Monetary Policy and Heterogeneity: An Analytical Framework[I]

Florin O. Bilbiie[II]

February 2019[III]

Abstract

An analytical (heterogeneous-agent New-Keynesian) HANK model allows a closed-form treatment of a wide range of NK topics: determinacy properties of interest-rate rules, resolving the forward guidance FG puzzle, amplification and fiscal multipliers, liquidity traps, and optimal monetary policy. The key channel shaping all the model's properties is that of *cyclical inequality*: whether the income of constrained agents moves less or more than proportionally with aggregate income. With countercyclical inequality, good news on aggregate demand gets compounded, making determinacy less likely and aggravating the FG puzzle (the resolution of which requires procyclical inequality)—a Catch-22, because countercyclical inequality is what HANK (and TANK) models need to deliver desirable amplification. The dilemma can be resolved if a distinct, "cyclical-risk" channel is procyclical enough. Even when both channels are countercyclical a Wicksellian rule of price-level targeting ensures determinacy and cures the puzzle. Optimal monetary policy is isomorphic to RANK and TANK but calls for less inflation stabilization. In a liquidity trap, even with countercyclical inequality and FG amplification, optimal policy does not imply larger FG duration because as FG power increases, so does its welfare cost.

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

Keywords: heterogenous agents; HANK; monetary policy; forward guidance puzzle; Taylor and Wicksellian rules; determinacy; liquidity traps; multipliers; optimal monetary policy.

[I] I am grateful to Sushant Acharya, Adrien Auclert, Jess Benhabib, Edouard Challe, John Cochrane, Davide Debortoli, Keshav Dogra, Axelle Ferrière, Xavier Gabaix, Gaetano Gaballo, Jordi Galí, Roger Guesnerie, Marcus Hagedorn, Greg Kaplan, Keith Kuester, Eric Leeper, Olivier Loisel, Alistair Macaulay, Virgiliu Midrigan, Kurt Mitman, Benjamin Moll, Emi Nakamura, Salvatore Nistico, Ricardo Reis, Xavier Ragot, Kenneth Rogoff, Gilles Saint-Paul, Emiliano Santoro, Christopher Sims, Jon Steinsson, Paolo Surico, Roman Sustek, Andrea Tambalotti, Gianluca Violante, Mirko Wiederholt, Michael Woodford, and participants in many seminars and conferences for useful comments. I gratefully acknowledge without implicating the support of Banque de France Chair at PSE, and of Institut Universitaire de France, as well as the hospitality of New York University and CREI.

[II] University of Lausanne and CEPR; florin.bilbiie@gmail.com. http://florin.bilbiie.googlepages.com.

[III] This paper supersedes the Dec. 2017 "A Catch-22 for HANK Models: No Puzzles, No Amplification" (then "Heterogeneity and Determinacy: Amplification without Puzzles"), and the previous Aug 2017 "The Puzzle, the Power, and the Dark Side: Forward Guidance Redux" that contained a subset of the material.

1 Introduction

The New Keynesian (NK) framework is the core of most models used for policy analysis since now decades; yet the post-2008 crisis and recession coupled with a liquidity trap (LT) raised the need for unconventional policy tools and challenged some of the model's predictions in those circumstances. One particular dimension for extending the model has been a focus on inequality and redistribution, with entire speeches by the world's most prominent policymakers (such as Bernanke, Yellen, or Draghi) devoted to the topic and asking for more research on its monetary policy implications.

Overwhelming evidence was since long available (and motivated an earlier literature reviewed below) that an aggregate Euler equation—one of the pillars of the benchmark RANK model—is not a good description of most households' behavior (starting with Hall 1988 and Campbell and Mankiw 1989 to more recent evidence such as Yogo 2004, Vissing-Jorgensen 2002, Canzoneri, Cumby, and Diba, 2007; Bilbiie and Straub, 2012); furthermore, micro evidence is plentiful indicating that a high fraction of households even in one of the world's most developed economies (the U.S.) has essentially zero net worth: Wolff (2000), Bricker et al (2014). Likewise, a plethora of studies have by now documented that, unlike what predicted by the permanent-income hypothesis, consumption responds to income for a large fraction of the population: see e.g. Johnson, Parker, Souleles 2006; Misra Surico 2014; Surico Trezzi 2016. These households are thus "hand-to-mouth", in that they have a high marginal propensity to consume MPC out of their current income.

Important recent empirical work shed light on the link between liquidity constraints and MPCs: according to this evidence, a large fraction of otherwise "wealthy" households behaves as hand-to-mouth because this wealth is illiquid (Kaplan and Violante (2014)), perhaps because it consists of a house subject to mortgage (Cloyne Ferreira and Surico (2015)), and even if housing is partially liquid (Gorea and Midrigan (2017)).

These elements triggered a significant change to the NK paradigm: the increasing use of heterogeneous-agent models for policy analysis. A burgeoning literature (that I review to some extent below) uses heterogeneous-agent New Keynesian models (labelled *HANK* by one of the main references, Kaplan, Moll and Violante 2018) for a multitude of topics.

In this paper, I build an analytical HANK model that, first, is a reasonable reduced-form description of some key channels at work in quantitative HANK models; and second, it allows a full-fledged NK analysis with pencil and paper of that framework's main topics of focus over the last decades: determinacy with interest rate rules; curing the FG puzzle (and other puzzling prediction of the NK model); fiscal multipliers and amplification; optimal monetary policy; and the analysis of all of the above in a liquidity trap.

My analytical HANK is a three-equation NK model isomorphic to RANK (which it nests). The difference is captured by its AD side, further explored in the companion paper Bilbiie (2017), where I underline the "New Keynesian Cross" that is at work in any HANK model, insofar as some households are constrained hand-to-mouth in equilibrium (while those who are not self-insure against the *risk* of becoming so using some liquid asset, whose return is controlled by the central bank). The key channel through which heterogeneity shapes equilibrium outcomes in my framework is that of *cyclical inequality*: how the distribution of income between constrained and unconstrained households changes over the cycle, e.g. who suffers more in recessions. This is also an admittedly very reduced and simplified version of one of the key channels emphasized by the important subsequent richer-heterogeneity, quantitative literature: the "earnings heterogeneity channel" that Auclert (2018) studies in conjunction with other channels, providing amplification when the covariance of MPCs and individual income elasticities is positive.

Furthermore, the model incorporates analytically a notion of uninsurable idiosyncratic uncertainty and the distinction between liquid and illiquid assets—both of which are staples of the main quantitative HANK papers. Other than this AD side, my model includes a standard AS side (a Phillips curve) and looks at monetary policy design, either as rules consisting of setting the nominal interest rate (the return on liquid assets), or as optimal policy found by maximizing aggregate welfare.

Under a further inconsequential simplification, the whole model can be boiled down to only *one* first-order difference equation whose root/eigenvalue governs the model's dynamics, determinacy properties, and dictates whether the model cures or not the *forward guidance (FG) puzzle* (Del Negro, Giannoni, and Patterson, 2012): that the later in the future an interest rate cut takes place, the larger an effect it will have today.

This root, which captures the equilibrium AD effect of future news, convolutes three channels: i. cyclical inequality, *the pivotal* channel of the TANK (two-agent NK) model in Bilbiie (2008) with constrained hand-to-mouth households—whose income elasticity to aggregate income is the key parameter χ; ii. self-insurance in face of idiosyncratic risk of becoming constrained, a HANK channel; and iii. a standard RANK supply channel.

As shown formally and discussed at length in text, AD-amplification occurs in this model when inequality is countercyclical ($\chi > 1$). An increase in demand leads to a more-than-proportional increase in constrained agents' income and a further demand expansion—the intertemporal version of which delivers *compounding* in the aggregate Euler equation. Conversely, when inequality is procyclical ($\chi < 1$), there is AD-dampening and *discounting* in the aggregate Euler equation.

The determinacy properties of Taylor rules reflect this dampening/amplification intuition. When inequality is countercyclical, the central bank needs to be (possibly much) more aggressive than the "Taylor principle" (increasing nominal interest rates more than one-to-one with inflation) to rule out indeterminacy and potential sunspot fluctuations. Whereas in the "discounting" case, with procyclical inequality, the Taylor principle is sufficient—but not necessary; indeed, for a large region of the "discounting" parameter subspace, determinacy occurs even under an interest rate peg, thus undoing the Sargent-Wallace result.

The paper shows in this context that determinacy under a peg is what is required to the solve the FG puzzle: therefore, heterogeneity *can* solve the puzzle if, in a nutshell, it generates enough discounting on the AD side to compensate for the compounding through the AS side that causes the puzzle in RANK.

An apparent Catch-22 occurs once we make the following uncomfortable observation: the conditions to rule out puzzles are the *opposite* of the condition needed for HANK models to generate *amplification* (relative to RANK) of shocks and policies—which is what much of the current literature uses this class of models for. I illustrate this by deriving analytically the conditions for fiscal multipliers: the model needs *countercyclical* inequality (*CI* for short $\chi > 1$); but by the same logic by which *procyclical* inequality (*PI*) leads to resolving the puzzles, CI implies their aggravation.[1]

A possible way out consists of considering the *distinct* HANK channel of cyclical risk. I extend my model to include a novel formalization of this channel emphasized previously by others (as reviewed in detail below). Procyclical risk can also give rise to discounting in the aggregate Euler equation and solve the puzzle, through a different mechanism: if an AD expansion leads to an increase in uninsurable risk, precautionary saving increases which leads agents to cut back demand. This channel is orthogonal to the cyclical-inequality channel that my work emphasizes: it operates even when inequality is acyclical. By the reverse logic though, *countercyclical* risk aggravates the puzzle—albeit through a different mechanism. A nagging policy implication is thus that when both inequality and risk are countercyclical (a far from empirically implausible property as discussed in the concluding section), both channels give amplification: all the puzzles are much aggravated and determinacy requirements with a Taylor rule become very stringent. I show that a simple policy remedy exists even under such extreme conditions: the *Wicksellian* interest rate rule proposed by Woodford (2003) and Giannoni (2014), which targets the price level rather than inflation. In HANK, this rule with *some*, no matter how small response of nominal interest rates to the price level ensures equilibrium determinacy and rule out the puzzles, while preserving amplification.

I then calculate analytically optimal monetary policy in a-HANK by approximating aggregate welfare to derive a second-order objective function for the central bank. In my benchmark model, the objective function is isomorphic to RANK and TANK (it requires stabilizing inflation and aggregate activity), but requires optimally tolerating more inflation volatility when more households are constrained (because inflation volatility is costly like a tax on financial assets). Optimal policy under commitment also ensures determinacy regardless of the degree of heterogeneity.

Finally, I analyze liquidity-trap equilibria and calculate in closed form the effects of forward guidance and its optimal duration. The general findings emphasized above apply: the severity of LT recessions and the power of FG are mitigated with PI and magnified with CI, with the FG puzzle resolved in the former case and aggravated in the latter. But the optimal-policy analysis shows that even in the latter case, the optimal FG duration is contained because amplification also applies to the welfare cost of FG and not only to its benefit.

[1] Recent empirical evidence provided by Patterson (2019) suggests that income inequality is indeed countercyclical; the paper also proposes a novel mechanism based on different cyclicalities of jobs in a matching framework.

1.1 Literature

Quantitative HANK models that model explicitly rich income risk heterogeneity and the feedback effects from equilibrium distributions to aggregates are being increasingly used to address a wide spectrum of issues in macroeconomic policy.[2]

The analytical HANK model proposed here can be viewed as an extension of the TANK model in Bilbiie (2008), which analyzed *monetary policy*, introducing the distinction between the two types based on asset markets participation (abstracting from physical investment, as done in previous two-agent studies):[3] H have no assets, while S own all the assets (price bonds and shares in firms through their Euler equation). That paper analyzed AD amplification of monetary policy and emphasized the key role of *profits* and their distribution, as well as of fiscal redistribution, for this—in an analytical 3-equation TANK model isomorphic to RANK. In recent work, Debortoli and Galí (2017) and Bilbiie (2017) both used this TANK model to argue that it can approximate reasonably well the aggregate implications of *some* HANK models: the authors' own, for the former paper, and several models from the literature, for the latter (Kaplan et al, McKay et al, Gornemann et al, Hagedorn et al, and Auclert and Rognlie). This suggests that the "cyclical inequality" channel plays an important role in HANK transmission in and of itself.

The first extension here pertains to introducing self-insurance to idiosyncratic uncertainty (the risk of becoming constrained in the future despite not being constrained today), a key mechanism in HANK models that is absent in TANK; doing so gives the model another margin to replicate the aggregate findings of quantitative HANK models, as shown in the companion paper Bilbiie (2017).[4]

Others studies also provide analytical frameworks different from this: both because they isolate *different* HANK mechanisms and focus on different *questions*. Werning (2015) studies monetary policy transmission, similarly emphasizing the possibility of AD amplification or dampening relative to RANK. My paper's subject is very different: a full analysis of NK topics. So is the mechanism, although some of its equilibrium implications pertaining to intertemporal amplification or dampening have a similar flavor. But the key here is *cyclical inequality:* the *distribution*

[2] Topics include the effects of transfer payments (Oh and Reis, 2012); deleveraging and liquidity traps (Guerrieri and Lorenzoni, 2017); job-uncertainty-driven recessions (Ravn and Sterk, 2017; den Haan, Rendahl, and Riegler, 2018); monetary policy transmission (Gornemann, Kuester, and Nakajima, 2016; Auclert, 2016; Debortoli and Gali, 2017); inequality and aggregate demand (Auclert and Ronglie, 2017); precautionary liquidity and portfolio composition (Bayer et al, 2016 and Luetticke, 2018); fiscal multipliers (Ferrière and Navarro, 2018, Hagedorn, Manovskii, and Mitman, 2018, and Auclert, Ronglie, and Straub, 2018); automatic stabilizers (McKay and Reis, 2016); and the FG puzzle (the focus of the McKay et al, 2016; Kaplan et al, 2017).

[3] Mankiw (2000) had used a growth model with this distinction, due to pioneerig work by Campbell and Mankiw (1989), to analyze long-run fiscal policy issues. Galí, Lopez-Salido and Valles (2007) embedded this same distinction in a NK model and studied numerically the business-cycle effects of government spending, with a focus on obtaining a positive multiplier on private consumption. They also analyzed numerically determinacy properties of interest rate rules, that Bilbiie (2008) then derived analytically.

[4] That paper also explains in detail the differences with earlier work using the switching between types to analyze monetary policy issues, such as Nistico (2016) and Curdia and Woodford (2016) in a related context. I spell out the differentiating assumptions below when presenting the model.

of income (between labor and "capital" understood as monopoly profits) and how it depends on aggregate income, as summarized through χ, the chief feature of my earlier TANK model in Bilbiie (2008). Whereas the intuition in Werning emphasizes the *cyclicality of income risk* (and/or of liquidity, which my model abstracts from): as *uninsurable* idiosyncratic income risk goes up in a recession, agents increase their precautionary savings and decrease their consumption, amplifying the initial recession which further increases idiosyncratic risk, and so on (a mechanism previously emphasized in the form of endogenous unemployment risk by Ravn and Sterk (2017) and Challe et al (2017)).

Therefore, my model's mechanism is instead an *intertemporal* extension of the cornerstone amplification (dampening) mechanism in TANK; in this extension, any agent can become constrained in any future period and self-insures (imperfectly) using liquid assets against the (acyclical) risk of doing so. This puts the *cyclicality of income of constrained* (and thus of inequality) at the core of transmission—whereas Werning emphasizes the cyclicality of income *risk of the unconstrained* (although the two *are* convoluted in the different, more general framework therein).

This separation is also clearly illustrated by a subsequent paper by Acharya and Dogra (2018), explicitly set to isolate the role of the risk channel: using CARA preferences to simplify heterogeneity, it shows that such an intertemporal amplification mechanism *may* occur *purely* as a result of uninsurable idiosyncratic income volatility going up in recessions. With this different mechanism, Acharya and Dogra also study determinacy and puzzles, making specific reference to the analysis in the previous version of this paper.

To incorporate this distinction, I extend the model with a (different) formalization of this separate cyclical-risk channel, assuming that the probability of becoming constrained is a function of aggregate demand-output. With this formalization, the two channels of cyclical inequality and risk are separate but related and coexist in shaping AD amplification. Not only *are* the two channels naturally separate: my analysis implies that the two channels *better be distinct*; for in order to eliminate the Catch-22 they *need* to go in opposite directions. Which channel prevails empirically is a very interesting and hitherto unexplored topic that is worth pursuing.

Additionally, my analysis is conducted within the context of a loglinearized NK model that nests as special cases not only the three-equation textbook RANK but also: TANK, a HANK model with cyclical inequality and acyclical risk, and a HANK version with cyclical risk and acyclical inequality. Since it is so simple and transparent and close to standard NK craft, it may be of independent interest to some researchers.

One implication of the analysis here consists of an analytical reinterpretation of the underpinnings of McKay et al's (2016) incomplete-markets based resolution of the FG puzzle, in particular in relation with the same authors' analytical "discounted Euler equation" in McKay et al (2017). My framework nests the latter as a special case and underscores the *procyclicality of inequality* as the keystone, necessary condition for delivering Euler-equation discounting in the presence of (albeit *acyclical*) idiosyncratic risk—a different, complementary interpretation to the framework emphasizing the procyclicality of risk such as Werning and Acharya and Dogra. Procyclicality of inequality can occur in my model through features such as labor market and fiscal redistribution making the income of constrained agents vary less than one-to-one with the cycle $\chi < 1$, whereas MNS (2017) consider the limit case with exogenous income of the constrained ($\chi = 0$). Solving the FG puzzle requires enough discounting to overturn the compounding of news inherent in RANK under a peg. If inequality is instead *countercyclical*, the prediction is overturned: the *compounded* Euler equation in my model implies an aggravation (rather than a resolution) of the FG puzzle. Furthermore, my paper addresses a wide range of NK topics as mentioned above.

Broer, Hansen, Krusell, and Oberg (2018) study a simplified HANK whose equilibrium has a two-agent representation, underscoring the implausibility of some of the model's implications for monetary transmission through income effects of profit variations on labor supply—and showing that a sticky-wage version features a more realistic transmission mechanism; Walsh (2017) provides another analytical model with heterogeneity emphasizing the role of sticky wages (see Colciago (2011), Ascari, Colciago, and Rossi (2017), and Furlanetto (2011) for earlier sticky-wage TANK).

Auclert, Rognlie, and Straub (2018) also use a version of a "Keynesian cross" to capture distinct and complementary amplification channel of HANK models. In particular, they abstract from the cyclical-inequality channel emphasized here and focus instead on the role of liquidity in the form of public debt; they unveil key summary statistics pertaining to the marginal propensities to consume out of past and future income (labelled iMPCs) and how they shape the responses of the economy to past and future income shocks. The quantitative HANK model with liquid and illiquid assets that they use can in fact be viewed as the closest generalization of my a-HANK model; or alternatively, it is among the wide spectrum of quantitative HANK models the one to which my a-HANK model is the closest reduced representation.

I show that my a-HANK model can match the iMPCs unveiled by Auclert et al and provide the *analytical* expressions for these; this affords insights into the important propagation mechanism emphasized by Auclert et al and, in the context of fiscal multipliers, also by Hagedorn et al (2018); as illustrated in both of those quantitative contributions, this propagation mechanism is absent in TANK models without idiosyncratic risk. Indeed, I show that self-insurance to idiosyncratic risk is necessary and sufficient (in the presence of liquidity) to generate the tent-shaped path of iMPCs; whereas cyclical inequality (which, combined with idiosyncratic uncertainty, delivers Euler equation discounting or compounding) is not of the essence to generate persistent iMPCs, but is likely important to deliver the right magnitudes under realistic calibrations.

Ravn and Sterk (2018) also study an analytical HANK but with search and matching (SaM), that is different from and complementary to my model and focus on a different (sub)set of the issues stdied here; while Challe (2018) uses a model very similar to Ravn and Sterk to study optimal monetary policy in normal times. Their models include *endogenous* unemployment risk (a feature of some HANK models) through labor SaM. Workers self-insure against this risk, which depends endogenously on aggregate outcomes. The simplifying assumptions used therein to maintain tractability, in particular pertaining to the asset market, are orthogonal to the ones used here.[5] Their framework delivers an interesting feedback loop from precautionary saving to aggregate demand (see also Challe et al (2017)) that is absent here. My model does much the opposite: it gains tractability from assuming *exogenous* transition probabilities (and a different asset market structure) but emphasizes the NK-cross feedback loop through the *endogenous* income of constrained agents that is absent in Ravn and Sterk and Challe (whereas my extension to cyclical risk can be viewed as an alternative, reduced-form formalization of their channel). Furthermore, the papers not only use complementary models, they also address a different set of NK topics; my paper emphasizes restoring determinacy under a peg and how that rules out the puzzles, points to the uncomfortable implication (Catch-22) that this also rules out amplification more generally, and offers a solution based on adopting a Wicksellian rule of price-level targeting. Furthermore, my paper analyzes optimal monetary policy both in normal times and in a liquidity trap.

Other modifications of the NK model have been proposed recently to solve NK puzzles: changing the information/expectations structure[6], pegging the interest on reserves (Diba and Loisel (2017)), introducing wealth in the utility function (Michaillat and Saez (2017), Hagedorn (2018)), or considering fiscalist equilibria with long-term debt (Cochrane (2017)).[7]

My paper is related to studies of optimal policy: in RANK (Clarida et al (1999), Woodford (2003) and many others), in TANKs (Bilbiie (2008), Ascari, Colciago, and Rossi (2017); Nistico (2016); Curdia and Woodford (2016)), and several varieties of HANKs: Bhandari Evans Golosov Sargent (2018); Challe (2018); McKay Reis (2017); Nuno Thomas (2017); Bilbiie Ragot (2017).

Finally, this paper is related to some of my own current work. The companion paper referred to above Bilbiie (2017) introduces the New Keynesian (NK) Cross as a graphical and analytical apparatus for the AD side of HANK models, expressing its key objects—MPC and multipliers—as functions of heterogeneity parameters. It studies the implications for monetary and fiscal multipliers, the link between MPC and multipliers with the "direct-indirect" decomposition of

[5] In my model savers hold and price the shares whose payoff (profits) they get. In Ravn and Sterk, hand-to-mouth workers get all the return on shares but do not price them (see also Broer et al (2018)). Ravn and Sterk's mechanism can create a third, "unemployment-trap" steady-state equilibrium, a breakup of the Taylor principle that is complementary to the one occurring here, and fix the puzzling NK effects of supply shocks in a LT, which I abstract from here.

[6] Kiley (2016) is an early example addressing the FG puzzle with sticky information à la Mankiw and Reis (2002). Other information imperfections can fix some puzzles, but do not generate the discounting necessary to solve the puzzles studied here (see Wiederholt (2016) and Andrade et al (2016) for models with dispersed information and heterogeneous beliefs). Euler-discounting occurs with deviations from rational expectations such as the reflective equilibrium considered by Garcia-Schmidt and Woodford (2015), the behavioral model with sparsity of Gabaix (2016), imperfect common knowledge as in Angeletos and Lian (2016), the combination of reflective equilibrium with incomplete markets in Farhi and Werning (2017), or the model with finite planning horizons in Woodford (2018).

[7] The price level can also be determined by the demand for nominal bonds by agents coupled with a supply rule for nominal bonds by the government responding to the price level, as discussed by Hagedorn (2017) in a different HANK model. This is related to (but different from, insofar as it requires passive fiscal policcy) the FTPL outlined e.g. in Leeper (1991), Sims (1994), Woodford (1996), and Cochrane (2005); it is also related to the Wicksellian rule proposed here as discussed in text.

KMV, and the ability of this simple model to replicate the aggregate equilibrium implications for quantitative, micro-calibrated HANK models.[8]

2 An Analytical HANK (a-HANK) Model

To study analytically whether and when heterogeneity cures the NK puzzles just described, I use a framework that fits the purpose: an analytical HANK model that captures several key channels of complicated HANK models: cyclical inequality, self-insurance in face of idiosyncratic uncertainty, and a distinction between liquid and illiquid assets. While related to several studies reviewed in the Introduction, the exact model is to the best of my knowledge novel to this and the companion paper Bilbiie (2017)—which focuses on the model's AD amplification of monetary and fiscal policies through a "New Keynesian Cross" and on using it as an approximation to richer HANK models.

Four key assumptions pertaining to the asset market structure render the equilibrium particularly simple and afford an analytical solution; I spell out the formal analysis in Appendix A.1. First, there are two states of the world—constrained hand-to-mouth H and unconstrained "savers" S—between which agents switch *exogenously* (idiosyncratic uncertainty). Second, there is *full insurance within* type (after idiosyncratic uncertainty is revealed), but *limited insurance across* types. Third, different assets have different *liquidity*: bonds are liquid (*can* be used to self-insure, before idiosyncratic uncertainty is revealed), while stocks are illiquid (cannot be used to self-insure). Fourth, I consider two cases: either zero-liquidity (assuming that in equilibrium there is no bond trading—see i.a. Krusell, Mukoyama and Smith (2011), Ravn and Sterk (2017), Werning (2015), McKay and Reis (2017), and Broer et al (2018)), or an equilibrium *with* (government-provided) liquidity.

That the unconstrained S may become constrained H can be interpreted as "risk", against which only one of the two assets—bonds—can be used to insure (is *liquid*). The exogenous change of state follows a Markov chain: the probability to *stay* type S is s, and to stay type H is h (with transition probabilities $1-s$ and $1-h$ respectively).

I focus on stationary equilibria whereby the mass of H is:

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

by standard results (as the steady state of $\lambda_{t+1} = h\lambda_t + (1-s)(1-\lambda_t)$). The requirement $s \geq 1-h$ insures stationarity and has a straightforward interpretation: the probability to stay S is larger than the probability to become S (the conditional probability is larger than the unconditional).[9]

[8] A separate paper Bilbiie and Ragot (2016) builds a different analytical HANK model with three assets, of which one ("money") is liquid and traded in equilibrium while the others (bonds and stock) are illiquid, and studies Ramsey-optimal monetary policy as liquidity provision.

[9] A general version of this condition appears e.g. in Huggett (1993); see also Challe et al (2016) for an interpre-

In the limit $s = 1 - h = 1 - \lambda$, idiosyncratic shocks are iid: being S or H tomorrow is independent on whether one is S or H today. At the other extreme stands TANK: idiosyncratic shocks are permanent ($s = h = 1$) and λ stays at its initial value (a free parameter).

To characterize the equilibrium in asset markets (detailed in Appendix A.1), start from H: in every period, those who happen to be H would like to borrow, but we assume that they cannot (for instance they face a borrowing limit of 0). Since the stock is illiquid, they cannot access that portfolio (owned entirely by S, whoever they happen to be in that period). We thus focus on an equilibrium where they are constrained hand-to-mouth, consuming all their (*endogenous*) income: like in TANK, $C_t^H = Y_t^H$; because transition probabilities are independent of history and with perfect insurance within type, all agents who are H in a given period have the same income and consumption.

S are also perfectly insured among themselves in every period by assumption, and would like to save in order to self-insure against the risk of becoming H. Because shares are illiquid, they can only use (liquid) bonds to do that. But since H cannot borrow, if there is no government-provided liquidity bonds are in zero supply (the no-trade equilibrium of Krusell, Mukoyama, and Smith). We will consider both such equilibria without liquidity, and equilibria with government-provided liquidity. An Euler equation prices these (possibly non-traded) bonds, just like in RANK and TANK, the aggregate Euler equation prices the possibly non-traded bond. But unlike in RANK and TANK (where there is no transition and no self-insurance), now the bond-pricing Euler equation takes into account the possible transition to the constrained H state.

Notice that in line with some key HANK contributions that emphasized the role of asset liquidity, such as Kaplan et al or Luetticke, my model distinguishes, albeit in an extreme way, between liquid (bonds) and illiquid (stock) assets: in equilibrium, there is infrequent (limited) participation in the stock market.

Given our four assumptions, the Euler equation governing the bond-holding decision of S self-insuring against the risk of becoming H is:

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

recalling that we focus on equilibria where the corresponding Euler condition for H holds with strict inequality (the constraint binds), while the Euler condition for stock holdings by S is standard: $\left(C_t^S\right)^{-\frac{1}{\sigma}} = \beta E_t \left[\left(1+r_{t+1}^S\right) \left(C_{t+1}^S\right)^{-\frac{1}{\sigma}} \right]$, merely defining the return on shares r_t^S.

The rest of the model is exactly like the TANK version in Bilbiie (2008, 2017), nested here when there is no idiosyncratic uncertainty. In every period λ households are "hand-to-mouth" H and *excluded* from asset markets (have no Euler equation)—but *do* participate in labor markets and make an optimal labor supply decision (their income is therefore *endogenous).* The rest of

tation in terms of job finding and separation rates, and Bilbiie and Ragot (2016).

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). 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.

All agents maximize present discounted utility, defined as previously, subject to the budget constraints. Utility maximization over hours worked delivers the standard intratemporal optimality condition for each j: $U_C^j\left(C_t^j\right) = W_t U_N^j\left(N_t^j\right)$. With σ^{-1} defined as before, $\varphi \equiv U_{NN}^j N^j / U_N^j$ denoting the inverse labor supply elasticity, and small letters log-deviations from steady-state (to be discussed below), we have the labor supply for each j: $\varphi n_t^j = w_t - \sigma^{-1} c_t^j$. Assuming for tractability that elasticities are identical across agents, the same holds on aggregate $\varphi n_t = w_t - \sigma^{-1} c_t$.

Firms The supply side is standard. All households consume an aggregate basket of individual goods $k \in [0,1]$, with constant elasticity of substitution $\varepsilon > 1$: $C_t = \left(\int_0^1 C_t(k)^{(\varepsilon-1)/\varepsilon} dk\right)^{\varepsilon/(\varepsilon-1)}$. Demand for each good is $C_t(k) = \left(P_t(k)/P_t\right)^{-\varepsilon} C_t$, where $P_t(k)/P_t$ is good k's price relative to the aggregate price index $P_t^{1-\varepsilon} = \int_0^1 P_t(k)^{1-\varepsilon} dk$. Each good is produced by a monopolistic firm with linear technology: $Y_t(k) = N_t(k)$, with real marginal cost W_t.

The profit function is: $D_t(k) = \left(1+\tau^S\right)\left[P_t(k)/P_t\right] Y_t(k) - W_t N_t(k) - T_t^F$ and I assume as a benchmark that the government implements the standard NK optimal subsidy inducing marginal cost pricing: with optimal pricing, the desired markup is defined by $P_t^*(k)/P_t^* = 1 = \varepsilon W_t^* / \left[\left(1+\tau^S\right)(\varepsilon-1)\right]$ and the optimal subsidy is $\tau^S = (\varepsilon-1)^{-1}$. Financing its total cost by taxing firms ($T_t^F = \tau^S Y_t$) gives total profits $D_t = Y_t - W_t N_t$. This policy is redistributive because it taxes the holders of firm shares: steady-state profits are zero $D = 0$, giving the "full-insurance" steady-state used here $C^H = C^S = C$. Loglinearizing around it (with $d_t \equiv \ln(D_t/Y)$), profits vary inversely with the real wage: $d_t = -w_t$ (an extreme form of the general property of NK models). This series of assumptions—optimal subsidy, steady-state consumption insurance, zero steady-state profits—is not necessary for the results and could be easily relaxed, but adopting them makes the algebra simpler and more transparent.

Under nominal rigidities, optimal pricing by firms delivers an "aggregate supply", Phillips curve written in loglinearized form as:

$$\pi_t = \beta_f E_t \pi_{t+1} + \kappa c_t, \tag{2}$$

and derived in the Appendix based on Rotemberg pricing. Closed-form results are particularly useful for my analytical approach. To obtain such tractability, I first focus on the simplest possible special case, used previously in a different context in Bilbiie (2016):

$$\pi_t = \kappa c_t, \tag{3}$$

nested in (2) above with $\beta_f = 0$. This is "microfounded" in the Appendix by assuming that monopolistic firms have to pay a Rotemberg price adjustment cost relative to yesterday's market average price index, rather than relative to their own individual price (the latter leading to the forward-looking version (2)). In other words, firms ignore the impact of today's choice of price on tomorrow's profits. While clearly over-simplified, this setup nevertheless captures a key mechanism of the NK model—the trade-off between inflation and real activity—and allows us to isolate and focus on the main topic and the essence of this paper: AD.[10] The results of this paper carry through reassuringly when considering the standard Phillips curve (2), as I show in Appendix C.

The government conducts *fiscal and monetary policy.* Other than the optimal subsidy discussed above, the former consists of a simple *endogenous redistribution* scheme: taxing profits at rate τ^D and rebating the proceedings lump-sum to H: $Transfer_t^H = \frac{\tau^D}{\lambda} D_t$; this is key here for the transmission of *monetary policy*, understood as changes in the nominal interest rate i_t.

Market clearing implies for equilibrium in the goods and labor market respectively $C_t \equiv \lambda C_t^H + (1-\lambda) C_t^S = \left(1 - \frac{\psi}{2}\pi_t^2\right) Y_t$ and $\lambda N_t^H + (1-\lambda) N_t^S = N_t$. With uniform steady-state hours ($N^j = N$) by normalization and the fiscal policy assumed above (inducing $C^j = C$) loglinearization around a zero-inflation steady state delivers $y_t = c_t = \lambda c_t^H + (1-\lambda) c_t^S$ and $n_t = \lambda n_t^H + (1-\lambda) n_t^S$.

2.1 Income Processes: Variance, Skewness and Kurtosis in a-HANK

Individual income in this model follows a two-state Markov chain with values Y_t^S and Y_t^H in the respective states. An analytical characterization of the key moments of this process is useful to both illustrate an extra dimension along which this model is a representation of complex models, and for calibration and quantitative analysis.

Consider first autocorrelation, this is given by (for any of the two states $j = S, H$):

$$corr\left(Y_{t+1}^j, Y_t^j\right) = s + h - 1 = 1 - \frac{1-s}{\lambda}; \tag{4}$$

thus, the condition for stability $s \geq 1 - h$ also ensures positive autocorrelation in income process.

To calculate **conditional variance**, a measure of *income risk*, observe that an S agent's expected income tomorrow is $E_t\left(Y_{t+1}^S | Y_t^S\right) = sY_{t+1}^S + (1-s) Y_{t+1}^H$. Using this, we immediately find**:**

$$var\left(Y_{t+1}^S | Y_t^S\right) = s(1-s)\left(Y_{t+1}^S - Y_{t+1}^H\right)^2. \tag{5}$$

[10] In a Calvo setup, this amounts to assuming that 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. Essentially, 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).

Conditional skewness and **kurtosis** are also easily calculated as:

$$\begin{aligned} skew\left(Y_{t+1}^{S}|Y_{t}^{S}\right) &= \frac{1-2s}{\sqrt{s\left(1-s\right)}}; \\ kurt\left(Y_{t+1}^{S}|Y_{t}^{S}\right) &= \frac{1}{s\left(1-s\right)}-3 \end{aligned} \tag{6}$$

As standard for Bernoulli distributions there is *negative skewness* for $s > .5$ and *leptokurtosis* (positive excess kurtosis $kurt\left(.\right)-3$) outside of the $\frac{1}{2}\pm\frac{1}{\sqrt{12}}$ interval, i.e. for s smaller than 0.21 or larger than 0.79. Notice that a value of $s > 0.79$ thus ensures both negative skewness and leptokurtosis.

Of special importance in order to fit key micro facts on income distribution in the cross-section (Guvenen et al; De Nardi et al) are the relative skewness and kurtosis of the two types. In particular, evidence suggests that income of (an empirical proxy of) S is more negatively skewed and more leptokurtic. It can be easily shown, comparing (6) with the equivalent formulae for H that both of these properties (income of S both more negatively skewed and more leptokurtic) are satisfied in the model if and only if:

$$s > h.$$

To summarize, this simple two-state model features, albeit in a stylized way, some key element of the HA literature pertaining to income heterogeneity and uncertainty, in particular conditional idiosyncratic variance that can be cyclical (or not), autocorrelated income processes with left-skewness and leptokurtosis. The combined conditions for matching the key micro facts are $s > 1-h$, $s > h$ and both s and h larger than .79. We will refer to this when calibrating the model below.

2.2 Cyclical Inequality and Aggregate Demand in a-HANK

We derive an *aggregate* Euler equation, or IS curve for this economy starting from the individual Euler equation that prices the asset whose return is the central bank's instrument, the self-insurance equation for bonds (1) loglinearized around the symmetric steady state $C^H = C^S$:

$$c_t^S = sE_t c_{t+1}^S + \left(1-s\right)E_t c_{t+1}^H - \sigma\left(i_t - E_t\pi_{t+1} - \rho_t\right). \tag{7}$$

where $E_t\pi_{t+1}$ is expected inflation. Note that the nominal interest rate i_t is expressed in *levels* (to allow dealing with the zero lower bound transparently later) and ρ_t an exogenous shock that is standard in the liquidity-trap literature (Eggertsson and Woodford, 2003) and captures 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, S households try to bring consumption into the present and "dis-save", and vice versa when it decreases.

To express this in terms of aggregates, we need individual c_t^j as a function of aggregate c_t. Take first the hand-to-mouth, who consume all *their* income (loglinearize the budget constraint) $c_t^H = y_t^H = w_t + n_t^H + \frac{\tau^D}{\lambda} d_t$. Substituting $w_t = (\varphi + \sigma^{-1}) c_t$ (the wage schedule derived using the economy resource constraint, production function, and aggregate labor supply), $d_t = -w_t$ and their labor supply, we obtain H's consumption function:

$$\begin{aligned} c_t^H &= y_t^H = \chi y_t, \\ \chi &\equiv 1 + \varphi \left(1 - \frac{\tau^D}{\lambda}\right) \lesseqgtr 1, \end{aligned} \tag{8}$$

H's consumption comoves one-to-one with *their* income, but *not necessarily* with *aggregate* income, and this is the model's keystone: the parameter χ—the elasticity of H's consumption (and income) to aggregate income y_t—which depends on fiscal redistribution and labor market characteristics.

Cyclical distributional effects make χ different from 1: the other agents (S, with income $y_t^S = w_t + n_t^S + \frac{1-\tau^D}{1-\lambda} d_t$) face an additional (relative to RANK) *income effect* of the real wage, which reduces their profits $d_t = -w_t$. Using this and S's labor supply, we obtain:

$$c_t^S = \frac{1 - \lambda\chi}{1 - \lambda} y_t, \tag{9}$$

so whenever $\chi < 1$ S's income elasticity to aggregate income is *larger* than one, and vice versa. This directly delivers the following definition of **cyclical inequality** (of income) γ_t as (the log deviation of $\Gamma_t \equiv Y_t^S / Y_t^H$):

$$\gamma_t \equiv y_t^S - y_t^H = (1 - \chi) \frac{y_t}{1 - \lambda},$$

which is *procyclical* ($\partial\gamma/\partial y > 0$) iff $\chi < 1$ and *countercyclical* ($\partial\gamma/\partial y < 0$) iff $\chi > 1$.

In RANK, there are by definition no such distributional considerations: one agent works and receives all the profits. When aggregate income goes up, labor demand goes up (sticky prices) and the real wage increases. This drives down profits (wage=marginal cost), but because the *same* agent incurs both the labor gain and the "capital" (monopolistic rents) loss, the distribution of income between the two is neutral.

Income distribution matters under heterogeneity, and to understand how start with no fiscal redistribution, $\tau^D = 0$ and $\chi > 1$. If demand goes up and (with upward-sloping labor supply $\varphi > 0$) the real wage goes up, H's income increases. Their demand increases proportionally, as they do not get hit by profits falling. Thus aggregate demand increases by *more* than initially, shifting labor demand and increasing the wage even further, and so on. In the new equilibrium, the extra demand is produced by S, whose decision to work more is optimal given the income loss from falling profits. Since the income of H goes up and down more than proportionally with aggregate income, inequality is *countercyclical (CI)*: it goes down in expansions and up in recessions.

Redistribution $\tau^D > 0$ dampens this channel, delivering a lower χ. As they receive a transfer,

H start internalizing the negative income effect of profits and do not increase demand by as much. The case considered by Campbell and Mankiw's (1989) seminal paper is $\chi = 1$, which I call the *Campbell-Mankiw benchmark* (see Bilbiie (2017) for an elaboration). This occurs when the distribution of profits is uniform, so the income effect disappears $\tau^D = \lambda$; or when labor is infinitely elastic $\varphi = 0$ (so that all households' consumption comoves perfectly with the wage); income inequality is then *acyclical.*

Finally, $\chi < 1$ occurs when H receive a disproportionate share of the profits $\tau^D > \lambda$. The AD expansion is now *smaller* than the initial impulse, as H recognize that this will lead to a fall in their income; while S, given the positive income effect from increased profits, optimally decide to work less.[11] As the income of H now moves less than proportionally with aggregate income, inequality is *procyclical* (PI).

Replacing the consumption functions of H (8) and S (9) in the self-insurance equation, we obtain the *aggregate Euler-IS*:

$$\begin{aligned} c_t &= \delta E_t c_{t+1} - \sigma \frac{1-\lambda}{1-\lambda\chi}\left(i_t - E_t\pi_{t+1} - \rho_t\right), \qquad (10) \\ \text{where } \delta &\equiv 1 + (\chi - 1)\frac{1-s}{1-\lambda\chi}. \end{aligned}$$

and the contemporaneous AD elasticity to interest rates is the TANK one, $\sigma\frac{1-\lambda}{1-\lambda\chi}$. This reflects the New Keynesian Cross logic described above: in particular and as analyzed in detail in Bilbiie (2017), even though the "direct effect" of a change in interest rates is scaled down by $(1-\lambda)$ (λ agents do not respond directly), the indirect effect, which amounts to the aggregate-MPC or slope of the planned-expenditure curve in the NK cross representation, is increasing with λ. The rate at which it does so depends on χ, and with CI the latter effect dominates the former, delivering amplification relative to RANK—while for PI the opposite is true, giving dampening.

The key property for our purpose and novelty relative to TANK is summarized by:[12]

[11] An alternative route to obtaining $\chi < 1$ is to assume *sticky wages*, as Colciago (2011) and Ascari, Colciago, and Rossi (2017) in TANK, and Broer et al (2018) or Walsh (2018) is simple-HANK; χ then becomes a decreasing function of wage stickiness: as wages become less cyclical so does the income of H.

[12] I restrict attention to $\lambda < \chi^{-1}$. When this condition fails the IS curve swivels ($\frac{\partial c_t}{\partial(-r_t)} < 0$), an "inverted Aggregate Demand" region occurs (Bilbiie, 2008; Bilbiie and Straub 2012, 2013). This is a *paradox of thrift* (Keynes, 1936): S want to consume more ("save" less) as r goes down, but we end up with *lower* aggregate consumption (aggregate saving goes up).When real interest rates fall, by the Euler equation, S's consumption goes *up,* proportionally (regardless of how many H there are). The income effect of S needs to agree with this intertemporal substitution effect, so something else needs to adjust for equilibrium. Evidently, consumption of H must go *down*, so the real wage must go down. We need to be moving *downwards* along the labor supply curve, so labor demand shifts *down* (which with non-horizontal AS will also give deflation)—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 strictly speaking a "paradox of thrift", for *individual* incentives to consume more (by savers) lead to equilibrium outcomes with lower *aggregate* consumption.This is different from the paradox of thrift occurring in a liquidity trap, see e.g. Eggertsson and Krugman (2012): there, AD is upward-sloping because the nominal interest rate is fixed. Here, it is upward sloping because of aggregation

Proposition 1 *The Aggregate Euler-IS equation of a-HANK (with $1-s>0$) is characterized by:*

$$\textbf{\textit{discounting}}\ (\delta < 1) \text{ iff inequality is procyclical } (\chi < 1) \text{ and}$$
$$\textbf{\textit{compounding}}\ (\delta > 1) \text{ iff inequality is countercyclical } (\chi > 1).$$

To understand this (echoing Proposition 3 in the companion paper Bilbiie (2017)), start with RANK, where good news about future income imply a one-to-one increase in aggregate demand today as the household wants to substitute consumption towards the present and (with no assets) income adjusts to deliver this. The same also holds in the TANK limit: with permanent idiosyncratic shocks ($s=h=1$), there is no discounting $\delta=1$; λ is then an arbitrary free parameter.

Consider then the case of PI which gives "discounting", generalizing MNS (nested for $\chi=0$, implying $\delta=s$, and iid idiosyncratic shocks $s=1-h=1-\lambda$). When good news about future *aggregate* income/consumption arrive, households recognize that in some states of the world they will be constrained and (because $\chi<1$) not benefit fully from it. They self-insure against this and increase their consumption less than they would if they were alone in the economy (or if there were no uncertainty). Like in RANK and TANK, this (now: self-insurance) increase in saving demand cannot be accommodated (there is no asset), so the household consumes less today and income adjusts accordingly to deliver this allocation. The interaction of idiosyncratic $(1-s)$ and aggregate uncertainty (news about y_t, and how they translate into individual income through $\chi-1$) thus determines the self-insurance channel. This channel is strengthened and the discounting is faster: the higher the risk $(1-s)$, the lower the χ, and the longer the expected hand-to-mouth spell (higher λ at given s implies higher h); these intuitive results follow immediately by calculating the respective derivatives of δ and noticing they are all proportional to $(\chi-1)$. In the iid, idiosyncratic-uncertainty special case $s=1-h$ (considered e.g. by Krusell Mukoyama Smith and MNS) we have $\lambda=h$ and the fastest discounting $\delta_{iid}=(1-\lambda)/(1-\lambda\chi)$.

The opposite logic holds with CI, implying *compounding* instead of discounting. The (there, contemporaneous) Keynesian-cross endogenous amplification that is the staple of TANK now extends *intertemporally*: good news about future aggregate income boost today's demand because they imply less need for self-insurance. Since future consumption in states where the constraint binds over-reacts to good aggregate news, households internalize this by demanding *less* "saving". But savings still need to be zero in equilibrium, so households consume more that one-to-one—while income increases more than it would without risk. By the same token as before (δ derivatives' being proportional to $(\chi-1)$), this effect is magnified with higher risk $(1-s)$, χ, and λ; the highest compounding is obtained in the iid case, because it corresponds to the strongest self-insurance

through the mechanism emphasized above, regardless of the zero lower bound. Note that such equilibria can be ruled out, if inequality is procyclical $\chi<1$ (changes in demand do not trigger over-compensating income effects on S no mater how large the share of H).

motive, with $\delta_{iid} = (1-\lambda) / (1-\lambda\chi)$.

Furthermore, the self-insurance channel is *complementary* with the (TANK) hand-to-mouth channel: compounding (discounting) is increasing with idiosyncratic risk at a higher rate when there are more λ $\left(\partial^2 \delta / (\partial\lambda\partial(1-s)) \sim \chi - 1\right)$: an increase in $(1-s)$ has a larger effect on self-insurance with a longer expected hand-to-mouth spell $(1-h)^{-1}$.

2.3 Liquidity and Inequality: New and Intertemporal Keynesian Crosses in a-HANK

The a-HANK model just described captures two key amplification channels of HANK models. The first is, as emphasized throughout, the TANK-originating cyclical-inequality channel which gives rise to a "New Keynesian Cross"-type of amplification mechanism as reviewed in detail in Bilbiie (2017): the aggregate MPC out of an increase in aggregate income is a convex combination of the MPCs of the two types out of their own incomes. That is, we need to sum the products of share in the population, MPCs out of own income, and elasticity of own income to aggregate income. For a transitory shock, this is thus

$$mpc = (1-\lambda)(1-\beta)\frac{1-\lambda\chi}{1-\lambda} + \lambda * 1 * \chi = 1 - \beta(1-\lambda\chi).$$

This aggregate MPC is the slope of a planned expenditure, consumption function as in Samuelson's Keynesian cross. It yields amplification whenever $\chi > 1$, for then the increase in slope corresponding to adding λ agents dominates the decrease in the shift of this curve (corresponding to λ agents being directly insensitive to policy changes).

The second key amplification channel is *orthogonal* to the former. It has been labelled "intertemporal Keynesian cross" by Auclert et al (2018) and is also discussed in the context of fiscal multipliers by Hagedorn et al (2018). My a-HANK model embeds this channel and in fact provides a novel analytical representation for its key summary statistics, the intertemporal MPCs (iMPCs) unveiled by Auclert et al.

To illustrate how the iMPCs can be calculated in my a-HANK model, we need to consider the equilibrium where liquidity is provided by the government.[13] The equivalent of the individual consumption functions (8) and (9) in this case with equilibrium liquidity are (derived in detail in the Appendix A.2):

$$\begin{aligned} c_t^H &= \hat{y}_t^H + \beta^{-1}\frac{1-s}{\lambda}b_t \\ c_t^S + \frac{1}{1-\lambda}b_{t+1} &= \hat{y}_t^S + \beta^{-1}\frac{s}{1-\lambda}b_t \end{aligned} \tag{11}$$

[13] I am particularly grateful to Adrien Auclert and to Matt Rognlie and Ludwig Straub for useful discussions and for generously sharing their own analytical results in a one-agent bonds-in-utility model.

where I again approximate around the no-inequality (and zero-liquidity) symmetric long-run steady-state whereby the real interest rate is $1 + r = \beta^{-1}$. Notice that b_t denotes the total quantity of private (liquid) assets demanded at the beginning of period t, expressed as shares of steady-state total income; while $\hat{y}_t^j \equiv y_t^j - t_t^j$ denotes the disposable income of agent j. The aggregation of (11) delivers:

$$c_t = \hat{y}_t + \beta^{-1} b_t - b_{t+1}. \tag{12}$$

The iMPCs as defined by Auclert et al are the partial derivatives of aggregate consumption c_t with respect to changes in aggregate disposable income $\hat{y}_{t+k}$ at different horizons k, keeping fixed everything else (in particular, taxes and public debt); to solve for these objects, we thus need to solve for the equilibrium dynamics of private liquid assets b_t. To that end, replace the individual budget constraints (11) into the loglinearized self-insurance equation for bonds (7) (assuming without loss of generality fixed interest rates $i_t = E_t \pi_{t+1} + \rho_t$) obtaining:

$$\begin{aligned} E_t b_{t+2} - \Theta b_{t+1} + \beta^{-1} b_t &= \frac{1-\lambda}{s} \left[s E_t \hat{y}_{t+1}^S + (1-s) E_t \hat{y}_{t+1}^H - \hat{y}_t^S \right]. \\ \text{where } \Theta &\equiv \frac{1}{s} + \beta^{-1} \left[1 + \frac{1-s}{s} \left(\frac{1-s}{\lambda} - 1 \right) \right] \end{aligned} \tag{13}$$

As clear from (13), finding the derivatives of b_{t+k} with respect to aggregate disposable income $\hat{y}_t$ requires a model of how individual disposable incomes are related to aggregate, such as this paper's. Furthermore, since the calculation of iMPCs keeps *fixed by definition* all the other variables when taking the derivative with respect to disposable income (in particular taxes, their distribution, and thus public debt), the partial derivatives of individual disposable incomes with respect to aggregate disposable income are respectively $d\hat{y}_t^H = \chi d\hat{y}_t$ and $d\hat{y}_t^S = \frac{1-\lambda\chi}{1-\lambda} d\hat{y}_t$.[14] Solving the asset dynamics equation taking this into account delivers:

$$db_{t+1} = x_b db_t + \frac{1-\lambda\chi}{s} \sum_{k=0}^{\infty} (\beta x_b)^{k+1} \left(d\hat{y}_{t+k} - \delta d\hat{y}_{t+k+1} \right), \tag{14}$$

where the roots of the characteristic polynomial of (13) are $x_b = \frac{1}{2}\left(\Theta - \sqrt{\Theta^2 - 4\beta^{-1}}\right)$ and $(\beta x_b)^{-1}$, with $0 < x_b < 1$ as required by stability whenever $\beta > 1 - \frac{1-s}{\lambda}$.

Substituting (14) in aggregate consumption (12) delivers the key equation for calculating the iMPCs, the consumption function:

$$dc_t = d\hat{y}_t + \beta^{-1} (1 - \beta x_b) db_t + \frac{1-\lambda\chi}{s} \sum_{k=0}^{\infty} (\beta x_b)^{k+1} \left(\delta d\hat{y}_{t+k+1} - d\hat{y}_{t+k} \right). \tag{15}$$

[14] In particular, any model would deliver a reduced-form expression $\hat{y}_t^H = \chi \hat{y}_t + \chi_{tax} t_t$, where χ_{tax} is an equilibrium elasticiy that depends among other things on the distribution of taxes, labor supply elasticity, etc. But for calculating iMPCs, we look at a partial equilibrium whereby $dt_t / \hat{y}_t = 0$.

The analytical expressions for the iMPCs are directly obtained from this and emphasized in Proposition 2.[15]

Proposition 2 *The intertemporal MPCs (iMPCs) for the a-HANK model, in response to a one-time shock to disposable income at any time T and for any $t \geq 0$: (i) are given by:*

$$\frac{dc_t}{d\hat{y}_T} = \begin{cases} \frac{1-\lambda\chi}{s}\frac{\delta-\beta x_b}{1-\beta x_b^2}\left(\beta x_b\right)^{T-t}\left(1 - x_b + x_b\left(1-\beta x_b\right)\left(\beta x_b^2\right)^t\right), & \text{if} \quad t \leq T-1; \\ 1 - \frac{1-\lambda\chi}{s}\beta x_b - \left(\delta - \beta x_b\right) x_b \frac{1-\lambda\chi}{s}\left(1-\beta x_b\right)\frac{1-\left(\beta x_b^2\right)^T}{1-\beta x_b^2}, & \text{if} \quad t = T; \\ \frac{1-\lambda\chi}{s}\frac{1-\beta x_b}{1-\beta x_b^2} x_b^{t-T}\left(1 - x_b\delta + x_b\left(\delta - \beta x_b\right)\left(\beta x_b^2\right)^T\right), & \text{if} \quad t \geq T+1. \end{cases}$$

and (ii) are increasing with the cyclicality of inequality χ when $t < T$ and decreasing with χ when $t \geq T > 0$ (keeping fixed the time-0 contemporaneous MPC $dc_0/d\hat{y}_0$).

It is useful, in order to isolate this *liquidity-amplification channel*, to start with the benchmark case (considered by Auclert et al's paper that discovered it) of acyclical inequality $\chi = 1$. This amounts to replacing individual disposable incomes with aggregate disposable income $\hat{y}_t^j = \hat{y}_t$, which delivers the same expressions as in Proposition 2 with $\chi = 1$ and $\delta = 1$. The intertemporal path of the iMPCs is apparent already in this special case: faced with a current income shock, agents optimally self-insure by saving in liquid wealth to maintain a higher level of consumption in the future. At the same time, when facing a higher future income shock agents consume in anticipation, decreasing their stock of liquid savings.

Adding the cyclical-inequality channel affects this through the mechanics described previously: when self-insuring against the risk of being constrained in the future, agents now take into account how the aggregate income shock affects their income in each respective state and change their demand for assets (and, with positive liquidity, their equilibrium liquidity holdings) consequently. Ceteris paribus, countercyclical inequality $\chi > 1$ leads to a higher contemporaneous MPC but to lower future MPCs (without affecting persistence as described by x_b which is independent of χ). Persistence is instead increasing with the share of hand-to-mouth and decreasing with the level of idiosyncratic risk (it can be directly verified that $\partial x_b / \partial \lambda > 0$ and $\partial x_b / \partial (1-s) < 0$).

A useful, sharp special case occurs in the limit as agents oscilate between the two states every other period, with the mass of half of the agents in each state in every period: $s = 0$ and $\lambda = 1/2$; this is the case studied by Woodford (1990). The asset accummulation equation (13) becomes:

$$b_{t+1} = \frac{\hat{y}_t^S - E_t \hat{y}_{t+1}^H}{2\left(1+\beta^{-1}\right)} = \frac{1}{2\left(1+\beta^{-1}\right)}\left(\frac{1-\lambda\chi}{1-\lambda}\hat{y}_t - \chi E_t \hat{y}_{t+1}\right), \tag{16}$$

[15] The iMPCs are by construction partial-equilibrium objects. It is nevertheless straightforward in my model to also compute the general-equilibrium effects of truly-exogenous income changes: that is, the effects of public-debt financed tax cuts (transfer increases).

with agents (dis-)saving when they expect (higher) lower income tomorrow. The consumption function follows immediately by substituting into (12):

$$c_t = \frac{2-\chi+\beta\chi}{2(1+\beta)}\hat{y}_t + \frac{2-\chi}{2(1+\beta)}\hat{y}_{t-1} + \frac{\beta\chi}{2(1+\beta)}\hat{y}_{t+1}; \tag{17}$$

higher income cyclicality in the constrained state χ makes agents want to consume more (save less) out of news of aggregate income and consume less (save more) out of past and current aggregate income. This expression allows the most transparent illustration of the dependence upon χ and, while not meant as a good quantitative approximation in other dimensions (for instance, higher moments of the income process), does deliver reasonable iMPCS: in the acyclical-inequality $\chi = 1$ case for example, the contemporaneous MPC $dc_0/d\hat{y}_0 = 0.5$ while the one-year-after MPC is $dc_1/d\hat{y}_0 = 0.256$.

An important observation is that TANK, the other extreme with $s = 1$, misses this intertemporal amplification altogether: without idiosyncratic risk, the roots of the debt dynamics equation are 1 and β^{-1}. The consumption function is thus (15) but with $s = 1$, $\delta = 1$, and $x_b = 1$, so the TANK iMPCs are:

$$\begin{aligned} \frac{dc_T}{d\hat{y}_T} &= \lambda\chi + (1-\lambda\chi)(1-\beta)\beta^T \text{ and} \\ \frac{dc_t}{d\hat{y}_T} &= (1-\lambda\chi)(1-\beta)\beta^T \ \forall t \neq T \end{aligned}$$

Therefore, TANK misses the intertemporal path of iMPCs, in particular when it comes to past income shocks, which are of the essence in the data and the very motivation for introducing these model objects, as argued forcefully by Auclert et al.

Figure 1 illustrates this by plotting the iMPCs for four cases: the acyclical-inequality case that is akin to Auclert et al's quantitative HANK model; the TANK model, and the a-HANK model encompassing both liquidity and cyclical inequality (for pro- and counter-cyclical inequality). In the upper panel, look at a date-0 shock and calibrate the a-HANK with acyclical inequality to closely follow Auclert et al, i.e. $\beta = 0.8$ and $\lambda = 0.5$; this requires $s = 0.84$ to match both the contemporaneous and next-year MPCs (0.55 and 0.15 respectively). The discount rate is very large, even for the yearly calibration adopted here; in the models with cyclical inequality (both TANK and a-HANK) I set $\beta = 0.95$ and match the two target MPCs with $\lambda = 0.33$, $s = 0.82$ and $\chi = 1.4$. This is remarkably close to the (quarterly) calibration used in Bilbiie (2017) to match other (aggregate, *general-equilibrium*) objects with the same model.

The intertemporal path of the iMPCs is remarkably in line with that documented by Auclert et al in the data; in particular, the effect of the income shock dies off a few years after; whereas the model with acyclical inequality implies unrealistically high persistence while TANK implies no persistence at all. The reverse side of it is that, as clear from the bottom panel that compares iMPCs out of current and future income shocks for the a-HANK with acyclical and countercyclical inequality,the latter implies larger iMPCs out of future income—an illustration of part (ii) of the Proposition; this is due, intuitively, to the same self-insurance forces that generate Euler-compounding in general equilibrium as illustrated in the previous section. It can in fact be shown directly by differentiation of the analytical expressions in Proposition 2 that the iMPCs out of future income (news) are increasing in χ while the iMPCs out of past income are decreasing in χ.

An important remark is that *counter*cyclical inequality is, nevertheless, *not necessary* for the a-HANK model to match the iMPCs. Indeed, the model with *pro*cyclical inequality $\chi < 1$ also does it. To illustrate this, consider the model with $\chi = 0.8$. Clearly, we need to re-calibrate the model for a lower χ implies, by the logic of the cyclical-inequality channel, a lower contemporaneous MPC and a higher MPC out of past income; matching the two MPCs thus requires re-calibrating $\lambda = 0.64$ and $s = 0.74$. The resulting path (the thin solid line in the Figure) illustrates our intuition: the MPC out of past income is virtually identical (not surprising given that we matched the one-period-ago MPC). But the whole path of the "forward" MPCs is below the countercyclical-inequality case (with the acyclical-inequality case between the two), which is a direct implication of the Euler discounting through δ discussed at length above. Notice, however, that discounting/compounding in the Euler equation is not of the essence for matching the iMPCs (although it certainly matters quantitatively)—but idiosyncratic risk *is*.

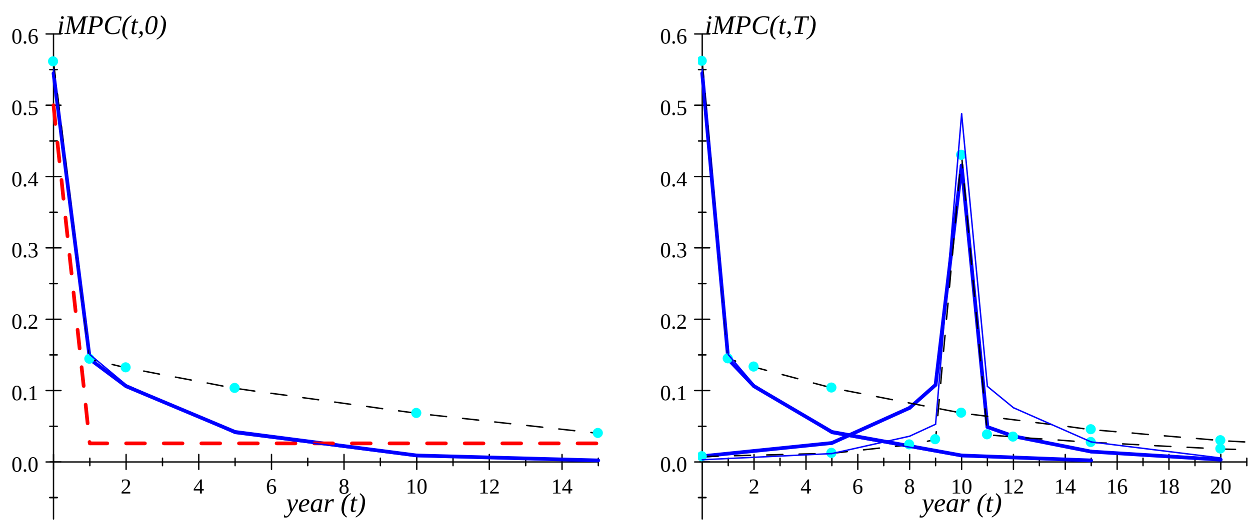


Figure 1: iMPCs in a-HANK with $\chi = 1$ (thin black dot-dash); TANK (red dash); a-HANK with counter- and pro-cyclical inequality (thick and thin blue solid). Left: $T = 0$; right: $T = 0; 10$

To conclude, my a-HANK model captures analytically in a realistic and flexible way a key amplification mechanism at work in quantitative HANK models and one that misses from TANK models. The baseline *annual* calibration that matches the empirical iMPCs under countercyclical inequality has $\lambda = 0.33$ and $s = 0.82$; this implies, using the formulas for income processes, the conditional skewness and excess kurtosis of S's income process of -1.66 and 0.77. In addition, this calibration implies a quarterly autocorrelation of idiosyncratic income processes of $(s + h - 1)^{\frac{1}{4}} =$

$\left(1-\frac{1-s}{\lambda}\right)^{\frac{1}{4}}=0.819$ (corresponding to the quarterly transition probability $1-s=1-0.82^{\frac{1}{4}}=1-0.952\simeq .04$). Given the very coarse two-state discretization employed here, it is no surprise that these moments are not perfectly aligned with their micro-data counterparts; they nevertheless capture the key feature of concomitant left-skewness and leptokurtosis (a standard discretization with more states usually matches these moments very well) as well as persistent and conditionally volatile idiosyncratic income.

3 NK Analytics with a-HANK: Determinacy, Puzzles, Amplification, and Optimal Policy

Hoping to have convinced the reader that it is a reasonable reduced-form representation of more complex, richer-heterogeneity quantitative HANK models, I now exploit the a-HANK's tractability to conduct—with pencil and paper—a full-fledged analysis of the main topics customarily analyzed in RANK: determinacy properties of interest-rate rules, solving the FG puzzle, conditions for amplification-multipliers, and optimal monetary policy (in normal times and in liquidity traps).

3.1 HANK, Taylor, and Sargent-Wallace

The model is completed by adding the simple aggregate-supply, Phillips-curve specification (3) (all the results carry through with the more familiar forward-looking NKPC (2) as I show in Appendix C) and a monetary policy rule. As a benchmark, the central bank sets the nominal rate i_t according to a Taylor rule:

$$i_t=\rho_t+i_t^*+\phi\pi_t, \tag{18}$$

where the intercept of the Taylor rule i_t^* is an exogenous (possibly persistent) process, and dealing with the zero lower bound ZLB amounts to adding the constraint $i_t\geq 0$.

With this simplified, RANK-isomorphic HANK we can similarly derive the classic determinacy results: a (HANK-)Taylor principle and the Sargent-Wallace issue of determinacy under a peg; further below, I study a Wicksellian rule of price-level targeting and then optimal policy. Under the assumed structure, the model is disarmingly simple: replacing the static Phillips curve (3) and Taylor rule (18) in the aggregate Euler equation, the *whole* analytical-HANK model boils down to *one* equation (using the notation $\bar{\sigma}\equiv\sigma\left(\frac{1-\lambda\chi}{1-\lambda}+\phi\kappa\sigma\right)^{-1}$):

$$\begin{aligned} c_t &= \nu E_t c_{t+1}-\bar{\sigma}i_t^*, \qquad (19)\\ \text{where } \nu &\equiv \frac{\delta+\kappa\sigma\frac{1-\lambda}{1-\lambda\chi}}{1+\phi\kappa\sigma\frac{1-\lambda}{1-\lambda\chi}} \end{aligned}$$

captures the effect of *good news* on AD, and the elasticity to interest rate shocks.[16]

There are three channels shaping this key summary statistic. First, the "pure AD" effect through δ discussed above (operating even when prices are fixed or if the central bank fixes the ex-ante real rate $i_t = E_t\pi_{t+1}$), coming from cyclical inequality.

The second term comes from a supply feedback *cum* intertemporal substitution: the inflationary effect (κ) of good news on income triggers, *ceteris paribus* (given nominal rates) a fall in the real rate and intertemporal substitution towards today—the magnitude of which depends on the within-the-period amplification/dampening resulting from cyclical inequality ($\frac{1-\lambda}{1-\lambda\chi}$).

Finally, through the monetary policy rule all this current demand amplification generates inflation and triggers movements in the real rate. When $\phi > 1$ ("active" policy in Leeper's (1991) terminology), inflation leads to an increase in the real rate, which has a contractionary effect today—the strength of which also depends on the "TANK" cyclical-inequality channel through $\frac{1-\lambda}{1-\lambda\chi}$. These considerations drive the main result concerning equilibrium determinacy and ruling out sunspot equilibria (a version of the Proposition for the standard case with forward-looking NKPC (2) is in Appendix C.1).

Proposition 3 ***The HANK Taylor Principle****: The HANK model under a Taylor rule (19) has a determinate, (locally) unique rational expectations equilibrium if and only if (as long as $\lambda < \chi^{-1}$):*

$$\nu < 1 \Leftrightarrow \phi > \phi_{HANK} \equiv 1 + \frac{\delta - 1}{\kappa\sigma\frac{1-\lambda}{1-\lambda\chi}}.$$

The ***Taylor principle*** *$\phi > 1$ is sufficient for determinacy if and only if there is Euler-IS discounting:*

$$\delta \leq 1.$$

The proposition follows by recalling that the requirement for a (locally) unique rational expectations equilibrium is that the root ν be inside the unit circle; in the discounting case $\delta < 1$, the threshold ϕ is evidently *weaker* than the Taylor principle, while in the compounding case it is *stronger*.

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 to rule out self-fulfilling sunspot equilibria. The higher the risk $(1-s)$ and the higher the

[16] This determinacy analysis generalizes to the case of endogenous liquidity (Section 2.3), insofar as we focus on the equilibrium locally approximated around a steady state with zero public debt (liquidity). The latter assumption implies de facto a passive (Leeper) or "Ricardian" (Woodford) fiscal policy rule and, under the additional assumption that the debt accumulation equation is stable, the AD side described by (19) is uncoupled from the debt-accumulation equation. There are potentially interesting implications for fiscal theory when relaxing these assumptions (with positive debt in the long run), that I pursue in current work; Hagedorn (2017) shows a novel route to determine prices in (incomplete-markets) economies with positive debt demand: a fiscal-monetary mix entailing the setting of nominal liabilities, the setting of nominal taxes so as to balance the budget intertmeporally, and a subordination of interest-rate policy.

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 demand is dampened, the Taylor principle is sufficient for determinacy.

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 $\chi = 1$ (acyclical inequality) or $s \to 1$ (no risk); or for $\kappa \to \infty$ (flexible prices). But the determinacy region for ϕ squeezes very rapidly with countercyclical inequality when prices are sticky, because of the complementarity between idiosyncratic and aggregate risk, as clear from the expression: $\phi_{HANK} = 1 + \frac{(\chi-1)(1-\tilde{s})}{\kappa\sigma(1-\lambda)}$. On the other hand, cyclical inequality can deliver sufficiency of the Taylor principle if procyclical enough.

Figure 2 illustrates these effects by plotting the Taylor coefficient threshold as a function of the hand-to-mouth share λ (the domain of which is $\lambda < \chi^{-1}$) for different idiosyncratic risk $(1-s)$, distinguishing between PI $\chi = 0.5$ in the left panel, and CI $\chi = 2$ in the right panel. The illustrative parametrization assumes $\kappa = 0.02$, $\sigma = 1$, and $\varphi = 1$.

Start with the right panel with CI ($\delta > 1$ and $\chi > 1$) whereby the Taylor principle is not sufficient for determinacy. The threshold increases with λ and (by complementarity) at a faster rate with higher idiosyncratic uncertainty $1-s$: the dotted line corresponds to highest possible level of idiosyncratic risk, the iid case $1-s = \lambda$, the solid line to $1-s = 0.04$ and the red dashed line to the TANK limit $1-s = 0$ (the same threshold as for RANK $\chi = 1$, the standard Taylor principle). The required response can be large: e.g. for the calibration used in Bilbiie (2017) to replicate the aggregate outcomes of KMV's quantitative HANK ($\chi = 1.48$, $\lambda = 0.37$, $1-s = 0.04$) the threshold is $\phi_{HANK} = 2.5$, while for the calibration replicating the aggregate implications of Debortoli and Gali's HANK model ($\chi = 2.38, \lambda = 0.21, 1-s = 0.04$) it is $\phi_{HANK} = 4.5$.

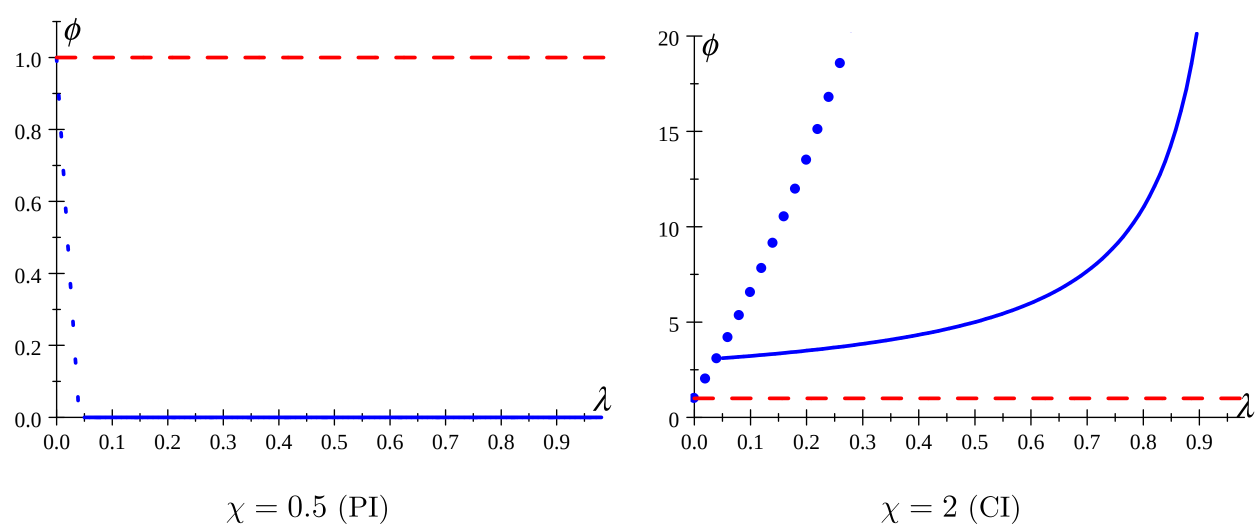


Fig. 2: Taylor threshold ϕ_{HANK} in TANK $1-s=0$ (dash); 0.04 (solid); λ (dots).

The left panel pertains to the PI, "discounting" region ($\chi < 1$), whereby the Taylor principle is *sufficient*—but *not necessary*—for determinacy: in fact, for a large subset of the region, there is determinacy even under a peg, an illustration of the following Corollary.

Corollary 1 ***Sargent-Wallace in HANK****: An interest rate* **peg** $\phi = 0$ *leads to a locally unique equilibrium (determinacy) if and only if*

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

With enough endogenous dampening, be it directly through Euler-equation discounting (the first term) or through mitigating the "expected inflation" channel (the second term), a pure expectation shock has no effects, even with a peg: the sunspot is ruled out inherently by the economy's endogenous forces (unlike in RANK where $\nu_0 = 1 + \kappa\sigma \geq 1$), as illustrated in the left panel. These considerations are intimately related to the FG puzzle to which we now turn.

3.2 When HA cures the FG Puzzle

Using our analytical framework, we are now in a position to provide closed-form conditions under which the HANK model solves NK puzzles, thus substantiating the mechanism at work in the quantitative papers that have focused on this previously—MNS (2016), KMV's (2017) note, as well as the more recent Hagedorn, Manovskii, and Mitman (2018).

Most of the (RA)NK model's properties pertaining to determinacy, amplification, and liquidity traps can be understood as stemming from one key composite parameter: the effects of news on AD under a peg. This is the root/eigenvalue in the simplified RANK model presented here, and is a fortiori on the "wrong" side of the unit circle, $\nu_0 > 1$: the Sargent-Wallace result of indeterminacy under a peg. Solving the RANK puzzles therefore boils down to introducing model features that bring this root inside the unit circle, so that news do not get compounded and there is determinacy under a peg; this is indeed how introducing heterogeneity solves the FG puzzle—but there is also a catch, that we will come back to after.

In a nutshell, HANK models solve the puzzle if and only if the HANK-AD channels emphasized above yield *enough* AD discounting to overturn the compounding through the AS side that is inherent in RANK and causes the trouble—as formalized in Proposition 4 (which pertains to the static Phillips Curve (3), but extends to the more familiar case with NKPC, the slightly more involved condition and the proof for which are outlined in Appendix C.2).

Proposition 4 *The analytical HANK model under a peg solves the FG puzzle ($\frac{\partial^2 c_t}{\partial\left(-i^*_{t+T}\right)\partial T} < 0$) if and only if:*

$$\nu_0 < 1.$$

Before proving the Proposition, it is worth discussing the necessary and sufficient condition $\nu_0 < 1$, which provides the main intuition. In light of our previous discussion in RANK, this requires that:

$$1-\delta > \kappa\sigma\frac{1-\lambda}{1-\lambda\chi},$$

i.e. that the novel, HANK-AD discounting on the left side dominate the AS-compounding of news (right side) that we identified as the source of trouble in RANK. In particular, Euler-equation discounting ($\delta < 1$) is a necessary, but *not sufficient* (unless prices are fixed) condition to solve the puzzle.

The necessary and sufficient conditions are, *jointly* (i) some idiosyncratic uncertainty $1-s > 0$, and (ii) procyclical enough inequality $\chi < 1 - \sigma\kappa\frac{1-\lambda}{1-s} < 1$, a clear manifestation of the *complementarity* between these two channels.

The Proposition's proof is immediate. Iterating forward the one equation that describes the entire HANK model (19) under a peg we obtain:

$$c_t = \nu_0 E_t c_{t+1} - \sigma\frac{1-\lambda}{1-\lambda\chi} i_t^* = \nu_0^{\bar{T}} E_t c_{t+\bar{T}} - \sigma\frac{1-\lambda}{1-\lambda\chi} E_t \textstyle\sum_{j=0}^{\bar{T}-1} \nu_0^j i_{t+j}^*$$

For any $T \in \left(t, \bar{T}\right)$ in response at time t to a one-time cut in interest rates at $t+T$ is

$$\frac{\partial c_t}{\partial\left(-i_{t+T}^*\right)} = \sigma\frac{1-\lambda}{1-\lambda\chi}\nu_0^T$$

which can now be decreasing in T if and only if $\nu_0 < 1$ (the derivative being $\sigma\frac{1-\lambda}{1-\lambda\chi}\nu_0^T \ln\nu_0$). Furthermore, since with $\nu_0 < 1$ the term $\nu_0^{\bar{T}} E_t c_{t+\bar{T}}$ vanishes when taking the limit as $\bar{T} \to \infty$, we can solve the equation forward for arbitrary i_t^* process and find a *unique* solution.[17]

One side implication of my results is an alternative interpretation of MNS's (2016) resolution of the FG puzzle, relative to that provided by Werning (2015)—that the power of FG is mitigated with incomplete markets through procyclical income risk. The independent channel that I emphasize dampens FG power through procyclical inequality, even though income risk is acyclical. Take for example acyclical income of H ($\chi = 0$), which gives $\delta = s$ and the effect of news is $\nu_0 = s + (1-\lambda)\sigma\kappa$; this is *not* necessarily smaller than 1: case in point, TANK, where it is larger than one since $s = 1$. To solve the FG puzzle, there needs to be enough idiosyncratic risk, namely $1 - s > (1-\lambda)\sigma\kappa$. It is worth noticing that MNS (2017) inherently satisfies these conditions because it assumes iid idiosyncratic risk ($s = 1-\lambda$) and exogenous income of H ($\chi = 0$). Notice that with fixed prices $\kappa = 0$ the requirement becomes $\delta < 1$: Euler-equation *discounting* and thus

[17] The same condition rules out neo-Fisherian effects in the a-HANK model, in the sense that $\frac{\partial c_t}{\partial i_t^*}$ is negative and uniquely determined. Under an AR(1) process for i^*, the (now *unique*) solution is $c_t = -\sigma\frac{1-\lambda}{1-\lambda\chi}\frac{1}{1-\nu_0\mu}i_t^*$; interest rate *increases* are short-run *contractionary* and *deflationary* (no neo-Fisherian effects). This was a theme of a previous version but is now studied in a separate paper.

$\chi < 1$ is then sufficient to solve the FG puzzle, as already shown in Bilbiie (2017).

Figure 3 provides a quantitative illustration of the findings, plotting the threshold level of redistribution that is sufficient to deliver determinacy under a peg and thus rule out the NK puzzles, for different values of idiosyncratic uncertainty and as a function of λ. Close to the TANK limit (small $1 - s$) there is no level of redistribution that delivers this; as idiosyncratic risk $1 - s$ increases, the region expands and is largest in the iid case. (The crosses represent the threshold above which the IS slope is positive $\lambda\chi < 1$).

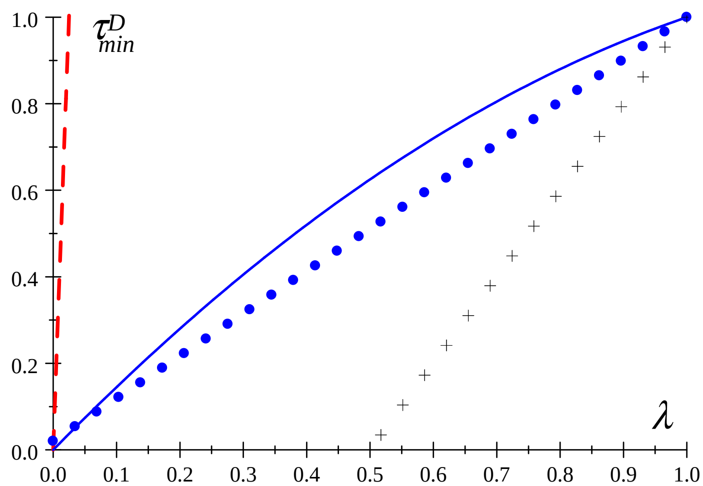


Fig. 3: Redistribution threshold $\tau^D_{\min}$ in TANK $1 - s \to 0$ (dash); 0.04 (solid); λ (dots).

3.3 Amplification Without Puzzles: A Catch-22?

To summarize the previous findings in one sentence: the a-HANK model *can* cure the FG puzzle but only when inequality is procyclical. Unfortunately, this is *the exact opposite* of the condition needed for this model to provide amplification ("multipliers") relative to RANK: conditional on this channel, that requires countercyclicality. But in that region, NK puzzles are in fact aggravated: multipliers multiply not only the good, but also the bad.

The majority of quantitative HANK studies reviewed in the Introduction use these models to deliver "amplification" of various shocks and policies with respect to the RANK benchmark. For example KMV use their HANK model to argue that it yields higher total effect of monetary policy changes (than RANK), and this is driven by "indirect", general-equilibrium forces; similar insights apply to Auclert (2016) and Gornemann et al (2015). Bilbiie (2017) compares the aggregate implications of the analytical HANK outlined here (and of TANK, also the focus of Debortoli and Galí, 2017) with that of KMV, and calibrates the simple model to match the aggregate predictions of the quantitative model. Particular values aside (see our discussion of Figure 2), a feature of the quantitative model necessary to yield that amplification is (some version of) $\chi > 1$.

Here, I use the analytical framework to illustrate the conditions for another form of amplification that has bene studied in this literature: fiscal multipliers, understood as the positive effect on private consumption of an increase in public spending. To illustrate this point, assume that the government buys an amount of goods G_t with zero steady-state value ($G = 0$) and taxes all agents uniformly in order to finance this;[18] straightforward derivation leads to the modified aggregate Euler-IS curve:

$$\begin{aligned} c_t &= \delta E_t c_{t+1} - \sigma \frac{1-\lambda}{1-\lambda\chi} \left(i_t - E_t \pi_{t+1} - \rho_t\right) + \frac{\lambda\zeta}{1-\lambda\chi} \left(\chi - 1\right) \left(g_t - E_t g_{t+1}\right) \\ & + \zeta \left(\delta - 1\right) E_t g_{t+1} \end{aligned} \tag{20}$$

where the new parameter $\zeta \equiv \left(1 + \varphi^{-1}\sigma^{-1}\right)^{-1}$ 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 (as such, it is also the elasticity of H consumption to a transfer).

The static Phillips curve becomes $\pi_t = \kappa c_t + \zeta\kappa g_t$, which together with (20) and the Taylor rule and using an AR(1) process for spending $E_t g_{t+1} = \mu_g g_t$ delivers the a-HANK fiscal multiplier emphasized in the Proposition.

Proposition 5 *The fiscal multiplier in HANK (the effect on private consumption of an increase in public spending financed by uniform lump-sum taxes) is:*

$$\frac{\partial c_t}{\partial g_t} = \frac{1}{1-\nu\mu_g} \frac{1}{1+\phi\kappa\sigma\frac{1-\lambda}{1-\lambda\chi}} \left[\underbrace{\zeta\left(\chi - 1\right) \frac{\lambda\left(1-\mu_g\right) + \left(1-s\right)\mu_g}{1-\lambda\chi}}_{\textit{TANK + HANK AD}} - \underbrace{\kappa\zeta\sigma\frac{1-\lambda}{1-\lambda\chi}\left(\phi - \mu_g\right)}_{\textit{RANK AS}} \right]. \tag{21}$$

The last term is the by now well-understood RANK channel: one the one hand, spending is inflationary (by creating extra demand), which via the Taylor rule leads to an increase in nominal rates today; on the other hand, if spending persists ($\mu_g > 0$) it creates expected inflation, which reduces the *real* rate, generating intertemporal substitution towards the present: But under an active Taylor rule $\phi > 1 > \mu_g$ the former effect always dominates the latter. Insofar as the interest-elasticity can be amplified or dampened in HANK and TANK, this AS-channel is correspondingly amplified or dampened through both $\frac{1-\lambda}{1-\lambda\chi}$ and ν.

But this is not the most important modification brought about by HANK and TANK; indeed, positive multipliers can occur regardless of the RANK AS-channel (fixed prices $\kappa = 0$). The necessary condition is, once more: *countercyclical inequality*—$\chi > 1$; thereby, an increase in G, even with zero persistence, has a demand effect that translates into an increase in labor demand,

[18] The implicit redistribution of the taxation scheme used to finance the spending is of the essence for the effect of the spending increase—see Bilbiie (2017) in the context of the analytical HANK: I abstract from that here by assuming uniform taxation to isolate the pure multiplier effect. See Bilbiie, Monacelli, and Perotti (2013) for a detailed analysis of the effects of redistribution/transfers in a TANK model, Oh and Reis (2012) for one of the earliest HANK models focusing on transfers, Ferrière and Navarro (2018) for a HANK model with tax progressivity and Hagedorn et al (2017) for fiscal multipliers in a HANK model.

wages, the income of H, and so on: the "new Keynesian cross" channel.[19] If the fiscal stimulus is expected to persist ($z > 0$), there is a multiplier due to self-insurance—as agents expect higher demand and higher aggregate income, with $\chi > 1$ they expect even higher income in the H state and thus less need to self-insure today.

To summarize, **amplification** relative to RANK (in the form of larger monetary or fiscal multipliers), which is what most quantitative HANK models have been used for, requires *necessarily*:

$$\chi > 1; \tag{22}$$

But, as we have shown above, this poses determinacy challenges and aggravates the FG puzzle—hence the "Catch-22".

3.4 Cyclical Inequality and Risk, and the Catch-22

The foregoing focuses on cyclical inequality and embeds a notion of idiosyncratic risk that is intimately related to whether liquidity constraints bind or not but is by construction *acyclical.* This key point can be formally illustrated, first, by referring to the standard measure of idiosyncratic risk, the (conditional) variance of idiosyncratic income for an agent S who contemplates self-insurance computed in (5). Its derivative with respect to aggregate income Y_{t+1}, *evaluated at the steady state*, is proportional to steady-state inequality $Y^S - Y^H$; thus, locally around a symmetric steady-state $Y^S = Y^H$ idiosyncratic risk as measured by the variance of idiosyncratic income is *acyclical.*

A further useful special case is the limit of my model that coincides with Woodford (1990):[20] $s = 0$, with agents oscilating between the two states every other period and the mass of each island being $\lambda = 1 - \lambda = \frac{1}{2}$; in this case too risk is acyclical: in fact, one could say that there is no risk at all for the conditional variance of individual income is nil, even though the *unconditional* variance of idiosyncratic income is positive $\lambda(1-\lambda)\left(Y_{t+1}^S - Y_{t+1}^H\right)^2 = \frac{1}{4}\left(Y_{t+1}^S - Y_{t+1}^H\right)^2$ and is still time-varying. Yet even in that extreme case my model implies discounting-compounding in the Euler equation with, replacing $s = 0$ and $\lambda = 1/2$ in (10):

$$\delta|_{s=0} = \frac{\chi}{2-\chi},$$

which implies discounting with procyclical inequality and compounding with countercyclical.

These two observations illustrate clearly that cyclical risk is *not necessary* for obtaining discounting/compounding in the Euler equation: cyclical inequality is sufficient, combined with idiosyncratic uncertainty even when risk is acyclical.

[19] This channel is at work in GLV's (2007) earliest quantitative model on this topic (where it was nevertheless convoluted with other channels), as well as in Bilbiie and Straub (2004), and Bilbiie, Meier and Mueller (2008)—all in TANK; it is also at play in Eggertsson and Krugman's (2012) borrower-saver model.

[20] I am grateful to Keshav Dogra for suggesting this special case.

In quantitative HANK models risk is most often cyclical, as in the data. Other analytical HANK frameworks model idiosyncratic risk in a way that is both *cyclical* and differently related (Challe et al, 2017; Ravn and Sterk, 2017; Werning, 2015) or unrelated (Acharya and Dogra, 2018) to constraints' being binding and thus to hand-to-mouth behavior. In this section, I propose an extension—inspired by Acharya and Dogra, although formally very different—that models *cyclical risk* separately and allows disentangling its role from *cyclical inequality*—thus clarifying the differences with the papers cited above.

Consider in particular that the probability of becoming constrained next period depends on the cycle, $1-s(Y_t)$, e.g. on today's aggregate consumption (in a model with endogenous unemployment risk like Ravn and Sterk's or Challe et al's, this happens in equilibrium through search and matching). If the first derivative of $1-s(.)$ is positive $-s'(Y_t)>0$, the probability to become constrained is higher in expansions: insofar as being constrained leads on average to lower income, income "risk" is then *procyclical* (it goes up in expansions). Conversely, when $-s'(Y_t)<0$ income risk is *countercyclical.*

With this small extension that captures a mechanism emphasized by the literature cited above, the Aggregate Euler-IS curve in loglinearized form, derived in detail in Appendix B, becomes:

$$\begin{aligned} c_t &= (\underbrace{\delta}_{\text{cycl.-ineq HANK}} + \underbrace{\eta}_{\text{cycl.-risk HANK}}) E_t c_{t+1} - \underbrace{\sigma\frac{1-\lambda}{1-\lambda\chi}\left(i_t - E_t\pi_{t+1} - \rho_t\right)}_{\text{cyclical-inequality TANK}} \qquad (23) \\ \text{with } \eta &\equiv \frac{s_Y Y}{1-s}\left(1-\Gamma^{-1/\sigma}\right)(1-\tilde{s})\,\sigma\frac{1-\lambda}{1-\lambda\chi}, \end{aligned}$$

where I denote by $1-\tilde{s}=\frac{(1-s)\Gamma^{1/\sigma}}{s+(1-s)\Gamma^{1/\sigma}}>1-s$ the inequality-weighted transition probability, the relevant inequality-adjusted measure of risk given steady-state inequality coming from financial income $\Gamma\equiv Y^S/Y^H\geq 1$. Notice that the discounting/compounding parameter due to cyclical inequality has a slightly different expression now $\delta\equiv 1+\frac{(\chi-1)(1-\tilde{s})}{1-\lambda\chi}$, generalized to the case with steady-state inequality; it follows that when approximated around the steady-state with inequality, idiosyncratic risk is (happens to be) cyclical too since the cyclicality of the variance is proportional to the cyclicality of inequality; but as clear from our preceding discussion, this is not essential for Euler discounting/compounding.

In this representation, the novel composite parameter η captures the aggregate implications of cyclical risk, a key determinant of which is the elasticity of idiosyncratic risk to the cycle $-s_Y Y/(1-s)$. This captures in a simple way the *different* channel emphasized by Werning (2015) and studied in isolation in a different simplified-HANK setup (with CARA preferences) by Acharya and Dogra (2018). As in those frameworks, dampening/amplification of future shocks (only) occurs depending on whether *risk* is pro- or counter-cyclical, i.e. on the sign of η—even in the Campbell-Mankiw acyclical-inequality benchmark $\chi=1$ whereby $\delta=1$. Procyclical risk (PR) implies *dampening* and Euler *discounting* $1+\eta<1$: a cut in interest rates or good news generate an expansion today—to start with. But this increases the probability of moving to the bad state, which triggers "precautionary" saving, thus containing the expansion. Conversely, countercyclical risk (CR, $s_Y > 0$) generates *amplification and compounding* $1 + \eta > 1$: an aggregate expansion reduces the probability of moving to the bad state and mitigates the need for insurance—thus amplifying the initial expansion.[21] This formalization of cyclical risk has thus similar reduced-form AD implications pertaining to the link between current and future consumption to the cyclical-inequality channel that my work emphasizes, even though the underlying economic mechanism is very different.

Notice that the risk channel operates *only if* there is long-run inequality $\Gamma > 1$, i.e. literally income *risk* of moving to a *lower income level*; whereas the cyclical-inequality channel (purposefully derived first for the case of no long-run inequality) relies on the idiosyncratic cyclicality of income χ (the cyclicality of inequality being $1 - \chi$).

This model thus captures three distinct AD channels relative to RANK: (i) the TANK channel of cyclical inequality without risk operating in e.g. Bilbiie (2008, 2017) and Debortoli and Galí (2018); (ii) the HANK-specific, cyclical-inequality component due to self-insurance—essentially adding *acyclical* idiosyncratic uncertainty to cyclical inequality, introduced in Bilbiie (2017) and in the previous section; and (iii) a second separate HANK-specific cyclical-risk channel that interacts with the previous but operates even in the limit cases with little to no risk $s \to 1$ or acyclical inequality ($\chi = 1$)—this channel is studied in isolation by Acharya and Dogra (2018) in a pseudo-RANK abstracting from heterogeneity and inequality to focus on cyclical risk (the *exact opposite* of TANK). My extension here is inspired by that analysis but provides a different formalization of cyclical risk that is intimately related to binding constraints and thus to inequality.[22]

The following proposition emphasizes the conditions under which this additional channel of cyclical risk can, by providing an additional and unrelated source of Euler-discounting, help the a-HANK model resolve the Catch-22—if inequality is countercyclical and risk procyclical enough.

Proposition 6 *The a-HANK model can deliver both amplification and solve the FG puzzle (resolving the Catch-22) if and only if:*

$$\begin{aligned} &\textit{(1).}\ \chi &>& \ 1 \ \textit{(countercyclical inequality) and} \\ &\textit{(2).}\ \eta &<& \ 1-\delta \leq 0 \ \textit{(procyclical enough risk).} \end{aligned}$$

[21] In the Appendix, I also consider a different setup whereby the probability (to be constrained next period) depends on current demand Y_t; this delivers contemporaneous amplification (multipliers) as the within-period AD elasticity to r depends of cycliclaity of risk. Notice that I assume throughout that the probability h also depends on Y in a compensating way, such that λ does not depend on the cycle.

[22] Acharya and Dogra also extend their pseudo-RANK, combining it with TANK by adding hand-to-mouth agents in a way that is *entirely orthogonal* to uninsurable risk. Something observationally equivalent can be recovered in my framework with a low *level* of idiosyncratic risk ($1 - s$ close to 1 so that $\delta \to 1$ even though $\chi > 1$, the TANK limit) but arbitrary cyclicality η.

The second condition amounts to requiring that the procyclicality of risk induced through the separate $s(Y)$ channel dominate the countercyclicality that is implied by $\chi > 1$, which instead requires enough inequality in steady-state *in levels*, i.e. a high Γ:

$$-s_Y Y \left(1 - \Gamma^{-1/\sigma}\right) \sigma > (1 - s) \frac{\chi - 1}{1 - \lambda} > 0.$$

This is a manifestation of the dependence of the cyclicality of risk channel upon the degree of heterogeneity or inequality; recall that without inequality in levels the cyclicality of risk is irrelevant (for Euler discounting/compounding), while without risk in levels ($1 - s = 0$) the cyclicality of inequality is instead irrelevant. When the two channels coexist, go in opposite directions, and have the right relative strengths (Proposition 6), the Catch-22 is resolved.

However, when the conditions of Proposition 6 are not met, cyclical risk (above and beyond the cyclicality of inequality) can aggravate the Catch-22; indeed, in case both risk and inequality are countercyclical (the more empirically plausible scenario) the model delivers amplification and aggravates the puzzles even further, while the determinacy conditions under a Taylor rule become even more stringent (it is "as if" δ was higher). The remainder of the paper studies the analytical-HANK version concentrating on the cyclical-inequality channel, bearing in mind that adding the other source of cyclical risk amounts, insofar as reduced-form dynamics are concerned, to changing δ.

3.5 The Virtues of a Wicksellian Rule in a-HANK

Can a HANK model calibrated to deliver "amplification" without also amplifying the NK puzzles—*even when* both inequality and risk are countercyclical? And what *can* the central bank do in such an economy to ensure equilibrium determinacy, given that the Taylor rule is usually a very bad prescription, according to our HANK-Taylor Principle in Proposition 3 (see the right panel of Figure 2)—when for a standard calibration, a central bank following the Taylor rule would need to change nominal rates by, say, 5 percent if inflation changed by one percent?

These questions are interrelated and one answer to both is the *"Wicksellian" policy rule* proposed by Woodford (2003) and Giannoni (2014), of price level targeting:

$$i_t = \rho_t + \phi_p p_t + i_t^* \text{ with } \phi_p > 0 \tag{24}$$

which the above authors originally demonstrated yields determinacy in RANK. This rule is especially powerful in HANK, as emphasized in the following Proposition.

Proposition 7 ***Wicksellian rule in HANK***: *In the HANK model with amplification and puzzles $\nu_0 > 1$, the Wicksellian rule (24) satisfying $\phi_p > 0$ leads to a locally unique rational-expectations equilibrium (determinacy) and eliminates the FG puzzle.*

The proof is simple and instructive and it is outlined in Appendix C.4 under static PC (3) and in Appendix C.5 with NKPC (2). The essence of it is that under the Wicksellian rule (24) the HANK model reduces, instead of one difference equation such as (19), to a *second-order* equation obtained (for the static PC case) by replacing (3) and the policy rule (24) in the aggregate Euler-IS (23), and then substituting in it the static PC rewritten in terms of the price level $p_t - p_{t-1} = \kappa c_t$. This delivers:

$$E_t p_{t+1} - \left[1 + \nu_0^{-1}\left(1 + \sigma \frac{1-\lambda}{1-\lambda\chi}\phi_p \kappa\right)\right] p_t + \nu_0^{-1} p_{t-1} = \sigma \frac{1-\lambda}{1-\lambda\chi}\kappa \nu_0^{-1} i_t^*. \tag{25}$$

It is easy to show by standard techniques that equation (25) has a unique solution if and only if $\phi_p > 0$; and, calculating that solution, to show that the effect of an interest rate cut now decreases with the horizon at which it takes place, i.e. $\partial c_t / \partial\left(-i_{t+T}^*\right)$ is decreasing in T: *the FG puzzle disappears*. These two elements prove the Proposition.

The intuition is that, as we discussed above, the source of these puzzles is indeterminacy under a peg; and a Wicksellian rule provides determinacy under a "quasi-peg". What is needed is "some" (no matter how small) response to the price level—which nevertheless anchors long-run expectations because agents know that under such a rule, bygones are *not* bygones and some inflation will a fortiori imply deflation in the future. This finding is particularly important HANK, for even under conditions whereby heterogeneity (HA) *aggravates* (instead of curing) NK puzzles, adopting this rule still works and restores standard logic, thus resolving the "Catch-22".[23]

A different option to obtain determinacy and potentially solve the puzzles is to resort to *fiscalist* equilibria—the same way one does in the standard model, by introducing (in the version with positive debt) 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. In an incomplete-markets economy, yet a further option to determine the price level exists, discovered by Hagedorn (2017, 2018). The self-insurance equation defines a demand for nominal debt. If the government supplies that nominal debt according to a rule that responds to the price level, the latter is determined without resorting to an interest-rate rule. That is similar to the Wicksellian rule I propose, which specifies $i = f(p)$ directly; it instead combines demand for bonds $B^d(i)$ with a supply rule $B^s(p)$ and requires a specific form of fiscal-monetary coordination: the government sets nominal debt responding to the price level, sets nominal taxes so as to balance the budget intertemporally (thus making policy passive-Ricardian and ruling out fiscal theory), while the central bank residually sets the nominal interest rate that clears the liquid-bond market; one could in principle adopt such a policy in (the version with positive long-run B of) my model too.

[23] An interesting and hitherto unnoticed to the best of my nowledge corrollary is that in RANK too, the FG puzzle disappears under a Wicksellian rule (recall that RANK is nested here for $\lambda = 0$ or $\chi = 1$, the Campbell-Mankiw benchmark).

3.6 Optimal Policy in a-HANK

This benchmark a-HANK economy is also easily amenable to studying optimal monetary policy analytically. Take second-order approximation of household welfare around a flexible-price equilibrium that is efficient: that is, perfect-insurance; this is due to profits being zero to first order in this efficient equilibrium, as in the TANK analysis in Bilbiie (2008). In fact, a second-order approximation to aggregate household welfare delivers Proposition 8.

Proposition 8 *Solving the welfare maximization problem is equivalent to solving:*

$$\min_{\{c_t,\pi_t\}} \frac{1}{2} E_0 \sum_{t=0}^{\infty} \beta^t \left\{ \alpha y_t^2 + \pi_t^2 \right\}, \tag{26}$$

where the optimal weight on output stabilization is:

$$\alpha \equiv \frac{\sigma^{-1}+\varphi}{\psi} \left[1 + \varphi^{-1}\sigma^{-1} \frac{\lambda}{1-\lambda} (\chi - 1)^2 \right].$$

Notice that the objective function is isomorphic to RANK and to TANK (Bilbiie, 2008 Proposition 5); in other words, the introduction of idiosyncratic risk does not introduce an additional stabilization motive for monetary policy–insofar as the economy is approximated around an efficient steady state, and the flexible-price equilibrium is *efficient*, including in its pefect-insurance dimension. Challe (2018) finds a similar isomorphism in an economy in which the cyclical-inequality channel emphasized here is absent but idiosyncratic risk is endogenous (and thus cyclical) because of search and matching, as in Ravn and Sterk (2018), under the further special assumption of full worker reallocation. Other studies have found additional stabilization motives using extensions of TANK or different reduced-heterogeneity models, e.g. Nistico (2016) and Curdia and Woodford (2016) for a financial-stability motive, and Bilbiie and Ragot (2016) for a liquidity-insurance motive arising because of approximating the economy around an imperfect-insurance equilibrium which gives rise to a linear term in the quadratic approximation.

In my benchmark a-HANK economy, the only modification relative to RANK is that the relative weight on output α increases with λ: HANK modifies the trade-off faced by the monetary authority. The intuition is simple and based, as in TANK, on the key role of profits—which are eroded by inflation volatility. With higher λ, less agents receives profit income and the welfare-weight on inflation falls (in the limit $\lambda \to 1$ inflation becomes completely irrelevant for welfare as there is no rationale for stabilizing profit income).

To solve for optimal policy, the constraints of the central bank are the private equilibrium conditions summarized by the aggregate IS curve ((10) or (23)) and a Phillips curve. Notice that, as in RANK, the former is not a constraint: it merely allows solving for the path of the policy instrument i_t once that optimal allocation (y_t, π_t) has been found. For the Phillips curve, we consider a more general version of the form:

$$\pi_t = \beta_f E_t \pi_{t+1} + \kappa y_t + u_t, \tag{27}$$

where u_t captures cost-push shocks that generate a meaningful trade-off between stabilizing inflation and real activity; in this framework, they can arise through any mechanism that makes the flexible-price inefficient—including e.g. inefficient fiscal redistribution—but we leave it unspecified for generality.

As in RANK, we can now study optimal policy under discretion (Markov-perfect equilibrium) or under commitment (time-inconsistent Ramsey equilibrium). The former is obtained by solving (26) by assuming that the central bank lacks a commitment technology and therefore treats all expectations parametrically (without recognizing the effect of its actions on expectations); therefore, this amounts to re-optimizing every period subject to the (27) constraint whereby all expectations at time t (when the policy is chosen) are treated parametrically. Since the problem is isomorphic to that in RANK (Clarida et al (1999), Woodford (2003)), we can go directly to the solution, in the form of a **targeting rule under discretion**:

$$\pi_t = -\frac{\alpha}{\kappa} y_t \tag{28}$$

The rule requires that the bank engineer an increase (decrease) in aggregate demand for a given increase (decrease) in inflation. Assuming an AR(1) process for the cost-push shock $E_t u_{t+1} = \mu_u u_t$, we obtain the equilibrium:

$$\pi_t^d = \alpha \Upsilon u_t; \quad y_t^d = -\kappa \Upsilon u_t, \tag{29}$$

where $\Upsilon \equiv \left[\kappa^2 + \alpha\left(1 - \beta_f \mu_u\right)\right]^{-1}$.

Optimal policy under discretion implies that both output and inflation deviate from their target values in response to cost-push shocks that thus creates a trade-off between inflation and output stabilization. Since α is increasing in λ as discussed above, optimal policy in a-HANK results in *greater inflation volatility and lower output volatility* than in RANK.

One instrument rule that implements this equilibrium is found by using the aggregate IS curve (10):

$$i_t = \rho_t + \left(1 + \frac{\kappa}{\alpha} \frac{\mu_u^{-1} - \delta}{\sigma \frac{1-\lambda}{1-\lambda\chi}}\right) E_t \pi_{t+1}$$

The implicit instrument rule is *passive* when the shock is persistent enough and there is enough compounding $\delta > \mu_u^{-1}$: the optimal response in HANK may require a cut of real rates when in RANK the optimal response requires an increase in real rates.[24]

[24] Notice that many other rules (indeed, an infinity) that implement the discretionary optimum would lead to indeterminacy: it suffices to replace the equilibrium value of $E_t \pi_{t+1}^d$ as a function of cost-push shocks to notice that the path of interest rates becomes exogenous—a peg which leads to equilibrium indeterminacy unless the conditions

By similar arguments as in RANK, (Ramsey-) optimal policy from a timeless perspective requires commitment to a different targeting rule, namely:

$$\pi_t = -\frac{\alpha}{\kappa}\left(y_t - y_{t-1}\right). \tag{30}$$

It is easy to show that commitment to this targeting rule delivers equilibrium determinacy regardless of the degree of heterogeneity; this is not surprising, since this is similar in spirit to the Wicksellian rule of price-level targeting. Furthermore, Bilbiie (2008, Proposition 6) showed this result in TANK.

The results on optimal policy in the benchmark a-HANK economy thus echo those found in TANK. This equivalence no longer applies to the analysis of optimal policy, which I pursue below using this analytical apparatus.

4 Liquidity Traps in a-HANK

Following the seminal paper of Eggertsson and Woodford (2003), I analyze liquidity traps assuming that the fundamental shock ρ_t follows a Markov chain with two states. The first is the good, "intended" steady state denoted by I, 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 $\rho_t = \rho_L < 0$ (which could be justified in a model with credit frictions by an increase in spreads as in Curdia and Woodford (2009)). The duration of the transitory state is a random variable T with expected value $E(T) = (1-z)^{-1}$.

Given this Markov chain structure and the Taylor rule subject to a zero lower bound:

$$i_t = \max\left(0, \rho_t + i_t^* + \phi\pi_t\right), \tag{31}$$

with $\phi > 1$, the LT equilibrium is found by conjecturing that it is time-invariant, denoting it by (c_L, π_L) which prevails for any time t between 0 and T (thereafter, it is straightforward to show that the model returns to the steady state). Equation (3) implies $\pi_L = \kappa c_L$ and, with a binding lower bound $i_L = 0$, the aggregate IS implies:

$$c_L = \frac{1}{1 - z\nu_0}\sigma\frac{1-\lambda}{1-\lambda\chi}\rho_L. \tag{32}$$

As clear from (32), this is an equilibrium with a recession and deflation ($c_L < 0$; $\pi_L < 0$) if

discussed above apply.

and only if:

$$z < \nu_0^{-1}. \tag{33}$$

Finally, verifying that ZLB binds indeed: $\rho_L + \phi\pi_L < 0$ implies $1 + \phi\frac{\nu_0 - 1}{1 - z\nu_0} > 0$, which always holds as long as (33) holds. Condition (33) rules out the occurrence of confidence-driven liquidity traps (Benhabib et al, 2002; Mertens and Ravn, 2014; ; see Bilbiie, 2018 for further discussion and analysis of these equilibria in connection with neo-Fisherian policies). In the "discounting" case with $\nu_0 < 1$, the restriction is *a fortiori* satisfied no matter how pessimistic agents are (how high the sunspot persistence), since z is a probability $z < 1 < \nu_0^{-1}$.[25]

The mechanism by which LT-recessions occur is similar to the one familiar from the RANK model; but in the simple HANK model, their magnitude (and whether they are larger or smaller than in RANK) depends on the key parameters λ, χ, $1-s$, through both the within-period demand elasticity to interest rates ($\sigma\frac{1-\lambda}{1-\lambda\chi}$) and through the AD effect of news under a peg parameter ν_0. Through the amplification mechanisms emphasized above, there can be a deep recession in a-HANK *even for fixed prices* $\kappa = 0$.

Amplification, understood as an LT recession deeper than in RANK, obtains if and only if inequality is countercyclical, $\chi > 1$ (the same amplification logic emphasized above for monetary or fiscal policies). Generally, this occurs through three forces. First, the within-the-period, TANK amplification of changes in interest rates through a New Keynesian Cross mechanism ($\frac{1-\lambda}{1-\lambda\chi}$).[26] Second, the HANK, intertemporal extension of that: the *self-insurance* channel yielding *compounding* in the aggregate Euler equation ($\delta > 1$) which amplifies the effect of "news". 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 saving. Since future consumption in states where the constraint binds over-reacts to bad "aggregate news" (countercyclical inequality), households internalize this by attempting to self-insure *more*. And since saving needs to be zero in equilibrium, households consume less and income falls to deliver this, thus magnifying the recession even further. Third, 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 proportionally to $\frac{1-\lambda}{1-\lambda\chi}$. Evidently, these conditions require the opposite of the no-puzzles condition;[27] in other words, in the region where the puzzles are resolved *all* these

[25] Notice, nevertheless, that a sunspot equilibrium may *always* be constructed, e.g. insofar as prices are flexible enough (or whatever makes $\nu_0 > 1$). In fact, they can always be constructed as long as the ZLB equilibrium is a steady state.

[26] This 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).

[27] The model has implications fo the **paradox of flexibility** (Eggertsson and Krugman, 2012), that an increase

channels imply, instead of amplification, *dampening*.[28]

4.1 FG Puzzle and Power in a Liquidity Trap

Forward guidance has been discussed in particular in the context of LTs, as a policy tool that remains available when the standard ones are not, and as a characteristic of optimal policy; see Eggertsson and Woodford (2003) for the original analysis, and Bilbiie (2016) for a more recent treatment and an up-to-date discussion of the literature.

To discuss the FG puzzle (and later optimal policy) in the context of LTs, I follow the latter paper and model FG stochastically through a Markov chain, as a state of the world with a probability distribution, as follows. 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:

$$E_t c_{t+1} = z c_L + (1-z)\, q c_F \tag{34}$$

and similarly for inflation. Evaluating the aggregate Euler-IS (10) and Phillips ($\pi_t = \kappa c_t$) 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:

$$\begin{aligned} c_F &= \frac{1}{1-q\nu_0}\sigma\frac{1-\lambda}{1-\lambda\chi}\rho; \\ c_L &= \frac{1-z}{1-z\nu_0}\frac{q\nu_0}{1-q\nu_0}\sigma\frac{1-\lambda}{1-\lambda\chi}\rho + \frac{1}{1-z\nu_0}\sigma\frac{1-\lambda}{1-\lambda\chi}\rho_L, \end{aligned} \tag{35}$$

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 (the solution with NKPC (2) is slightly more involved and included in Appendix C.3).

---

in price flexibility κ makes the ZLB recession worse; in the HANK model: $\partial\left(\frac{\partial c_L}{\partial \rho_L}\right)/\partial\kappa = z\left(\frac{1}{1-z\nu_0}\sigma\frac{1-\lambda}{1-\lambda\chi}\right)^2 > 0$. The paradox is merely *mitigated* (the derivative above decreases) with λ iff $\chi < 1$ (the proof follows immediately as $\sigma\frac{1-\lambda}{1-\lambda\chi}$ and δ, and hence ν_0, are decreasing with λ iff $\chi < 1$). Conversely, the paradox is aggravated with λ if $\chi > 1$.

[28] Turning 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).

It is immediately apparent that the future expansion c_F is increasing in the degree of FG q regardless of the model. In the CI case ($\chi > 1$), the future expansion is also *increasing* with the H share λ, and with risk $1 - s$; whereas with PI the opposite holds.

Figure 4 illustrates these findings: Distinguishing between $\chi < 1$ (left) and $\chi > 1$ (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. Other than the parameter values used for Figure 2, it uses $z = 0.8$ and a spread shock of 4 percent per annum ($\rho_L = -0.01$). This delivers a recession of 5 percent and annualized inflation of 1 percent in RANK without FG ($q = 0$). The domain is such that $q < \nu_0^{-1}$. The RANK model is with green solid lines, the TANK limit ($s = h = 1$) with red dashed lines, and the other extreme, iid limit of the HANK model ($1 - s = h = \lambda$) with blue dots.

The pictures illustrate dampening and amplification (respectively) in a LT: at given q, low future rates have a lower effect (on both c_F and c_L) in TANK, and an even lower one in HANK, with PI. The last point illustrates the complementarity: the dampening is magnified when moving towards higher risk $1 - s$, with the fastest discounting in the limit when $1 - s = h = \lambda$ (blue dots). Whereas with CI (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 HANK model is substantial: the recession is three 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—indeed, with $\beta = 0.99$ in (2), the recession is *10* (ten) times larger.

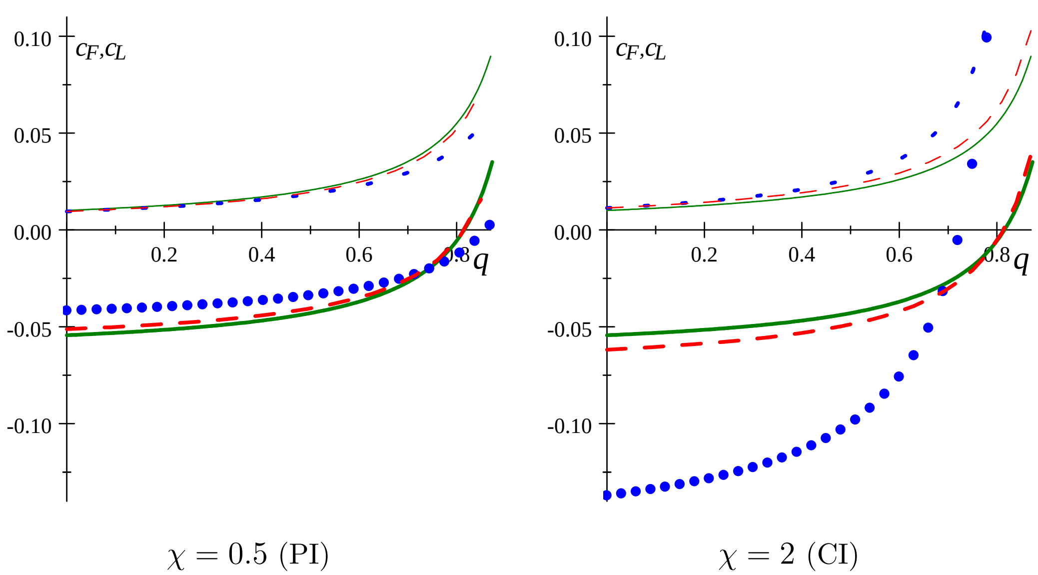


Fig. 4: c_L (thick) and c_F (thin) in RANK (green solid), TANK (red dashed) and iid-HANK (blue dots)

We can now define *FG power*, denoted by $\mathcal{P}_{FG}$, formally as 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 4, this is much larger in HANK with countercyclical inequality. The properties of amplification and dampening of FG power follow the same logic as those applying to any demand shock. Since $\mathcal{P}_{FG}$ is increasing with ν (and hence with both δ and $\sigma\frac{1-\lambda}{1-\lambda\chi}$), in the CI case it increases with idiosyncratic risk $1-s$ and with the share of hand-to-mouth λ (while it decreases with PI). Furthermore, the complementarity between self-insurance and hand-to-mouth also applies to FG power.

The *FG puzzle* is then in this context that $\mathcal{P}_{FG}$ increases with the persistence (and thus expected duration) of the trap z:

$$\frac{d\mathcal{P}_{FG}}{dz} \geq 0.$$

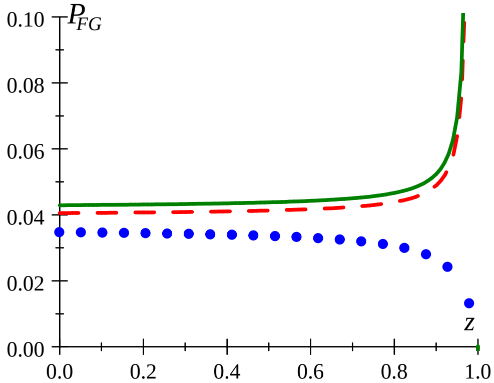
When does the model resolve the FG puzzle in a LT? The general insight found in Proposition 4 reassuringly applies in a LT, as emphasized in the following.

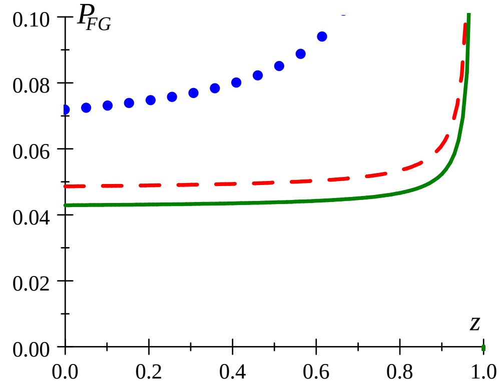
Corollary 2 *The analytical HANK model* ***solves*** *the FG puzzle in a LT equilibrium* ($\frac{d\mathcal{P}_{FG}}{dz} < 0$) *if and only if:*

$$\nu_0 < 1.$$

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.

To further illustrate how the FG puzzle operates and how the complementarity between cyclical inequality and idiosyncratic (albeit acyclical) risk helps eliminate it, consider Figure 5; it plots $\mathcal{P}_{FG}$ 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 HANK. This shows most clearly that it is the interaction of procyclical inequality (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 PI 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 idiosyncratic risk (the self-insurance channel by itself) added to the CI, "amplification" case magnifies power even further, thus *aggravating* the puzzle (blue dots in the right panel for the iid HANK model).





$\chi < 1$ (PI) $\chi > 1$ (CI)

Fig. 5: FG power in RANK (green solid), TANK (red dashed) and iid-HANK (blue dots)

Evidently, the puzzle is aggravated at higher values of ν_0 ($\frac{d\mathcal{P}_{FG}}{dz}$ 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 PI case; but worsens with idiosyncratic risk $1-s$ and with λ in the CI case.

4.2 Optimal Policy in a Liquidity Trap in a-HANK

Optimal monetary policy in a liquidity trap is easy to compute with this setup for FG, because equilibria are a smooth function of q: we can find the optimal FG duration by maximizing lifetime welfare with respect to q. This is shown in Bilbiie (2016) for RANK (including that it is very close to the full Ramsey-optimal monetary policy taking the ZLB as a constraint calculated by Eggertsson and Woodford (2003), Jung Teranishi and Watanabe (2005) and several others since).

We thus look for q that maximizes the aggregate welfare function that can be represented as a quadratic loss function and, given the Markov chain structure, it is of the form:[29]

$$W = \frac{1}{1-\beta z}\frac{1}{2}\left[c_L^2 + \omega(q)\, c_F^2\right],$$

where $\omega(q)$ is the appropriate discount factor for the FG state.[30] The central bank chooses FG duration (persistence probability q) by solving the optimization problem $\min_q W$ taking as constraints the equilibrium values c_F and c_L given in (35) above. The first-order condition of this problem is:

[29] Since the equilibrium solution is time-invariant in each of the three states, the per-period loss function is, for any state $j = \{L, F, S\}$: $\pi_j^2 + \chi c_j^2 = \left(\chi + \kappa^2\right) c_j^2$. Recall that in state S the economy is back to steady state, so the loss there is zero.

[30] The optimal weight is $\omega(q) = \frac{1-\beta z+\beta(1-z)q}{1-\beta q}$ and counts for the times the process spends in state F when starting from F (given by $(1-\beta q)^{-1}$); as well as for all the times spent in time F when starting from L, before being absorbed into S (given by $\beta(1-p)\, q / \left((1-\beta p)(1-\beta q)\right)$. $\omega(q)$ is increasing in q, which is intuitive: the longer the economy spends in the F state, the larger the total welfare cost of consumption variability in that state. See Bilbiie (2016) for the details, including the second-order sufficient conditions for the RANK model (that apply here too).

$$c_L \frac{dc_L}{dq} + \omega(q) c_F \frac{dc_F}{dq} + \frac{1}{2}\frac{d\omega(q)}{dq} c_F^2 = 0 \quad (36)$$

and has a clear intuitive interpretation.

The first term is the welfare *benefit* of more forward guidance, through remedying the LT-caused recession and hence minimizing consumption volatility in the trap. This is proportional to the *level* of consumption in the trap: the larger the initial recession, the higher the marginal utility of an extra unit of consumption, and the larger the welfare scope of any policy that can deliver it—such as FG. The last two terms are the *total cost* of forward guidance: the former is the direct cost, a future consumption boom being associated with inefficient volatility; the latter is the discounting effect discussed above: the longer the time spent under FG, the larger the cost (which is proportional to consumption volatility in the F state).

The basic **analytical insights** can be obtained by focusing first on a simpler case, assuming in addition that the central bank attaches equal weights to future and present: $\omega(q) = 1, \omega'(q) = 0$. This provides an *upper bound* on optimal FG because it ignores the second-order discounting costs.[31] The optimal duration can then be solved in closed-form: (36) becomes $c_L \frac{dc_L}{dq} = -c_F \frac{dc_F}{dq}$, which replacing c_F and c_L from (35) delivers the following.

Proposition 9 *The optimal FG duration is $q = 0$ if $\Delta_L < \frac{(1-z\nu_0)^2}{1-z}$ and $q^* > 0$ otherwise, with:*

$$q^* = \frac{1}{\nu_0} \frac{\Delta_L - \frac{(1-z\nu_0)^2}{1-z}}{1 - z + \Delta_L},$$

where $\Delta_L \equiv -\rho_L/\rho > 0$ is the financial disruption causing the ZLB.

It is optimal to refrain from FG altogether ($q^* = 0$) when there is not enough "news-amplification": when ν_0 is *smaller* than a certain threshold $\tilde{\nu}$.[32] Thus in the amplification case ($\chi > 1$ and $\nu_0 > 1$) the region of λ for which FG is optimal will be ceteris paribus smaller than in the "dampening ($\chi < 1$ and ν_0) case. Moreover, since in the former case ν_0 is increasing both with λ and with $1-s$, an increase in either restricts the case for optimal FG.[33] The reason is that more amplification also brings about a higher welfare cost of FG. Conversely of course, in the latter (dampening) case the opposite is true: an increase in either λ or $1 - s$ pushes up the threshold and enlarges the region for which FG is optimal (ν_0 is decreasing in both parameters).

Optimal FG duration depends on the key heterogeneity parameters λ, χ, and $1 - s$ through

[31] See Bilbiie, 2016 for an analysis of the accuracy of this in a RANK model.

[32] Specifically, $\tilde{\nu} \equiv \left(1 - \sqrt{(1-z)\Delta_L}\right)/z$ which under the baseline calibration is 0.86.

[33] The derivatives are $\frac{d\nu_0}{d(1-s)} = \frac{d\delta}{d(1-s)} = \frac{\chi-1}{1-\lambda\chi}$ and $\frac{d\nu_0}{d\lambda} = \frac{d\delta}{d\lambda} + \frac{d\eta}{d\lambda}\kappa = (\chi - 1)\frac{\chi(1-s)+\kappa\sigma}{(1-\lambda\chi)^2}$.

the key composite parameter elasticity-to-news ν_0:

$$\frac{dq^*}{d\nu_0} = \frac{1}{\nu_0^2}\left(\frac{1-(z\nu_0)^2}{1-z} - \Delta_L\right).$$

When the disruption causing the liquidity trap is lower than a certain threshold $\Delta_L < (1-z)^{-1}$ (the more empirically plausible case),[34] q^* is increasing in ν_0 if $\nu_0 < \bar{\nu} \equiv \sqrt{1-\Delta_L(1-z)}/z$ and decreasing otherwise. Notice that this threshold is larger than the threshold needed for FG to be optimal at all ($q^* > 0$) derived above: $\bar{\nu} > \tilde{\nu}$. We have $dq^*/d\nu_0 > 0$ when $\tilde{\nu} < \nu_0 < \bar{\nu}$ and $dq^*/d\nu_0 < 0$ when $\tilde{\nu} < \bar{\nu} < \nu$. It is useful to again distinguish the two cases depending on χ.

In the dampening case ($\chi < 1$) ν_0 is decreasing in λ and $1-s$; if we start with $\nu_0 > \bar{\nu}$, optimal FG duration first increases, then decreases as ν_0 crosses the threshold. Whereas if we start below the threshold, optimal FG duration decreases uniformly (this is the case shown in the Figure below). The effect is mitigated by idiosyncratic risk which, because it reduces both the power of FG and the scope for it (the LT recession is smaller) implies uniformly lower optimal duration.

With amplification ($\chi > 1$), ν_0 is increasing in both λ and $1-s$; therefore, if we start below the threshold $\bar{\nu}$, optimal FG first increases up to a maximum level (reached at the threshold) and then decreases abruptly. Furthermore, it increases faster and reaches its maximum sooner when there is idiosyncratic risk, because of the complementarity: amplification itself is in that case magnified—by the same token, the welfare cost of FG suffers from the same amplification, so the point where FG ceases to be optimal is reached sooner than without risk $s = 1$.

Figure 6 illustrates these findings by plotting the optimal FG duration, the solution of (36)), as a function of λ, under our baseline parameter values, distinguishing again the dampening ($\chi < 1$, left) and amplification ($\chi > 1$, right) cases. In the dampening case, the degree of optimal FG is *decreasing* with the share of H; the more so, the higher idiosyncratic risk; this result holds generally and was shown analytically above in the simplest version (notice that the Figure plots the general case with $\omega(q)$ whereby a closed-form solution is no longer possible.). The intuition is that all forces work in the same direction: the recession is lower to start with (which gives *less* scope for using FG) and the power of FG is monotonically decreasing with λ: because the elasticity to interest rates is decreasing, in TANK, and in addition because of discounting in a-HANK.

The amplification case is, in view of our previous results, more surprising: the optimal degree of FG is almost invariant to λ in TANK (albeit initially mildly increasing) because there are two counterbalancing forces. On the one hand, the benefit component is higher: the recession is larger (c_L more negative, and a wider output gap gives more welfare reason to use FG), and the power of FG $\frac{dc_L}{dq}$ is higher. But on the other hand, the cost of FG is also increasing (the last two terms in

[34] If instead $\Delta_L > (1-z)^{-1}$, q^* is uniformly *decreasing* in ν_0: that is, it is decreasing in χ , λ, and $1-s$ in the "amplification" case $\chi > 1$. The reason is that the contractionary effect coming from the steeper recession dominates the expansionary effect of increased FG effectiveness; the opposite is of course true with $\chi < 1$: q^* is increasing in λ and $1-s$.

(36)). At some threshold λ level, the cost of FG is no longer worth bearing: the implied volatility during the FG state is so high that the optimal degree of FG drops rapidly towards zero. In a-HANK, these affects are further amplified by idiosyncratic risk: the higher share of H makes the recession larger and accelerates the increase in FG power, making optimal FG initially increasing; but the same amplification also holds for the welfare cost of future volatility, which kicks in earlier (at a lower λ) and makes optimal FG drop abruptly towards zero. This sharp increase in the welfare cost occurs precisely when FG power is large: the "dark side" of FG power.

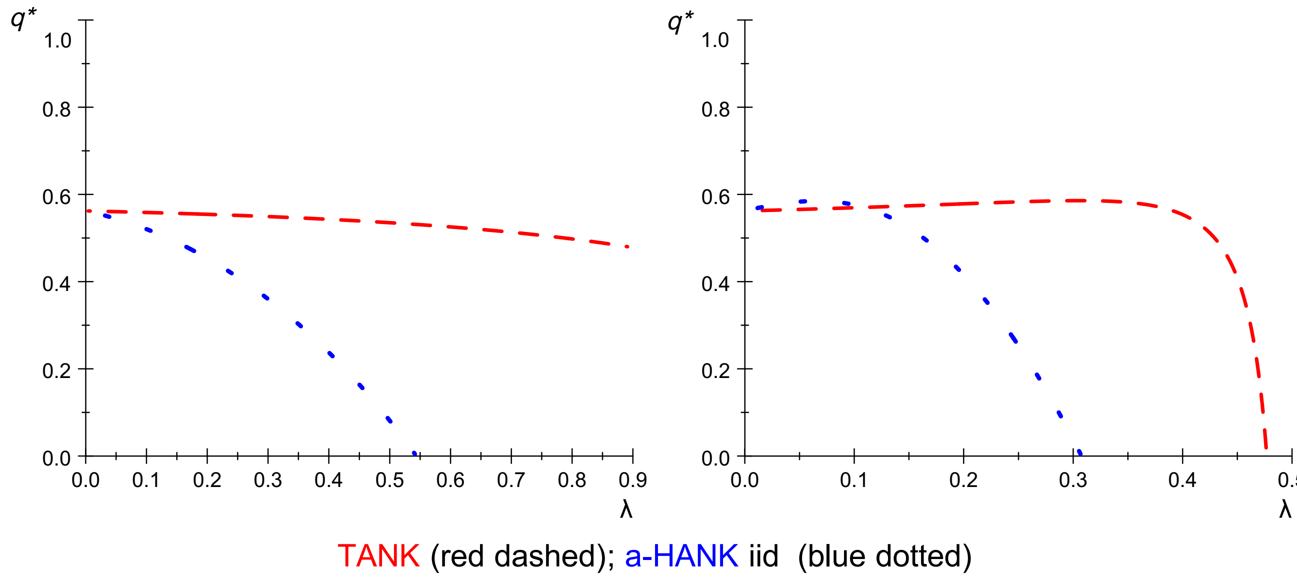


Figure 6 Optimal FG persistence as a function of λ for $\chi < 1$ (left) and $\chi > 1$ (right)

In both models, it becomes optimal to do no FG at all beyond a certain threshold λ. The underlying reason is, however, very different. With dampening it is because a higher λ implies *low* FG power, and because it also implies a weaker welfare-scope for using FG (the recession is lower). With amplification, it happens because a high λ implies a *large* AD elasticity to interest rates, and hence a high FG cost; even though FG too becomes more powerful, this effect is dwarfed by the increase in cost.

5 Conclusions

I propose an analytical HANK (a-HANK) model that captures some key channels of richer, quantitative HANK models and use it for a full-NK analysis (with pencil and paper) of the main themes of the NK literature of the past decades: determinacy properties of interest rate rules, resolving the forward guidance puzzle, amplification and multipliers, liquidity traps, and optimal monetary policy.

The key channel is that of cyclical inequality: whether the income of constrained hand-to-mouth agents comoves more (countercyclical inequality) or less (procyclical) with aggregate income. This channel is already operating and is the main focus of earlier TANK (Bilbiie (2008)) but interacts with idiosyncratic uncertainty and self-insurance in the a-HANK, as it does in quantitative HANK models. Thus, procyclical inequality delivers discounting in the aggregate Euler equation which makes the Taylor principle not necessary for equilibrium determinacy. If this channel is strong enough to overturn the compounding of news that generates the NK puzzles in RANK, rules out the FG puzzle by generating equilibrium discounting of news shocks.

Conversely, however, countercyclical inequality makes the Taylor principle insufficient for determinacy (the response necessary to ensure determinacy can become very large indeed) and aggravates the FG puzzle. That is because it generates compounding in the Euler equation and thus even more compounding of news shocks than in RANK. This is a Cattch-22, for countercyclicality is precisely the condition needed for HANK models to deliver "amplification", or multipliers—which is what the majority of quantitative studies have used them for, exploiting a New Keynesian cross that is inherent in these models. The news seems even worse: with countercyclicality (of either inequality, or risk) the NK puzzles are in fact aggravated, and the Taylor Principle is vastly insufficient for determinacy.

The paper provides two resolutions (amplification without puzzles). First, keeping policy fixed, if a distinct HANK channel delivers enough discounting without affecting amplification it can resolve this tension: I illustrate how a notion of *cyclical risk* previously emphasized by others (but formalized here in a novel way) can deliver that independently of the cyclicality of inequality. Second, at this next level, there is a further uncomfortable observation: when both inequality and risk are countercyclical, the puzzles are aggravated further and the requirement for a central bank to ensure determinacy with a Taylor rule is significantly more stringent than merely being "active". Yet I show that if the central bank adopts a Wicksellian rule of price-level targeting (shown by Woodford (2003) and Giannoni (2014) to deliver determinacy in RANK), this tension disappears: The HANK model is determinate and suffers from no puzzles, even in the "amplification" region with countercyclical inequality and risk.[35]

In the benchmark a-HANK that isolates the role of cyclical inequality, optimal monetary policy requires stabilizing inflation and a measure of real activity (around an efficient, perfect-insurance equilibrium) and is thus similar to RANK and TANK—but implies tolerating more inflation volatility in response to cost-push shocks.

In a liquidity-trap driven by a fundamental increase in aggregate desired savings, the severity of LT recessions and the power of FG are mitigated (and the FG puzzle resolved) with procyclical inequality but magnified with countercyclical (and the puzzle aggravated). Yet optimal policy implies that even with countercyclical inequality the same amplification increasing FG power also magnifies its welfare cost , thus containing its optimal duration.

My results can guide empirical work as to what are the key parameters that empirical evidence

[35] Other possible solutions consist of extending the model by adding either a "discounting" feature that independently solves the puzzles to an "amplifying" HANK: e.g. changing the information-expectation structure or otherwise, as reviewed in the Introduction.

should shed light on in the realm of models with heterogeneity. In particular, given that existing evidence suggests that idiosyncratic risk is likely countercyclical (Storesletten, Telmer, and Yaron, (2004); Guvenen, Ozkan, and Song (2014)), the paramount parameter pertains to the cyclicality of inequality χ. Early empirical evidence (Heathcote, Perri, and Violante (2010)) seems to suggest that inequality in the US is likely countercyclical, too; more recent evidence provided in Patterson (2019), estimating something closer to this paper's χ parameter, is consistent with that view. This points on the one hand, for the quantitative macroeconomist, to the urgency of estimating a model where all these channels can be identified and disentangled, of which this paper's is a simple example. And on the other hand, for the policymaker, to the importance for finding optimal policy prescriptions (such as the ones derived here) that can anchor expectations even when the heterogeneity channels deliver amplification and make the economy very unstable.

References

[1] Acharya, S. and K. Dogra, 2018 "Understanding HANK: Insights from a PRANK," Mimeo

[2] Aiyagari, R., 1994, "Uninsured Idiosyncratic Risk and Aggregate Saving," Quarterly Journal of Economics, 109 (1994), 659–684.

[3] Andrade, P., G. Gaballo, E. Mengus, and B. Mojon, 2016, "Forward Guidance with Heterogeneous Beliefs," American Economic Journal: Macroeconomics

[4] Angeletos, G.M. and C. Lian, 2017, Forward Guidance without Common Knowledge, American Economic Review

[5] Ascari, G., A. Colciago and L. Rossi, 2017, "Limited Asset Market Participation, Sticky Wages, and Monetary Policy", Economic Inquiry

[6] Auclert, A. 2016, "Monetary Policy and the Redistribution Channel," American Economic Review

[7] Auclert, A. and M. Rognlie 2016, "Inequality and Aggregate Demand," Mimeo

[8] Auclert, A., M. Rognlie, and L. Straub 2018, "The Intertemporal Keynesian Cross," Mimeo

[9] Bayer, C., R. Luetticke, L. Pham-Dao, and V. Tjaden, 2018, "Precautionary Savings, Illiquid Assets, and the Aggregate Consequences of Shocks to Household Income Risk," forthcoming Econometrica

[10] Bhandari, A., D. Evans, M. Golosov, and T. Sargent, 2017, "Optimal Fiscal-Monetary Policy with Redistribution", Mimeo

[11] Benhabib, J., Schmitt-Grohe, S., and Uribe, M., 2001, "The Perils of Taylor Rules," Journal of Economic Theory

[12] Benhabib, J., Schmitt-Grohe, S., and Uribe, M., 2002, "Avoiding Liquidity Traps", Journal of Political Economy

[13] Bewley, T., 1977, “The Permanent Income Hypothesis: A Theoretical Formulation,” Journal of Economics Theory, 16, 252–292.

[14] Bilbiie F. 2008. "Limited Asset Market Participation, Monetary Policy, and (Inverted) Aggregate Demand Logic". Journal of Economic Theory 140, 162-196.

[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., Monacelli T. and R. Perotti 2013, “Public Debt and Redistribution with Borrowing Constraints", The Economic Journal 2013, vol. 0, pages F64-F98, 02

[18] Bilbiie, F. 2016, "Optimal Forward Guidance", American Economic Journal: Macroeconomics

[19] Bilbiie, F. and X. Ragot, 2016, "Optimal Monetary Policy and Liquidity with Heterogeneous Households", CEPR DP

[20] Bilbiie, 2017, "The New Keynesian Cross," Journal of Monetary Economics

[21] Bilbiie, F. O. 2018, "Neo-Fisherian Policies and Liquidity Traps," CEPR DP

[22] 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).

[23] Broer, T., N. J. Hansen, P. Krusell, and E. Oberg, "The New Keynesian transmission mechanism: a heterogeneous agent perspective," Review of Economic Studies

[24] 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.

[25] Carlstrom, C., T. Fuerst, and M. Paustian, 2015. "Inflation and output in New Keynesian models with a transient interest rate peg," Journal of Monetary Economics, 76(C), 230-243

[26] Challe, E., 2019 “Uninsured Unemployment Risk and Optimal Monetary Policy in a Zero-Liquidity Economy,” Mimeo

[27] Challe, E. and X. Ragot, 2016 “Precautionary Saving Over the Business Cycle,” The Economic Journal

[28] Challe, E, J. Matheron, X. Ragot and J. Rubio-Ramirez, 2017, "“Precautionary Saving and Aggregate Demand", Quantitative Economics

[29] Christiano L., M. Eichenbaum and S. Rebelo, 2011, "When Is the Government Spending Multiplier Large?," Journal of Political Economy, 119(1), pages 78 - 121.

[30] Clarida, R., J. Galí, and M. Gertler, 1999, ’The science of monetary policy: a New Keynesian perspective’, Journal of Economic Literature, 37, 1661-1707.

[31] Cloyne, J., C. Ferreira and Surico, P., 2015 "Monetary Policy when Households have Debt: New Evidence on the Transmission Mechanism" CEPR Discussion Paper

[32] Cochrane, J., 2005, "Money as Stock," Journal of Monetary Economics

[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," NBER Macroeconomics Annual.

[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 2016, “Credit Frictions and Optimal Monetary Policy,” Journal of Monetary Economics

[38] Debortoli, D., and J. Galí, 2017, "Monetary Policy with Heterogeneous Agents: Insights from TANK models"; Mimeo

[39] Del Negro, M., M. Giannoni, and C. Patterson 2012: “The Forward Guidance Puzzle,” Mimeo

[40] Den Haan, W. J.; P. Rendahl, and M. Riegler 2016, "Unemployment (Fears) and Deflationary Spirals" Journal of the European Economic Association

[41] Diba, B. and O. Loisel, 2017, "Pegging the Interest Rate on Bank Reserves: A Resolution of New Keynesian Puzzles and Paradoxes," Mimeo

[42] Eggertsson G. and M. Woodford 2003, "The Zero Bound on Interest Rates and Optimal Monetary Policy", Brookings Papers on Economic Activity, No. 1.

[43] Eggertsson, G. 2010, "What Fiscal Policy Is Effective at Zero Interest Rates?", in: D. Acemoglu and M. Woodford, eds: NBER Macroeconomics Annual University of Chicago Press

[44] 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.

[45] Eser, F. 2009. "Monetary Policy in a Currency Union with Heterogeneous Limited Asset Markets Participation," Mimeo

[46] Fagereng, A., M. Holm and G. Natvik, 2018, "MPC Heterogeneity and Household Balance Sheets" Mimeo

[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

[49] Gabaix, X., 2016, "A Behavioral New Keynesian Model," Mimeo

[50] Galí, J. 2008, "Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework and Its Applications", Princeton University Press

[51] 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.

[52] Garcia-Schmidt, M. and M. Woodford, 2014. "Are Low Interest Rates Deflationary? A Paradox of Perfect Foresight Analysis," forthcoming, American Economic Review

[53] Gorea, D. and V. Midrigan, 2017, "Liquidity Constraints in the U.S. Housing Market" Mimeo

[54] Gornemann, N., K. Kuester, and M. Nakajima, 2016, "Doves for the Rich, Hawks for the Poor? Distributional Consequences of Monetary Policy," Mimeo

[55] Guerrieri, V. and G. Lorenzoni, 2017, "Credit Crises, Precautionary Savings, and the Liquidity Trap" Forthcoming, Quarterly Journal of Economics

[56] Guvenen, F., S. Ozkan, and J. Song, "The Nature of Countercyclical Income Risk," Journal of Political Economy, 122 (3):621-660.

[57] Hagedorn, M., 2017, "A Demand Theory of the Price Level", Mimeo

[58] Hagedorn, M., 2018, "Prices and Inflation when Government Bonds are Net Wealth," mimeo

[59] Hagedorn, M., I. Manovskii, and K. Mitman, 2017, "The Fiscal Multiplier," Mimeo

[60] Hagedorn, M., J. Luo, I. Manovskii, and K. Mitman, 2018, "Forward Guidance," Journal of Monetary Economics

[61] Hall, R. E. 2011, "The Long Slump," American Economic Review, 101, 431-469

[62] Heathcote, J., F. Perri, and G. L. Violante, 2010. "Unequal We Stand: An Empirical Analysis of Economic Inequality in the United States: 1967-2006," 13(1): 15-51, Review of Economic Dynamics.

[63] Huggett, M., 1993, "The Risk-Free Rate in Heterogeneous-Agent Incomplete-Insurance Economies," Journal of Economic Dynamics and Control, 17, 953–969.

[64] Iacoviello, M. 2005, "House Prices, Borrowing Constraints, and Monetary Policy in the Business Cycle" American Economic Review

[65] Johnson, D., J. Parker and N. Souleles, 2006. "Household Expenditure and the Income Tax Rebates of 2001." American Economic Review, 96(5): 1589-1610.

[66] Jones. C., V. Midrigan, and T. Philippon, 2018, " Household Leverage and the Recession," Mimeo

[67] Kaplan, G. and G. Violante, 2014 "A Model of the Consumption Response to Fiscal Stimulus Payments," Econometrica

[68] Kaplan, G. and G. Violante, 2018. "Microeconomic Heterogeneity and Macroeconomic Shocks," Journal of Economic Perspectives, 32(3), 167-194

[69] Kaplan, G., G. Violante, and J. Weidner, 2014 "The Wealthy Hand-to-Mouth," Brookings Papers on Economic Activity.

[70] Kaplan, G., B. Moll and G. Violante, 2018, "Monetary Policy According to HANK", American Economic Review

[71] Kaplan, G., B. Moll and G. Violante, 2017, "A Note on Unconventional Monetary Policy in HANK", Mimeo Princeton and Chicago

[72] Kiley, M. T., 2016, "Policy Paradoxes in the New Keynesian Model," Review of Economic Dynamics, 21, pp. 1-15

[73] Krusell, P., T. Mukoyama, and A. Smith. 2011. "Asset Prices in a Huggett Economy," Journal of Economic Theory, 146 (3), 812-844

[74] Leeper, E., 1991, "Equilibria Under "Active" and "Passive" Monetary and Fiscal Policies", Journal of Monetary Economics

[75] 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

[76] Lubik, T. A., and F. Schorfheide. 2004. "Testing for Indeterminacy: An Application to U.S. Monetary Policy." American Economic Review, 94 (1): 190-217

[77] Lucas R. E. Jr., 1990 "Liquidity and Interest Rates." *Journal of Economic Theory* 50, 237 - 264.

[78] Luetticke, R., 2018 "Transmission of Monetary Policy with Heterogeneity in Household Portfolios," Mimeo

[79] Mankiw, N. G. 2000. "The Savers-Spenders Theory of Fiscal Policy," American Economic Review, Vol. 90 (2) pp. 120-125.

[80] Mankiw, N. G. and R. Reis. 2002 "Sticky Information Versus Sticky Prices: A Proposal To Replace The New Keynesian Phillips Curve," Quarterly Journal of Economics, 117(4): 1295-1328

[81] McCallum, B. T. 1998 "Solutions To Linear Rational Expectations Models: A Compact Exposition," Economics Letters, 61(2), 143-147.

[82] McKay, A. and R. Reis, 2016 "The Role of Automatic Stabilizers in the U.S. Business Cycle," Econometrica, Vol. 84, 1

[83] McKay, A. and R. Reis, 2017 "Optimal Automatic Stabilizers," Mimeo

[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] Mertens K. and M. Ravn 2013. "Fiscal Policy in an Expectations Driven Liquidity Trap," Review of Economic Studies

[87] Michaillat, P. and E. Saez, 2018, "A New Keynesian Model with Wealth in the Utility Function", Mimeo

[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] Nakata, T., R. Ogaki, S. Schmidt and P. Yoo, 2018, "Attenuating the Forward Guidance Puzzle: Implications for Optimal Monetary Policy," Mimeo

[90] Nistico, S., 2015, "Optimal monetary policy and financial stability in a non-Ricardian economy." Journal of the European Economic Association

[91] Nuno, G. and C. Thomas, 2017, "Optimal Monetary Policy with Heterogeneous Agents" Mimeo Bank of Spain

[92] Oh, H. and R. Reis, 2012, "Targeted transfers and the fiscal response to the great recession." Journal of Monetary Economics, 59(S), S50-S64

[93] Patterson, C., 2019, "The Matching Multiplier and the Amplification of Recessions," Mimeo

[94] Ravn, M. and V. Sterk, 2017 "Job Uncertainty and Deep Recessions," Journal of Monetary Economics.

[95] Ravn, M. and V. Sterk, 2018, "Macroeconomic Fluctuations with HANK and SAM: an analytical approach", mimeo

[96] Schmitt-Grohe, S. and Uribe, M., 2017, "Liquidity Traps And Jobless Recoveries," American Economic Journal: Macroeconomics 9, 165-204

[97] Sargent, T. and N. Wallace, 1975, "Rational" Expectations, the Optimal Monetary Instrument, and the Optimal Money Supply Rule," Journal of Political Economy, 83(2), 241-254

[98] Sims, C. A., 1994. "A simple model for the determination of the price level and the interaction of monetary and fiscal policy." Economic Theory 4, 381–399.

[99] Storesletten, K., C. Telmer, and A. Yaron, 2004. "Cyclical Dynamics in Idiosyncratic Labor Market Risk," Journal of Political Economy, 112(3), 695-717.

[100] Surico, P. and R. Trezzi, 2016, "Consumer Spending and Fiscal Consolidation: Evidence from a Housing Tax Experiment", Mimeo

[101] Walsh, C., 2008, "Monetary Theory and Policy", MIT Press

[102] Walsh, C., 2016, "Workers, Capitalists, Wages, Employment and Welfare", Mimeo

[103] Werning, I., 2015, "Incomplete markets and aggregate demand", Mimeo

[104] Wiederholt, M. 2015, "Empirical Properties of Inflation Expectations at the Zero Lower Bound," Mimeo

[105] Woodford, M., 1990, "Public Debt as Private Liquidity," American Economic Review Papers and Proceedings, 80 (2), pp. 382-388

[106] Woodford, M., 1996, "Controlling the public debt: a requirement for price stability?" Mimeo

[107] Woodford, M., 2003, Interest and prices: foundations of a theory of monetary policy, Princeton University Press

[108] Woodford, M. 2011. "Simple Analytics of the Government Expenditure Multiplier," American Economic Journal: Macroeconomics, 3(1), p 1-35.

[109] Woodford, M. 2018. "Monetary Policy when Planning Horizons are Finite," NBER Macroeconomics Annual

A Model Details

This Appendix presents in detail the equilibrium conditions of the model.

A.1 Aggregate Demand: Asset Markets Details

There is a mass 1 of households, indexed by $j \in [0,1]$, who discount the future at rate β and derive utility from consumption C_t^j and disutility from labor supply N_t^j. Households have access to two assets: a government-issued riskless bond (with nominal return $i_t > 0$), and shares in monopolistically competitive firms.

Households participate infrequently in financial markets. When they do, they can freely adjust their portfolio and receive dividends from firms. When they do not, they can use only bonds to smooth consumption. Denote by s the probability to keep participating in period $t+1$, conditional upon participating at t (hence, the probability to switch to not participating is $1-s$). Likewise, call h the probability to keep non-participating in period $t+1$, conditional upon not participating at t (hence, the probability to become a participant is $1-h$). The fraction of *non-participating* households is $\lambda = (1-s)/(2-s-h)$, and the fraction $1-\lambda$ participates.

Furthermore, households belong to a family whose head maximizes the intertemporal welfare of family members using a utilitarian welfare criterion (all households are equally weighted), but faces some limits to the amount of risk sharing that it can do. Households can be thought of as being in two states or "islands"[36]. All households who are participating in financial markets are on the same island, called S. All households who are not participating in financial markets are on the same island, called H. The family head can transfer *all* resources across households *within* the island, but cannot transfer *some* resources *between* islands.

T*iming*: At the beginning of the period, the family head pools resources within the island. The aggregate shocks are revealed and the family head determines the consumption/saving choice for each household in each island. Then households learn their next-period participation status and have to move to the corresponding island accordingly, taking *only bonds* with them. There are no transfers to households *after* the idiosyncratic shock is revealed, and this taken as a constraint for the consumption/saving choice.

The *flows across islands* are as follows. The total measure of households leaving the H island each period is the number of households who participate next period: $\lambda(1-h)$. The measure of households staying on the island is thus λh. In addition, a measure $(1-s)(1-\lambda)$ leaves the S island for the H island at the end of each period.

Total welfare maximization implies that the family head pools resources at the beginning of the period in a given island and implements symmetric consumption/saving choices for all households in that island. Denote as B_{t+1}^S the per-capita *beginning-of-period-*$t+1$ bonds of S: after the

[36] This follows e.g. Challe et al (2017) and Bilbiie and Ragot (2016).

consumption-saving choice, and also after changing state and pooling. The *end-of-period-t* per capita real values (*after* the consumption/saving choice but *before* agents move across islands) are Z_{t+1}^S. Denote as B_t^H the per capita beginning-of-period bonds in the H island (where the only asset is bonds). The end-of-period values (*before* agents move across islands) are Z_{t+1}^H. We have the following relations:

$$\begin{aligned}(1-\lambda)\, B_{t+1}^S &= (1-\lambda)\, sZ_{t+1}^S + (1-\lambda)(1-s)\, Z_{t+1}^H \\ \lambda B_{t+1}^H &= \lambda (1-h)\, Z_{t+1}^S + \lambda h Z_{t+1}^H.\end{aligned} \tag{37}$$

or rescaling by the relative population masses and using $\lambda = \frac{1-s}{1-s+1-h}$:

$$\begin{aligned}B_{t+1}^S &= sZ_{t+1}^S + (1-s)\, Z_{t+1}^H \\ B_{t+1}^H &= (1-h)\, Z_{t+1}^S + hZ_{t+1}^H.\end{aligned} \tag{38}$$

(as stocks do not leave the S island, we can ignore them).

The *program of the family head* is (with π_t denoting the net inflation rate):

$$\begin{aligned}W\left(B_t^S, B_t^H, \omega_t\right) \;\; = & \max_{\{C_t^S, Z_{t+1}^S Z_{t+1}^H, C_t^H, \omega_{t+1}\}} (1-\lambda)\, U\left(C_t^S\right) + \lambda U\left(C_t^H\right) \\ & +\beta E_t W\left(B_{t+1}^S, B_{t+1}^H, \omega_{t+1}\right)\end{aligned}$$

subject to:

$$C_t^S + Z_{t+1}^S + v_t\omega_{t+1} = Y_t^S + \frac{1+i_{t-1}}{1+\pi_t} B_t^S + \omega_t\left(v_t + D_t\right),$$

$$C_t^H + Z_{t+1}^H = Y_t^H + \frac{1+i_{t-1}}{1+\pi_t} B_t^H \tag{39}$$

$$Z_{t+1}^S, Z_{t+1}^H \geq 0 \tag{40}$$

and the laws of motion for bond flows relating the Zs to the Bs, (38). S-households (who own all the firms) receive dividends D_t, and the real return on bond holdings. With these resources they consume and save in bonds and shares. Equation (39) is the budget constraint of H. Finally (40) are positive constraints on bond holdings. Using the first-order and envelope conditions, we have:

$$U'\left(C_t^S\right) \geq \beta E_t\left\{\frac{v_{t+1}+D_{t+1}}{v_t}U'\left(C_{t+1}^S\right)\right\} \text{ and } \omega_{t+1}=\omega_t=(1-\lambda)^{-1}; \tag{41}$$

$$U'\left(C_t^S\right) \geq \beta E_t\left\{\frac{1+i_t}{1+\pi_{t+1}}\left[sU'\left(C_{t+1}^S\right)+(1-s)\,U'\left(C_{t+1}^H\right)\right]\right\} \tag{42}$$

$$\text{and } 0 = Z_{t+1}^S\left[U'\left(C_t^S\right)-\beta E_t\left\{\frac{1+i_t}{1+\pi_{t+1}}\left[sU'\left(C_{t+1}^S\right)+(1-s)\,U'\left(C_{t+1}^H\right)\right]\right\}\right]$$

$$U'\left(C_t^H\right) \geq \beta E_t\left\{\frac{1+i_t}{1+\pi_{t+1}}\left[(1-h)\,U'\left(C_{t+1}^S\right)+hU'\left(C_{t+1}^H\right)\right]\right\} \tag{43}$$

$$\text{and } 0 = Z_{t+1}^H\left[U'\left(C_t^H\right)-\beta E_t\left\{\frac{1+i_t}{1+\pi_{t+1}}\left[(1-h)\,U'\left(C_{t+1}^S\right)+hU'\left(C_{t+1}^H\right)\right]\right\}\right]$$

The first Euler equation corresponds to the choice of stock: there is no self-insurance motive, for they cannot be carried to the H state: the equation is the same as with a representative agent.[37]

The bond choice of S-island agents is governed by (42), which takes into account that bonds can be used when moving to the H island. The third equation (43) determines the bond choice of agents in the H island; both bond Euler conditions are written as complementary slackness conditions.

With this market structure, the Euler equations (42) and (43) of the same form as in fully-fledged incomplete-markets model of the Bewely-Huggett-Aiyagari type. In particular, the probability $1-s$ measures the uninsurable risk to switch to a bad state next period, risk for which only bonds can be used to self-insure—thus generating a demand for bonds for "precautionary" purposes.

Two more assumptions deliver our simple equilibrium representation. First, we focus on equilibria where (whatever the reason) the constraint of H agents always binds and their Euler "equation" is in fact a strict inequality (for instance, because the shock is a "liquidity" or impatience shock making them want to consume more today, or because their average income in that state is lower enough than in the S state—as would be the case if average profits were high enough; or simply because of a technological constraint preventing them from accessing any asset markets).

For the most part, we work with the **zero-liquidity limit**. That is, we assume that even though the demand for bonds from S is well-defined (the constraint is not binding), the supply of bonds is zero—so there are no bonds traded in equilibrium. Under these assumptions the only equilibrium condition from this part of the model is the Euler equation for bonds of S. The Euler equation of shares simply determines the share price v_t, and the fact that H's constrain binds

[37] As households pool resources when participating (which would be optimal with t=0 symmetric agents and t=0 trading), they perceive a return conditional on participating next period. This exactly compensates for the probability of not participating next period, thus generating the same Euler equation as with a representative agent.

implies that they are hand-to-mouth $C_t^H = Y_t^H$.

A.2 Liquidity: the a-HANK as a representation of intertemporal MPCs

Assume that liquidity is supplied by the government through issuing a bond: denote by B_{t+1}^N the total nominal quantity of bonds outstanding at the end of each period. In nominal terms, $B_{t+1}^N = (1 + i_{t-1}) B_t^N - \mathcal{P}_t T_t$, and in real terms:

$$B_{t+1} = (1 + r_t) B_t - T_t \tag{44}$$

where $1 + r_t = \frac{1+i_{t-1}}{1+\pi_t}$.

The bond market clears $B_{t+1} = \lambda Z_{t+1}^H + (1 - \lambda) Z_{t+1}^S$. Denoting the disposable (net of taxes) income of agent j by $\hat{Y}_t^j$ (where how this is determined in equilibrium depends on the particular model), we have for H

$$C_t^H + Z_{t+1}^H = \hat{Y}_t^H + (1 + r_t) B_t^H$$

Recall now that $Z_{t+1}^H = 0$, so that $B_{t+1} = (1 - \lambda) Z_{t+1}^S$; using the flow definitions:

$$B_{t+1}^H = (1 - h) Z_{t+1}^S = \frac{1-h}{1-\lambda} B_{t+1} = \frac{1-s}{\lambda} B_{t+1}$$

Replacing

$$C_t^H = \hat{Y}_t^H + \frac{1-s}{\lambda} (1 + r_t) B_t$$

Similarly for S we obtain (using $B_{t+1}^S = s Z_{t+1}^S = \frac{s}{1-\lambda} B_{t+1}$):

$$C_t^S + \frac{1}{1-\lambda} B_{t+1} = \hat{Y}_t^S + \frac{s}{1-\lambda} (1 + r_t) B_t.$$

Loglinearizing around a long-run steady-state with zero public debt (and thus zero liquidity) $B = 0$—one form of specifying a Ricardian, passive fiscal policy—we obtain:

$$\begin{aligned} c_t^H &= \hat{y}_t^H + \frac{1-s}{\lambda} \beta^{-1} b_t \\ c_t^S + \frac{1}{1-\lambda} b_{t+1} &= \hat{y}_t^S + \frac{s}{1-\lambda} \beta^{-1} b_t, \end{aligned}$$

where we used that in a steady-state with zero liquidity and no inequality $C^H = C^S$, the self-insurance Euler equation for bonds implies $1 + r = \beta^{-1}$. Loglinearizing the self-insurance equation we have the equivalent of (7):

$$c_t^S = s E_t c_{t+1}^S + (1 - s) E_t c_{t+1}^H - \sigma r_t. \tag{45}$$

Derivation of analytical iMPCs (proof of Proposition 2)

The solution of the asset-accumulation equation implies the following recursions for the responses of assets to income shocks:

$$
\begin{aligned}
t &\leq T-1: \frac{db_{t+1}}{d\hat{y}_T} = x_b \frac{db_t}{d\hat{y}_T} + \frac{1-\lambda}{s} \left(\beta x_b\right)^{T-t} \left(\beta x_b - \delta\right) \\
t &= T: \frac{db_{t+1}}{d\hat{y}_T} = x_b \frac{db_t}{d\hat{y}_T} + \frac{1-\lambda}{s} \beta x_b \\
t &> T: \frac{db_{t+1}}{d\hat{y}_T} = x_b \frac{db_t}{d\hat{y}_T}
\end{aligned}
$$

The solutions of these equations are (setting initial debt equal to steady-state without loss of generality):

$$
\begin{aligned}
t &\leq T-1: \frac{db_{t+1}}{d\hat{y}_T} = \left(\beta x_b\right)^{T-t} \frac{1-\lambda\chi}{s} \left(\beta x_b - \delta\right) \frac{1-\left(x_b \beta x_b\right)^{t+1}}{1-x_b \beta x_b} \\
t &= T: \frac{db_{T+1}}{d\hat{y}_T} = x_b \beta x_b \frac{1-\lambda\chi}{s} \left(\beta x_b - \delta\right) \frac{1-\left(x_b \beta x_b\right)^{T}}{1-x_b \beta x_b} + \frac{1-\lambda\chi}{s} \beta x_b \\
t &\geq T+1: \frac{db_{t+1}}{d\hat{y}_T} = x_b^{t-T} \frac{db_{T+1}}{d\hat{y}_T}
\end{aligned}
$$

Taking derivatives of the consumption function 15, we have:

$$
\begin{aligned}
t &\leq T-1: \frac{dc_t}{d\hat{y}_T} = \beta^{-1} \left(1-\beta x_b\right) \frac{db_t}{d\hat{y}_T} + \frac{1-\lambda\chi}{s} \left(\beta x_b\right)^{T-t} \left(\delta - \beta x_b\right) \\
t &= T: \frac{dc_t}{d\hat{y}_T} = 1 + \beta^{-1} \left(1-\beta x_b\right) \frac{db_t}{d\hat{y}_T} - \frac{1-\lambda\chi}{s} \beta x_b \\
t &> T: \frac{dc_t}{d\hat{y}_T} = \beta^{-1} \left(1-\beta x_b\right) \frac{db_t}{d\hat{y}_T}
\end{aligned}
$$

Replacing the solution for assets:

$$
\begin{aligned}
t &\leq T-1: \frac{dc_t}{d\hat{y}_T} = \beta^{-1} \left(1-\beta x_b\right) \left(\beta x_b\right)^{T-t+1} \frac{1-\lambda\chi}{s} \left(\beta x_b - \delta\right) \frac{1-\left(x_b \beta x_b\right)^{t}}{1-x_b \beta x_b} + \frac{1-\lambda\chi}{s} \left(\beta x_b\right)^{T-t} \left(\delta - \beta x_b\right) \\
t &= T: \frac{dc_t}{d\hat{y}_T} = 1 + \beta^{-1} \left(1-\beta x_b\right) \beta x_b \frac{1-\lambda\chi}{s} \left(\beta x_b - \delta\right) \frac{1-\left(x_b \beta x_b\right)^{T}}{1-x_b \beta x_b} - \frac{1-\lambda\chi}{s} \beta x_b \\
t &\geq T+1: \frac{dc_t}{d\hat{y}_T} = \beta^{-1} \left(1-\beta x_b\right) \frac{db_t}{d\hat{y}_T} = \beta^{-1} \left(1-\beta x_b\right) x_b^{t-T-1} \frac{db_{T+1}}{d\hat{y}_T} \\
&= \beta^{-1} \left(1-\beta x_b\right) x_b^{t-T-1} \left(x_b \beta x_b \frac{1-\lambda\chi}{s} \left(\beta x_b - \delta\right) \frac{1-\left(x_b \beta x_b\right)^{T}}{1-x_b \beta x_b} + \frac{1-\lambda\chi}{s} \beta x_b \right)
\end{aligned}
$$

Rewriting and simplifying, we obtain the expressions in Proposition 2. Notice that, as argued by Auclert et al, the present discounted sum of the iMPCs needs to be 1 (the increase in income is consumed entirely, sooner or later). To prove that the iMPCs in a-HANK derived here satisfy this property, replace the respective solution into the sum:

$$\sum_{t=0}^{T-1} \beta^{t-T} \frac{dc_t}{d\hat{y}_T} + \frac{dc_T}{d\hat{y}_T} + \sum_{t=T+1}^{\infty} \beta^{t-T} \frac{dc_t}{d\hat{y}_T}$$

obtaining

$$\begin{aligned}
& \frac{1-\lambda\chi}{s} \frac{\delta - \beta x_b}{1-\beta x_b^2} \beta^T x_b \left[1 - \left(\beta x_b^2\right)^T\right] \\
& + 1 - \frac{1-\lambda\chi}{s} \beta x_b - \left(\delta - \beta x_b\right) x_b \frac{1-\lambda\chi}{s} \left(1 - \beta x_b\right) \frac{1 - \left(x_b \beta x_b\right)^T}{1 - x_b \beta x_b} \\
& + \frac{1-\lambda\chi}{s} \frac{\beta x_b}{1-\beta x_b^2} \left(1 - x_b \delta + x_b \left(\delta - \beta x_b\right) \left(\beta x_b^2\right)^T\right) \\
= \; & 1
\end{aligned}$$

The **calibration** in text following Auclert et al concerns two iMPCs, $\frac{dc_0}{d\hat{y}_0} = 1 - \frac{1-\lambda\chi}{s} \beta x_b$ and $\frac{dc_1}{d\hat{y}_0} = \frac{1-\lambda\chi}{s} \left(1 - \beta x_b\right) x_b$.

Part (ii) of the Proposition concerns the dependence on χ (and δ Euler-compounding), keeping fixed the time-0 contemporaneous MPC $\frac{dc_0}{d\hat{y}_0}$; denote this by:

$$m_{00} \equiv \frac{dc_0}{d\hat{y}_0} = 1 - \frac{1-\lambda\chi}{s} \beta x_b$$

Replacing in the Proposition and rewriting the iMPCs, taking the derivative with respect to the cyclicality of inequality χ we obtain:

$$\frac{\partial \frac{dc_t}{d\hat{y}_T}}{\partial \chi} |_{\overline{m}_{00}} = \frac{\partial}{\partial \chi} \begin{cases} \frac{1-m_{00}}{\beta x_b} \frac{\delta - \beta x_b}{1-\beta x_b^2} \left(\beta x_b\right)^{T-t} \left(1 - x_b + x_b \left(1 - \beta x_b\right) \left(\beta x_b^2\right)^t\right), & \text{if} \quad t \leq T-1; \\ 1 - \frac{1-m_{00}}{\beta x_b} \beta x_b - \left(\delta - \beta x_b\right) \frac{1-m_{00}}{\beta} \left(1 - \beta x_b\right) \frac{1-\left(\beta x_b^2\right)^T}{1-\beta x_b^2}, & \text{if} \quad t = T; \\ \frac{1-m_{00}}{\beta x_b} \frac{1-\beta x_b}{1-\beta x_b^2} x_b^{t-T} \left(1 - x_b \delta + x_b \left(\delta - \beta x_b\right) \left(\beta x_b^2\right)^T\right), & \text{if} \quad t \geq T+1. \end{cases}$$

It follows directly that "anticipation iMPCs" ($t < T$) are increasing in χ (using $\frac{\partial \delta}{\partial \chi} = (1-s) \frac{1-\lambda}{(1-\lambda\chi)^2} > 0$); iMPCs out of past income ($t > T$) are decreasing in χ (the derivative is proportional to $-x_b \left(1 - \left(\beta x_b^2\right)^T\right) \frac{\partial \delta}{\partial \chi} < 0$), and decrease the contemporaneous MPC at given T.

A.3 Aggregate Supply: New Keynesian Phillips Curve

The intermediate goods producers solve:

$$\max_{P_t(k)} E_0 \sum_{t=0}^{\infty} Q_{0,t}^S \left[\left(1 + \tau^S\right) P_t(k) Y_t(k) - W_t N_t(k) - \frac{\psi}{2} \left(\frac{P_t(k)}{P_{t-1}^{**}} - 1\right)^2 P_t Y_t\right],$$

where I consider two possibilities for the reference price level P^{**}_{t-1}, with respect to which it is costly for firms to deviate. In the first scenario, this is the aggregate price index P_{t-1} which small atomistic firms take as given—this delivers the static Phillips curve. In the second, P^{**}_{t-1} is firm k's own individual price as in standard formulations. $Q^S_{0,t} \equiv \beta^t \left(P_0 C^S_0 / P_t C^S_t\right)^{\sigma^{-1}}$ is the marginal rate of intertemporal substitution of participants between times 0 and t, and τ^S the sales subsidy. Firms face demand for their products from two sources: consumers and firms themselves (in order to pay for the adjustment cost); the demand function for the output of firms z is $Y_t(z) = (P_t(z)/P_t)^{-\varepsilon} Y_t$. Substituting this into the profit function, the first-order condition is, after simplifying, for each case:

Static PC case $P^{**}_{t-1} = P_{t-1}$

$$0 = Q_{0,t}\left(\frac{P_t(k)}{P_t}\right)^{-\varepsilon} Y_t \left[\left(1+\tau^S\right)(1-\varepsilon) + \varepsilon \frac{W_t}{P_t}\left(\frac{P_t(k)}{P_t}\right)^{-1}\right] - Q_{0,t}\psi P_t Y_t \left(\frac{P_t(k)}{P_{t-1}} - 1\right)\frac{1}{P_{t-1}}$$

In a symmetric equilibrium all producers make identical choices (including $P_t(k) = P_t$); defining net inflation $\pi_t \equiv (P_t/P_{t-1}) - 1$, this becomes:

$$\pi_t\left(1+\pi_t\right) = \frac{\varepsilon - 1}{\psi}\left[\frac{\varepsilon}{\varepsilon - 1} w_t - \left(1+\tau^S\right)\right],$$

loglinearization of which delivers the static PC in text (3).

Dynamic PC case $P^{**}_{t-1} = P_{t-1}$; the first-order condition is

$$\begin{aligned} 0 = {} & Q_{0,t}\left(\frac{P_t(k)}{P_t}\right)^{-\varepsilon} Y_t \left[\left(1+\tau^S\right)(1-\varepsilon) + \varepsilon \frac{W_t}{P_t}\left(\frac{P_t(k)}{P_t}\right)^{-1}\right] \\ & - Q_{0,t}\psi P_t Y_t \left(\frac{P_t(k)}{P_{t-1}(k)} - 1\right)\frac{1}{P_{t-1}(k)} + \\ & + E_t\left\{Q_{0,t+1}\left[\psi P_{t+1} Y_{t+1}\left(\frac{P_{t+1}(k)}{P_t(k)} - 1\right)\frac{P_{t+1}(k)}{P_t(k)^2}\right]\right\} \end{aligned}$$

In a symmetric equilibrium, using again the definition of net inflation π_t, and noticing that $Q_{0,t+1} = Q_{0,t}\beta\left(C^S_t / C^S_{t+1}\right)^{\sigma^{-1}}\left(1+\pi_{t+1}\right)^{-1}$, this becomes:

$$\begin{aligned} \pi_t\left(1+\pi_t\right) = {} & \beta E_t\left[\left(\frac{C^S_t}{C^S_{t+1}}\right)^{\sigma^{-1}} \frac{Y_{t+1}}{Y_t}\pi_{t+1}\left(1+\pi_{t+1}\right)\right] + \\ & + \frac{\varepsilon - 1}{\psi}\left[\frac{\varepsilon}{\varepsilon - 1} w_t - \left(1+\tau^S\right)\right], \end{aligned}$$

the loglinearization of which delivers the NKPC in text (2). Notice that this nests the static PC when the discount factor of firms $\beta = 0$.

B Cyclical Idiosyncratic Risk

The self-insurance equation when the probability depends on aggregate demand (today) is

$$\left(C_t^S\right)^{-\frac{1}{\sigma}} = \beta E_t \left\{ \frac{1+i_t}{1+\pi_{t+1}} \left[s\left(C_t\right) \left(C_{t+1}^S\right)^{-\frac{1}{\sigma}} + \left(1 - s\left(C_t\right)\right) \left(C_{t+1}^H\right)^{-\frac{1}{\sigma}} \right] \right\}. \tag{46}$$

We loglinearize this around a steady-state with inequality; in the context of our model, that requires assuming that stead-state fiscal redistribution is imperfect and that a sales subsidy does not completely undo market power (generating zero profits). In particular, we focus on a steady state with no subsidy, so that the profit share is $D/C = 1/\varepsilon$ and the labor share $WN/C = (\varepsilon - 1)/\varepsilon$. Under the same arbitrary redistribution scheme, the consumption shares of each type are respectively

$$\begin{aligned}
\frac{C^H}{C} &= \frac{WN + \frac{\tau^D}{\lambda} D}{C} = 1 - \frac{1}{\varepsilon}\left(1 - \frac{\tau^D}{\lambda}\right) \\
\frac{C^S}{C} &= \frac{WN + \frac{1-\tau^D}{1-\lambda} D}{C} = 1 + \frac{1}{\varepsilon}\frac{\lambda}{1-\lambda}\left(1 - \frac{\tau^D}{\lambda}\right) > \frac{C^H}{C} \text{ iff } \tau^D < \lambda.
\end{aligned}$$

Denoting steady-state inequality $\frac{C^S}{C^H} \equiv \Gamma$ we loglinearize around a steady state:

$$1 = \beta\left(1 + r\right)\left[s\left(C\right) + \left(1 - s\left(C\right)\right)\Gamma^{\frac{1}{\sigma}}\right], \tag{47}$$

where I restrict attention to cases with positive real interest-rate r (the topic of "secular stagnation" in this framework is interesting in its own right—it can occur for high enough risk and high enough inequality). Loglinearization delivers, denoting by r_t the ex-ante real interest rate for brevity, and the steady-state value of the probability by $s\left(C\right) = s$ and its elasticity relative to the cycle (consumption) is $-\frac{s'(C)C}{1-s(C)}$:

$$c_t^S = -\sigma r_t + \frac{s}{s + (1-s)\Gamma^{1/\sigma}} \mathrm{E}_t c_{t+1}^S + \frac{(1-s)\Gamma^{1/\sigma}}{s + (1-s)\Gamma^{1/\sigma}} \mathrm{E}_t c_{t+1}^H + \left(-\frac{s'(C)C}{1-s(C)}\right) \frac{\sigma(1-s)\left(1-\Gamma^{1/\sigma}\right)}{s+(1-s)\Gamma^{1/\sigma}} \mathrm{E}_t c_{t+1}$$

which replacing individual consumption levels as function of aggregate becomes

$$c_t^S = -\sigma\frac{1-\lambda}{1-\lambda\chi} r_t + \left(1 + \frac{1-\tilde{s}}{1-\lambda\chi}(\chi - 1) - \left(-\frac{s'(C)C}{1-s(C)}\right)(1-\tilde{s})\frac{\sigma(1-\lambda)}{1-\lambda\chi}\left(1-\Gamma^{-1/\sigma}\right)\right)\mathrm{E}_t c_{t+1}$$

denote by $1 - \tilde{s} = \frac{(1-s)\Gamma^{1/\sigma}}{s+(1-s)\Gamma^{1/\sigma}} > 1 - s$ the inequality-weighted transition probability, the relevant inequality-adjusted measure of risk given steady-state inequality coming from financial income $\Gamma \equiv Y^S/Y^H \geq 1$. There can be discounting as long as risk is procyclical enough $\eta > \frac{\Gamma^{1/\sigma}(\chi-1)}{\sigma(1-\lambda)\left(\Gamma^{1/\sigma}-1\right)}$. But the contemporary AD elasticity to interest rates is unaffected by the cyclicality of risk (as in

Acharya and Dogra (2018)).

B.1 Current aggregate demand

For the case where the probability depends on current aggregate demand, the aggregate Euler-IS is

$$c_t^S = -\sigma r_t + \beta (1+r) s \mathrm{E}_t c_{t+1}^S + \beta (1+r)(1-s) \Gamma^{\frac{1}{\sigma}} \mathrm{E}_t c_{t+1}^H + \sigma \beta (1+r) \left(-\frac{s'(C) C}{1-s(C)} \right) (1-s) \left(1 - \Gamma^{\frac{1}{\sigma}} \right) c_t$$

Replacing $\beta(1+r)$

$$c_t^S = -\sigma r_t + \frac{s}{s+(1-s)\Gamma^{\frac{1}{\sigma}}} \mathrm{E}_t c_{t+1}^S + \frac{(1-s)\Gamma^{\frac{1}{\sigma}}}{s+(1-s)\Gamma^{\frac{1}{\sigma}}} \mathrm{E}_t c_{t+1}^H + \left(-\frac{s'(C) C}{1-s(C)} \right) \frac{\sigma (1-s) \left(1 - \Gamma^{\frac{1}{\sigma}} \right)}{s+(1-s)\Gamma^{\frac{1}{\sigma}}} c_t$$

Replace the consumption functions of H and S we obtain:

$$\begin{aligned} c_t &= \theta \delta E_t c_{t+1} - \theta \sigma \frac{1-\lambda}{1-\lambda\chi} \left(i_t - E_t \pi_{t+1} - \rho_t \right) \qquad (48) \\ \text{with } \theta &\equiv \left[1 + \left(-\frac{s'(C) C}{1-s(C)} \right) \left(1 - \Gamma^{-1/\sigma} \right) (1-\tilde{s}) \sigma \frac{1-\lambda}{1-\lambda\chi} \right]^{-1}, \end{aligned}$$

where I denote by $1-\tilde{s} = \frac{(1-s)\Gamma^{1/\sigma}}{s+(1-s)\Gamma^{1/\sigma}} > 1-s$ the inequality-weighted transition probability, the relevant inequality-adjusted measure of risk given steady-state inequality coming from financial income $\Gamma \equiv Y^S/Y^H \geq 1$. Notice that the discounting/compounding parameter due to cyclical inequality has a slightly different expression now $\delta \equiv 1 + \frac{(\chi-1)(1-\tilde{s})}{1-\lambda\chi}$, generalized to the case with steady-state inequality. Notice that now the two channels (cyclical inequality and cyclical risk via $s'(.)$) are intertwined for both the amplification/dampening of current interest rates and for future consumption. A previous working paper version contained a full analysis of this version of the model and its implications for curing puzzles and the Catch-22.

C The analytical-HANK 3-equation model with NKPC

This section derives the same results as in text but with the forward-looking NKPC (2).

C.1 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

(2003) Proposition C.1. With the Taylor rule (18), the system becomes $\begin{pmatrix} E_t\pi_{t+1} & E_t c_{t+1} \end{pmatrix}' = A \begin{pmatrix} \pi_t & c_t \end{pmatrix}'$ with transition matrix:

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

with determinant $\det A = \beta^{-1}\delta^{-1}\left(1+\kappa\sigma\frac{1-\lambda}{1-\lambda\chi}\phi_\pi\right)$ and trace $\mathrm{tr}A = \beta^{-1}+\delta^{-1}\left(1+\sigma\frac{1-\lambda}{1-\lambda\chi}\beta^{-1}\kappa\right)$.

Determinacy can obtain in either of two cases. Case 2. ($\det A - \mathrm{tr}A < -1$ and $\det A + \mathrm{tr}A < -1$) can be ruled based on sign restrictions. Case 1. requires three conditions to be satisfied jointly:

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

The third condition is always satisfied under the sign restrictions, so the necessary and sufficient conditions are:

$$\phi_\pi > 1 + \frac{(\delta-1)(1-\beta)}{\kappa\sigma\frac{1-\lambda}{1-\lambda\chi}}$$

$$\phi_\pi > \max\left(\frac{\beta\delta-1}{\kappa\sigma\frac{1-\lambda}{1-\lambda\chi}},\ 1+\frac{(1-\beta)(\delta-1)}{\kappa\sigma\frac{1-\lambda}{1-\lambda\chi}}\right) \tag{49}$$

The second term is larger than the first iff $\delta < \frac{\kappa\sigma\frac{1-\lambda}{1-\lambda\chi}+\beta}{2\beta-1}$. Condition (49) thus generalize the HANK Taylor principle to the case of forward-looking Phillips curve.

C.2 Ruling out FG Puzzle and neo-Fisherian Effects

The analogous of Proposition 4 for the case with NKPC (2) is:

Proposition 10 *The analytical HANK model (with (2)) under a peg is locally determinate and solves the FG puzzle ($\frac{\partial^2 c_t}{\partial\left(-i^*_{t+T}\right)\partial T} < 0$) if and only if:*

$$\delta + \sigma\frac{1-\lambda}{1-\lambda\chi}\frac{\kappa}{1-\beta} < 1,$$

Notice that the condition nests the one of Proposition 4 when $\beta \to 0$. Indeed, it has exactly the same interpretation with $\delta + \sigma\frac{1-\lambda}{1-\lambda\chi}\frac{\kappa}{1-\beta}$ being the "long-run" effect of news, and $\frac{\kappa}{1-\beta}$ being the slope of the long-run NKPC.

Point 1. (determinacy under a peg with NKPC) follows directly from (49): a peg is sufficient if both $\delta < \beta^{-1}$ and $1 + \frac{(1-\beta)(\delta-1)}{\kappa\sigma\frac{1-\lambda}{1-\lambda\chi}} < 0$, the latter implying $\delta < 1 - \frac{\kappa}{1-\beta}\sigma\frac{1-\lambda}{1-\lambda\chi} < \beta^{-1}$, which delivers the threshold in the Proposition.

Point 2 requires solving the model; focusing therefore on the case where the condition holds, and the model is determinate under a peg, we rewrite the model in forward (matrix) form as:

$$\begin{pmatrix} \pi_t \\ c_t \end{pmatrix} = A^{-1} \begin{pmatrix} E_t \pi_{t+1} \\ E_t c_{t+1} \end{pmatrix} - \sigma \frac{1-\lambda}{1-\lambda\chi} \begin{pmatrix} \kappa \\ 1 \end{pmatrix} i_t^* \tag{50}$$

where

$$A^{-1} = \begin{pmatrix} \beta + \kappa\sigma \frac{1-\lambda}{1-\lambda\chi} & \kappa\delta \\ \sigma \frac{1-\lambda}{1-\lambda\chi} & \delta \end{pmatrix}$$

is the inverse of matrix A defined above under a peg $\phi = 0$. To find the elasticity of $\begin{pmatrix} \pi_t & c_t \end{pmatrix}'$ to an interest rate cut at T, $-i_{t+T}^*$ we iterate forward (50) to obtain $\sigma \frac{1-\lambda}{1-\lambda\chi} \left(A^{-1}\right)^T \begin{pmatrix} \kappa \\ 1 \end{pmatrix}$. But notice that we know by point 1 that the eigenvalues of A are both outside the unit circle; it follows by standard linear algebra results that the eigenvalues of A^{-1} are both inside the unit circle and therefore $\left(A^{-1}\right)^T$ is decreasing with T. (the eigenvalues to the power of T appear in the Jordan decomposition used to compute the power of A^{-1}). This proves that the FG puzzle is eliminated.

Point 3 requires computing the equilibrium given an AR1 interest rate with persistence μ as before $E_t i_{t+1}^* = \mu i_t^*$; since we are in the determinate case, the equilibrium is unique and there is no endogenous persistence, so the persistence of endogenous variables is equal to the persistence of the exogenous process. Replacing $E_t c_{t+1} = \mu c_t$ and $E_t \pi_{t+1} = \mu \pi_t$ in (50) we therefore have:

$$\begin{pmatrix} \pi_t \\ c_t \end{pmatrix} = -\sigma \frac{1-\lambda}{1-\lambda\chi} \left(I - \mu A^{-1}\right)^{-1} \begin{pmatrix} \kappa \\ 1 \end{pmatrix} i_t^*.$$

Computing the inverse we obtain

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

where $\det \equiv \mu^2 \beta\delta - \mu \left(\delta + \sigma \frac{1-\lambda}{1-\lambda\chi}\kappa + \beta\right)\mu + 1$. Replacing in the previous equation, differentiating, and simplifying, the effects are:

$$\begin{pmatrix} \frac{\partial \pi_t}{\partial i_t^*} \\ \frac{\partial c_t}{\partial i_t^*} \end{pmatrix} = -\sigma \frac{1-\lambda}{1-\lambda\chi} \frac{1}{\det} \begin{pmatrix} \kappa \\ 1 - \mu\beta \end{pmatrix}$$

Therefore, neo-Fisherian effects are ruled out iff $\det > 0$, i.e.:

$$\delta < \frac{1 - \beta\mu - \sigma \frac{1-\lambda}{1-\lambda\chi}\kappa\mu}{\mu\left(1 - \beta\mu\right)}.$$

But this is always satisfied under the condition in the proposition (for determinacy under a peg) $\delta < 1 - \frac{\sigma\frac{1-\lambda}{1-\lambda\chi}\kappa}{1-\beta} \leq \frac{1-\beta\mu-\sigma\frac{1-\lambda}{1-\lambda\chi}\kappa\mu}{\mu(1-\beta\mu)}$ where the second inequality can be easily verified (it implies $[(1-\beta\mu)(1-\beta)+\beta\sigma\kappa\mu](1-\mu)\geq 0$).

C.3 Liquidity trap and FG

Under the Markov chain structure used in text, we can use the same solution method to obtain the LT equilibrium under forward guidance (which evidently nests the LT equilibrium without FG) . Using the notations:

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

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

the equilibrium is:

$$\begin{aligned}
c_F &= \frac{1}{1-q\nu_{0q}}\sigma\frac{1-\lambda}{1-\lambda\chi}\rho; \\
c_L &= \frac{(1-p)\,q\nu_{0zq}}{(1-q\nu_{0q})(1-z\nu_{0z})}\sigma\frac{1-\lambda}{1-\lambda\chi}\rho + \frac{1}{1-z\nu_{0z}}\sigma\frac{1-\lambda}{1-\lambda\chi}\rho_L,
\end{aligned} \tag{51}$$

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

C.4 Ruling out puzzles with Wicksellian rule and Contemporaneous PC

Replacing (3) and the policy rule (24) in the aggregate Euler-IS (23) we have

$$c_t = \nu_0 E_t c_{t+1} - \sigma\frac{1-\lambda}{1-\lambda\chi}\left(\phi_p p_t + i_t^*\right); \tag{52}$$

the static PC rewritten in terms of the price level is:

$$p_t - p_{t-1} = \kappa c_t. \tag{53}$$

Combining, we obtain:

$$E_t p_{t+1} - \left[1+\nu_0^{-1}\left(1+\sigma\frac{1-\lambda}{1-\lambda\chi}\phi_p\kappa\right)\right]p_t + \nu_0^{-1}p_{t-1} = \sigma\frac{1-\lambda}{1-\lambda\chi}\kappa\nu_0^{-1}i_t^*. \tag{54}$$

Notice that the RANK model is nested here for $\lambda = 0$ (or $\chi = 1$, the Campbell-Mankiw benchmark), which would yield a simplified version of Woodford and Giannoni's analyses.

Recall that we are interested in the case whereby $\nu_0 \geq 1$ (as the paper shows, for $\nu_0 < 1$ there there is determinacy under a peg in HANK and thus no puzzles). The model has a locally unique equilibrium (is determinate) when the above second-order equation has one root inside and one outside the unit circle. The characteristic polynomial is $J(x) = x^2 - \left[1 + (\nu_0)^{-1}\left(1 + \sigma\frac{1-\lambda}{1-\lambda\chi}\phi_p\kappa\right)\right]x + \nu_0^{-1}$ where by standard results, the roots' sum is $1+\nu_0^{-1}\left(1 + \sigma\frac{1-\lambda}{1-\lambda\chi}\phi_p\kappa\right)$ and the product is $\nu_0^{-1} < 1$. So at least one root is inside the unit circle, and we need to rule out that both are; Since we have $J(1) = -\nu_0^{-1}\sigma\frac{1-\lambda}{1-\lambda\chi}\phi_p\kappa$ and $J(-1) = 2 + 2\nu_0^{-1} + \nu_0^{-1}\sigma\frac{1-\lambda}{1-\lambda\chi}\phi_p\kappa$, the necessary and sufficient condition for the second root to be outside the unit circle is precisely $\phi_p > 0$—coming from $J(1) < 0$ and $J(-1) > 0$. This completes the proof of Proposition 7.

To find the solution, denote the roots of the polynomial by $x_+ > 1 > x_- > 0$; the difference equation is solved by standard factorization: The roots of the characteristic polynomial are

$$\begin{aligned} x_\pm &= \frac{1 + \nu_0^{-1}\left(1 + \sigma\frac{1-\lambda}{1-\lambda\chi}\phi_p\kappa\right) \pm \sqrt{\left[1 + \nu_0^{-1}\left(1 + \sigma\frac{1-\lambda}{1-\lambda\chi}\phi_p\kappa\right)\right]^2 - 4\nu_0^{-1}}}{2} \\ x_+ &> 1 > x_- > 0 \end{aligned}$$

Factorizing the difference equation (25):

$$\left(L^{-1} - x_-\right)\left(L^{-1} - x_+\right)p_{t-1} = \sigma\frac{1-\lambda}{1-\lambda\chi}\kappa\nu_0^{-1}i_t^*$$

we obtain:

$$\begin{aligned} p_t &= x_- p_{t-1} - \sigma\frac{1-\lambda}{1-\lambda\chi}\kappa\nu_0^{-1}x_+^{-1}\frac{1}{1-(x_+L)^{-1}}i_t^* \\ &= x_- p_{t-1} - \sigma\frac{1-\lambda}{1-\lambda\chi}\kappa\nu_0^{-1}x_+^{-1}\sum_{j=0}^{\infty} x_+^{-j} i_{t+j}^* \end{aligned}$$

Let $\Delta_{t+j} \equiv -\sigma\frac{1-\lambda}{1-\lambda\chi}\kappa\nu_0^{-1}x_+^{-1}i_{t+j}^*$ denote the rescaled interest rate *cut*:

$$p_t = x_-^{t+1}p_{-1} + \left[\sum_{j=0}^{\infty}(x_+)^{-j}\Delta_{t+j} + x_-\sum_{j=0}^{\infty}(x_+)^{-j}\Delta_{t-1+j} + ... + x_-^{t-1}\sum_{j=0}^{\infty}(x_+)^{-j}\Delta_{1+j} + x_-^t\sum_{j=0}^{\infty}(x_+)^{-j}\Delta_j\right]$$

Normalizing initial value to zero (since $x_- < 1$ it vanishes when t goes to infinity), the solution is made of a forward and a backward component:

$$p_t = \frac{1-\left(x_- x_+^{-1}\right)^{t+1}}{1 - x_- x_+^{-1}}\sum_{j=0}^{\infty}\left(x_+^{-1}\right)^j\Delta_{t+j} + \sum_{k=0}^{t-1} x_-^{1+k}\frac{1-\left(x_- x_+^{-1}\right)^{t-k}}{1-x_- x_+^{-1}}\Delta_{t-1-k}$$

Lagging it once and taking the first difference we obtain the solution for inflation:

$$\begin{aligned}
\pi_t &= \frac{1-\left(x_- x_+^{-1}\right)^{t+1}}{1-x_- x_+^{-1}} \sum_{j=0}^{\infty} \left(x_+^{-1}\right)^j \Delta_{t+j} - \frac{1-\left(x_- x_+^{-1}\right)^{t}}{1-x_- x_+^{-1}} \sum_{j=0}^{\infty} \left(x_+^{-1}\right)^j \Delta_{t-1+j} \\
&\quad + \sum_{k=0}^{t-1} x_-^{1+k} \frac{1-\left(x_- x_+^{-1}\right)^{t-k}}{1-x_- x_+^{-1}} \Delta_{t-1-k} - \sum_{k=0}^{t-2} x_-^{1+k} \frac{1-\left(x_- x_+^{-1}\right)^{t-1-k}}{1-x_- x_+^{-1}} \Delta_{t-2-k} \\
&= A\left(t\right) \sum_{j=0}^{\infty} \left(x_+^{-1}\right)^j \Delta_{t+j} + \Psi_{t-1}.
\end{aligned}$$

where $A\left(t\right) \equiv \frac{1-\left(x_+^{-1}\right)+\left(x_-\right)^t\left(x_+^{-1}\right)^{t+1}-\left(x_- x_+^{-1}\right)^{t+1}}{1-x_- x_+^{-1}}$ (if we put ourselves at time 0 this simply becomes $A\left(0\right) = \sigma \frac{1-\lambda}{1-\lambda\chi}\nu_0^{-1}$), while in Ψ_{t-1} we grouped all terms that consist of lags of Δ_t (Δ_{t-1} and earlier) which are predetermined at time t and will not be used in any of the derivations of interest here—where we consider shocks occurring at t or thereafter. This delivers, for consumption:

$$c_t = -A\left(t\right) E_t \sum_{j=0}^{\infty} \left(x_+^{-1}\right)^{j+1} i_{t+j}^* + \Psi_{t-1} \tag{55}$$

where Ψ_{t-1} is a weighted sum of past realizations of the shock and $A\left(t\right) > 0$ is a function only of calendar date; both Ψ_{t-1} and $A\left(t\right)$ are irrelevant for our purpose because they are invariant to current and future shocks.

The effect of a one-time interest rate cut at $t+T$ is now:

$$\frac{\partial c_t}{\partial\left(-i_{t+T}^*\right)} = A\left(t\right)\left(x_+^{-1}\right)^{T+1}$$

which, since $A\left(.\right) > 0$ and $x_+ > 1$, is a decreasing function of T: *the FG puzzle disappears.*[38] Notice that the Wicksellian rule *also* cures the FG puzzle in the (nested) RANK model (this follows immediately by replacing $\lambda = 0$ or $\chi = 1$ above).

C.5 Determinacy with Wicksellian rule and NKPC

Rewrite the system made of (10), (2) and the definition of inflation as (ignoring shocks):

[38] Likewise for neo-Fisherian effects: take an AR(1) process for i_t^* with persistence μ as before; the solution is now both 1. uniquely determined (by virtue of determinacy proved above) and 2. in line with standard logic—an increase in interest rates leads to a fall in consumption and deflation in the short run: $\frac{\partial c_t}{\partial i_t^*} = -A\left(t\right)\frac{1}{x_+ - \mu}$,which is negative as $A\left(.\right) > 0$ and $x_+ > 1 > \mu$. Notice that in the long-run, i.e. if there is a permanent change in interest rates, the economy moves to a new steady-state and the uncontroversial. long-run Fisher effect applies as usual.

$$
\begin{aligned}
c_t &= \delta E_t c_{t+1} - \sigma \frac{1-\lambda}{1-\lambda\chi}\phi_p p_t + \sigma \frac{1-\lambda}{1-\lambda\chi} E_t \pi_{t+1} \\
\pi_t &= \beta E_t \pi_{t+1} + \kappa c_t \\
p_t &= \pi_t + p_{t-1}
\end{aligned}
$$

Substituting and writing in canonical matrix form $\begin{pmatrix} E_t c_{t+1} & E_t \pi_{t+1} & p_t \end{pmatrix}' = A \begin{pmatrix} c_t & \pi_t & p_{t-1} \end{pmatrix}'$ with transition matrix A given by

$$
A = \begin{pmatrix} \delta^{-1}\left(1+\beta^{-1}\sigma \frac{1-\lambda}{1-\lambda\chi}\kappa\right) & \delta^{-1}\sigma \frac{1-\lambda}{1-\lambda\chi}\left(\phi_p - \beta^{-1}\right) & \delta^{-1}\sigma \frac{1-\lambda}{1-\lambda\chi}\phi_p \\ -\beta^{-1}\kappa & \beta^{-1} & 0 \\ 0 & 1 & 1 \end{pmatrix}.
$$

We can apply Proposition C.2 in Woodford (2003, Appendix C): determinacy requires two roots outside the unit circle and one inside. The characteristic equation of matrix A is:

$$
J(x) = x^3 + A_2 x^2 + A_1 x + A_0 = 0
$$

with coefficients:

$$
\begin{aligned}
A_2 &= -\frac{1}{\beta} - \frac{1}{\delta}\left(\frac{\sigma\kappa}{\beta}\frac{1-\lambda}{1-\lambda\chi} + 1\right) - 1 < 0 \\
A_1 &= \frac{1}{\beta} + \frac{1}{\delta}\left[\frac{\sigma\kappa}{\beta}\frac{1-\lambda}{1-\lambda\chi}\left(1+\phi_p\right) + 1 + \frac{1}{\beta}\right] > 0 \\
A_0 &= -\frac{1}{\beta\delta}
\end{aligned}
$$

To check the determinacy conditions, we first calculate:

$$
\begin{aligned}
J(1) &= 1 + A_2 + A_1 + A_0 = \frac{1}{\delta}\frac{\sigma\kappa}{\beta}\frac{1-\lambda}{1-\lambda\chi}\phi_p > 0 \\
J(-1) &= -1 + A_2 - A_1 + A_0 \\
&= -2 - \frac{2}{\beta} - \frac{1}{\delta}\left[2\frac{\sigma\frac{1-\lambda}{1-\lambda\chi}\kappa}{\beta} + \frac{\sigma\frac{1-\lambda}{1-\lambda\chi}\kappa}{\beta}\phi_p + 2 + \frac{2}{\beta}\right] < 0
\end{aligned}
$$

Since $J(1) > 0$ and $J(-1) < 0$ we are either in case Case II or Case III in Woodford Proposition C.2;

Case III in Woodford implies that $\phi_p > 0$ is sufficient for determinacy if the additional condition is satisfied:

$$
A_2 < -3 \rightarrow \delta < \frac{\sigma\frac{1-\lambda}{1-\lambda\chi}\kappa + \beta}{2\beta - 1}. \tag{56}
$$

This is a fortiori satisfied in RANK (and delivers determinacy there), but not here with $\delta > 1$. Therefore, we also need to check Case II in Woodford and to that end we need to check the additional requirement (C.15) therein:

$$A_0^2 - A_0 A_2 + A_1 - 1 > 0,$$

which replacing the expressions for the A_is delivers:

$$\phi_p > \frac{(1-\beta)(\delta-1) + \sigma\frac{1-\lambda}{1-\lambda\chi}\kappa}{\sigma\frac{1-\lambda}{1-\lambda\chi}\kappa\delta\beta}(1-\delta\beta)$$

Since the ratio is positive, this requirement is only stronger than the already assumed $\phi_p > 0$ when

$$\delta < \beta^{-1}; \tag{57}$$

It can be easily checked that the δ threshold 57 is always smaller than the threshold 56; therefore, whenever $\delta < \beta^{-1}$, Case III applies and $\phi_p > 0$ is sufficient for determinacy. While when 56 fails (for large enough δ), Case II applies and $\phi_p > 0$ is still sufficient for determinacy.

D Optimal Policy in a-HANK

We approximate the economy around an efficient equilibrium, defined as an equilibrium with both flexible prices and perfect insurance; this is the case in our baseline economy under the assumed steady-state fiscal policy, because the optimal subsidy inducing zero profits in steady state implies that consumption shares are equalized across agents. In particular, since the fiscal authority subsidize sales at the constant rate τ^S and redistribute the proceedings in a lump-sum fashion T^S such that in steady-state there is marginal cost pricing, and profits are zero. The profit function becomes $D_t(k) = (1+\tau^S) P_t(k)Y_t(k) - W_t N_t(k) - \frac{\psi}{2}\left(\frac{P_t(k)}{P_{t-1}^{**}} - 1\right)^2 P_t Y_t + T_t^S$ where by balanced budget $T_t^S = \tau^S P_t(k) Y_t(k)$. Efficiency requires $\tau^S = (\varepsilon - 1)^{-1}$, such that under flexible prices $P_t^*(k) = W_t^*$ and hence profits are $D_t^* = 0$ (evidently, with sticky prices profits are not zero as the mark-up is not constant). Under this assumption we have that in steady-state:

$$\frac{U_N(N^H)}{U_C(C^H)} = \frac{U_N(N^S)}{U_C(C^S)} = \frac{W}{P} = 1 = \frac{Y}{N},$$

where $N^j = N = Y$ and $C^j = C = Y$.

Suppose further that the social planner maximizes a convex combination of the utilities of the two types, weighted by the mass of agents of each type: $U_t(.) \equiv \lambda U^H(C_t^H, N_t^H) + [1-\lambda] U^S(C_t^S, N_t^S)$. The second-order approximation to type $j's$ utility around the **efficient flex-price equilibrium** delivers:

$$\begin{aligned}\hat{U}_{j,t} &\equiv U_j\left(C_{j,t},N_{j,t}\right)-U_j\left(C_{j,t}^{*},N_{j,t}^{*}\right)= \\ &= U_C C^j\left[c_t^j+\frac{1-\sigma^{-1}}{2}\left(c_t^j\right)^2\right]-U_N N^j\left[n_t^j+\frac{1+\varphi}{2}\left(n_t^j\right)^2\right]+t.i.p+O\left(\parallel\zeta\parallel^3\right),\end{aligned}\tag{58}$$

where we used that flex-price values are equal to steady-state values (because of our assumption of no shocks to the natural rate) $c_t^{j*}\left(\equiv\log\frac{C_t^{j*}}{C}\right)=c_t^*=0$ and $n_t^{j*}\left(\equiv\log\frac{N_t^{j*}}{N}\right)=n_t^*=0$.

Approximating the goods market clearing condition to second order delivers:

$$\begin{aligned}\lambda C_{H,t}+(1-\lambda)C_{S,t} &\simeq \lambda c_{H,t}+(1-\lambda)c_{S,t}+\frac{1}{2}\left(\lambda c_{H,t}^2+(1-\lambda)c_{S,t}^2\right) \\ &= \lambda N_{H,t}+(1-\lambda)N_{S,t}\simeq\lambda n_{H,t}+(1-\lambda)n_{S,t}+\frac{1}{2}\left(\lambda n_{H,t}^2+(1-\lambda)n_{S,t}^2\right)\end{aligned}$$

The linearly-aggregated first-order term is thus found from this second-order approximation of the economy resource constraint as:

$$\lambda c_{H,t}+(1-\lambda)c_{S,t}-\lambda n_{H,t}-(1-\lambda)n_{S,t}+\frac{1}{2}\left(\lambda c_{H,t}^2+(1-\lambda)c_{S,t}^2-\left(\lambda n_{H,t}^2+(1-\lambda)n_{S,t}^2\right)\right)=0 \tag{59}$$

The economy resource constraint is

$$C_t=\left(1-\frac{\psi}{2}\pi_t^2\right)Y_t=\left(1-\frac{\psi}{2}\pi_t^2\right)N_t$$

which approximated to second order is:

$$c_t=n_t-\frac{\psi\pi}{1-\frac{\psi}{2}\pi^2}\pi_t-\frac{1}{2}\frac{\psi}{1-\frac{\psi}{2}\pi^2}\pi_t^2$$

It is straightforward to show that the optimal long-run inflation target in this economy is, just like in RANK, $\pi=0$. Replacing, we obtain the second-order approximation of the resource constraint around the optimal long-run equilibrium:

$$c_t=n_t-\frac{\psi}{2}\pi_t^2, \tag{60}$$

where the second term captures the welfare cost of inflation.

Note that since $U_C C^j$ and $U_N N^j$ are equal across agents we can aggregate the approximations of individual utilities above (58), using (59) and (60) to eliminate linear terms, into:

$$\begin{aligned}\hat{U}_t &= -U_C C\left\{\frac{\sigma^{-1}}{2}\left[\lambda\left(c_t^H\right)^2+(1-\lambda)\left(c_t^S\right)^2\right]+\frac{\varphi}{2}\left[\lambda\left(n_t^H\right)^2+(1-\lambda)\left(n_t^S\right)^2\right]+\frac{\psi}{2}\pi_t^2\right\} \\ &\quad +t.i.p+O\left(\parallel\zeta\parallel^3\right).\end{aligned}$$

Quadratic terms can be expressed as a function of aggregate consumption (output). Notice that in evaluating these quadratic terms we can use first-order approximations of the optimality conditions (higher order terms imply terms of order $O(\| \zeta \|^3)$). Recall that up to first order, we have that $c_t^H = \chi y_t$ and $c_t^S = \frac{1-\lambda\chi}{1-\lambda} y_t$ and (after straightforward manipulation for hours worked):

$$
\begin{aligned}
n_t^H &= \left(1+\varphi^{-1}\sigma^{-1}(1-\chi)\right) y_t \\
n_t^S &= \left(1+\varphi^{-1}\sigma^{-1}\frac{\lambda}{1-\lambda}(\chi-1)\right) y_t
\end{aligned}
$$

To second order we thus have

$$
\begin{aligned}
\left(c_t^H\right)^2 &= \chi^2 y_t^2 + O\left(\| \zeta \|^3\right) \\
\left(n_t^H\right)^2 &= \left[1+\varphi^{-1}\sigma^{-1}(1-\chi)\right]^2 y_t^2 + O\left(\| \zeta \|^3\right) \\
\left(c_t^S\right)^2 &= \left(\frac{1-\lambda\chi}{1-\lambda}\right)^2 y_t^2 + O\left(\| \zeta \|^3\right) \\
\left(n_t^S\right)^2 &= \left[1+\varphi^{-1}\sigma^{-1}\frac{\lambda}{1-\lambda}(\chi-1)\right]^2 y_t^2 + O\left(\| \zeta \|^3\right)
\end{aligned}
$$

Replacing, the aggregate *per-period* welfare function is thus up to second order, ignoring terms independent of policy and of order larger than 2 and after straightforward algebra to simplify the relative weight on consumption/output stabilization denoted by α:

$$
\alpha \equiv \frac{\sigma^{-1}+\varphi}{\psi}\left[1+\varphi^{-1}\sigma^{-1}\frac{\lambda}{1-\lambda}(\chi-1)^2\right]
$$

we obtain

$$
\hat{U}_t = -\frac{U_C C \psi}{2}\left\{\alpha y_t^2 + \pi_t^2\right\}
$$

D.1 Optimal FG in a Liquidity Trap: A Caveat

A caveat is worth mentioning: when FG is less effective, shouldn't optimal policy imply doing more (rather than less) of it? Nakata, Schmidt, and Yoo (2018) follow this line of reasoning in a calibrated model with a discounted Euler equation that delivers FG power mitigation. They show that, if instead of keeping the size of the disturbance fixed—as we did above—one fixes the *size of the recession* (itself a function of other structural parameters), one obtains the opposite conclusion to this paper's with $\chi < 1$: the optimal duration of FG becomes *increasing* in the share of constrained households. The reason is that, as λ increases, the shock necessary to generate the given recession gets larger and larger, which adds a force calling for more optimal FG. If this force is strong enough, it can overturn the conclusion obtained above for a given shock.

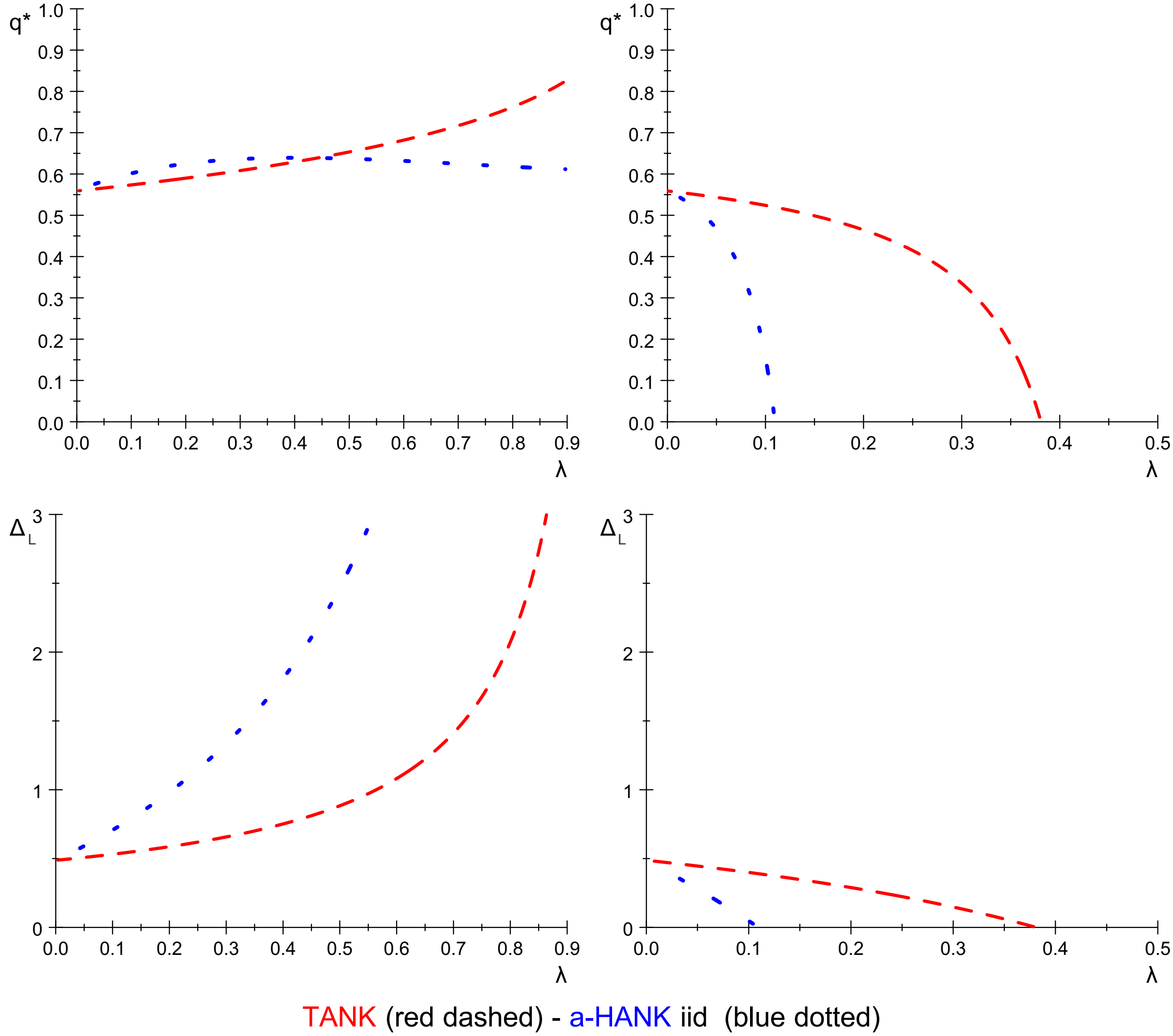


Figure 7 q^* as a function of λ: fixed recession, Δ_L adjusts endogenously

This also holds in my model with procyclical inequality ($\chi < 1$) and little or no idiosyncratic risk, i.e. the TANK model (red dash in the upper left panel in Figure 7): the optimal duration becomes increasing with λ. There is, however, an important qualification as the level of idiosyncratic risk increases: the blue dotted line in the same panel (corresponding to the a-HANK model with the strongest self-insurance motive) is increasing only slightly initially, and decreasing thereafter. The reason is that idiosyncratic risk delivers more dampening overall—so while the shock necessary to reproduce a given recession is increasing in λ at an even faster rate, the power of FG also goes down very fast. The FG puzzle and having optimal FG increase with the share of constrained households are two sides of the same coin: in this simple model at least, you cannot throw one and keep the other.[39]

[39] The other qualification pertaining to this case refers to the implied shock, plotted in the lower left panel. With so much dampening as implied by the a-HANK model, the shock necessary to replicate an even modest recession

Moreover, the very same logic that generates increasing FG duration is now turned on its head in the amplification, $\chi > 1$ case: as λ gets larger, a smaller shock is needed to generate a given recession (lower right panel). This adds a force calling for *less* optimal FG, so the optimal duration is *lower* (and more rapidly decreasing) than in the "fixed-shock" case. And since amplification is so powerful in a-HANK, self-insurance makes optimal FG duration decrease even faster.

The general message is that keeping fixed the observable recession (rather than the unobservable disturbance) is a useful optimal policy exercise but does not necessarily imply a stronger case for longer optimal FG duration. Indeed, in some cases such as the "amplification" case whereby FG power is highest (and the puzzle at its most extreme) it unambiguously implies an even weaker case.

(4 percent here) becomes very large indeed (several times larger than the normal-times interest rate); while the shock is unobservable, this type of configuration seems unlikely.