

A Catch-22 for HANK Models: No Puzzles, No Amplification^I

Florin O. Bilbiie^{II}

December 2017 (First draft April 2016)^{III}

Abstract

New Keynesian (NK) models with heterogeneous agents (HA) *can* deliver aggregate-demand amplification (monetary-fiscal multipliers and deep recessions) through a New Keynesian Cross—but *only if* hand-to-mouth's income elasticity to aggregate is larger than one. This "hand-to-mouth channel" gives static amplification through the within-the-period elasticity of aggregate demand to policies. A dynamic, complementary "self-insurance channel" magnifies the effects when households are hand-to-mouth only occasionally: the aggregate Euler equation features discounting when the elasticity is lower than one, but *compounding* when larger. The former channel matters most for short-lived shocks, the latter for persistent or future news. Yet amplification has a dark side: the very same condition that delivers it also aggravates the NK troubles (the forward guidance puzzle, neo-Fisherian effects, and the paradox of flexibility) and creates new ones (insufficiency of the Taylor principle and a paradox of thrift). This tension also holds in a quantitative, empirically-relevant DSGE version of the model that I build. Phrased positively: adding HA *can* solve all these NK troubles, *but* that also rules out amplification and multipliers—a Catch-22.

JEL Codes: E21, E31, E40, E44, E50, E52, E58, E60, E62

Keywords: hand-to-mouth; heterogenous agents; aggregate demand; Keynesian cross; monetary policy; fiscal multipliers; liquidity trap; deep recessions; missing deflation; forward guidance; paradox of flexibility; neo-Fisherian effects; Taylor principle; paradox of thrift.

^II am grateful to Jess Benhabib, Edouard Challe, John Cochrane, Daniel Cohen, Davide Debortoli, Axelle Ferrière, Gaetano Gaballo, Jordi Galí, Roger Guesnerie, Olivier Loisel, Virgiliu Midrigan, Benjamin Moll, Emi Nakamura, Salvatore Nistico, Xavier Ragot, Kenneth Rogoff, Gilles Saint-Paul, Jon Steinsson, Paolo Surico, Gianluca Violante, Mirko Wiederholt, Michael Woodford, and participants in several seminars and conferences for useful comments. I gratefully acknowledge without implicating the support of Banque de France via the eponymous Chair at PSE, and of Institut Universitaire de France, as well as the hospitality of New York University and CREI during part of writing this paper.

^{II}Paris School of Economics, UP1 Panthéon Sorbonne, and CEPR; 48 Boulevard Jourdan 75014 Paris. florin.bilbiie@gmail.com. <http://florin.bilbiie.googlepages.com>.

^{III}This paper originates in notes for a keynote lecture at ERMAS 2015 Cluj; it supersedes "The Puzzle, the Power, and the Dark Side: Forward Guidance Redux" that contained an earlier version of a restricted subset of the material. Bilbiie, 2017 contains some overlapping material with parts of Sections 3 and 4 here.

1 Introduction

One decade after the 2008 crisis, the domain of study of macroeconomic stabilization policies seems to witness nothing short of a change of paradigm: the increasing use of heterogeneous-agent models to address monetary and fiscal policy questions while paying attention to concerns for inequality and redistribution.¹ The very nature of these models (their complexity—necessary to address the micro-data evidence—and the mathematical sophistication needed to solve them) also inherently implies a high barrier to entry in this field.

In this paper, I propose a simple version of this class of models that nevertheless contains some of its main ingredients, clarifies some of its key mechanisms, and allows a full-fledged analysis in the "New Keynesian" (NK) tradition *lato sensu*—with pencil and paper. The model can be extended to include other relevant features and frictions to build larger quantitative DSGE models for estimation, and to analyze optimal policy: it can be used to do everything (positive and normative) that researchers have been doing with standard NK models in order inform real policy analysis. As a bonus, it is also useful for teaching and an effective tool for communicating effectively some insights of the more complicated models; as such, I view this "simple model" approach in no way as a substitute for building more complicated, larger and quantitative models, but as complementary and hence as a step towards a true synthesis in this field.

The starting point of this literature is this observation: the baseline New Keynesian model lacks a "Keynesian cross"; the slope of aggregate demand, or the *planned expenditure* (PE) curve is very close to zero—consumption is almost insensitive to current income. This is blatantly in contrast with mounting evidence—reviewed in great detail elsewhere—obtained using a wide spectrum of (micro and macro) data and econometric techniques.² The second relevant observation follows immediately: there is no fiscal multiplier on private spending in the standard, baseline NK model—the multiplier is at most 1;³ yet empirical evidence tends to find a range of output

¹The increasing use of such models went hand in hand with the increasing use of micro data, and has been made possible by technological advances pertaining to both numerical resolution of the models (building on the seminal contributions of Krusell and Smith, 1998 and Den Haan, 1997) and advances in mathematical theory useful for a rigorous analysis of these models (the "mean field games" theory of Lions and Lasry, thoroughly summarized by Achdou et al, 2017). There has also been increased demand for such research by prominent policymakers—Bernanke, Yellen, and Draghi all devoted entire speeches to this issue alone, explicitly calling for more such research

²Some reviews can be found in Mankiw (2000), and Bilbiie and Straub (2012, 2013). Five streams of evidence emerge: (i) direct evidence on zero net worth (i.a. Wolff and Caner, 2004; Bricker et al, 2014); (ii) estimated *zero* elasticity of intertemporal substitution in aggregate consumption Euler equations (Hall, 1988; Campbell and Mankiw, 1989, 1990, 1991; Yogo 2003; Bilbiie and Straub, 2012 and many others); (iii) the sensitivity of consumption to fiscal transfers (Parker, 1999; Johnson, Parker and Souleles 2006); (iv) the recent literature on "wealthy hand to mouth"—a significant share of agents with illiquid wealth behave as H (e.g. Kaplan and Violante 2014; Misra and Surico, 2014; Surico and Trezzi; Cloyne, Ferreira, and Surico 2017); (v) worldwide evidence (World Bank, 2014).

³Theories do exist that deliver higher multipliers, other than the heterogeneous-agent based ones analyzed here. They include complementarity between consumption and hours (Bilbiie, 2011) and Monacelli and Perotti (2008) and fiscalist equilibria with passive monetary policy (Davig and Leeper, 2011). Relatedly, multipliers can be much higher in a liquidity trap: see Eggertsson (2010), Christiano, Eichenbaum, and Rebelo (2011), and Woodford (2011); but those high multipliers rely upon creating high expected inflation.

multipliers between 0.5 and 2, with "local multipliers" (see Nakamura and Steinsson, 2014) at the higher end of the range.⁴ The third observation is also an implication of this lack of Keynesian cross: deep recessions in a liquidity trap in the standard NK model can only occur if accompanied by large deflations; the model thus seems ill suited to describe the post-2008 experience and the "missing disinflation" (Hall, 2011). The final observation is that, by definition, a one-agent model (like the baseline NK model) is a fortiori silent about distributional issues such as those taking center stage in policy debates recently.

We can summarize this as follows: baseline NK models do not deliver "multipliers"—understood as 1. amplification of monetary policy effects through endogenous mechanisms; 2. fiscal multipliers on private spending; and 3. deep recessions in a liquidity trap without necessarily relying on sharp deflation.

A separate (but, I will argue, intimately related) battery of results in the NK realm is this: the standard model is subject to several by now well-known *puzzles and paradoxes*; let us also mention three and come back momentarily to discuss them further: 1. The forward guidance puzzle (Del Negro, Giannoni, and Patterson, 2012) refers to the notion that the further an interest rate cut is pushed into the future, the larger an effect it will have today; 2. neo-Fisherian effects (Benhabib, Schmitt-Grohe, and Uribe, 2002; Cochrane, 2017) refer to the property of the model that under an interest rate peg an increase in interest rates can be inflationary—both in the long and in the short run; and 3. the paradox of flexibility, that in a liquidity trap an increase in price flexibility makes matters worse, i.e. it leads to a larger deflation (Eggertsson and Krugman, 2012).

Heterogeneous-agent New Keynesian models (labelled *HANK* by one of the main references in this literature, Kaplan, Moll and Violante 2015—hereinafter KMV) can and have been used to remedy **both** of these problems: 1. deliver amplification and multipliers;⁵ and 2. resolve *one* of the outstanding NK problems, the FG puzzle. This paper is to the best of my knowledge the first to provide a unifying treatment of amplification of all policies (and other demand) shocks, and address the other puzzles with this class of models.⁶

I first propose a "New Keynesian Cross" for the analysis of heterogeneous-agent models that consists of a Planned Expenditure curve, PE for short (pictured in Figure 1 further below); the *slope* of this curve is the share of the *indirect effect* in total, while the *shift* of the curve (in

⁴See i.a. Blanchard and Perotti (2003), Perotti (2007), Hall (2011), Auerback and Gorodnichenko (2012), Ramey and Zubairy (2017), and Leeper, Traum, and Walker (2017) for a variety of approaches.

⁵KMV focus on monetary amplification, as do Auclert (2016), Werning (2015), Gornemann, Kuester, and Nakajima (2013); Bilbiie and Ragot (2016) study optimal policy in a model with liquidity. Oh and Reis (2012), McKay and Reis (2016), Ferrière and Navarro (2016) and Hagedorn, Manovskii, and Mitman (2017) use HANK-type models to look at different aspects of topic 2: fiscal multipliers. Guerrieri and Lorenzoni (2017) use a model in this class to generate a deep recession in a liquidity trap (in response to deleveraging) without deflation. Challe et al (2016) estimate a model with endogenous unemployment risk; see also Ravn and Sterk (2012) and Den Haan et al (2016).

⁶McKay, Nakamura and Steinsson (2015, 2016) used a HANK-type model to point out the resolution of the FG puzzle; See also Kaplan, Moll and Violante (2017). Other papers looked at the puzzle using this class of models subsequently, see the discussion in text.

response to monetary or fiscal policy changes) is the *direct effect*. I analyze several heterogeneous-agent models through this lens and calculate in closed form these effects and decompositions—starting from the representative-agent NK benchmark whereby, as emphasized by KMV already, the indirect effect is virtually zero, and the fiscal multiplier on consumption is exactly zero.

The analysis then shows that an earlier, 2000s vintage of simpler models (labelled *TANK* by KMV, label that I will embrace here too) contains some of the key mechanisms of the richer HANK models; this draws on a revised and extended version of the earlier analytical TANK framework introduced by Bilbiie (2008) to study aggregate demand amplification and monetary policy (building on the earlier Galí, Lopez-Salido and Valles (2007) quantitative model of fiscal multipliers in a different setup—see below).⁷

This part elucidates a key mechanism that I call "*the hand-to-mouth channel*": the necessary and sufficient condition to have *amplification* ("*multipliers*"), and for *indirect*, general equilibrium effects to be of the essence. This is summarized by one parameter denoted by χ , a useful notation to recall the catch-phrase "the key is χ " and to remember what it stands for: the elasticity of hand-to-mouth, constrained agents income to aggregate income. The TANK model implies a steeper PE curve—much like the old Keynesian cross implies a steep PE curve when the marginal propensity to consume (MPC) increases. This holds true here too: when we add households with unit MPC (out of *their own* income), aggregate MPC increases. The keystone for this mechanism of indirect-effect driven amplification is the "their own" qualification in the previous bracket. For what delivers this amplification, in addition to the mere addition of hand-to-mouth agents, is that their income respond to the cycle *more than one-to-one*: $\chi > 1$. In short, the NK-cross amplification occurs by the interaction of *hand-to-mouth* behavior (liquidity constraints) and an *income distribution* such that hand-to-mouth income rises more than one-to-one with aggregate. The latter requires that there be not too much endogenous redistribution in favor of the hand-to-mouth, i.e. the tax system not be too progressive. This gives a prominent role for income distribution (through labor market and fiscal redistribution) and how it changes over the cycle: just as in Keynes' quotes at the outset. The "hand-to-mouth channel" summarized is thus: amplification through λ occurs if and only if χ is larger than one.

A similar conclusion applies when I extend this to include another key ingredient in HANK models: self-insurance in face of idiosyncratic risk. I label this "simple HANK" the $\lambda\chi\delta$ model—from the three Greek letters denoting its key parameters, of which the last one (δ) is novel with respect to the previous framework.⁸ This δ is the coefficient in front of future consumption in the

⁷Bilbiie and Straub (2012, 2013) present empirical evidence consistent with the "Keynesian" region since the 1980s and with a non-Keynesian region during the Great Inflation. Colciago (2012) and Ascari, Colciago, and Rossi (2015) extend this to the case of sticky wages. Nistico (2016) generalizes this to Markov switching between types and studies financial stability as an objective of monetary policy. Eggertsson and Krugman (2012) use a similar aggregate demand structure but with savers and borrowers (instead of spenders). They show that a deleveraging shock can throw the economy into a liquidity trap; the amplification mechanism emphasized here operates in that framework too.

⁸I am grateful to John Cochrane for suggesting the "three Greek letters" abbreviation.

loglinearized aggregate Euler equation, and depends in a very transparent way on idiosyncratic risk and its interaction with aggregate uncertainty—which is instead summarized by the hand-to-mouth channel described earlier. The conclusion here is that self-insurance magnifies whatever the hand-to-mouth channel gives: when there is dampening ($\chi < 1$), self-insurance implies more dampening because it implies a form of "discounting" in the aggregate Euler equation ($\delta < 1$, a version of which has first been obtained in an incomplete-markets model by McKay, Nakamura, and Steinsson 2016 for the special case where income of the constrained is fixed). But when hand-to-mouth gives amplification, self-insurance magnifies that amplification because it implies the inverse of discounting: *compounding* in the aggregate Euler equation.⁹ The intuition for this compounding is that good news about future aggregate income mean disproportionately more good news in the hand-to-mouth state, and thus dis-saving (less demand for self-insurance). With zero equilibrium savings, today's consumption must go up and income adjusts upwards to deliver this.

The self-insurance and hand-to-mouth channels are *complementary*: the former is the larger, the more the latter is expected to matter (the longer the expected hand-to-mouth spell). The former is thus chiefly important to explain *short-term* issues: short lived shocks and monetary and fiscal policies, for which the TANK and the $\lambda\chi\delta$ (simple HANK) models deliver very similar conclusions. But when it comes to the medium and long run, i.e. to persistent shocks and "news" (such as forward guidance) or other policy announcements, the difference between the two models can become very large: the $\lambda\chi\delta$ model can deliver much more amplification—or dampening, depending on whether χ is larger or smaller than one.

I then build a complete NK model by adding an aggregate supply side, and looking at the determinacy properties of interest rate rules, and at liquidity traps—all in closed form. This is essentially the three-equation, textbook NK model extended to include the above-mentioned channels through three extra parameters $\lambda\chi\delta$, and can be of independent interest to other researchers. The message of this "part 1" is: **yes**, HANK models *can* give "multipliers" under a certain condition that is common for all forms of desirable amplification: monetary, fiscal, and deep demand recessions without deflation. That condition is, for reasons alluded to but that will become clear below: $\chi > 1$.

The simple version of the model used here also allows us to characterize analytically and transparently the necessary and sufficient conditions for NK puzzles to be ruled out in HANK models. I study the power of forward guidance in a liquidity trap and the conditions to solve the FG puzzle; neo-Fisherian effects; and the paradox of flexibility. Alas, the very same condition that is necessary for multipliers also leads to multiplying trouble: **all** these puzzles are *aggravated*

⁹Some version of this amplification has been obtained by Werning (2015) in a related model, with more general income processes: amplification occurs when income risk is countercyclical or liquidity procyclical; I compare and contrast my results to that paper in detail in text. My analysis helps understand the relationship between that novel amplification and the earlier "hand-to-mouth" amplification mechanism.

when $\chi > 1$; worse still, two new puzzles occur: the Taylor principle fails, and the door is open for a version of the paradox of thrift with an inverted AD curve (without the zero bound).

Table 1: *The Amplification-Puzzles Dilemma*

	Amplification	No Amplification
Puzzles	TANK & HANK $\chi > 1$	(RA)NK
No Puzzles		HANK $\chi < 1$ (& $\delta < 1$)

This leaves us with a Catch-22 summarized in Table 1: amplification applies to the effects of policies and shocks, but also to trouble. Or we can take the reverse side of that argument: we can use this new vintage of models to eliminate outstanding NK puzzles and rule out potential new ones—but then we also have to give up hope for amplification in these models. This is itself a puzzle because obtaining endogenous amplification beyond mere intertemporal substitution is one of the main reasons why we want to build these models in the first place. I embed this in an empirically-relevant, state-of-the art DSGE model with several distortions (habits, investment in physical capital subject to adjustment costs, sticky wages, price and wage indexation). I solve this model with heterogeneous agents using nonlinear methods under a binding zero lower bound and illustrate that the same tension identified in the simple model also occurs there: for parameter values that correspond, roughly speaking, to $\chi > 1$ there is amplification (deeper liquidity-trap recessions), but also an aggravation of the FG puzzle. Conversely, with " $\chi < 1$ " there is dampening and the FG puzzle is mitigated; thus, the "catch-22" also applies to a more complicated, quantitative "HANK-DSGE" model. The concluding Section contains several possible avenues for escaping this dilemma, and another possible interpretation of these results.

2 The New Keynesian Cross and Direct-Indirect Effects: the Representative-Agent Benchmark

To set the stage and introduce the key concepts, consider the representative-agent model first. I show in Appendix A using standard intertemporal budget constraint algebra that the "consumption function" for an agent j who takes as given the interest rate and her disposable (net of taxes) income is, loglinearized around a steady-state equilibrium:¹⁰

$$c_t^j = (1 - \beta) \hat{y}_t^j - \sigma \beta r_t + \beta E_t c_{t+1}^j. \quad (1)$$

The other key equation is the Euler equation, or IS curve, obtained by further imposing market

¹⁰The first examples I know of of such loglinearized intertemporal budget constraints leading to consumption functions are Campbell and Mankiw (1989, 1990, 1991) and Gali (1991)—although the idea has an illustrious history the details of which can be found in Cambell and Mankiw, 1989. For other recent uses in different contexts see Garcia-Schmidt and Woodford (2014), Gali (2016), and Farhi and Werning (2017).

clearing—which with a representative agent is also the definition of income, $c_t^j = \hat{y}_t^j \equiv y_t^j - t_t^j$:

$$c_t^j = E_t c_{t+1}^j - \sigma r_t \quad (2)$$

A major theme of the HANK model is the decomposition of the effect of monetary policy proposed by KMV: into a "direct" effect, driven essentially by intertemporal substitution, and an "indirect effect" consisting of the endogenous amplification on output through general-equilibrium effects. KMV show that while in the representative-agent model most of the total effect of monetary policy is driven by the former component, in their HANK model a large portion (as much as 80 percent) is due to the latter.¹¹ Following KMV I compute the total effect, and the decomposition between direct and indirect effects, of an exogenous change in monetary policy summarized by a *decrease* in the real interest rate r_t meant to capture for instance more expansionary monetary policy.¹² I extend this by computing the benchmark *fiscal multiplier*: the effect on total gross income (output) of a one-percent increase in government spending financed by an increase in taxes (balanced-budget), $g_t = t_t^j$. I assume that both shocks have exogenous persistence p and denote by Ω the total effect of an interest rate cut, by $\mathcal{M} \equiv \frac{dy_t^j}{dg_t}$ the output multiplier and by ω the indirect effect share. The effects are:¹³

$$\begin{aligned} \Omega &= \frac{\sigma}{1-p}; \quad \omega = \frac{1-\beta}{1-\beta p} \\ \mathcal{M} &= 1. \end{aligned} \quad (3)$$

The first line is essentially a discrete-time version of KMV's decomposition in the representative-agent model. The total effect, denoted by Ω , is obtained by imposing market clearing, or in other words directly from the Euler-IS equation, as: $\Omega \equiv \frac{dc_t^j}{d(-r_t)}$ which leads to the above expression. The direct effect (Ω_D) is the partial derivative of the consumption function, keeping y_t^j fixed: $\Omega_D \equiv \frac{dc_t^j}{d(-r_t)}|_{y_t^j=\bar{y}} = \frac{\sigma\beta}{1-\beta p}$. Conversely, the indirect effect (Ω_I) is the derivative along the path where $c_t^j = y_t^j$, but the interest rate is kept fixed: $\Omega_I \equiv \frac{dc_t^j}{d(-r_t)}|_{r_t=\bar{r}} = \frac{1-\beta}{1-\beta p} \frac{\sigma}{1-p}$; naturally, this is also given by the difference between total and direct, $\Omega_I = \Omega - \Omega_D$. Finally, the *relative share of the indirect effect* is $\omega \equiv \frac{\Omega_I}{\Omega}$ given above. Notice that as p increases, the indirect effect becomes stronger. The multiplier is 1, which means the multiplier on consumption is nil $\frac{dc_t}{dg_t} = 0$; this is the *best* that the representative agent model can do. Allowing some prices to be flexible leads

¹¹Auclert (2016) performs a different but related decomposition into three channels that account for households' financial positions.

¹²Below, I solve the complete forward-looking model with a Phillips curve and Taylor rule. Here, I abstract from the exact equilibrium mechanism by which the real interest rate is determined. An exogenous interest rate corresponds to the case of a Taylor rule that neutralizes expected inflation; or to the case of fixed prices—an interpretation that I used before, e.g. in Bilbiie (2011).

¹³Since there is no endogenous state variable (no endogenous persistence) p is obviously also the persistence of any endogenous variable. The fiscal multiplier is computed for the case of fixed- r_t , for example with fixed prices (drawing on Bilbiie, 2011, in a different context).

to a multiplier lower than one, because future inflation leads to an increase in the real interest rate and intertemporal substitution towards tomorrow, so lower consumption today. We will see a manifestation of this in the more general model below; notice that consumption-hours complementarity is necessary and sufficient to deliver positive multipliers in this fixed-price case (Bilbiie, 2011, Proposition 3).

A useful benchmark is that of iid shocks—which allows to abstract from the effects of persistence and use these concepts in order to gauge endogenous amplification. When $p = 0$, the total effect of MP is $\Omega = \sigma$, and the indirect share $\omega = 1 - \beta$; this is the first result emphasized by KMV: with discount rate close to 1, the indirect effect is almost absent in the representative-agent NK model.

Consider then the following picture, a familiar-looking Keynesian cross. The key equation throughout the paper is the one delivering the upward sloping line labelled PE: like the *planned expenditure line* from the standard textbook ("old") Keynesian cross diagram, it expresses consumption (aggregate demand) as a function of current income, for a given real interest rate (and for arbitrary persistence of shocks p), as follows:

$$c_t = \omega \hat{y}_t - (1 - \omega) \Omega r_t + (1 - \omega) (\mathcal{M} - 1) g_t, \quad (4)$$

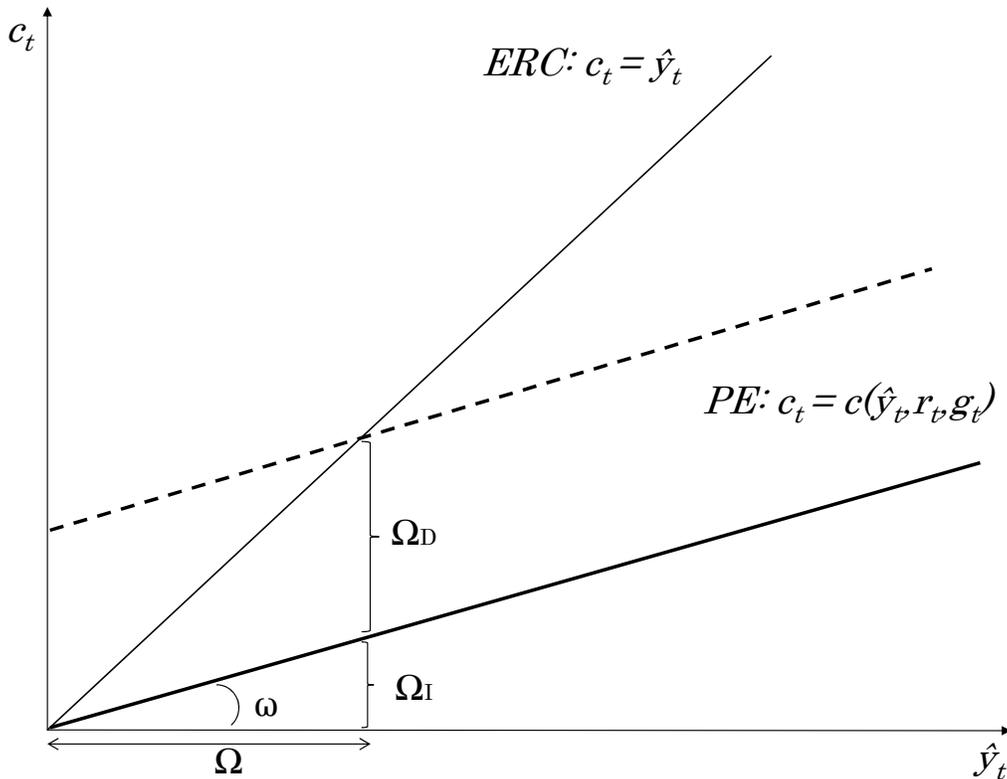


Figure 1: The New Keynesian Cross: Indirect and Direct Effects of Monetary Policy

I will show that not only in the baseline RANK model just studied but also in several heterogeneous-agent models reducible to this form, Ω is generally the *total effect* of an interest rate change on aggregate demand, while the *slope* ω captures the *share* of what KMV call the *indirect effect* in total Ω . The *shift* of the PE curve will hence be the *direct effect* $(1 - \omega)\Omega$. While \mathcal{M} will be the "fiscal multiplier": the increase in output following an increase in public spending. A cut in interest rates translates the PE curve upwards (by $(1 - \omega)\Omega$) and the equilibrium moves from the origin to the intersection of the dashed PE curve and the 45 degree line. Similarly, an increase in government spending, insofar as it has a positive multiplier on consumption, shifts the PE curve up. The rest of the paper is devoted to the analysis of the key objects ω , Ω , and \mathcal{M} , their key determinants in a series of two-agent models, and the implications for several puzzles and paradoxes for the NK model.

The results for the RANK model are naturally interpreted through the lens of this NK cross: because the slope of PE is very close to zero, almost all the effect of monetary policy comes from the direct shift of the PE curve. In addition, the fiscal multiplier is one—there is no effect on private spending. We can conclude that there is very little Keynesian about the representative-agent NK model. We now move to a model that has a very Keynesian flavor.

3 TANK: A Keynesian Model with Amplification and (through) Indirect Effect

The earlier vintage analyzing heterogeneous agents and aggregate demand in sticky-price models used a simpler, *two-agent* setup (*TANK*) where a fraction of households are "hand-to-mouth": they consume all their disposable income. This section consists of revisiting and extending that framework in light of the new HANK literature. The exposition here draws heavily on Bilbiie (2008)—the reader can consult that paper for a slightly different setup where the same key mechanism occurs, and for complementary discussion.

Households There are two types: one class of households with total mass λ is *excluded* from asset markets (and hence has no Euler equation) but *does* participate in labor markets and makes an optimal labor supply decision (their income is therefore *endogenous*); these are the "hand-to-mouth" H . The rest of the agents $1 - \lambda$ also work and trade a full set of state-contingent securities, including shares in monopolistically competitive firms (thus receiving their profits from the assets that they price).¹⁴

¹⁴Galí, Lopez-Salido and Valles (2007) focused on government spending multipliers and were the first to study a sticky-price model with two agents but using a different setup: the distinction between optimizing and rule-of-thumbs comes from access (or lack thereof) to *physical capital*. Bilbiie (2008) studied *monetary policy* by modelling the distinction between agents based on participation in asset markets (savers hold shares in firms) and the *distribution of profits*. The novel element afforded by this simplification was an analytical, closed-form expression for the

Savers' dynamic problem is exactly as outlined in Appendix A replacing j with S and recognizing that in equilibrium their portfolio of shares is now (by stock-market clearing) $(1 - \lambda)^{-1}$. The budget constraint of H is $C_t^H = W_t N_t^H + Transfer_t^H$ where C is consumption, w the real wage, N^H hours worked and $Transfer_t^H$ net fiscal transfers to be spelled out as government policies below.

All agents maximize present discounted utility, defined as separable over consumption and hours and satisfying the usual Inada conditions, subject to the budget constraints. Utility maximization over hours worked delivers the standard intratemporal optimality condition for each agent $j = S, H : U_C^j(C_t^j) = W_t U_N^j(N_t^j)$. Denoting $\sigma^{-1} \equiv -(U_{CC}^j C^j) / U_C^j$ as relative risk aversion and $\varphi \equiv (U_{NN}^j N^j) / U_N^j$ the inverse elasticity of labor supply, and letting small letters denote log-deviations from a steady-state (to be discussed below), we obtain the constant-consumption labor supply for each agent $\varphi n_t^j = w_t - \sigma^{-1} c_t^j$. Assuming for tractability that elasticities are identical across agents, the same equation also holds on aggregate with the same elasticity and income effect,

$$\varphi n_t = w_t - \sigma^{-1} c_t.$$

The Euler equation of S (the only households who do have one, because they participate in all asset markets and get the payoffs from holding such assets) is, for nominal assets with gross nominal return I_t and denoting gross inflation by Π_{t+1} : $I_t U_C(C_t^S) = \beta E_t [U_C(C_{t+1}^S) / \Pi_{t+1}]$. Loglinearized (and denoting by small letters absolute deviations when they pertain to rates of return or inflation), this is the standard

$$c_t^S = E_t c_{t+1}^S - \sigma r_t,$$

with r_t the ex-ante real interest rate $r_t \equiv i_t - E_t \pi_{t+1}$.

Firms The supply side of the model is completely standard: All households (regardless of whether S or H) consume an aggregate basket of individual goods $k \in [0, 1]$, with constant elasticity of substitution ε : $C_t = \left(\int_0^1 C_t(k)^{(\varepsilon-1)/\varepsilon} dk \right)^{\varepsilon/(\varepsilon-1)}$, $\varepsilon > 1$, and so does the government G_t (specified below). Total demand for each good is $C_t(k) + G_t(k) = (P_t(k) / P_t)^{-\varepsilon} (C_t + G_t)$, where $P_t(k) / P_t$ is good k 's relative price and $C_t + G_t$ aggregate consumption. The aggregate price index is $P_t^{1-\varepsilon} = \int_0^1 P_t(k)^{1-\varepsilon} dk$. Each individual good is produced by a monopolistic firm using a linear technology: $Y_t(k) = N_t(k)$. Cost minimization taking the wage as given, implies that real marginal cost is W_t .

The profit function in real terms is given by: $D_t(k) = (1 + \tau^S) [P_t(k) / P_t] Y_t(k) - W_t N_t(k) - T_t^F$. I assume that the government (whose full set of policies is described in detail below) implements the optimal subsidy that is standard in the NK literature, eliminating the markup distortion in

aggregate Euler-IS curve, which clarified aggregate demand amplification with hand-to-mouth. The paper also derived *optimal* monetary policy in this framework (a three-equation NK model with a different IS curve), and studied the determinacy properties of interest rate rules.

steady-state; the pricing condition under flexible prices (denoted with a star) gives the desired markup: $P_t^*(k)/P_t^* = 1 = \varepsilon W_t^* / [(1 + \tau^S)(\varepsilon - 1)]$ and the value of the optimal subsidy (inducing marginal cost pricing) is $\tau^S = (\varepsilon - 1)^{-1}$. Financing the total cost of this by taxing firms lump-sum ($T_t^F = \tau^S Y_t$) gives total profits $D_t = Y_t - W_t N_t$; notice that this policy is redistributive because it is as if it was taxing savers—the holders of firm shares. Indeed, under this policy profits (and hence dividend income) are zero in steady state $D = 0$ which generates the "full-insurance" steady-state used here $C^H = C^S = C$ for tractability. Loglinearizing around this, profits vary inversely with the real wage:¹⁵

$$d_t = -w_t,$$

where d_t is expressed as a share of steady-state output.

We leave arbitrary for now how firms set their prices but return to this below when adding an aggregate supply side. For now, notice that for several price schemes with nominal rigidity, the aggregate production function of this economy loglinearized around a steady state with zero inflation delivers simply, as a first order approximation: $y_t = n_t$.¹⁶

Government. The government does fiscal and monetary policy. The former is made of three dimensions: first, the optimal sales subsidy classic in NK models but with a twist described above: it is financed by taxing firms (and thus, implicitly, savers) which further generates zero steady-state profits and *full insurance* in steady state. The second fiscal policy is an *endogenous redistribution* scheme that is simple and transparent (borrowed from Bilbiie, 2008, Section 4.3, which first analyzed the link between the redistribution of profits and Keynesian logic): taxing profits at rate τ^D and rebating the proceedings in a lump-sum fashion to H ; the Appendix outlines a more general redistribution, see also the discussion below. These dimensions govern automatic stabilizers in this economy, as will become clear below.

The third dimension of fiscal policy consists of discretionary changes in government spending: I assume thus that the government spends an exogenous amount G_t every period, whose steady-state value is normalized to zero for simplicity ($G = 0$). This wasteful spending is financed by lump-sum taxes on agents, where each agent pays T_t^j ; to analyze the role of this *exogenous redistribution*, I assume that H agents pay on aggregate an arbitrary share of total tax revenues, i.e. $\lambda T_t^H = \alpha T_t$ (evidently, $(1 - \lambda) T_t^S = (1 - \alpha) T_t$).¹⁷ We thus have in percentage deviations (as

¹⁵This is not strictly necessary for any of the results but it simplifies the algebra.

¹⁶Around zero inflation, relative price dispersion (if it were Calvo) or the price adjustment cost (if it were Rotemberg) are second-order and do not matter for loglinearized dynamics.

¹⁷Monacelli and Perotti (2012) also study the link between this type of redistribution and the multiplier in a borrower-saver model. Bilbiie, Monacelli, and Perotti (2013) study the effects of pure redistribution and the link with public debt.

share of steady-state output Y , e.g. $t_{H,t} \approx T_t^H/Y$ and $g_t \approx G_t/Y$):

$$\begin{aligned} t_{H,t} &= \frac{\alpha}{\lambda} t_t = \frac{\alpha}{\lambda} g_t \\ &= \underbrace{g_t}_{\text{balanced-budget}} - \underbrace{\left(1 - \frac{\alpha}{\lambda}\right) t_t}_{\text{exog. redistrib.}} \end{aligned}$$

where the second line decomposes the tax on H into a uniform tax corresponding to the balanced-budget G increase and a term capturing the exogenous redistribution: a transfer to H whenever $\alpha < \lambda$, and a transfer from H otherwise. This captures in a crude way (with two agents) the notion of progressivity of the tax changes used to finance spending. Ferrière and Navarro (2017) provide empirical evidence that in the US G increases are indeed accompanied by changes in tax progressivity.¹⁸ Notice thus that there is a simple distinction between the systematic, *endogenous redistribution* captured by τ^D/λ (a proxy for automatic stabilizers), and the *exogenous redistribution* captured by α/λ —the implicit transfer. The total transfer to an H agent will be under this policy $Transfer_t^H = \frac{\tau^D}{\lambda} D_t - T_t^H = \frac{\tau^D}{\lambda} D_t - \frac{\alpha}{\lambda} T_t$, and is zero by construction in steady state.

Market clearing and equilibrium Labor and goods market clearing in turn imply $\lambda C_t^H + (1 - \lambda) C_t^S = C_t$ and $\lambda N_t^H + (1 - \lambda) N_t^S = N_t$. Assume further (by normalization) that preferences are such that both agents work the same hours in steady state $N^H = N^S = N$. Since fiscal policy also makes consumption equal across agents in steady-state $C^H = C^S = C$; loglinearization of the aggregation equations delivers $c_t = \lambda c_t^H + (1 - \lambda) c_t^S$ and $n_t = \lambda n_t^H + (1 - \lambda) n_t^S$.

All output is consumed by either private consumers or the government, and is produced only by labor with constant returns. Approximated around the zero- G steady-state, we have the economy resource constraint

$$c_t + g_t = y_t = n_t.$$

3.1 Aggregate Euler-IS and Aggregate PE Curve

Combining the economy resource constraint cum production function with the aggregate labor supply delivers the equilibrium wage schedule: $w_t = (\varphi + \sigma^{-1}) c_t + \varphi g_t$.

Hand-to-mouth thus consume all *their* after-tax income $c_t^H = \hat{y}_t^H$, and the key word is "*their*": for while their consumption comoves one-to-one with their income, it comoves *more or less than one-to-one* with *aggregate* income—and this is the keystone for dynamics in this model. To understand this, start from the loglinearized budget constraint of H: $c_t^H = \hat{y}_t^H = w_t + n_t^H + \frac{\tau^D}{\lambda} d_t - t_t^H$ combined with their labor supply and with the equilibrium wage schedule, to obtain

¹⁸That paper builds a heterogeneous-agent model with a general tax scheme capturing progressivity (based on Heathcote, Storesletten, and Violante) in a finer way. I do not use that here because I want to have different progressivity in steady state (maintaining the assumption of full insurance that simplifies the algebra) which is useful for maintaining the distinction between the endogenous redistribution through χ and the exogenous redistribution through α .

the consumption function of H:

$$c_t^H = \hat{y}_t^H = \chi \hat{y}_t + \zeta_H (\chi g_t - t_{H,t}) \quad (5)$$

where $\chi \equiv 1 + \varphi \left(1 - \frac{\tau^D}{\lambda}\right) \leq 1$

will be the *key parameter* throughout the paper: it denotes the elasticity of H agents' consumption (income) to aggregate disposable income $\hat{y}_t \equiv y_t - t_t$.

As will become clear momentarily, χ governs amplification effects of monetary policy and fiscal multipliers in this model, and depends chiefly on fiscal redistribution and labor market characteristics. It is also the main distinguishing feature of my model from earlier analyses such as Campbell and Mankiw (1989) and the literature that followed it—where it is assumed that hand-to-mouth (or, in their terminology, "rule-of-thumb") agents consume a fraction (< 1) of *aggregate* income; the implications of this are discussed in more detail below. The other new parameter is the elasticity of H consumption to a transfer $\zeta_H \equiv (1 + \varphi^{-1}\sigma^{-1})^{-1}$; this governs the strength of the income effect relative to substitution: it is 0 when labor supply is infinitely elastic and 1 (largest) when it is inelastic, or when the income effect σ^{-1} is nil.

For S agents, we need to worry about distributional effects. The after-tax income of savers is: $\hat{y}_t^S = w_t + n_t^S + \frac{1-\tau^D}{1-\lambda}d_t - t_t^S$, recognizing that they hold all the shares and thus internalizing the effect of profit income. Savers S face an extra income effect of the real wage (which for them counts as marginal cost and reduces profits) that is the keystone for monetary and fiscal transmission. Replacing $d_t = -w_t$ and S agents' labor supply schedule into their income definition, we obtain:

$$c_t^S = \frac{1 - \lambda\chi}{1 - \lambda} \hat{y}_t + \zeta_H \left(\frac{1 - \lambda\chi}{1 - \lambda} g_t - t_t^S \right). \quad (6)$$

The additional negative income effect of wages captures the externality imposed by H agents on S agents: when demand goes up, the real wage goes up (because prices are sticky), income of H agents goes up, and so does their demand—crucially, more than proportionally to aggregate when $\chi > 1$. Total demand goes up, thus amplifying the initial expansion; S agents "pay" for this by working more, which is an equilibrium outcome because their income goes down as profits fall (marginal cost goes up and, insofar as labor is not perfectly elastic $\varphi > 0$, sales do not increase by as much). By this intuition, income of savers is less procyclical the more H agents there are—and the more so, the more inelastic is labor supply; while clearly, redistributing profits from *S* to *H* dampens this channel.

Using the consumption function for savers, which is of the form (1) with $j = S$, we are in a

position to write the *aggregate* consumption function:

$$c_t = [1 - \beta(1 - \lambda\chi)] \hat{y}_t - (1 - \lambda)\beta\sigma r_t + \beta(1 - \lambda\chi) E_t c_{t+1} + \beta\lambda\zeta_H \left(\chi - \frac{\alpha}{\lambda}\right) (g_t - E_t g_{t+1}) \quad (7)$$

It is important to notice that this is *very different* from the equation obtained by Campbell and Mankiw (1989, 1990, 1991) in their model with savers and spenders (and no fiscal policy). The spenders in their model consume a constant fraction of *aggregate* income; this is equivalent to assuming, in my model, that $\chi = 1$, either because labor is infinitely elastic or because fiscal redistribution perfectly insures agents (see below).¹⁹

The aggregate Euler-IS equation is obtained by imposing good market clearing $c_t = y_t$ in (7):

$$c_t = E_t c_{t+1} - \frac{1 - \lambda}{1 - \lambda\chi} \sigma r_t + \frac{\lambda\zeta_H}{1 - \lambda\chi} \left(\chi - \frac{\alpha}{\lambda}\right) (g_t - E_t g_{t+1}). \quad (8)$$

The aggregate elasticity of intertemporal substitution—the elasticity of aggregate demand to interest rates—is *increasing* with the share of H agents, as long as $\lambda < \chi^{-1} < 1$. The reason is the Keynesian spiral already emphasized above: an interest rate cut implies an aggregate demand expansion, through intertemporal substitution of S agents; with sticky prices, this translates into a labor demand shift, which increases the real wage. Since the wage is the H agents' income, this increases their demand further, which amplifies the initial demand expansion. This is an equilibrium: the extra output is optimally produced by S agents, who face a negative income effect coming from profits (recalling that the real wage is marginal cost).²⁰

3.2 Amplification, Dampening, and Redistribution: The Key is χ

The intuition outlined above suggests that this model can, but *need not*, deliver amplification and multipliers—meaning that both the *total effect* of an interest rate cut and of fiscal stimulus are increasing with the share of H agents. In terms of the NK cross in Figure 1, a higher share of H agents λ or a higher elasticity of their income to aggregate income χ increase the slope of the PE curve—just like an increase in the marginal propensity to consume does in old-Keynesian models—thus leading to higher multipliers for both monetary and fiscal policy. The following proposition spells out the exact expressions for the total effect and indirect share in this model,

¹⁹In their latest paper, Campbell and Mankiw (1991) do acknowledge, in a different context (of utility costs from following rule-of-thumb behavior—see their footnote 26), that under the assumption that spenders consume their own income the model behaves differently; That is the only mention of this alternative assumption (maintained throughout this paper) that is crucial for the amplification emphasized here. See Bilbiie and Straub (2012, 2013) for the implications of this different assumption for empirical estimates of λ .

²⁰Equation (8) and this amplification mechanism are analyzed for the first time in Bilbiie (2008). As λ increases the amplification gets larger and larger: when $\lambda > \chi^{-1}$ an expansion cannot be an equilibrium any longer, as the income effect on S agents starts dominating. We return to that "non-Keynesian" region (analyzed in detail in the paper cited above) in Section 6.4.

and describes the necessary and sufficient condition for amplification.

Proposition 1 *Monetary and Fiscal Multipliers and Indirect Effects in TANK.* *The total effect of an interest rate cut, the output multiplier of government spending, and the indirect effect share are:*

$$\begin{aligned}\Omega &= \frac{\sigma}{1-p} \frac{1-\lambda}{1-\lambda\chi}, \\ \mathcal{M} &= 1 + \frac{\lambda\zeta_H}{1-\lambda\chi} \left(\chi - \frac{\alpha}{\lambda} \right), \\ \omega &= \frac{1-\beta(1-\lambda\chi)}{1-\beta p(1-\lambda\chi)},\end{aligned}$$

where p is exogenous persistence common to both r_t and g_t .

The condition that the elasticity of hand-to-mouth income to aggregate be greater than unity

$$\chi > 1 \tag{9}$$

is necessary and sufficient ("iff") for: **monetary amplification** ($\frac{\partial\Omega}{\partial\lambda} > 0$), a **fiscal multiplier** $\mathcal{M} > 1$, and **fiscal amplification** ($\frac{\partial\mathcal{M}}{\partial\lambda}$) for stimulus financed with uniform taxes $\alpha = \lambda$.

The indirect effect share ω is increasing with the share of hand-to-mouth λ regardless of χ ; in particular $\frac{\partial\omega}{\partial\lambda} > 0$ even when there is dampening $\frac{\partial\Omega}{\partial\lambda} < 0$ ($\chi < 1$).²¹

For arbitrary exogenous redistribution α , the multiplier is an increasing function of χ , λ , and of the transfer ($-\alpha$).

Proposition 1 provides a complete picture of the landscape of monetary and fiscal amplification in the TANK model. This has a distinctly Keynesian flavor, in particular when we recall Keynes' view that the marginal propensity to consume (the equivalent of ω in our framework) depends on income distribution, as clear in the quotes from the General Theory provided at the outset. The point that income distribution matters is a general one, as captured through χ . We now discuss the monetary and fiscal multipliers in turn.

The total effect Ω and the indirect share ω are all (potentially very) large in the TANK model, and *increasing* with the share of constrained agents iff $\chi > 1$ (and decreasing otherwise). Notice that the total effect Ω under zero persistence is nothing else than the within-period *aggregate demand elasticity to interest rates*, or an aggregate measure of the elasticity of intertemporal substitution: it is increasing with λ and χ as long as $\lambda < \chi^{-1} < 1$ and decreasing with λ when $\chi < 1$. The explanation lies in the NK cross logic emphasized above: the slope of the PE curve is increasing with λ and χ . The indirect effect share is also increasing with both λ and χ ; the direct and indirect effects are given by $\Omega_D = (1-\omega)\Omega$ and $\Omega_I = \omega\Omega$ respectively. Notice that the direct

²¹The derivatives are $\frac{\partial\Omega}{\partial\lambda} = \frac{\sigma}{1-p} \frac{\chi-1}{(1-\lambda\chi)^2}$ and $\frac{\partial\omega}{\partial\lambda} = \frac{\beta\chi(1-p)}{(1-\beta p(1-\lambda\chi))^2}$.

effect *decreases* with λ —even though the total effect is increasing, the share ω increases faster. In other words, as λ increases the PE curve gets steeper and steeper.

These effects depend crucially on χ , so we turn to its role and key determinants. The setup here is simple but still captures a key insight that is likely to hold in any heterogeneous agent model: this key parameter (the elasticity of hand-to-mouth, constrained agents' income to aggregate income) depends on the income distribution through the labor market structure and fiscal redistribution. χ is lower than $1 + \varphi$ but higher than 1 inasmuch as $\tau^D < \lambda$; in other words, if there is not too much redistribution, amplification still occurs. When $\tau^D = \lambda$, all the endogenous redistributive effects emphasized here are undone, and the economy is back to the perfect-insurance, representative-agent case. Finally, when H get a share of profits higher than their share in the population $\tau^D > \lambda$ (an example of progressive taxation) we have $\chi < 1$ and there is *dampening* instead of amplification; indeed, in the limit when $\chi = 0$ the total effect is scaled down by $1 - \lambda$ and the indirect share is the same as in the RANK model. Notice that as emphasized in the Corollary, while the total effect is decreasing in λ when $\chi < 1$, the indirect effect is still increasing in λ : monetary policy "works" less, but it does so disproportionately more through the general equilibrium response of H agents' income, made of labor income and fiscal redistribution. All these effects are illustrated in Figure 2 which plots the total effect and indirect share for the TANK model under the two different assumptions concerning χ (> 1 and < 1) and distinguishing transitory and persistent policy changes.

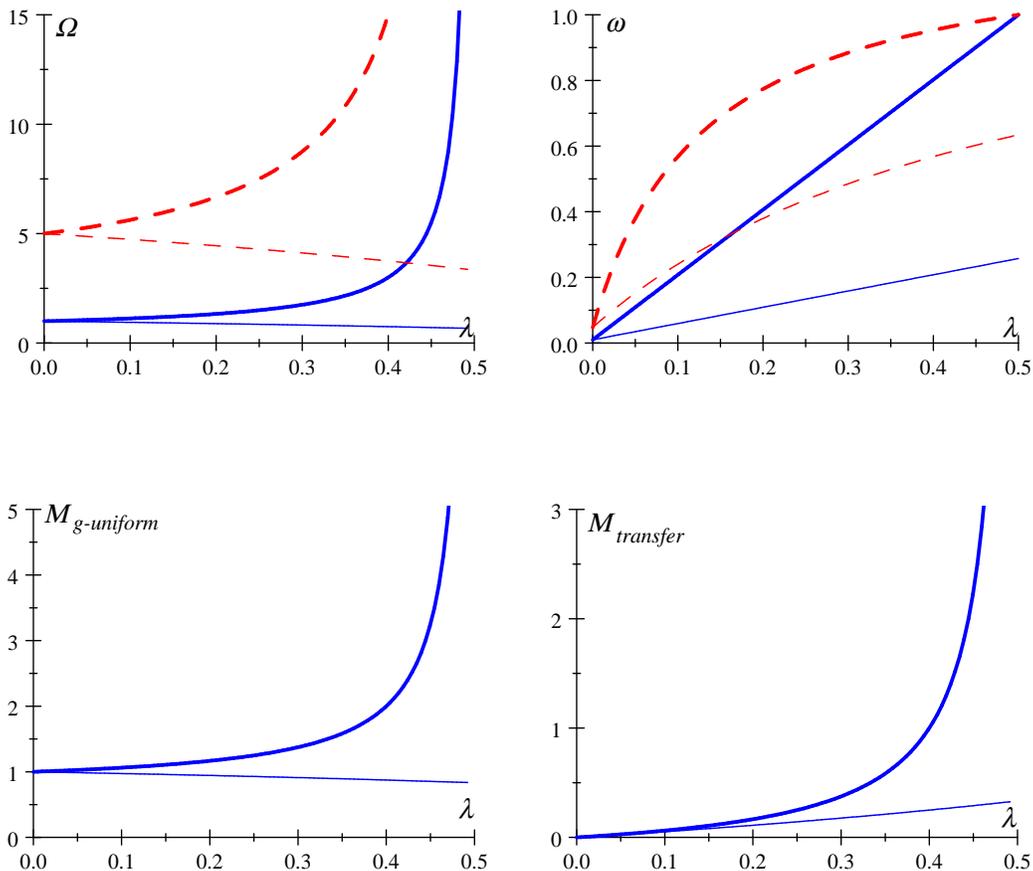


Fig 2: $\chi = 2$ (thick), 0.5 (thin), $p = 0$ (solid) and 0.8 (dash).

3.3 Fiscal Multipliers

Now to fiscal multipliers, a topic that has of course been extensively studied in TANK models, starting with GLV's (2007) seminal contribution. That paper showed for the first time, in a version of this model with physical capital and solved numerically, that government spending may have a positive multiplier on private consumption if enough agents are H (and, in their quantitative model, the labor market is non-Walrasian and spending is deficit financing). As the analytical approach here allows us to illustrate clearly, those additional ingredients are not necessary to obtain multipliers.²² Furthermore, the simple closed-form solutions for the multipliers (decomposing the

²²Other work extended GLV: Bilbiie and Straub (2004) analyzed a model with distortionary taxation and competitive labor market. Bilbiie Meier, and Mueller (2008) showed that an estimated version of the GLV model can reproduce the decrease in consumption multipliers during the Great Moderation period through a combination of more widespread participation in financial markets (lower hand-to-mouth share) and more active monetary policy. Monacelli and Perotti (2012) studied the role of redistribution for the multiplier in a borrower-saver model, and Bilbiie, Monacelli, and Perotti (2013) at the effects of public debt and redistribution (transfers)—see also Mehrotra (2017) and Giambattista and Pennings (2017).

roles of pure spending and redistribution) make the intuition very transparent and pave the way for linking this to the NK puzzles.

The fiscal multiplier on output can be written as:

$$\mathcal{M} = 1 + \frac{\lambda \zeta_H}{1 - \lambda \chi} \left[\underbrace{(\chi - 1)}_{\text{NK cross}} + \underbrace{\left(1 - \frac{\alpha}{\lambda}\right)}_{\text{exog. redistrib.}} \right]$$

The first component inside the square bracket captures the limit under uniform taxation $\alpha = \lambda$ and is due to the NK cross mechanism: the demand effect of G combined with the *endogenous redistribution* through profits. Higher demand for goods leads to an expansion of labor demand, higher wage, and higher consumption for H agents, which amplifies the initial demand expansion, and so on. Equilibrium is reached when the negative income effect on S through profits is just enough to make them willing to work to produce this extra demand. The condition for this mechanism to lead to a multiplier larger than one is (9) i.e. that "automatic stabilizers" (the endogenous, automatic redistribution of profits through the tax system) not be too strong. Evidently, when this condition fails there is instead a negative multiplier on consumption $\mathcal{M} < 1$.

The second component is due to the transfer: *exogenous redistribution* ($\alpha < \lambda$) operates even if endogenous redistribution makes the first mechanism inoperative (the Campbell-Mankiw $\chi = 1$ benchmark). A transfer works through pretty much the same mechanism as described above: being a direct income transfer to H it leads to an increase in demand, a shift in labor demand, higher wage, and so on.

Both effects (and thus the multiplier) are proportional to the term outside the square bracket: the income effect on H ζ_H times their share λ , multiplied by the "Keynesian multiplier" encountered everywhere in this model, coming for the NK cross: $(1 - \lambda \chi)^{-1}$. Indeed, the effect of an increase in public spending can be represented graphically using the NK cross Figure 1 in exactly the same way as a monetary shock but replacing Ω with \mathcal{M} . The fiscal multiplier \mathcal{M} is then the "total effect", and the same decomposition as before applies (note: persistence is irrelevant). The multiplier is increasing with χ because this increases the PE slope (indirect effect) and also increases the PE shift (direct effect)—the latter *only if* (9) holds: $\chi > 1$. It is increasing with the implicit transfer (decreases with α) which increases the PE shift (direct effect)—*only if this is indeed a transfer* (a progressive taxation shock) $\alpha < \lambda$. Finally, the multiplier increases with λ at given α ; but if taxation is uniform, the multiplier is increasing with λ if and only if, again, $\chi > 1$.

Another useful decomposition of the multiplier is additive, between the multiplier of a spending increase financed with uniform taxation $\mathcal{M}_{g\text{-uniform}} = 1 + \zeta_H \frac{\lambda(\chi-1)}{1-\lambda\chi}$, and the multiplier of a pure redistribution (a transfer from S to H) denoted $\mathcal{M}_{transfer} = \frac{\lambda \zeta_H}{1 - \lambda \chi}$:

$$\mathcal{M} = \mathcal{M}_{g\text{-uniform}} + \left(1 - \frac{\alpha}{\lambda}\right) \mathcal{M}_{transfer} \quad (10)$$

These components are plotted in the second row of Figure 2, respectively; $\mathcal{M}_{g-uniform}$ is increasing with λ in the $\chi > 1$ case and decreasing otherwise. In contrast, the transfer multiplier is increasing with λ even in the "dampening" $\chi < 1$ case, albeit at a smaller rate. Fiscal stimulus in the form of transfers (or public debt, which is equivalent here to a pure redistribution, see Bilbiie et al, 2013 for a formalization of the argument) is thus the one policy instrument that can stimulate the economy in the "dampening" case.

3.4 Indirect Amplification

An important observation is that, when the elasticity of H' income to the cycle is *higher than one*, the indirect effect is potentially *much larger* than in the RA model, even at small λ . Take for example a purely transitory interest rate shock, so that in the representative-agent model the indirect effect share is $1 - \beta$; with KMV's calibration ($\beta = 0.95$), this indirect share is merely 0.05, while the calibrated HANK model gives an indirect share of 0.8. What is the TANK model's indirect effect share? The key point is that it is *not only* proportional to the share of hand-to-mouth in the population λ .²³ Instead, as I have just shown, the indirect share is also proportional to their *income's elasticity to aggregate income*: $\omega = 1 - \beta(1 - \lambda\chi)$; thus, the TANK model delivers KMV's HANK model's $\omega = 0.8$ for (at $\tau^D = 0$): $\lambda = 0.4$ if labor elasticity is 1; for $\lambda = 0.26$ if $\varphi = 2$; and for a mere $\lambda = 0.13$ if $\varphi = 5$. Figure 3 illustrates this further by plotting a "HANK surface": the combination of λ and φ that delivers $\omega = 0.8$ for given τ^D . The solid line is under no fiscal redistribution (the largest amplification) and the dotted line for $\tau = 0.5$; for reasons by now clear, more redistribution of the type that makes H agents' income less cyclical implies a lower indirect effect share (and hence a higher λ at given φ to get the same indirect share). In the limit as $\tau^D = \lambda$ we have the Campbell-Mankiw benchmark $\chi = 1$ and λ is invariant to φ and roughly equal to ω (plotted with dots). If the redistribution is stronger, the implied λ necessary for a given ω becomes an increasing function of φ .

²³This is true with Campbell and Mankiw's assumption that hand-to-mouth consume a proportional share of aggregate income, but not in the TANK model where they consume *their* income.

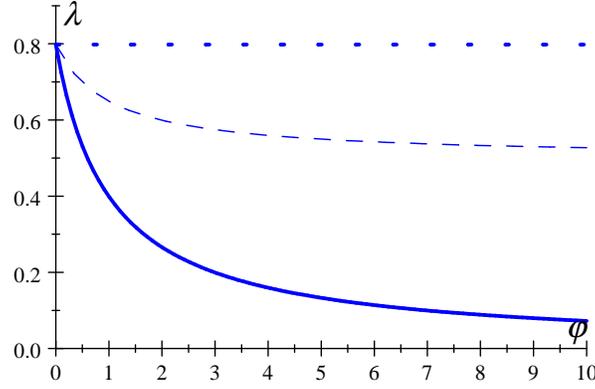


Fig. 3: (φ, λ) s.t. $\omega = 0.8$, $\tau^D = \lambda$ (dots: Campbell-Mankiw); 0.5 (thin); 0 (thick)

A related way to understand the inherently indirect-effect-driven amplification of TANK models is emphasized in the following Corollary.

Corollary 1 "Indirect amplification". *If a TANK model gives \mathcal{A} times higher total effect than the RANK model (i.e. **amplification**), $\Omega(\lambda) = \mathcal{A} * \Omega(0)$, then the **indirect** share is at least (for iid shocks):*

$$\omega \geq 1 - \frac{1}{\mathcal{A}}.$$

In other words, if the total effect of a TANK model is *twice* as much as that of a RANK model, at least *half* of it is indirect;²⁴ if it is four times, then at least three quarters is indirect, and so on. Note that the above is a *lower bound*, and is invariant to λ and χ . The proof is immediate: with iid shocks the ratio of the two total effects is $\mathcal{A} = \frac{1-\lambda}{1-\lambda\chi}$. Replacing in the indirect share we have $\omega = 1 - \beta \frac{1-\lambda}{\mathcal{A}} > 1 - \frac{1-\lambda}{\mathcal{A}} \geq 1 - \frac{1}{\mathcal{A}}$.²⁵

3.5 Other Determinants of χ : Fiscal Redistribution and Sticky Wages

The intuition for these redistributive effects through both the labor market and fiscal policy in shaping the marginal propensity to consume, and hence the aggregate effects of policies, can be traced back to Keynes' General Theory.²⁶ The above outlined a simple fiscal policy that

²⁴I am grateful to Davide Debortoli who suggested this interpretation for the useful special case $M = 2$.

²⁵For persistent shocks, the lower bound is $\omega \geq (1 - \frac{1}{\mathcal{A}}) / (1 - p\frac{1}{\mathcal{A}})$.

²⁶To keep with the catchphrases, one could recall this as " χ as in Keynes" (1936, Ch. 8, Books I and III, emphasis added): "The amount that the community spends on *consumption* obviously depends on [...] the principles on which *income is divided* between the individuals composing it (which *may suffer modification as output is increased*)." "[...] we may have to make an allowance for the possible *reactions of aggregate consumption* to the *change in the distribution* of a given real *income* between *entrepreneurs and rentiers* resulting from a change in the *wage-unit*." "If *fiscal policy* is used as a deliberate instrument for the more equal *distribution of incomes*, its effect in increasing the *propensity to consume* is, of course, all the greater."

captures endogenous redistribution (automatic stabilizers) as one key determinant of χ and hence of monetary and fiscal multipliers. What else can determine whether χ is larger or smaller than one? Two avenues seem possible to explore, corresponding to each of the two determinants already present in the simpler setup above.

The first is a more general, arbitrary redistribution scheme (I describe one example in the Appendix in detail). The general point is that, given an income function of aggregate income for H agents, say $C_t^H = \Gamma(Y_t) + \mathcal{T}$, a transfer will always reduce the elasticity of their after-tax income to aggregate income. In particular, the loglinearized consumption function is now, letting χ^0 denote the elasticity to aggregate income without the transfer $\chi^0 = \frac{\Gamma_Y Y}{\Gamma}$ and $\chi^\tau = \frac{\Gamma_Y Y}{\Gamma + \mathcal{T}}$ the elasticity with a transfer, it follows immediately that as long as the transfer is positive $\chi^\tau < \chi^0$. If the transfer is high enough, it can bring the model to the "dampening" region even if without a transfer it were in the "amplification region: $\chi^\tau < 1 < \chi^0$ (See also footnote 3). Redistribution through fiscal policy is of course a key component of some of the most prominent analyses of MP transmission in more complicated HANK model (e.g. KMV, Auclert, 2015; and Werning, 2015). Mine is a simple and transparent way to formalize that dependence directly addressing the key element in this model: the (re)distribution of profits that was already one of the major themes of Bilbiie (2008).

A recent paper by Debortoli and Gali (2017) also uses the version of the TANK model introduced in Bilbiie (2008)—whereby the distinction between the two types is due to access to asset markets, including shares of monopolistic firms—combining it with a more general transfer scheme than the one considered there, and in this paper. One common insight of all these papers is that the degree of redistribution governs aggregate amplification and dampening to monetary policy changes, relative to the RANK model. That paper takes a different direction and conducts a very useful comparison of the aggregate implications of TANK and HANK models, through their implied transfers; it shows that a TANK model can deliver the same "total effect" as a HANK model for an appropriately calibrated degree of fiscal redistribution, all other (common) parameters being equal. The same paper also extends the analysis of optimal monetary policy in TANK first conducted by Bilbiie (2008) for a particular fiscal redistribution scheme, to this case of a general transfer scheme and decreasing returns.²⁷

The second example of χ determinants pertains to the labor market: the transmission mechanism relies crucially on flexible wages. Colciago (2011) and Ascari, Colciago, and Rossi (2016) extended to sticky nominal wages. The main intuition is straightforward: ceteris paribus, sticky wages lead to dampening because, in this paper's notation, they reduce χ by making the wage less responsive to the cycle. It then follows immediately that sticky wages will also lead to *smaller* monetary and fiscal multipliers (the latter point has been made by Furlanetto, 2011 in a quantitative TANK model).

²⁷Previous extensions (in other directions) of the optimal policy analysis in Bilbiie (2008) include Ascari et al (2016), Nistico (2016), and Bilbiie and Ragot (2016).

4 The $\lambda\chi\delta$ (Simple-HANK) Model: Amplification and Dampening, Magnified

I now provide a generalization of this model that can be seen as a simple HANK: I label it $\lambda\chi\delta$ (from its key parameters). It is inspired by the recent influential contribution of McKay, Nakamura, and Steinsson (2015, 2016—hereinafter MNS), although it is different in some key respects: in particular, income of constrained depends on aggregate, as in the TANK model. Households are subject to idiosyncratic risk against which they self-insure against by using liquid bonds; to simplify and obtain analytical solutions, I then assume that these bonds are not traded in equilibrium.²⁸ Thus in equilibrium a fraction of agents are hand-to-mouth (as in the TANK model), while the others are savers (and stockholders) and have an Euler equation. This Euler equation now takes into account the possible transition to the "constrained", hand-to-mouth state—unlike the TANK model (nested here when idiosyncratic shocks become permanent, which eliminates self-insurance). And like the HANK model, my model distinguishes between liquid assets (bonds) and illiquid assets (stock). In equilibrium, there is thus infrequent (limited) participation in the stock market.²⁹

4.1 Discounting and Compounding through Self-Insurance in the $\lambda\chi\delta$ Model

There are two states, as in the TANK model: savers S and hand-to-mouth H . But unlike in the TANK model, there is now idiosyncratic risk: agents switch states following a Markov chain. The probability to **stay** type S is s and the probability to stay type H is h (while the transition probabilities are respectively $1 - s$ and $1 - h$), and by standard results the mass of H is:

$$\lambda = \frac{1 - s}{2 - s - h}.$$

The stability condition:

$$s \geq 1 - h$$

insures stationarity and has a straightforward economic interpretation: the probability to stay a saver/participant is larger than the probability to become one (the conditional probability is larger

²⁸The original contribution for this simple way of modelling self-insurance to idiosyncratic risk is Krusell, Mukoyama, and Smith (2011). Others have employed this no-trade equilibrium, e.g. Ravn and Sterk (2013), MNS (2015, 2016), and Werning 2015. See also Challe et al (2016) for an estimated quantitative model, and Ravn and Sterk (2016) for analytical results in a model that abstracts from the channels considered here but focuses on what that is absent: endogenous unemployment risk.

²⁹Curdia and Woodford (2009) and Nistico (2016) also study NK models with this "infrequent participation" metaphor due to Lucas (1990); their focus is different, and in their models constrained agents borrow in equilibrium subject to a spread. This is also related to Bilbiie and Ragot (2016), who build a model with three assets—of which one ("money") is liquid and is traded in equilibrium while the others are accessed only infrequently—and focuses on equilibrium liquidity to study optimal monetary policy in that framework.

than the unconditional).³⁰ Notice that this nests the TANK model when idiosyncratic shocks are permanent, $s = h = 1$: the share of H λ stays at its initial value and is a free parameter. At the other extreme, idiosyncratic shocks are iid when $s = 1 - h$: the probability for a household to be S or H tomorrow is independent on whether it is S or H today.

There are two assets: liquid public bonds (that will not be traded) and illiquid stock that can only be accessed when S . S households can thus infrequently become H and self-insure through bonds (liquidity), leaving their illiquid stock portfolio temporarily. The price for self-insurance is the interest rate on bonds that are *not traded*. The following Euler equation governs the bond-holding decision of S households who self-insure against the risk of becoming H :

$$(C_t^S)^{-\frac{1}{\sigma}} = \beta E_t \left\{ (1 + r_t) \left[s (C_{t+1}^S)^{-\frac{1}{\sigma}} + (1 - s) (C_{t+1}^H)^{-\frac{1}{\sigma}} \right] \right\}.$$

I therefore assume that in the H state the equivalent Euler equation holds with strict inequality: households are constrained, or impatient, and become hand-to-mouth thus consuming all their income $C_t^H = Y_t^H$.³¹ Abstract for now from any exogenous fiscal shocks (we reintroduce them below) to focus on the endogenous dynamics.

Loglinearizing around the same symmetric steady state $C^H = C^S$ as in the TANK model, the self-insurance equation is:

$$c_t^S = s E_t c_{t+1}^S + (1 - s) E_t c_{t+1}^H - \sigma r_t.$$

Replacing the consumption function of H that is identical to previously (5): $c_t^H = y_t^H = \chi y_t$ (whatever the redistribution scheme determining χ) we obtain the the aggregate Euler-IS:

$$c_t = \delta E_t c_{t+1} - \sigma \frac{1 - \lambda}{1 - \lambda \chi} r_t, \tag{11}$$

where $\delta \equiv 1 + (\chi - 1) \frac{1 - s}{1 - \lambda \chi}$.

Several remarks are in order. First, in the TANK limit of the previous section (permanent idiosyncratic shocks $s = h = 1$) we have no discounting $\delta = 1$, and λ is then an arbitrary free parameter. In the other extreme (the iid idiosyncratic uncertainty special case $s = 1 - h$ of e.g. Krusell Mukoyama Smith and McKay, Nakamura and Steinsson) we have $\lambda = h$ and $\delta = \frac{1 - \lambda}{1 - \lambda \chi}$. The striking implications for the aggregate Euler equation are summarized in the following Proposition.

Proposition 2 *Discounting, Compounding, and Complementarity.* *In the $\lambda \chi \delta$ model, the aggregate Euler equation features **compounding** under idiosyncratic risk ($s < 1$) if and only if*

³⁰An general version of this condition appears e.g. in Huggett (1993); see also Challe et al (2016) for an interpretation in terms of job finding and separation rates, and Bilbiie and Ragot (2016).

³¹One justification for this could be that the idiosyncratic shock is a preference shock to β rendering households impatient "enough" to make the constraint bind.

condition (9) holds:

$$\delta > 1 \iff \chi > 1,$$

the elasticity of hand-to-mouth income to aggregate is greater than one. Compounding ($\delta > 1$) is magnified by idiosyncratic risk $\partial\delta/\partial(1-s) > 0$ and by hand-to-mouth $\partial\delta/\partial\lambda > 0$.

The Euler equation features **discounting** $\delta < 1$ when the reverse of (9) holds: $\chi < 1$. The discounting effect is also magnified by idiosyncratic risk $\partial\delta/\partial(1-s) < 0$ and by share of H $\partial\delta/\partial\lambda < 0$.³²

Complementarity between the hand-to-mouth and self-insurance channels: When there is amplification ($\chi > 1$) the compounding effect is increasing with λ at a higher rate when there is more idiosyncratic risk: $\partial^2\delta/(\partial\lambda\partial(1-s)) = (\chi-1)\chi/(1-\lambda\chi)^2 > 0$. When there is dampening ($\chi < 1$), the same applies to the discounting effect (it is magnified at an increasing rate, i.e. δ decreases faster).

To understand these findings, let us start with the easier case of "discounting", which generalizes the finding of MNS (something like the case they considered is nested here for $\chi = 0$ —implying $\delta = s$ —and iid idiosyncratic shocks with $s = 1 - h = 1 - \lambda$). When good news about future *aggregate* income/consumption arrive, households recognize that in some states of the world they will be constrained (and not benefit from this good aggregate news, because $\chi = 0$). They seek to self-insure against this idiosyncratic risk, but this "precautionary" increase in saving demand cannot be accommodated (there is no asset), so the household consumes less today than it would if it were alone in the economy (or if there were no risk). Income adjusts accordingly to give the household the right incentives for this allocation. The same intuition holds (but the incentives to self-insure are mitigated) when χ is between 0 and 1. It is the interaction of idiosyncratic ($1-s$) and aggregate uncertainty (news about y_t , and how they translate into individual income through $\chi-1$) that determines the self-insurance channel. The higher the risk ($1-s$) and the lower the χ , the more the discounting (the lower is δ); furthermore, the higher the expected hand-to-mouth spell (higher λ at given s implies higher h), the stronger the self-insurance channel and the larger the discounting. In the limit as idiosyncratic shocks become permanent the self-insurance channel disappears and we recover the TANK model $\delta \rightarrow 1$.

The opposite holds in case 2, when $\chi > 1$: the effect of monetary policy is amplified—on the one hand through the elasticity to interest rate (as previously emphasized in Bilbiie, 2008 and above) but also, more surprisingly, through overturning the "discounting" effect discovered by MNS. The endogenous amplification through the Keynesian cross now holds not only contemporaneously, but also for the future: good news about future aggregate income increase today's demand because they imply less need for self-insurance, precautionary saving. Since future consumption in states where the constraint binds over-reacts to good "aggregate news", households internalize this by

³²More precisely: $\partial\delta/\partial(1-s) = \frac{\chi-1}{1-\lambda\chi}$; $\partial\delta/\partial\lambda = (\chi-1)\frac{(1-s)\chi}{(1-\lambda\chi)^2}$.

attempting to self-insure *less*. But the precautionary saving still needs to be zero in equilibrium, so households consume more than one-to-one and income increases to deliver this, thus delivering amplification. This effect is now magnified with higher risk (higher $1-s$): the highest compounding is obtained in the iid case, because this corresponds to the strongest self-insurance motive (at given χ).

The intuition for the complementarity is that an increase in idiosyncratic risk (higher $1-s$) has a larger effect on self-insurance when there is a longer expected hand-to-mouth spell $(1-h)^{-1}$, which is the same as a higher hand-to-mouth share λ ; evidently, the highest compounding for a given λ (and at given χ) is obtained in the iid case $1-s = h = \lambda$.

Figure 4 represents these effects by plotting the discounting/compounding coefficient δ as a function of the share of hand-to-mouth λ , for different degrees of idiosyncratic risk ($1-s$) and χ .³³ The figure illustrates clearly the amplification obtained under idiosyncratic risk when $\chi > 1$, and how it depends on the level of risk: the dots correspond to the highest possible level of risk, the iid case $1-s = \lambda$, the solid line to a low level of risk $1-s = 0.05$ and the red dashed line to the TANK limit with no risk $1-s = 0$ (the same value of $\delta = 1$ obtains for $\chi = 1$). Under that line, we go to the "discounting" region with $\chi < 1$: discounting is very mild for $1-s = 0.05$ (the thin solid line) and rapid with $1-s = \lambda$. The complementarity is illustrated by the curves shifting away from the horizontal dashed line as risk $1-s$ increases, in both cases (with $\chi > 1$ that implies more compounding, and with $\chi < 1$ more discounting).

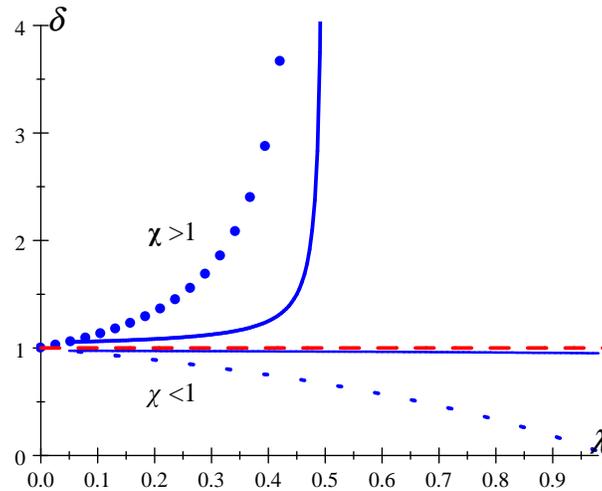


Fig 4: $1-s = \lambda$ (dots); 0.05 (solid); 0 (dash)

Lastly, note that like the TANK model, this model too trivially nests the representative-agent NK model when the constrained agents' income elasticity to aggregate income is unitary, $\chi = 1$ —

³³I left out the domain of λ where, in the $\chi > 1$ case, the aggregate elasticity of demand to interest rates changes sign—and so does δ . This case is thoroughly analyzed in Bilbiie (2008) and Bilbiie and Straub (2013); we return to it in Section 6.4 below.

for instance because agents are perfectly insured through either labor supply ($\varphi = 0$) or the tax system ($\tau^D = \lambda$). This result, as well as the finding emphasized in the previous paragraph (of more amplification with more constrained agents, partly because of the intertemporal, self-insurance dimension) are also related to important results recently obtained by Werning (2015) with a more general specification of income processes, where amplification arises when income *risk* is countercyclical and liquidity procyclical. My simple framework abstracts from the latter, and can be viewed as providing a simple and transparent formalization of the former, capturing it through one additional parameter in the aggregate Euler-IS curve. The framework that I develop is closer to the textbook NK model, and thus simpler in what concerns the modelling of market incompleteness—but allows a full-fledged NK analysis of a simple general equilibrium model (including a supply side, see below). It also allows a transparent comparison to the earlier analyses emphasizing amplification through the hand-to-mouth channel (e.g. Bilbiie, 2008) and thus isolating the role of self-insurance. One property that can be seen clearly due to this simplified framework is that it is exactly the same condition (9) $\chi > 1$ that governs whether there is amplification or not from the hand-to-mouth channel (Proposition 1) that also governs whether there is discounting or compounding in the Euler equation through the self-insurance channel (Proposition 2): once more, the key is χ . Perhaps most importantly, this simplification allows this paper to analyze all the puzzles and paradoxes of the NK model, and elucidate the inherent link between them and amplification in these models, and thus the dilemma giving the paper’s title.

4.2 PE Curve and NK Cross in the $\lambda\chi\delta$ Model

While the Euler-equation representation seems particularly useful to understand the possibility of compounding, in this model too we can recover the PE curve, or consumption function (whose cumbersome derivation is in the Appendix):

$$c_t = [1 - \beta(1 - \lambda\chi)]\hat{y}_t - (1 - \lambda)\beta\sigma r_t + \beta\delta(1 - \lambda\chi)E_t c_{t+1} + \beta\lambda\zeta_H\left(\chi - \frac{\alpha}{\lambda}\right)(g_t - E_t g_{t+1}) + \beta\zeta_H(1 - s)\left(\chi - \frac{\alpha}{\lambda}\right)E_t g_{t+1} \quad (12)$$

Relative to the TANK model there are two differences: the discounting/compounding through δ emphasized above, and a self-insurance motive for news about fiscal shocks, through their effect on aggregate future income (last term). Using this PE curve or the aggregate Euler-IS curve (obtained as before by imposing the economy resource constraint $c_t = \hat{y}_t$),

$$c_t = \delta E_t c_{t+1} - \sigma \frac{1 - \lambda}{1 - \lambda\chi} r_t + \frac{\lambda\zeta_H}{1 - \lambda\chi} \left(\chi - \frac{\alpha}{\lambda}\right) (g_t - E_t g_{t+1}) + \frac{(1 - s)\zeta_H}{1 - \lambda\chi} \left(\chi - \frac{\alpha}{\lambda}\right) E_t g_{t+1} \quad (13)$$

the effects of monetary policy (an interest rate cut) and fiscal stimulus (an increase in g), both

with persistence p , are found in the $\lambda\chi\delta$ model as:

$$\begin{aligned}\Omega &= \frac{\sigma}{1-\delta p} \frac{1-\lambda}{1-\lambda\chi}, \\ \mathcal{M}^* &= 1 + (\mathcal{M} - 1) \frac{1-p + \frac{1-s}{\lambda}p}{1-\delta p} > \mathcal{M} \\ \omega &= \frac{1-\beta(1-\lambda\chi)}{1-\delta\beta p(1-\lambda\chi)}.\end{aligned}\tag{14}$$

The results are very intuitive: when it comes to the effects of monetary and fiscal policy changes, the difference between the $\lambda\chi\delta$ and TANK models matters only for persistent shocks, or shocks that are about the future in some way (news shocks, or forward guidance); this is only natural, because self-insurance (the one channel that the $\lambda\chi\delta$ adds) is about future shocks. The main difference is thus that in the "compounding" case of the $\lambda\chi\delta$ model, there are two sources of amplification: the first is as in the TANK model, through increasing the contemporaneous elasticity of aggregate demand to interest rates (conversely, the slope of the PE curve in the NK cross is unchanged). The second is through the compounding effect δ , which only applies to future changes (i.e. if policy changes are persistent). For the fiscal multiplier in addition, fiscal shocks also matter through their influence on future income, and the self-insurance incentives that this triggers. Notice that a multiplier larger than one $\mathcal{M}^* > 1$ can be consistent with discounting $\delta < 1$ under self-insurance if and only if:

$$\frac{\alpha}{\lambda} < \chi < 1,$$

i.e. if there is enough endogenous redistribution (second inequality) and even more exogenous redistribution (first inequality). Under self-insurance the persistence of fiscal shocks p matters, through its interaction with idiosyncratic risk $1-s$; in particular (recall $1-s \leq \lambda$): $\partial\mathcal{M}^*/\partial p \geq 0$; $\partial\mathcal{M}^*/\partial(1-s) \geq 0$.³⁴ This suggests potential novel insights of this model (relative to TANK) into the role of public debt that are left for future work. Finally, note that the decomposition of the multiplier into a uniform-taxation spending multiplier and a transfer multiplier (10) still applies, with $\mathcal{M}_{g-uniform}^* = 1 + (\mathcal{M}_{g-uniform} - 1) \frac{1-p+p\frac{1-s}{\lambda}}{1-\delta p}$, and the multiplier of a pure redistribution (a transfer from S to H) $\mathcal{M}_{transfer}^* = \mathcal{M}_{transfer} \frac{1-p+p\frac{1-s}{\lambda}}{1-\delta p}$.

³⁴Specifically, the derivatives are $\partial\mathcal{M}^*/\partial p = \frac{1-s}{\lambda(1-\delta p)^2} \frac{1-\lambda}{1-\lambda\chi}$ and $\partial\mathcal{M}^*/\partial(1-s) = \frac{(1-p)p}{\lambda(1-\delta p)^2} \frac{1-\lambda}{1-\lambda\chi}$.

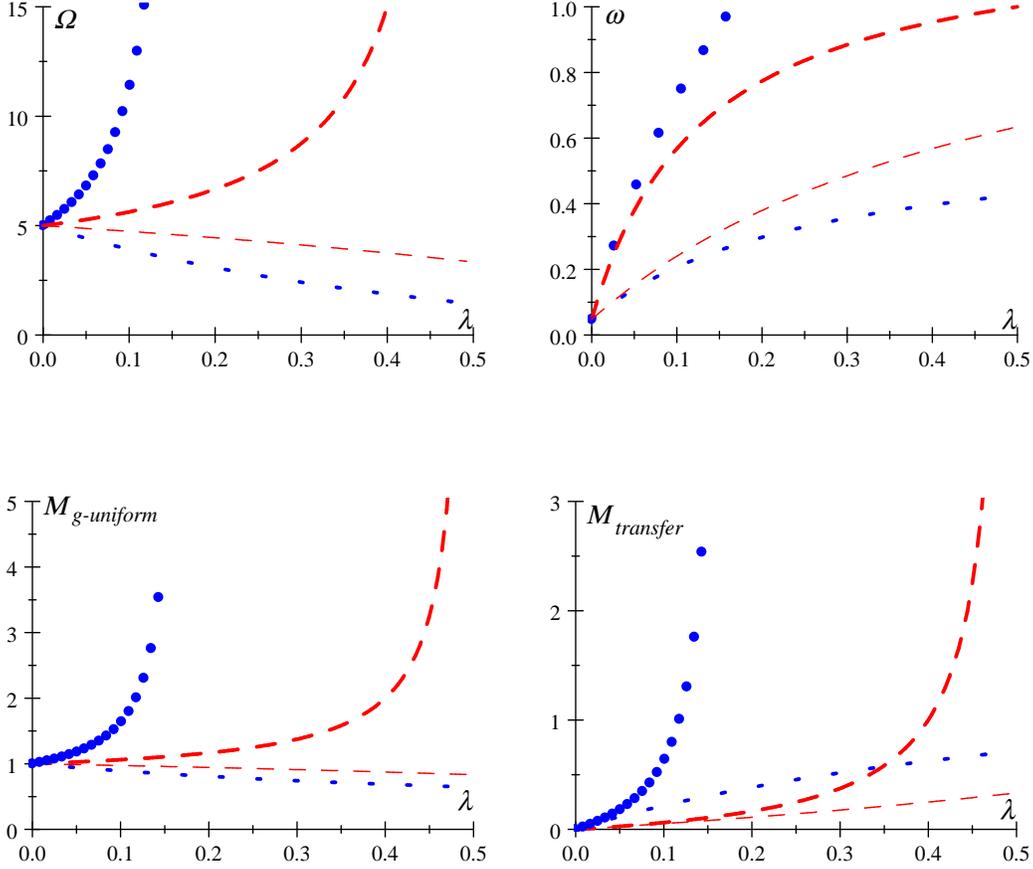


Fig 5: $\chi = 2$ (thick), 0.5 (thin), $s = 1 - \lambda$ (dots: iid $\lambda\chi\delta$) and 1 (dash: TANK)

Figure 5 illustrates and summarizes these findings, compared to the TANK model as described in Proposition 1; it plots the total and indirect effects of MP and the two multiplier components for FP in the $\lambda\chi\delta$ model as a function of the share of hand-to-mouth, for several cases, assuming that the persistence of the policy change is $p = 0.8$ (with iid monetary policy shocks the two models trivially coincide; while persistence does not matter for fiscal shocks in the TANK model). With red dashed line we have the TANK limit of the $\lambda\chi\delta$ model ($s = h = 1$), distinguishing between $\chi > 1$ and $\chi < 1$: as we saw above, in the former case there is amplification and in the latter dampening, and the share of the indirect effect increases with λ . These effects are amplified when moving towards higher risk (higher $1 - s$). In the limit when $1 - s = h = \lambda$, represented by blue dots, we have the highest compounding and the fastest discounting. It is worth noticing that the transfer multiplier is the only one that does not flip sign as χ crosses 1; indeed, its limit as λ tends to 1 is still well-defined for $\chi < 1$, even in the iid (largest-multiplier) case (it is: $\zeta_H / (1 - \chi) > 0$). Therefore, as in the TANK model transfers to the liquidity-constrained are the only policy instrument that works in stimulating the economy when $\chi < 1$.

4.3 Using TANK and $\lambda\chi\delta$ to understand HANK?

The bottomline of the analysis in this section and last is that the TANK model contains the seeds of heterogeneous-agent models that add self-insurance—such as HANK: in particular, the very same condition that gives amplification versus dampening in the TANK model also governs whether there is amplification or dampening in HANK models. In the former, this is due exclusively to the "static" hand-to-mouth channel operating within every period, while the latter adds a dynamic dimension: self-insurance against the risk of becoming hand-to-mouth in the future. Almost tautologically then, the hand-to-mouth channel summarized by $\lambda\chi$ is sufficient to understand the effects of short-lived, transitory shocks; while the self-insurance channel described by δ is necessary in order to understand the role of the medium- and long-run: the effects of persistent shocks, or of news about the future.

These results raise a cautionary note concerning the interpretation of TANK (without idiosyncratic risk) models' amplification as capturing HANK models' amplification, such as Debortoli and Gali (2017). These authors compare a TANK model similar to the one outlined in Section 3 and Bilbiie (2008, 2017)—but with a more general fiscal transfer scheme—and compare it with a HANK model à la McKay, Nakamura, and Steinsson (2015). In particular, they find the fiscal transfer scheme that gives the same total effect of a persistent monetary policy shock in the TANK model as in the HANK model, for a given calibration of the other parameters. In this paper's notation, they find a value of χ such that the Ω of the TANK model is the same as the Ω of a HANK model that they solve under the same parameterization. The exercise is therefore similar in spirit to the exercise in Section 3 in this paper and in Figure 3, whereby I match the indirect effect share ω computed by Kaplan, Moll and Violante (2016) in their HANK model to be 0.8. Both exercises are subject to a potentially important bias, for as shown in this Section's $\lambda\chi\delta$ model the self-insurance channel (that *is* present in HANK models) implies a complementarity between aggregate (as captured by χ) and idiosyncratic (as captured by $1 - s$) risk. As illustrated by Figure 5, one can get the same total effect Ω and indirect share ω for a much lower λ or χ (or both) than in a TANK model by adding sufficient idiosyncratic risk $1 - s$. In other words shutting down the self-insurance channel is asking the hand-to-mouth channel to do too much (all) of the job.

To understand the nature of the bias just described, denote by χ_0 the value of the composite parameter χ that makes $\Omega_{TANK} = \Omega_{HANK}$. Exploiting the analytical expressions provided in Proposition 1 and above for the $\lambda\chi\delta$ model, we find that to equate the total effect of the latter with the total effect of the TANK model under χ_0 : $\Omega_{\lambda\chi\delta} = \Omega_{TANK}(\chi_0)$ for an otherwise identical calibration (in particular, same λ) we need:

$$\chi + \frac{p}{1-p}(\chi - 1)\frac{1-s}{\lambda} = \chi_0,$$

which describes a surface for χ and $1 - s$ that clearly illustrates the complementarities discussed

above. A similar discussion applies if one tried to instead match only the indirect effect share ω : in the presence of self-insurance, the TANK model necessarily overstates the role played by (the determinants of) χ in order to match a given level of amplification.

An alternative approach to matching the amplification obtained from a given HANK model by using the $\lambda\chi\delta$ model is to use as a metric both the total and indirect effect and use them to indirectly infer both χ and s . Denote by $\bar{\Omega}$ and $\bar{\omega}$ the values that we are trying to match, perhaps because they were obtained from a fully-fledged HANK model: e.g. in KMV, the total effect $\bar{\Omega}$ is 50 percent higher than in the RANK model (which under $\sigma = 1$ and for $p = 0.8$ is equal to 5, meaning $\bar{\Omega} = 7.5$) and, as discussed previously, $\bar{\omega} = 0.8$. We can then invert the corresponding expressions in (14) to solve for the χ and δ (and therefore s) that deliver these values in the $\lambda\chi\delta$ model; simple algebra delivers the expressions:

$$\begin{aligned}\chi &= \lambda^{-1} \left(1 - \beta^{-1} + \frac{\sigma(1-\lambda)}{\bar{\Omega}(1-\bar{\omega})} \bar{\omega} \right); \\ \delta &= \frac{1 - \beta\sigma(1-\lambda) - \bar{\Omega}(1-\bar{\omega})}{p\bar{\omega}\beta\sigma(1-\lambda) - \bar{\Omega}(1-\bar{\omega})},\end{aligned}$$

which for a calibration with $\beta = 0.99$ and $\lambda = 0.2$ implies the values $\chi = 2.0828$ and $\delta = 1.0215$, or in terms of idiosyncratic risk $1 - s = 0.01158$. A similar "indirect inference" exercise can be conducted for any other calibrated HANK model, insofar as the equilibrium objects Ω, ω and/or \mathcal{M} are calculated in those models.

5 The "Simple HANK" 3-Equation $\lambda\chi\delta$ Model: Adding AS

We now move closer to the standard NK model and visit classic and important topics with the familiar ingredients. Aggregate demand is still given by the aggregate Euler-IS equation above (13), with $r_t \equiv i_t - E_t\pi_{t+1}$ the ex-ante real rate and i_t the nominal interest rate controlled by the central bank, as in Woodford's celebrated 2003 Wicksellian analysis. The central bank sets the nominal rate i_t according to a Taylor rule:

$$i_t = i_t^* + \phi E_t\pi_{t+1} \tag{15}$$

i_t is the nominal interest rate set by the central bank and expressed in levels (i.e. the ZLB is $i_t \geq 0$), $E_t\pi_{t+1}$ expected inflation, and ρ_t an exogenous shock that is standard in the liquidity-trap literature, see below. The intercept of the Taylor rule i_t^* is an exogenous (possibly persistent) process.

We add a supply side in the simplest possible way, a standard Phillips curve:

$$\pi_t = \beta_f E_t\pi_{t+1} + \kappa c_t + \zeta_H \kappa g_t, \tag{16}$$

with a twist. Closed-form results are particularly useful here in order to shed light on the role of each amplification channel and analyze determinacy conditions as well as NK puzzles and paradoxes. To obtain such analytical tractability, I first focus on a special case: the simplest possible *aggregate supply* curve whereby each period a fraction of firms f keep their price fixed, while the rest can re-optimize their price freely *but* ignoring that this price affects future demand. This delivers the Phillips curve with $\beta_f = 0$: $\pi_t = \kappa c_t$, where now $\kappa = (\varphi + \sigma^{-1})(1 - f)/f$. While clearly over-simplified, this setup nevertheless captures a key mechanism of the NK model, i.e. a trade-off between inflation and real activity; results are conceptually very similar when considering a more standard Phillips curve adding future expected inflation, as I show in the Appendix. Using the simpler AS block allows us to isolate and focus on the main topic of this paper, and the essence: AD.³⁵

5.1 The (HANK) Taylor Principle and Interest Rate Pegs

Under the assumed structure, the model is disarmingly simple: replacing the Phillips curve and Taylor rule in the aggregate Euler equation, the $\lambda\chi\delta$ model boils down to **one** (!) equation:

$$c_t = \nu E_t c_{t+1} - \sigma \frac{1 - \lambda}{1 - \lambda\chi} i_t^*, \quad (17)$$

where the newly defined parameter:

$$\nu \equiv \delta - (\phi - 1) \kappa \sigma \frac{1 - \lambda}{1 - \lambda\chi}$$

has a very useful interpretation: it captures the effect of **good news** on AD (the notation purposefully agrees with that intuition); part of this news has an effect through δ , as explained when discussing proposition 2. The other term comes from the inflationary effect (κ) of future good news on income; this future expected inflation triggers movements in the real rate ($\phi - 1$); when $\phi > 1$ and policy is "active" in Leeper's 1991 terminology, future inflation leads to an increase in the real rate, which has a contractionary effect today. This channel is standard in the NK model, and is magnified here because of the "TANK" amplification through the hand-to-mouth channel $\lambda\chi$ when $\chi > 1$ as the elasticity of demand to interest rates is increasing. The opposite is true, of course, if we are in the dampening region $\chi < 1$ with, for example, enough endogenous redistribution or sticky enough prices. It is precisely these considerations that deliver the main result concerning equilibrium determinacy and ruling out sunspot equilibria.

Proposition 3 *The HANK Taylor Principle: The model has a (locally) unique rational ex-*

³⁵Essentially, such a setup reduces to assuming $\beta_f = 0$ *only* in the firms' problem (they do not recognize that today's reset price prevails with some probability in future periods). See Bilbiie (2016) for an extension to the case $\beta_f > 0$, and a comparison in the context of optimal forward guidance in the baseline NK model.

expectations equilibrium, i.e. "determinacy" if and only if (as long as $\lambda < \chi^{-1}$):

$$\nu < 1 \Leftrightarrow \phi > 1 + \frac{\delta - 1}{\kappa\sigma \frac{1-\lambda}{1-\lambda\chi}},$$

with two implications worth spelling out:

a. The **Taylor principle** $\phi > 1$ is sufficient for determinacy if and only if:

$$\chi \leq 1 \rightarrow \delta \leq 1.$$

b. **Sargent-Wallace revisited:** An interest rate **peg** $\phi = 0$ leads to a locally unique equilibrium if and only if

$$\nu_0 \equiv \delta + \kappa\sigma \frac{1-\lambda}{1-\lambda\chi} < 1.$$

The proposition follows immediately by recalling that the requirement for existence of a (locally) unique rational expectations equilibrium is that the root, here equal to ν , be inside the unit circle. It is evident that in the discounting case $\delta < 1$, the threshold ϕ is *weaker* than the Taylor principle, while in the compounding case it is *stronger*. Written differently:

$$\phi > 1 + (\chi - 1) \frac{(1-s)}{\kappa\sigma(1-\lambda)}$$

The intuition is the same as for other "demand shocks": in the *compounding* case, there is a more powerful demand amplification to sunspot shocks, which raises the need for a more aggressive response in order to rule out sunspot equilibria. The higher the risk $(1-s)$ and the higher the elasticity of H income to aggregate χ the higher this endogenous amplification, and the higher the threshold. The opposite is true in the *discounting* case, since the transmission of sunspot shocks on current demand is then dampened. Recall that this demand amplification is increasing with the degree of price stickiness (which governs the labor demand expansion that sets off the Keynesian spiral, as opposed to the direct inflationary response): thus, the threshold is also increasing with price stickiness (decreasing with κ). The Taylor threshold $\phi > 1$ is recovered for either of $\chi = 1$, $s = 1$, or $\kappa \rightarrow \infty$ (flexible prices). But the determinacy region for ϕ squeezes very rapidly with idiosyncratic risk when prices are sticky, as clear from the expression.

The second immediate implication, useful for future reference, is that an interest-rate peg may deliver determinacy in the $\lambda\chi\delta$ model, thus revisiting the classic Sargent-Wallace result and its more modern formulations, e.g. Woodford, 2003. With enough endogenous dampening, be it directly through the self-insurance channel or through mitigating the "expected inflation" channel, a pure expectation shock has no effects even without any feedback from expectations to the policy rate; the sunspot is ruled out inherently by the economy's endogenous forces. As a side note and a useful reminder, this is utterly impossible in the standard NK model, where $\nu_0 = 1 + \kappa\sigma \geq 1$.

These considerations are intimately related to several of the puzzles and paradoxes at work in NK models, and we return to them below.

What *can* the central bank do when the Taylor principle fails miserably—which happens in this economy at reasonable calibrations.³⁶ The central bank can still ensure equilibrium determinacy by adopting what Woodford (2003) proposed as a *Wicksellian rule*: responding to the price level instead of the inflation rate. Such a rule of the form $i_t = \phi_p p_t$ ensures determinacy here too, like in the standard model, for any $\phi_p > 0$ —I prove this formally in Appendix C.2. Another option to obtain determinacy is to resort to fiscalist equilibria—the same way one does in the standard model, by introducing nominal government debt and a fiscal rule that is "active" in the sense of Leeper (1991), i.e. it does not ensure that debt is eventually repaid for any possible price level (i.e., that the government debt equation is a constraint)—see also Woodford (1996), and Cochrane (2017) for further implications.³⁷

Lastly, notice that adding an AS side has straightforward implications for fiscal multipliers. Using the Euler-IS curve (13) and the Taylor and Phillips curves, the multiplier is:

$$\mathcal{M}^{**} = \mathcal{M}^* - \frac{p\zeta_H}{1 - \nu p} \kappa \sigma \frac{1 - \lambda}{1 - \lambda\chi} (\phi - 1).$$

The last term is the novel part with respect to the fixed- r model; it is standard in that, as in RANK models, G has demand effect which, if expected to persist ($p > 0$), leads to higher expected inflation; if the monetary policy rule is active $\phi > 1$, this leads to an increase in real interest rates and intertemporal substitution towards the future—thus to a contraction today. This effect is amplified by hand-to-mouth when $\chi > 1$, because as we saw above this increases the elasticity of aggregate demand to interest rates. It is also amplified, under the same condition, by δ through the mechanism explained at length above for persistent shocks. Lastly, this channel is only operative if $\zeta_H > 0$, that is if the income and substitution effects interplay such that some of the negative income effect of taxation is accommodated though reducing consumption: else ($\varphi = \zeta_H = 0$), it all goes into the infinitely-elastic labor supply.

5.2 Liquidity Traps: Deep Recessions without Deflation?

To study liquidity traps in this economy, we need to introduce a shock that triggers the zero lower bound on nominal interest rates to bind. A simple and standard way to do this is (ignoring any other shocks momentarily) is to consider that the ex ante real rate r_t in the aggregate Euler-IS equation above (13) is given now by $r_t \equiv i_t - E_t \pi_{t+1} - \rho_t$ where i_t is the nominal interest rate set

³⁶For example with $\chi = 2$, $\kappa = 0.02$, $\sigma = 1$ and small idiosyncratic risk $1 - s = 0.95$, the threshold ϕ is still around 4 at the lowest value of λ , 0.05, and increases sharply with λ ; the problem is of course compounded at higher values of the risk.

³⁷Since ours is an incomplete-markets economy, a further option to determine the price level exists, discovered by Hagedorn (2017): the self-insurance equation defines a demand for nominal debt, which if supplied by the government can determine the price level on the asset market, without resorting to a Taylor rule.

by the central bank and *expressed in levels* (i.e. the ZLB is $i_t \geq 0$), and ρ_t an exogenous shock that is standard in the liquidity-trap literature (Eggertsson and Woodford, 2003). This disturbance governs impatience, or the urgency to consume in the present (its steady-state value is the normal-times discount rate $\rho = \beta^{-1} - 1$): when it increases, households try to bring consumption into the present and "dis-save", and vice versa when it decreases.

A **liquidity trap** is triggered by the zero lower bound binding as in the seminal paper of Eggertsson and Woodford (2003): ρ_t follows a Markov chain with two states. The first is the steady state denoted by S , with $\rho_t = \rho$, and is absorbing: once in it, there is a probability of 1 of staying. The other state is transitory and denoted by L : $\rho_t = \rho_L < 0$ with persistence probability z (conditional upon starting in state L , the probability that $\rho_t = \rho_L$ is z , while the probability that $\rho_t = \rho$ is $1 - z$). At time t , there is a negative realization of $\rho_t = \rho_L < 0$, meant to capture in this reduced-form model an increase in spreads as in Curdia and Woodford (2009). I assume for further simplification that the monetary authority tracks the natural interest rate of this economy (which absent other shocks is equal to ρ_t) whenever feasible, meaning $i_t = \max(\rho_t, 0)$. It follows that the ZLB will bind when $\rho_t = \rho_L < 0$, while the flexible-price efficient equilibrium will be achieved whenever $\rho_t = \rho$.

Since the shock is unexpected, we can solve the model in the liquidity trap state, denoting by subscript L the time-invariant equilibrium values of consumption and inflation therein:

$$c_L = \frac{1}{1 - z\nu_0} \sigma \frac{1 - \lambda}{1 - \lambda\chi} \rho_L, \quad (18)$$

where $z\nu_0 < 1$ is needed to rule out expectations-driven liquidity traps,³⁸ ν_0 is again the effect of news on aggregate demand under a peg (ν for $\phi = 0$) defined above. Inflation is $\pi_L = \kappa c_L$.

Why an increase in the desire to save generates a recession with a binding zero lower bound in the standard NK model is much-researched territory since more than a decade: it causes excess saving and, with zero saving in equilibrium, income has to adjust downwards to give the income effect consistent with that equilibrium outcome. And if prices are not entirely fixed, there is also deflation, which—because it causes an increase in real rates when the zero bound is binding—leads to a further contraction, and so on.

In the simple HANK model, the magnitude of the liquidity trap recession depends on the three parameters s , λ , and χ that are key for transmission more generally. In particular, (18) elucidates how this will be different from the representative-agent model through three channels: the within-period demand elasticity to interest rates $\sigma \frac{1-\lambda}{1-\lambda\chi}$ (this holds even for transitory shocks $z = 0$ and for fixed prices $\kappa = 0$); and—for persistent shocks $z > 0$ only—the rate of discounting in the Euler equation δ (even with fixed prices) and again $\sigma \frac{1-\lambda}{1-\lambda\chi}$ but through its interaction with the Phillips curve slope κ . It is worth spending some time now understanding each channel. As

³⁸See Bilbiie (2016) for further discussion, Benhabib, Schmitt-Grohe and Uribe (2001, 2002). for the original point regarding sunspot ZLB equilibria, and Schmitt-Grohe and Uribe (2016) for a recent application.

for monetary policy shocks, the main distinction is between two cases, according to whether H 's income elasticity to aggregate income is lower or larger than one: once more, the key is χ .

When condition (9) holds ($\chi > 1$) there is **amplification** (an LT recession deeper than in the representative-agent NK model), to which there are two sides. The first is the *hand-to-mouth* channel operating through $\sigma \frac{1-\lambda}{1-\lambda\chi}$, by Proposition 1. The second is the *self-insurance* channel, through which there is *compounding* in the aggregate Euler equation ($\delta > 1$) by Proposition 2. Conversely, when (9) does not hold ($\chi < 1$) **dampening** occurs through both channels.

Let us briefly discuss the intuition for the amplification case—the dampening case being merely the mirror image. Take first the *hand-to-mouth channel* operating through $\lambda\chi$: the aggregate elasticity of intertemporal substitution—the elasticity of aggregate demand to interest rates *within the period*—is *increasing* with the share of H agents, as long as $\lambda < \chi^{-1}$. This is again the "New Keynesian cross" at work: a fall in the natural interest rate implies an aggregate demand contraction, through intertemporal substitution of S agents; with sticky prices, this translates into a labor demand contraction, which compresses the real wage. Since the wage is the H agents' income, this reduces their demand further, which magnifies the initial demand contraction³⁹.

Second, the *self-insurance channel* operating through δ . The endogenous amplification through the Keynesian cross now holds not only contemporaneously, but also for the future—insofar as the liquidity trap is expected to persist: bad news about future aggregate income reduce today's demand because they imply *more* need for self-insurance, precautionary saving. Since future consumption in states where the constraint binds over-reacts to bad "aggregate news", households internalize this by attempting to self-insure *more*. But since precautionary saving needs to be zero in equilibrium, households consume less and income falls to deliver this, thus magnifying the recession even further. As understood from our discussion of the transmission mechanism and emphasized in Proposition 2, the two channels are complementary.

Lastly, the *expected deflation channel*. A shock that is expected to persist with z triggers self-insurance because of expected deflation ($\kappa\sigma \frac{1-\lambda}{1-\lambda\chi}$), which at the ZLB means an increase in interest rate—so more saving and, since equilibrium saving is zero, less consumption and less income. This last effect operates in the standard representative-agent model too, but here it is amplified by the hand-to-mouth channel (it is proportional to $\sigma \frac{1-\lambda}{1-\lambda\chi}$).⁴⁰

Recall now that in the standard NK model, a "deep recession" can occur in response to a financial disruption only if accompanied by a large deflation: $c_L = \sigma\rho_L / (1 - z(1 + \sigma\kappa))$ can only

³⁹This mechanism is also at play in Eggertsson and Krugman's deleveraging-based model of a liquidity trap, where it compounds with a debt-deflation channel. The borrowers whose constraint is binding at all times are, effectively, hand-to-mouth (even though their income then comprises nominal financial income that I abstract from and is at the core of Eggertsson and Krugman's analysis). A related channel is also at work in Guerrieri and Lorenzoni (2017).

⁴⁰Turning the above logic over its head, in the *dampening* case ($\chi < 1$) the LT-recession is *decreasing* with λ and $1 - s$: the more H agents and the more risk, the lower the elasticity to interest rates within the period, and the lower the discount factor of the Euler equation δ —both of which lead to dampening (and increasingly so when taken together, through the complementarity).

be large in absolute value if κ is large enough. Not in the heterogeneous-agent model: due to the specific amplification mechanisms emphasized above (hand-to-mouth and self-insurance), there can be a deep recession driven by a negative ρ_L shock even for fixed prices $\kappa = 0$. A version of this mechanism has been analyzed in a heterogeneous-agent model by Guerrieri and Lorenzoni (2017) to analyze deleveraging-induced deep recessions in a liquidity trap.

Like in the textbook NK model, and as emphasized by Eggertsson (2010), Christiano, Eichenbham, and Rebelo (2011), and Woodford (2011), fiscal multipliers are larger in a liquidity trap; in particular, solving the model formed by (13) under $i_t = 0$ and using the Phillips curve, the multiplier is given by the second term in:

$$c_L = \frac{1}{1 - z\nu_0} \sigma \frac{1 - \lambda}{1 - \lambda\chi} \rho_L + \left(\mathcal{M}^* + \frac{z\zeta_H}{1 - z\nu_0} \kappa \sigma \frac{1 - \lambda}{1 - \lambda\chi} \right) g_L.$$

As is well understood, the additional "expected inflation" channel that reduces multipliers away from the zero lower bound with an active monetary policy (through higher interest rates) now *increases* multipliers in a liquidity trap. For higher expected inflation now means a *fall* in the real rate, and incentives to bring consumption towards the present through intertemporal substitution (the second term in brackets). This positive effect is increasing with λ when $\chi > 1$ because the elasticity of aggregate demand to interest rates is higher;⁴¹ it is, of course, also increasing with δ .

These amplification (or dampening) channels shape the effect of shocks to the natural rate of interest (and the ensuing recessions) and fiscal stimulus, but they also shape the effects of news about future interest rates, aka forward guidance FG, that we attack further below.

6 The Catch-22: Amplification, or Puzzles and Paradoxes?

If you were to summarize the previous sections in one sentence, it would be that heterogeneous-agent NK models can deliver monetary and fiscal multipliers (amplification) only if $\chi > 1$: the key is χ . Unfortunately, this very same condition is what gets us in trouble: under $\chi > 1$, the troubles (puzzles and paradoxes) of the NK models are amplified, and new ones occur. The multiplier multiplies not only the good, but also the bad.

That market incompleteness can help solve one of the NK puzzles (the forward guidance one) has been the subject of several papers; while other papers doing other extensions of the standard NK model noticed the connection with other puzzles.⁴² Here, I provide an unifying treatment that makes possible to derive the analytical necessary and sufficient conditions for the incomplete-markets model to solve the FG puzzle and the other puzzles (some of which are discussed in other

⁴¹This amplification also holds for the multipliers derived by Eggertsson and Krugman (2012) in a deleveraging-driven LT.

⁴²See Garcia-Schmidt and Woodford (2014) for an application of reflective equilibrium; Gabaix (2016) for a behavioral model; Angeletos and Lian (2017) for imperfect common knowledge; Diba and Loisel (2017) for pegging the interest on reserves; and Cochrane (2017) for a fiscalist solution with long-term debt.

papers, others not).

6.1 Trouble 1: Forward Guidance Power and Puzzle

I model FG stochastically through a Markov chain, as a state of the world with a probability distribution, as follows.⁴³ Recall that the (stochastic) expected duration of the LT is $T_L = (1 - z)^{-1}$, the stopping time of the Markov chain. After this time T_L , the central bank commits to keep the interest rate at 0 while $\rho_t = \rho > 0$, with probability q . Denote this state by F , and let $T_F = (1 - q)^{-1}$ denote the expected duration of FG. The Markov chain implied by this structure has three states: liquidity trap L ($i_t = 0$ and $\rho_t = \rho_L$), forward guidance F ($i_t = 0$ and $\rho_t = \rho$) and steady state S ($i_t = \rho_t = \rho$), of which the last one is absorbing. The probability to transition from L to L is, as before, z , and from L to F it is $(1 - z)q$. The persistence of state F is q , and the probability to move back to steady state from F is hence $1 - q$.

Under this stochastic structure, expectations are determined by:

$$E_t c_{t+1} = z c_L + (1 - z) q c_F \quad (19)$$

and similarly for inflation. Evaluating the aggregate Euler-IS (11) and Phillips ($\pi_t = \kappa c_t + \zeta \kappa g_t$) curves during state F and L and solving for the time-invariant equilibria delivers equilibrium consumption (and inflation) during the forward guidance state F and the liquidity trap state L respectively as:

$$\begin{aligned} c_F &= \frac{1}{1 - q\nu_0} \sigma \frac{1 - \lambda}{1 - \lambda\chi} \rho; \\ c_L &= \frac{1 - z}{1 - z\nu_0} \frac{q\nu_0}{1 - q\nu_0} \sigma \frac{1 - \lambda}{1 - \lambda\chi} \rho + \frac{1}{1 - z\nu_0} \sigma \frac{1 - \lambda}{1 - \lambda\chi} \rho_L, \end{aligned} \quad (20)$$

and $\pi_F = \kappa c_F$, $\pi_L = \kappa c_L$; ν_0 is again the response of consumption in a liquidity trap to news about future income/consumption.

It is immediately apparent that the future expansion c_F is increasing in the degree of forward guidance q regardless of the model. In the amplification case ($\chi > 1$), the future expansion is also *increasing* with the share of hand-to-mouth λ and with risk $1 - s$; whereas in the dampening case, the opposite holds.

Figure 6 illustrates these findings: Distinguishing between dampening $\chi < 1$ (left) and amplification $\chi > 1$ (right), it plots in both panels consumption in the liquidity trap (thick) and in the FG state (thin), as a function of the FG probability q .⁴⁴ We represent the RANK model by green

⁴³This was introduced by Bilbiie (2016) in a representative-agent model; I refer the reader to that paper for details, robustness, an application to optimal monetary policy subject to the zero lower bound, and a "simple FG rule" implementation of that optimal policy.

⁴⁴The illustrative parametrization used in the Figure has $\kappa = 0.02$, $\sigma = 1$, $\varphi = 1$, $p = 0.8$ and a spread shock of $\rho_L = -0.01$, i.e. 4 percent per annum. This delivers a recession of 5 percent and annualized inflation of 1 percent

solid lines, the TANK limit of the $\lambda\chi\delta$ model ($s = h = 1$) with red dashed lines, and the $\lambda\chi\delta$ model in the iid case $1 - s = h = \lambda$ with blue dots.

The pictures illustrate the dampening and respectively amplification at work: at given q , low future rates have a lower effect (on both c_F and c_L) in the TANK model, and an even lower one in the $\lambda\chi\delta$ model, in the *dampening* case. The last point illustrates the complementarity: the dampening is magnified when moving towards higher risk $1 - s$ and in the limit when $1 - s = h = \lambda$ (blue dots) we have the fastest discounting. Whereas in the amplification case (right panel), the opposite is true: low rates have a higher effect in the TANK model, and through complementarity an even higher one under self-insurance: the pictured iid case represents the highest compounding. Indeed, even though $\chi = 2$ is a rather conservative number and the share of H is very small ($\lambda = 0.1$)—which makes amplification in the TANK version rather limited—amplification in the $\lambda\chi\delta$ model is tremendous: the recession is 3 times larger than in the RANK model. This number goes up steeply when we use the forward-looking Phillips curve, or when we increase either λ or χ if only slightly.⁴⁵

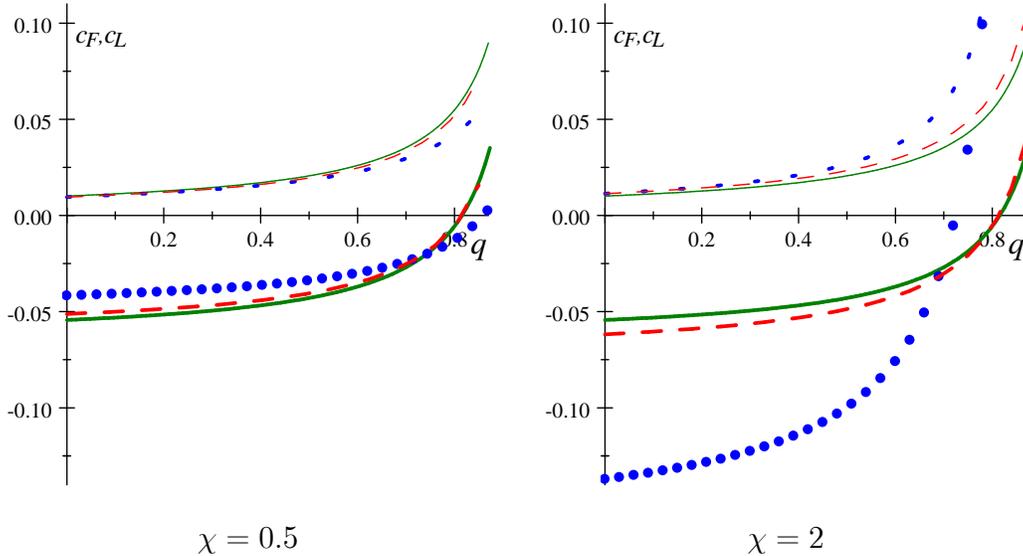


Figure 6: c_L (thick) and c_F (thin) in RANK (green solid), TANK (red dashed) and $\lambda\chi\delta$ -iid (blue dots)

We can now define the FG power and puzzle formally as follows. *FG power*, denoted by \mathcal{P}_{FG} , is the derivative of consumption during the trap c_L with respect to q , dc_L/dq :

$$\mathcal{P}_{FG} \equiv \frac{dc_L}{dq} = \left(\frac{1}{1 - q\nu_0} \right)^2 \frac{(1 - z)\nu_0}{1 - z\nu_0} \sigma \frac{1 - \lambda}{1 - \lambda\chi} \rho.$$

As we can already see in Figure 6, this is much larger in the $\lambda\chi\delta$ model in the "amplification" case. The properties of amplification and dampening of FG power follow the same logic as those applying

in the RANK model absent FG ($q = 0$). The domain is such that $q < \nu_0^{-1}$.

⁴⁵Indeed, with $\beta_f = 0.99$ the recession is 10 times larger.

to the natural rate shock and LT recessions. Since \mathcal{P}_{FG} is increasing with ν (and hence with both δ and $\sigma \frac{1-\lambda}{1-\lambda\chi}$), in the amplification case $\chi > 1$ it increases with idiosyncratic risk $1 - s$ and with the share of hand-to-mouth λ (while it decreases in the dampening case $\chi < 1$). Furthermore, the complementarity between self-insurance and hand-to-mouth also applies to FG power. The *FG puzzle* is the property (thus labelled by Del Negro, Giannoni and Patterson, 2012) that the power of FG increases, the further it is pushed into the future (i.e., in our context, with the persistence of the trap z): $\frac{d\mathcal{P}_{FG}}{dz} \geq 0$. When does the model resolve the FG puzzle?

Proposition 4 *The (simple HANK) $\lambda\chi\delta$ model **solves** the FG puzzle ($\frac{d\mathcal{P}_{FG}}{dz} < 0$) if and only if:*

$$\nu_0 < 1,$$

*i.e. **both**:*

$$1 - s > 0 \text{ \textit{and}} \\ \chi < 1 - \sigma\kappa \frac{1 - \lambda}{1 - s} < 1;$$

The model needs both the self-insurance and hand-to-mouth channels, which is a manifestation of complementarity.

The result follows directly calculating the derivative $d\mathcal{P}_{FG}/dz = \frac{(\nu_0-1)\nu_0}{[(1-q\nu_0)(1-z\nu_0)]^2} \sigma \frac{1-\lambda}{1-\lambda\chi} \rho$ and then replacing the expression for ν_0 . The Proposition emphasizes that, to solve the FG puzzle, the model needs *two conditions*: some idiosyncratic uncertainty $1 - s > 0$, and a cyclical income of H 's income that is lower than the threshold defined above. This is a clear illustration of the *complementarity* emphasized in Proposition 1. In other words, having discounting in the aggregate Euler equation ($\delta < 1$) is a necessary, but not sufficient condition to solve the puzzle.⁴⁶ Rewriting the condition, we have $\sigma\kappa < \frac{(1-s)(1-\chi)}{1-\lambda}$: this is more stringent when prices are more flexible (κ larger) and λ smaller (at given s and χ).

To consider an even simpler example, consider the case of acyclical income of H , $\chi = 0$. Under that assumption the discount factor in the Euler equation is equal to the probability s , and the effect of news is $\nu_0 = s + (1 - \lambda) \sigma\kappa$; this is not necessarily smaller than 1—for example, in the TANK model it is larger than one since $s = 1$. To solve the FG puzzle, there needs to be enough idiosyncratic risk, namely in this case $1 - s > (1 - \lambda) \sigma\kappa$. It is worth noticing that MNS's 2016

⁴⁶Farhi and Werning (2017) emphasize a different complementarity, between market incompleteness and "k-level thinking": an informational imperfection related to Garcia-Schmidt and Woodford's notion of reflective equilibrium, that leads to mitigation of FG—also through discounting in the Euler equation. In their framework, market incompleteness magnifies the mitigation of FG effects obtained with k-level thinking. The complementarity I emphasize is between two different channels, and can work both ways—generating more mitigation, or more amplification. Indeed, it affects not only the quantitative properties, but the qualitative insights: it changes the sign of a key derivative, as illustrated in Figure 7 below, needed to solve the FG puzzle as defined formally here.

simple model (with iid idiosyncratic risk and exogenous income of H) inherently satisfies these conditions: essentially, to $\chi = 0$ it adds $s = 1 - \lambda$.

Notice that with fixed prices $\kappa = 0$ the requirement becomes $\delta < 1$: Euler-equation *discounting* (MNS) and thus $\chi < 1$ is then sufficient to solve the FG puzzle. I provide an alternative illustration (working with the more familiar setup where FG takes place a given number of periods in the future, and how its power changes with this horizon) of this in the Appendix.

To further illustrate how the FG puzzle operates, and how the complementarity between the two channels helps eliminate it, consider Figure 7; it plots FG power as a function of z , for the same calibration as before (fixing in addition $q = 0.5$) in the two cases $\chi < 1$ and $\chi > 1$ for the three models RANK, TANK, and iid $\lambda\chi\delta$. It illustrates clearly that it is the interaction of dampening through $\chi < 1$ and idiosyncratic risk (which, as shown above, magnifies that dampening through discounting) that leads to resolving the FG puzzle: the power of FG becomes decreasing in the duration of the trap. The dampening channel by itself (TANK model with $\chi < 1$, red dashed line on the left panel) is not enough—although it alleviates the puzzle relative to the RANK model, it does not make the power decrease with the horizon z . While the self-insurance channel by itself added to the "amplification" case magnifies power even further, thus aggravating the puzzle (blue dots in the right panel for the $\lambda\chi\delta$ iid model in the amplification case).

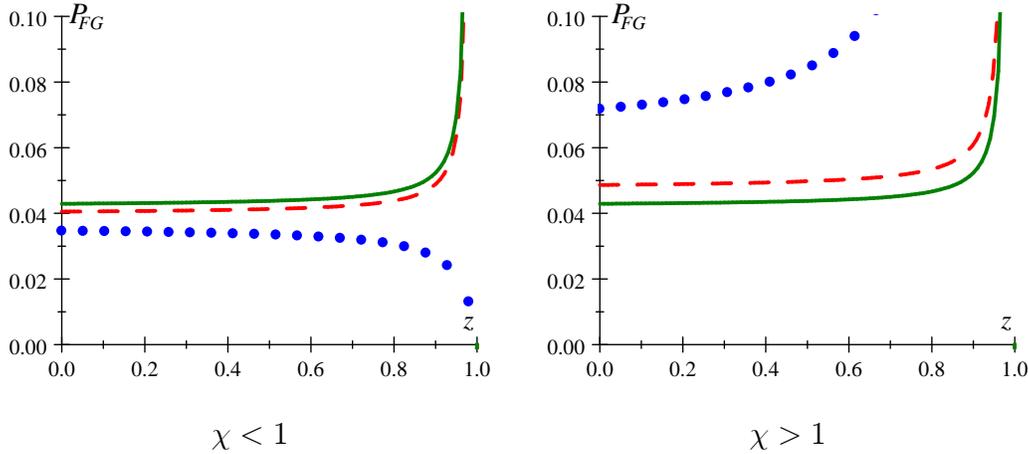


Figure 7: FG power in RANK (green solid), TANK (red dashed) and $\lambda\chi\delta$ -iid (blue dots)

Finally, it is immediate to see that the puzzle is aggravated at higher values of ν_0 ($\frac{dP_{FG}}{dz}$ is increasing in ν_0). It follows from the monotonicity of ν_0 that the puzzle is alleviated with higher idiosyncratic risk $1 - s$ and with λ in the dampening case; but worsens with idiosyncratic risk $1 - s$ and with λ in the amplification case $\chi > 1$.

6.2 Trouble 2: The Paradox of Flexibility

A separate liquidity-trap "trouble" of the NK model is that an increase in price flexibility can make things worse, i.e. be destabilizing; early discussions of this issue include Tobin, 1975 and De Long and Summers, 1986. Eggertsson and Krugman (2012) provide a very clear discussion

and illustration of this property that they dub "the paradox of flexibility". This also holds in our model; an increase in price flexibility summarized by an increase in the Phillips curve slope κ makes the ZLB recession worse:⁴⁷

$$\partial \left(\frac{\partial c_L}{\partial \rho_L} \right) / \partial \kappa = z \left(\frac{1}{1 - z\nu_0} \sigma \frac{1 - \lambda}{1 - \lambda\chi} \right)^2 > 0. \quad (21)$$

The paradox is aggravated—in the sense that the derivative in (21) increases—by adding hand-to-mouth agents if and only if $\chi > 1$ (the proof follows immediately by noticing that both $\sigma \frac{1-\lambda}{1-\lambda\chi}$ and δ , and hence also ν_0 , are increasing with λ iff $\chi > 1$). Conversely, the paradox of flexibility is *mitigated* by hand-to-mouth if and only if their income elasticity is lower than one, i.e. once again condition (9): $\chi < 1$.

6.3 Troubles 3 and 4: Neo-Fisherian Effects, Taylor Principle and Pegs Revisited

The "Neo-Fisherian" view holds that an increase in nominal interest rates can lead to inflation and, with a Phillips curve, also to a real expansion. Cochrane (2017) summarizes and reviews the subject clearly and exhaustively. Benhabib, Schmitt-Grohe, and Uribe (2002) contain an early formalization of such neo-Fisherian effects in a liquidity trap (with flexible and sticky prices), and Schmitt-Grohe and Uribe (2017) a more recent treatment in a model with sticky wages.

In our model, neo-Fisherian effects can be understood as follows. Recall the one difference equation (17) that summarizes the simple HANK model, obtained by combining the aggregate Euler-IS with the PC and Taylor rule). With $\nu > 1$ there are two neo-Fisherian effects: first, *in the long run*, a permanent increase in i^* leads to an increase in consumption and inflation:

$$\bar{c} = \frac{1}{\nu - 1} \sigma \frac{1 - \lambda}{1 - \lambda\chi} i^*;$$

for example under a peg in the standard NK model, $\bar{c} = \kappa^{-1} i^*$.

The other nagging effect is that it also leads (or, strictly speaking, *may* lead) to an expansion and inflation in *the short run*. When ν is larger than 1, the equation cannot be solved forward; we would like to solve it backward, but we have no initial condition to iterate from—the classic problem of indeterminacy. We can still pick one equilibrium by imposing restrictions on the structure of sunspots and on how fundamental uncertainty determines expectation errors: I describe this in detail in the Appendix. Pick such a "reasonable" (minimum-state variable MSV) equilibrium which, assuming persistence p^* for the interest rate shock and picking the solution with the same endogenous persistence (the MSV solution implies we rule out the additional endogenous

⁴⁷In Eggertsson and Krugman this holds more generally (even with pure transitory shocks) through a Fisherian debt-deflation channel that I abstract from here.

persistence that indeterminacy customarily induces) $E_t c_{t+1} = p^* c_t$ we have:

$$c_t = -\frac{1}{1 - \nu p^*} \sigma \frac{1 - \lambda}{1 - \lambda \chi} i_t^*.$$

An increase in interest rates would thus lead to an expansion and inflation (neo-Fisherian effects) whenever:

$$\nu > (p^*)^{-1};$$

under a peg, this reassuringly delivers the same condition we discovered for the occurrence of sunspot-driven liquidity traps, a manifestation of Benhabib, Schmitt-Grohe, and Uribe (2002).

How can we rule out such neo-Fisherian equilibria? The answer is very transparent here: embrace a policy that makes it possible to solve the equation (17) forward, i.e. the Taylor Principle: Proposition 3 applies. The condition:

$$\nu < 1$$

makes it impossible to satisfy $\nu > (p^*)^{-1}$.

What is needed to rule out such nagging effects, instead, in a liquidity trap? It is the same condition that gives determinacy under a peg, that is, the second part of Proposition 3 applies: $\nu_0 < 1$. Notice that this is exactly the same condition needed to solve the FG puzzle!⁴⁸ It seems useful to recall, finally, that once again the key is χ : with fixed prices, the necessary and sufficient condition to rule out neo-Fisherian effects is (9) $\chi < 1$.

6.4 Trouble 5: Inverted AD Logic and a Paradox of Thrift

Having $\chi > 1$ also opens up to a different kind of trouble: in the TANK model, it can make the IS curve swivel and bring the economy to an "inverted Aggregate Demand" region; this is a manifestation of the paradox of thrift described among others in Keynes (1936). The first analysis of this in the TANK model, as well as a detailed discussion of how to rule it is in Bilbiie (2008). Bilbiie and Straub (2012, 2013) study the empirical plausibility of this hypothesis for estimates of the aggregate Euler equation and for explaining the Great Inflation without relying on indeterminacy, respectively.

Recalling the Euler-IS equation of the TANK model (8) abstracting from fiscal shocks, we see that "inverted AD logic" (contractionary interest rate cuts) occurs when the share of hand-to-

⁴⁸This connection has been discussed in a different context by Cochrane (2017), who offers a resolution of neo-Fisherian effects relying upon the fiscal theory of the price level with long-term debt: nominal interest changes the market value and composition of the current portfolio of (long- and short-term) public debt and can lead to deflation. Like the solution discussed in text, this maintains rational expectations. Woodford (2017) also discusses the connection and offers a different resolution that does rely on stepping away from rational expectations, based on the notion of "reflective equilibrium"; other deviations from rational expectations discussed in the context of the FG puzzle would likewise lead to a resolution of this puzzle too.

mouth is beyond a certain threshold:

$$\lambda > \chi^{-1} \longrightarrow \frac{\partial c_t}{\partial (-r_t)} < 0.$$

This is manifestation of the *paradox of thrift*: S households start by wanting to consume more ("save" less) because interest rates go down; but we end up with lower aggregate consumption (aggregate saving goes up). How can that be? The intuition for this inversion is as follows. Consider an *exogenous* fall in real interest rates. The Euler equation of the saver already gives us her consumption: it goes *up*, proportionally. This is true regardless of how many *H* there are: no matter what else happens, it goes **up**—by an amount that is invariant to the number of *H*. The key point is that the income effect of the savers will need to agree with this intertemporal substitution effect, so something else needs to adjust for equilibrium. How can there be a recession following an interest rate cut? Evidently, consumption of *H* must go *down*, which means that the real wage must go down. We need to be moving *downwards* along the labor supply curve, so labor demand shifts down (which in an equilibrium with non-horizontal AS will also give deflation). By how much does it shift down? By as much as necessary to precisely strike the balance between the implied movement in real wage (marginal cost) and hours (and hence sales, output, and ultimately profits), and thus the income effect on savers, on the one hand. And the intertemporal substitution effect that we started off with, on the other hand.

This is a case of the "paradox of thrift", for *individual* incentives to consume more (by savers) lead to equilibrium outcomes with lower *aggregate* consumption. It is very different from the version of the paradox of thrift occurring in a liquidity trap, see e.g. Eggertsson and Krugman (2012): there, the aggregate demand curve is upward-sloping because the nominal interest rate is fixed. Here, it is upward sloping because of aggregation through the mechanism emphasized above, for given interest rates; in particular, there is no need for the zero lower bound to bind.

Can such "paradoxical" equilibria be ruled out? Yes, and the restriction is by now familiar: only if (9) holds. Then, if $\chi < 1$, changes in demand do not trigger over-compensating income effects on savers such as the ones causing the trouble here, no matter how large the share of *H*.

6.5 Rejoinder: The HANK Catch-22

Let us summarize what we now know about amplification and puzzles in NK models, using Table 2.

Table 2: <i>A Catch-22 for HANK</i>	
Amplification	Puzzles
MP, Fiscal mult.; deep recessions w/o deflation	FG; neo-Fisher; flex paradox; Taylor breaks; thrift (flip AD)
YES	worse + more
$\chi > 1$ ↑	$\chi < 1$ (& $\delta < 1$) ↓
no	ALL resolved

Under the condition ($\chi > 1$) delivering amplification with HA (the North-West corner of the Table) NK puzzles are magnified: the FG puzzle is aggravated; neo-Fisherian effects occur even with fixed prices; the paradox of flexibility is worsened. Worse still, new troubles emerge: the Taylor Principle is not sufficient for determinacy; and a paradox of thrift (inverted AD logic) occurs with enough H .

The flip side is, phrased positively, that HANK models can take us to the South-East corner: solve the NK puzzles (FG puzzle, neo-Fisherian effects, paradox of flexibility) and rule out potential additional HANK puzzles. But the necessary condition for this is $\chi < 1$, which also implies dampening—so no amplification (multipliers).

Sufficient conditions for solving the puzzles are provided and discussed at length above, and can be summarized as follows. Solving the FG puzzle and ruling out neo-Fisherian effects requires (in addition to $\chi < 1$) self-insurance to idiosyncratic risk; in addition, $\chi < 1$ is also sufficient if prices are fixed. If prices are not fixed, the upper bound for χ is lower than 1 and depends on price flexibility. $\chi < 1$ is necessary and sufficient to rule out the paradox of thrift in the TANK model, and to make the Taylor principle a sufficient requirement for determinacy.

The good news is thus that the model *can* solve all the puzzles. The bad news is that the condition needed to solve the puzzle is precisely the opposite of the condition needed to deliver amplification, and MP and FP and multipliers. This leaves us with the Catch-22: we can use HANK models to get amplification and multipliers and all these things we may want to explain in the data. But doing so has a high price, for we need to give up everything else: all the properties of the NK model we customarily describe as "puzzles" or paradoxes are aggravated by the same condition. Worse still, some more trouble occurs: new puzzles and paradoxes.

Conversely, we can use HANK models to cure puzzles and paradoxes; but then we need to give up amplification and multipliers—which, if one recalls that we invoke using heterogeneity precisely in order to obtain aggregate amplification beyond intertemporal substitution, is itself a puzzle.

7 HANK Meets DSGE (and the ZLB)

The simple approach to heterogeneity employed here is useful for obtaining all the analytical results above, in closed-form. But it is also useful in order to complicate the model in other

dimensions. In this section, I do this: build an empirically realistic DSGE model à la Smets and Wouters, and in particular a version that is close to Justiniano, Primiceri and Tambalotti (2013), also used by Bilbiie, Monacelli, and Perotti (2016) in the context of optimal government spending in a liquidity trap. I extend this "RANK-DSGE" model by introducing two types of households with idiosyncratic risk, just as above. The model features, in addition to the three-equation model above: consumption habits, investment in physical capital subject to an adjustment cost, sticky wages (I follow Colciago and Ascari Colciago, and Rossi to extend this block of the model to two types), and price and wage indexation. I solve this model under perfect foresight using global methods, taking into account an occasionally binding zero lower bound on interest rates.

The model, outlined in detail in the Appendix, has six endogenous state variables. Evidently, solving the equivalent HANK model with full heterogeneity and six endogenous state variables is, to the best of my knowledge, unfeasible. Thus, the simplified approach to heterogeneity helps build a quantitative model usable for policy analysis and estimation (see Bilbiie and Straub 2013 for a first Bayesian estimation of a TANK model, and Challe et al 2016 for a different model), as well as for optimal policy—exercises that I pursue at the moment in follow-up work. Here, I use the model to illustrate that the two sides of "Catch-22" survive in this quantitatively realistic model: more amplification to demand shocks (liquidity trap and deep recession) and FG puzzle.

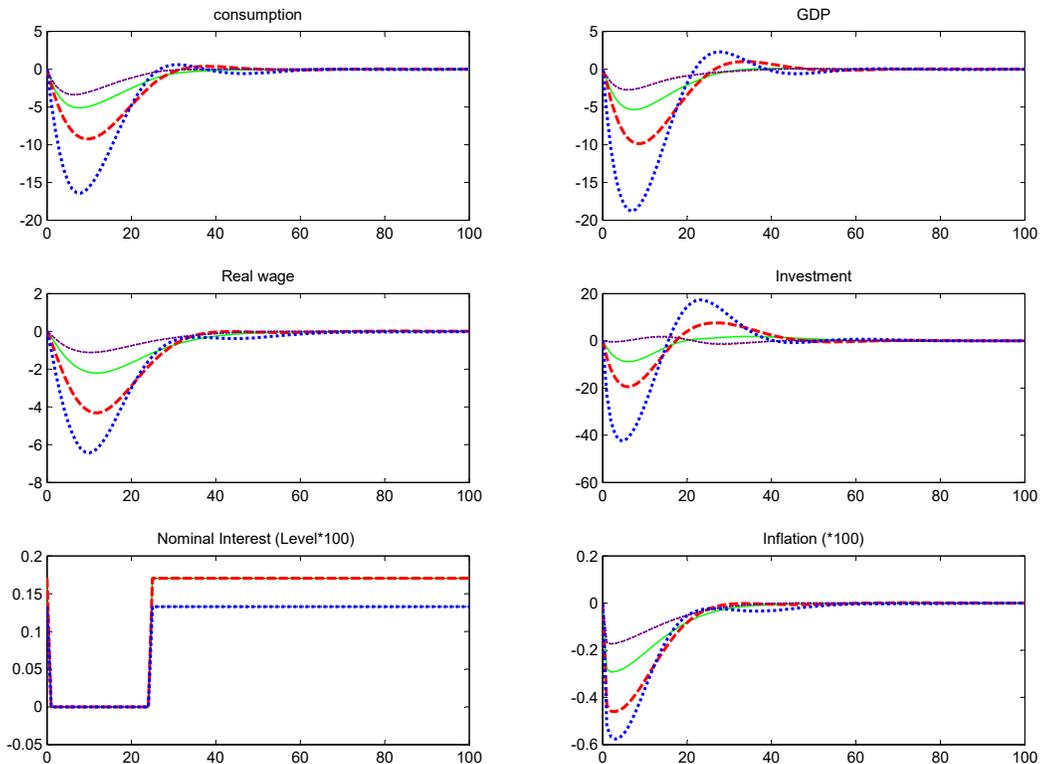


Fig. 8 Amplification vs Dampening: Deep LT recessions in (DSGE-) RANK, TANK, and simple HANK

The first lesson from this exercise, that would apply in any quantitative heterogeneous-agent model with these features, is that a standard calibration (I consider the estimates of Justniano et al) imposes a straightjacket for λ to avoid the "inverted AD thrift paradox" discussed above and in length in Bilbiie (2008) and Bilbiie and Straub (2013); indeed, under that baseline calibration $\lambda = 0.2$ already brings the economy to that "inverted AD" region (even with high wage stickiness).

Figure 8 illustrates the "amplification" part by plotting the response of the economy to a discount factor shock that lasts for 24 quarters under five alternative calibrations of λ , χ and δ . I take a small value of $\lambda = 0.15$ that is in line with the empirical evidence in Bilbiie and Straub (2013) and with the values considered by Kaplan Moll Violante, McKay Nakamura Steinsson, and others. The DSGE-equivalent of parameter χ is in this model a complicated function of all the deep parameters, and in particular wage stickiness in addition to labor elasticity and fiscal redistribution through profit taxation (but also indexation, habits, and so on); I keep all parameters fixed except for the profit tax rate which determines the fiscal redistribution, and consider two alternative values: no redistribution $\tau^D = 0$ ($\chi > 1$) and progressive taxation $\tau^D = 0.2 > \lambda$ (which in the simple HANK model implies $\chi < 1$). Finally, I consider both cases of no idiosyncratic risk, i.e. TANK ($s = 0$ and $\delta = 1$) and a small degree of idiosyncratic risk of 2 percent per annum ($s = 0.005$). This gives five calibrations RANK-DSGE (green solid), TANK with $\chi > 1$ (thick red dash), TANK with $\chi < 1$ (thin red dash), HANK with $\chi > 1$ (thick blue dots) and HANK with $\chi < 1$ (thin blue dots).

We recover the same conclusions as in the simple model: there is more amplification under the TANK model (the trough of the recession is double that of RANK), and even more so when adding idiosyncratic risk (the trough is doubled again), but only for " $\chi > 1$ "; in the other region, there is dampening instead of amplification and, because the values of both λ and idiosyncratic risk $1 - s$ are small, the latter does not make much difference.

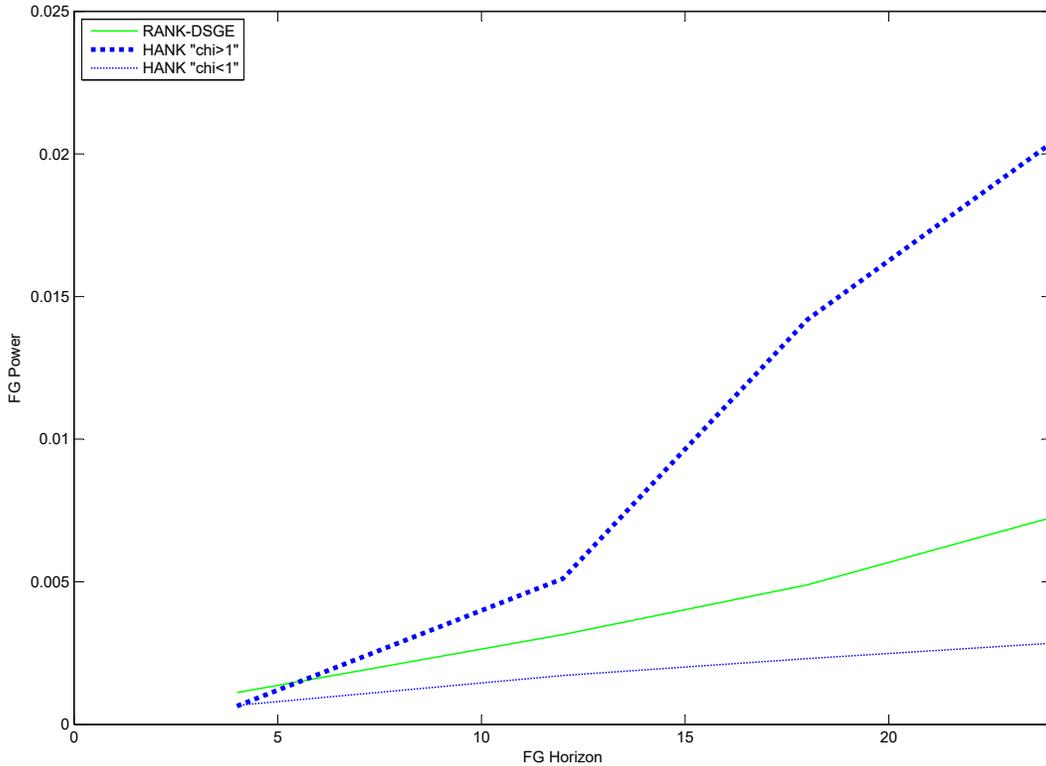


Fig. 9 FG Power and Puzzle in (DSGE-) RANK and simple HANK

Figure 9 illustrates the "puzzles" part; it plots the power of FG, defined much like in the simple model, as the effect of a 1% (annualized) cut in interest rates at a certain (varying) horizon, but this time on the present discounted value of consumption. The green solid line illustrates the "FG puzzle": the power in increasing with the horizon of the interest rate cut. In the HANK $\chi < 1$ case (with $\tau = 0.2$) there is dampening but not enough to make the power decreasing with the horizon. While in the HANK $\chi > 1$ case, the puzzle is aggravated: FG power increases with the horizon.

8 Conclusions

What can we learn about the workings of modern heterogeneous-agent (HANK) models through simplified versions based on earlier two-agent (TANK) models? This paper starts by a New Keynesian cross, centered on a planned expenditure PE curve, that captures aggregate demand in this class of models. The slope of PE is the indirect effect share (the part of the total effect that is due to the endogenous response of output and income), and its shift in response to either monetary or fiscal policy is the direct effect.

This representation unveils an important amplification mechanism when hand-to-mouth households' income is endogenous and responds to aggregate income more than one-to-one: the more constrained agents there are, the higher the slope of the planned expenditure line, and the larger the "multipliers" (the effects of monetary and fiscal policies, and of demand shocks). Such amplification is thus driven by the indirect effect (in Kaplan, Moll and Violante's terminology). The mechanism is overturned—and thus there is dampening instead of amplification—when the income of hand-to-mouth agents responds to the cycle less than one-to-one.⁴⁹ Whether that key elasticity (of hand-to-mouth income to the cycle) is larger or smaller than one depends chiefly on the details of the labor market (how much of an aggregate expansion goes to labor income) and on fiscal redistribution (how progressive is the tax system). This is an example of a more general insight on the role of redistribution, which has its origins in Keynes' General Theory as illustrated by the quotes: that the marginal propensity to consume (the slope of planned expenditure in the old Keynesian cross) depends on the distribution of income.

Adding a self-insurance channel (another defining feature of HANK models), magnifies these effects further. When income of hand-to-mouth responds to aggregate income less than proportionally, there is further dampening through discounting in the Euler equation (of the type first identified in this type of models by McKay, Nakamura and Steinsson). But when income of hand-to-mouth responds more than one-to-one to the cycle, amplification is magnified—there is now *compounding* in the aggregate Euler equation, for future aggregate expansions imply an incentive to dis-save (less self-insurance) and thus consume disproportionately more today.⁵⁰ While this further amplification does not change the effects of transitory, unanticipated policy shocks significantly (relative to a TANK model with only the hand-to-mouth channel), it does imply increasingly different effects as the persistence of aggregate shocks increases—because this is when self-insurance matters most. The starkest difference pertains to the effects of announcements of future policy changes (such as forward guidance). It is worth emphasizing the logic for fiscal transfers: the (positive) multiplier found in the absence of idiosyncratic risk is now amplified even in the "discounting" case.

Overall, the first part of the analysis elucidates how the one key requirement for obtaining monetary and fiscal multipliers in heterogeneous-agent models where some households are hand-to-mouth is that the elasticity of their income to aggregate be larger than one. The second part of the paper shows the dark side of this: the same condition that delivers multipliers also multiplies trouble; several well-known "puzzles" of the NK model are aggravated, and new ones occur. The

⁴⁹For fiscal multipliers this is strictly true with spending financed via uniform taxation. But there is a qualification: multipliers can still occur with $\chi < 1$ when spending is financed with progressive taxation, because of an implicit transfer.

⁵⁰This captures in a simple and intuitive way (through one parameter in the aggregate Euler equation) an amplification mechanism that holds in more general HANK models (and was also emphasized in a more general framework by Werning, 2015). This also clarifies the connection between that mechanism and the earlier two-agent analyses: the very same condition that delivers amplification in a TANK model also delivers compounding with self-insurance.

former category includes the forward-guidance puzzle, neo-Fisherian effects, and the paradox of flexibility; the latter includes the breakup of the Taylor principle and the inversion of aggregate demand logic—a version of another Keynesian topic: the paradox of thrift.

To look at the landscape differently: heterogeneous-agent NK models can help resolve the existing and avoid the novel puzzles of the NK model; the second part of the paper derives necessary and sufficient conditions for this. But the common denominator for having a puzzle-free NK model turns out to be a necessary condition that income of constrained hand-to-mouth vary with aggregate income less than one-to-one: that is the exact opposite of the requirement for multipliers. Hence the Catch-22: can there be amplification without puzzles in the NK model?

Two solutions to this Cornelian dilemma appear, but they are orthogonal to the model structure: they consist in adding an additional ingredient. The first is about doing something to reduce the δ without reducing the χ , and the second is about increasing the χ without reducing the δ . In particular, the former consists of deviating from either rational expectations, or from perfect information; some version of either is explored in several recent papers discussed in text, and has proven successful at solving the first two puzzles discussed here. But those resolutions rely on some form of Euler-equation discounting and hence do not (by themselves) lead to multipliers and amplification—such models need something like the hand-to-mouth channel in addition for delivering within-the-period amplification.⁵¹

The latter solution relies on yet another ingredient that does the opposite: it increases the within-the-period amplification without affecting the intertemporal. In particular, multipliers occur even in the "dampening" region of a heterogeneous-agent NK model ($\delta < 1$) if household preferences feature complementarity between consumption and hours, as in Bilbiie (2011)—e.g. GHH preferences also considered by Monacelli and Perotti (2008) in the context of fiscal multipliers. This creates a different (preference-based) feedback loop between income and output: any demand shock that leads to an increase in income also leads to an increase in hours worked and output if the cross-derivative between consumption and hours is positive.

Finally, even within the context of the model discussed here and with no extra ingredients, there does exist one policy discussed throughout the paper that delivers positive multipliers and amplification: transfers to the hand-to-mouth financed by taxing savers. Perhaps the conclusion of HANK-type models in this respect is that these models are useful for resolving puzzles (that the relevant region is $\chi < 1$), and that indirect effects are important but without amplification. While the policy conclusion would be that the only stimulus that "works" are transfers, or public debt:⁵² a redistributive transfer is expansionary even when labor market structure and endogenous redistribution (automatic stabilizers) are such that the elasticity of hand-to-mouth income

⁵¹Other puzzle resolutions that do not relax rational expectations or perfect information, such as Cochrane (2017) or Diba and Loisel (2017) may also deliver multipliers—but those studies do not focus on this question.

⁵²That the latter is equivalent to (an intertemporal version of) the former is shown in a TANK-type model by Bilbiie, Monacelli, and Perotti (2013).

to aggregate is lower than unity, if its demand effect through the Keynesian cross nevertheless dominates its negative income effect (if $\chi > \alpha/\lambda$). Such stimulus though raises obvious political economy issues and the need for a careful study of optimal policy.

Either way, the analysis suggests that three parameters are paramount for assessing the effects of monetary and fiscal policies, those providing the abbreviation of the simple model: $\lambda\chi\delta$. How many households are liquidity-constrained? How does their income vary with aggregate income, and how does that depend on labor market structure and automatic stabilizers? And do households self-insure in face of idiosyncratic risk? All these questions are of utmost importance for designing macro policies; the answer to the second is the keystone and one hopes that empirical research will soon provide it.⁵³

References

- [1] Achdou, Y., J. Han, J.-M. Lasry, P.-L. Lions, and B. Moll, 2017, "Income and Wealth Distribution in Macroeconomics: A Continuous-Time Approach," Mimeo Université Pierre et Marie Curie
- [2] Aiyagari, R., 1994, "Uninsured Idiosyncratic Risk and Aggregate Saving," *Quarterly Journal of Economics*, 109 (1994), 659–684.
- [3] Alvarez F., Lucas Jr., R. E., Weber W., 2001, "Interest Rates and Inflation," *American Economic Review*, Vol. 91(2), 219-225
- [4] Angeletos, G.M. and C. Lian, 2017, Forward Guidance without Common Knowledge, Mimeo MIT
- [5] Ascari, G., A. Colciago and L. Rossi, 2017, "Limited Asset Market Participation, Sticky Wages, and Monetary Policy", forthcoming *Economic Inquiry*
- [6] Auclert, A. 2016, "Monetary Policy and the Redistribution Channel" Mimeo Stanford
- [7] Auerbach, A. and Y. Gorodnichenko, 2013. "Fiscal Multipliers in Recession and Expansion." In *Fiscal Policy After the Financial Crisis*, edited by Alberto Alesina and Francesco Giavazzi, pp. 63–98. University of Chicago Press
- [8] Benhabib, J., Schmitt-Grohe, S., and Uribe, M., 2002, "Avoiding Liquidity traps", *Journal of Political Economy*
- [9] Bewley, T., 1977, "The Permanent Income Hypothesis: A Theoretical Formulation," *Journal of Economics Theory*, 16, 252–292.

⁵³Recent research by Cloyne, Ferreira, and Surico (2017) seems to support the transmission mechanism discussed in this paper while suggesting that the more plausible scenario is $\chi < 1$.

- [10] Bilbiie F. 2008. "Limited Asset Market Participation, Monetary Policy, and (Inverted) Aggregate Demand Logic". *Journal of Economic Theory* 140, 162-196.
- [11] Bilbiie, F. 2009 " Non-Separable Preferences, Fiscal Policy Puzzles and Inferior Goods, *Journal of Money, Credit and Banking*, 41(2-3), pp 443-450.
- [12] Bilbiie F. 2011. "Non-Separable Preferences, Frisch Labor Supply and the Consumption Multiplier of Government Spending: One Solution to a Fiscal Policy Puzzle, *Journal of Money, Credit and Banking*, 43(1), 221-251
- [13] Bilbiie, F., Mueller G. and A. Meier 2008, "What Accounts for the Change in U.S. Fiscal Policy Transmission?", *Journal of Money, Credit and Banking*, 2008
- [14] Bilbiie F. and R. Straub, 2004, "Fiscal Policy, Labor Markets, and Business Cycle Fluctuations," MNB Working Paper
- [15] Bilbiie F. and R. Straub, 2012, "Changes in the Output Euler Equation and Asset Markets Participation" *Journal of Economic Dynamics and Control* 36(11), 1659-1672
- [16] Bilbiie F. and R. Straub, 2013, "Asset Market Participation, Monetary Policy Rules and the Great Inflation", *Review of Economics and Statistics*, 95(2), pp 377-392
- [17] Bilbiie, F., T. Monacelli, and R. Perotti 2012, "Tax Cuts vs. Spending Increases: Welfare at the Zero Lower Bound" Mimeo PSE and Bocconi
- [18] Bilbiie F., Monacelli T. and R. Perotti 2013, "Public Debt and Redistribution with Borrowing Constraints", *The Economic Journal* 2013, vol. 0, pages F64-F98, 02
- [19] Bilbiie, F. 2016, "Optimal Forward Guidance", CEPR DP
- [20] Bilbiie, F. and X. Ragot, 2016, "Optimal Monetary Policy and Liquidity with Heterogeneous Households", CEPR DP
- [21] Bilbiie, 2017, "The New Keynesian Cross: Understanding Monetary Policy with Hand-to-Mouth Households," CEPR DP
- [22] Blanchard, O. and R. Perotti, 2002, "An Empirical Characterization of the Dynamic Effects of Changes in Government Spending and Taxes on Output", *Quarterly Journal of Economics*, 117(4), 1329-1368
- [23] Bricker, J., L. Dettling, A. Henriques, J. Hsu, K. Moore, J. Sabelhaus, J. Thompson, and R. Windle. 2014. "Changes in U.S. Family Finances from 2010 to 2013: Evidence from the Survey of Consumer Finances," *Federal Reserve Bulletin*, 100(4).

- [24] Broer, T., N. J. Hansen, P. Krusell, and E. Oberg, "The New Keynesian transmission mechanism: a heterogeneous agent perspective", Mimeo IIES Stockholm
- [25] Campbell, J. Y. and Mankiw, N. G. 1989. "Consumption, Income, and Interest Rates: Reinterpreting the Time Series Evidence," in Blanchard, O. and Fisher, S., eds, NBER Macroeconomics Annual 185-216, MIT Press.
- [26] Campbell, J. Y. and Mankiw, N. G. 1990, "Permanent Income, Current Income, and Consumption" *Journal of Business and Economic Statistics*, Vol. 8, No. 3
- [27] Campbell, J. Y. and Mankiw, N. G. 1991, "The response of consumption to income A cross-country investigation", *European Economic Review*
- [28] Challe, E. and X. Ragot, 2014 "Precautionary Saving Over the Business Cycle," *The Economic Journal*
- [29] Challe, E, J. Matheron, X. Ragot and J. Rubio-Ramirez, 2016, "Precautionary Saving and Aggregate Demand", *Quantitative Economics*
- [30] Christiano L., M. Eichenbaum and S. Rebelo (2011). "When Is the Government Spending Multiplier Large?," *Journal of Political Economy*, 119(1), pages 78 - 121.
- [31] Clarida, R., J. Galí, and M. Gertler, 1999, 'The science of monetary policy: a New Keynesian perspective', *Journal of Economic Literature*, 37, 1661-1707.
- [32] Cloyne, J., C. Ferreira and Surico, P., , 2015 "Monetary Policy when Households have Debt: New Evidence on the Transmission Mechanism" *CEPR Discussion Paper*
- [33] Cochrane, J. 2011 "Determinacy and Identification with Taylor Rules," *Journal of Political Economy*, 119 (3), pp. 565-615
- [34] Cochrane, J. 2016, "The New-Keynesian Liquidity Trap", *Forthcoming, Journal of Monetary Economics*
- [35] Cochrane, J., 2017 "Michelson-Morley, Fisher and Occam: The Radical Implications of Stable Inflation at the Zero Bound," *Mimeo*.
- [36] Colciago, A. 2011, Rule-of-thumb consumers meet sticky wages, *Journal of Money Credit and Banking* 43(2), 325-353.
- [37] Curdia, V. and M. Woodford 2009, "Credit Frictions and Optimal Monetary Policy," *Journal of Monetary Economics*, forthcoming
- [38] Davig, T. and E. Leeper 2011, "Monetary-fiscal policy interactions and fiscal stimulus," *European Economic Review* 55 (2), 211-227

- [39] De Long, J. B., and L. H. Summers, 1986 "Is Increased Price Flexibility Stabilizing?" *American Economic Review*, 76 (1986), 1031–1044.
- [40] Debortoli, D., and J. Galí, 2017, "Monetary Policy with Heterogeneous Agents: Insights from TANK models"; Mimeo CREI
- [41] Del Negro, M., M. Giannoni, and C. Patterson 2012: "The Forward Guidance Puzzle," Federal Reserve Bank of New York
- [42] Den Haan, W. J., 1997, "Solving dynamic models with aggregate shocks and heterogeneous agents," *Macroeconomic Dynamics* 1 (2), 355-386
- [43] Den Haan, W. J.; P. Rendahl, and M. Riegler 2016, "Unemployment (Fears) and Deflationary Spirals" Mimeo LSE
- [44] Eggertsson G. and M. Woodford 2003, "The Zero Bound on Interest Rates and Optimal Monetary Policy", *Brookings Papers on Economic Activity*, No. 1.
- [45] Eggertsson, G. 2010, "What Fiscal Policy Is Effective at Zero Interest Rates?", in: D. Acemoglu and M. Woodford, eds: *NBER Macroeconomics Annual 2010*, University of Chicago Press
- [46] Eggertsson G. and P. Krugman 2012, "Debt, Deleveraging, and the Liquidity Trap: A Fisher-Minsky-Koo Approach," *Quarterly Journal of Economics*, 127(3): 1469-1513.
- [47] Farhi, E. and I. Werning, 2017, "Monetary policy with bounded rationality and incomplete markets"
- [48] Ferrière, A. and G. Navarro, 2017 "The Heterogeneous Effects of Government Spending: It's All About Taxes," Mimeo EUI and Fed Board
- [49] Furlanetto, F. 2011 "Fiscal Stimulus and the Role of Wage Rigidity" *Journal of Economic Dynamics and Control* 35, 512-527
- [50] Gabaix, X., 2016, "A Behavioral New Keynesian Model," Mimeo Harvard
- [51] Galí, J. 1991, "Budget Constraints and Time Series Evidence on Consumption," *American Economic Review*, vol. 81, no. 5, 1991, 1238-1253
- [52] Galí, J. 2008, "Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework and Its Applications", Princeton University Press
- [53] Galí, J., D. López-Salido, and J. Vallés, 2007. "Understanding the Effects of Government Spending on Consumption". *Journal of the European Economic Association*, March, vol. 5 (1), 227-270.

- [54] Garcia-Schmidt, M. and M. Woodford, 2014. "Are Low Interest Rates Deflationary? A Paradox of Perfect Foresight Analysis," Mimeo Columbia.
- [55] Giambattista, M. and S. Pennings, 2017, "When is the Government Transfer Multiplier Large?" *European Economic Review*
- [56] Gornemann, N., K. Kuester, and M. Nakajima, 2012 "Monetary Policy with Heterogeneous Agents," Working Papers 12-21, Federal Reserve Bank of Philadelphia
- [57] Guerrieri, V. and G. Lorenzoni, 2017, "Credit Crises, Precautionary Savings, and the Liquidity Trap" Forthcoming, *Quarterly Journal of Economics*
- [58] Hagedorn, M., 2017, "A Demand Theory of the Price Level", Mimeo
- [59] Hagedorn, M., I. Manovskii, and K. Mitman, 2017, "The Fiscal Multiplier," Mimeo
- [60] Hall, R. E., 1978, "Stochastic implications of the life cycle-permanent income hypothesis: Theory and evidence", *Journal of Political Economy* 86, 971-987.
- [61] Hall, R. E., 1988, "Intertemporal Substitution in Consumption," *Journal of Political Economy*, 96(2), 339-57.
- [62] Hall, R. E., 2009, "By How Much Does GDP Rise if the Government Buys More Output," *Brookings Papers on Economic Activity*, 2, 183-231
- [63] Hall, R. E. 2011, "The Long Slump," *American Economic Review*, 101, 431-469
- [64] Heathcote, J., K. Storesletten, and G. Violante, 2017, "Optimal Tax Progressivity: An Analytical Framework" Forthcoming, *Quarterly Journal of Economics*
- [65] Huggett, M., 1993, "The Risk-Free Rate in Heterogeneous-Agent Incomplete-Insurance Economies," *Journal of Economic Dynamics and Control*, 17, 953-969.
- [66] Hurst, E., 2004, Grasshoppers, Ants and Pre-Retirement Wealth: A Test of Permanent Income Consumers, NBER Working Paper 10098.
- [67] Iacoviello, M. 2005, "House Prices, Borrowing Constraints, and Monetary Policy in the Business Cycle" *American Economic Review*
- [68] Jappelli, T and Pistaferri, L, 2014, "Fiscal Policy and MPC Heterogeneity", *American Economic Journal: Macroeconomics*, 6(4), 107-36
- [69] Johnson, D., Parker J. and Souleles N., 2006. "Household Expenditure and the Income Tax Rebates of 2001." *American Economic Review*, 96(5): 1589-1610.

- [70] Justiniano, A., G. Primiceri, and A. Tambalotti, 2013, "Is There a Trade-Off Between Inflation and Output Stabilization?" *American Economic Journal: Macroeconomics*, 5(2),1-31
- [71] Keynes, J.M. 1936 "The General Theory of Employment, Interest and Money", MacMillan, Cambridge.
- [72] Kaplan, G. and G. Violante, 2014 "A Model of the Consumption Response to Fiscal Stimulus Payments," *Econometrica*
- [73] Kaplan, G., G. Violante, and J. Weidner, 2014 "The Wealthy Hand-to-Mouth," *Brookings Papers on Economic Activity*.
- [74] Kaplan, G., B. Moll and G. Violante, 2015, "Monetary Policy According to HANK", Mimeo Princeton and Chicago
- [75] Kaplan, G., B. Moll and G. Violante, 2016, "A Note on Unconventional Monetary Policy in HANK", Mimeo Princeton and Chicago
- [76] Krusell, P. and A. Smith. 1998. "Income and wealth heterogeneity in the macroeconomy." *Journal of Political Economy* 106(5), 867-896.
- [77] Krusell, P., T. Mukoyama, and A. Smith. 2011. "Asset Prices in a Huggett Economy," *Journal of Economic Theory*, 146 (3), 812-844
- [78] Leeper, E., 1991, "Equilibria Under "Active" and "Passive" Monetary and Fiscal Policies", *Journal of Monetary Economics*
- [79] Leeper, E. and C. Leith, 2016, "Understanding Inflation as a Joint Monetary-Fiscal Phenomenon", *Handbook of Macroeconomics*, vol. 2B, John B. Taylor and Harald Uhlig, eds., Elsevier, 2305–2416
- [80] Leeper, E., N. Traum, and T. Walker, 2017, "Clearing Up the Fiscal Multiplier Morass" *Forthcoming American Economic Review*
- [81] Lucas R. E. Jr., 1990 "Liquidity and Interest Rates." *Journal of Economic Theory* 50, 237 - 264.
- [82] Mankiw, N. G. 2000. "The Savers-Spenders Theory of Fiscal Policy," *American Economic Review*, Vol. 90 (2) pp. 120-125.
- [83] McKay, A. and R. Reis, 2016 "The Role of Automatic Stabilizers in the U.S. Business Cycle," *Econometrica*, Vol. 84, 1
- [84] McKay, A., E. Nakamura, and J. Steinsson 2015: "The Power of Forward Guidance Revisited," *American Economic Review*

- [85] McKay, A., E. Nakamura and J. Steinsson 2016: "The Discounted Euler Equation", *Economica*
- [86] Mehrotra, N., 2017, "Fiscal Policy Stabilization: Purchases or Transfers?," Forthcoming *International Journal of Central Banking*
- [87] Mertens K. and M. Ravn 2013. "Fiscal Policy in an Expectations Driven Liquidity Trap," *Review of Economic Studies*
- [88] Misra, K. and P. Surico, 2014, "Consumption, Income Changes, and Heterogeneity: Evidence from Two Fiscal Stimulus Programs" *American Economic Journal: Macroeconomics*, 6(4)
- [89] Monacelli T. and R. Perotti 2008, "Fiscal Policy, Wealth Effects, and Markups", NBER Working Paper No. 14584
- [90] Monacelli, and R. Perotti 2012, "Redistribution and the Multiplier," *IMF Economic Review*
- [91] Nakamura, E. and J. Steinsson, 2014. "Fiscal Stimulus in a Monetary Union: Evidence from US Regions," *American Economic Review*, 104(3), 753-792
- [92] Nistico, S., 2015, "Optimal monetary policy and financial stability in a non-Ricardian economy." Forthcoming, *Journal of the European Economic Association*
- [93] Oh, H. and R. Reis, 2012, "Targeted transfers and the fiscal response to the great recession." *Journal of Monetary Economics*, 59(S), S50-S64
- [94] Parker, J.A., 1999. "The Reaction of Household Consumption to Predictable Changes in Social Security Taxes," *American Economic Review*, vol. 89(4), 959-973
- [95] Parker, J. A., N. S. Souleles, D. S. Johnson and R. McClelland. 2013. "Consumer Spending and the Economic Stimulus Payments of 2008." *American Economic Review*, 103(6): 2530-53.
- [96] Perotti, R., 2007, "In Search of the Transmission Mechanism of Fiscal Policy," *NBER Macroeconomics Annual 2007*, Volume 22.
- [97] Ramey, V. and S. Zubairy, 2017, "Government Spending Multipliers in Good Times and in Bad: Evidence from U.S. Historical Data," Forthcoming *Journal of Political Economy*
- [98] Ravn, M. and V. Sterk, 2017 "Job Uncertainty and Deep Recessions," Forthcoming, *Journal of Monetary Economics*.
- [99] Ravn, M. and V. Sterk, 2016, "Macroeconomic Fluctuations with HANK and SAM: an analytical approach", mimeo
- [100] Schmitt-Grohe, S., and Uribe, M., 2017, "Liquidity Traps And Jobless Recoveries," *American Economic Journal: Macroeconomics* 9, 165-204

- [101] Surico, P. and R. Trezzi, 2016, "Consumer Spending and Fiscal Consolidation: Evidence from a Housing Tax Experiment", Mimeo LBS and FRB
- [102] Tobin, J., 1975 "Keynesian Models of Recession and Depression," *American Economic Review*, 65, 195–202
- [103] Vissing-Jorgensen, A. 2002 "Limited Asset Market Participation and the Elasticity of Intertemporal Substitution' *Journal of Political Economy*
- [104] Walsh, C., 2008, "Monetary Theory and Policy", MIT Press
- [105] Walsh, C., 2016, "Workers, Capitalists, Wages, Employment and Welfare", Mimeo UCSC
- [106] Werning, I., 2015, "Incomplete markets and aggregate demand", Mimeo MIT
- [107] Wiederholt, M. 2015, "Empirical Properties of Inflation Expectations at the Zero Lower Bound," Mimeo Goethe University Frankfurt.
- [108] Wolff, E. and Caner, A. 2002, "Asset Poverty in the United States, 1984-1999: Evidence from the Panel Study of Income Dynamics," WP 356, The Levy Institute
- [109] Woodford, M., 1996, "Controlling the public debt: a requirement for price stability?" Mimeo
- [110] Woodford, M., 2003, *Interest and prices: foundations of a theory of monetary policy*, Princeton University Press
- [111] Woodford, M. 2011. "Simple Analytics of the Government Expenditure Multiplier," *American Economic Journal: Macroeconomics*, 3(1), p 1-35
- [112] World Bank, 2014, *Measuring Financial Inclusion around the World*", The Global Findex Database
- [113] Yogo Y., 2004, "Estimating the Elasticity of Intertemporal Substitution When Instruments Are Weak," *Review of Economic and Statistics*, August 2004, 86(3): 797-810.

A Intertemporal Budget Constraint, Euler Equation, and Consumption Function

An agent j chooses consumption, asset holdings and leisure solving the standard intertemporal problem: $\max E_0 \sum_{t=0}^{\infty} \beta^t U(C_{j,t}, N_{j,t})$, subject to the sequence of constraints:

$$B_{j,t} + \Omega_{j,t+1} V_t \leq Z_{j,t} + \Omega_{j,t} (V_t + P_t D_t) + W_t N_{j,t} - P_t C_{j,t}.$$

$C_{j,t}, N_{j,t}$ are consumption and hours worked, $B_{j,t}$ is the nominal value at *end of period* t of a portfolio of all state-contingent assets held, except for shares in firms—likewise for $Z_{j,t}$, beginning of period wealth.⁵⁴ V_t is average market value at time t of shares, D_t their real dividend payoff and $\Omega_{j,t}$ are share holdings. Absence of arbitrage implies that there exists a stochastic discount factor $Q_{t,t+1}$ such that the price at t of a portfolio with uncertain payoff at $t+1$ is (for state-contingent assets and shares respectively):

$$\frac{B_{j,t}}{P_t} = E_t \left[Q_{t,t+1} \frac{Z_{j,t+1}}{P_{t+1}} \right] \text{ and } \frac{V_t}{P_t} = E_t \left[Q_{t,t+1} \left(\frac{V_{t+1}}{P_{t+1}} + D_{t+1} \right) \right], \quad (22)$$

which iterated forward gives the fundamental pricing equation: $\frac{V_t}{P_t} = E_t \sum_{i=1}^{\infty} Q_{t,t+i} D_{t+i}$. The riskless gross short-term REAL interest rate R_t is a solution to:

$$\frac{1}{R_t} = E_t Q_{t,t+1} \quad (23)$$

Note for nominal assets we have the nominal interest rate $\frac{1}{I_t} = E_t \frac{P_t}{P_{t+1}} Q_{t,t+1}$

Substituting the no-arbitrage conditions (22) into the wealth dynamics equation gives the flow budget constraint. Together with the usual 'natural' no-borrowing limit for *each* state, and anticipating that in equilibrium all agents will hold a constant fraction of the shares (there is no trade in shares) Ω_j (whose integral across agents is 1), this implies the usual intertemporal budget constraint:

$$E_t \left[\frac{P_t}{P_{t+1}} Q_{t,t+1} X_{j,t+1} \right] \leq X_{j,t} + W_t N_{j,t} - P_t C_{j,t}.$$

$$\begin{aligned} X_{j,t} &= Z_{j,t} + \Omega_j (V_t + P_t D_t) \\ &= Z_{j,t} + \Omega_j \left(E_t \sum_{i=0}^{\infty} P_t Q_{t,t+i} D_{t+i} \right) \end{aligned}$$

⁵⁴We distinguish shares from the other assets explicitly since their distribution plays a crucial role in the rest of the analysis.

$$\begin{aligned}
E_t \sum_{i=0}^{\infty} Q_{t,t+i} C_{j,t+i} &\leq \frac{X_{j,t}}{P_t} + E_t \sum_{i=0}^{\infty} Q_{t,t+i} \frac{W_{t+i}}{P_{t+i}} N_{j,t+i} \\
&= E_t \sum_{i=0}^{\infty} Q_{t,t+i} Y_{j,t+i}
\end{aligned} \tag{24}$$

where

$$Y_{j,t+i} = \Omega_j D_{t+i} + \frac{W_{t+i}}{P_{t+i}} N_{j,t+i} \tag{25}$$

is income of agent j . Maximizing utility subject to this constraint gives the first-order necessary and sufficient conditions at each date and in each state:

$$\beta \frac{U_C(C_{j,t+1})}{U_C(C_{j,t})} = Q_{t,t+1}$$

along with (24) holding with equality (or alternatively flow budget constraint holding with equality and transversality conditions ruling out Ponzi games be satisfied: $\lim_{i \rightarrow \infty} E_t [Q_{t,t+i} Z_{j,t+i}] = \lim_{i \rightarrow \infty} E_t [Q_{t,t+i} V_{t+i}] = 0$). Using (24) and the functional form of the utility function the short-term nominal interest rate must obey:

$$\frac{1}{R_t} = \beta E_t \left[\frac{U_C(C_{j,t+1})}{U_C(C_{j,t})} \right].$$

A.1 Loglinearized equilibrium

Denote by small letter log deviations from SS. Notice

$$Q_{t,t+i} = \beta^i \frac{U_C(C_{j,t+i})}{U_C(C_{j,t})}$$

and in SS: $Q_i = \beta^i$. Thus we have

$$q_{t,t+i} = \ln \frac{Q_{t,t+i}}{Q_i} = \ln \frac{U_C(C_{j,t+i})}{U_C(C_{j,t})} = -\gamma (c_{t+i} - c_t),$$

where we notice

$$q_{t,t+i} = q_{t,t+1} + q_{t+1,t+2} + \dots + q_{t+i-1,t+i}$$

The Euler equation is:

$$r_t = -E_t q_{t,t+1}$$

Rewrite, with $\sigma = \gamma^{-1}$

$$c_t = E_t c_{t+1} - \sigma r_t$$

and iterate forward, using $q_{t,t+i} = -\sum_{k=0}^{i-1} r_{t+k}$

$$c_t = E_t c_{t+i} + \sigma E_t q_{t,t+i} \tag{26}$$

Now loglinearize intertemporal budget constraint

$$E_t \sum_{i=0}^{\infty} \beta^i (q_{t,t+i}^j + c_{t+i}^j) = E_t \sum_{i=0}^{\infty} \beta^i (q_{t,t+i}^j + y_{t+i}^j)$$

Add to each side $(\sigma - 1) \sum_{i=0}^{\infty} \beta^i E_t q_{t,t+i}^j$

$$E_t \sum_{i=0}^{\infty} \beta^i (\sigma q_{t,t+i}^j + c_{t+i}^j) = E_t \sum_{i=0}^{\infty} \beta^i (\sigma q_{t,t+i}^j + y_{t+i}^j)$$

By virtue of the Euler equation the LHS simplifies

$$\frac{1}{1-\beta} c_t^j = \sigma \sum_{i=0}^{\infty} \beta^i E_t q_{t,t+i}^j + \sum_{i=0}^{\infty} \beta^i E_t y_{t+i}^j$$

Develop RHS, use $q_{t,t} = 0$

$$\begin{aligned} \sum_{i=0}^{\infty} \beta^i E_t q_{t,t+i}^j &= 0 - \sum_{i=1}^{\infty} \beta^i E_t \sum_{k=0}^{i-1} r_{t+k} \\ &= -\frac{\beta}{1-\beta} \sum_{i=0}^{\infty} \beta^i E_t r_{t+i} \end{aligned}$$

And replace to obtain (multiplying by $1 - \beta$)

$$\begin{aligned} c_t^j &= -\sigma \beta \sum_{i=0}^{\infty} \beta^i E_t r_{t+i} + (1-\beta) \sum_{i=0}^{\infty} \beta^i E_t y_{t+i}^j \\ &= -\sigma \beta r_t + (1-\beta) y_t^j - \sigma \beta \sum_{i=1}^{\infty} \beta^i E_t r_{t+i} + (1-\beta) \sum_{i=1}^{\infty} \beta^i E_t y_{t+i}^j \end{aligned}$$

Now replace expression for expected consumption tomorrow

$$\beta c_{t+1}^j = -\sigma \beta \sum_{i=0}^{\infty} \beta^{i+1} E_t r_{t+1+i} + (1-\beta) \sum_{i=0}^{\infty} \beta^{i+1} E_t y_{t+1+i}^j$$

to obtain the consumption function in text 1.

B More general fiscal redistribution in TANK, and χ

Consider a different redistribution scheme, as follows. An arbitrary sales subsidy at rate x is financed by arbitrarily distributed lump-sum taxes T_x^H and T_x^S . Pricing with the arbitrary subsidy delivers the steady-state real wage $w = \frac{1+x}{1+\mu} \leq 1$. The government thus spends xY in subsidy and

needs to gather $T_x = xY$ taxes. Taxes per agent to pay for this subsidy are thus distributed as

$$T_x^H = \frac{\theta}{\lambda}xY \text{ and } T_x^S = \frac{1-\theta}{1-\lambda}xY$$

with θ the share of taxes paid by H. Assume that there is also an arbitrary redistribution ϱ for each agent H financed by taxing S agents $-\frac{\lambda}{1-\lambda}\varrho$. The steady-state consumption shares for each agent are thus respectively, denoting the share of redistribution in total consumption by $\rho = \varrho/C$:

$$\begin{aligned} \frac{C^H}{C} &= \frac{1+x}{1+\mu} - \frac{\theta}{\lambda}x + \rho \\ \frac{C^S}{C} &= \frac{1+x}{1+\mu} \left(1 + \frac{\mu}{1-\lambda}\right) - \frac{1-\theta}{1-\lambda}x - \frac{\lambda}{1-\lambda}\rho \end{aligned}$$

This arbitrary redistribution nests the one assumed in the baseline when $x = \mu$ and all taxes to pay this are paid by savers $\theta = 0$. But perfect redistribution can also be attained with no subsidy $x = 0$ if $\rho = \frac{\mu}{1+\mu}$.

Letting η denote the share of non-labor income (i.e. fiscal transfer) to labor income $\eta = \frac{\rho - \frac{\theta}{\lambda}x}{\frac{1+x}{1+\mu}}$, the loglinearized budget constraint of H is: $c_t^H = \frac{1}{1+\eta} (w_t + n_t^H)$ which combined with their labor supply delivers $c_t^H = \frac{1+\varphi^{-1}}{1+\eta+\varphi^{-1}\sigma^{-1}}w_t$. It is straightforward to show that the wage-hours locus is still $w_t = (\varphi + \sigma^{-1})y_t$, which implies the consumption function for H agents: $c_t^H = \chi y_t$ where under this redistribution scheme we have

$$\chi = 1 + \frac{\varphi + \sigma^{-1} - \eta}{1 + \eta + \varphi^{-1}\sigma^{-1}}$$

It follows that the elasticity χ is larger than 1 iff $\eta < \varphi + \sigma^{-1}$ or replacing η :

$$\rho - \frac{\theta}{\lambda}x < \frac{1+x}{1+\mu} (\varphi + \sigma^{-1}).$$

C Analysis of the (Simple-HANK) $\lambda\chi\delta$ Model

C.1 Consumption Function and PE Curve

Using the stochastic discount factor notation used above for the Euler equation (11), we have

$$\sigma q_{t,t+1}^S = c_t^S - sE_t c_{t+1}^S - (1-s)E_t c_{t+1}^H$$

Iterating forward (note: we no longer have $q_{t,t+i}^j = -\sum_{k=0}^{i-1} r_{t+k}$)

$$\begin{aligned} c_t^S &= s^i E_t c_{t+i}^S - \sigma \sum_{k=0}^{i-1} s^k (r_{t+k} - (1-s) E_t c_{t+k}^H) \\ \sigma q_{t,t+i}^S &= s^i E_t c_{t+i}^S + \sigma E_t \sum_{k=0}^{i-1} s^k (q_{t,t+k}^S + (1-s) E_t c_{t+k}^H) \end{aligned}$$

Using the definition of stochastic discount factor:

$$\begin{aligned} \sigma q_{t,t+i}^S &= c_t^S - s E_t c_{t+1}^S - (1-s) E_t c_{t+1}^H + c_{t+1}^S - s E_t c_{t+2}^S - (1-s) E_t c_{t+2}^H + \\ &\quad \dots + c_{t+i-1}^S - s E_t c_{t+i}^S - (1-s) E_t c_{t+i}^H \\ \sigma q_{t,t+i}^S + c_{t+i}^S &= c_t^S + (1-s) E_t \sum_{k=1}^i (c_{t+k}^S - c_{t+k}^H) \end{aligned}$$

Now loglinearize intertemporal budget constraint

$$E_t \sum_{i=0}^{\infty} \beta^i (q_{t,t+i}^S + c_{t+i}^S) = E_t \sum_{i=0}^{\infty} \beta^i (q_{t,t+i}^S + y_{t+i}^S)$$

Add to each side $(\sigma - 1) \sum_{i=0}^{\infty} \beta^i E_t q_{t,t+i}^S$

$$E_t \sum_{i=0}^{\infty} \beta^i (\sigma q_{t,t+i}^S + c_{t+i}^S) = E_t \sum_{i=0}^{\infty} \beta^i (\sigma q_{t,t+i}^S + y_{t+i}^S)$$

By virtue of the Euler equation the LHS simplifies

$$\begin{aligned} \frac{1}{1-\beta} c_t^S + (1-s) E_t \sum_{i=0}^{\infty} \beta^i \sum_{k=1}^i (c_{t+k}^S - c_{t+k}^H) &= \sigma \sum_{i=0}^{\infty} \beta^i E_t q_{t,t+i}^S + \sum_{i=0}^{\infty} \beta^i E_t y_{t+i}^S \\ \frac{1}{1-\beta} c_t^S + \frac{1-s}{1-\beta} E_t \sum_{i=1}^{\infty} \beta^i (c_{t+i}^S - c_{t+i}^H) &= \sigma \sum_{i=0}^{\infty} \beta^i E_t q_{t,t+i}^S + \sum_{i=0}^{\infty} \beta^i E_t y_{t+i}^S \end{aligned}$$

Develop RHS $\sum_{i=0}^{\infty} \beta^i E_t q_{t,t+i}^e$ using $q_{t,t} = 0$, this is as above in general case and replace to obtain (multiplying by $1 - \beta$)

$$\begin{aligned} c_t^S &= -(1-s) E_t \sum_{i=1}^{\infty} \beta^i (c_{t+i}^S - c_{t+i}^H) - \sigma \beta \sum_{i=0}^{\infty} \beta^i E_t r_{t+i} + (1-\beta) \sum_{i=0}^{\infty} \beta^i E_t y_{t+i}^S \\ &= -\sigma \beta r_t + (1-\beta) y_t^S - (1-s) E_t \sum_{i=1}^{\infty} \beta^i (c_{t+i}^S - c_{t+i}^H) - \sigma \beta \sum_{i=1}^{\infty} \beta^i E_t r_{t+i} + (1-\beta) \sum_{i=1}^{\infty} \beta^i E_t y_{t+i}^S \end{aligned}$$

Now replace expression for expected consumption tomorrow

$$\beta c_{t+1}^S = -(1-s) E_t \sum_{i=1}^{\infty} \beta^{i+1} (c_{t+i+1}^S - c_{t+i+1}^H) - \sigma \beta \sum_{i=0}^{\infty} \beta^{i+1} E_t r_{t+1+i} + (1-\beta) \sum_{i=0}^{\infty} \beta^{i+1} E_t y_{t+1+i}^S$$

to obtain the consumption function:

$$c_t^S = -\sigma \beta r_t + (1-\beta) y_t^S - (1-s) E_t \sum_{i=1}^{\infty} \beta^i (c_{t+i}^S - c_{t+i}^H) - \sigma \beta \sum_{i=1}^{\infty} \beta^i E_t r_{t+i} + (1-\beta) \sum_{i=1}^{\infty} \beta^i E_t y_{t+i}^S$$

or in recursive form:

$$\begin{aligned} c_t^S &= -\sigma \beta r_t + (1-\beta) y_t^S - (1-s) \beta (E_t c_{t+1}^S - E_t c_{t+1}^H) + \beta E_t c_{t+1}^S \\ &= -\sigma \beta r_t + (1-\beta) y_t^S + \beta s E_t c_{t+1}^S + \beta (1-s) E_t c_{t+1}^H \end{aligned}$$

Aggregate and use $c_t^H = y_t^H = \chi y_t$ to obtain (using the notation for $\delta = \frac{s+(1-\lambda-s)\chi}{1-\lambda\chi}$)

$$c_t = [1 - \beta(1 - \lambda\chi)] y_t - (1 - \lambda) \sigma \beta r_t + \beta \delta (1 - \lambda\chi) E_t c_{t+1}.$$

C.2 Virtues of a Wicksellian rule

The central bank can avoid indeterminacy if it follows the Wicksellian rule proposed by Woodford (2003) for the standard NK model:

$$i_t = \phi_p p_t.$$

To see how, take first differences to obtain $i_t - i_{t-1} = \phi_p \pi_t$; replacing the contemporaneous Phillips Curve both in this and in the Euler-IS curve the system to be solved is

$$\begin{aligned} i_t &= i_{t-1} + \phi_p \kappa c_t \\ c_t &= \left(\delta + \kappa \sigma \frac{1-\lambda}{1-\lambda\chi} \right) E_t c_{t+1} - \sigma \frac{1-\lambda}{1-\lambda\chi} i_t \end{aligned}$$

The system is determinate if one eigenvalue is inside and one outside the unit circle. Replace the second equation in the first using the notation used in text $\nu_0 = \delta + \kappa \sigma \frac{1-\lambda}{1-\lambda\chi}$) to reduce to a second-order difference equation:

$$E_t c_{t+1} - [1 + \nu_0^{-1} + \phi_p (1 - \delta \nu_0^{-1})] c_t + \nu_0^{-1} c_{t-1} = 0$$

The characteristic polynomial is

$$J(x) = x^2 - [1 + \nu_0^{-1} + \phi_p (1 - \delta \nu_0^{-1})] x + \nu_0^{-1}$$

There are two cases. First, if $\nu_0 < 1$ we have that determinacy occurs even under a peg $\phi_p = 0$ (this is consistent with our result in text). What we are most interest in is Case 2: the Wicksellian rule giving determinacy even when a Taylor rule fails to do so: $\nu_0 > 1$. In this case the product of the eigenvalues (ν_0^{-1}) is smaller than 1 so at least one is inside the unit circle. To rule out that both are, all we need is $J(-1)J(1) < 0$. Since $J(-1)$ is always positive under our sign restrictions, the necessary and sufficient condition for determinacy comes from $J(1) < 0$ and is:

$$\phi_p > 0$$

just as in the standard NK model.

C.3 Neo-Fisherian Effects

We want to solve the equation (17) with $\nu \equiv \delta - (\phi - 1)\kappa\sigma\frac{1-\lambda}{1-\lambda\chi} > 1$ (example: peg in the RANK model).

$$c_t = \nu E_t c_{t+1} - \sigma \frac{1-\lambda}{1-\lambda\chi} i_t^* \quad (27)$$

We cannot solve it forward, and to solve it backward we miss an initial condition (c is not a state variable); I follow Lubik and Schorfheide (2004) and define the new expectation variable $\mathbb{F}_t \equiv E_t c_{t+1}$ and the expectation (forecast) error as: $\eta_t \equiv c_t - \mathbb{F}_{t-1}$ indicating how far off the prediction using yesterday's information set is from the actual, realized value. Using these definitions, we can rewrite our equation as:

$$\mathbb{F}_t = \nu^{-1} \mathbb{F}_{t-1} + \nu^{-1} \eta_t + \nu^{-1} \sigma \frac{1-\lambda}{1-\lambda\chi} i_t^* \quad (28)$$

We can try to solve equation (28) backwards (use repeated substitution or lag operators L , or whatever else) to get:

$$\mathbb{F}_t = \frac{\nu^{-1}}{1 - \nu^{-1}L} \left(\eta_t + \sigma \frac{1-\lambda}{1-\lambda\chi} i_t^* \right) = \sum_{j=0}^{\infty} \nu^{-j-1} \left(\eta_{t-j} + \sigma \frac{1-\lambda}{1-\lambda\chi} i_{t-j}^* \right). \quad (29)$$

But, of course, we have not really solved for anything: expectations \mathbb{F}_t are a function of past and present expectation errors η_{t-j} . The problem is that when $\nu > 1$ and c_t is not a predetermined variable, we have no restrictions on either expectations or expectation errors that we can use so solve our equation: the classic problem of **equilibrium indeterminacy** (the 'solution' (29) expresses an endogenous variable, \mathbb{F}_t as a function of another endogenous variable η_t). There is an infinity of equilibria, indexed by the expectation errors. Since expectation errors are not determined, sunspots (shocks that are completely extrinsic to the model) can have real effects.

Since there is nothing to pin down expectation errors η_t , we can assume that it takes the

arbitrary (but linear, since the model is linear) form:

$$\eta_t = mi_t^* + s_t \tag{30}$$

i.e. that expectation errors are an arbitrary combination of fundamental uncertainty (i_t^*) and purely non-fundamental uncertainty: sunspots s_t . Notably, m is an arbitrary constant. Picking one particular equilibrium path among the infinite possibilities boils down to: (i) specifying the stochastic properties of s_t and (ii) picking a value for m . The latter emphasizes that indeterminacy affects the propagation of fundamental shocks in an arbitrary way dictated by the value of m even when sunspot shocks are absent, $s_t = 0$.

One equilibrium advocated by McCallum is obtained by the minimum-state variable MSV criterion; in this simple example, this amounts to setting $s_t = 0$ and ruling out endogenous persistence (this is what Lubik and Schorfheide call the "continuity" solution: impulse response functions to fundamental shocks are continuous when crossing between the determinacy and indeterminacy regions). Under this restriction we have that if the fundamental shock persistence is p^* , so is the endogenous persistence, $E_t c_{t+1} = p^* c_t$; to see what this requires in our context, rewrite the equation using the definition of η :

$$c_{t+1} = \nu^{-1} c_t + \eta_{t+1} + \nu^{-1} \sigma \frac{1 - \lambda}{1 - \lambda \chi} i_t^* \tag{31}$$

It is immediately apparent that the restriction $m = \sigma \frac{1 - \lambda}{1 - \lambda \chi}$ gives the same impulse response as under determinacy. Under these assumptions, we recover the particular solution given in text for a peg with persistence p .

C.4 The simple HANK 3-equation model with NKPC

A full solution of the NK model can be obtained by standard methods. Consider the model with a forward-looking Phillips curve, i.e. (13), (15) and:

$$\pi_t = \beta E_t \pi_{t+1} + \kappa c_t + \zeta_H \kappa g_t,$$

in other words assume $\beta_f = \beta$ in (16).

The HANK Taylor Principle: Equilibrium Determinacy with Interest Rate Rules

Determinacy can be studied by standard techniques, extending the result in text (there will now be two eigenvalues). Necessary and sufficient conditions are provided i.a. in Woodford, Proposition C.1. With the forward-looking rule (15), the system becomes $\begin{pmatrix} E_t \pi_{t+1} & E_t c_{t+1} \end{pmatrix}' = A_1 \begin{pmatrix} \pi_t & c_t \end{pmatrix}'$

with transition matrix:

$$A = \begin{bmatrix} \beta^{-1} & -\beta^{-1}\kappa \\ \beta^{-1}\sigma(\phi-1)\delta^{-1}\frac{1-\lambda}{1-\lambda\chi} & \delta^{-1}\left(1 - \beta^{-1}\sigma\kappa(\phi-1)\frac{1-\lambda}{1-\lambda\chi}\right) \end{bmatrix}$$

with $\det A = (\beta\delta)^{-1}$ and $\text{tr}A = \beta^{-1} + \delta^{-1}\left(1 - \beta^{-1}\sigma\kappa(\phi-1)\frac{1-\lambda}{1-\lambda\chi}\right)$. Determinacy can obtain in either of two cases. Case 2. ($\det A - \text{tr}A < -1$ and $\det A + \text{tr}A < 1$) can be ruled based on sign restrictions. Case 1. requires three conditions to be satisfied jointly:

$$\det A > 1; \det A - \text{tr}A > -1; \det A + \text{tr}A > -1$$

Replacing the expressions for determinant and trace, we obtain:

$$\begin{aligned} \delta &< \beta^{-1} \text{ and} \\ \phi &\in \left(1 + (\delta - 1)\frac{1 - \beta}{\sigma\kappa}\frac{1 - \lambda\chi}{1 - \lambda}, 1 + (1 + \delta)\frac{1 + \beta}{\sigma\kappa}\frac{1 - \lambda\chi}{1 - \lambda}\right) \end{aligned} \quad (32)$$

The interval for ϕ is non-empty iff $1 + \delta\beta > 0$ which is always satisfied (the same condition that rules out Case 1). Notice that for $\delta > \beta^{-1}$ the determinant is less than one so there is at least one root inside the unit circle, rendering determinacy impossible. But it is still possible with a **contemporaneous Taylor rule**

$$i_t = \phi_\pi \pi_t$$

Under this rule, the transition matrix becomes

$$A = \begin{bmatrix} \beta^{-1} & -\beta^{-1}\kappa \\ \delta^{-1}\sigma\frac{1-\lambda}{1-\lambda\chi}(\phi_\pi - \beta^{-1}) & \delta^{-1}\left(1 + \sigma\frac{1-\lambda}{1-\lambda\chi}\beta^{-1}\kappa\right) \end{bmatrix}$$

with $\det A = \beta^{-1}\delta^{-1}\left(1 + \kappa\sigma\frac{1-\lambda}{1-\lambda\chi}\phi_\pi\right)$ and $\text{tr}A = \beta^{-1} + \delta^{-1}\left(1 + \sigma\frac{1-\lambda}{1-\lambda\chi}\beta^{-1}\kappa\right)$. Case 2 can again be ruled out by sign restrictions, leaving us with case 1 (same conditions as above). The third condition is always satisfied under the sign restrictions, so the necessary and sufficient conditions are:

$$\phi_\pi > \max\left(\frac{\beta\delta - 1}{\kappa\sigma}\frac{1 - \lambda\chi}{1 - \lambda}, 1 + \frac{(1 - \beta)(\delta - 1)}{\kappa\sigma}\frac{1 - \lambda\chi}{1 - \lambda}\right) \quad (33)$$

The second term is larger than the first iff $\delta < \frac{1+\beta}{2\beta-1}$. Notice that determinacy now obtains even when $\delta > \beta^{-1}$.

Conditions (15) and (15) thus generalize the HANK Taylor principle to the case of NKPC and forward-looking and contemporaneous rule.

Closed-form Solution

Assuming that determinacy occurs because the modified Taylor principle is satisfied and assuming persistence p for the exogenous processes we can solve the model in closed form: there is no endogenous state variable (so, because of determinacy, no endogenous persistence) and p is also the persistence of endogenous variables. Defining the composite parameter

$$\kappa_p \equiv \frac{\kappa}{1 - \beta p}$$

we can thus rewrite the Phillips curve and real rate under this structure as:

$$\begin{aligned}\pi_t &= \kappa_p c_t + \zeta_H \kappa_p g_t \\ r_t &= (\phi - 1) \kappa_p p c_t + \zeta_H (\phi - 1) \kappa_p p g_t + i_t^*.\end{aligned}$$

Substituting in (13) using the definition

$$\nu_p = \delta - \sigma \frac{1 - \lambda}{1 - \lambda \chi} (\phi - 1) \kappa_p$$

we obtain the solution:

$$\begin{aligned}c_t &= -\frac{1}{1 - \nu_p p} \sigma \frac{1 - \lambda}{1 - \lambda \chi} i_t^* \\ &\quad + \frac{1}{1 - \nu_p p} \frac{\lambda \zeta_H}{1 - \lambda \chi} \left[\left(\chi - \frac{\alpha}{\lambda} \right) \left(1 - p + \frac{1 - s}{\lambda} p \right) - \sigma \frac{1 - \lambda}{\lambda} (\phi - 1) \kappa_p p \right] g_t\end{aligned}$$

Notice that this is exactly of the same form as the solution in text, with κ_p and ν_p replacing κ and ν .

Liquidity trap and FG

Under the Markov chain structure used in text and for the general NKPC case $\beta_f = \beta$ we can use the same solution method to obtain the liquidity-trap solution under forward guidance (abstracting from fiscal shocks). Using the notations

$$\kappa_z \equiv \frac{\kappa}{1 - \beta z}; \kappa_q \equiv \frac{\kappa}{1 - \beta q}; \kappa_{zq} = \frac{\kappa}{(1 - \beta q)(1 - \beta z)}$$

$$\begin{aligned}\nu_{0z} &= \delta + \sigma \frac{1 - \lambda}{1 - \lambda \chi} \kappa_z; \nu_{0q} = \delta + \sigma \frac{1 - \lambda}{1 - \lambda \chi} \kappa_q \\ \nu_{0zq} &= \delta + \sigma \frac{1 - \lambda}{1 - \lambda \chi} \kappa_{zq}\end{aligned}$$

the equilibrium is:

$$\begin{aligned} c_F &= \frac{1}{1 - q\nu_{0q}} \sigma \frac{1 - \lambda}{1 - \lambda\chi} \rho; \\ c_L &= \frac{(1 - p) q\nu_{0zq}}{(1 - q\nu_{0q})(1 - z\nu_{0z})} \sigma \frac{1 - \lambda}{1 - \lambda\chi} \rho + \frac{1}{1 - z\nu_{0z}} \sigma \frac{1 - \lambda}{1 - \lambda\chi} \rho_L, \end{aligned} \quad (34)$$

and $\pi_F = \kappa_q c_F$, $\pi_L = \beta(1 - z) q\kappa_{zq} c_F + \kappa_z c_L$.

The fiscal multiplier during the trap obeys the same formula as in normal times but with $\phi = 0$ and replacing $p = z$, namely

$$\frac{1}{1 - \nu_z z} \frac{\lambda\zeta_H}{1 - \lambda\chi} \left[\left(\chi - \frac{\alpha}{\lambda} \right) \left(1 - z + \frac{1 - s}{\lambda} z \right) + \sigma \frac{1 - \lambda}{\lambda} \kappa_z z \right].$$

D Forward Guidance with Fixed Prices

I briefly characterize the implications for the effects of *forward guidance with time* taking the iid case for simplicity: at $t + T$ there is a shock that lasts for one period. To find the effect of FG, we iterate the PE curve or consumption function of this model (12) to obtain:

$$c_t = - (1 - \lambda) \sigma \beta \sum_{i=0}^{\infty} [\beta \delta (1 - \lambda\chi)]^i E_t r_{t+i} + [1 - \beta (1 - \lambda\chi)] \sum_{i=0}^{\infty} [\beta \delta (1 - \lambda\chi)]^i E_t y_{t+i}. \quad (35)$$

Direct differentiation with respect to a one-time interest rate cut at $t + T$ delivers that the response to FG (an interest rate cut in T periods) the total effect and indirect effect share are:

$$\begin{aligned} \Omega^F &= \sigma \frac{1 - \lambda}{1 - \lambda\chi} \delta^T, \\ \omega^F &= 1 - [\beta (1 - \lambda\chi)]^{1+T}. \end{aligned} \quad (36)$$

Specifically, for any k from 0 to T the total effect is (by direct differentiation of the forward-iterated Euler equation (11)) $\Omega^{F(k)} \equiv \frac{dc_{t+k}}{d(-r_{t+T})} = \frac{1-\lambda}{1-\lambda\chi} \sigma \delta^{T-k}$, for any k from 0 to T . The direct FG effect Ω_D^F corresponds to the derivative of the first sum in (35): $\Omega_D^F \equiv \frac{dc_{t+k}}{d(-r_{t+T})} \Big|_{y_{t+k}=y} = \beta \sigma (1 - \lambda) [\delta \beta (1 - \lambda\chi)]^T$. The indirect FG effect corresponds to the second term in (35): $\Omega_I^F \equiv \frac{dc_{t+k}}{d(-r_{t+T})} \Big|_{r_{t+k}=r} = \frac{1-\lambda}{1-\lambda\chi} \sigma \delta^T \left\{ 1 - [\beta (1 - \lambda\chi)]^{1+T} \right\}$, which delivers the indirect share.⁵⁵

In the RANK limit ($s = 1$ and $\lambda = 0$) the total effect of one-time FG is invariant to time (one instance of the FG puzzle: the interest rate cut has the same effect regardless of whether it

⁵⁵Garcia-Schidt and Woodford (2014) also use a version of the forward-iterated consumption function to compute the effects of FG under finite planning horizon using a notion of "reflective equilibrium". See also Farhi and Werning (2017) for combining incomplete markets with a version of that information imperfection, i.e. "k-level thinking", that delivers a complementarity. The last paper also derived independently the analytical expressions found here for the simple RANK case.

takes place next period, in one year, or in one century). Furthermore, the indirect effect's share *increases*, the further FG is pushed into the future (ω^F is increasing with T).

Take now the TANK special case ($s = h = 1$, λ arbitrary). As for within-period policy changes, the total effect Ω^F is larger—but it is still time-invariant, i.e. it is the same for any k from 0 to T . The same insights as for iid monetary policy shocks apply: higher λ results in higher total effect, higher indirect effect and lower direct effect, and higher indirect effect share. In addition, the indirect effect share is increasing with time, just as—but at a faster rate than—in the RANK model. The key point is that in the TANK model forward guidance is *more* powerful than in the RANK model, but this has no impact on the way in which the total effect depends (*not*) on the *horizon* of FG.

The main novel insight from the $\lambda\chi\delta$ model, as found by MNS is to break this invariance: the effect of *forward guidance* is no longer time-invariant, because of discounting. However, as holds true in Werning's (2015) more general setting, this insight is overturned if, as in the TANK model, the income of hand-to-mouth covaries with aggregate income more than one-to-one. Direct inspection of the expressions in (36) unveils that when $\chi < 1$ and $\delta < 1$ (and decreasing with λ), the total, direct, and indirect effect are *all decreasing* with λ ; furthermore, the total effect is lower when pushed far into the future, thus resolving the FG puzzle as discussed above. The indirect share increases when the horizon T increases, but at the same rate as in the TANK model. Matters are different when $\chi > 1$ because of the two mechanisms: the contemporaneous (TANK) amplification, and the compounding discussed above. As a consequence, the total effect of FG now **increases** with T , which this delivers a novel side of (and thus aggravates) the "FG puzzle".

Table A1: Summary of MP and FG effects

		TANK	$\lambda\chi\delta$
MP	Total effect Ω	$\frac{\sigma}{1-p} \frac{1-\lambda}{1-\lambda\chi}$	$\frac{\sigma}{1-\delta p} \frac{1-\lambda}{1-\lambda\chi}$
	Indirect share ω	$\frac{1-\beta(1-\lambda\chi)}{1-\beta p(1-\lambda\chi)}$	$\frac{1-\beta(1-\lambda\chi)}{1-\beta\delta p(1-\lambda\chi)}$
FG iid	Total effect Ω^F	$\sigma \frac{1-\lambda}{1-\lambda\chi}$	$\sigma \frac{1-\lambda}{1-\lambda\chi} \delta^T$
	Indirect share ω^F	$1 - [\beta(1-\lambda\chi)]^{1+T}$	$1 - [\beta(1-\lambda\chi)]^{1+T}$

Figure A1 illustrates the findings. The left column plots the total effect and the right column the indirect share, as a function of λ (top, for $T = 10$) and T (bottom, for $\lambda = 0.2$). I distinguish the two cases according to whether χ is larger (thick) or lower (thin) than unity, and plot for each case the TANK model with dash and the iid $\lambda\chi\delta$ model with dots (recall that the former is the limit as $s = h = 1$ and the latter the limit as $s = 1 - h = 1 - \lambda$; the analysis and discussion in the previous section still apply). The total effect of FG increases steeply with λ , relatively more when there is more idiosyncratic risk ($1 - s$ higher), when $\chi > 1$; and it decreases with λ —relatively more when there is more idiosyncratic risk—when $\chi < 1$.

The same is true with respect to the horizon of FG: the further FG is pushed into the future, the more powerful it is. The more risk, the larger is this amplification (it disappears with no risk,

i.e. in the TANK model). Conversely, when $\chi < 1$, there is dampening: the total effect decreases with the horizon, and the more so the higher the risk (it is again invariant in the TANK limit, even though $\chi < 1$ makes the effect lower in levels). The share of the indirect effect, on the other hand, is invariant to the level of idiosyncratic risk: it is increasing with both λ and T ; the speed with which it does so depends on χ , as noted before.

Fig. A1: Ω^F and ω^F : $\chi = 2$ (thick), 0.5 (thin), TANK (dash) and iid $\lambda\chi\delta$ (dots)

E HANK meets DSGE and ZLB: Model and Calibration

To be completed