear from the figure that the cyclical sensitivity to aggregate income fluctuations is higher for low permanent income individuals.

We build up on this strategy to estimate the elasticities along the entire distribution of permanent income. Let y_{it}^z be labor or transfer income of individual i , belonging to quantile z of permanent income in year t .¹⁰ To recover elasticities $\gamma(z)$, we estimate the following (constrained) system of equations:

$$\log y_{it}^z = \beta_0(z) + \beta_1(z)t + \gamma(z) \log Y_t + \epsilon_{it}, \quad \forall z \in \{1, \dots, 50\} \quad (3)$$

7. Other years show a similar pattern.

8. To see this, assume that type z is discrete with p.d.f. $\Pi(z)$, which is constant through time. In this case, the share of income flowing to group z at date t , defined as $s_t(z)$, is equal to $\Pi(z)v_z(z)$. Since we are using quantiles to define the different groups, all group have the same size— $\Pi(z)$ does not depend on z . Therefore, shares $s_t(z)$ are proportional to coefficients $v(z)$.

9. Remember from our discussion on incidence functions that we are interested in the household's income sensitivity to cyclical fluctuations. We allow the trend to be quantile specific to capture the differential secular evolution of labor income at different points in the distribution related to the well-documented widening in U.S. earnings inequality.

10. With a slight abuse of notation, we use z to index both the level of permanent income and its quantile.

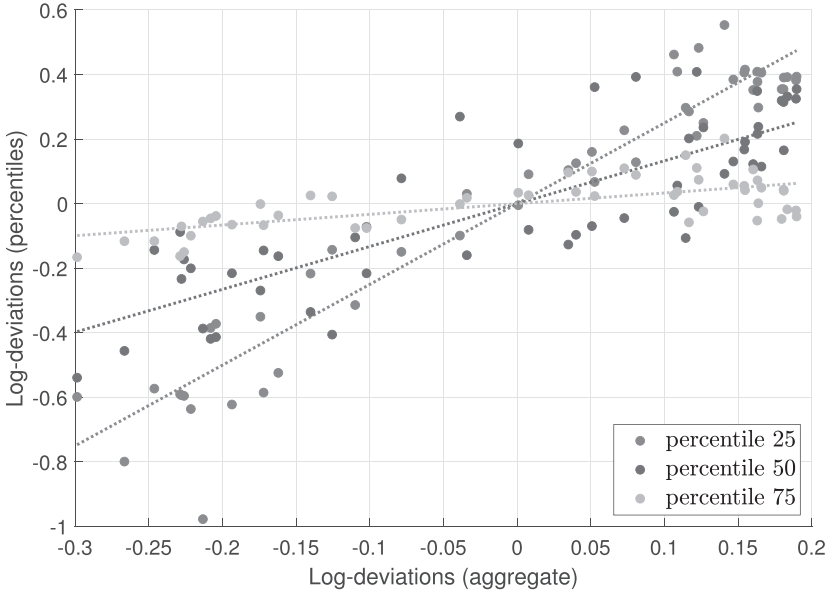


Fig 2. Scatterplot of Aggregate Income Log-Deviations (x-Axis) against Log-Deviations of Average Income for Different Percentiles of the Permanent Income Distribution (y-Axis).

NOTE: The dotted lines are best fitted OLS estimates.

$$\text{s.t. } \sum_{z=1}^{50} \bar{s}(z)\gamma(z) = 1.$$

While all coefficients are indexed by z , the 50 equations are still related by the constraint in the second line.¹¹ Figure 3 plots the estimated elasticities. The left panel confirms the results of Figure 2 and shows that the elasticity for low permanent income workers is two to three times larger than those of high permanent income.¹²

The log formulation in equation (3) presents a potential problem, however. Individual earnings are frequently zero in the data: more than 20% of all labor income observations in our sample are zeros and are concentrated in the lowest quantiles of

11. The constraint guarantees that the elasticities $\gamma(z)$ satisfy the consistency condition in (1). Going back to the notation in the previous section, note that we can rewrite the expectation as

$$\mathbb{E}_t[\bar{v}_y(z_t)\gamma(z_t)] = \sum_z \Pi(z) \frac{\bar{y}(z)}{\bar{Y}} \gamma(z) = \sum_z \bar{s}(z)\gamma(z),$$

which corresponds to the restriction in (3).

12. Our Online Appendix shows that the elasticity estimates are largely robust to a quadratic specification of the trend.

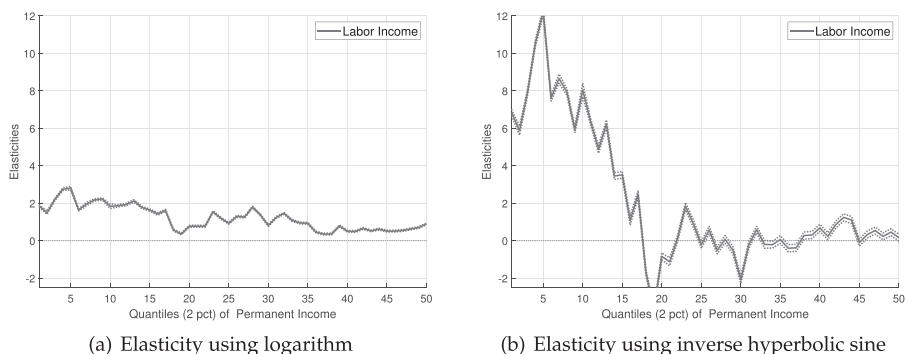


Fig 3. Estimated Elasticities of Individual Earnings to Aggregate Earnings as a Function of Permanent Income Quantile.

NOTE: Dotted lines are the 95% confidence bands.

SOURCE: ASEC 1967–2017.

of the distribution. For example, in the bottom decile, over half of the observations are zero, whereas at the top this fraction is less than 10%.¹³

When estimating elasticities using log earnings, these zero observations are dropped. Since the zeros are more likely to occur at times when aggregate earnings are low, one would expect this selection to produce a negative bias in $\gamma(z)$, especially at the low end of the permanent income distribution. To assess this bias, we replace the log operator in equation (3) with the inverse hyperbolic sine (asinh):

$$\operatorname{asinh}(y) = \log\left(y + \sqrt{y^2 + 1}\right).$$

This approach, which allows retaining zero-valued observations, is common in statistics (Bellemare and Wichman 2018). The right panel of Figure 3 reestimates equation (3) with this approach. The conclusions are quite stark. While above the 30th percentile the estimates of regression coefficient on aggregate earnings are roughly unchanged, at the bottom of the distribution the exposure to the cycle appears much stronger than in the left panel, and the hockey stick shape is much more pronounced.¹⁴

13. See Heathcote, Perri, and Violante (2020) for a discussion on how excluding zero earnings observations masks the cyclicity of inequality at the bottom of the earnings distribution.

14. There is the question of how to interpret the estimated coefficients from the regression that uses the asinh transformation. Let these coefficient be $\tilde{\gamma}(z)$ and the true elasticity be $\gamma(z)$. Bellemare and Wichman (2018) show that the elasticity is given by

$$\gamma(z|y_{it}, Y_t) = \tilde{\gamma}(z) \cdot \sqrt{\frac{y_{it}^2 + 1}{y_{it}}} \cdot \sqrt{\frac{Y_t}{Y_t^2 + 1}}.$$

Obviously, this elasticity cannot be computed at $y_{it} = 0$. Bellemare and Wichman (2018) suggest to evaluate it instead at the mean value for y_{it} . In practice, even for the lowest percentiles, the mean is large enough that $\gamma(z|y_{it}, Y_t) = \tilde{\gamma}(z)$.

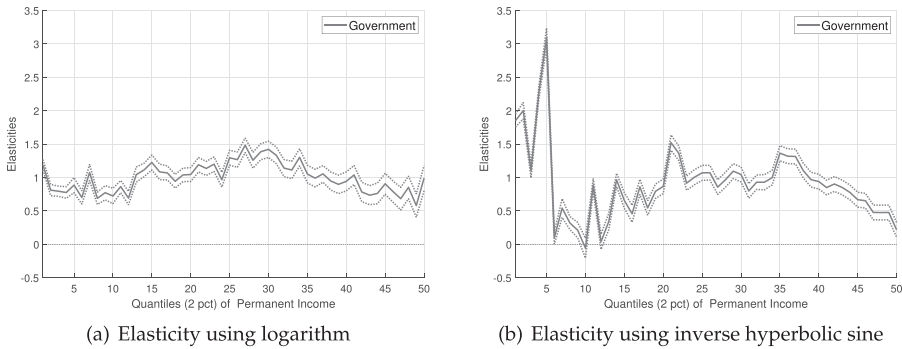


Fig 4. Estimated Elasticities of Individual Government Transfer Income to Aggregate Transfers as a Function of Permanent Income Quantile.

NOTE: Dotted lines are the 95% confidence bands.
 SOURCE: ASEC 1967–2017.

Figure 4 repeats the exercise with government transfers. Also for this case, the log and asinh are quite similar in shape except at the lowest percentiles of permanent income. Once again, when measured with the asinh transformation, the exposure in the bottom decile appears much higher. Overall, the shape of the incidence function is not monotonic.

One major caveat with the labor income estimates is that we dropped top-coded observations from the sample. In addition, the CPS is known to undersample individuals at the very top of the earnings distribution. Therefore, these estimates do not reveal how sensitive the very high-income households are to the cycle. Guvenen et al. (2017) estimate “workers’ betas” (i.e. systematic risk exposure) with respect to GDP using data from the Master earnings File of the SSA. Although these data are annual, cover a shorter period of time (1981–2009), and have only information on earnings (not transfers), they have the key advantage of a much better coverage of the top end of the income distribution. Here, we use these same data, but estimate incidence functions with respect to aggregate earnings (instead of GDP), consistently with our framework.¹⁵

We report our findings in Figure 5. The key difference with Figure 3 is the fact that exposure increases significantly again for the very top earners, that is, above the top 5%–10% and markedly for the top 1%. A natural interpretation of these findings is that the high exposure at the bottom of the distribution is associated with nonemployment risk, whereas at the top it is due to the fact that a large share

15. Because of the format in which these data are available, the specification we use is not exactly the same as in (3). The measure of permanent income is the mean of the previous 5 years of earnings. Moreover, we estimate the equation in first differences, without any quantile-specific trend, that is, the dependent variable is the log of the average change in earnings between t and $t + 1$ across all the individuals who were in quantile z at t .

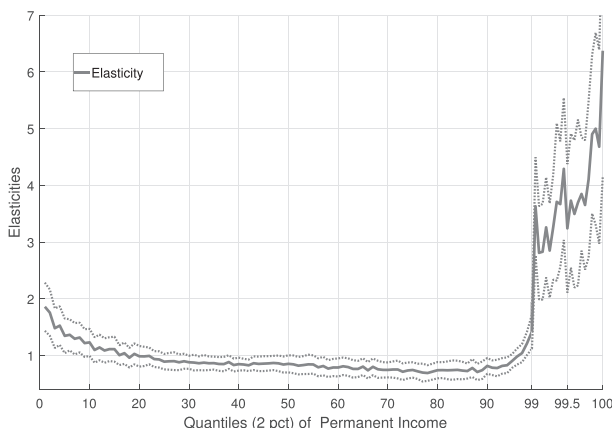


Fig 5. Estimated Elasticities of Individual Earnings to Aggregate Earnings as a Function of Permanent Income Quantile.

NOTE: Dotted lines are the 95% confidence bands.

SOURCE: Master Earnings File of the SSA 1979–2011.

of the compensation of high earners is made of performance-related bonuses and commissions.

We also explored whether the incidence functions for earnings and transfers are asymmetric with respect to positive and negative shocks, but found evidence of only very small sign dependence and thus, in what follows, we assume effects are symmetric.

2. THE MODEL

Households. Time is continuous. The economy is populated by a continuum of households who face an exogenous death rate $\xi \geq 0$. Households receive a utility flow u from consuming $c_t \geq 0$, where u is strictly increasing and strictly concave in consumption. Preferences are time separable and, conditional on surviving, the future is discounted at rate $\rho \geq 0$:

$$\mathbb{E}_0 \int_0^\infty e^{-(\rho+\xi)t} u(c_t) dt, \quad (4)$$

where the expectation is taken over realization of idiosyncratic earnings shocks (ζ, z) , where z is a permanent and ζ is a transitory component. The pair follows an exogenous stationary Markov process—which we describe in detail in Section 2.2—and determines household earnings through the incidence functions. Because of the law of large numbers, and the absence of aggregate shocks, there is no economy-wide uncertainty.

Households can hold non-negative positions in two types of real assets: a liquid asset b , which pays a rate of return r_t^b , and an illiquid asset a . Assets of type a are illiquid in the sense that households need to pay a cost for depositing into or withdrawing from their illiquid account. Let d_t be a household's deposit rate (with $d_t < 0$ corresponding to withdrawals) and $\chi(d_t, a_t)$ be the flow cost of depositing at a rate d_t for a household with illiquid holdings a_t . As a consequence of this transaction cost, in equilibrium the illiquid asset pays a higher real return than the liquid asset, that is, $r_t^a > r_t^b$.

Households are indexed by their holdings of liquid assets b , illiquid assets a , and by their idiosyncratic earnings shock pair (z, ζ) . At each instant in time t , the state of the economy is the joint distribution $\mu_t(da, db, dz, d\zeta)$. Upon death, households give birth to an offspring with zero wealth and a pair (z, ζ) equal to a random draw from its ergodic distribution.¹⁶ There are perfect annuity markets so that the estates of the deceased are redistributed to other individuals in proportion to their asset holdings.¹⁷

A household's asset holdings evolve according to

$$\begin{aligned} \dot{b}_t = & (1 - \tau_t)\Gamma_n(z_t, \zeta_t, w_t N_t) + r_t^b(b_t)b_t + \Gamma_T(z_t, T_t) \\ & + \Gamma_\pi(z_t, \zeta_t, \Pi_t) - d_t - \chi(d_t, a_t) - c_t \end{aligned} \tag{5}$$

$$\dot{a}_t = r_t^a a_t + d_t \tag{6}$$

$$b_t \geq 0, \quad a_t \geq 0. \tag{7}$$

Savings in liquid assets \dot{b}_t equal the household's income stream (composed of labor earnings taxed at rate τ_t , interest payments on liquid assets, and government transfers) net of deposits into or withdrawals from the illiquid account d_t , transaction costs $\chi(d_t, a_t)$, and consumption expenditures c_t . The functions Γ_n , Γ_T , and Γ_π are incidence functions that capture how aggregate labor earnings, government transfers, and profits of intermediary producers are distributed across households as a function of their idiosyncratic earnings states (z, ζ) and of aggregate levels of income.¹⁸ In Section 2.2, we describe these functions in more detail.

Net savings in illiquid assets \dot{a}_t equal interest payments on illiquid assets plus net deposits from the liquid account d_t . Note that while we distinguish between liquid and illiquid wealth, we net out gross positions within the two asset classes.

16. We allow for stochastic death to help in generating a sufficient number of households with zero illiquid wealth relative to the data. This is not a technical assumption that is needed to guarantee the existence of a stationary distribution, which exists even in the case $\xi = 0$.

17. The assumption of perfect annuity markets is implemented by making the appropriate adjustment to the asset returns faced by surviving households. To ease notation, we fold this adjustment directly into the rates of return, which should therefore be interpreted as including the return from the annuity.

18. More generally, the incidence functions could depend on the entire vector or individual states (a, b, z, ζ) or even the identity of each individual. Here, we instead restrict it to depend only on the exogenous states, as we did for our empirical counterparts.

The functional form for the transaction cost $\chi(d, a)$ is given by

$$\chi(d, a) = \chi_1 \left| \frac{d}{a} \right|^{\chi_2} a. \quad (8)$$

The convexity ($\chi_1 > 0$, $\chi_2 > 1$) ensures that deposit rates are finite, $|d_t| < \infty$ and hence household's holdings of assets never jump. Finally, scaling the convex term by illiquid assets a delivers the desirable property that marginal costs $\chi_d(d, a)$ are homogeneous of degree zero in the deposit rate d/a so that the marginal cost of transacting depends on the fraction of illiquid assets transacted, rather than the raw size of the transaction.¹⁹

Households maximize (4) subject to (5)–(8). They take as given equilibrium paths for the real wage $\{w_t\}_{t \geq 0}$, the real return to liquid assets $\{r_t^l\}_{t \geq 0}$, the real return to illiquid assets $\{r_t^a\}_{t \geq 0}$, and taxes and transfers $\{\tau_t, \Gamma_T(\cdot, T_t)\}_{t \geq 0}$.

Final-goods producers. A competitive representative final-good producer aggregates a continuum of intermediate inputs indexed by $j \in [0, 1]$

$$Y_t = \left(\int_0^1 y_{j,t}^{\frac{\varepsilon-1}{\varepsilon}} dj \right)^{\frac{\varepsilon}{\varepsilon-1}},$$

where $\varepsilon > 0$ is the elasticity of substitution across goods. Cost minimization implies that demand for intermediate good j is

$$y_{j,t}(p_{j,t}) = \left(\frac{p_{j,t}}{P_t} \right)^{-\varepsilon} Y_t, \quad \text{where} \quad P_t = \left(\int_0^1 p_{j,t}^{1-\varepsilon} dj \right)^{\frac{1}{1-\varepsilon}}.$$

Intermediate goods producers. Each intermediate good j is produced by a monopolistically competitive producer using effective units of capital $k_{j,t}$ and effective units of labor $n_{j,t}$ according to the production function

$$y_{j,t} = k_{j,t}^\alpha n_{j,t}^{1-\alpha}. \quad (9)$$

Intermediate producers rent capital at rate r_t^k in a competitive capital market and hire labor at wage w_t (we discuss details of the labor market below). Cost minimization implies that the marginal cost is common across all producers and given by

$$m_t = \left(\frac{r_t^k}{\alpha} \right)^\alpha \left(\frac{w_t}{1-\alpha} \right)^{1-\alpha}, \quad (10)$$

where factor prices equal their respective marginal revenue products.

19. Because the transaction cost at $a = 0$ is infinite, in computations we replace the term a with $\max\{a, \underline{a}\}$, where the threshold $\underline{a} > 0$ is a small value (always corresponding to \$500 in all calibrations) that guarantees costs remain finite even for households with $a = 0$.

Each intermediate producer chooses its price to maximize profits subject to price adjustment costs as in Rotemberg (1982). These adjustment costs are quadratic in the rate of price change \dot{p}_t/p_t and expressed as a fraction of aggregate output Y_t as

$$\Theta_t \left(\frac{\dot{p}_t}{p_t} \right) = \frac{\theta}{2} \left(\frac{\dot{p}_t}{p_t} \right)^2 Y_t, \tag{11}$$

where $\theta > 0$. Suppressing notational dependence on j , each intermediate producer chooses $\{p_t\}_{t \geq 0}$ to maximize

$$\int_0^\infty e^{-\int_0^t r_s^\alpha ds} \left\{ \tilde{\Pi}_t(p_t) - \Theta_t \left(\frac{\dot{p}_t}{p_t} \right) \right\} dt,$$

where

$$\tilde{\Pi}_t(p_t) = \left(\frac{p_t}{P_t} - m_t \right) \left(\frac{p_t}{P_t} \right)^{-\varepsilon} Y_t \tag{12}$$

are flow profits before price adjustment costs. The choice of r_t^α for the rate at which firms discount future profits is justified by a no-arbitrage condition that we explain below.

As proved in Kaplan, Moll, and Violante (2018), the combination of a continuous-time formulation of the problem and quadratic price adjustment costs yields a simple equation (the New Keynesian Phillips curve) characterizing the evolution of inflation $\pi_t = \dot{P}_t/P_t$ without the need for log-linearization:

$$\left(r_t^\alpha - \frac{\dot{Y}_t}{Y_t} \right) \pi_t = \frac{\varepsilon}{\theta} (m_t - m^*) + \dot{\pi}_t, \quad m^* = \frac{\varepsilon - 1}{\varepsilon}. \tag{13}$$

Equation (13) can be also written in present-value form as

$$\pi_t = \frac{\varepsilon}{\theta} \int_t^\infty e^{-\int_t^s r_\tau^\alpha d\tau} \frac{Y_s}{Y_t} (m_s - m^*) ds. \tag{14}$$

The marginal gain to a firm from increasing its price at time s is $\Pi'_s(p_s) = \varepsilon Y_s (m_s - m^*)$. Firms raise prices when their markup $1/m_s$ is below the flexible price optimum $1/m^* = \frac{\varepsilon}{\varepsilon-1}$. Inflation in (14) is the rate of price changes that equates the discounted sum of all future marginal payoffs from changing prices this period to its marginal cost $\theta \pi_t Y_t$ obtained from (11).

Investment fund. Illiquid assets are equity claims on an investment fund. Thus, the value of the fund equals households' aggregate stock of illiquid assets $A_t = \int a d\mu_t$. The investment fund owns the economy's capital stock K_t and shares in the intermediate producers X_t . The fund makes the economy's investment decision subject to an adjustment cost $\Phi(\iota_t)$, where ι_t is the investment rate, that is, investment as a fraction of the capital stock. The shares X_t represent a claim on a fraction ψ of the entire

future stream of monopoly profits net of price adjustment costs, $\Pi_t := \tilde{\Pi}_t - \frac{\theta}{2} \pi_t^2 Y_t$. Let q_t^x denote the share price. The remaining fraction $1 - \psi$ of profits flows directly into households' liquid asset account.

The investment fund solves the problem

$$A_0 := \max_{\{l_t, X_t\}_{t \geq 0}} \int_0^\infty e^{-\int_0^t r_s^a ds} \{ [r_t^k - l_t - \Phi(l_t)] K_t + \psi \Pi_t X_t - q_t^x \dot{X}_t \} dt$$

subject to

$$\dot{K}_t = (l_t - \delta) K_t,$$

with K_0 and X_0 given.

LEMMA 1. *The optimal investment rate l_t satisfies $1 + \Phi'(l_t) = q_t^k$ where $q_t^k := \frac{dA_t}{dK_t}$ is the fund's shadow value of capital. The value of the fund is given by $A_t = q_t^k K_t + q_t^x X_t$. And the return to illiquid assets r_t^a satisfies*

$$r_t^a = \frac{r_t^k - l_t - \Phi(l_t) + q_t^k (l_t - \delta) + \dot{q}_t^k}{q_t^k} = \frac{\psi \Pi_t + \dot{q}_t^x}{q_t^x}. \tag{15}$$

The proof of the Lemma 1 can be found in the Online Appendix. Note that the arbitrage condition (15) pins down the return on the illiquid asset r_t^a . Finally, (15) implies that $q_t^x = \psi \int_t^\infty e^{-\int_t^\tau r_s^a ds} \Pi_\tau d\tau$, which justifies the use of r_t^a as the rate at which future profits are discounted by the intermediate firms and, thus, as the discount rate appearing in equation (13), the Phillips curve.

Labor market. Our modeling of the labor market is nonstandard. As already mentioned, we assume that aggregate effective units of labor N_t and wages w_t are determined from firms' labor demand together with an exogenous wage-setting rule. The labor demand schedule comes from intermediate firms' profit maximization and pins down aggregate labor as a function of wages and a number of demand shifters. To determine wages we assume an exogenous wage-setting rule

$$w_t = \bar{w} \left(\frac{N_t}{\bar{N}} \right)^{\epsilon_w}, \tag{16}$$

where \bar{w} and \bar{N} are steady-state values. For instance, if $\epsilon_w = 0$, wages are perfectly rigid and employment is simply determined by the location of firms' labor demand schedule. If $\epsilon_w > 0$, there is downward pressure on wages whenever employment is below its steady-state value. The labor demand together with the wage-setting rule pin down payments to labor $w_t N_t$, which are distributed across households according to the incidence function $\Gamma(z_t, \zeta_t, w_t N_t)$.

In Kaplan, Moll, and Violante (2018), we adopted the assumption of the basic New Keynesian model that prices are sticky while wages are flexible. As a result, markups

are countercyclical under a monetary shock. In practice, this typically also implies that profits decrease sharply after a monetary expansion. It is by now well understood in the HANK literature that the distribution of profits can have large effects on the model’s cyclical properties of aggregate consumption and output (e.g., Werning 2015, Broer et al. 2016). Falling profits in response to expansionary monetary shocks is counterfactual. The advantage of our assumption is that we can control the degree of wage rigidity in the economy. When wages are rigid, intermediate firms’ marginal costs and markups move less in response to shocks. Therefore the dynamics of profits, dividends, and equity prices are more in line with the data.²⁰

Monetary authority. The monetary authority sets the nominal interest rate on liquid assets i_t according to a Taylor rule. We consider two alternative specifications. First, we consider a Taylor rule that reacts to current inflation only

$$i_t = \bar{r}^b + \phi_\pi \pi_t \tag{17}$$

with $\phi_\pi > 1$. Rule (17) has the unappealing feature, which is worse in continuous time, that the nominal rate reacts instantaneously to variations in inflation. So we also consider a Taylor rule with partial-adjustment dynamics for the nominal rate.²¹

$$\frac{di_t}{dt} = -\rho_i(i_t - \bar{r}^b - \phi_\pi \pi_t). \tag{18}$$

Parameter ϕ_π still captures the response of the interest rate to inflation, but this now occurs with a certain delay, which is controlled by the value of ρ_i . Large values imply a smaller delay.²²

Our main experiment in the paper studies the economy’s response to an unexpected monetary shock ϵ . In the case of the first specification, the shock is persistent ($\epsilon_t =$

20. The Online Appendix lists all sources of differences between our current environment relative to Kaplan, Moll, and Violante (2018).

21. Notice that the policy rate goes from being a jump variable to a state variable. This is the continuous time analogue of the discrete time Taylor rule:

$$i_t = (1 - \rho^i)i_{t-1} + \rho^i(\phi_\pi \pi_t).$$

See Sims (2004) and Cochrane (2017) for more details on this specification.

22. To see this, note that we can solve the differential equation (18) backward and write it as

$$i_t - \bar{r}^b = \rho_i \int_0^\infty e^{-\rho_i s} (\phi_\pi \pi_{t-s}) ds.$$

If we let $\bar{\pi}_t \equiv \rho_i \int_0^\infty e^{-\rho_i s} \pi_{t-s} ds$ denote the exponential moving average of past inflation rates, and substitute that in our integral equation above, we come up to the following representation

$$i_t = \bar{r}^b + \phi_\pi \bar{\pi}_t.$$

Note that this is analogous to our baseline Taylor rule, except that monetary policy now reacts to the smoothed inflation rate $\bar{\pi}_t$ instead of π_t . Moreover, the smaller the value of ρ_i , the bigger the weight of past inflation on $\bar{\pi}_t$, hence bigger the delay.

$\epsilon_0 \exp(-\eta t)$) and enters additively in the Taylor rule

$$i_t = \bar{r}^b + \phi_\pi \pi_t + \epsilon_t.$$

In the case of the partial-adjustment rule, the monetary shock takes the form of a single innovation $d\epsilon_0$ at time zero that makes the nominal rate jump from \bar{r}^b to $\bar{r}^b + d\epsilon_0$. After that initial perturbation, nominal rate dynamics follow (18).

Government. The government faces exogenous government expenditures G_t and administers a progressive tax and transfer scheme on household labor income that consists of a lump-sum transfer T_t and a proportional tax rate τ_t . The government is the sole issuer of liquid assets in the economy, which are real bonds of infinitesimal maturity B_t^g , with negative values denoting government debt. Its sequential budget constraint is

$$\dot{B}_t^g + G_t + T_t = \tau_t w_t N_t + r_t^b B_t^g. \quad (19)$$

It is useful to define the government's *primary surplus* $S_t \equiv \tau_t w_t N_t - T_t - G_t$ and rewrite the budget constraint²³ as

$$\dot{B}_t^g = r_t^b B_t^g + S_t.$$

Notice that the *primary surplus* depends both on government's decision with respect to transfers T_t , tax rates τ_t , and government expenditures G_t as well as on overall economic activity through labor income $w_t N_t$. In the steady state, the primary surplus \bar{S} is just enough to cover the interest payments on debt $-r^b B_t^g$. Outside of steady state, fiscal policy determines a path of *primary surplus* as well as the instrument (τ_t , T_t , or G_t) used to achieve it.²⁴

2.1 Equilibrium

An equilibrium in this economy is defined as paths for individual household and firm decisions $\{a_t, b_t, c_t, d_t, n_t, k_t\}_{t \geq 0}$, input prices $\{w_t, r_t^k\}_{t \geq 0}$, returns on liquid and illiquid assets, $\{r_t^b, r_t^a\}_{t \geq 0}$, the value of the fund $\{A_t\}_{t \geq 0}$, the inflation rate $\{\pi_t\}_{t \geq 0}$, fiscal variables $\{\tau_t, T_t, G_t, B_t^g\}_{t \geq 0}$, distributions $\{\mu_t\}_{t \geq 0}$, and aggregate quantities such that, at every t : (i) households and firms maximize their objective functions taking as given equilibrium prices, taxes, and transfers; (ii) the sequence of distributions satisfies aggregate consistency conditions; (iii) the government budget constraint holds; and

23. Note that if the government held a portfolio of bonds of different maturity, then an analogous government budget constraint would hold, with the real market value of outstanding government debt taking the place of B_t^g , and a no-arbitrage condition between bonds of different maturity would ensure that the instantaneous return on the portfolio would be r_t^b . For a *given path* of r_t^b , the set of feasible paths for primary surpluses S_t does not depend on the composition of the government's bond portfolio. However the value of alternative bond portfolios with different durations would exhibit different sensitivity to *changes in the path* of r_t^b . We return to this point in the Conclusions.

24. This fiscal policy takes the form of a simple rule for the evolution of primary surplus as a function of other aggregates. We specify the rule and explore its implications in Section 3.5.

(iv) the liquid asset (bond) market, the illiquid asset (shares of the fund) market, and the goods market all clear.

The liquid asset market clears when

$$B_t^h + B_t^g = 0, \tag{20}$$

where B_t^g is the stock of outstanding government debt and $B_t^h = \int b d\mu_t$ are total household holdings of liquid bonds. In equilibrium the investment fund holds all the shares in intermediary producers, which we normalize to one so that $X_t = 1$. From Lemma 1 this implies that households' holdings of illiquid assets $A_t = \int a d\mu_t$ equals

$$A_t = q_t^k K_t + q_t^x. \tag{21}$$

The goods market clearing condition is

$$Y_t = C_t + I_t + G_t + \Theta_t + \Phi_t + \chi_t. \tag{22}$$

Here, Y_t is aggregate output, C_t is total consumption expenditures, I_t is gross additions to the capital stock K_t , G_t is government spending, Θ_t and Φ_t are total price and capital adjustment costs, and the last term reflects transaction costs (to be interpreted as financial services).

As explained, the labor market is not competitive. The aggregation of intermediate producers' labor demand determines N_t and, given N_t , equation (16) determines the wage.

2.2 Calibration

The model period is one quarter. Our calibration is divided into three main steps. First, we calibrate the exogenous stochastic process (z, ζ) determining household's earnings. Second, we target a realistic distribution of liquid and illiquid assets and the fraction of households with low liquid wealth as this directly maps to the distribution of MPCs, which is key to consumption response as highlighted in Section 1.1. Finally, we calibrate parameters of the production and monetary side of the model to standard values of the New Keynesian literature. The list of parameter values is in Table 1.

Continuous time earnings. We take the processes (z_{it}, ζ_{it}) from Kaplan, Moll, and Violante (2018). The process is estimated to replicate the higher order moments of the distribution of earnings changes estimated by Guvenen et al. (2015) from SSA data. Each component is modeled as a "jump-drift" process in logarithms and the expectation of their product $\mathbb{E}[z\zeta]$ is normalized to unity. Let the logarithm of the permanent component be $\tilde{z}_{it} \equiv \log z_{it}$. Jumps arrive at some Poisson intensity λ_z and upon their realization a new value for the state \tilde{z}'_{it} is drawn from a normal distribution with mean zero and variance v_z , $\tilde{z}'_{it} \sim \mathcal{N}(0, v_z)$. Between jumps, the process simply

TABLE 1
LIST OF PARAMETER VALUES AND TARGETED MOMENTS

Description		Value	Target
Preferences			
ξ	Death rate	1/180	Avg. lifespan of 45 years
σ^{-1}	Intertemporal elasticity of substitution	1.00	—
ρ	Discount rate (p.a.)	7.2%	See Table 2
Transaction cost function			
a	Min a in denominator	\$500.00	See Table 2
χ_1	Level component	0.395	See Table 2
χ_2	Convex component	1.326	See Table 2
Production			
ε	Demand elasticity	10	Profit share of 10%
θ	Price adjustment cost	100	Slope of Phillips curve $\varepsilon/\theta = 0.1$
α	Capital share	0.33	National Accounts
δ	Depreciation rate (p.a.)	5.75%	National Accounts
ϕ_0	Capital adj. cost	[0,25]	VAR evidence
Labor market			
ϵ_w	Wage elasticity	0.10	VAR evidence
Government			
τ	Proportional labor tax	0.30	National Accounts
T	Lump-sum transfer (rel GDP)	0.027	Transfer GDP share of 3%
ϕ_π	Taylor rule coefficient to inflation	1.25	—

reverts to zero at some rate β_z . Formally, the process for \tilde{z}_{it} is

$$d\tilde{z}_{it} = -\beta_z \tilde{z}_{it} dt + dJ_{z,it},$$

where $dJ_{z,it}$ captures the jumps in the process. The description of the transitory component is analogous. The Poisson shock of the permanent component z occurs on average once every 38 years and the process has a half-life of around 18 years. The transitory component ζ jumps on average once every 3 years and the process has a half-life of around one quarter.²⁵

Demographics and preferences. We set the quarterly death rate ξ to 1/180 so that the average lifespan is of 45 years. Households have CRRA utility over consumption with risk aversion parameter σ set to 1.

Wealth distribution. We set steady-state nominal return on liquid asset at 2% per year and inflation at zero. The steady-state return on the illiquid asset is endogenously determined by the market clearing condition.

Taking as given the process for (z, ζ) and the level of risk aversion, households' incentives to accumulate liquid and illiquid assets depend mainly the discount rate ρ and the parameters of the transaction cost function χ_1, χ_2 (recall that we have assumed

25. See table 3 in Kaplan, Moll, and Violante (2018) for the fit and the exact parameter values. Overall, the fitted earnings process matches the variance and kurtosis of 1 and 5 year earnings changes, as well as fraction of small changes. Consistent with cross-sectional earnings distribution in the data, our earnings process features a large amount of right-tail inequality. The top 10, 1, and 0.1 shares of *gross* household labor earnings in the steady state are 46%, 14%, and 4%, respectively.

TABLE 2

TARGETED EMPIRICAL MOMENTS FOR THE WEALTH DISTRIBUTION (RATIOS OF NET ASSET POSITIONS TO ANNUAL GDP) AND THE SHARE OF HAND TO MOUTH HOUSEHOLDS (RELATIVE TO THE TOTAL POPULATION), WITH THEIR MODEL COUNTERPART

	Data	Model
Mean illiquid asset	2.92	2.88
Mean liquid asset	0.20	0.21
Frac. with $b \approx 0$ and $a = 0$	0.10	0.06
Frac. with $b \approx 0$ and $a > 0$	0.20	0.25

NOTES: Approximately 0 stands for $b(0, b]$ where we set b to 5% of quarterly labor income or around \$800.

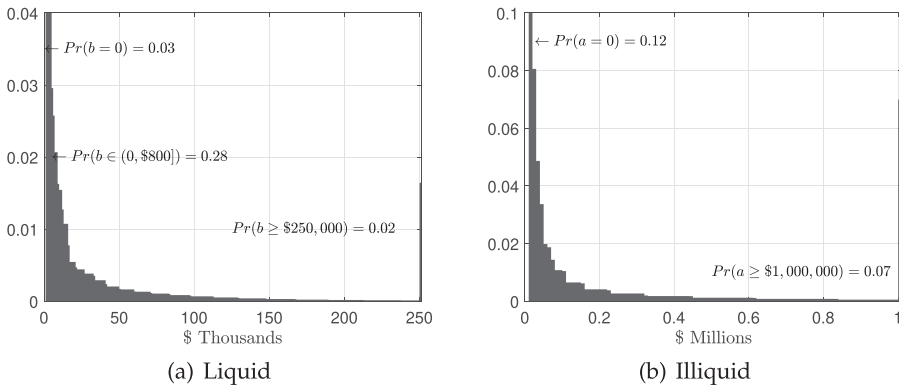


Fig 6. Distribution of Liquid and Illiquid Assets in the Model.

away unsecured borrowing). We choose these parameters to match four moments of the household wealth distribution: (i)–(ii) the mean of liquid and illiquid wealth over annual GDP from Kaplan, Moll, and Violante (2018), (iii)–(iv) the fraction of poor and wealthy hand-to-mouth (HtM) households.²⁶ Table 2 shows the fit of the model with respect to these targets. The implied steady-state return on illiquid assets r^a is 6.6 percent per annum.

Figure 6 displays the steady-state distributions of liquid and illiquid wealth for this calibration. The Gini coefficients in the model are 0.75 and 0.79 for the liquid and illiquid wealth distributions, respectively, which imply a Gini coefficient for net worth very close to its empirical counterpart of 0.81.²⁷

26. The definition of poor and wealthy HtM follows the one adopted by Kaplan and Violante (2014) and Kaplan, Violante, and Weidner (2014), that is, it is based on the ratio between liquid wealth holdings and income.

27. In the resulting ergodic distribution, roughly 85% of households are adjusting at any point in time. Conditional on making a deposit or withdrawal, the mean absolute quarterly transaction as a fraction of the stock of illiquid assets is 2.3%. The transaction cost associated with a transaction this size is 11% of the transaction. In steady state, the equilibrium aggregate transaction costs, which one can interpret as financial services, amount to less than 3% of GDP.

Production and labor market. The elasticity of substitution for final goods producers ε is set to 10. In the production function of intermediate goods producers we set to $\alpha = 0.33$, which yields a capital share of 29% and labor share of 60%. The price adjustment cost parameter θ is set to 100 so that the slope of the Phillips curve ε/θ is 0.10.

When we solve the model with capital adjustment costs, we adopt the following specification for the function $\Phi(\cdot)$:

$$\Phi(\iota) = \frac{\phi_0}{2}(\iota - \delta)^2, \quad (23)$$

where δ is the depreciation rate. We set ϕ_0 to 25 so that when the economy is hit by a monetary shock, at its peak, the ratio of investment to output is around 2, in line with VAR evidence presented by Christiano, Eichenbaum, and Trabandt (2016).²⁸

The wage elasticity to aggregate hours ϵ_w in the wage setting rule is set to 0.10.²⁹

Fiscal and monetary policy. We set the proportional labor income tax rate τ to 0.30 and the lump-sum transfer T to be 3% of output. Since the government is the only provider of liquid assets, government debt is 21% of annual GDP—the target in Table 2. Government expenditures are determined residually from the government budget constraint. The fiscal rule outside steady state is detailed in Section 3.5. The coefficient ϕ_π in the Taylor rule is set to 1.25. The persistence of the monetary shock η (for the standard rule) and in the nominal rate ρ_i (for the partial-adjustment rule) are discussed on the result section.

Distribution of monopoly profits. In our two-asset model, we need to take a stand on whether profits paid out as dividends end up in a household's liquid or illiquid accounts. This matters because the MPC out liquid resources is much larger than the MPC out of illiquid resources, due to the transaction cost.

In our model, monopolistic profits Π_t are split between dividends paid to the illiquid investment fund and dividends paid directly into liquid accounts in proportion to $(\psi, 1 - \psi)$, respectively. In our baseline, we set $\psi = \alpha$ (capital share)—as discussed in Kaplan, Moll, and Violante (2018), this particular choice “neutralizes” the distributional consequences (with respect to aggregate liquidity) of countercyclical profits. The profits received by the investment fund end up in illiquid wealth and their distribution across individuals is pinned down by households' endogenous accumulation of illiquid assets. Profits flowing into the liquid account are distributed across house-

28. We note that, often, the literature on estimated DSGE models uses a different specification of adjustment costs which penalizes changes in investment as opposed to investment rates. There, the aim is to obtain hump-shaped IRF's. Here, instead, the aim is to study amplification at impact and transmission mechanism, thus we opted for a more traditional specification of adjustment costs.

29. Taking into account the confidence intervals in Figure 1 of Christiano, Eichenbaum, and Trabandt (2016), the elasticity of wage to hours in response to a monetary shock can be placed anywhere between 0.0 and 1.00. As explained, we choose a value closer to the lower bound to reduce the movement in marginal cost, and hence the movement in profits.

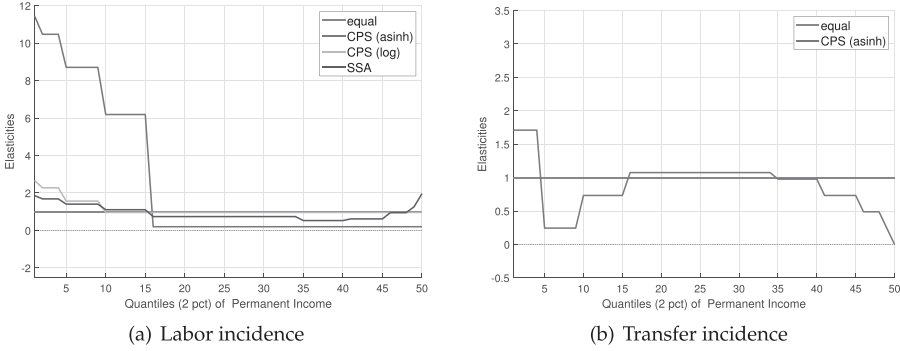


Fig. 7. Approximation of the Estimated Unequal Incidence Functions to Be Used in Simulation.

holds through Γ_π in proportion individuals' labor income, that is,

$$\Gamma_\pi(z_{it}, \zeta_{it}, \Pi_t) = z_{it} \zeta_{it} (1 - \alpha) \Pi_t. \quad (24)$$

This specific distribution rule reflects the fact that a sizable share of labor compensation is in terms of bonuses and commissions linked to firm's performance.

Incidence functions. Our incidence functions Γ_n and Γ_T that enter households' budget constraint (5) follow the specification introduced in Section 1.2

$$\Gamma_n(z, \zeta, wN) = \frac{z\zeta (wN/\bar{w}\bar{N})^{\gamma_n(z)}}{\int z'\zeta'(wN/\bar{w}\bar{N})^{\gamma_n(z')} d\mu_t} wN,$$

$$\Gamma_T(z, T) = \frac{\bar{v}_T(z)(T/\bar{T})^{\gamma_T(z)}}{\int \bar{v}_T(z')(T/\bar{T})^{\gamma_T(z')} d\mu_t} T,$$

where \bar{N} , \bar{w} , and \bar{T} are steady-state aggregates. The parameters $\gamma_n(z)$ and $\gamma_T(z)$ are the elasticities at quantile z for earnings and transfers estimated in Section 1.2, while $z\zeta$ and $\bar{v}_T(z)$ are the steady-state shares of labor earnings and transfers accruing to each household type, calibrated based on Figure 1, right panel.³⁰

The left panel of Figure 7 plots the incidence functions for labor earnings that we use in our experiments. *Equal* refers to the neutral baseline where individuals have an equal exposure to shocks. *SSA* approximates the incidence function estimated on the SSA data following Guvenen et al. (2017). *CPS (log)* and *CPS (asinh)* approximates our estimated incidence function using the ASEC data. The right panel of Figure 7 plots the incidence function for government transfers that we use in our experiments, which approximates the right panel of Figure 4.

30. The model reproduces very precisely the share of labor income and transfers by permanent income in steady state. The share of transfers at each level of z is generated exactly in calibration. The share of labor earnings is not exact because our process for individual labor earnings is estimated with data from the Master Earnings File of the SSA, rather than CPS. However, the correspondence is very close.

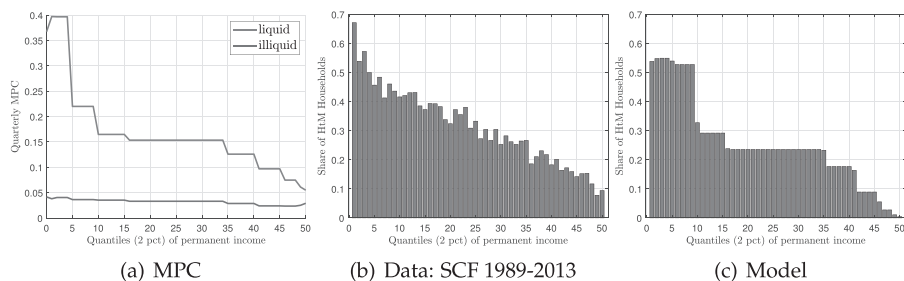


Fig 8. Panel (a): Quarterly MPC Out of Liquid and Illiquid Wealth. Panels (b) and (c): Share of Hand-to-Mouth Households in the Data and in the Model.

Recall from Section 1.1 that the impact of unequal incidence on the aggregate consumption response depends on the covariance between MPC and the elasticity across the z distribution. Therefore, to quantitatively assess the amplification generated by this channel it is essential that the distribution of MPCs (and not only its first moment) is in line with the data. Panel (a) of Figure 8 reports the model implied quarterly MPCs out of liquid and illiquid wealth for each quantile of permanent income. The MPC out of liquid wealth decreases with permanent income, ranging from 0.40% to 0.05%. Compared to that, the MPC out of illiquid wealth is fairly stable, averaging 3% quarterly (five times smaller than the MPC out transitory income). How does those numbers compare with the data? In order to evaluate this, we take an indirect approach by noting that there is a very close correspondence between the MPC and the share of HtM in the model. This correlation is useful because, while estimating MPCs is an arduous empirical task, the HtM status is observable in the microdata.³¹

Panels (b) and (c) of Figure 8 plot the share of HtM households in the data and in the model for each quantile of permanent income. Not surprisingly, the share of HtM households is declining in permanent income in the data (left panel).³² Notice that the model replicates this empirical pattern quite well, which gives us some confidence on the empirical validity of the distribution of MPC by z implied by our model.³³

3. RESULTS

We study the transitional dynamics to a one-time unexpected expansionary monetary shock. In the standard Taylor rule case (17), there is a time zero quarterly innovation to the Taylor rule shock ϵ_t of $\epsilon_0 = -0.25\%$ (i.e., -1% annually) that

31. This strategy is an important difference between our work and Patterson (2018). While she estimates the distribution of MPCs out of unexpected transitory income changes, we rely on the model.

32. The data source we used for these calculations is the SCF 1989–2016. The sample selection is the same as in the CPS, and so is the Mincer regression to impute permanent income.

33. The large flat region corresponds to the mid point in the permanent income distribution which, in the discretization, has a large share of the total mass—a consequence of its kurtosis.

mean-reverts at rate $\eta = 0.5$.³⁴ In the partial-adjustment case (18), the innovation is also of 25 bp quarterly ($d\epsilon_0 = -0.25\%$) and we set the coefficient ρ_i to the same persistence of the shock.³⁵

We are interested in identifying the model features that matter, quantitatively, for the amplification/dampening of the monetary policy shock. Our baseline model is one where: (i) capital adjustment costs are zero; (ii) the Taylor rule is the one specified in equation (17); (iii) exposure of labor income and government transfers is uniform (equal incidence case); (iv) a share α (the capital share) of profits is paid out in liquid form proportionately to individuals' labor income, while the rest is directly reinvested in the illiquid account; (v) the government pays off its debt in the short run, and adjusts transfers in medium term.

We start by analyzing the role of the two generalizations of the baseline HANK model, capital adjustment costs, and partial-adjustment Taylor rule. Next, we study the three channels of potential shock amplification that rely on how the change in aggregate income (labor, financial, and government) induced by the shock is distributed across households.

3.1 Capital Adjustment Costs

Figure 9 plots the impulse response functions (IRFs) for output, consumption, investment, and the value of the fund for the cases with and without capital adjustment costs. In the absence of adjustment costs, aggregate investment, and output react strongly to the shock, but the share price of the fund barely moves: if anything, it falls slightly at impact because the profits of the intermediaries decrease at impact.³⁶ In the presence of adjustment costs, the investment response is much weaker, but there is a strong positive reaction in the value of the fund. This result highlights an important shortcoming of this first generation of HANK models: in the data both investment and asset prices react strongly after a monetary shock, but in the model large movements in prices can only be achieved with small changes in quantities, and viceversa. In contrast to the behavior of investment and output, aggregate consumption is largely unaffected by the presence of adjustment costs.

Figure 10 decomposes the IRF for aggregate consumption into direct and indirect effects, following Kaplan, Moll, and Violante (2018). Direct and indirect effects are computed by counterfactuals. To compute the direct impact of a monetary shock, we let the real liquid rate change as in the baseline, but freeze all other prices and government transfers at their steady-state value. Indirect effects are computed in a similar way, varying one price at a time. The figure splits indirect effects between

34. We set η to 0.5, corresponding to a quarterly autocorrelation of $e^{-\eta} = 0.61$, a value consistent with the VAR-based empirical evidence.

35. This ensures that the direct impact of the shock on the nominal rate is the same across the two specifications. The equilibrium nominal path may yet be different because the two rules respond differently to inflation.

36. The aggregate investment response in the case without adjustment cost is around 7% in the first quarter. We chose a smaller value for the upper limit of the vertical axis to better compare the consumption and output response.

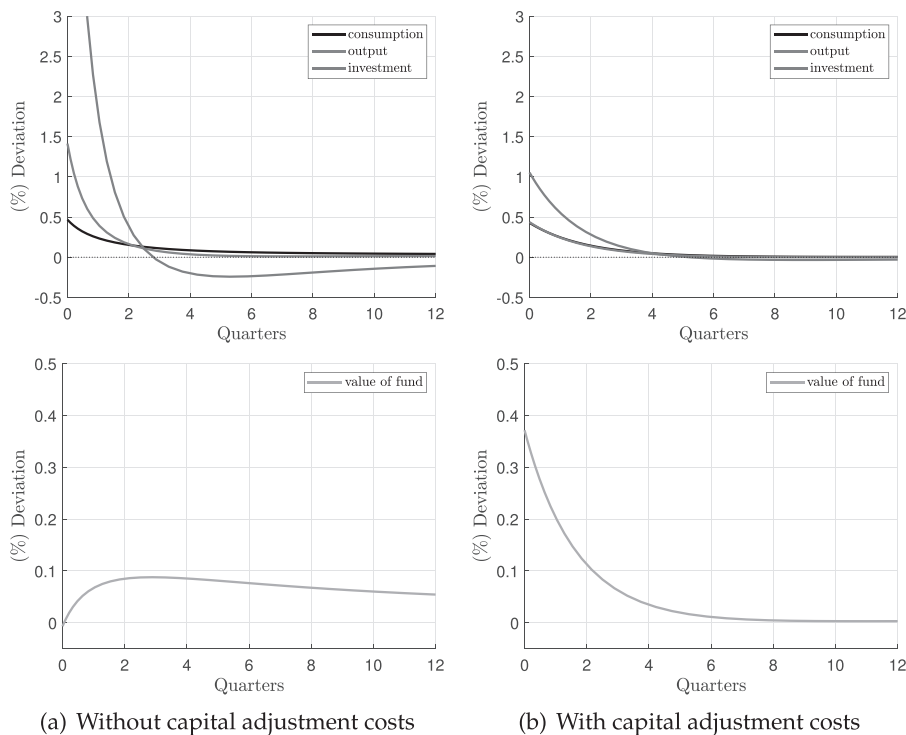


Fig 9. Aggregate Responses to a Monetary Policy Shock.

NOTES: Top panels: output, investment, and consumption. Bottom panels: fund share price.

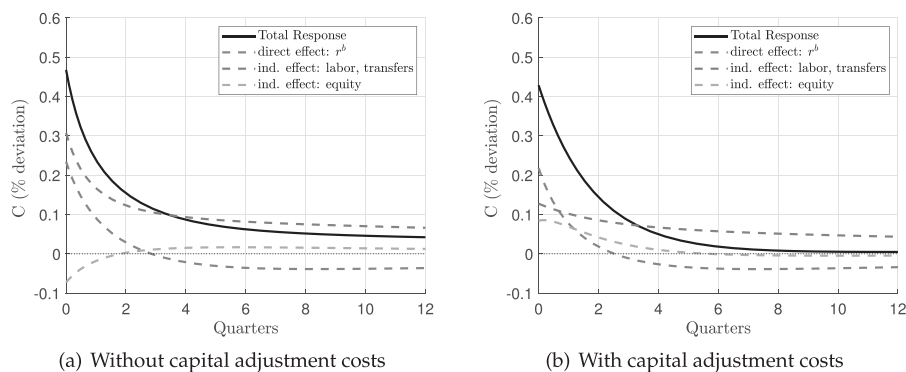


Fig 10. Decomposition of the Consumption IRF to a Monetary Shock into Direct and Indirect Components.

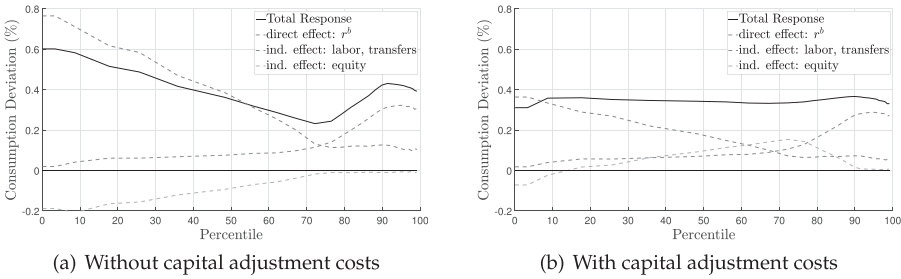


Fig 11. Decomposition of the Impact Effect of a Monetary Shock on Consumption across the Distribution of Liquid Wealth (with and without Capital Adjustment Costs).

the impact on consumption caused by the change in disposable labor income and the change in the equity value.

As in Kaplan, Moll, and Violante (2018), we find in both scenarios that the indirect general equilibrium channel accounts for about half of the total increase in aggregate consumption at impact. After a year, indirect effects account virtually for all of the consumption response. This stands in stark contrast with the representative agent version of the New Keynesian model—where intertemporal substitution dominates the transmission mechanism at all frequencies.

Even though the decomposition between direct and indirect channels is similar in the two cases, the relative importance of labor income versus equity prices in the overall indirect effect changes with the introduction of the adjustment costs. To understand why, notice that the initial impulse to aggregate demand always comes from the direct channel, that is, the effect of the real rate cut on the consumption of non-HtM households and on the investment decisions made by the fund. This initial aggregate demand response pushes up employment and labor income, to which the consumption of HtM households strongly responds, leading to a second round of demand, employment and income expansion. This demand to income feedback, which eventually reaches its equilibrium, is the main driver of consumption response in the case without adjustment costs.

In the presence of adjustment costs, the investment response to the real rate change is milder, which reduces the power of the aggregate demand channel. However, there is now another channel contributing to aggregate consumption response: equity prices rise in response to the decrease in the real rate, a change that households also react to by increasing their consumption (recall our comment on the MPC out of illiquid wealth in Figure 8). This extra consumption coming from the behavior of equity prices offsets the smaller expansion of labor income, explaining why aggregate consumption moves by similar amounts in the two economies.

There is one important consequence of this different transmission mechanism. The identities of those households who gain from the expansion in asset prices are not the same of those who benefit from the raise in employment and labor income. Figure 11 illustrates this distributional impact by showing the consumption reaction across the

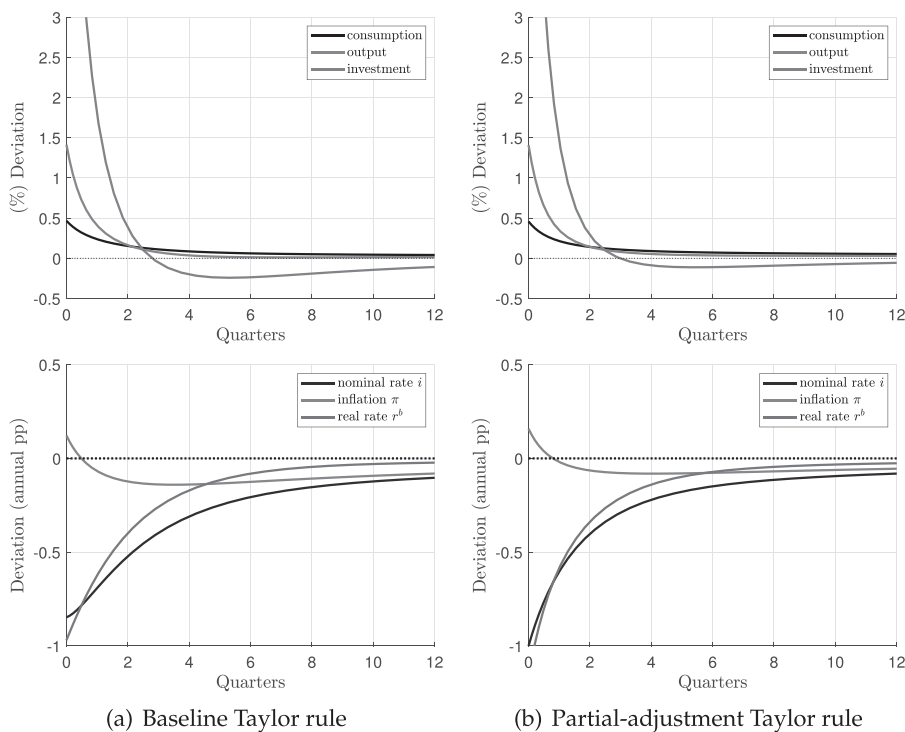


Fig 12. Aggregate Responses to a Monetary Policy Shock.

NOTES: Top panels: output, investment, and consumption. Bottom panels: nominal, inflation, and real rate.

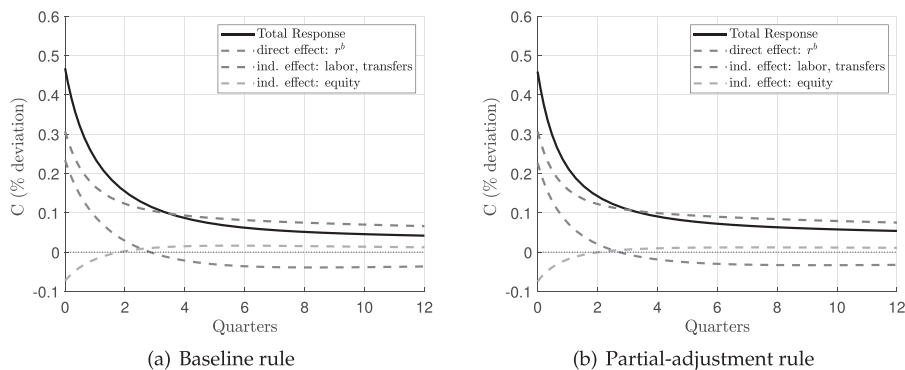


Fig 13. Decomposition of the Consumption IRF to a Monetary Shock into Direct and Indirect Components under Different Taylor Rules.

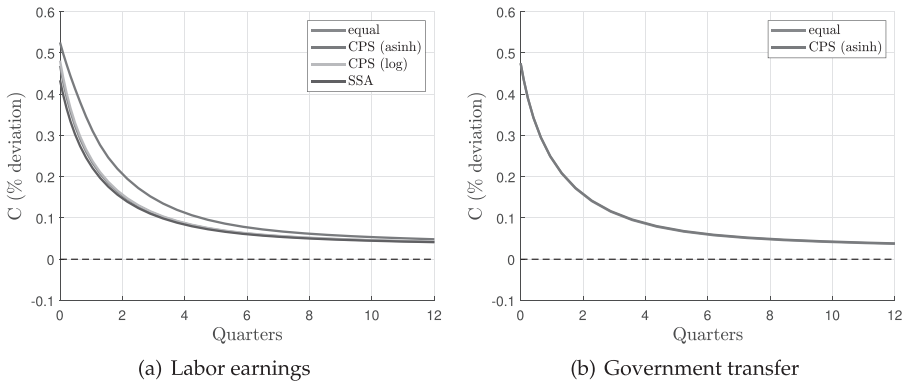


Fig 14. Consumption Response to Monetary Shock in the Model with Unequal Incidence of Income across the Household Distribution.

percentiles of the liquid asset. Without capital adjustment costs, poor households increase their consumption the most as labor income, the major driver of aggregate dynamics, plays a larger role on their response. With high capital adjustment costs, the overall consumption response is almost flat across the distribution because wealthier households now benefit disproportionately more from the capital gains.

3.2 Taylor Rule

Figure 12 plots the IRFs to a monetary shock in the baseline and partial-adjustment Taylor rule. Given our calibration, the partial-adjustment rule has only a minor impact on the equilibrium path of the nominal and real rate (see bottom panels). As a consequence of that, aggregate quantities respond similarly in both cases (see upper panels). In contrast to the effects of capital adjustment costs, the decomposition between direct and indirect effects is also unaffected, as seen in Figure 13.

3.3 Unequal Income Incidence

Figure 14 plots the impulse response of aggregate consumption under the different parameterizations for the labor incidence function Γ_n presented in Figure 7.

The model with the “CPS (log)” unequal incidence function generates only a tiny amount of amplification relative to the equal incidence case. The estimated incidence increases the cumulative aggregate consumption response over the first quarter from 0.35% to 0.36%.³⁷ Under the more extreme “CPS (asinh)” estimate, however, we find stronger amplification, with first quarter aggregate consumption rising from 0.35% to 0.43%.

Perhaps surprisingly, the SSA calibration of the incidence function yields a small dampening relative to the equal incidence case. To understand why, recall that the

37. The IRFs in the figure are the continuous time ones. To obtain the average of the first quarter it is enough to integrate that impulse response in the figure from $t = 0$ to $t = 1$.

TABLE 3
FIRST QUARTER AGGREGATE CONSUMPTION RESPONSE TO MONETARY SHOCK RELATIVE TO THE BASELINE

Experiment	HANK	HANK Capital Adj. Cost	HANK Partial Adj. Taylor
Baseline	1.000	1.000	1.000
-	(1.000)	(0.991)	(0.941)
Unequal labor incidence			
CPS (asinh)	1.234	1.188	1.232
CPS (log)	1.024	1.010	1.023
SSA	0.934	0.973	0.937
Profit distribution			
Profit all liq.	1.199	1.089	1.066
Profit all illiq.	0.904	0.943	0.931
Fiscal reaction			
$\rho^R = 1, \rho^N = 0$	1.573	1.265	1.571
$\rho^R = 0, \rho^N = 0$	-	-	-
(Transfers adjust)	2.140	1.819	2.358
(Labor tax adjust)	1.593	1.436	1.773
(Government exp. adjust)	1.538	1.379	1.751

NOTES: Values in each column are normalized by the consumption in the first row of the corresponding column (baselines). Values in brackets in the second row denote the relative consumption of the baseline specification relative to the baseline in HANK without capital adjustment costs and with the standard Taylor rule (first column, first row). See Section 3.3 for a description of the unequal labor incidence exercises, Section 3.4 for profit distribution, and Section 3.5 for fiscal adjustment.

SSA incidence function is U-shaped, meaning that incomes at both the bottom and the top of the distribution are more exposed to fluctuations in aggregate incomes than those in the middle. There are therefore two offsetting forces at work. More exposure at the bottom, where MPCs are higher than average, leads to amplification; but more exposure at the top, where MPCs are lower than average, leads to dampening. Furthermore, recall from Section 1.1 that it is the *income-weighted* covariance between MPCs and the elasticities, $\mathbb{C}\tilde{\mathcal{O}}\mathbb{V}_i(\text{MPC}_i, \gamma_i)$, that matters for amplification. Since individuals at the top of the distribution receive a higher share of aggregate income, the upward-sloping part of the SSA incidence function receives higher weight than the downward-sloping part. The net effect is that the SSA incidence function yields a slightly smaller consumption response than our baseline with equal incidence.

How does this magnitude of amplification compares with the work by Patterson (2018)? Patterson expresses her main amplification result in terms of the consumption multiplier: her estimated unequal income incidence function increases the general equilibrium multiplier from 1.3 to 1.42 (see her Table 3). She reports this as a 40% increase in the *net multiplier* (from 0.3 to 0.42). In contrast, we measure amplification in terms of the overall consumption response. Applying this metric to Patterson's findings, the unequal distribution of labor income leads to a consumption response that is 1.42/1.3 times larger, which corresponds to a 9% increase, and hence in line with our findings.³⁸

38. Formally, returning to simple model of Section 1.1, the total effect of a monetary shock on aggregate consumption can be written as $dC = \mathbb{M}\mathbb{P}\mathbb{C}dY + \mathbb{C}\tilde{\mathcal{O}}\mathbb{V}dY + \mathbb{D}dr$, where the first two terms represent the general equilibrium component and $\mathbb{D}dr$ represents the direct effect of the shock at impact. Using the equilibrium condition $C = Y$, it is immediate that the total effect can be written as

Our empirical analysis in Section 1.2 highlighted that government transfers are also unequally distributed over the cycle. In the right panel of Figure 14 we compare the consumption response for our baseline of equal incidence of government transfers with the one approximated from our CPS estimates. For transfers, the impact of unequal incidence on the aggregate consumption response is even smaller than for labor income.³⁹

3.4 Profit distribution

Our next candidate for the amplification of monetary shocks is the distribution of profits outside of steady state. Recall that our baseline distribution rule (24) assumes that a fraction $1 - \alpha$ of profits is paid out to the liquid account proportionately to individuals' labor income $z_{it}\zeta_{it}$. We now let profit distributions into the liquid account be given by

$$\Gamma_{\pi}(z_t, \zeta_t, \Pi_t) = z_t \zeta_t [(1 - \alpha)\bar{\Pi} + (1 - \omega)(\Pi_t - \bar{\Pi})],$$

where $\bar{\Pi}$ denotes steady-state monopoly profits and $(1 - \omega)$ denotes the deviations of profits from steady state that are paid out as liquid dividends instead of flowing to the investment fund.

We consider three different scenarios corresponding to three different values of ω . Our baseline model corresponds to the case $\omega = \alpha$ so the same rule holds both in and out of steady state. Second, we consider the case $\omega = 0$ meaning that all deviations from steady state profit are paid out in liquid form proportionately to earnings. Third, we set $\omega = 1$ meaning that all deviations from steady state profits end up in the illiquid account and are distributed to individuals according to their holdings of shares in the investment fund.

Figure 15 summarizes the results for the economy without capital adjustment costs. The left panel shows the consumption response for the three alternative profit distribution rules based on the different values for ω . When a large share of profits is paid into the liquid account (low ω), the fall in profits after an expansionary monetary shock directly depresses household disposable income. When it is, instead, retained into the illiquid account (high ω), it drags down investment (right panel) which, in turn, lowers demand for labor and therefore also curtails the expansion of household disposable income. In our experiments it is the second of these two offsetting forces that dominates: aggregate consumption increases more when profits are retained in the liquid account.

$dC/dr = 1/(1 - \text{MPC} - \text{COV})D$. In Patterson (2018), the GE multiplier with equal incidence (with $\text{COV} = 0$) is estimated to be 1.3 and with unequal incidence 1.42. Therefore, adding unequal incidence amplifies the rise of C at impact by $(1.42 - 1.3)/1.3 \times 100$ percent, i.e. 9 percent.

39. We have experimented with alternative specifications for the transfer incidence function as well. For example, we have assumed that deviations of transfers from steady state are distributed equally (rather than proportionately to steady state transfers) across the entire population. This assumption leads to a slight dampening compared to the "equal incidence" case.

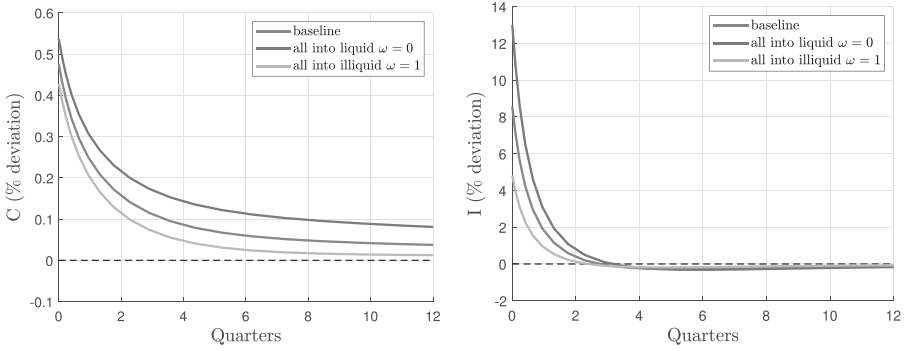


Fig 15. Response to Monetary Shock for Different Profit Distributions.

NOTE: Left panel: consumption. Right panel: investment.

3.5 Fiscal Adjustment to the Monetary Shock

An interest rate cut by the monetary authority affects the government budget constraint (19) through two channels. First, the lower interest rate raises the present value of primary surpluses (or alternatively, lowers the present value of primary deficits) for a given path of tax revenues, transfers, and government expenditure. Second, the expansionary effect on output raises revenues from labor taxation, hence raises current and future surpluses. As a result, a fiscal adjustment is required in order for the present value government budget constraint to remain balanced. This is true whether or not we dealing with a RANK or HANK economy, but the magnitude of the required adjustment depends on the maturity structure of government debt (see the Conclusions for further discussion of this latter point) However, in RANK models, the details of the fiscal adjustment are irrelevant to the determination of other equilibrium variables, because of Ricardian equivalence. That is not the case in HANK models. Borrowing constraints and the heterogeneity in MPCs across households breaks Ricardian equivalence, making the assumptions about how the government balances its budget constraint crucial for the overall aggregate consumption response.⁴⁰

We let fiscal policy take the form of a rule for the government *primary* surplus S_t together with specification of fiscal instruments used to achieve it. Namely, deviations of the primary surplus from steady state are a linear function of deviations of interest payments on debt $r_t^b B_t^g - \bar{r}^b \bar{B}^g$, of aggregate labor income $w_t N_t - \bar{w} \bar{N}$, and of the level of debt $B_t^g - \bar{B}^g$ itself:

$$S_t = \bar{S} - (1 - \rho^R)(r_t^b B_t^g - \bar{r}^b \bar{B}^g) + \rho^N(w_t N_t - \bar{w} \bar{N}) - \rho^B(B_t^g - \bar{B}^g), \quad (25)$$

40. This point is also highlighted in Kaplan, Moll, and Violante (2018).

where parameters (ρ^R, ρ^N, ρ^B) control the primary surplus sensitivity to these deviations.⁴¹ Substituting the fiscal rule (25) into the government budget constraint (19), the implied evolution of government assets can be written as

$$\dot{B}_t^g = \rho^R (r_t^b B_t^g - \bar{r}^b \bar{B}^g) + \rho^N (w_t N_t - \bar{w} \bar{N}) - \rho^B (B_t^g - \bar{B}^g). \tag{26}$$

To understand the implications of different fiscal rules for the evolution of government debt, consider the following two extreme cases: $(\rho^R = \rho^N = 0)$ and $(\rho^R = 1, \rho^N = \tau)$. When $(\rho^R = \rho^N = 0)$, government debt is kept fixed at its steady-state value independently of changes in revenue and borrowing costs. Thus, the primary surplus matches the interest payments on debt at all times (i.e., $S_t = -r_t B_t$). When $(\rho^R = 1, \rho^N = \tau)$, the (initial) primary surplus varies proportionally with government tax revenues and government debt adjusts to accommodate the overall surplus.⁴²

While (25) defines the response of the primary surplus to changes in the government budget, it does not pin down the way in which the surplus is achieved. There are numerous alternatives here: the same surplus can be obtained by either appropriately adjusting one of the instruments (T_t , τ_t , and G_t), or any combination of them. In our baseline we have assumed that: (i) government debt absorbs all of the short-run fiscal imbalance $(\rho^R = 1, \rho^N = \tau)$ and (ii) transfers T_t adjust in the medium term to bring the debt back to steady state, while government expenditures and labor tax rate are kept fixed at their steady-state values.⁴³ In this section, we consider alternatives to this

41. See Leith and Leeper (2016) for a similar formulation.

42. We assume that $\rho^B > \rho^R \bar{r}^b$ so that the future surpluses response to movements in real debt guarantees $\lim_{t \rightarrow \infty} B_t^g = \bar{B}^g$ for arbitrary deviations of debt from steady state. To see how this condition is sufficient to guarantee a converging debt path, note that we can rewrite (26) as

$$\begin{aligned} \dot{B}_t^g &= \rho^R (r_t^b B_t^g - \bar{r}^b \bar{B}^g) + \rho^N (w_t N_t - \bar{w} \bar{N}) - \rho^B (B_t^g - \bar{B}^g), \\ \dot{B}_t^g &= (\rho^R r_t^b - \rho^B) (B_t^g - \bar{B}^g) + \rho^R \bar{B}^g (r_t^b - \bar{r}^b) + \rho^N (w_t N_t - \bar{w} \bar{N}). \end{aligned}$$

If we restrict attention to sequences $\{r_t^b, N_t, w_t\}$ that converge back to their steady state as $t \rightarrow \infty$, then the dynamics of real debt for sufficiently large t are

$$\dot{B}_t^g = (\rho^R \bar{r}^b - \rho^B) (B_t^g - \bar{B}^g),$$

which imply that $B_t^g \rightarrow \bar{B}^g$ if $\rho^B > \rho^R \bar{r}^b$.

43. To see the implications of the rule for the behavior of transfers in the case that the labor tax rate and government expenditures are fixed, note that we can write the surplus as

$$\begin{aligned} S_t &= \tau w_t N_t - T_t - \bar{G} \\ S_t &= -(T_t - \bar{T}) + (\tau w_t N_t - \bar{T} - \bar{G}) \\ S_t &= -(T_t - \bar{T}) + \tau (w_t N_t - \bar{w} \bar{N}) + (\tau \bar{w} \bar{N} - \bar{T} - \bar{G}) \end{aligned}$$

or in terms of transfer

$$\begin{aligned} T_t - \bar{T} &= -(S_t - \bar{S}) + \tau (w_t N_t - \bar{w} \bar{N}) \\ T_t - \bar{T} &= -(-(1 - \rho^R)(r_t^b B_t^g - \bar{r}^b \bar{B}^g) + \rho^N (w_t N_t - \bar{w} \bar{N}) - \rho^B (B_t^g - \bar{B}^g)) + \tau (w_t N_t - \bar{w} \bar{N}) \end{aligned}$$

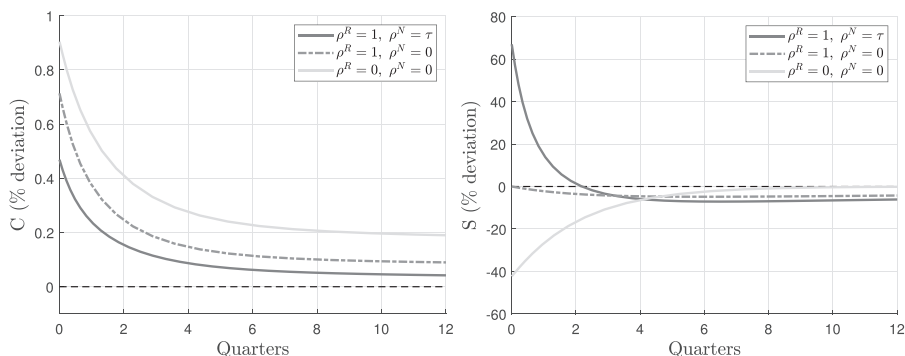


Fig 16. Consumption and Surplus Response to Monetary Shock under Three Different Types of Fiscal Adjustments
 NOTES: (i) $\rho^R = 1, \rho^N = \tau$: government debt responds to changes in interest payments as well as to changes in tax revenue (the baseline); (ii) $\rho^R = 1, \rho^N = 0$: government debt responds to changes in interest payments, but transfers compensate for changes in tax revenues; (iii) $\rho^R = 0, \rho^N = 0$: transfers adjust to keep government debt fixed. In all cases $\rho^B = 0.05$.

baseline that differ (i) in the values of the parameters ρ^R, ρ^N measuring the reaction of primary surplus to deviations in government’s revenues and interest expenses; and (ii) in the fiscal instrument used to adjust the surplus.

Figure 16 shows consumption and surplus for the first experiment. A fiscal rule that keeps the value of outstanding debt constant ($\rho^R = 0, \rho^N = 0$) leads to a first quarter consumption response twice as large as that of our baseline ($\rho^R = 1, \rho^N = \tau$) where government debt is used to smooth out the initial fluctuations. The corresponding surplus deviation is shown in the right panel. While the government increases its primary surplus in the baseline by saving the extra tax revenue $\tau(w_t N_t - \bar{w}\bar{N})$, the fixed debt case has the government running down its surplus as it increases transfers to households in the amount of $-(r_t^b B_t^g - \bar{r}^b \bar{B}^g) + \tau(w_t N_t - \bar{w}\bar{N})$.

The intuition is simple: households’ disposable income rises with the increase in transfers and generates an additional impulse to aggregate consumption. We also explore the impact of an intermediate policy ($\rho^R = 1, \rho^N = 0$) that keeps primary surplus constant at impact. In this case, the government adjusts transfers in proportion to the rising revenues, but uses the lower borrowing cost to pay off its debt.⁴⁴ The aggregate consumption response under this alternative rule is between the two previous cases, as one should expect.

In Figure 17, we compare the fiscal instrument that is used to adjust primary surpluses. We concentrate on the case with debt fixed ($\rho^R = 0, \rho^N = 0$) and experiment

$$T_t - \bar{T} = (1 - \rho^R)(r_t^b B_t^g - \bar{r}^b \bar{B}^g) + (\tau - \rho^N)(w_t N_t - \bar{w}\bar{N}) + \rho^B(B_t^g - \bar{B}^g).$$

So in the baseline ($\rho^R = 1, \rho^N = \tau$), transfers increase only when debt deviates from steady state.

44. In this case the behavior of transfers is given by

$$T_t - T = \tau(w_t N_t - wN) + \rho^B(B_t^g - \bar{B}^g).$$

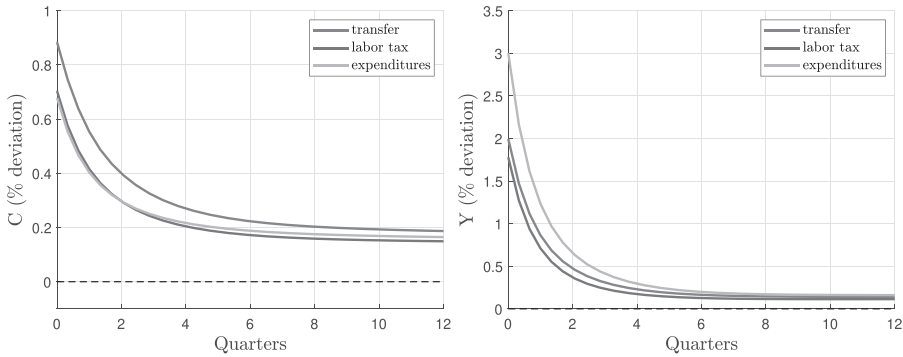


Fig 17. Consumption and Output Response When Different Instruments Are Used to Achieve the Desired Fiscal Surplus.

with different alternative fiscal instruments to deliver the desired surplus. We compare three alternatives: we adjust transfers T_t only, the labor tax τ_t only, or government expenditures G_t only. Among these different instruments, using transfers to adjust the government budget leads to the largest aggregate consumption response, while changing government expenditures promotes the largest increase in output. Transfers directly affect household disposable income, and because of its lump-sum nature, this bump in income is especially potent for low-income HtM households. Adjusting government expenditures leads to a one-to-one impact on aggregate demand, but some crowding out in private consumption. Cutting the labor tax distributes resources to households proportionally to their labor income. This force tends to primarily benefit high-income low-MPC individuals, which explains why consumption goes up by less in this scenario.

To sum up, the different assumptions about timing of the fiscal response explored in Figure 16 and instrument choice explored in Figure 17 generate substantial differences in the aggregate consumption response. In the Conclusions, we discuss why these results on the sensitivity of monetary policy to the fiscal reaction are not a consequence of restricting the government to only issue short-term debt.

3.6 Summary

Table 3 summarizes our different experiments. The first column (HANK) refers to the baseline model specification. The other columns report results for the models with aggregate capital adjustment costs and a more sophisticated Taylor rule. The values in the table denote the difference in first quarter consumption following an expansionary monetary shock, relative to the baseline from the same column.

The presence of adjustment cost lessens the impact of all the elements we have considered. This can be traced to the smaller impact of the shock on investment, which curtails the aggregate demand response. As a result, all channels that rely on the general equilibrium feedback are weakened. Results under the partial-adjustment

rule are much closer to the baseline, except for the profit distribution exercises that are dampened under the different rule. Quantitatively, however, the main takeaway is about the importance of the fiscal response relative to other elements considered: its impact on aggregate consumption response is much larger than that of unequal income incidence or the distribution of profits.⁴⁵

4. CONCLUSIONS

HANK models contain several channels that amplify or dampen the response of aggregate consumption to a monetary policy shock which are either absent or less relevant in RANK models. Our goal in this paper has been to provide some guidance on how a subset of these channels compare in terms of their quantitative strength. We performed these exercises in a rich two-asset HANK model calibrated to be consistent with microevidence on household earnings, wealth distributions, and MPCs.

The model elements that we have considered include: unequal incidence of fluctuations in aggregate labor income, capital income, and government transfers across households; capital adjustment costs; and different rules for fiscal and monetary policy. Our findings suggest that of these elements, assumptions about how the government budget constraint adjusts in response to the monetary shock is by far the most important.

In our model, all government debt and household liquid assets are of infinitely short duration with an instantaneously adjusting interest rate. In reality, however, a substantial fraction of government debt and household assets are long term, with coupon rates that were set in the past and that do not adjust to changes in short rates. In HANK models, these assumptions about the maturity structure of government debt are not innocuous.

The extent to which a change in the path of interest rates affects the intertemporal government budget constraint depends on the difference between the duration of the government's assets and its liabilities. The governments' only asset is its claim on future primary surpluses, which, in the absence of general equilibrium effects, delivers a fixed constant stream of resources. The government could thus perfectly hedge its exposure to interest rate changes by issuing a constant coupon perpetual bond, in which case, monetary policy would not require any fiscal adjustment (although general equilibrium effects that alter primary surpluses would still necessitate a fiscal adjustment).

There is thus a sense in which our assumption of infinitely short duration debt leaves the government maximally exposed to monetary policy, and thus requires the largest fiscal adjustment among alternative possible assumptions about the maturity structure of government debt. In RANK models, none of this matters, because of Ricardian Equivalence. But in HANK models, where Ricardian Equivalence fails, a

45. Check the Online Appendix for an equivalent of Table 3 for the representative agent counterpart of our model, together with a discussion.

shorter maturity structure translates into a larger required fiscal adjustment and hence a wider set of possible indirect effects of the monetary shock.

We leave the systematic analysis of the effects of monetary policy with long-term government debt in HANK for future research—see Auclert, Rognlie, and Straub (2020) for a recent example. We note that while increasing the duration of government debt may mitigate the effects of fiscal policy for monetary transmission in HANK, it introduces additional effects of monetary policy that are absent in RANK and in HANK models with short-term debt. This is because even though the government is better hedged against interest rate changes with longer term debt, the households who hold the other side of these claims as long-duration assets are now more exposed to changes in interest rates. This introduces additional redistributive effects of monetary policy.

Consistent with previous findings from simpler, analytic HANK models, we find that unequal incidence of labor income, profit income and transfers can all either amplify or dampen the effect of a monetary shock. For incidence functions implied by our estimates from U.S. data, the degree of amplification, or dampening is relatively small. With respect to this finding, the main caveat is that our estimated incidence functions are not conditional on a monetary shock. An interesting avenue of research is the analysis of whether heterogeneous exposure across households varies with respect to the shock and how this affect the propagation of the main aggregate shocks that account for U.S. fluctuations. See Broer, Kramer, and Mitman (2020) for recent progress on this question.

With respect to the distribution of capital income across households, we emphasized that whether profits end up in liquid or illiquid accounts is important for the dynamics of investment. In our model, investment is financed by inflows into illiquid assets (equity, housing) whereas liquid assets are instead invested in government bonds. In a more general model in which households deposit their liquid wealth in the banking sector and banks intermediate funds to firms, this distinction may be less sharp, depending on the transaction cost faced by banks vis-à-vis individuals.

Finally, we emphasized that capital adjustment costs modify the transmission mechanism by muting movements in the quantities of capital and amplifying movements in equity prices. While this feature of the model improves the macrofinance side of the model relative to, for example, Kaplan, Moll, and Violante (2018), its predictions remain very far from the basic asset pricing facts documented in the literature. Going forward, this is another dimension where progress should be made. Recent work by Lenel and Kekre (2019) is an important step in this direction.

LITERATURE CITED

- Acharya, S., and K. Dogra. (2018) “Understanding HANK: Insights from a PRANK,” Staff Reports 835, Federal Reserve Bank of New York.
- Auclert, A. (2019) “Monetary Policy and the Redistribution Channel,” *American Economic Review*, 109, 2333–67.

- Auclert, A., and M. Rognlie. (2018) "Inequality and Aggregate Demand," NBER Working Papers 24280, National Bureau of Economic Research, Inc.
- Auclert, A., M. Rognlie, and L. Straub. (2020) "Micro Jumps, Macro Humps: Monetary Policy and Business Cycles in an Estimated HANK Model," NBER Working Papers w26647, National Bureau of Economic Research, Inc.
- Bayer, C., R. Luetticke, L. Pham-Dao, and V. Tjaden. (2019) "Precautionary Savings, Illiquid Assets, and the Aggregate Consequences of Shocks to Household Income Risk," *Econometrica*, 87, 255–90.
- Bellemare, M., and C. Wichman. (2018) "Elasticities and the Inverse Hyperbolic Sine Transformation," University of Minnesota.
- Berger, D., L. Bocola, and A. Dovis. (2019) "Imperfect Risk-Sharing and the Business Cycle," Discussion paper, Stanford University.
- Bernstein, J. (2019) "Redistribution and the Asymmetric Transmission of Aggregate Shocks," Discussion paper, Princeton University.
- Bilbiie, F. (2017) "The New Keynesian Cross: Understanding Monetary Policy with Hand-to-Mouth Households," Discussion paper, Center for Economic Policy Research.
- Bilbiie, F. O., D. R. Känzig, and P. Surico. (2019) "Capital, Income Inequality, and Consumption: The Missing Link," Discussion paper, CEPR Discussion Paper No. DP14118.
- Broer, T., N.-J. Harbo Hansen, P. Krusell, and E. Oberg. (2016) "The New Keynesian Transmission Mechanism: A Heterogenous-Agent Perspective," CEPR Discussion Papers 11382, C.E.P.R. Discussion Papers.
- Broer, T., J. Kramer, and K. Mitman. (2020) "The Curious Incidence of Shocks Along the Income Distribution," Discussion paper, IIES, Stockholm.
- Christiano, L. J., M. S. Eichenbaum, and M. Trabandt. (2016) "Unemployment and Business Cycles," *Econometrica*, 84, 1523–69.
- Cochrane, J. H. (2017) "The New-Keynesian Liquidity Trap," *Journal of Monetary Economics*, 92, 43.
- Cui, W., and V. Sterk. (2019) "Quantitative Easing," Discussion paper, UCL.
- Debortoli, D., and J. Gali. (2018) "Monetary Policy with Heterogeneous Agents: Insights from Tank Models," Working paper, CREI.
- Den Haan, W., P. Rendahl, and M. Riegler. (2015) "Unemployment (Fears) and Deflationary Spirals," Discussion paper, Centre for Macroeconomics (CFM).
- Gornemann, N., K. Kuester, and M. Nakajima. (2014) "Doves for the Rich, Hawks for the Poor? Distributional Consequences of Monetary Policy," Discussion paper, Federal Reserve Bank of Philadelphia.
- Guerrieri, V., and G. Lorenzoni. (2011) "Credit Crises, Precautionary Savings, and the Liquidity Trap," NBER Working Papers 17583, National Bureau of Economic Research.
- Guvenen, F., F. Karahan, S. Ozkan, and J. Song. (2015) "What Do Data on Millions of U.S. Workers Reveal about Life-Cycle Earnings Risk?" NBER Working papers 20913, National Bureau of Economic Research.
- Guvenen, F., S. Schulhofer-Wohl, J. Song, and M. Yogo. (2017) "Worker Betas: Five Facts About Systematic Earnings Risk," *American Economic Review*, 107, 98–403.
- Heathcote, J., F. Perri, and G. L. Violante. (2020) "The Rise in Earnings Inequality: Does the Cycle Drive the Trend?," Discussion paper, Princeton University.

- Kaplan, G., B. Moll, and G. L. Violante. (2018) "Monetary Policy According to HANK," *American Economic Review*, 108, 697–743.
- Kaplan, G., and G. L. Violante. (2014) "A Model of the Consumption Response to Fiscal Stimulus Payments," *Econometrica*, 82, 1199–1239.
- Kaplan, G., G. L. Violante, and J. Weidner. (2014) "The Wealthy Hand-to-Mouth," *Brookings Papers on Economic Activity*, 48(1 (Spring), 77–153.
- Kekre, R. (2019) "Unemployment Insurance in Macroeconomic Stabilization," Discussion paper, Booth School of Business, University of Chicago.
- Leith, C., and E. Leeper. (2016) "Understanding Inflation as a Joint Monetary - Fiscal Phenomenon," Working papers 2016-01, Business School - Economics, University of Glasgow.
- Lenel, M., and R. Kekre. (2019) "Redistribution, Risk Premia, and the Macroeconomy," Discussion paper, Princeton University.
- Luetticke, R. (2015) "Transmission of Monetary Policy and Heterogeneity in Household Portfolios," Discussion paper, University of Bonn.
- McKay, A., E. Nakamura, and J. Steinsson. (2016) "The Power of Forward Guidance Revisited," *American Economic Review*, 106, 3133–58.
- McKay, A., and R. Reis. (2016) "The Role of Automatic Stabilizers in the U.S. Business Cycle," *Econometrica*, 84, 141–94.
- Oh, H., and R. Reis. (2012) "Targeted Transfers and the Fiscal Response to the Great Recession," *Journal of Monetary Economics*, 59, S50–S64.
- Patterson, C. (2018) "The Matching Multiplier and the Amplification of Recessions," Discussion paper, MIT.
- Ravn, M. O., and V. Sterk. (2017) "Job Uncertainty and Deep Recessions," *Journal of Monetary Economics*, 90, 125–41.
- Rotemberg, J. J. (1982) "Sticky Prices in the United States," *Journal of Political Economy*, 90, 1187–1211.
- Sims. (2004) "The Inflation-Targeting Debate," Discussion paper, Princeton University.
- Sims, C. A. (1980) "Macroeconomics and Reality," *Econometrica*, 48, 1–48.
- Werning, I. (2015) "Incomplete Markets and Aggregate Demand," Working Paper 21448, National Bureau of Economic Research.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.