DISCUSSION PAPER SERIES

DP12231

THE PUZZLE, THE POWER, AND THE DARK SIDE: FORWARD GUIDANCE REDUX

Florin Ovidiu Bilbiie

MONETARY ECONOMICS AND FLUCTUATIONS

ISSN 0265-8003

THE PUZZLE, THE POWER, AND THE DARK SIDE: FORWARD GUIDANCE REDUX

Florin Ovidiu Bilbiie

Discussion Paper DP12231
Published 21 August 2017
Submitted 21 August 2017

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

THE PUZZLE, THE POWER, AND THE DARK SIDE: FORWARD GUIDANCE REDUX

Abstract

Forward guidance (FG) is phenomenally powerful in New Keynesian models---a feature that earned the label "FG puzzle". This paper shows formally how two channels are jointly necessary to reduce the power enough to resolve the puzzle: hand-to-mouth constrained households' income respond to aggregate less than one-to-one; and unconstrained households self-insure idiosyncratic risk. These channels are complementary: if the former condition fails, FG power is instead amplified and the puzzle aggravated. Yet optimal policy does not imply larger FG duration even with such puzzling amplification, because FG power has a dark side: when it increases, so does its welfare cost.

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

Keywords: forward guidance; hand-to-mouth; heterogenous households; aggregate demand; self-insurance; optimal monetary policy; liquidity trap; Keynesian cross.

Florin Ovidiu Bilbiie - florin.bilbiie@gmail.com
Paris School of Economics and CEPR

Acknowledgements
I am grateful to Jess Benhabib, Davide Debortoli, Marco Del Negro, Gaetano Gaballo, Jordi Galí, Alejandro Justiniano, Virgiliu Midrigan, Benjamin Moll, Emi Nakamura, Taisuke Nakata, Xavier Ragot, Jon Steinsson, Gianluca Violante, Mirko Wiederholt, Michael Woodford, and participants in several seminars and conferences for comments and discussions, and Ezgi Ozsogut for research assistance. I gratefully acknowledge without implicating the support of Banque de France via the eponymous Chair at PSE, and of Institut Universitaire de France, as well as the hospitality of New York University and CREi during part of writing this paper.

The Puzzle, the Power, and the Dark Side: Forward Guidance Redux[I]

Florin O. Bilbiie[II]
PSE and CEPR

August 2017 (First preliminary draft April 2016)

Abstract

Forward guidance (FG) is phenomenally powerful in New Keynesian models—a feature that earned the label "FG puzzle". This paper shows formally how two channels are jointly necessary to reduce the power enough to resolve the puzzle: *hand-to-mouth* constrained households' income respond to aggregate *less than one-to-one*; and unconstrained households *self-insure* idiosyncratic risk. These channels are complementary: if the former condition fails, FG power is instead amplified and the puzzle aggravated. Yet optimal policy does not imply larger FG duration even with such puzzling amplification, because FG power has a dark side: when it increases, so does its welfare cost.

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

Keywords: forward guidance; hand-to-mouth; heterogenous households; aggregate demand; self-insurance; optimal monetary policy; liquidity trap; Keynesian cross.

[I] I am grateful to Jess Benhabib, Davide Debortoli, Marco Del Negro, Gaetano Gaballo, Jordi Galí, Alejandro Justiniano, Virgiliu Midrigan, Benjamin Moll, Emi Nakamura, Taisuke Nakata, Xavier Ragot, Jon Steinsson, Gianluca Violante, Mirko Wiederholt, Michael Woodford, and participants in several seminars and conferences for comments and discussions, and Ezgi Ozsogut for research assistance. I gratefully acknowledge without implicating the support of Banque de France via the eponymous Chair at PSE, and of Institut Universitaire de France, as well as the hospitality of New York University and CREi during part of writing this paper.

[II] Paris School of Economics and UP1 Panthéon Sorbonne; 48 Boulevard Jourdan 75014 Paris. florin.bilbiie@parisschoolofeconomics.eu. http://florin.bilbiie.googlepages.com.

1 Introduction

Forward guidance in the standard New Keynesian NK model is nothing short of miraculous: its *power*, defined by the expansion in aggregate activity today induced by a marginal increase in FG duration, is increasing with said duration. More miraculously still, the power increases the further FG is pushed into the future—a feature that Del Negro, Giannoni and Patterson (2012) famously called the *FG puzzle*.[1] If the economy worked thus, getting out of a liquidity trap would be very easy: announce a long enough period of low interest rates, far enough into the future. Real-life experience suggests that this is not a good description of how things work: Japan's liquidity trap duration is measured in decades, and the recent experience of most OECD countries in years. If it were so easy to end a liquidity trap, we would have found out by now. Furthermore, empirical evidence on the effects of FG does not render support to there being such power.[2]

The literature focused on finding model features that imply dampening of FG power. Several such solutions exist by now, and a prominent one, that this paper takes as a starting point, is McKay, Nakamura and Steinsson's (2015, 2016—hereinafter MNS) solution based on incomplete markets and uninsurable risk. This is also confirmed in their different heterogeneous-agent NK (HANK) model by Kaplan, Moll, and Violante (2016—hereinafter KMV). In a nutshell, incomplete markets induce dampening because they imply a form of discounting in the Euler equation: a cut in interest rates far into the future increases demand and income in the future, but has only a modest effect on current demand when it is discounted, because of self-insurance—households save part of the good income news.[3] A crucial element of MNS' framework is that the income of constrained households covaries with aggregate income *less than one-to-one* in equilibrium: either because of an underlying fiscal redistribution (more below), or because—if the constrained are unemployed receiving benefits or home production, as in MNS' 2016 paper—their income is simply exogenous (hence, the elasticity to aggregate income is zero).

I first show that a model where the income of constrained covaries with aggregate income *more than one-to-one* has the *opposite* prediction: FG power is amplified, rather than dampened, and the FG puzzle is aggravated, rather than solved. There are two parts to the

[1] See Carlstrom, Fuerst, and Paustian (2015), Cochrane (2015) and Kiley (2015) for further discussions of this and other puzzling implications.

[2] See Campbell et al (2016), Del Negro et al (2012); Gurkaynak et al (2005) were among the first to study the effect of monetary policy announcements. See also Filardo and Hoffman (2012) for an account of FG experiences in various countries.

[3] Other solutions include: OLG perpetual-youth models (Del Negro et al, 2012; see Piergallini 2006 and Nistico, 2010); reflective equilibrium (Garcia-Schmidt and Woodford 2014; Farhi and Werning 2017); other behavioural assumptions (Gabaix, 2016), heterogeneous beliefs (Andrade et al, 2015; Bilbiie, 2016), dispersed information (Wiederholt, 2015).

mechanism, and they both rely on a "New Keynesian cross" logic.

The first component, that I will refer to as the *hand-to-mouth channel,* generates amplification of monetary policy changes at all horizons, uniformly, because it implies an elasticity of aggregate demand to interest rates within the period that is increasing with the share of hand-to-mouth households. This mechanism, first unveiled and discussed in Bilbiie (2008) but also operating the earlier model of hand-to-mouth with sticky prices of Gali, Lopez-Salido and Valles (2007), works as follows. An interest rate cut makes savers want to bring consumption forward by standard intertemporal substitution, which with sticky prices has an expansionary effect on labor demand bidding up the real wage. But the hand-to-mouth consume all their income, and this income is essentially the real wage; if the fiscal redistribution and labor market parameters are such that hand-to-mouth's income is related to aggregate income more than one-to-one, this amplifies the initial demand increase further, which bids up the real wage further, and so on.

The second component, that I will refer to as the *self-insurance channel*, is that when the above feature coexists with idiosyncratic risk it further generates amplification of news—and forward guidance *is* news. When a household receives good news about future aggregate income, *and* its own income covaries with the aggregate more than one-to-one in states where it becomes liquidity-constrained, it has an incentive to dis-save for precautionary self-insurance reasons, i.e. consume more. This overturns the "discounting" in the aggregate Euler equation that MNS unveiled as the key mechanism shaping the power of FG in their model.[4] There is now *compounding* in the aggregate Euler equation, and FG becomes one order of magnitude more powerful. The FG puzzle is then, of course, much aggravated.

Furthermore, the two channels (hand-to-mouth, and self-insurance) are *complementary* in generating amplification, or dampening;[5] I show formally that, because of this complementarity, *both* conditions are needed to solve the FG puzzle: some risk inducing self-insurance, and income respond sufficiently little to aggregate. In other words, if either of the conditions does not hold the FG puzzle is alive and well; and worse still, if the former condition holds but the latter does not, we are in an amplification where the complementarity also applies: FG power is *magnified*, and the FG puzzle *aggravated*.

Yet despite this terrific amplification, I end on a more positive and constructive note for

[4] This second channel and its interaction with the first in shaping a "New Keynesian cross" are discussed in Bilbiie (2017) as capturing two key mechanisms of richer, more general HANK models such as KMV and MNS. Werning (2015) derives results on amplification in a more general framework with cyclical income risk and liquidity.

[5] A different complementarity—between incomplete markets and k-level thinking—is discussed by Farhi and Werning (2017); there, differently from the mechanism I emphasize, both channels imply *mitigation* of FG power, and the addition of the two even more so. Here, each channel by itself leads to *amplification* of FG power, but the addition of the two leads to *dampening.* See the discussion in text.

NK models: optimal policy does *not* imply longer FG, even (and in fact especially) when the implied FG power is miraculous; indeed, I show that the optimal policy in a liquidity trap implies an FG duration that is little affected by (and it often is in fact decreasing with) the share of constrained households. The reason is what I call the *dark side* of FG: when its power increases, so does its *welfare cost*—the inefficient consumption volatility once the trap is over (first emphasized in a representative-agent model in Eggertsson and Woodford's seminal 2003 paper). This sharp increase in the welfare cost whenever there is amplification (the dark side) precludes a welfare-maximizing central bank from using more FG even when FG is puzzlingly powerful.

2 A Simple HANK Model for Forward Guidance Analysis

Based on Bilbiie (2017), I outline a simplified HANK model, isolating two of the channels operating in richer quantitative HANK models such as the mentioned KMV and MNS papers.[6] The model nests two simpler models existing in the literature: an earlier two-agent NK (TANK) model discussed in Bilbiie (2008, 2017) and Gali, Lopez-Salido and Valles (2007) and the "discounted Euler equation" model that MNS (2016) proposed as a simplified version of their important 2015 paper. This simplified model with incomplete markets and idiosyncratic (unemployment) risk implies a form of "discounting" in the aggregate Euler equation and can be solved analytically. [7]

Here, I outline a simple HANK model that builds on those contributions and uses an "infrequent participation" structure similar to Bilbiie and Ragot (2016) but, as in the other papers cited in the previous footnote, with no trade (no equilibrium liquidity)—even though it distinguishes, like the HANK model, between liquid assets (bonds) and illiquid assets (stock). In equilibrium, there is thus infrequent (limited) participation in the *stock* market.

There are two states: savers S and hand-to-mouth H. The source of idiosyncratic risk is that agents switch states following a Markov chain. The probability to **stay** type S is s and the probability to stay type H is h (while the transition probabilities are respectively $1 - s$

[6] For other quantitative HANK models containing these mechanisms among others, see i.a. Auclert (2016) and Gornemann, Kuester, and Nakajima (2015).

[7] To achieve this, the authors use a set of assumptions first used by Krusell Mukoyama Smith (2011) for asset pricing and Ravn and Sterk (2012) in the context of a New Keynesian model with incomplete markets and endogenous unemployment risk. Werning (2015) uses a similar assumption to derive an aggregate Euler equation under a more general risk structure. Curdia and Woodford (2009) and Nistico (2016) are two other examples of the use of the "infrequent participation" device (introduced by Lucas, 1990) in models with nominal rigidities, albeit in a different context and for different questions. Bilbiie and Ragot (2016) build a model with three assets—of which one ("money") is liquid and is traded in equilibrium while the others are accessed only infrequently—and study optimal monetary policy in that framework.

and $1-h$), and by standard results the mass of H is:

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

with the stability condition $s \geq 1-h$. This condition has an intuitive interpretation: it implies that the probability of remaining a participant/saver is higher than the probability of becoming a participant. By consequence and given the definition of λ we also have:

$$1-s \leq \lambda;$$

the conditional probability of becoming hand-to-mouth is lower than the unconditional probability (the share of hand-to-mouth).

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

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

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

I therefore assume that in the H state the equivalent Euler equation holds with strict inequality: households are constrained, or impatient, and become hand-to-mouth thus consuming all their income $C_t^H = Y_t^H$.[8] Loglinearizing around a symmetric steady state $C^H = C^S$ (which is instead achieved by assuming a steady-state redistribution scheme), the self-insurance equation is:

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

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

[8] One justification for this could be that the idisyncratic shock is a preference shock to β rendering households impatient "enough" to make the constraint bind.

The other key ingredient is that all agents make an optimal labor supply decision: their income is therefore labor income, plus any fiscal transfer. Because they are standard, I present the loglinearized equilibrium conditions directly; in particular, I approximate around a "full-insurance" steady-state whereby an optimal sales subsidy induces marginal-cost pricing and is financed by taxing firms (and thus, implicitly, savers). This further generates zero steady-state profits and, hence, full insurance—a convenient but largely innocuous assumption for the results derived here. Both types' labor supply decision $j = S, H$ is governed by: $\varphi n_t^j = w_t - \sigma^{-1} c_t^j$, with σ^{-1} relative risk aversion, φ the inverse elasticity of labor supply, and n are hours worked, w the real wage, and c consumption, with everything expressed in percentage deviations of steady-state aggregates. Assuming that elasticities are identical across agents, the same holds on aggregate: $\varphi n_t = w_t - \sigma^{-1} c_t$. All output is consumed and produced using (only) labor with constant returns $c_t = y_t = n_t$, which implies $w_t = (\varphi + \sigma^{-1}) c_t$. Finally, around the steady state with zero profits,[9] we have $d_t = -w_t$: profits vary inversely with the real wage (d_t is expressed as a share of steady-state output).

Let us now add one example of fiscal redistribution, following Bilbiie (2008, 2017): taxing profits at rate τ^D and rebating the proceedings in a lump-sum fashion to hand-to-mouth agents.[10] The model being otherwise unchanged, we have in loglinearized form that per-capita transfers to hand-to-mouth are $t_t^H = \frac{\tau^D}{\lambda} d_t$. H agents' income is then $y_t^H = w_t + n_t^H + \frac{\tau^D}{\lambda} d_t$; while for S agents, it is $y_t^S = w_t + n_t^S + \frac{1-\tau^D}{1-\lambda} d_t$. This shows that savers S face an extra income effect of the real wage (which for them counts as marginal cost and reduces profits) that is the keystone for monetary transmission.

Hand-to-mouth thus consume all *their* income $c_t^H = y_t^H$, the key word being "*their*": for while their consumption comoves one-to-one with their income, it comoves *more or less than one-to-one* with *aggregate* income. In particular, combining the labor supply and budget constraint of H agents with the other equations, we immediately obtain the consumption function:

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

denotes the *elasticity of H's income to aggregate income,* which depends on redistribution; this captures intuitively the point that the marginal propensity to consume depends on income distribution—a point that goes back to Keynes (1936). But this dependence holds

[9] This is not strictly necessary for any of the results but it simplifies the algebra.

[10] See Section 4.3 and Proposition 3 in Bilbiie (2008) for a first analysis of the link between the redistribution of profits and Keynesian logic. See the Appendix in Bilbiie (2017) for a more general redistribution scheme with similar equilibrium implications.

more generally for arbitrary fiscal redistribution schemes. As will become clear momentarily, this parameter is key for the effects of forward guidance in this model.[11] It is lower than $1+\varphi$ but higher than 1 inasmuch as $\tau^D < \lambda$ (in other words, if there is not too much redistribution). When $\tau^D = \lambda$, all the endogenous redistributive effects emphasized here are undone, and the economy is back to the perfect-insurance, representative-agent case. Finally, when H get a share of profits higher than their share in the population $\tau^D > \lambda$ (an example of progressive taxation) we have $\chi < 1$—indeed, in the limit when $\chi = 0$ the income of H is exogenous: this happens for example if they are unemployed and benefits (or home production) do not depend on the cycle, as in MNS (2016).

The mirror image is of course what happens to savers' consumption, which follows immediately as:

$$c_t^S = \frac{1-\lambda\chi}{1-\lambda} y_t. \tag{2}$$

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

Replacing in the loglinearized Euler equation, we obtain the aggregate Euler-IS curve:

$$\begin{aligned} c_t &= \delta E_t c_{t+1} - \eta \left(i_t - E_t \pi_{t+1} - \rho_t\right), \qquad (3) \\ \text{where } \delta &\equiv \frac{s + (1-\lambda-s)\chi}{1-\lambda\chi} \text{ and } \eta \equiv \sigma \frac{1-\lambda}{1-\lambda\chi} \end{aligned}$$

In the companion paper Bilbiie (2017) I analyze the properties of this model of aggregate demand for dynamics and transmission of monetary policy shocks, focusing on a "New Keynesian cross" interpretation centered around a planned expenditure curve, or consumption function; In this paper, I study the implications for forward guidance in a liquidity trap, including its optimal design. Some of the insights discussed in the companion paper naturally apply in this context too, with important nuances emphasized below; I postpone this discussion to solving for the liquidity trap equilibrium.

[11] This distinguishes my model (and its earlier 2008 incarnations) from earlier analyses such as Campbell and Mankiw (1989) and the literature that followed it—where it is assumed that hand-to-mouth (or, in their terminology, "rule-of-thumb") agents consume a fraction (<1) of *aggregate* income; the implications of this are discussed in more detail below.

To complete the model, add a standard aggregate supply side, a "New Keynesian Phillips curve" coming from a forward-looking pricing model à la Calvo-Yun or Rotemberg model:

$$\pi_t = \beta E_t \pi_{t+1} + \kappa c_t. \tag{4}$$

with slope $\kappa = \psi (\varphi + \sigma^{-1})$, where $\psi = (1 - \zeta)(1 - \beta\zeta)/\zeta$ with ζ the probability to be unable to change one's price in the Calvo-Yun model.[12] The specification of monetary policy defined as a choice of i_t follows below.

3 The Power of Forward Guidance Redux

Consider first the properties of a **liquidity trap** in this economy, where the zero lower bound is triggered as in the seminal paper of Eggertsson and Woodford (2003): ρ_t follows a Markov chain with two states. The first is the steady state denoted by S, with $\rho_t = \rho$, and is absorbing: once in it, there is a probability of 1 of staying. The other state is transitory and denoted by L: $\rho_t = \rho_L < 0$ with persistence probability p (conditional upon starting in state L, the probability that $\rho_t = \rho_L$ is p, while the probability that $\rho_t = \rho$ is $1 - p$). At time t, there is a negative realization of $\rho_t = \rho_L < 0$, meant to capture in this reduced-form model an increase in spreads as in Woodford (2011) and Curdia and Woodford (2010). I assume for further simplification that the monetary authority tracks the natural interest rate of this economy ($r_t^n = \rho_t$) whenever feasible, meaning $i_t = \max(\rho_t, 0)$. It follows that the ZLB will bind when $\rho_t = \rho_L < 0$, while the flexible-price efficient equilibrium will be achieved whenever $\rho_t = \rho$.

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

$$c_L = \frac{\eta}{1 - p\left(\delta + \eta\frac{\kappa}{1-\beta p}\right)}\rho_L; \tag{5}$$

where $p\left(\delta + \eta\frac{\kappa}{1-\beta p}\right) < 1$ is needed to rule out expectations-driven liquidity traps.[13] Inflation is $\pi_L = \frac{\kappa}{1-\beta p} c_L$.

Why an increase in the desire to save generates a recession with a binding zero lower

[12] The slope of the Phillips curve is independent of the demand parameters λ and s because of the simplifying assumption of perfect consumption insurance in steady-state—see Bilbiie (2008) for elaboration in a more general case.

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

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

In the simple HANK model, the magnitude of the liquidity trap recession depends on the three parameters s, λ, and χ that are key for transmission more generally. In particular, (5) elucidates how this will be different from the representative-agent model through three channels: the denominator, η (this holds even for transitory shocks $p = 0$ and for fixed prices $\kappa = 0$); and—for persistent shocks $p > 0$ only—the rate of discounting in the Euler equation δ (even with fixed prices) and again η but through its interaction with the Phillips curve slope. It is worth spending some time now understanding each channel.[14] The main distinction is between two cases, according to whether *H's* income elasticity to aggregate income of is lower or larger than one.

Case 1 ***Amplification:*** $\chi > 1$. *There are two sides to amplification (understood as an LT recession that is deeper than in the RANK model). The first is the "**hand-to-mouth**" channel operating through* η, *with* $\partial\eta/\partial\lambda = (\chi - 1)\,\sigma/\left(1 - \lambda\chi\right)^2 > 0$. *The second is the "**self-insurance**" channel, through which there is **compounding** in the aggregate Euler equation (*$\delta > 1$*); the compounding effect is magnified by idiosyncratic risk* $\partial\delta/\partial\left(1 - s\right) = \left(\chi - 1\right)/\left(1 - \lambda\chi\right) > 0$ *and by the share of* H $\partial\delta/\partial\lambda = \left(\chi - 1\right)\left(1 - s\right)\chi/\left(1 - \lambda\chi\right)^2$.

Case 2 ***Dampening:*** $\chi < 1$. *The "**hand-to-mouth**" channel operating through* η *implies dampening as* $\partial\eta/\partial\lambda < 0$. *The "**self-insurance**" channel implies **discounting** in the aggregate Euler equation (*$\delta < 1$*); the discounting effect is magnified by idiosyncratic risk* $\partial\delta/\partial\left(1 - s\right) < 0$ *and by share of* H $\partial\delta/\partial\lambda < 0$.

Let us briefly discuss the intuition for the "amplification" case—the "dampening" case being merely the mirror image. Take first the hand-to-mouth channel operating through η: the aggregate elasticity of intertemporal substitution—the elasticity of aggregate demand to interest rates *within the period*—is *increasing* with the share of H agents, as long as $\lambda < \chi^{-1}$. The reason is the "New Keynesian cross" already emphasized above and analyzed in detail in Bilbiie (2017): a fall in the natural interest rate implies an aggregate demand contraction, through intertemporal substitution of S agents; with sticky prices, this translates into a labor demand contraction, which compresses the real wage. Since the wage is the H agents'

[14] Some of the intuition here echoes some of the results for moetary policy transmission in Bilbiie (2017). A different channel is emphasized by Ravn and Sterk (2016) in an analytical HANK-type model abstracting from the channels emphasized here but with limited insurance of endogenous unemployment risk.

income, this reduces their demand further, which magnifies the initial demand contraction.[15] 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 same channel also amplifies fiscal multipliers in a liquidity trap in their framework. 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).

Second, the self-insurance channel operating through δ. The endogenous amplification through the Keynesian cross now holds not only contemporaneously, but also for the future—insofar as the liquidity trap is expected to persist: bad news about future aggregate income reduce today's demand because they imply *more* need for self-insurance, precautionary saving. Since future consumption in states where the constraint binds over-reacts to bad "aggregate news", households internalize this by attempting to self-insure *more*. But since precautionary saving needs to be zero in equilibrium, households consume less and income falls to deliver this, thus magnifying the recession even further.[16]

What is more, these two channels are in fact complementary—a property worth emphasizing in the following Proposition.

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

The reason for this complementarity is that the higher the risk (higher $1-s$), the stronger the self-insurance motive: the highest compounding is obtained in the iid case (at given χ).

Below, we will focus on two extreme special cases that are useful for intuition and already present in the literature. The first is the **TANK** limit with permanent idiosyncratic shocks $s = h = 1$: in that case, there is no discounting $\delta = 1$, and λ is then an arbitrary free parameter. The hand-to-mouth channel is the only one operating. The other extreme is

[15] This is an equilibrium: S households face a positive income effect from profits (recal that the real wage is marginal cost). This amplification mechanism is analyzed for the first time in Bilbiie (2004, 2008) for monetary policy shocks in normal times; it also holds in GLV's (2004, 2007) framework but is somewhat hidden because convoluted with physical capital, which in itself affects monetary transmission non-trivially (Dupor, 2001) and because the model then needs to be solved numerically. As λ increases the amplification gets larger and larger: when $\lambda > \chi^{-1}$ an expansion cannot be an equilibrium any longer, as the income effect on S agents starts dominating. That "non-Keynesian" region, whereby interest rate cuts are contractionary, is analyzed in detail in Bilbiie (2008); here we concentrate on the standard, Keynesian region throughout.

[16] A related amplification channel (for interest rate shocks) also holds and is discussed in a more general framework in Werning (2015), through the interaction of countercyclical income risk and procyclical liquidity.

that **iid idiosyncratic uncertainty** with $s = 1 - h$ studied by Krusell Mukoyama Smith (2011), and used by McKay, Nakamura and Steinsson (2015, 2016) and others to study monetary policy and forward guidance. In that case, we have $\lambda = 1 - s$ and $\delta = \frac{1-\lambda}{1-\lambda\chi}$. In light of the previous Proposition, this is the case that delivers the highest amplification (or dampening) because it implies the highest level of idiosyncratic risk that satisfies the restrictions. Notice that strictly speaking, the case studied by MNS (2016) is that of $\chi = 0$, which delivers most discounting.

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

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).

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

3.1 Forward guidance

I model FG stochastically through a Markov chain, as a state of the world with a probability distribution, as follows.[17] Recall that the (stochastic) expected duration of the LT is $T_L = (1-p)^{-1}$, 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, p, and from L to F it is $(1-p)\,q$. The persistence of state F is q, and the probability to move back to steady state from F is hence $1 - q$.

[17] This was introduced by Bilbiie (2016) in a representative-agent model; I refer the reader to that paper for details, robustness, and for an assessment of this optimality concept relative to Ramsey policy. The modelling of FG as a "state" is inspired by Woodford's (2011) modelling of government spending stimulus after a liquidity trap probabilistically (in contrast to that paper, this concentrates on FG, solves for optimal duration, and adds heterogeneous households).

Under this stochastic structure, expectations are determined by:

$$E_t c_{t+1} = p c_L + (1-p)\, q c_F \tag{6}$$

and similarly for inflation. Evaluating the aggregate Euler-IS (3) and Phillips (4) 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:

$$\begin{aligned} c_F &= \frac{\eta}{1-q\left(\delta+\eta\frac{\kappa}{1-\beta q}\right)}\rho; \\ c_L &= \frac{(1-p)\, q\left(\delta+\eta\frac{\kappa}{(1-\beta q)(1-\beta p)}\right)\eta}{\left[1-q\left(\delta+\eta\frac{\kappa}{1-\beta q}\right)\right]\left[1-p\left(\delta+\eta\frac{\kappa}{1-\beta p}\right)\right]}\rho+\frac{\eta}{1-p\left(\delta+\eta\frac{\kappa}{1-\beta p}\right)}\rho_L, \end{aligned} \tag{7}$$

and $\pi_F = \frac{\kappa}{1-\beta q} c_F$, $\pi_L = \frac{\kappa\beta(1-p)q}{(1-\beta q)(1-\beta p)} c_F + \frac{\kappa}{1-\beta p} c_L$.

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

Figure 1 illustrates these findings and the next section derives analytical results useful to understand the mechanisms. Distinguishing between dampening $\chi < 1$ (left) and amplification $\chi > 1$ (right), it plots in both panels consumption in the liquidity trap (thick) and in the FG state (thin), as a function of the FG probability q.[18] We represent the RANK model by green solid lines, the TANK limit of the SHANK model ($s = h = 1$) with red dashed lines, and the SHANK model in the iid case $1-s = h = \lambda$ with blue dots.

The pictures illustrate the dampening and respectively amplification at work: at given q, low future rates have a lower effect (on both c_F and c_L) in the TANK model, and an even lower one in the SHANK model, in the *dampening* case. The last point illustrates the complementarity: the dampening is magnified when moving towards higher risk $1-s$ and in the limit when $1-s = h = \lambda$ (blue dots) we have the fastest discounting. Whereas in the amplification case (right panel), the opposite is true: low rates have a higher effect in the TANK model, and through complementarity an even higher one under self-insurance: the pictured iid case represents the highest compounding. Indeed, even though $\chi = 2$ is a rather conservative number and the share of H is very small (0.1)—which makes amplification in

[18] The illustrative parametrization used in the Figure has $\beta = 0.99, \psi = 0.01, \sigma = 1, \varphi = 1, p = 0.8$ and a spread shock of $\rho_L = -0.005$, i.e. 2 percent per annum. This delivers a recession of 4 percent and annualized inflation of 1 percent in the absence of FG ($q = 0$). The domain is such that $\Gamma_q > 0$.

the TANK version rather limited—amplification in the SHANK model is *phenomenal*: the recession is 10 times (!) larger than in the RANK model (the scale on the right panel had to be changed to accommodate that).

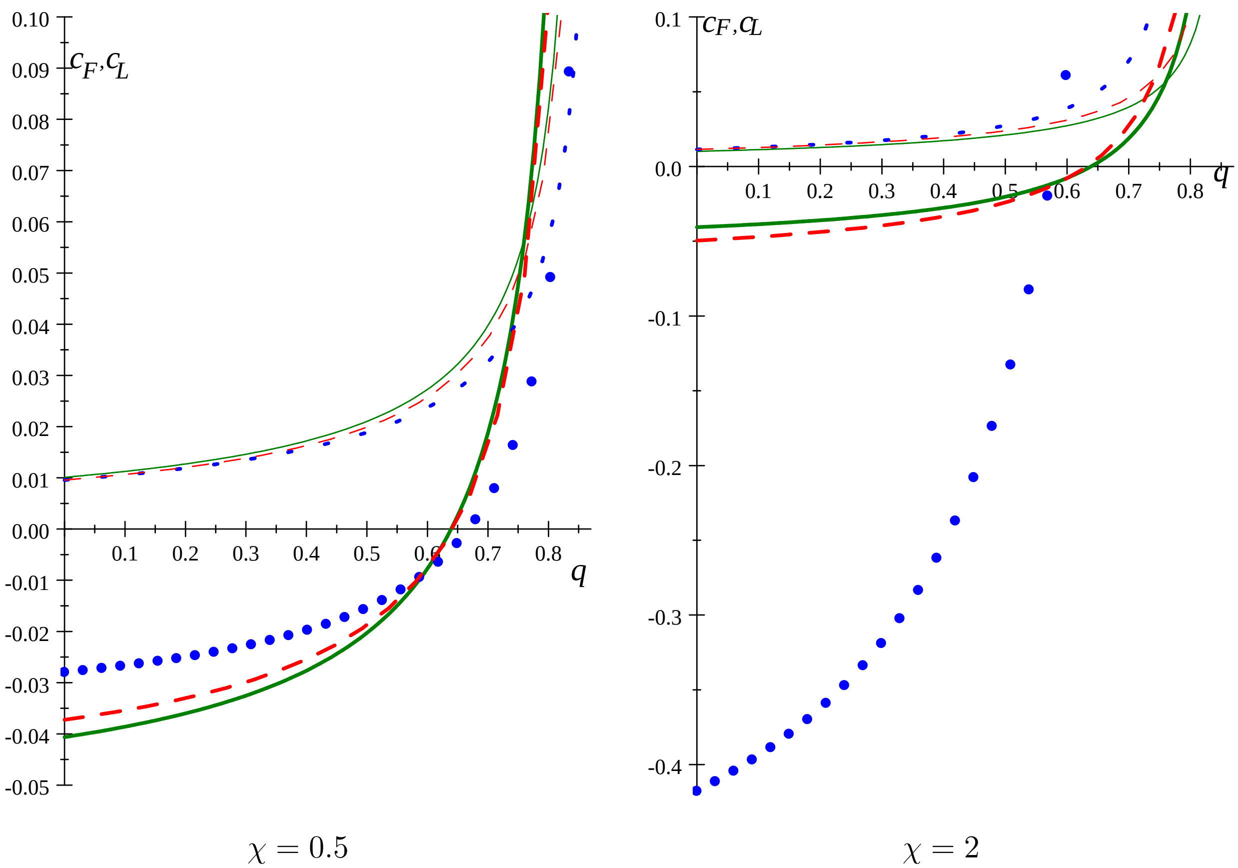


Figure 1: c_L (thick) and c_F (thin) in RANK (green solid), TANK (red dashed) and SHANK-iid (blue dots)

Furthermore, **FG power** (defined as the derivative of consumption during the trap c_L with respect to q, dc_L/dq) is much larger in the SHANK model in the "amplification" case. This is related to the aforementioned "FG puzzle": the higher the persistence of the trap p (the further into the future FG starts), the higher the power dc_L/dq. Our results suggest that while a HANK model with $\chi = 0$ (or $\chi < 1$), as shown by MNS, solves this puzzle, the version with amplification $\chi > 1$ aggravates it. The next section substantiates this point analytically.

3.2 Simple Analytics of FG Power and Puzzle

Closed-form results are particularly useful here in order to shed light on the role of each amplification channel in determining FG power and potential resolutions of the "puzzle"; later, they are also useful for understanding the properties of optimal policy. Consider thus a special case: the simplest possible *aggregate supply* curve whereby each period a fraction of firms ζ keep their price fixed, while the rest can re-optimize their price freely *but* ignoring that this price affects future demand. This delivers the Phillips curve: $\pi_t = \kappa c_t$, where now $\kappa = (\varphi + \sigma^{-1})(1-\zeta)/\zeta$. In Bilbiie (2016) I show that, in the context of a representative-agent NK model, this special case delivers conclusions for FG and optimal policy in a LT that are very close to those obtained using the more general NKPC (4); the insight being that FG is chiefly about aggregate demand, and not so much about supply.[19]

The equilibrium found in (7) above simplifies to:

$$\begin{aligned} c_F &= \frac{\eta}{1-q\nu}\rho; \\ c_L &= \frac{1-p}{1-p\nu}\frac{q\nu}{1-q\nu}\eta\rho + \frac{\eta\rho_L}{1-p\nu}; \end{aligned} \tag{8}$$

I defined the composite parameter:

$$\nu \equiv \delta + \eta\kappa,$$

which has an intuitive interpretation: it captures the response of consumption in a liquidity trap to news about future income/consumption.[20]

We can now define **FG power** formally as:

$$\mathcal{P}_{FG} \equiv \frac{dc_L}{dq} = \left(\frac{1}{1-q\nu}\right)^2 \frac{(1-p)\,\nu\eta}{1-p\nu}\rho.$$

The properties of amplification and dampening of FG power follow the same logic as those applying to the natural rate shock and LT recessions. They are worth emphasizing in the following Corollary (which follows directly by noticing that $\mathcal{P}_{FG}$ is increasing with ν, and hence with both δ and η):

[19] Essentially, such a setup reduces to assuming $\beta = 0$ *only* in the firms' problem (they do not recognize that today's reset price prevails with some probability in future periods). See Bilbiie (2016) for an extension to the case $\beta > 0$, which makes the algebra more involved without affecting the results. In particular, Figure 2 in that paper shows that optimal FG varies very little with the discount factor of firms.

[20] In a "regular" equilibrium whereby the zero bound does not bind and monetary policy follows an active interest rate rule, this parameter is obviously less than one (because news about future income bring about higher real rates, through higher inflation); in particular, $\nu = 1 - \sigma\kappa(\phi - 1)$, where $\phi > 1$ is the response of nominal interest rates to expected inflation.

Corollary 3 *In the amplification case $\chi > 1$ FG power $\mathcal{P}_{FG}$ increases with idiosyncratic risk $1-s$ and with the share of hand-to-mouth λ (while it decreases in the dampening case $\chi < 1$). Furthermore, the* ***complementarity*** *between self-insurance and hand-to-mouth also applies to FG power.*

Defining formally the **FG puzzle** as the property (thus labelled by Del Negro, Giannoni and Patterson, 2012) that the power of FG increases, the further it is pushed into the future (i.e., in our context, with the persistence of the trap p) we obtain the following Proposition:

Proposition 2 *The simple HANK model* ***solves*** *the FG puzzle ($\frac{d\mathcal{P}_{FG}}{dp} < 0$) if and only if:*

$$\nu < 1.$$

Equivalently, the model needs both the self-insurance and hand-to-mouth channels:

$$\begin{aligned} 1-s &> 0 \textit{ and} \\ \chi &< 1-\sigma\kappa\frac{1-\lambda}{1-s} < 1, \end{aligned}$$

which is a manifestation of complementarity.

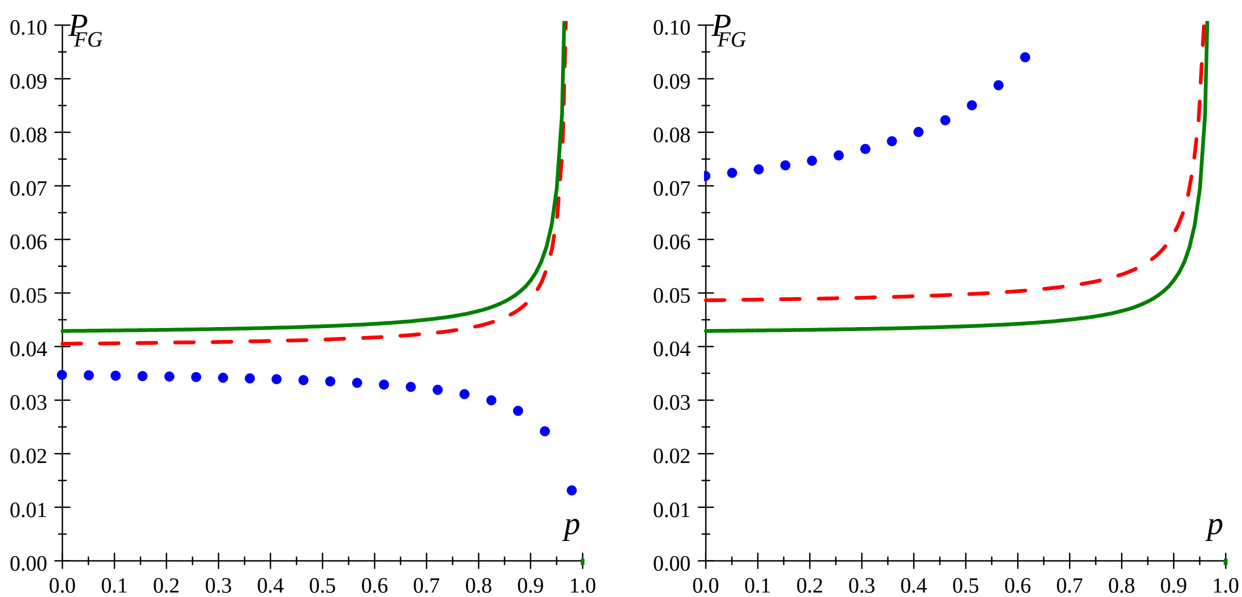
The result follows directly calculating the derivative $d\mathcal{P}_{FG}/dp = (\nu-1)\frac{\nu\eta\rho}{(1-q\nu)^2(1-p\nu)^2}$ and then replacing the expression for ν. The Proposition emphasizes that, to solve the FG puzzle, the model needs *two conditions:* some idiosyncratic uncertainty $1-s>0$, and a cyclicality of H's income that is lower than the threshold defined above. This is a clear illustration of the *complementarity* emphasized in Proposition 1. In other words, having discounting in the aggregate Euler equation ($\delta < 1$) is a necessary, but not sufficient condition to solve the puzzle.[21] Rewriting the condition, we have $\sigma\kappa < \frac{(1-s)(1-\chi)}{1-\lambda}$: this is more stringent when prices are more flexible (κ larger) and λ smaller (at given s and χ).

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

[21] Farhi and Werning (2017) emphasize a related but different complementarity, between market incompletenss and "k-level thinking": an informational imperfection related to Garcia-Schmidt and Woodford's notion of reflective equilibrium, that leads to mitigation of FG—also through discounting in the Euler equation. In their framework, market incompleteness magnifies the mitigation of FG effects obtained with k-level thinking. The complementarity I emphasize is between two different channels, and can work both ways—generating more mitigation, or more amplification. Indeed, it affects not only the quantitative properties, but the qualitative insights: it changes the sign of a key derivative, as illustrated in Figure 2 below, needed to solve the FG puzzle as defined formally here.

needs to be enough idiosyncratic risk, namely in this case $1 - s > (1 - \lambda)\,\sigma\kappa$. It is worth noticing that MNS's 2016 simple model (with iid idiosyncratic risk and exogenous income of H) inherently satisfies these conditions: essentially, to $\chi = 0$ it adds $s = 1 - \lambda$.

To further illustrate how the FG puzzle operates, and how the complementarity between the two channels helps eliminate it, consider Figure 2; it plots FG power as a function of p, for the same calibration as before (fixing in addition $q = 0.5$) in the two cases $\chi < 1$ and $\chi > 1$ for the three models RANK, TANK, and iid SHANK. It illustrates clearly that it is the interaction of dampening through $\chi < 1$ and idiosyncratic risk (which, as shown above, magnifies that dampening through discounting) that leads to resolving the FG puzzle: the power of FG becomes decreasing in the duration of the trap. The dampening channel by itself (TANK model with $\chi < 1$, red dashed line on the left panel) is not enough—although it alleviates the puzzle relative to the RANK model, it does not make the power decrease with the horizon p. While the self-insurance channel by itself added to the "amplification" case magnifies power even further, thus aggravating the puzzle (blue dots in the right panel for the SHANK iid model in the amplification case).



$\chi < 1$ $\qquad\qquad$ $\chi > 1$

Figure 2: FG power in RANK (green solid), TANK (red dashed) and SHANK-iid (blue dots)

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

Since the power of FG is increased (and the puzzle aggravated) in the latter case, does it follow that more FG is always a good thing? That is, does this strengthen the welfare scope for FG as a feature of optimal policy emphasized since the celebrated paper of Eggertsson and Woodford (2003)? No, because there is a dark side to FG power: whenever it is high, the welfare costs of FG, as measured by future volatility, are also high. The following optimal policy exercise formalizes this intuition.

4 The Dark Side of FG Power and Optimal Policy

Optimal monetary policy in a liquidity trap is very 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 the representative-agent model, and turns out to also be 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 an aggregate welfare function that can be represented as a quadratic loss function and, given the Markov chain structure, it is of the form:[22]

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

where $\omega(q)$ is the appropriate discount factor for the FG state.[23] 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 (7) above. The first-order condition of this problem is:

$$c_L \frac{dc_L}{dq} + \omega(q) c_F \frac{dc_F}{dq} + \frac{1}{2} \frac{d\omega(q)}{dq} c_F^2 = 0 \tag{9}$$

and has a clear intuitive interpretation.

The first term is the welfare *benefit* of more forward guidance, through remedying the

[22] Bilbiie (2008, 2017) derives this welfare function under certain conditions that are fulfilled here (an optimal subsidy makes the steady-state efficient). Then, since the equilibrium solution is time-invariant in each of the three states, the per-period loss function is, for any state $j = \{L, F, S\}$: $\pi_j^2 + \chi c_j^2 = (\chi + \kappa^2) c_j^2$. Recall that in state S the economy is back to steady state, so the loss there is zero.

[23] The optimal weight is $\omega(q) = \frac{1-\beta p+\beta(1-p)q}{1-\beta q}$ 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 $\beta(1-p) q / ((1-\beta p)(1-\beta q))$. $\omega(q)$ 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).

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).

Figure 3 plots the optimal FG duration (the solution of equation (9)) 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 can be shown analytically in the simplified version of the model—see next Section. 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 the TANK model, and in addition because of the discounting effect of MNS, in the idiosyncratic risk case.

The amplification case is, in view of our previous results, more surprising: the optimal degree of FG is almost invariant to λ in the TANK case (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 (9)). 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. With idiosyncratic risk, these affects are further amplified: 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 share of H) and makes optimal FG drop abruptly towards zero. It is this sharp increase in the welfare cost that occurs precisely when FG power is large that I refer to as the "dark side" of FG power.

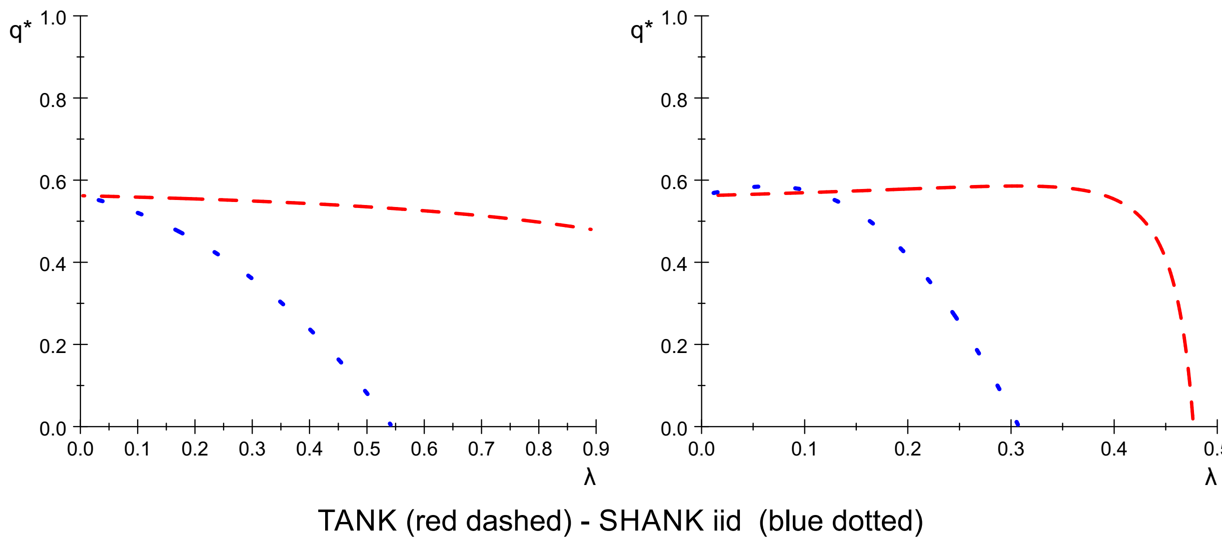


Figure 3 Optimal FG persistence as a function of λ for $\chi < 1$ (left) and $\chi > 1$ (right)

In both models, it becomes optimal to do no FG at all beyond a certain threshold share of hand-to-mouth. The underlying reason is, however, very different. With dampening it is because (as MNS discovered), a higher share of H 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 share implies a *large* elasticity of aggregate demand 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.

4.1 Simple Analytics of Optimal FG

Further insights can be obtained by focusing again on the simpler case studied in Section 3.2 and assuming in addition that the central bank attaches equal weights to future and present: $\omega(q) = 1, \omega'(q) = 0$ (in other words, its discount factor is also 0). This provides an *upper bound* on optimal FG because it ignores the second-order discounting costs.[24] The optimal duration can now be solved in closed-form from (9) as:

$$q^* = \max\left\{0, \frac{1}{\nu}\frac{\Delta_L - \frac{(1-p\nu)^2}{1-p}}{1-p+\Delta_L}\right\},$$

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

A first insight allowed by the closed-form solution is that it is optimal to refrain from FG altogether ($q^* = 0$) when ν is *smaller* than a certain threshold $\nu^0 \equiv \left(1 - \sqrt{(1-p)\Delta_L}\right)/p$.[25]

[24] See Bilbiie, 2016 for an analysis of the accuracy of this in a RANK model.
[25] Under the baseline calibration, the threshold is 0.86.

Recalling the expression for ν, it follows immediately that (regardless of self-insurance) in the "amplification" case ($\chi > 1$) the region of λ for which FG is optimal will be ceteris paribus smaller than in the "dampening $\chi < 1$ case (ν is increasing with χ because both δ and η are). Moreover, in the "amplification" region $\chi > 1$ since ν is increasing both with the share of H λ and with idiosyncratic risk $1-s$, it follows that an increase in either of these parameters restricts the case for optimal FG.[26] The reason is the dark side of FG: more amplification brings more welfare cost of FG. Conversely of course, in the "dampening" region $\chi < 1$ the opposite is true: ν is decreasing in both λ and $1-s$, and an increase in either of those parameters pushes up the threshold for FG to be optimal.

How does the optimal FG duration depend on the key structural parameters? Since we know the derivatives of ν with respect to λ, s, and χ, it suffices to calculate the derivative of q^* with respect to ν, which is:

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

When the disruption causing the liquidity trap is lower than a certain threshold $\Delta_L < (1-p)^{-1}$, which is the more empirically plausible case,[27] then q^* is increasing in ν if $\nu < \bar{\nu} \equiv \sqrt{1-\Delta_L(1-p)}/p$ and decreasing otherwise. Notice that this threshold is larger than the threshold needed for FG to be optimal at all ($q^* > 0$) derived above: $\bar{\nu} > \nu^0$. We have $dq^*/d\nu > 0$ when $\nu^0 < \nu < \bar{\nu}$ and $dq^*/d\nu < 0$ when $\nu^0 < \bar{\nu} < \nu$. It is useful to again distinguish the two cases depending on χ.

With dampening $\chi < 1$ more H and more risk imply that ν is decreasing; if we start with $\nu > \bar{\nu}$, optimal FG duration first increases, then decreases as ν crosses the threshold. Whereas if we start below the threshold, optimal FG duration decreases uniformly (this is the case shown in the Figure). 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 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

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

[27] If instead $\Delta_L > (1-p)^{-1}$, q^* is uniformly *decreasing* in ν: 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 $1-s$.

is in that case magnified—by the same token, the dark side (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 $s = 1$.

4.2 A Caveat

Is not conceivable that, if FG is less effective, optimal policy should imply doing more (rather than less) of it? Nakata, Schmidt, and Yoo (2017) follow this line of reasoning in a calibrated model with a discounted Euler equation à la MNS' that delivers FG power mitigation. The authors 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-policy-implied duration of FG becomes *increasing* in the share of constrained households. The reason is that, as the share of constrained 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.

I confirm Nakata et al's conclusion in my simpler model, with $\chi < 1$ and little idiosyncratic risk, i.e. the TANK model (red dash in the upper left panel in Figure 4): the optimal duration becomes increasing with the share of hand-to-mouth. There is an important qualification though, afforded by the analytical framework studied here. First, notice that in the same panel the blue dotted line is increasing only slightly initially, and still decreasing thereafter: in the SHANK model (with most idiosyncratic risk and strongest self-insurance motive) there is more dampening—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 seem to be two sides of the same coin: in this simple model at least, you cannot throw one and keep the other.[28]

[28] 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 SHANK model, the shock necessary to replicate an even modest recession (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.

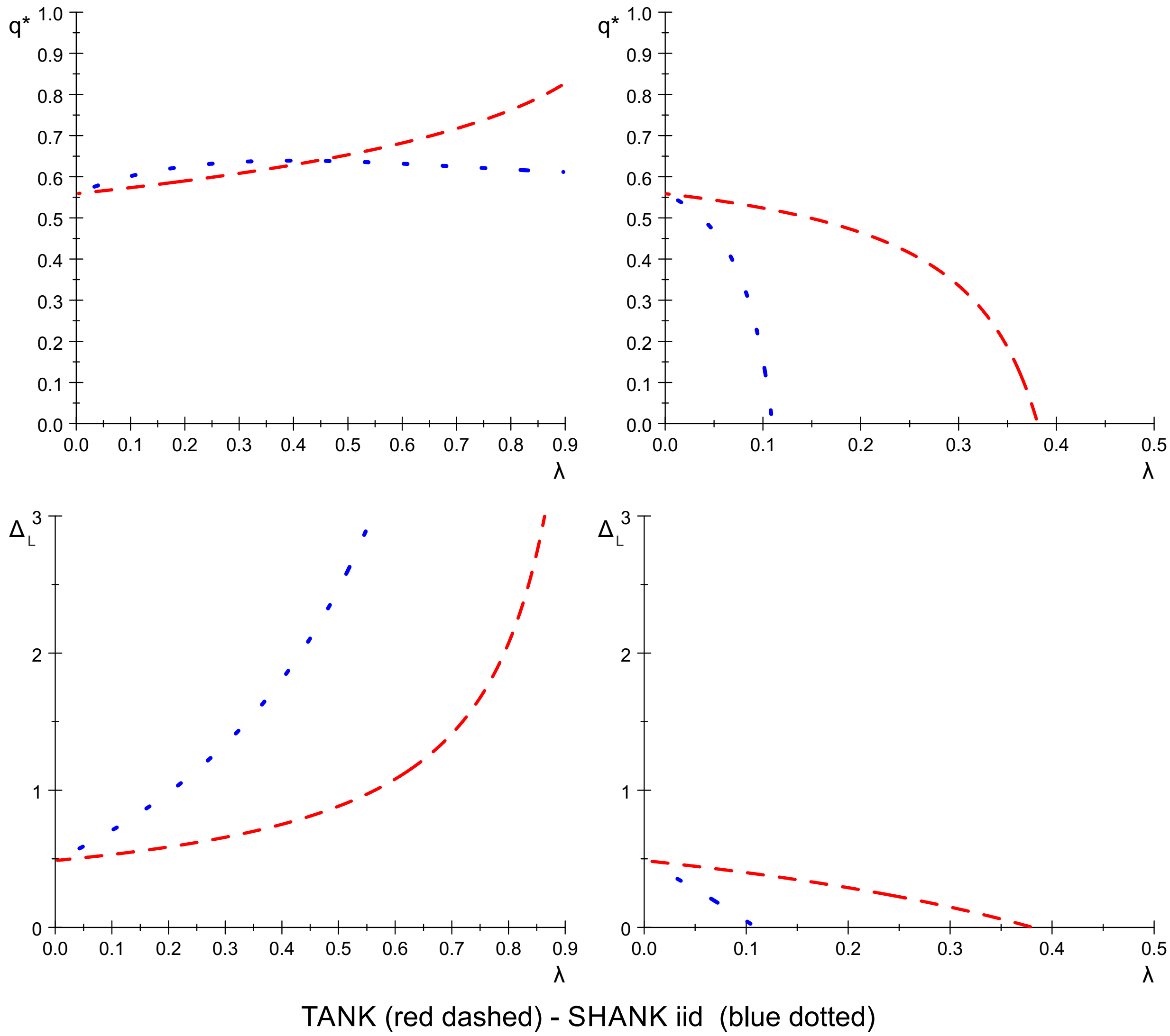


Figure 4 q^* as a function of λ: fixed recession, Δ_L adjusts endogenously

Moreover, the very same logic that generates increasing FG duration in the previous case is turned on its head in the amplification, $\chi > 1$ case: as the share of constrained 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 the SHANK model, self-insurance makes optimal FG duration decrease even faster. The general message is that keeping the observable recession fixed (rather than the unobservable disturbance) is a useful optimal policy exercise; but it 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.

5 Conclusions

How to make forward guidance less powerful in NK models? This paper has concentrated on one stream of solutions, of the several that have been proposed over the recent years. Namely, McKay, Nakamura and Steinsson (2015, 2016), have shown that the "forward guidance puzzle" (thus labelled by Del Negro, Giannoni and Patterson, 2012) is alleviated in heterogeneous-agent New Keynesian (HANK) models. Their models lead to a dampening of the power of forward guidance; the same holds in the different HANK model of Kaplan, Moll, and Violante (2015, 2016).

In this paper, I show that it is the interaction of two complementary channels that delivers this solution: constrained (*hand-to-mouth*) households' income depends on aggregate income less than one-for-one, and there is idiosyncratic risk inducing a *self-insurance* motive for unconstrained households. The former channel by itself does deliver dampening relative to the representative-agent model, but uniformly at all horizons and thus does not solve the puzzle. While the latter channel, because of the complementarity, magnifies the amplification inherent when hand-to-mouth income elasticity to aggregate is *higher* than one: FG power becomes phenomenal, and the puzzle is much aggravated.

Yet even when this (puzzling indeed) amplification of FG's power is a model feature, and despite there being more scope for using FG—because the same amplification also makes ZLB recessions deeper and increases the marginal utility of consumption—optimal policy does not imply a correspondingly higher duration of FG; indeed, it at least eventually implies that the optimal FG duration decrease with the share of hand-to-mouth. That is because there is a *dark side* to FG power: the welfare cost of inefficient volatility once the trap is over (first unveiled in Eggertsson and Woodford's celebrated 2003 analysis in a representative-agent model). This cost becomes very large too precisely when FG power does, thus making it optimal to contain the optimal duration of extra accommodation.

Three novel policy inputs are key, in light of this analysis, for a central bank's optimal policy in a liquidity trap: how many households are constrained; how their income is related to aggregate income (through employment or redistribution); and a measure of idiosyncratic risk.[29]

References

[1] Ascari, G., A. Colciago and L. Rossi, 2016, "Limited Asset Market Participation and Optimal

[29] I abstract from FG issues pertaning to commitment and communication in environments with information imperfections, emphasized i.a. by Bassetto (2016), Wiederholt (2016), and Garcia-Schmidt and Woodford (2014).

Monetary Policy" Mimeo Oxford University

[2] Auclert, A. 2016, "Monetary Policy and the Redistribution Channel" Mimeo Stanford

[3] Bassetto, M., 2016, "Forward Guidance: Communication, Commitment, or Both?", Mimeo Federal Reserve Bank of Chicago

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

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

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

[7] Bilbiie F. 2011. "Non-Separable Preferences, Frisch Labor Supply and the Consumption Multiplier of Government Spending: One Solution to a Fiscal Policy Puzzle, Journal of Money, Credit and Banking, 43(1), 221-251

[8] Bilbiie, F., Mueller G. and A. Meier 2008, "What Accounts for the Change in U.S. Fiscal Policy Transmission?", Journal of Money, Credit and Banking, 2008

[9] 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

[10] 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

[11] Bilbiie, F., T. Monacelli, and R. Perotti 2012, "Tax Cuts vs. Spending Increases: Welfare at the Zero Lower Bound" Mimeo PSE and Bocconi

[12] 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

[13] Bilbiie, F. 2016, "Optimal Forward Guidance", CEPR Discussion Paper

[14] Bilbiie, F. and X. Ragot, 2016, "Optimal Monetary Policy and Liquidity with Heterogeneous Households", Mimeo Paris School of Economics

[15] Broer, T., N. J. Hansen, P. Krusell, and E. Oberg, "The New Keynesian transmission mechanism: a heterogeneous agent perspective", Mimeo IIES Stockholm

[16] 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.

[17] Campbell, J. R., C. L. Evans, J. D. Fisher, and A. Justiniano, 2012: "Macroeconomic Effects of Federal Reserve Forward Guidance," Brookings Papers on Economic Activity, 44, 1–80.

[18] Campbell, J. R., J. D. Fisher, A. Justiniano, and L. Melosi, 2016: "Forward Guidance and Macroeconomic Outcomes Since the Financial Crisis", NBER Macroeconomics Annual

[19] Carlstrom, C. T., T. S. Fuerst, and M. Paustian, 2015, "Inflation and Output in New Keynesian Models with a Transient Interest Rate Peg," Journal of Monetary Economics, 76, 230-243

[20] Cochrane, J., 2015, "The New Keynesian Liquidity Trap", Mimeo Hoover Institution

[21] Challe, E. and X. Ragot, 2014 "Precautionary Saving Over the Business Cycle," The Economic Journal

[22] Challe, E, J. Matheron, X. Ragot and J. Rubio-Ramirez, 2016, ""Precautionary Saving and Aggregate Demand", forthcoming Quantitative Economics

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

[24] Colciago, A. 2011, Rule-of-thumb consumers meet sticky wages, Journal of Money Credit and Banking 43(2), 325-353.

[25] Curdia, V. and M. Woodford 2009, "Credit Frictions and Optimal Monetary Policy," Journal of Monetary Economics, forthcoming

[26] Del Negro, M., M. Giannoni, and C. Patterson 2012: "The Forward Guidance Puzzle," Federal Reserve Bank of New York

[27] 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.

[28] Eggertsson, G. and M. Woodford, 2003, "The Zero Bound on Interest Rates and Optimal Monetary Policy," Brookings Papers on Economic Activity

[29] Farhi, E. and I. Werning, 2017, "Monetary policy, bounded rationality, and incomplete markets" Mimeo Harvard and MIT

[30] Filardo, A. and B. Hofmann, 2014, “Forward guidance at the zero lower bound,” BIS Quarterly Review.

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

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

[33] 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.

[34] Garcia-Schmidt, M. and M. Woodford, 2014. “Are Low Interest Rates Deflationary? A Paradox of Perfect Foresight Analysis,” Mimeo Columbia.

[35] Gürkaynak, R. S., B. Sack, and E. Swanson, “The sensitivity of long-term interest rates to economic news: Evidence and implications for macroeconomic models,” American Economic Review, 2005, pp. 425–436

[36] Kaplan, G., B. Moll and G. Violante, 2015, "Monetary Policy According to HANK", Mimeo Princeton and NYU

[37] Kaplan, G., B. Moll and G. Violante, 2016, "A Note on Unconventional Monetary Policy in HANK", Mimeo Princeton and NYU

[38] Keynes, J.M. 1935 ”The General Theory of Employment, Interest and Money”, MacMillan, Cambridge.

[39] Kiley, M., 2016, 2016 “Policy Paradoxes in the New Keynesian Model,” Review of Economic Dynamics, 21, 1–15.

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

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

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

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

[44] McKay, A., E. Nakamura, and J. Steinsson 2015: “The Power of Forward Guidance Revisited,” American Economic Review

[45] McKay, A., E. Nakamura and J. Steinsson 2016: "The Discounted Euler Equation", Economica

[46] Nakata, T., 2014 "Reputation and Liquidity Traps”, Mimeo Federal Reserve Board

[47] Nakata, T. and S. Schmidt, 2016 "Conservatism and Liquidity Traps" mimeo Federal Reserve Board and European Central Bank

[48] Nakata, T., S. Schmidt, and P. Yoo, 2017, "Attenuating the Forward Guidance Puzzle: Implications for Optimal Monetary Policy" mimeo Federal Reserve Board and European Central Bank

[49] Nistico, S., 2012 “Monetary policy and stock-price dynamics in a DSGE framework,” Journal of Macroeconomics, 34 (1), 126–146.

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

[51] Piergallini, A., 2006, “Real balance effects and monetary policy,” Economic Inquiry, 44 (3), 497–511.

[52] Ravn, M. and V. Sterk, “Job Uncertainty and Deep Recessions,” Mimeo, University College London 2013.

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

[54] Walsh, C., 2016, "Sustainable Forward Guidance at the ELB", Mimeo UCSB

[55] Werning, I, 2016, "Incomplete markets and aggregate demand", Mimeo MIT

[56] Wiederholt, M. 2015, “Empirical Properties of Inflation Expectations at the Zero Lower Bound,” Mimeo Goethe University Frankfurt.

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

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