

Monetary Policy According to HANK[†]

By GREG KAPLAN, BENJAMIN MOLL, AND GIOVANNI L. VIOLANTE*

We revisit the transmission mechanism from monetary policy to household consumption in a Heterogeneous Agent New Keynesian (HANK) model. The model yields empirically realistic distributions of wealth and marginal propensities to consume because of two features: uninsurable income shocks and multiple assets with different degrees of liquidity and different returns. In this environment, the indirect effects of an unexpected cut in interest rates, which operate through a general equilibrium increase in labor demand, far outweigh direct effects such as intertemporal substitution. This finding is in stark contrast to small- and medium-scale Representative Agent New Keynesian (RANK) economies, where the substitution channel drives virtually all of the transmission from interest rates to consumption. Failure of Ricardian equivalence implies that, in HANK models, the fiscal reaction to the monetary expansion is a key determinant of the overall size of the macroeconomic response. (JEL D31, E12, E21, E24, E43, E52, E62)

A prerequisite for the successful conduct of monetary policy is a satisfactory understanding of the monetary transmission mechanism—the ensemble of economic forces that determine how the actions of the monetary authority affect the aggregate performance of the economy. This paper follows the tradition of treating the short-term nominal interest rate as the primary monetary policy instrument and is concerned with its transmission to the largest component of GDP, household consumption.

Changes in interest rates influence household consumption through both direct and indirect effects. Direct effects are those that operate even in the absence of any change in household disposable labor income. The most important direct effect is intertemporal substitution: when real rates fall, households save less or borrow more and, therefore, increase their demand for consumption. In general equilibrium, additional indirect effects on consumption arise from the expansion in labor demand, and thus in labor income, that emanates from the direct impact of the original interest

* Kaplan: Department of Economics, University of Chicago, Saieh Hall, 5757 S. University Avenue, Chicago, IL 60637, and NBER (email: gkaplan@uchicago.edu); Moll: Department of Economics, Princeton University, Julis Romo Rabinowitz Building, Princeton, NJ 08542, and NBER (email: moll@princeton.edu); Violante: Department of Economics, Princeton University, Julis Romo Rabinowitz Building, Princeton, NJ 08542, CEPR, and NBER (email: glv2@princeton.edu). This paper was accepted to the *AER* under the guidance of John Leahy, Coeditor. We thank Yves Achdou, Mark Aguiar, Fernando Alvarez, Adrien Auclert, Jess Benhabib, Luca Dedola, Emmanuel Farhi, Mark Gertler, Narayana Kocherlakota, Keith Kuester, David Lagakos, Emi Nakamura, Larry Schmidt, Jón Steinsson, Mirko Wiederholt, and seminar participants at various institutions. Felipe Alves, Damien Capelle, and Julia Fonseca provided superb research assistance.

[†]Go to <https://doi.org/10.1257/aer.20160042> to visit the article page for additional materials and author disclosure statement(s).

rate cut. The relative magnitude of the direct and indirect channels is determined by how strongly household consumption responds to changes in real interest rates given income, and to changes in disposable income given the real rate.

Our first result concerns Representative Agent New Keynesian (RANK) models. In these commonly used benchmark economies, the aggregate consumption response to a change in interest rates is driven entirely by the Euler equation of the representative household. Therefore, for any reasonable parameterization, monetary policy in RANK models works almost exclusively through intertemporal substitution: direct effects account for nearly the entire impact of interest rate changes on the macroeconomy and indirect effects are negligible.

The strong response of aggregate consumption to movements in real rates that accounts for the large direct effects in RANK is questionable in light of empirical evidence. Macroeconometric analysis of aggregate time-series data finds a small sensitivity of consumption to changes in the interest rate after controlling for income (Campbell and Mankiw 1989; Yogo 2004; Canzoneri, Cumby, and Diba 2007). Crucially, this finding does not necessarily imply that the individual intertemporal elasticity of substitution is small, as other offsetting direct effects can be powerful. First, micro survey data on household portfolios show that a sizable fraction of households (between one-quarter and one-third) hold close to zero liquid wealth and face high borrowing costs (Kaplan, Violante, and Weidner 2014). Since these households are at a kink in their budget set, they are insensitive to small changes in interest rates (consistent with evidence in Vissing-Jorgensen 2002 that non-asset-holders do not react to interest rate cuts). Moreover, the possibility of hitting a kink in the future effectively shortens the time horizon and dampens the substitution effect even for those households with positive holdings of liquid wealth. Second, standard consumption theory implies that an interest rate cut has negative income effects on the consumption of rich households. Third, these same survey data reveal vast inequality in wealth holdings and composition across households (Díaz-Giménez, Glover, and Ríos-Rull 2011). Some households may react to a short-term rate cut by rebalancing their asset portfolio rather than by saving less and consuming more.

The small indirect effects in RANK models follow from the property that the representative agent is, in essence, a permanent income consumer and so is not responsive to transitory income changes. This type of consumption behavior is at odds with a vast macro and micro empirical literature (Jappelli and Pistaferri 2010). The most convincing corroboration of this behavior is the quasi-experimental evidence that uncovers (i) an aggregate quarterly marginal propensity to consume (MPC) out of small transitory government transfers of around 25 percent (Johnson, Parker, and Souleles 2006; Parker et al. 2013) and (ii) a vast heterogeneity in consumption responses across the population which is largely driven by the level of liquid wealth and by the composition of household balance sheets (Misra and Surico 2014; Cloyne and Surico 2016; Broda and Parker 2014).¹

In light of this empirical evidence, we argue that the relative strength of the direct and indirect channels of monetary policy can be properly gauged only

¹ A recent body of work estimating the marginal propensity to consume out of changes in housing net worth also documents consumption responses that are very heterogeneous and heavily dependent on portfolio composition (e.g., Mian, Rao, and Sufi 2013).

within a framework that offers a better representation of household consumption and household finances than RANK. To this end, we develop a quantitative Heterogeneous Agent New Keynesian (HANK) model that combines two leading workhorses of modern macroeconomics. On the household side, we build on the standard Aiyagari-Huggett-İmrohoroğlu incomplete market model, with one important modification: as in Kaplan and Violante (2014), households can save in two assets, a low-return liquid asset and a high-return illiquid asset that is subject to a transaction cost. This extended model has the ability to be consistent with the joint distribution of earnings, liquid wealth and illiquid wealth, as well as with the sizable aggregate MPC out of small windfalls. The remaining blocks of the model follow the New Keynesian tradition. On the supply side, prices are set by monopolistically competitive producers who face nominal rigidities. We close the model by assuming that monetary policy follows a Taylor rule.

Our main finding is that in stark contrast to RANK economies, the direct effects of interest rate shocks in our HANK model are always small, while the indirect effects can be substantial. Monetary policy is effective only to the extent that it generates a general equilibrium response in household disposable income. In our framework, by virtue of this indirect channel, overall consumption responses can be large, even though the strength of the direct channel is modest.

The sharply different consumption behavior between RANK and HANK lies at the heart of these results. Uninsurable risk, combined with the coexistence of liquid and illiquid assets in financial portfolios, leads to the presence of a sizable fraction of poor and wealthy hand-to-mouth households, as in the data. These households are highly sensitive to labor income shocks but are not responsive to interest rate changes. Moreover, the vast inequality in liquid wealth implies that even for non-hand-to-mouth households, a cut in liquid rates leads to strong offsetting income effects on consumption. Finally, with this multiple asset structure, to the extent that the spread between asset returns widens after a monetary expansion, household portfolios adjust away from liquid holdings and toward more lucrative assets rather than toward higher consumption expenditures. All these economic forces counteract the intertemporal substitution effect and lower the direct channel of monetary policy in HANK.

A second important finding is that in HANK the consequences of monetary policy are intertwined with the fiscal side of the economy, because of the failure of Ricardian equivalence. Since the government is a major issuer of liquid obligations, a change in the interest rate necessarily affects the intertemporal government budget constraint and generates some form of fiscal response that affects household disposable income. Unlike in RANK models, the details of this response matter a great deal for the overall macroeconomic impact of a monetary shock and for its split between direct and indirect channels, both in terms of its timing and distributional burden across households.²

Why is it important to correctly quantify the direct and indirect channels of the monetary transmission mechanism? To give a concrete answer to this question, we compare RANK and HANK along two key trade-offs that policymakers face in the

²The importance of government debt for the monetary transmission mechanism is also emphasized by Sterk and Tenreyro (2015) in a model with flexible prices and heterogeneous households where open market operations have distributional wealth effects and by Eusepi and Preston (2017) in a model in which Ricardian equivalence fails because of imperfect knowledge.

conduct of monetary policy. First, when attempting to stimulate the macroeconomy, the monetary authority faces a choice between large but transitory versus small but persistent nominal rate cuts. In RANK models, transitory rate cuts and persistent rate cuts are equally powerful, as long as the cumulative interest rate deviations are the same. Instead, in HANK a less persistent but larger rate cut can be more effective at expanding aggregate consumption because it leads to a more immediate reduction in interest payments on government debt that translate into additional fiscal stimulus. Second, we analyze the inflation-activity trade-off. The slope of this trade-off is not too different in the two economies because it is the common New Keynesian side of the models that largely pins down the relationship. However, in HANK the slope depends on the type of fiscal adjustment: more passive adjustment rules, where government debt absorbs the change in interest payments, are associated with a more favorable trade-off for the monetary authority.

Taking a broader perspective, there are additional reasons why it is important to develop a full grasp of the monetary transmission. First, as economists, we strive to gather well-identified and convincing empirical evidence on all the policy experiments we contemplate. However, this is not always feasible, as demonstrated by the recent experience of central banks that were forced to deal with a binding zero lower bound by turning to previously unused policy instruments. In these circumstances, well-specified structural models are especially useful to extrapolate from the evidence we already have.

Moreover, the relative size of direct versus indirect effects determines the extent to which central banks can precisely target the expansionary impact of their interventions. When direct effects are dominant, as in a RANK model, for the monetary authority to boost aggregate consumption it is sufficient to influence real rates: intertemporal substitution then ensures that expenditures will respond. In a HANK model, instead, the monetary authority must rely on equilibrium feedbacks that boost household income in order to influence aggregate consumption. Reliance on these indirect channels means that the overall effect of monetary policy may be more difficult to fine-tune by manipulating the nominal rate. The precise functioning of complex institutions, such as labor and financial markets, and the degree of coordination with the fiscal authority play an essential role in mediating the way that monetary interventions affect the macroeconomy.

We are not the first to integrate incomplete markets and nominal rigidities, and there is a burgeoning literature on this topic.³ Relative to this literature, our paper adds an empirically realistic model of the consumption side of the economy by exploiting state-of-the-art ideas for modeling household consumption and the joint distribution of income and wealth. The combination of uninsurable earnings risk and a two-asset structure is at the root of our finding that most of the monetary transmission is due to indirect general equilibrium effects. In the paper, we show that the one-asset model explored by the whole literature up to this point faces a daunting challenge when used to study monetary policy. If calibrated to match total wealth in the economy, it implies a very small MPC (similar to the one in RANK) and

³ See Guerrieri and Lorenzoni (2017); Oh and Reis (2012); Ravn and Sterk (2017); McKay and Reis (2016); Gornemann, Kuester, and Nakajima (2014); Auclert (2016); McKay, Nakamura, and Steinsson (2016); Den Haan, Rendahl, and Riegler (2017); Bayer et al. (2015); Luetticke (2015); and Werning (2015).

enormous income effects on consumption because all wealth is liquid. If calibrated to match only *liquid* wealth, it features a large aggregate MPC out of transitory income and reasonable income effects. However, because such calibration misses over 95 percent of the wealth in the economy, the model must completely abstract from some key sources of indirect effects of monetary policy, such as those originating from firm investment and from movements in the price of capital.

Additionally, the focus of our paper differs from that of earlier papers studying monetary policy in the presence of incomplete markets (Gornemann, Kuester, and Nakajima 2014; McKay, Nakamura, and Steinsson 2016) in that we inspect the transmission mechanism of monetary policy and decompose it into direct and indirect general equilibrium effects. Our emphasis on general equilibrium effects is shared by Werning (2015), who develops a useful theoretical benchmark where direct and indirect channels exactly offset each other so that the overall effect of interest rate changes on consumption is unchanged relative to the RANK benchmark. Since Werning's assumptions do not hold in our economy, the presence of heterogeneity and incomplete markets affects both the decomposition and the overall effect of monetary policy in our model. Conceptually, our decomposition is similar to the one proposed by Auclert (2016).

Our paper is also related to the literature that studies New Keynesian models with limited heterogeneity, building on the spender-saver model of Campbell and Mankiw (1989).⁴ The "spenders" in these models consume their entire income every period and therefore share some similarities with our hand-to-mouth households in that they do not respond to interest rate changes. However, these Two-Agent New Keynesian (TANK) models also feature "savers" who engage in intertemporal substitution and are highly responsive to interest rate changes. In contrast, in our model even high liquid-wealth households do not increase consumption much in response to an interest rate cut because the risk of receiving negative income shocks and binding liquidity constraints in the future truncates their effective time horizon. We show that, when the fraction of spenders reflects the share of hand-to-mouth households in the data, also TANK models feature a monetary transmission mechanism with large direct effects. Our emphasis on indirect channels is shared by Caballero and Farhi (2018), who propose an alternative framework where the transmission of monetary policy works through its general equilibrium impact on asset values.

Finally, we solve the model in continuous time building on Achdou et al. (2017). In addition to imparting some notable computational advantages, continuous time provides a natural and parsimonious approach to modeling an individual earnings process with leptokurtic annual income growth, as recently documented by Guvenen et al. (2015): random (Poisson) arrival of normally distributed jumps generates kurtosis in data observed at discrete time intervals. This process, estimated by matching targets from Social Security Administration data, may prove useful in other contexts where an empirically realistic representation of household earning dynamics is vital.

The rest of the paper proceeds as follows. Section I introduces the idea of decomposing the monetary transmission mechanism into direct and indirect effects, and applies it to small- and medium-scale RANK models and spender-saver models.

⁴See Iacoviello (2005); Galí, López-Salido, and Vallés (2007); Bilbiie (2008); and Challe et al. (2017).

Section II lays out our HANK framework and Section III describes how we take it to the data. Section IV contains our quantitative analysis of monetary policy in HANK, and Section V examines the implications of our findings for some key trade-offs faced by policymakers in the conduct of monetary policy. Section VI concludes.

I. Monetary Policy in Benchmark New Keynesian Models

In this section, we introduce a formal decomposition of the consumption response to a one-time unexpected interest rate shock into direct and indirect effects.⁵ Since this decomposition is instrumental to our analysis of the transmission of monetary policy in our larger quantitative model, we begin by applying it to a series of stylized New Keynesian models. We first demonstrate that in representative agent economies, conventional monetary policy works almost exclusively through direct intertemporal substitution and that indirect general equilibrium effects are unimportant. Next, we illustrate how the monetary transmission mechanism is affected by the presence of non-Ricardian hand-to-mouth households: (i) the introduction of hand-to-mouth households increases the relative share of indirect general equilibrium effects; (ii) the overall effect of monetary policy depends on the fiscal response that necessarily arises because monetary policy affects the government budget constraint. Finally, we show that these insights carry over to richer representative agent economies, such as typical medium-scale New Keynesian dynamic stochastic general equilibrium (DSGE) models. Online Appendix A contains proofs of the results in this section.

A. Representative Agent Model

Setup.—A representative household has constant relative risk aversion (CRRA) utility from consumption C_t with parameter $\gamma > 0$, and discounts the future at rate $\rho \geq 0$. A representative firm produces output using only labor, according to the production function $Y = N$. Both the wage and final goods price are perfectly rigid and normalized to 1. The household commits to supplying any amount of labor demanded at the prevailing wage so that its labor income equals Y_t in every instant. The household receives (pays) lump-sum government transfers (taxes) $\{T_t\}_{t \geq 0}$ and can borrow and save in a riskless government bond at rate r_t . Its initial bond holdings are B_0 . In the absence of aggregate uncertainty, household optimization implies that the time path of consumption satisfies the Euler equation $\dot{C}_t/C_t = \frac{1}{\gamma}(r_t - \rho)$. The government sets the path of taxes in a way that satisfies its intertemporal budget constraint.

Since prices are fixed, the real interest rate r_t also equals the nominal interest rate, so effectively the monetary authority sets an exogenous time path for real rates $\{r_t\}_{t \geq 0}$. We restrict attention to interest rate paths with the property that $r_t \rightarrow \rho$ as $t \rightarrow \infty$ so that the economy converges to an interior steady state. Our results place no additional restrictions on the path of interest rates. However, clean and intuitive formulae can be obtained for the special case

$$(1) \quad r_t = \rho + e^{-\eta t}(r_0 - \rho), \quad t \geq 0,$$

⁵This section benefited greatly from detailed comments by Emmanuel Farhi and some of the results directly reflect those comments.

where the interest rate unexpectedly jumps at $t = 0$ and then mean reverts at rate $\eta > 0$. In equilibrium, the goods market clears $C_t(\{r_t, Y_t, T_t\}_{t \geq 0}) = Y_t$, where $C_t(\{r_t, Y_t, T_t\}_{t \geq 0})$ is the optimal consumption function for the household. We assume that the economy returns to its steady-state level in the long-run, $C_t \rightarrow \bar{C} = \bar{Y}$ as $t \rightarrow \infty$.^{6,7}

Overall Effect of Monetary Policy.—We can analyze the effects of a change in the path of interest rates on consumption using only two conditions: the household Euler equation and our assumption that consumption returns back to its steady-state level. It therefore follows that $C_t = \bar{C} \exp\left(-\frac{1}{\gamma} \int_t^\infty (r_s - \rho) ds\right)$. When the path of interest rates satisfies (1), this formula collapses to a simple expression for the elasticity of initial consumption to the initial change in the interest rate

$$(2) \quad \frac{d \log C_0}{dr_0} = -\frac{1}{\gamma \eta}.$$

The response of consumption is large if the elasticity of substitution $1/\gamma$ is high, and if the monetary expansion is persistent (η is low).

Note that if initial government debt is positive, $B_0 > 0$, then a drop in interest rates necessarily triggers a fiscal response. This is because the time path of taxes must satisfy the government budget constraint, and therefore depends on the path of interest rates: $T_t = T_t(\{r_s\}_{s \geq 0})$. The government pays less interest on its debt and so will eventually rebate this income gain to households. However, Ricardian equivalence implies that the particular path of taxes chosen by the government does *not* affect the consumption response to monetary policy. In present value terms, the government’s gain from lower interest payments is exactly offset by the household’s loss from lower interest receipts.

Decomposition into Direct and Indirect Effects.—We begin with the case of zero government debt, $B_t = 0$ (and $T_t = 0$) for all t . We use a perturbation argument around the steady state. Assume that initially $r_t = \rho$ for all t so that $Y_t = \bar{Y}$ for all t . Now consider a small change to the path of interest rates $\{dr_t\}_{t \geq 0}$, while holding the path of income $\{Y_t\}_{t \geq 0}$ constant. The effect of this change in interest rates on consumption is the *direct effect*. In equilibrium, the consumption change induces changes in labor income $\{dY_t\}_{t \geq 0}$ which lead to further changes in consumption. This is the *indirect effect*. Formally, these two effects are defined by totally differentiating the initial consumption function $C_0(\{r_t, Y_t\}_{t \geq 0})$:

$$(3) \quad dC_0 = \underbrace{\int_0^\infty \frac{\partial C_0}{\partial r_t} dr_t dt}_{\text{direct response to } r} + \underbrace{\int_0^\infty \frac{\partial C_0}{\partial Y_t} dY_t dt}_{\text{indirect effects due to } Y}.$$

⁶There are multiple equilibria in this economy. We select an equilibrium by anchoring the economy in the long run and focusing only on paths for which $Y_t \rightarrow \bar{Y}$ as $t \rightarrow \infty$ for some fixed $0 < \bar{Y} < \infty$. For any value of steady-state output \bar{Y} , the equilibrium is then unique. Since we are only concerned with deviations of consumption and output from steady state, the level of \bar{Y} is not important for any of our results.

⁷Rather than assuming that wages and prices are perfectly rigid, our equilibrium could be viewed as a “demand-side equilibrium” as in Werning (2015). In this interpretation, we characterize the set of time paths $\{r_t, Y_t\}_{t \geq 0}$ that are consistent with optimization on the demand (household) side of the economy without specifying the supply (firm) side. Our results thus apply in richer environments such as the textbook three-equation New Keynesian model.

The income deviations $\{dY_t\}_{t \geq 0}$ are equilibrium outcomes induced by the changes in interest rates, which satisfy $d \log Y_t = -\frac{1}{\gamma} \int_t^\infty dr_s ds$.⁸

The key objects in the decomposition (3) are the partial derivatives of the consumption function $\partial C_0 / \partial r_t$ and $\partial C_0 / \partial Y_t$, i.e., the household's responses to interest rate and income changes. In this simple model, these two derivatives can be computed analytically which leads to the main result of this section.⁹

PROPOSITION 1: *Consider small deviations dr_t of the interest rate from steady state. The overall effect on initial consumption $d \log C_0 = -\frac{1}{\gamma} \int_0^\infty dr_t dt$ can be decomposed as*

$$(4) \quad d \log C_0 = \underbrace{-\frac{1}{\gamma} \int_0^\infty e^{-\rho t} dr_t dt}_{\text{direct response to } r} - \underbrace{\frac{\rho}{\gamma} \int_0^\infty e^{-\rho t} \int_t^\infty dr_s ds dt}_{\text{indirect effects due to } Y}$$

The decomposition is additive, i.e., the two components sum to the overall effect.

This decomposition of the initial consumption response holds for *any* time path of interest rate changes $\{dr_t\}_{t \geq 0}$. The relative importance of each effect *does not* depend on the intertemporal elasticity of substitution $1/\gamma$.

When the interest rate path follows (1), the decomposition becomes

$$(5) \quad -\frac{d \log C_0}{dr_0} = \frac{1}{\gamma \eta} \left(\underbrace{\frac{\eta}{\rho + \eta}}_{\text{direct response to } r} + \underbrace{\frac{\rho}{\rho + \eta}}_{\text{indirect effects due to } Y} \right).$$

The split between direct and indirect effect depends only on the discount rate ρ and the rate of mean reversion η . A higher discount rate implies a smaller direct effect and a larger indirect general equilibrium effect. This reflects the fact that: (i) in this model, the marginal propensity to consume out of current income is equal to the discount rate; and (ii) the lower is η the larger is the impact of the interest rate change on the permanent component of labor income.

One important implication of equation (5) is that, for any reasonable parameterization, the indirect effect is very small, and monetary policy works almost exclusively through the direct channel. For example, a quarterly steady-state interest rate of 0.5 percent (2 percent annually, as we assume in our quantitative analysis later in the paper) implies $\rho = 0.5\%$. Suppose the monetary policy shock mean reverts at rate $\eta = 0.5$, i.e., a quarterly autocorrelation of $e^{-\eta} = 0.61$.¹⁰ Then, the direct effect accounts for $\eta/(\rho + \eta) = 99\%$ of the overall effect. Even with a quarterly autocorrelation of 0.95 ($\eta = 0.05$) (an implausibly persistent monetary shock,

⁸ Adjustments in income dY_t can themselves be further decomposed into direct effects and indirect general equilibrium effects. We nevertheless find this version of the decomposition especially useful. In particular, it allows us to distinguish whether, following a change in interest rates, individual households primarily respond through intertemporal substitution in and of itself or to changes in their labor income.

⁹ See Theorem 3 in Auclert (2016) for a related decomposition.

¹⁰ This value implies the shock is fully reabsorbed after around six quarters. This speed of mean-reversion is consistent with the dynamics of a shock to the federal funds rate commonly estimated by vector autoregressions (VARs). See Christiano, Eichenbaum, and Evans (2005) and Gertler and Karadi (2015).

from an empirical standpoint), the contribution of the direct effect would still be above 90 percent.¹¹

This result extends to the case where government debt is nonzero, $B_0 > 0$. When the government issues debt, in equilibrium a monetary expansion must trigger a fiscal response $T_t = T_t(\{r_s\}_{s \geq 0})$ in order to satisfy the government budget constraint. Because household consumption $C_t(\{r_t, Y_t, T_t\}_{t \geq 0})$ depends on taxes/transfers, the direct-indirect decomposition becomes

$$(6) \quad dC_0 = \underbrace{\int_0^\infty \frac{\partial C_0}{\partial r_t} dr_t dt}_{\text{direct response to } r} + \underbrace{\int_0^\infty \left(\frac{\partial C_0}{\partial Y_t} dY_t + \frac{\partial C_0}{\partial T_t} dT_t \right) dt}_{\text{indirect effects}}$$

Thus, in the special case (1) where interest rates mean-revert at rate η , we have

$$(7) \quad -\frac{d \log C_0}{dr_0} = \underbrace{\frac{1}{\gamma \eta} \left(\frac{\eta}{\rho + \eta} \left(1 - \rho \gamma \frac{B_0}{\bar{Y}} \right) \right)}_{\text{direct response to } r} + \underbrace{\frac{\rho}{\rho + \eta}}_{\text{indirect effects due to } Y} + \underbrace{\frac{\eta}{\rho + \eta} \rho \gamma \frac{B_0}{\bar{Y}}}_{\text{indirect effects due to } T}$$

As already noted, due to Ricardian neutrality, the overall effect of monetary policy is independent of fiscal policy. Relative to (5), though, the presence of government debt reduces the direct effect. This is because households now own some wealth and hence experience a negative (capital) income effect following an interest rate cut. Under Ricardian equivalence, this reduction in the direct component is exactly offset by an additional indirect effect due to the corresponding increase in transfers.

The relative share of these two components depends on the debt-to-GDP ratio B_0/\bar{Y} . With large enough government debt, direct effects can be small even in RANK. However, for plausible debt levels, the decomposition is hardly affected relative to (5). For instance, with log utility ($\gamma = 1$), only with a quarterly debt-to-GDP ratio B_0/\bar{Y} above 20, would the direct component $\frac{\eta}{\rho + \eta} \left(1 - \rho \gamma \frac{B_0}{\bar{Y}} \right)$ fall below 90 percent of the total.

B. Non-Ricardian Hand-to-Mouth Households

We now introduce “rule-of-thumb” households as in Campbell and Mankiw (1989, 1991) and Bilbiie (2008, 2017). The setup is identical, except that we assume that a fraction Λ of households consume their entire current income, i.e., per capita consumption of these “spenders” is given by $C_t^{sp} = Y_t + T_t^{sp}$, where T_t^{sp} is a lump-sum transfer to spenders. Spenders therefore have a marginal propensity to consume out of labor income and transfers equal to 1. The remaining fraction $1 - \Lambda$ of households optimize as before, yielding a consumption function for these “savers” $C_t^{sa}(\{r_t, Y_t, T_t^{sa}\}_{t \geq 0})$. Aggregate consumption is given by $C_t = \Lambda C_t^{sp} + (1 - \Lambda)C_t^{sa}$. In equilibrium $C_t = Y_t$.

¹¹As suggested by John Cochrane (<https://johnhcochrane.blogspot.com/2015/08/whither-inflation.html>), a better name for the standard New Keynesian model may therefore be the “sticky-price intertemporal substitution model.”

The results from RANK extend in a straightforward fashion to this Two-Agent New Keynesian (TANK) economy. Consider first the case in which $B_t = 0$ for all t . For brevity, we only analyze the generalization of (5):

$$(8) \quad -\frac{d \log C_0}{dr_0} = \frac{1}{\gamma\eta} \left(\underbrace{\left(1 - \Lambda\right) \frac{\eta}{\rho + \eta}}_{\text{direct response to } r} + \underbrace{\left(\left(1 - \Lambda\right) \frac{\rho}{\rho + \eta} + \Lambda\right)}_{\text{indirect effects due to } Y} \right).$$

Note first that the total aggregate effect of monetary policy is exactly as in RANK. The contribution of the direct effect and indirect effects are each a weighted average of the corresponding quantities for spenders and savers, with the weights equal to each group's population share. Since the direct effect for spenders is 0 and the indirect effect is 1, the overall share of the indirect effect approximately equals the population fraction of spenders Λ . A reasonable estimate for the proportion of hand-to-mouth households in the United States is 0.3 (Kaplan, Violante, and Weidner 2014). Thus, in TANK the share of direct effects is roughly 0.7.

The overall effect in TANK is the same as in RANK because the addition of hand-to-mouth households decreases direct effects and increases indirect effects by the same magnitude. To see this, note that aggregate consumption is given by $C_t = Y_t = \Lambda Y_t + (1 - \Lambda) C_t^{sa}$ where consumption of savers is pinned down from the time path of interest rates $C_t^{sa} = \bar{C} \exp(-\frac{1}{\gamma} \int_t^\infty (r_s - \rho) ds)$. Equivalently, $C_t = M \times (1 - \Lambda) C_t^{sa}$ where $M = \frac{1}{1 - \Lambda} > 1$ is a multiplier. The presence of hand-to-mouth households scales down direct effects by a factor $1 - \Lambda$, but these then get scaled up again, through equilibrium feedbacks, by an exactly offsetting factor $\frac{1}{1 - \Lambda}$. This is the same logic that lies behind a result of Werning (2015), who showed that in a particular sticky-price economy with heterogeneous agents and incomplete markets, direct and indirect channels exactly offset so that the overall effect of interest rate changes on consumption is unchanged relative to the representative agent complete markets benchmark. In Werning's economy, as well as in our toy model, labor is demand-determined and, therefore, labor supply plays no role. Bilbiie (2008, 2017) studies the monetary transmission mechanism in a TANK model with endogenous labor supply. His analysis implies that this "as-if" result holds only in the knife-edge case of infinite labor supply elasticity.¹²

Next, we consider the case where the government issues debt $B_0 > 0$. As in Section IA, a change in the path of interest rates affects the government budget constraint and induces a fiscal response. Because Ricardian equivalence need not hold in the spender-saver economy, the effect of monetary policy depends on the specifics of this fiscal response. As long as the fiscal response entails increasing transfers to the hand-to-mouth households, this will increase the overall response of aggregate consumption to monetary policy. This mechanism can be seen most clearly in the case of the exponentially decaying interest rate path (1). Let us assume that the government keeps debt constant at its initial level, $B_t = B_0$ for all t , and transfers a

¹²The equivalence result between TANK and RANK derived in (8) also depends on the identity $C_t = Y_t$ and hence on the fact that this model does not feature capital and investment. In the presence of investment, the introduction of hand-to-mouth households has ambiguous effects on the elasticity of aggregate consumption. In particular, we would have $C_t + I_t = Y_t = M \times ((1 - \Lambda) C_t^{sa} + I_t)$ and hence $C_t = C_t^{sa} + \frac{\Lambda}{1 - \Lambda} I_t$ so that the elasticity may be larger or smaller depending on the magnitude of $\Lambda/(1 - \Lambda)$ as well as other factors determining the size of the investment response.

fraction Λ^T of the income gains from lower interest payments to spenders (and the residual fraction to savers) so that $\Lambda T_t^{sp}(\{r_s\}_{s>0}) = -(r_t - \rho) \Lambda^T B_0$.¹³ Then, the response of aggregate consumption at impact is

$$(9) \quad -\frac{d \log C_0}{dr_0} = \frac{1}{\gamma \eta} + \frac{\Lambda^T B_0}{1 - \Lambda \bar{Y}}.$$

Note the presence of the additional term $\Lambda^T (B_0/Y)$. The overall effect of monetary policy differs from RANK only if there is *both* a debt-issuing government ($B_0 > 0$) and non-Ricardian hand-to-mouth households who receive a positive share of the transfers ($\Lambda^T > 0$). It is only under this scenario that the indirect component of the transmission mechanism could be much larger in TANK models compared to RANK models (for the decomposition corresponding to equation (9), see equation (A.25) in the online Appendix).

C. Richer RANK and TANK Models

Is our finding that conventional monetary policy works almost exclusively through direct intertemporal substitution special to these simple models? Compared to typical medium-scale New Keynesian DSGE models used in the literature, the RANK model in the present section is extremely stylized. For instance, state-of-the-art medium-scale DSGE models typically feature investment subject to adjustment costs, variable capital utilization, habit formation, and prices and wages that are partially sticky as opposed to perfectly rigid. We therefore conducted a decomposition analogous to that in equation (4) in one such state-of-the-art framework, the Smets and Wouters (2007) model (see online Appendix A.4 for details). The result confirms our findings: 99 percent of the consumption response to an expansionary monetary policy shock is accounted for by direct intertemporal substitution effects. The reason is that none of the additional features of this richer model change the property that the consumption of the representative agent is insensitive to the transitory income changes resulting from monetary shocks.¹⁴

We have also solved numerically versions of RANK and TANK models which, like the HANK model that follows, feature government debt and capital in positive supply, a New Keynesian production side with Rotemberg-style price adjustment costs, and a Taylor rule. These models, which are fully described in online Appendix A.5, are designed to be as close as possible to HANK, except for the nature of household heterogeneity. Comparing column 4 of Table 1 with columns 1 and 2 (for RANK), and comparing column 7 with columns 5 and 6 (for TANK) illustrates that the simple models of Sections IA and IB approximate well these richer economies both in terms of the size of the total consumption response and its decomposition into direct and indirect shares.

¹³This is equivalent to assuming that the government maintains budget balance by adjusting lump-sum transfers, which is the baseline assumption we make in our full quantitative model.

¹⁴With Smets and Wouters' baseline parameterization, the total elasticity for consumption at impact is -0.74 , which is substantially smaller than that of our stylized models. The key reason is that their model features habit formation in consumption which mutes the consumption response at impact. We conducted a number of robustness checks, particularly with respect to the habit formation parameter which directly enters the representative agent's Euler equation, and found that the share due to direct effects never drops below 90 percent.

TABLE 1—ELASTICITY OF AGGREGATE CONSUMPTION AND SHARE OF DIRECT EFFECTS IN SEVERAL VERSIONS OF THE RANK AND TANK MODELS

	RANK				TANK		
	$B = 0$	$B > 0$	S-W	$B, K > 0$	$B = 0$	$B > 0$	$B, K > 0$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Elasticity of C	-2.00	-2.00	-0.74	-2.07	-2.00	-2.43	-2.77
PE. elast. of C	-1.98	-1.96	-0.73	-1.95	-1.38	-1.39	-1.39
Direct effects (%)	99	98	99	94	69	57	50

Notes: “ $B = 0$ ” denotes the simple models of Section I with wealth in zero net supply. “ $B > 0$ ” denotes the extension of these models with government bonds in positive net supply. In RANK, we set $\gamma = 1, \eta = 0.5, \rho = 0.005$, and $B_0/Y = 1$. In addition, in TANK we set $\Lambda = \Lambda^T = 0.3$. “S – W” is the medium-scale version of the RANK model described in online Appendix A.4 based on Smets-Wouters. “ $B, K > 0$ ” denotes the richer version of the representative-agent and spender-saver New Keynesian model featuring a two-asset structure, as in HANK. See online Appendix A.5 for a detailed description of this model and its calibration. In all economies with bonds in positive supply, lump-sum transfers adjust to balance the government budget constraint. “PE. elast. of C ” is the partial equilibrium (or direct) elasticity computed as total elasticity times the share of direct effects.

II. HANK: A Framework for Monetary Policy Analysis

We now turn to our paper’s main contribution: the development and analysis of our Heterogeneous Agent New Keynesian (HANK) model. Our main innovation is a rich representation of household consumption and saving behavior. Households face uninsurable idiosyncratic income risk which they can self-insure through two savings instruments with different degrees of liquidity. The rest of the model is purposefully kept simple and as close as possible to the New Keynesian literature: there is price stickiness and a monetary authority that operates a Taylor rule, and we analyze the economy’s response to an innovation to this Taylor rule. For simplicity, we consider a deterministic transition following a one-time zero-probability shock.

A. The Model

Households.—The economy is populated by a continuum of households indexed by their holdings of liquid assets b , illiquid assets a , and their idiosyncratic labor productivity z . Labor productivity follows an exogenous Markov process that we describe in detail in Section IIIB. Time is continuous. At each instant in time t , the state of the economy is the joint distribution $\mu_t(da, db, dz)$. Households die with an exogenous Poisson intensity ζ , and upon death give birth to an offspring with zero wealth and labor productivity equal to a random draw from its ergodic distribution.¹⁵ There are perfect annuity markets so that the estates of the deceased are redistributed to other individuals in proportion to their asset holdings.¹⁶

¹⁵We 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 $\zeta = 0$.

¹⁶The 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.

Households receive a utility flow u from consuming $c_t \geq 0$ and a disutility flow from supplying labor ℓ_t , where $\ell_t \in [0, 1]$ are hours worked as a fraction of the time endowment, normalized to 1. The function u is strictly increasing and strictly concave in consumption, and strictly decreasing and strictly convex in hours worked. Preferences are time-separable and, conditional on surviving, the future is discounted at rate $\rho \geq 0$:

$$(10) \quad E_0 \int_0^\infty e^{-(\rho+\zeta)t} u(c_t, \ell_t) dt,$$

where the expectation is taken over realizations of idiosyncratic productivity shocks. Because of the law of large numbers, and the absence of aggregate shocks, there is no economy-wide uncertainty.

Households can borrow in liquid assets b up to an exogenous limit \underline{b} at the real interest rate of $r_t^{b-} = r_t^b + \kappa$, where $\kappa > 0$ is an exogenous wedge between borrowing and lending rates. With a slight abuse of notation, $r_t^b(b_t)$ summarizes the full interest rate schedule.

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. We use d_t to denote a household’s deposit rate (with $d_t < 0$ corresponding to withdrawals) and $\chi(d_t, a_t)$ to denote 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, i.e., $r_t^a > r_t^b$. Short positions in illiquid assets are not allowed.

A household’s asset holdings evolve according to

$$(11) \quad \dot{b}_t = (1 - \tau_t)w_t z_t \ell_t + r_t^b(b_t)b_t + T_t - d_t - \chi(d_t, a_t) - c_t,$$

$$(12) \quad \dot{a}_t = r_t^a a_t + d_t,$$

$$(13) \quad b_t \geq -\underline{b}, \quad a_t \geq 0.$$

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 T_t) net of deposits into or withdrawals from the illiquid account d_t , transaction costs $\chi(d_t, a_t)$, and consumption expenditures c_t . 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 assets, we net out assets and liabilities within the two asset classes. That is, ours is not a model of gross positions.

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

$$(14) \quad \chi(d, a) = \chi_0 |d| + \chi_1 \left| \frac{d}{a} \right|^{x_2} a.$$

This transaction cost has two components that play distinct roles. The linear component generates an inaction region in households’ optimal deposit policies because for some households the marginal gain from depositing or withdrawing the first dollar is smaller than the marginal cost of transacting $\chi_0 > 0$. The convex component

($\chi_1 > 0, \chi_2 > 1$) 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.¹⁷

Households maximize (10) subject to (11)–(14). They take as given equilibrium paths for the real wage $\{w_t\}_{t \geq 0}$, the real return to liquid assets $\{r_t^b\}_{t \geq 0}$, the real return to illiquid assets $\{r_t^a\}_{t \geq 0}$, and taxes and transfers $\{\tau_t, T_t\}_{t \geq 0}$. As we explain below, $\{r_t^b\}_{t \geq 0}$ will be determined by monetary policy and a Fisher equation, and $\{w_t\}_{t \geq 0}$ and $\{r_t^a\}_{t \geq 0}$ will be determined by market clearing conditions for capital and labor. In online Appendix B.1 we describe the household's problem recursively with a Hamilton-Jacobi-Bellman equation. In steady state, the recursive solution to this problem consists of decision rules for consumption $c(a, b, z; \Gamma)$, deposits $d(a, b, z; \Gamma)$, and labor supply $\ell(a, b, z; \Gamma)$, with $\Gamma := (r^b, r^a, w, \tau, T)$.¹⁸ These decision rules imply optimal drifts for liquid and illiquid assets and, together with a stochastic process for z , they induce a stationary joint distribution of illiquid assets, liquid assets, and labor income $\mu(da, db, dz; \Gamma)$. In the online Appendix, we also describe the Kolmogorov forward equation that characterizes this distribution. Outside of steady state, each of these objects is time-varying and depends on the time path of prices and policies $\{\Gamma_t\}_{t \geq 0} := \{r_t^b, r_t^a, w_t, \tau_t, T_t\}_{t \geq 0}$.

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

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

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

$$(15) \quad y_{j,t} = k_{j,t}^\alpha n_{j,t}^{1-\alpha}.$$

¹⁷ 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 less than \$500 in all calibrations) that guarantees costs remain finite even for households with $a = 0$.

¹⁸ In what follows, when this does not lead to confusion, we suppress the explicit dependence of decision rules on the vector of prices and policies Γ .

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

$$(16) \quad m_t = \left(\frac{r_t^k}{\alpha}\right)^\alpha \left(\frac{w_t}{1-\alpha}\right)^{1-\alpha},$$

where factor prices equal their respective marginal revenue products.

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 :

$$(17) \quad \Theta_t\left(\frac{\dot{p}_t}{p_t}\right) = \frac{\theta}{2}\left(\frac{\dot{p}_t}{p_t}\right)^2 Y_t,$$

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

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

where

$$(18) \quad \tilde{\Pi}_t(p_t) = \left(\frac{p_t}{P_t} - m_t\right) \left(\frac{p_t}{P_t}\right)^{-\varepsilon} Y_t$$

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

Lemma 1 (proof available in online Appendix B.2) characterizes the solution to the pricing problem and derives the exact New Keynesian Phillips curve in our environment. The combination of a continuous-time formulation of the problem and quadratic price adjustment costs yields a simple equation characterizing the evolution of inflation without the need for log-linearization.

LEMMA 1: *The aggregate inflation rate $\pi_t = \dot{P}_t/P_t$ is determined by the New Keynesian Phillips curve*

$$(19) \quad \left(r_t^a - \frac{\dot{Y}_t}{Y_t}\right) \pi_t = \frac{\varepsilon}{\theta}(m_t - m^*) + \dot{\pi}_t, \quad m^* = \frac{\varepsilon - 1}{\varepsilon}.$$

The expression in (19) can be usefully written in present-value form as

$$(20) \quad \pi_t = \frac{\varepsilon}{\theta} \int_t^\infty e^{-\int_t^s r_\tau^a d\tau} \frac{Y_s}{Y_t} (m_s - m^*) ds.$$

Note that the marginal payoff to a firm from increasing its price at time s is $\tilde{\Pi}'_s(p_s) = \varepsilon Y_s(m_s - m^*)$. Firms raise prices when their markup $1/m_s$ is below the flexible price optimum $1/m^* = \frac{\varepsilon}{\varepsilon - 1}$. Inflation in (20) 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 (17).

Composition of Illiquid Wealth.—Illiquid savings can be invested in two assets: (i) capital k_t , and (ii) equity shares of the aggregate portfolio of intermediate firms, which we denote by s_t . This equity represents a claim on 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 denote the share price. An individual's illiquid assets can thus be expressed as $a_t = k_t + q_t s_t$. The dynamics of capital and equity then satisfy

$$(21) \quad \dot{k}_t + q_t \dot{s}_t = (r_t^k - \delta) k_t + \Pi_t s_t + d_t.$$

We assume that within the illiquid account, resources can be costlessly shifted between capital and shares. Hence, a no-arbitrage condition must hold for the two assets, implying that the return on equity equals the return on capital,

$$(22) \quad \frac{\Pi_t + \dot{q}_t}{q_t} = r_t^k - \delta =: r_t^a.$$

We can therefore reduce the dimensionality of the illiquid asset space and consider only the combined illiquid asset a with rate of return r^a given by any of the two returns in (22) and with law of motion as in (12).¹⁹ Finally, note that (22) implies that $q_t = \int_t^\infty e^{-\int_t^\tau r_s^a ds} \Pi_\tau d\tau$ 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 the Phillips curve.

Monetary Authority.—The monetary authority sets the nominal interest rate on liquid assets i_t according to a Taylor rule,

$$(23) \quad i_t = \bar{r}^b + \phi \pi_t + \epsilon_t,$$

where $\phi > 1$ and $\epsilon_t = 0$ in steady state. Our main experiment studies the economy's adjustment after an unexpected temporary monetary shock ϵ_t .²⁰

Given inflation and the nominal interest rate, the real return on the liquid asset is determined by the Fisher equation $r_t^b = i_t - \pi_t$. The real liquid return r_t^b needs also to be consistent with equilibrium in the bond market, which we describe in Section IIB.

Government.—The government faces exogenous government expenditures G_t and administers a progressive tax and transfer scheme on household labor income $w_t z \ell_t$ that consists of a lump-sum transfer T_t and a proportional tax rate τ_t , with $\tau_t, T_t > 0$. The government is the sole issuer of liquid assets in the economy, which are real

¹⁹The no-arbitrage condition that allows us to reduce the illiquid portfolio to a single state variable, holds only in the absence of jumps in share price q , for example in steady state. In this case, each individual's illiquid asset portfolio composition between capital and equity is indeterminate, even though the aggregate composition is determined.

²⁰We assume that the monetary authority responds only to inflation. Generalizing the Taylor rule (23) to also respond to output gaps is straightforward and does not substantially affect our conclusions. Since our focus is on understanding the transmission mechanism of conventional monetary policy in normal times, we do not consider cases in which the zero-lower bound on nominal interest rates becomes binding.

bonds of infinitesimal maturity B_t^g , with negative values denoting government debt. Its intertemporal budget constraint is

$$(24) \quad \dot{B}_t^g + G_t + T_t = \tau_t \int w_t z \ell_t(a, b, z) d\mu_t + r_t^b B_t^g.$$

Outside of steady state, the fiscal instrument that adjusts to balance the budget can be either τ_t , T_t , or G_t . In our experiments, we consider various alternatives.

B. Equilibrium

An equilibrium in this economy is defined as paths for individual household and firm decisions $\{a_t, b_t, c_t, d_t, \ell_t, n_t, k_t\}_{t \geq 0}$, 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 share price $\{q_t\}_{t \geq 0}$, the inflation rate $\{\pi_t\}_{t \geq 0}$, fiscal variables $\{\tau_t, T_t, G_t, B_t^g\}_{t \geq 0}$, measures $\{\mu_t\}_{t \geq 0}$, 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 (iv) all markets clear. There are five markets in our economy: the liquid asset (bond) market, markets for capital and shares of the intermediate firms (that can be folded into a single illiquid asset), the labor market, and the goods market.

The liquid asset market clears when

$$(25) \quad B_t^h + B_t^g = 0,$$

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. The illiquid asset market clears when physical capital K_t plus the equity value of monopolistic producers q_t (with the total number of shares normalized to 1) equals households' holdings of illiquid assets $A_t = \int a d\mu_t$,

$$(26) \quad K_t + q_t = A_t.$$

The labor market clears when

$$(27) \quad N_t = \int z \ell_t(a, b, z) d\mu_t.$$

Finally, the goods market clearing condition is

$$(28) \quad Y_t = C_t + I_t + G_t + \Theta_t + \chi_t + \kappa \int \max\{-b, 0\} d\mu_t.$$

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 are total price adjustment costs, and the last two terms reflect transaction and borrowing costs (to be interpreted as financial services).

C. Monetary Transmission in HANK

We are interested in analyzing the response of the economy to a one-time unexpected expansionary monetary shock. We assume that the economy is initially in steady state with monetary policy following the Taylor rule (23) with $\epsilon_t = 0$. At time $t = 0$ there is an innovation to the Taylor rule $\epsilon_0 < 0$ with some deterministic decay back to zero. To examine the economy's response to this shock, we generalize the methodology proposed in Section I to decompose the total effect of a monetary shock into direct/partial equilibrium and indirect/general equilibrium effects. Our focus is on the transmission mechanism of the shock on the dynamics of aggregate consumption at impact, but it is clear that our decomposition can be extended to any other aggregate variable at any horizon.

Let us begin by writing aggregate consumption C_t explicitly as a function of the sequence of equilibrium prices, taxes, and transfers $\{\Gamma_t\}_{t \geq 0}$, with $\Gamma_t = \{r_t^b, r_t^a, w_t, \tau_t, T_t\}$, induced by the path of the monetary shock $\{\epsilon_t\}_{t \geq 0}$ from its initial innovation until its full reversal back to zero:

$$(29) \quad C_t(\{\Gamma_t\}_{t \geq 0}) = \int c_t(a, b, z; \{\Gamma_t\}_{t \geq 0}) d\mu_t.$$

Here $c_t(a, b, z; \{\Gamma_t\}_{t \geq 0})$ is the household consumption policy function and $\mu_t(da, db, dz; \{\Gamma_t\}_{t \geq 0})$ is the joint distribution of liquid and illiquid assets and idiosyncratic income.²¹

Totally differentiating (29), we decompose the consumption response at $t = 0$ as

$$(30) \quad dC_0 = \underbrace{\int_0^\infty \frac{\partial C_0}{\partial r_t^b} dr_t^b dt}_{\text{direct effect}} + \underbrace{\int_0^\infty \left(\frac{\partial C_0}{\partial w_t} dw_t + \frac{\partial C_0}{\partial r_t^a} dr_t^a + \frac{\partial C_0}{\partial \tau_t} d\tau_t + \frac{\partial C_0}{\partial T_t} dT_t \right) dt}_{\text{indirect effects}}.$$

The first term in the decomposition reflects direct effects of a change in the path of the liquid return, holding the wage, the illiquid return, and fiscal policy constant.²² Since the path of liquid rates enters the budget constraint (11), households respond directly to interest rate changes. This direct effect itself consists of both intertemporal substitution and income effects. The latter effects arise for two reasons: (i) aggregate liquid assets are in positive net supply; and (ii) liquid asset positions are unequal in the cross section and covary with MPCs.

The remaining terms in the decomposition reflect the indirect effects of changes in wages, the illiquid return, and the government budget constraint that arise in general equilibrium. There are three separate indirect channels at work in response to

²¹ Strictly speaking, because households are forward-looking, the consumption policy function at time t is only a function of the sequence of prices from time t onward $\{\Gamma_s\}_{s \geq t}$. Similarly, the distribution is backward-looking and is only a function of the sequence of prices up to time t , $\{\Gamma_s\}_{s < t}$. We chose the somewhat less precise notation above for simplicity.

²² We define the direct effect of a monetary policy with respect to changes in r_t^b because this is the relevant price from the point of view of households. Alternatively, we could define it "even more directly" with respect to the monetary policy shock ϵ_t . With this alternative decomposition, the direct effect in (30) would be split further into a direct effect due to ϵ_t and an indirect effect due to inflation π_t . This follows because $r_t^b = \bar{r}^b + (\phi - 1)\pi_t + \epsilon_t$ from the Taylor rule and the Fisher equation. Panel A of Figure 3 in our quantitative analysis shows that the drops in r_t^b and ϵ_t are almost equal so that the two decompositions are quantitatively similar.

an expansionary monetary policy shock. First, when the liquid return falls, intertemporal substitution causes non-hand-to-mouth households to increase consumption. In order to meet this additional demand for goods, intermediate firms increase their demand for labor, which pushes up wages. Households respond to the increase in labor income by further increasing their consumption expenditures.

Second, when the illiquid return changes in response to the change in the liquid return, consumption may be further affected as households choose to rebalance their asset portfolio with deposits into or withdrawals from the illiquid account.²³

Third, there is a fiscal response to changes in the liquid rate through the government budget constraint. A fall in r^b reduces government’s interest payments on its debt and results in higher tax revenues because of the additional labor income from the economic expansion. Both forces loosen the government budget constraint, and lead to an adjustment in one of the fiscal instruments. As will become clear from our numerical experiments, both the total size of the macroeconomic effects of monetary policy and the split between direct and indirect components depends on the type of fiscal response, a consequence of the non-Ricardian nature of this class of HANK economies.

In practice, we need to compute each of these components numerically. For example, the formal definition of the first term in (30), the direct effect of changes in the liquid return $\{r_t^b\}_{t \geq 0}$, is

$$(31) \quad \int_0^\infty \frac{\partial C_0}{\partial r_t^b} dr_t^b dt = \int_0^\infty \left(\int \frac{\partial c_0(a, b, z; \{r_t^b, \bar{r}^a, \bar{w}, \bar{\tau}, \bar{T}\}_{t \geq 0})}{\partial r_t^b} d\mu_0^{r^b} \right) dr_t^b dt,$$

where $\mu_0^{r^b} = \mu_0(da, db, dz; \{r_t^b, \bar{r}^a, \bar{w}, \bar{\tau}, \bar{T}\}_{t \geq 0})$. That is, this term is the aggregate *partial-equilibrium* consumption response of a continuum of households that face a time-varying interest rate path $\{r_t^b\}_{t \geq 0}$ but paths for illiquid asset return \bar{r}^a , wage \bar{w} , and taxes and transfers $(\bar{\tau}, \bar{T})$ that are held constant at their steady-state values. We calculate this term from the model by feeding these time paths into the households’ optimization problem, computing c_0 for each household, and aggregating across households using the corresponding distribution.

The other terms in the decomposition are computed in a similar fashion. We discuss the similarities between our decomposition and the one proposed by Auclert (2016) in Section IVC.

III. Taking the Model to the Data

In this section, we explain how we map our framework to the data. We begin by presenting a small extension of HANK that allows the model to generate more volatile and procyclical investment in response to a monetary shock. Next, we describe

²³ At impact, the share price q_0 jumps, reflecting the change in expected future profits and induces a revaluation of illiquid wealth. With a slight abuse of notation, the derivative of C_0 with respect to r_0^b in (30) embeds this effect (and all the results we report later in the paper take this initial instantaneous jump into account). For $t > 0$, changes in q_t are already embedded in r_t^a through the no-arbitrage condition (22). In order to quantify the effect of this price movement on the value of households’ illiquid assets, we need to make an assumption about the portfolio composition between shares and capital. We simply assume that every agent has the same portfolio composition as the aggregate.

our calibration strategy and illustrate our parameterization. Finally, we demonstrate that the model offers a realistic representation of microeconomic consumption behavior.

A. Distribution of Monopoly Profits

In models of monopolistic competition with price rigidities only, countercyclical markups are at the heart of economic fluctuations. Our framework is no exception. Because prices are sticky but nominal marginal costs are not, expansionary monetary shocks shrink markups, causing firm profits to fall.²⁴ In RANK models these fluctuations in profits are typically borne lump sum by the representative household. But in HANK models, additional assumptions are needed about how profits are distributed across households. In two-asset HANK models, yet further assumptions are needed about how profits are distributed between liquid and illiquid assets. These assumptions can have a large effect on the volatility and cyclicity of investment.²⁵

For example, in our baseline HANK model of Section IIA, the fluctuations in profits manifest as movements in the share price q_t . Since equity is a component of *illiquid* assets, rather than *liquid* assets, the fall in profits associated with an expansionary monetary shock creates a downward pull on investment at a time when output is expanding.²⁶ This feature is in stark contrast with the data where, quantitatively, investment is the most volatile and procyclical component of output.

To avoid this counterfactual implication of sticky prices and restore an empirically realistic comovement between output, consumption, and investment, we make a simple modification to the baseline HANK model: we add one parameter $\omega \in [0, 1]$ that controls the fraction of profits reinvested directly into the illiquid account. Then, if we aggregate the total illiquid income flow across all households, we obtain

$$(32) \quad (r_t^k - \delta)K_t + \omega\Pi_t = \alpha m_t Y_t + \omega(1 - m_t)Y_t.$$

The profit distribution scheme that fully sterilizes the impact of fluctuating markups corresponds to $\omega = \alpha$, the capital share of output. In this case, the total income flow accruing into the illiquid account becomes αY_t , which is independent of m_t and is always procyclical.

We assume that the residual share of profits $(1 - \omega)\Pi_t$ is paid in liquid form to every individual i as a lump-sum transfer in proportion to household productivity, i.e., $\pi_{it}^b = \frac{z_{it}}{\bar{z}}(1 - \omega)\Pi_t$, where \bar{z} is average productivity. We interpret this additional income as the profit-sharing component of worker compensation from

²⁴Since output increases, it is theoretically possible that profits increase in response to an expansionary shock. However, for plausible parameterizations it is typically the case that markups respond more than output and profits fall.

²⁵In one-asset RANK and TANK models, assumptions about the distribution of profits between liquid and illiquid assets are, of course, moot. In two-asset RANK and TANK models these assumptions also matter. See online Appendix A.5.

²⁶Aggregating (21) across households and using the relevant market clearing conditions we have $\dot{K}_t = (r_t^k - \delta)K_t + \Pi_t + D_t$ or equivalently that aggregate investment is $I_t = r_t^k K_t + \Pi_t + D_t$ where Π_t is the profit flow and D_t are aggregate net deposits. When net deposits D_t do not move much over the business cycle as is the case in practice, countercyclical fluctuations in profits Π_t have a large effect on investment I_t .

bonuses, commissions, and gains from exercising stock options, and in what follows the term labor income should be interpreted as the sum of wage payments $w_t z_{it} \ell_{it}$ and bonuses π_{it}^b , whenever $\omega < 1$. Online Appendix B.4 contains more details of this extension of the model.²⁷

B. Calibration Strategy

We have four broad goals in choosing parameters for the model. First, we need to develop a mapping between our aggregated two-asset (liquid-illiquid) structure and data on the complex balance sheet of the US household sector. Second, we seek a calibration of the exogenous stochastic process for labor earnings, which is the ultimate source of inequality in the model. Third, in order to obtain quantitatively realistic consumption behavior at the microeconomic level, our model must generate realistic distributions of liquid and illiquid assets. Of particular importance is the skewness of liquid wealth holdings: matching the fraction of households with low liquid wealth bears directly on the sensitivity of consumption to income changes, whereas matching the top of the liquid wealth distribution is key to generate plausible redistributive effects of interest rate changes. Finally, for the production and monetary sides of the model, we stay as close as possible to the parameterization that is well accepted in the New Keynesian literature.

Categorization of Assets into Liquid and Illiquid.—Mapping the model to data requires classifying assets held by US households as liquid versus illiquid. We label an asset as liquid or illiquid based on the extent to which buying or selling the asset involves transaction costs. We define net liquid assets B^h as all deposits in financial institutions (checking, saving, call, and money market accounts), government bonds, and corporate bonds net of revolving consumer credit. We define illiquid assets A as real estate wealth net of mortgage debt, consumer durables net of non-revolving consumer credit, plus equity in the corporate and non-corporate business sectors. We have chosen to include equity among illiquid assets because nearly three-quarters of total equity is either indirectly held (in tax-deferred retirement accounts) or held in the form of private businesses. Both of these assets are significantly less liquid than all the other asset classes included in our definition of B^h .²⁸

We measure the aggregate size of each category of assets and liabilities using data from the Flow of Funds (FoF) and the Survey of Consumer Finances (SCF). We use data from 2004, since this is the last SCF survey year before the Great Recession. In online Appendix C, we undertake a comprehensive comparison between these two data sources for each component of the balance sheet. Based on this analysis, we choose to use FoF measures for all assets and liabilities except

²⁷The importance of countercyclical profits has recently been highlighted by Broer et al. (2016), who argue that the resulting income effects on labor supply greatly amplify the New Keynesian monetary transmission mechanism in RANK. As we discuss in the next section, the parameter ω also allows us to discipline the strength of such income effects which, under standard balanced-growth preferences, in our model are present whenever $\omega < 1$.

²⁸In a former version of the paper (Kaplan, Moll, and Violante 2016a), we also separated illiquid assets between productive (equity) and nonproductive (housing and durables). In that version of the model, the productive assets served as input into production and paid the rate of return r^a , whereas the nonproductive ones yielded a utility flow to households. Our key quantitative findings are largely invariant to adopting this richer classification, once the model is properly recalibrated.

TABLE 2—SUMMARY OF TAXONOMY OF ASSETS

Liquid (B^h)		Illiquid ($A = K + q$)	Total
Revolving consumer debt	-0.03	Net housing	1.09
Deposits	0.23	Net durables	0.22
Corporate bonds	0.04	Corporate equity	1.02
Government bonds	0.02	Private equity	0.59
Total	0.26		2.92 3.18

Notes: Categorization of assets into liquid versus illiquid. Values are expressed as a multiple of 2004 GDP (\$12,300B). See online Appendix C for details of all calculations.

for the three main categories of liquid assets—deposits, government bonds, and corporate bonds—for which we use estimates from the SCF. Table 2 summarizes our preferred estimates, expressed as fractions of annual 2004 GDP (\$12,300B). The total quantity of net liquid assets B^h amounts to \$2,700B (26 percent of annual GDP). The total quantity of net illiquid assets A amounts to \$36,000B (2.92 times annual GDP).

Continuous-Time Earnings Dynamics.—When households have a choice between saving in assets with different degrees of liquidity, as in our model, the frequency of earnings shocks is a crucial input for determining the relative holdings of the two assets. Households who face small, but frequent, shocks have a strong incentive to hold low-return liquid assets to smooth consumption, while households who face large infrequent shocks would prefer to hold high-return illiquid assets that can be accessed at a cost in the unlikely event of a sizable windfall or a severe income loss.

In standard discrete-time error component models (e.g., the classic persistent-transitory model), the frequency of arrival of earnings shocks is dictated by the assumed time period. In continuous-time models, the frequency at which shocks arrive is a property of the stochastic process, and must be estimated alongside the size and persistence of shocks. Empirically, the challenge in estimating the frequency of earnings shocks is that almost all high-quality panel earnings data are available only at an annual (or lower) frequency. It is thus challenging to learn about the dynamics of earnings at any higher frequency. Our strategy to overcome this challenge is to infer high frequency earnings dynamics from the *high-order moments of annual earnings changes*. To understand why this identification strategy has promise, consider two possible distributions of annual earnings changes, each with the same mean and variance, but with different degrees of kurtosis. The more leptokurtic distribution (i.e., the distribution with more mass concentrated around the mean and in the tails) is likely to have been generated by an earnings process that is dominated by large infrequent shocks; the more platykurtic distribution (i.e., the distribution with more mass in the shoulders) by a process that is dominated by small frequent shocks.

Motivated by these observations, we model log-earnings as the sum of two independent components:

$$(33) \quad \log z_{it} = z_{1,it} + z_{2,it}$$

where each component $z_{j,it}$ evolves according to a “jump-drift” process. Jumps arrive at a Poisson rate λ_j . Conditional on a jump, a new log-earnings state $z'_{j,it}$ is

drawn from a normal distribution with mean zero and variance σ_j^2 , $z'_{j,it} \sim N(0, \sigma_j^2)$. Between jumps, the process drifts toward zero at rate β_j . Formally, the process for $z_{j,it}$ is

$$(34) \quad dz_{j,it} = -\beta_j z_{j,it} dt + dJ_{j,it}$$

where $dJ_{j,it}$ captures jumps in the process.²⁹

The process for each component is closely related to a discrete-time AR(1) process.³⁰ The key difference is that in our continuous-time formulation, the arrival of each innovation is stochastic, and hence each process has an additional parameter, λ_j , which captures the frequency of arrival.³¹

Estimation with Male Earnings Data.—We estimate the earnings process in (33)–(34) by Simulated Method of Moments using Social Security Administration (SSA) data on male earnings from Guvenen et al. (2015).³² These authors report eight key moments that we target in the estimation (see Table 3).³³ Moments of the distribution of earnings changes at multiple durations are needed to separately identify the two components. Since these data refer to annual earnings, we simulate earnings from the model at a high frequency, aggregate to annual earnings, and compare moments from model and data.

The fitted earnings process matches the eight targeted moments well. The estimated parameter values, reported in Table 4, are consistent with the existence of a transitory and a persistent component in earnings. The transitory component ($j = 1$) arrives on average once every 3 years and has a half-life of around one quarter. The persistent component ($j = 2$) arrives on average once every 38 years and has a half-life of around 18 years. Both components are subject to relatively large, similarly sized innovations. In the context of an infinite horizon model, the estimated process thus has the natural interpretation of a large and persistent “career” shock that is perturbed by periodic temporary shocks. Note that relative to a discrete-time model, our estimated transitory shock is both less frequent, and more temporary than an i.i.d. annual shock.

²⁹Even more formally, the infinitesimal generators $\mathcal{A}_j f(z) := \lim_{t \downarrow 0} \frac{E[f(z_t)] - f(z)}{t}$ of the two components $j = 1, 2$ are given by $\mathcal{A}_j f(z) = -\beta_j z f'(z) + \lambda_j \int_{-\infty}^{\infty} (f(x) - f(z)) \phi_j(x) dx$ where ϕ_j is the density of a normal distribution with mean zero and variance σ_j^2 .

³⁰In particular, if the earnings innovations always arrived at regular intervals (say, annually), rather than stochastically at rate λ_j , then each component would follow an AR(1) process. The drift parameter β_j would correspond to (one minus) the discrete-time autoregressive parameter and the innovation variance σ_j^2 would describe the size of innovations. In this sense, the model is only a minimal departure from the familiar persistent-transitory process used to model discrete-time earnings data.

³¹Schmidt (2015) models earnings dynamics as a discrete-time compound Poisson process, using a similar logic.

³²The main benefits of targeting moments from administrative earnings data such as the SSA are that they are based on a very large sample and so are less prone to measurement error than survey data, and that they are not top-coded. Both features are important: the sample size and absence of measurement error allows a precise estimate of higher-order moments, and the absence of top-coding allows for an accurate portrayal of the right-tail of the income distribution, which is important for capturing the skewness in wealth holdings.

³³We restrict attention to a symmetric process since Guvenen et al. (2015) find only a small amount of negative skewness in 1-year and 5-year annual changes. It is possible to generate skewness in annual changes by allowing the drift parameters β_j to differ based on the sign of $z_{j,it}$.

TABLE 3—EARNINGS PROCESS ESTIMATION FIT

Moment	Data	Model
Variance: annual log earns	0.70	0.70
Variance: 1-year change	0.23	0.23
Variance: 5-year change	0.46	0.46
Kurtosis: 1-year change	17.8	16.5
Kurtosis: 5-year change	11.6	12.1
Frac. 1-year change < 10%	0.54	0.56
Frac. 1-year change < 20%	0.71	0.67
Frac. 1-year change < 50%	0.86	0.85

TABLE 4—EARNINGS PROCESS PARAMETER ESTIMATES

Parameter	Component	
	$j = 1$	$j = 2$
Arrival rate λ_j	0.080	0.007
Mean reversion β_j	0.761	0.009
Standard deviation of innovations σ_j	1.74	1.53

Note: Rates expressed as quarterly values.

In our model, flow earnings are given by $y_{it} := w_t z_{it} \ell_{it}$ and are thus determined by both the realization of productivity shocks z_{it} and the choice of labor supply ℓ_{it} . In online Appendix D.1 we explain how we convert the estimated process for individual male earnings to a discrete-state process for idiosyncratic productivity that is consistent with our assumption of a household as the unit of observation.³⁴ Relative to typical earnings process calibrations based on survey data, and consistent with the cross-sectional earnings distribution in SSA data, the resulting earnings process features a large amount of right-tail inequality. The top 10, 1, and 0.1 percent shares of gross household labor earnings in the steady state are 32 percent, 7 percent, and 2 percent respectively. This skewed earnings distribution is an important factor in the model's ability to generate skewed distributions of liquid and illiquid assets. However, unlike much of the existing literature that has generated wealth concentration at the top of the distribution from ad hoc skewed earnings distributions, here both inequality and dynamics of earnings are disciplined directly by high-quality data.³⁵

Adjustment Cost Function and Wealth Distribution.—We set the steady-state real return on liquid assets \bar{r}^b at 2 percent per annum and steady-state inflation to zero. Given values for the capital share, demand elasticity, and depreciation rate (all set externally as described below) and for the unsecured borrowing limit, our target for

³⁴In online Appendix D.1 we describe the discretization process in detail and report further statistics from the discretized distribution, including plots of the Lorenz curves for the ergodic distributions from the continuous and discretized processes.

³⁵The existing literature reverse-engineers a process for earnings risk in order to match data on wealth inequality. This approach typically requires an implausibly extreme characterization of risk, with a top income state around 500 times as large as the median, and a high probability of a dramatic fall in earnings once the top state is reached. See Benhabib and Bisin (2016) and De Nardi, Fella, and Pardo (2016) for a discussion of this issue. In our discretized process, instead, the highest productivity realization is around 13 times as large as the median, and is realized by only 0.03 percent of the population.

the illiquid assets of 2.9 times output yields a steady-state return to illiquid assets r^a of 5.7 percent per annum.

Given these returns, and the exogenous process for idiosyncratic labor income, the key parameters that determine the incentives for households to accumulate liquid and illiquid assets are the borrowing limit \underline{b} , the discount rate ρ , the intermediation wedge κ , and the three parameters of the adjustment cost function χ_0 , χ_1 , and χ_2 .

Borrowing in the model should be interpreted as unsecured credit, so we set the borrowing limit \underline{b} exogenously at 1 times quarterly average labor income.³⁶ We then choose the remaining five parameters (ρ , κ , χ_0 , χ_1 , χ_2) to match five moments of the distribution of household wealth: (i)–(ii) the mean of the illiquid and liquid wealth distributions from Table 2; (iii)–(iv) the fraction of poor and wealthy hand-to-mouth households from Kaplan, Violante, and Weidner (2014), since these are the most important moments of the liquid wealth distribution for determining household consumption responses to income shocks; and (v) the fraction of households with negative net liquid assets, which serves to identify the borrowing wedge, also from Kaplan, Violante, and Weidner (2014).³⁷

The calibrated annual discount rate ρ is 5.1 percent, and the annual wedge κ is 6.0 percent (implying an annual borrowing rate of 8.0 percent). The calibrated transaction cost function is displayed in Figure D.3 in online Appendix D. In the resulting ergodic distribution, roughly 80 percent 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 1.7 percent. The quarterly transaction cost for a transaction this size is 23 percent of the transaction. In steady state, the equilibrium aggregate transaction costs, which one can interpret as financial services, amount to less than 4 percent of GDP.

The model replicates the five targeted moments well (left panel of Table 5).³⁸ Figure 1 displays the distributions of liquid and illiquid wealth in the model. Despite only targeting a handful of moments of each distribution, the model successfully matches the distributions of liquid and illiquid wealth up to the very top percentiles, as is clear from the right panel of Table 5, which reports top wealth shares from the model and data. Both Gini coefficients in the model are close to their data counterparts. The reason for this success is a combination of the realistically skewed earnings distribution and the heterogeneity in effective returns on wealth because of the two-asset structure: a fraction of households ends up spending a long time in high earnings states, hold high-return illiquid assets, and accumulate a lot of wealth. These households populate the upper tail of the wealth distribution.³⁹

³⁶In the steady-state ergodic distribution only 0.02 percent of households are at the limit.

³⁷The targets of 10 percent and 20 percent correspond to the fraction of households with net liquid wealth $\in [-\$1,000, \$1,000]$ with zero and positive illiquid assets, respectively. The target of 15 percent of households with negative liquid wealth corresponds to the fraction of households with net liquid wealth less than $-\$1,000$.

³⁸Besides matching the share of hand-to-mouth agents in the population, the model also does well with respect to their relative importance in terms of consumption share (20 percent versus 25 percent in the Panel Study of Income Dynamics (PSID)) and wealth share (2.5 percent versus 4.4 percent in the SCF).

³⁹To put this result in the context of the literature, see the survey by Benhabib and Bisin (2016) on the sources of skewed wealth distributions in macroeconomic models. Note that our model is not able to match the extreme right tail of the liquid wealth distribution and also does not feature a Pareto tail for the distribution of total wealth as arguably observed in the data. It is notoriously challenging to match the extreme right tail of wealth distributions with labor income risk alone and our model is no exception.

TABLE 5

	Data	Model	Moment	Liquid wealth		Illiquid wealth	
				Data	Model	Data	Model
Mean illiquid assets	2.92	2.92	Top 0.1 percent share	17	2.3	12	7
Mean liquid assets	0.26	0.23	Top 1 percent share	47	18	33	40
Frac. with $b = 0$ and $a = 0$	0.10	0.10	Top 10 percent share	86	75	70	88
Frac. with $b = 0$ and $a > 0$	0.20	0.19	Bottom 50 percent share	-4	-3	3	0.1
Frac. with $b < 0$	0.15	0.15	Bottom 25 percent share	-5	-3	0	0
			Gini coefficient	0.98	0.86	0.81	0.82

Notes: Left panel: moments targeted in calibration and reproduced by the model. Means are expressed as ratios to annual output. Right panel: statistics for the top and bottom of the wealth distribution not targeted in the calibration.

Source: SCF 2004

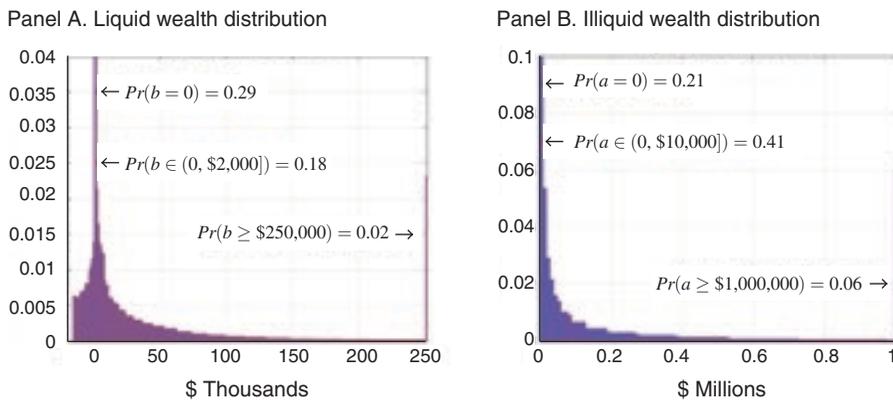


FIGURE 1. DISTRIBUTIONS OF LIQUID AND ILLIQUID WEALTH

Remaining Model Parameters.—

Demographics: We set the quarterly death rate ζ to $1/180$ so that the average lifespan of a household is 45 years.

Preferences: Households have instantaneous utility that is separable over consumption and hours worked:

$$(35) \quad u(c, \ell) = \frac{c^{1-\gamma}}{1-\gamma} - \varphi \frac{\ell^{1+\nu}}{1+\nu},$$

with $\gamma \geq 0$ and $\nu \geq 0$. We set $1/\gamma$, the intertemporal elasticity of substitution (IES), to 1 and $1/\nu$, the Frisch elasticity of labor supply at the household level, to 1.⁴⁰ The weight on the labor supply component of utility, φ , is set so that average hours worked are equal to one half in steady state.

⁴⁰This number is slightly higher than the 0.82 identified by Chetty et al. (2011) as the representative estimate from existing studies of the micro elasticity at the individual level, accounting for intensive and extensive margins of adjustment. At the household level though, the marginally attached worker is often the wife and a Frisch labor supply elasticity of one is in line with the estimates of Blundell, Pistaferri, and Saporta-Eksten (2016) for married women.

Production: The elasticity of substitution for final goods producers is set to $\varepsilon = 10$, implying a steady-state markup $1/(\varepsilon - 1)$ of 11 percent. Intermediate goods producers have a weight on capital of $\alpha = 0.33$, which yields a capital share of 29 percent and a labor share of 60 percent. We set the constant θ in the price adjustment cost function to 100, so that the slope of the Phillips curve in (19) is $\varepsilon/\theta = 0.1$.⁴¹ The fraction ω of aggregate profits reinvested into the illiquid accounts is set to α , as explained in Section IIIA. This choice which, as explained, neutralizes the countercyclicality of mark-ups, happens to be also roughly in line with the data. In 2004, the sum of undistributed corporate profits (the empirical counterpart of profits reinvested in the illiquid account in the model) and dividend income (the counterpart of profits paid to households) was \$946B. Of these, undistributed profits amounted to \$384B, thus about 40 percent of the total.⁴²

Government Policy: We set the proportional labor income tax rate τ to 0.30 and the lump-sum transfer T to be 6 percent of output (equivalent to around \$7,000 per year). In steady state just over 9 percent of households receive a net transfer from the government. In our model, the government is the only provider of liquid assets. Given our calibration of household liquid holdings, government debt amounts to 23.3 percent of annual GDP. Expenditures are then determined residually from the government budget constraint (24).

Monetary Policy: We set the Taylor rule coefficient ϕ to 1.25 which is in the middle of the range commonly used for New Keynesian models.

Table 6 summarizes our parameter values. In Section IVB, we verify the robustness of our results with a series of sensitivity analyses.

C. Micro Consumption Behavior

How successful is the calibrated model at generating empirically realistic distributions of household responses to changes in labor income? Some of the most convincing empirical evidence on marginal propensities to consume (MPCs) comes from household consumption responses to the tax rebates of 2001 and fiscal stimulus payments of 2008 (see, e.g., Johnson, Parker, and Souleles 2006; Parker et al. 2013; Misra and Surico 2014; Broda and Parker 2014). While the estimates are often imprecise because of the small sample size, this collective quasi-experimental evidence concludes that households spend approximately 15–25 percent of these payments (which average between \$500 and \$1,000 depending on the episode) on nondurables in the quarter that they are received.

Let $MPC_{\tau}^x(a, b, z)$ be the MPC over a period of length τ quarters out of a one-time inflow of x additional dollars of liquid wealth. This is the notion of an MPC that is comparable to the empirical evidence cited above (as opposed to the slope of the consumption function with respect to liquid wealth). In online Appendix B.3 we

⁴¹ See Schorfheide (2008) who surveys many studies using the labor share as a proxy to measure marginal costs, an approach suggested by Galí and Gertler (1999).

⁴² See NIPA Tables 2.1 and 5.1. Also over the period 1990–2016, for example, this fraction fluctuates around 40 percent.

TABLE 6—LIST OF CALIBRATED PARAMETER VALUES

Description	Value	Target/source
<i>Preferences</i>		
ζ Death rate	1/180	Avg. lifespan 45 years
$1/\gamma$ Intertemporal elasticity of subst.	1	
$1/\nu$ Frisch elasticity of labor supply	1	
φ Disutility of labor	2.2	Avg. hours worked equal to 1/2
ρ Discount rate (p.a.)	5.1%	Internally calibrated
<i>Production</i>		
ε Demand elasticity	10	Profit share of 10 percent
θ Price adjustment cost	100	Slope of Phillips curve, $\varepsilon/\theta = 0.1$
α Capital share	0.33	
$\bar{\delta}$ Steady-state depreciation rate (p.a.)	7%	
<i>Government</i>		
τ Proportional labor tax	0.30	
T Lump-sum transfer (rel. GDP)	0.06	40% hh with net govt. transfer
<i>Monetary Policy</i>		
ϕ Taylor rule coefficient	1.25	
\bar{r}^b Steady-state real liquid return (p.a.)	2%	
<i>Unsecured borrowing</i>		
r^{borr} Borrowing rate (p.a.)	8%	Internally calibrated
\underline{b} Borrowing limit	\$16,500	$1 \times$ quarterly labor income
<i>Adjustment cost function</i>		
χ_0 Linear component	0.0438	Internally calibrated
χ_1 Convex component	0.956	Internally calibrated
χ_2 Convex component	1.402	Internally calibrated
a Min a in denominator	\$1,000	

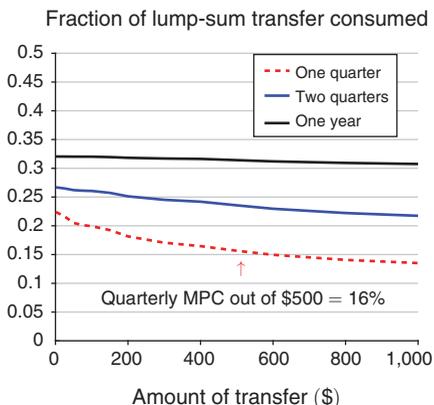
state the formal definition and explain how to compute it directly from households' consumption policy functions by using the Feynman-Kac formula, as in Achdou et al. (2017).

The average quarterly MPC out of a \$500 transfer is 16 percent in the model, which is within the range of typical empirical estimates. As seen in panel A of Figure 2, the fraction consumed decreases with the size of the transfer, and increases sharply as the horizon increases.

The average MPCs in panel A mask important heterogeneity across the population. This heterogeneity can be seen in panel B, which plots the function $MPC_{\tau}^x(a, b, z)$ for an $x = \$500$ payment over one quarter as a function of liquid and illiquid assets, averaged across labor productivity z . The figure illustrates the strong source of bimodality in the distribution of consumption responses in the population. In the model, the average response of 16 percent is composed of a group of households with positive net liquid wealth and very low consumption responses, and another group of hand-to-mouth households with no liquid wealth who display strong consumption responses. Of these hand-to-mouth households, roughly two-thirds have positive illiquid wealth.

Several recent empirical papers have documented patterns of the distribution of MPCs that are consistent with Figure 2. Broda and Parker (2014) find much stronger consumption responses to the 2008 fiscal stimulus payments among households with low easily accessible liquid funds. Misra and Surico (2014) use quantile regression

Panel A. $\int MPC_{\tau}^x(a, b, z) d\mu$ by τ, x



Panel B. $MPC_1^{\$500}(a, b, z)$

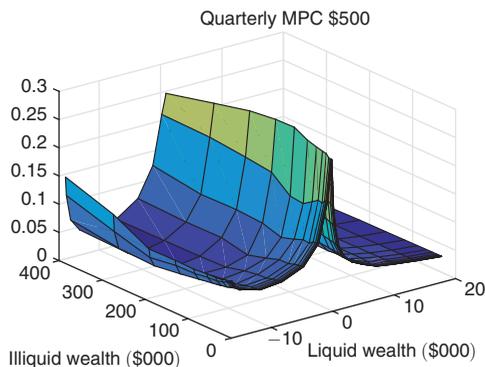


FIGURE 2. MPC HETEROGENEITY

techniques to study the consumption responses in the tax rebate episodes of 2001 and 2008 and document the presence of high-income households both in the low MPC and the high MPC group, a fact consistent with the presence of wealthy hand-to-mouth (HtM) households. Kaplan, Violante, and Weidner (2014) use the method proposed by Blundell, Pistaferri, and Saporta-Eksten (2016) to estimate the consumption response to transitory income shocks on PSID data for three groups—non HtM, poor HtM, and wealthy HtM—and uncover much higher MPCs for both types of HtM households. Baker (2016) finds that households with similar net asset positions behave differently with respect to income shocks if they hold varying shares of liquid and illiquid wealth. Fagereng, Holm, and Natvik (2016) examine MPCs out of lottery prizes using Norwegian administrative data. They find that MPCs vary with the amount of households’ liquid assets and that households with close to zero liquid assets have high MPCs even if they are wealthy in terms of their illiquid asset positions.

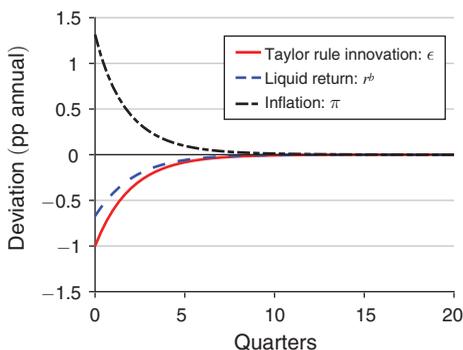
This striking heterogeneity in MPCs underlines the importance of obtaining a realistic distribution of both wealth components. With such distributions in hand, we now turn to the monetary transmission mechanism.

IV. Monetary Transmission: Quantitative Results

Our main results concern the response of the economy to a one-time unexpected monetary shock. We consider an experiment in which at time $t = 0$, there is a quarterly innovation to the Taylor rule (23) of $\epsilon_0 = -0.25$ percent (i.e., -1 percent annually) that mean-reverts at rate η , i.e., $\epsilon_t = e^{-\eta t} \epsilon_0$. We set $\eta = 0.5$, corresponding to a quarterly autocorrelation of $e^{-\eta} = 0.61$, a value consistent with the VAR-based empirical evidence, as argued in Section IA.

In our baseline specification, lump-sum transfers T_t adjust so as to keep the budget balanced, with expenditures and debt fixed at their steady-state level, as we did in Section IB. In Section IVB we provide results under alternative assumptions.

Panel A. Monetary shock, interest rate, inflation



Panel B. Aggregate quantities

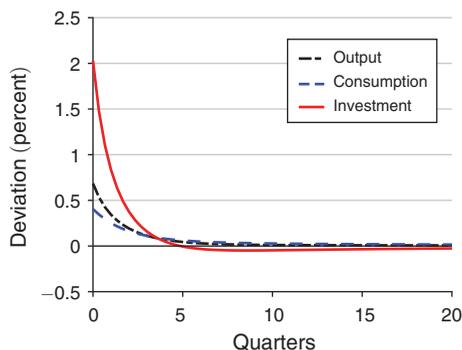


FIGURE 3. IMPULSE RESPONSES TO A MONETARY POLICY SHOCK
(A Surprise, Mean-Reverting Innovation to the Taylor Rule)

A. Impulse Response to a Monetary Shock

Panel A of Figure 3 displays the exogenous time path for the innovation ϵ and the implied changes in the liquid interest rate and rate of inflation. Panel B displays the corresponding impulse responses for aggregate quantities.

In response to an expansionary monetary policy shock, the real return on liquid assets r_t^b falls, which stimulates consumption and investment, and leads to an increase in both output and inflation. The magnitudes of these responses are, at least qualitatively, consistent with empirical evidence from VARs: consumption increases by less than output and by much less than investment, with an elasticity to the change in r^b over the first year after the shock equal to -2.9 .⁴³

How does this magnitude compare to the corresponding response in the RANK models analyzed in Section IA? Table 1 shows that, across RANK models, the total elasticity is always around -2 . Thus, in our baseline specification of HANK (with lump-sum transfers adjusting) the elasticity is almost 50 percent higher than in RANK. Notably, this discrepancy implies that the “as if” result of Werning (2015) does not hold in our framework.⁴⁴

In the next section, we decompose this total effect of the monetary shock on aggregate consumption into direct and indirect components through the lens of our methodology developed in Section IIC.

B. The Size of Direct and Indirect Effects

The equilibrium time paths for prices and government transfers induced by the monetary shock that we feed into the household problem to compute each element

⁴³ See Figure 1 in Christiano, Eichenbaum, and Evans (2005). Our model cannot generate hump-shaped impulse responses since we abstract from the modeling ingredients in typical medium-scale DSGE models that generate these dynamics, such as external habits and investment-rate adjustment costs.

⁴⁴ Werning (2015) studies deviations from his benchmark incomplete markets economy and argues that, in plausible cases, consumption becomes more sensitive than in RANK to current and future interest rate changes, as we find.

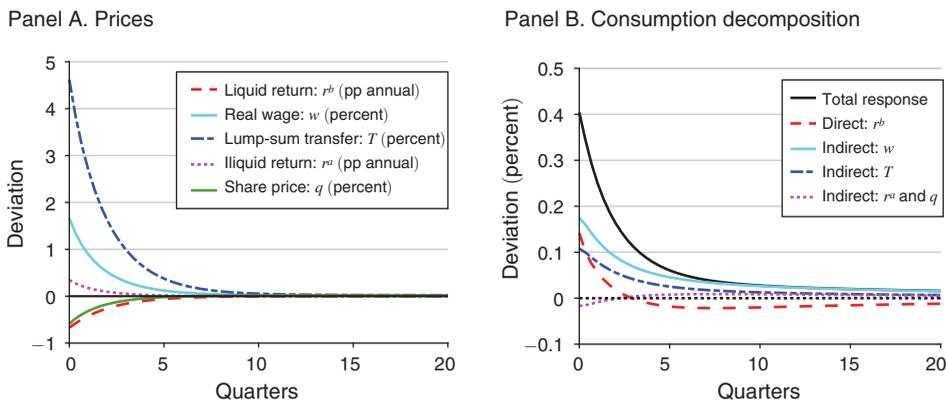


FIGURE 4. DIRECT AND INDIRECT EFFECTS OF MONETARY POLICY IN HANK

Notes: Returns are shown as annual percentage point deviations from steady state. Real wage and lump-sum transfers are shown as log deviations from steady state.

TABLE 7—DECOMPOSITION OF THE EFFECT OF MONETARY SHOCK ON AGGREGATE CONSUMPTION

	Baseline (1)	$\omega = 1$ (2)	$\omega = 0.1$ (3)	$\frac{\varepsilon}{\theta} = 0.2$ (4)	$\phi = 2.0$ (5)	$\frac{1}{\nu} = 0.5$ (6)
Change in r^b (pp)	-0.28	-0.34	-0.16	-0.21	-0.14	-0.25
Elasticity of Y	-3.96	-0.13	-24.9	-4.11	-3.94	-4.30
Elasticity of I	-9.43	7.83	-105	-9.47	-9.72	-9.79
Elasticity of C	-2.93	-2.06	-6.50	-2.96	-3.00	-2.87
Partial eq. elasticity of C	-0.55	-0.45	-0.99	-0.57	-0.59	-0.62
<i>Component of percent change in C due to</i>						
Direct effect: r^b	19	22	15	19	20	22
Indirect effect: w	51	56	51	51	51	38
Indirect effect: T	32	38	19	31	31	45
Indirect effect: r^a and q	-2	-16	15	-2	-2	-4

Notes: Average responses over the first year. Column 1 is the baseline specification. In column 2, profits are all reinvested into the illiquid account. In column 3, 10 percent of profits are reinvested in the illiquid account. In column 4, we reduce the stickiness of prices by lowering the cost of price adjustment θ . In column 5, we increase ϕ , which governs the responsiveness of the monetary policy rule to inflation. In column 6, we lower the Frisch elasticity of labor supply from 1 to 0.5.

of the direct and indirect effects are displayed in panel A of Figure 4, alongside the resulting decomposition in panel B. In the bottom panel of Table 7 we explicitly report the contribution of each component to the overall consumption response over the first year following the shock.⁴⁵

The decomposition reveals our first novel quantitative insight into the monetary transmission mechanism. The combined indirect effects of an unexpected shock are much larger than the direct effect. In the HANK model, the indirect components account for 80 percent of the consumption response while the direct component

⁴⁵In principle, the contribution of the components need not add to 100 percent, since the exact decomposition holds only for infinitesimal changes in prices, as in Proposition 1 for the stylized model of Section I. In practice, though, they almost exactly do.

accounts for only 20 percent of the response. This is in stark contrast to typical RANK models, as argued in Section I. This finding is very robust, as is evident from the remaining columns of Table 7 that report analogous results from alternative model specifications.

In the baseline model we allocate a fraction $\omega = \alpha$ of profits to illiquid equity in order to neutralize the effect of countercyclical profits on investment, as explained in Section IIIA. Columns 2 and 3 of Table 7 show that this assumption does not impact the decomposition, but it is important for generating procyclical investment and a positive output response. When all profits are allocated to equity in the illiquid account ($\omega = 1$), the fall in profits following the monetary shock substantially dampens the response of investment and thus reduces the total consumption response (line 4 in the table) because with lower investment there is a smaller increase in labor demand and wages. When profits are nearly all paid as dividends to households ($\omega = 0.1$), consumption responds very aggressively due to the huge reaction of investments.⁴⁶

Columns 4 and 5 of Table 7 show that the two key parameters that determine the strength of the New Keynesian elements in the model, the Taylor rule coefficient ϕ and the degree of price stickiness θ , do not substantially affect either the overall size of the consumption elasticity or its decomposition between direct and indirect effects. Rather, these elements primarily affect the inflation response for a given monetary shock and, hence, the extent of movements in the real interest rate.⁴⁷

In column 6 of Table 7, we set the Frisch elasticity to 0.5, one-half of its baseline value. The effect of lowering the Frisch elasticity is merely to shift the composition of the indirect effects away from the wage component toward the transfer component.

In a previous version of this paper (Kaplan, Moll, and Violante 2016a), we showed that replacing the separable preference specification with GHH utility (Greenwood, Hercowitz, and Huffman 1988) yields an elasticity of aggregate consumption C to r^b that is almost twice as large as in the baseline.⁴⁸ Moreover, the indirect general equilibrium effects account for over 90 percent of the total effect. The key feature of GHH utility that drives these results is the strong complementarity between hours worked and consumption. As aggregate demand and the wage rate increase, households raise their labor supply. Because of this complementarity, desired consumption for all households rises very sharply, even for non-hand-to-mouth households with low marginal propensities to consume.

In Tables E.1 and E.2 in online Appendix E, we report results from a comprehensive robustness analysis of the key parameters that govern behavior in the “heterogeneous agent block” of the model. We show that changes in the tightness of borrowing limits, the cost of borrowing, and the adjustment cost function can have large effects on the level of liquid wealth holdings and the fraction of poor and wealthy hand-to-mouth households. However, in all cases the share of indirect effects remains around 80 percent.

⁴⁶When $\omega = 0.1$, the role of the illiquid return component among the indirect effects is much more important. We come back to the reasons in Section IVC.

⁴⁷Table 7 reports results from a more aggressive monetary policy rule and lower price stickiness; a less aggressive policy rule or higher price stickiness have similarly sized opposite effects.

⁴⁸The intra-period utility function is specified as $\log(c - \psi \frac{h^{1+\nu}}{1+\nu})$, with $\nu = 1$ and ψ set to replicate average hours equal to one-third of time endowment.

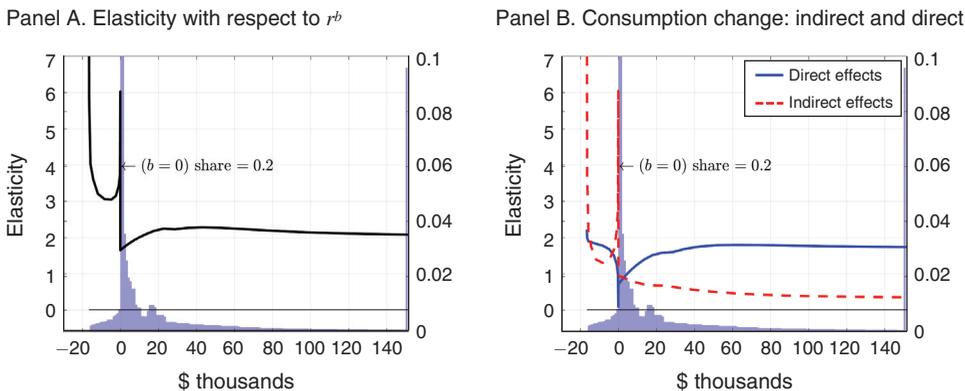


FIGURE 5. CONSUMPTION RESPONSES BY LIQUID WEALTH POSITION

These experiments (including those done under alternative fiscal adjustments that we discuss below) reveal another, related, robust feature of HANK models. The partial-equilibrium (or direct) elasticity of aggregate consumption, the consumption response to changes in the liquid rate, keeping all other prices and taxes/transfers unchanged, never deviates too much from 0.55 (its baseline value) even across configurations where the total elasticity differs greatly.⁴⁹ This magnitude is considerably lower than in all the versions of RANK, where the direct elasticity is always above 1.9 (see Table 1).

C. The Distribution of the Monetary Transmission

To better understand why the direct effects of an unexpected reduction in interest rates are small and the indirect effects are large in HANK relative to RANK, it is instructive to inspect the consumption response to the monetary policy shock across the entire distribution of liquid wealth holdings.

Panel A of Figure 5 shows the elasticity of average consumption of households with a given liquid wealth level to the change in the interest rate at each point in the liquid wealth distribution (black line, left axis), along with the corresponding consumption shares of each liquid wealth type (light shaded histogram, right axis).⁵⁰ The distribution of consumption responses displays big spikes at the borrowing constraint $b = \underline{b}$ and at $b = 0$. Our model features few households at the borrowing limit, but those with zero liquid wealth account for 20 percent of total consumption and have an elasticity of around 6. Because many of these households have moderate

⁴⁹We here purposely exclude the result in column 3 of Table 7 for $\omega = 0.1$ where the aggregate partial equilibrium elasticity is 0.99 because it obtains from an extremely unrealistic parameterization with an investment elasticity of 105 and an output elasticity of 25.

⁵⁰Note that the figure reports the elasticity of consumption on impact of the monetary policy shock, in contrast to the numbers in Table 7 which report elasticities over the first year. This is because the impact elasticities are considerably easier to compute. Integrating the elasticities in the figure weighted by the consumption shares yields (the negative of) the overall impact elasticity to the monetary shock, which is -2.81 . The average consumption of households with a given liquid wealth level b is defined as $C_t(b) = \int c_t(a, b, z) \mu_t(da, b, dz)$ so that aggregate consumption satisfies $C_t = \int_{\underline{b}}^{\infty} C_t(b) db$. Therefore, the overall elasticity is a consumption-weighted average of the elasticities at each level of liquid wealth.

income and own illiquid assets (i.e., they are wealthy hand-to-mouth), their consumption share is much larger than in models where hand-to-mouth are income- and wealth-poor. All other households with positive liquid assets, representing around 80 percent of total consumption expenditures, contribute with an elasticity of around 2. A back-of-the-envelope calculation yields $0.2 \times 6 + 0.8 \times 2 = 2.8$, which is roughly the overall impact elasticity.

Panel B separates the total elasticity into the direct and indirect elasticities. These two additive components measure the strength of the direct and general equilibrium channels of monetary policy. We now examine each of them separately.

Why Are Direct Effects Small?—Panel B of Figure 5 reveals that the direct effects are highest for households close to the borrowing constraint, decline to zero for households with no liquid wealth and, as liquid wealth grows, increase until the direct elasticity peaks just below a value of 2. Beyond a sufficiently high level of liquid wealth, direct effects start to slowly decline. The aggregate partial-equilibrium elasticity is only about 0.55 (see column 1 of Table 7) because most of the population and consumption distribution is between 0 and \$20,000 of liquid wealth, and in that range the direct elasticity is quite small.

To better understand these cross-sectional patterns, we further split the direct elasticity into a substitution effect, an income effect, and a portfolio reallocation effect.⁵¹ Our decomposition is conceptually identical to the one in Auclert (2016), but we build on recent work by Olivi (2017), who substantially generalizes Auclert's approach to allow for persistent price changes and a more general stochastic process for idiosyncratic risk. Online Appendix F contains the exact expression for this decomposition and its interpretation.

Panel A of Figure 6 implements this decomposition. The solid blue line plots the direct effect, i.e., the same line as in panel B of Figure 5. The figure then breaks it down into its three components. Households at the borrowing limit and those with zero holdings of liquid wealth do not substitute intertemporally because they are at a kink in their budget constraint. For those with positive liquid wealth, the substitution effect gradually increases until leveling off at 1.95 for high liquid holdings. This is the size of the substitution effect for a household who is fully insured against idiosyncratic income risk, as in RANK.

The income effect is a monotonically decreasing function of liquid wealth. It is positive for borrowers since lower interest payments on their debt translate into higher consumption and is negative for lenders with positive liquid wealth.⁵² For households with sufficiently high holdings of liquid wealth (outside the range plotted in the graph), the income effect becomes so strong that the direct elasticity becomes negative. As noted above, the direct response of households with positive, but moderate, amounts of liquid wealth (\$1,000 to \$20,000) is small and this is what accounts for most of the small aggregate direct elasticity.

⁵¹ More precisely, what we call the income effect is a combination of a classic income effect and a wealth/endowment effect.

⁵² Di Maggio, Kermani, and Ramcharan (2014), Flodén et al. (2016) study borrowers with adjustable rate mortgages who faced changes in monthly interest payments, and find evidence of a positive consumption response to a drop in monthly payments. In addition, Cloyne, Ferreira, and Surico (2015) offer supporting evidence that the direct channel is small relative to indirect effects occurring through changes in household labor income.

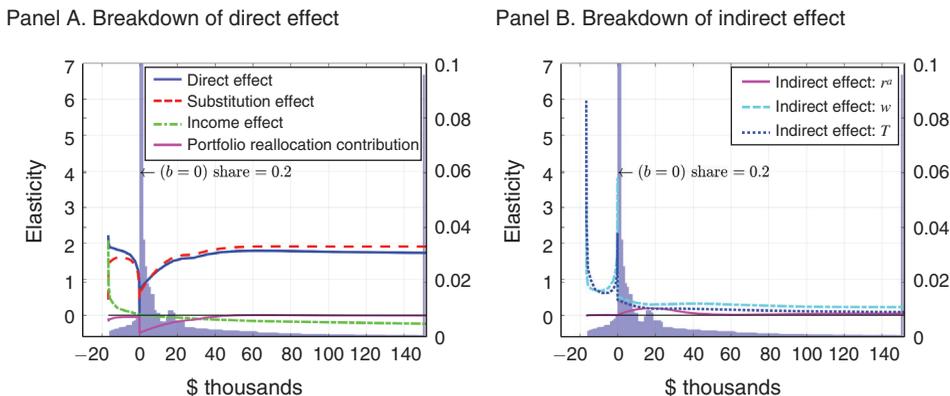


FIGURE 6. CONSUMPTION RESPONSES BY LIQUID WEALTH POSITION

Why do these low MPC households not respond more strongly to the reduction in interest rates? The decomposition in panel A of Figure 6 suggests that the reason is twofold. First, although these households currently have low MPCs, they still face the possibility of having higher MPCs in the future because of the occasionally binding borrowing limits, a point emphasized by McKay, Nakamura, and Steinsson (2016). Finally, portfolio rebalancing (solid pink line) has a considerably negative effect for households with positive but moderate amounts of liquid wealth.

Relationship to Auclert (2016): As already noted, our decomposition is closely related to the one proposed by Auclert (2016, Theorems 1 and 3). Auclert’s Theorem 1 breaks down households’ micro consumption response into income and substitution effects. He refers to the income effect of the interest rate change as “unhedged interest rate exposure.” Auclert’s Theorem 3 then aggregates the micro decomposition into an aggregate decomposition as we do in equation (30). The combination of indirect effects from wages and fiscal policy can be interpreted as his “aggregate income” and “earnings heterogeneity” channels, expanded to include income from the government besides labor income.

At the same time, there are also some important differences between the two decompositions. First, our model does not feature his Fisher channel because all assets in our model are real. Second, Auclert emphasizes heterogeneity in asset maturities, whereas our emphasis is on asset liquidity. Third, our decomposition can handle persistent dynamics of the economy. This is important because in models like ours and Auclert’s, even purely transitory one-time shocks typically lead to endogenous persistence through movements in the wealth distribution, a case to which Auclert’s decomposition does not apply. Finally, his decomposition relies heavily on being able to collapse period budget constraints into a single present-value constraint: as a result, his approach cannot handle binding borrowing limits, a wedge between borrowing and savings rates, or illiquid assets, features that are at the heart of our model.

Why Are Indirect Effects Large?—Panel B of Figure 5 reveals that indirect effects are very large for households with zero liquid wealth. The presence of hand-to-mouth

households is thus a key determinant of the transmission mechanism of monetary policy on the macroeconomy.

Panel B of Figure 6, which offers a breakdown of the indirect effect among its three components, shows that these households respond sharply to changes in both labor income and government transfers that occur in equilibrium in the wake of a monetary shock. The rise in labor income is a consequence of an expansionary monetary shock that increases demand for final goods. Transfers rise because the interest payments on government debt fall and because the rise in aggregate income increases tax revenues. This mechanism shares similarities with TANK models with government debt where, like in HANK, the presence of non-Ricardian households means that the fiscal response can play an important role in the indirect effects of monetary policy.

Finally, the combined indirect effect due to changes in r^a and q is slightly negative, but very small, everywhere in the distribution. Our model inherits the typical feature of the standard New Keynesian model that markups and profits fall in a monetary expansion. Since the stock price q is the present discounted value of future profits, q drops as well. A sizable literature examines the response of equity prices to monetary policy shocks and finds positive, but only weakly significant, responses of stock prices to expansionary monetary policy shocks.⁵³ There are a number of potential strategies for generating procyclical profits, and hence stock prices, in New Keynesian models that would also apply in HANK. Chief among these is the introduction of sticky wages.⁵⁴

While capital gains on equity are always countercyclical, our model is capable of generating both procyclical and countercyclical returns through their dividend component, depending on the assumption made on the fraction of profits ω reinvested in illiquid accounts. As shown in columns 2 and 3 of Table 7, this parameter changes the share of the indirect effect due to the illiquid return r^a .⁵⁵ Sharper empirical evidence on the response of various asset prices to monetary policy shocks as well as the design of a HANK model that is consistent with this evidence should be a priority for future research.⁵⁶

D. *The Role of the Fiscal Response to a Monetary Shock*

We now discuss some important implications of Ricardian non-neutrality in HANK. In Table 8 we report the overall response and decomposition for alternative assumptions about how the government satisfies its intertemporal budget constraint after a monetary shock.

⁵³ See Rigobon and Sack (2004), Bernanke and Kuttner (2005), and Gürkaynak, Sack, and Swanson (2005). In the words of Rigobon and Sack (2004, p. 1568), the literature has been “somewhat inconclusive about the significance of the response of stock prices to monetary policy actions.”

⁵⁴ Indeed, this is the route taken by Challe and Giannitsarou (2014) in the context of a RANK model.

⁵⁵ Some readers may argue that the indirect effects due to illiquid returns r^a should be counted as direct effects because all effects working through changes in asset returns are intimately linked due to arbitrage considerations. Note that even in the case of column 3 in Table 7 where the indirect effects due to r^a are large, the combined effect due to r^a and r^b is still only 30 percent of the overall effect.

⁵⁶ That correctly modeling asset price movements is potentially important is also consistent with a result in Werning (2015), who shows that the consumption response to monetary policy depends on the cyclicity of asset prices.

TABLE 8—IMPORTANCE OF FISCAL RESPONSE TO MONETARY SHOCK

	<i>T</i> adjusts (1)	<i>G</i> adjusts (2)	τ adjusts (3)	B^g adjusts (4)
Change in r^b (pp)	-0.28	-0.23	-0.33	-0.34
Elasticity of <i>Y</i>	-3.96	-7.74	-3.55	-2.17
Elasticity of <i>I</i>	-9.43	-14.44	-8.80	-5.07
Elasticity of <i>C</i>	-2.93	-2.80	-2.75	-1.68
Partial eq. elasticity of <i>C</i>	-0.55	-0.60	-0.56	-0.71
<i>Component of percent change in C due to</i>				
Direct effect: r^b	19	21	20	42
Indirect effect: <i>w</i>	51	81	62	49
Indirect effect: <i>T</i>	32	-	-	9
Indirect effect: τ	-	-	18	-
Indirect effect: r^a and <i>q</i>	-2	-2	0	0

Notes: Average responses over the first year. Column 1 is the baseline specification in which transfers *T* adjust to balance the government budget constraint. In column 2 government expenditure *G* adjusts, and in column 3 the labor income tax τ adjusts. In column 4 government debt adjusts, as described in the main text.

Column 1 contains the baseline case, in which government expenditures and debt are held constant, and transfers adjust in every instant. When, instead, government expenditures adjust, the overall impact of monetary policy on aggregate output is stronger (column 2). This is because when transfers adjust, only high MPC households increase consumption, and by less than one-for-one with the transfer; when government expenditures adjust, the reduced interest payments on debt translate one-for-one into an increase in aggregate demand, which contributes directly to an increase in output. The elasticity of private consumption is similar to the baseline, but the bulk of the indirect effects are accounted for by higher labor income rather than a combination of labor income and transfers.

In column 3, we let the tax rate adjust. Compared to the case where transfers rise, here there are offsetting forces: on the one hand, a lower tax rate expands labor supply across the board, whereas the higher transfers have a small negative impact on hours worked; on the other hand, lowering taxes is less redistributive than more generous lump-sum transfers and, thus, spurs a smaller demand for private consumption. Overall, the results are similar to the baseline.

The remaining alternative is to let government debt absorb the majority of the fiscal imbalance in the short run. In the economies of columns 1 and 2, a sizable fraction of the overall effect of monetary policy is due to additional government transfers or expenditures from reduced debt payments. Without this additional stimulus to aggregate demand, labor income does not increase as much and indirect effects account for a smaller share of the total (60 percent compared to 80 percent in the other two scenarios, but still an order of magnitude larger than in RANK). As a result, when government debt absorbs the slack, the monetary shock has a much smaller impact on the economy, roughly one-half of the baseline value.⁵⁷

⁵⁷In this experiment, we assume that lump-sum transfers jump by a very small amount on impact and then decay back to their steady-state level at a slow exogenous rate. Given the assumed rate of decay, the initial jump is

In conclusion, our second quantitative insight into the transmission mechanism of monetary policy is that the type of fiscal adjustment following the monetary shock matters for the effectiveness of policy.⁵⁸ This result represents another important deviation from RANK and from versions of the HANK model, such as the one developed by Werning (2015) where the overall efficacy of monetary policy does not depend on liquidity constraints and incomplete markets.

E. *The Role of Two Assets and Micro Heterogeneity*

At this stage of our analysis, two questions naturally arise: what do we gain from the two-asset version of HANK relative to the one-asset versions that have been studied in the existing literature? What do we gain from a realistic model of household heterogeneity relative to the simpler spender-saver structure of TANK models?

Two-Asset versus One-Asset HANK Models.—We choose the version of the one-asset HANK model as in McKay, Nakamura, and Steinsson (2016) in which all wealth is held as liquid government bonds, and we let transfers adjust to balance the government budget constraint following the monetary shock.⁵⁹

Recall that in our calibrated two-asset HANK model the wealth-to-output ratio was over 3 (Table 2) and the average quarterly MPC out of \$500 was 0.16 (Figure 2). In one-asset HANK models, however, there is a well-known tension between matching the high observed aggregate wealth-to-output ratios and generating a large average MPC. Panel A of Figure 7, which plots aggregate wealth and the average quarterly MPC out of \$500 in the one-asset HANK model for values of the discount rates ρ between 2.5 percent p.a. to 7.5 percent p.a., illustrates this tension: the one-asset model can generate high average wealth or a high MPC, but not both simultaneously.⁶⁰

Notwithstanding this failure of the one-asset model, panel B of Figure 7 plots the direct and total elasticities of aggregate consumption following a monetary policy shock (the analogues for the two-asset model are in Table 7). For low discount rates in which there is a large amount of liquid wealth in the economy, the direct elasticity becomes negative because of the strong wealth effects that pull down the direct channel. For high discount rates, the direct elasticity rises but is always a small share of the overall elasticity. This is because even though wealth effects are now modest due to the smaller amount of wealth in the economy, there is a larger fraction of hand-to-mouth households and so the intertemporal substitution channel is muted. The total elasticity is hump-shaped with respect to ρ because of two

chosen so that the government's budget constraint holds in present value terms. In column 4 of Table 8 the transfer decays at a quarterly rate of 0.02. We experimented with smaller and bigger decay rates and our main conclusions are unchanged.

⁵⁸Clearly, this insight applies more generally in HANK models: how the fiscal authority responds to any shock, not just monetary, that affects the government budget constraint is bound to shape the aggregate impact of that shock.

⁵⁹The model can be thought of as the limit of our two-asset model as the capital share α goes to zero, implying that the marginal product of capital is zero. Because we assume that a fraction $\omega = \alpha$ of firm profits are reinvested directly into illiquid assets (see Section IIIA), this also implies that all profits are paid out in liquid form. As a result, the illiquid return r_t^i goes to zero and all wealth is held as liquid assets.

⁶⁰The rest of the calibration is identical to the one in our baseline model, except for $\alpha = \omega = 0$.

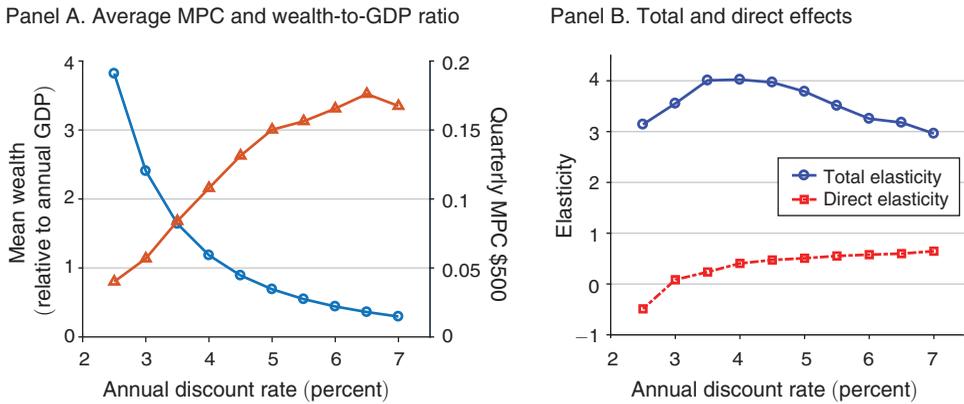


FIGURE 7. KEY FEATURES OF ONE-ASSET MODEL FOR DIFFERENT CALIBRATIONS

offsetting forces: as the discount rate increases, MPCs rise (panel A) resulting in larger indirect effects; at the same time, the amount of liquid wealth (all of which is government debt) decreases, resulting in a weaker fiscal response to the monetary shock, and hence smaller indirect effects.

Perhaps surprisingly, one calibration of the one-asset model replicates many features of our two-asset model closely: with a discount rate of 7 percent, the average MPC is 0.16, the direct elasticity is just above 0.5, and the overall elasticity is just below 3. The one moment of the data that this calibration misses completely is the total amount of wealth in the economy (0.25 versus 3). But if one interprets wealth as liquid wealth only, this calibration performs well in that dimension (a value of 0.25 just as in Table 2). This calibration can therefore be thought of as the “liquid-wealth-only calibration,” advocated by Carroll, Slacalek, and Tokuoka (2017).

This result then raises the question of what is to be gained from studying monetary policy through the lens of a two-asset HANK model, rather than a one-asset HANK model calibrated only to liquid wealth. The answer is that this latter model completely abstracts from capital, and the responses of quantity and price of capital greatly matter for the monetary transmission, through the indirect channel. To illustrate this point, recall columns 2 and 3 of Table 7. When $\omega = 0.1$, strong procyclical movements in the illiquid rate of return r^a , stock price q , and investment I result in the total consumption response more than doubling relative to the baseline. Instead, when $\omega = 1$, the illiquid price falls sharply in the wake of a monetary expansion and the total consumption elasticity shrinks to two-thirds of the baseline. While these two alternative calibrations are counterfactual in some dimensions (in particular, with regard to the extreme investment responses), they illustrate qualitatively an important point: the effects of monetary policy depend strongly on how investment and equity returns move in equilibrium. It is hard to see how the liquid-wealth-only calibration of the one-asset model could ever accommodate these forces.

Micro Heterogeneity versus Spender-Saver Structure.—In this section, we compare our model to the TANK models analyzed in Section IB. The total consumption elasticity in TANK models is somewhat smaller, but the share of direct effects

is roughly three times larger than in HANK. As a result, the direct elasticity of aggregate consumption in TANK is always around 1.38, which is 2.5 times higher than in HANK.

The direct elasticity in TANK is entirely determined by the savers and the MPC of the savers is always very small, equal to ρ , the discount rate. This observation has two implications. First, in HANK the negative income effect from an expansionary monetary shock is much stronger than in TANK. Comparing the versions of TANK with wealth in positive supply with the version without wealth (columns 6 and 7 versus column 5 in Table 1) reveals that the direct elasticity is basically unchanged, which implies income effects are negligible precisely because of the low MPC. Second, the substitution effect is much weaker in HANK, even for low MPC households, because the prospect of hitting a kink in the budget constraint and having a high MPC in the future shortens their effective planning horizon.

This result underscores the importance of modeling heterogeneity through uninsurable earnings shocks as opposed to via built-in differences in preferences.

V. Monetary Policy Trade-Offs in HANK

We have thus far emphasized two main results. First, in our HANK model the indirect effects of monetary policy on aggregate consumption far outweigh the direct effects that are dominant in RANK models. Second, the overall response of aggregate consumption to a cut in interest rates may be larger or smaller than in RANK models, depending on a number of factors that are neutral in RANK, in particular the fiscal reaction to the monetary expansion.

We now highlight two implications of these differences for some key trade-offs that policymakers face in the conduct of monetary policy. First, we study the choice between sharper but more transitory versus smaller but more persistent interest rate cuts. Second, we analyze the most classical trade-off of monetary policy: the one between inflation and real activity.

A. Trade-Off between Size and Persistence of Monetary Shocks

Our aim is to compare, within RANK and HANK models, a transitory drop in the interest rate of a given size and persistence with a smaller but more persistent drop.

Recall, from Section I, that the aggregate Euler equation in RANK implies

$$(36) \quad C_0 = \bar{C} \exp\left(-\frac{1}{\gamma} \int_0^{\infty} (r_s - \rho) ds\right).$$

The integral, which we hereafter denote as R_0 , is the cumulative deviation of the real interest rate from the natural rate ρ . The elasticity of aggregate consumption at impact with respect to R_0 , $-d \log C_0 / dR_0$, is always equal to $1/\gamma$ and equal to 1 under our calibrated value for the IES. Crucially, this *cumulative elasticity* is independent of the particular path of the real rate. More or less persistent paths with the same cumulative deviation R_0 have the same impact on aggregate consumption. Put

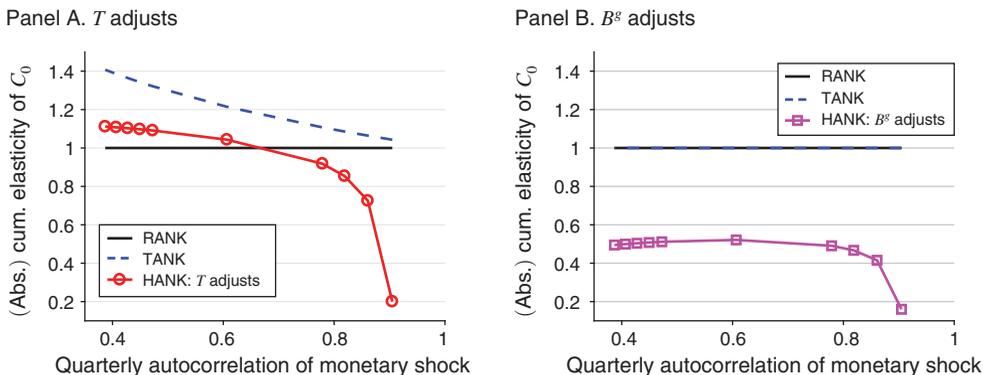


FIGURE 8. CUMULATIVE ELASTICITY OF AGGREGATE CONSUMPTION BY PERSISTENCE OF THE SHOCK

differently, RANK models feature a neutrality property with respect to the timing of monetary policy and do not feature a size-persistence trade-off.

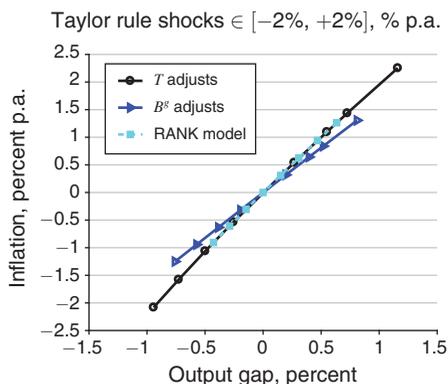
This neutrality property of RANK models does not hold in HANK. Panel A of Figure 8 plots the cumulative elasticity of aggregate consumption at impact with respect to different values for the persistence of the innovation $e^{-\eta}$ for our baseline fiscal policy scenario in which transfers adjust. Intuitively, persistence is irrelevant for the non-hand-to-mouth households, as in RANK, but it does affect the response of hand-to-mouth households. When shocks are persistent, a large portion of the interest rate cut, and the associated relaxation of the government budget constraint, occurs in the future. Hence, the hand-to-mouth households receive a smaller increase in transfers upon the impact of the shock and so their consumption response is weaker. As a result, the cumulative elasticity, which is invariant to persistence in RANK, declines sharply with persistence in HANK. The failure of Ricardian equivalence implies that not only the timing of fiscal policy matters but also that of monetary policy. For comparison, we also plot the cumulative elasticity $-d \log C_0/dR_0$ for the simple TANK model of Section IB. As in HANK, the timing of monetary policy matters. This is again due to the failure of Ricardian equivalence. However, the difference is much smaller and, in contrast to HANK, the consumption response in TANK is always weakly larger than that in RANK.

Panel B of Figure 8 repeats the exercise for the case where government debt adjusts. As explained in the context of Table 8, in this case the consumption response in HANK is considerably diminished because of the lack of transfers accruing to hand-to-mouth households. Even in the absence of transfers though, highly persistent interest rate cuts are considerably less potent than sharp but transitory ones.

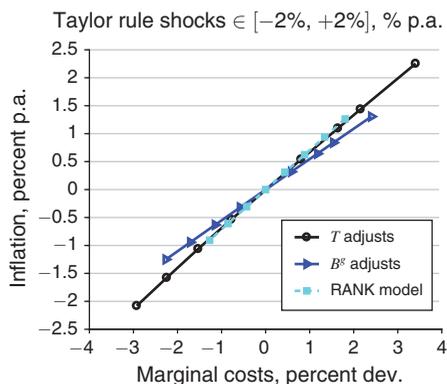
B. Inflation-Activity Trade-Off

In New Keynesian models, any desired increase in aggregate output can be achieved by an appropriate choice of the size of the monetary innovation. A relevant question is about the cost of such monetary stimulus in terms of inflation. That is, the proper conduct of monetary policy requires knowledge of the trade-off between inflation and real activity.

Panel A. Inflation-output gap



Panel B. Inflation-marginal cost



Panel C. Marginal cost-output

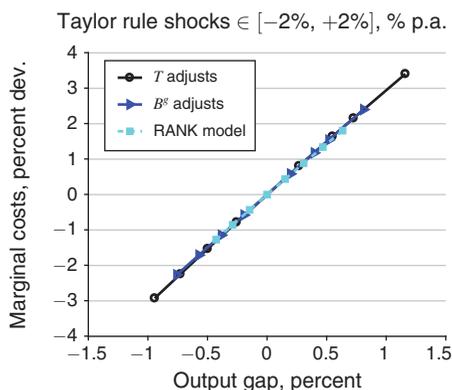


FIGURE 9. ANALYSIS OF INFLATION-ACTIVITY TRADE-OFF

Figure 9 graphically examines this trade-off in RANK and HANK.⁶¹ Panel A plots the inflation-output trade-off, panel B the inflation-marginal cost relationship, and panel C the marginal cost-output relationship. For each model, we feed monetary policy shocks ϵ_0 ranging from -2 percent to $+2$ percent annually into the Taylor rule.

We begin by comparing the T-adjusts case in HANK, our baseline, with RANK. The main result, visible in panel A, is that the inflation-activity trade-off is similar. Panels B and C illustrate that the way movements in marginal costs induced by policy shocks translate into movements in inflation and output is basically identical across models. The reason is that the inflation-activity relationship is largely determined by the New Keynesian side (summarized by the Phillips curve and Taylor rule), which is the same in RANK and HANK.

Although the slopes are the same across the two models, the length of the lines in panel A differs sharply. This is a reflection of the different elasticities of economic

⁶¹For these sets of results, we use the richer RANK model outlined in Section IC and in online Appendix A.5.

activity to the monetary shock in the two models. As explained in Section IV, the elasticity of C in HANK under the T-adjusts case is higher than in RANK. As a consequence, the same expansionary policy shock generates more inflation and a larger output gap in HANK.

An examination of the inflation-activity trade-off across different types of fiscal adjustments in HANK (T-adjusts versus B-adjusts) reveals an additional finding. As opposed to RANK, where Ricardian neutrality holds, in HANK the type of fiscal adjustment matters for the slope of this trade-off. Panel A implies that a more passive fiscal adjustment rule, where debt absorbs the change in interest payments following the monetary shock, is associated to a more favorable trade-off, i.e., a flatter line. The reason, which can be seen in panel B, is the different marginal cost-inflation equilibrium relationship: in the B-adjusts case, changes in marginal costs are spread out over a longer period of time and, as a result, inflation reacts less at impact.

Finally, note that in our version of HANK the mapping from the output gap to marginal costs is the same as in RANK (see panel C). An interesting avenue for future research is to examine whether this is true more generally. In principle, this mapping may depend on the heterogeneity in the economy, which would lead to a different inflation-output trade-off (panel A) between HANK and RANK.

VI. Conclusion

In our Heterogeneous Agent New Keynesian (HANK) framework, monetary policy affects aggregate consumption primarily through indirect effects that arise from a general equilibrium increase in labor demand. This finding is in stark contrast to Representative Agent New Keynesian (RANK) economies, where intertemporal substitution drives virtually all of the transmission from interest rates to consumption. Throughout the paper, we argued that this difference between HANK and RANK matters a great deal for the conduct of monetary policy.

Our model's ability to match the cross section of household portfolios, wealth distribution, and microeconomic consumption behavior lies at the heart of this set of results. Nonetheless, the household side of the model could be improved in a number of dimensions. The model lacks a distinction between net and gross positions, which would be necessary to assess the role of household leverage on monetary transmission. The model also lacks a distinction between real and nominal assets, which is a consequence of all assets in our economy being of infinitely short duration. This distinction would be necessary to study revaluation (or Fisher) effects of monetary policy as in Doepke and Schneider (2006) and Auclert (2016). Together these two abstractions mean that our model cannot generate a commonly observed household portfolio: illiquid housing assets together with long-term nominal mortgage debt with either fixed or variable nominal coupon payments. Such a balance-sheet configuration brings additional channels of monetary transmission: recent progress in this area has been made, for example, by Garriga, Kydland, and Šustek (2017) and Wong (2016).

There are several other open areas for the next generation of HANK models to address. First, in our version of HANK, the price of illiquid assets (which can be interpreted as stock and house prices) comoves slightly negatively with a monetary shock. The empirical evidence on this correlation is inconclusive, but if anything

it points to a positive correlation. Getting this comovement right is important for the size of the portfolio reallocation effect that mutes the intertemporal substitution channel in HANK. On a similar note, more direct evidence is needed on the size of transaction costs and on portfolio rebalancing behavior following shocks.

Second, we have only studied deterministic transitional dynamics following a one-time monetary shock. The computational method recently developed by Ahn et al. (forthcoming) will allow future HANK models to incorporate aggregate fluctuations into the economic environment.

Third, in HANK the way that fiscal policy responds to an interest rate change profoundly affects the overall effectiveness of monetary policy, a result that is also at odds with the Ricardian nature of standard RANK economies. Currently, there is no empirical evidence that reveals what type of fiscal adjustment is the most likely to occur in practice, following a monetary shock. We view this as a fruitful area of research.

Fourth, we have focused on the macroeconomic effects of conventional monetary policy, i.e., shocks to the Taylor rule, in economies that are far from the zero lower bound on nominal interest rates. When the lower bound is binding, the relevant monetary instrument switches from short-term rates to forward guidance and asset purchases. Our experiments on monetary shocks with different levels of persistence suggest that in HANK models, forward guidance may be less effective than conventional monetary policy, providing a possible solution to the forward guidance puzzle (Del Negro, Giannoni, and Patterson 2012). In Kaplan, Moll, and Violante (2016b) we fully articulate this point following the lead of McKay, Nakamura, and Steinsson (2016) and Werning (2015). The presence of assets with different degrees of liquidity also makes the framework a natural one to analyze the macroeconomic effects of large-scale asset purchases (quantitative easing).

Finally, in RANK models there is a clear pecking order between monetary and fiscal policies: in economies that are away from the zero lower bound, monetary policy can by itself restore the first-best equilibrium allocation (what Blanchard and Galí 2007 have termed the “divine coincidence”). An important question that remains unanswered is the design of optimal policy in HANK economies where the presence of incomplete markets and distributional concerns, in addition to nominal rigidities, breaks such “divine coincidence.”

REFERENCES

- Achdou, Yves, Jiequn Han, Jean-Michel Lasry, Pierre-Louis Lions, and Benjamin Moll. 2017. “Income and Wealth Distribution in Macroeconomics: A Continuous-Time Approach.” Unpublished.
- Ahn, SeHyoun, Greg Kaplan, Benjamin Moll, Thomas Winberry, and Christian Wolf. Forthcoming. “When Inequality Matters for Macro and Macro Matters for Inequality.” In *NBER Macroeconomics Annual 2017*, Vol. 32, edited by Martin S. Eichenbaum and Jonathan Parker. Chicago: University of Chicago Press.
- Auclert, Adrien. 2016. “Monetary Policy and the Redistribution Channel.” Unpublished.
- Baker, Scott R. 2016. “Debt and the Consumption Response to Household Income Shocks.” Unpublished.
- Bayer, Christian, Ralph Lütticke, Lien Pham-Dao, and Volker Tjaden. 2015. “Precautionary Savings, Illiquid Assets, and the Aggregate Consequences of Shocks to Household Income Risk.” Unpublished.
- Benhabib, Jess, and Alberto Bisin. 2016. “Skewed Wealth Distributions: Theory and Empirics.” National Bureau of Economic Research Working Paper 21924.

- Bernanke, Ben S., and Kenneth N. Kuttner.** 2005. "What Explains the Stock Market's Reaction to Federal Reserve Policy?" *Journal of Finance* 60 (3): 1221–57.
- Bilbiie, Florin O.** 2008. "Limited Asset Markets Participation, Monetary Policy and (Inverted) Aggregate Demand Logic." *Journal of Economic Theory* 140 (1): 162–96.
- Bilbiie, Florin O.** 2017. "The New Keynesian Cross: Understanding Monetary Policy with Hand-to-Mouth Households." Centre for Economic Policy Research Discussion Paper 11989.
- Blanchard, Olivier, and Jordi Galí.** 2007. "Real Wage Rigidities and the New Keynesian Model." *Journal of Money, Credit and Banking* 39 (S1): 35–65.
- Blundell, Richard, Luigi Pistaferri, and Itay Saporta-Eksten.** 2016. "Consumption Inequality and Family Labor Supply." *American Economic Review* 106 (2): 387–435.
- Broda, Christian., and Jonathan A. Parker.** 2014. "The Economic Stimulus Payments of 2008 and the Aggregate Demand for Consumption." *Journal of Monetary Economics* 68 (S): S20–S36.
- Broer, Tobias, Niels-Jakob Harbo Hansen, Per Krusell, and Erik Öberg.** 2016. "The New Keynesian Transmission Mechanism: A Heterogeneous-Agent Perspective." Centre for Economic Policy Research Discussion Paper 11382.
- Caballero, Ricardo J., and Emmanuel Farhi.** 2018. "The Safety Trap." *Review of Economic Studies* 85 (1): 223–74.
- Campbell, John Y., and N. Gregory Mankiw.** 1989. "Consumption, Income and Interest Rates: Reinterpreting the Time Series Evidence." In *NBER Macroeconomics Annual 1989*, Vol. 4, edited by Jonathan A. Parker and Michael Woodford, 185–246. Cambridge, MA: MIT Press.
- Campbell, John Y., and N. Gregory Mankiw.** 1991. "The Response of Consumption to Income: A Cross-Country Investigation." *European Economic Review* 35 (4): 723–56.
- Canzoneri, Matthew B., Robert E. Cumby, and Behzad T. Diba.** 2007. "Euler Equations and Money Market Interest Rates: A Challenge for Monetary Policy Models." *Journal of Monetary Economics* 54 (7): 1863–81.
- Carroll, Christopher D., Jiri Slacalek, and Kiichi Tokuoka.** 2017. "The Distribution of Wealth and the Marginal Propensity to Consume." *Quantitative Economics* 8 (3): 977–1020.
- Challe, Edouard, and Chryssi Giannitsarou.** 2014. "Stock Prices and Monetary Policy Shocks: A General Equilibrium Approach." *Journal of Economic Dynamics and Control* 40: 46–66.
- Challe, Edouard, Julien Matheron, Xavier Ragot, and Juan F. Rubio-Ramirez.** 2017. "Precautionary Saving and Aggregate Demand." *Quantitative Economics* 8 (2): 435–78.
- Chetty, Raj, Adam Guren, Day Manoli, and Andrea Weber.** 2011. "Are Micro and Macro Labor Supply Elasticities Consistent? A Review of Evidence on the Intensive and Extensive Margins." *American Economic Review* 101 (3): 471–75.
- Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans.** 2005. "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy." *Journal of Political Economy* 113 (1): 1–45.
- Cloyne, James, Clodomiro Ferreira, and Paolo Surico.** 2015. "Monetary Policy When Households Have Debt: New Evidence on the Transmission Mechanism." Unpublished.
- Cloyne, James, and Paolo Surico.** 2016. "Household Debt and the Dynamic Effects of Income Tax Changes." *Review of Economic Studies* 84 (1): 45–81.
- Congressional Budget Office.** 2013. "The Distribution of Federal Spending and Taxes in 2006." Washington, DC: Congressional Budget Office.
- Del Negro, Marco, Marc Giannoni, and Christina Patterson.** 2012. "The Forward Guidance Puzzle." Federal Reserve Bank of New York Staff Report 574.
- De Nardi, Mariacristina, Giulio Fella, and Gonzalo Paz Pardo.** 2016. "The Implications of Richer Earnings Dynamics for Consumption and Wealth." National Bureau of Economic Research Working Paper 21917.
- Den Haan, Wouter J., Pontus Rendahl, and Markus Riegler.** 2017. "Unemployment (Fears) and Deflationary Spirals." *Journal of the European Economic Association* <http://dx.doi.org/10.1093/jeaa/jvx040> (accessed February 6, 2018).
- Díaz-Giménez, Javier, Andrew Glover, and José-Víctor Ríos-Rull.** 2011. "Facts on the Distributions of Earnings, Income, and Wealth in the United States: 2007 Update." *Federal Reserve Bank of Minneapolis Quarterly Review* 34 (1): 2–31.
- Di Maggio, Marco, Amir Kermani, and Rodney Ramcharan.** 2014. "Monetary Policy Pass-Through: Household Consumption and Voluntary Deleveraging." Unpublished.
- Doepke, Matthias, and Martin Schneider.** 2006. "Inflation and the Redistribution of Nominal Wealth." *Journal of Political Economy* 114 (6): 1069–97.
- Eusepi, Stefano, and Bruce Preston.** 2017. "Fiscal Foundations of Inflation: Imperfect Knowledge." Unpublished.
- Fagereng, Andreas, Martin B. Holm, and Gisle J. Natvik.** 2016. "MPC Heterogeneity and Household Balance Sheets." Statistics Norway Research Department Discussion Paper 852.

- Flodén, Martin, Matilda Kilström, Jósef Sigurdsson, and Roine Vestman.** 2016. "Household Debt and Monetary Policy: Revealing the Cash-Flow Channel." Unpublished.
- Gali, Jordi, and Mark Gertler.** 1999. "Inflation Dynamics: A Structural Econometric Analysis." *Journal of Monetary Economics* 44 (2): 195–222.
- Gali, Jordi, J. David López-Salido, and J. Vallés.** 2007. "Understanding the Effects of Government Spending on Consumption." *Journal of the European Economic Association* 5 (1): 227–70.
- Garriga, Carlos, Finn E. Kydland, and Roman Sustek.** 2017. "Mortgages and Monetary Policy." *Review of Financial Studies* 30 (10): 3337–75.
- Gertler, Mark, and Peter Karadi.** 2015. "Monetary Policy Surprises, Credit Costs, and Economic Activity." *American Economic Journal: Macroeconomics* 7 (1): 44–76.
- Gornemann, Nils, Keith Kuester, and Makoto Nakajima.** 2014. "Doves for the Rich, Hawks for the Poor? Distributional Consequences of Monetary Policy." Unpublished.
- Greenwood, Jeremy, Zvi Hercowitz, and Gregory W. Huffman.** 1988. "Investment, Capacity Utilization, and the Real Business Cycle." *American Economic Review* 78 (3): 402–17.
- Guerrieri, Veronica, and Guido Lorenzoni.** 2017. "Credit Crises, Precautionary Savings, and the Liquidity Trap." *Quarterly Journal of Economics* 132 (3): 1427–67.
- Gürkaynak, Refet S., Brian Sack, and Eric T. Swanson.** 2005. "Do Actions Speak Louder Than Words? The Response of Asset Prices to Monetary Policy Actions and Statements." *International Journal of Central Banking* 1 (1): 55–93.
- Guvenen, Fatih, Fatih Karahan, Serdar Ozkan, and Jae Song.** 2015. "What Do Data on Millions of U.S. Workers Reveal about Life-Cycle Earnings Risk?" National Bureau of Economic Research Working Paper 20913.
- Iacoviello, Matteo.** 2005. "House Prices, Borrowing Constraints, and Monetary Policy in the Business Cycle." *American Economic Review* 95 (3): 739–64.
- Jappelli, Tullio, and Luigi Pistaferri.** 2010. "The Consumption Response to Income Changes." *Annual Review of Economics* 2: 479–506.
- Johnson, David S., Jonathan A. Parker, and Nicole S. Souleles.** 2006. "Household Expenditure and the Income Tax Rebates of 2001." *American Economic Review* 96 (5): 1589–1610.
- Kaplan, Greg, Benjamin Moll, and Giovanni L. Violante.** 2016a. "Monetary Policy According to HANK." National Bureau of Economic Research Working Paper 21897.
- Kaplan, Greg, Benjamin Moll, and Giovanni L. Violante.** 2016b. "A Note on Unconventional Monetary Policy in HANK." Unpublished.
- Kaplan, Greg, Benjamin Moll, and Giovanni L. Violante.** 2018. "Monetary Policy According to HANK: Dataset." *American Economic Review*. <https://doi.org/10.1257/aer.20160042>.
- Kaplan, Greg, and Giovanni L. Violante.** 2014. "A Model of the Consumption Response to Fiscal Stimulus Payments." *Econometrica* 82 (4): 1199–1239.
- Kaplan, Greg, Giovanni L. Violante, and Justin Weidner.** 2014. "The Wealthy Hand-to-Mouth." *Brookings Papers on Economic Activity* 48 (Spring): 77–153.
- Luetticke, Ralph.** 2015. "Transmission of Monetary Policy with Heterogeneity in Household Portfolios." Unpublished.
- McKay, Alisdair, Emi Nakamura, and Jón Steinsson.** 2016. "The Power of Forward Guidance Revisited." *American Economic Review* 106 (10): 3133–58.
- McKay, Alisdair, and Ricardo Reis.** 2016. "The Role of Automatic Stabilizers in the U.S. Business Cycle." *Econometrica* 84 (1): 141–94.
- Mian, Atif, Kamalesh Rao, and Amir Sufi.** 2013. "Household Balance Sheets, Consumption, and the Economic Slump." *Quarterly Journal of Economics* 128 (4): 1687–1726.
- Misra, Kanishka, and Paolo Surico.** 2014. "Consumption, Income Changes, and Heterogeneity: Evidence from Two Fiscal Stimulus Programs." *American Economic Journal: Macroeconomics* 6 (4): 84–106.
- Oh, Hyunseung, and Ricardo Reis.** 2012. "Targeted Transfers and the Fiscal Response to the Great Recession." *Journal of Monetary Economics* 59 (S): S50–64.
- Olivi, A.** 2017. "Sufficient Statistics for Heterogeneous Agent Models." Unpublished.
- Parker, Jonathan A., Nicholas S. Souleles, David S. Johnson, and Robert McClelland.** 2013. "Consumer Spending and the Economic Stimulus Payments of 2008." *American Economic Review* 103 (6): 2530–53.
- Ravn, Morten O., and Vincent Sterk.** 2017. "Job Uncertainty and Deep Recessions." *Journal of Monetary Economics* 90 (C): 125–141.
- Rigobon, Roberto, and Brian Sack.** 2004. "The Impact of Monetary Policy on Asset Prices." *Journal of Monetary Economics* 51 (8): 1553–75.
- Rotemberg, Julio J.** 1982. "Sticky Prices in the United States." *Journal of Political Economy* 90 (6): 1187–1211.

- Schmidt, Lawrence.** 2015. "Climbing and Falling Off the Ladder: Asset Pricing Implications of Labor Market Event Risk." Unpublished.
- Schorfheide, Frank.** 2008. "DSGE Model-Based Estimation of the New Keynesian Phillips Curve." *Economic Quarterly* (Fall): 397–433.
- Smets, Frank, and Rafael Wouters.** 2007. "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach." *American Economic Review* 97 (3): 586–606.
- Sterk, Vincent, and Silvana Tenreyro.** 2015. "The Transmission of Monetary Policy through Redistributions and Durable Purchases." Unpublished.
- Vissing-Jorgensen, Annette.** 2002. "Limited Asset Market Participation and the Elasticity of Intertemporal Substitution." *Journal of Political Economy* 110 (4): 825–53.
- Werning, Iván.** 2015. "Incomplete Markets and Aggregate Demand." National Bureau of Economic Research Working Paper 21448.
- Wong, Arlene.** 2016. "Population Aging and the Transmission of Monetary Policy to Consumption." Unpublished.
- Yogo, Motohiro.** 2004. "Estimating the Elasticity of Intertemporal Substitution When Instruments Are Weak." *Review of Economics and Statistics* 86 (3): 797–810.