DISCUSSION PAPER SERIES

DP14159
(v. 4)

MONETARY POLICY AND REDISTRIBUTION: A LOOK UNDER THE HATCH WITH TANK

Lilia Maliar and Christopher Naubert

MONETARY ECONOMICS AND FLUCTUATIONS

ISSN 0265-8003

MONETARY POLICY AND REDISTRIBUTION: A LOOK UNDER THE HATCH WITH TANK

Lilia Maliar and Christopher Naubert

Discussion Paper DP14159
First Published 30 November 2019
This Revision 24 August 2020

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

MONETARY POLICY AND REDISTRIBUTION: A LOOK UNDER THE HATCH WITH TANK

Abstract

We provide an under-the-hatch look at the mechanisms at work in two-agent new Keynesian (TANK) models. A distinctive feature of our analysis is novel closed-form solutions for general Taylor-type monetary-policy rules. Our analysis of TANK models informs the quantitative heterogeneous-agent new Keynesian literature about potential mechanisms at play when evaluating the effects of monetary policy and fiscal redistribution between asset-market participants and non-participants. We find that, in conventional times, forward guidance is less effective at stimulating today's economy when asset-market participation is low. Additionally, future monetary-policy shocks may work mostly through direct channels even though contemporaneous shocks work largely through indirect channels. Finally, both consumption and income inequality in the model are fully determined by markups. Introducing capital adjustment costs overturns this result and leads to consumption inequality being more countercyclical than income inequality, as is in the data.

JEL Classification: C61, C63, C68, E31, E52

Keywords: forward guidance, New Keynesian Model, TANK, redistribution

Lilia Maliar - lmaliar@gc.cuny.edu
The Graduate Center, CUNY and CEPR

Christopher Naubert - cnaubert@gradcenter.cuny.edu
The Graduate Center, CUNY

Acknowledgements
The earlier version of the paper circulated under the title "TANK Models with Amplification and no Puzzles: the Magic of Output Stabilization and Capital". We thank participants at the 2019 CEPR conference "Heterogeneous Agents or Heterogeneous Information" in Paris, seminars at the CUNY Graduate Center and Bank of England, and the 2020 Econometric Society World Congress for useful comments.

Monetary Policy and Redistribution: A Look under the Hatch with TANK*

Lilia Maliar† and Christopher Naubert‡

December 29, 2019

Abstract

We provide an under-the-hatch look at the mechanisms at work in two-agent new Keynesian (TANK) models. A distinctive feature of our analysis is novel closed-form solutions for general Taylor-type monetary-policy rules. Our analysis of TANK models informs the quantitative heterogeneous-agent new Keynesian literature about potential mechanisms at play when evaluating the effects of monetary policy and fiscal redistribution between asset-market participants and non-participants. We find that, in conventional times, forward guidance is less effective at stimulating today's economy when asset-market participation is low. Additionally, future monetary-policy shocks may work mostly through direct channels even though contemporaneous shocks work largely through indirect channels. Finally, both consumption and income inequality in the model are fully determined by markups. Introducing capital adjustment costs overturns this result and leads to consumption inequality being more countercyclical than income inequality, as is in the data.

JEL classification : C61, C63, C68, E31, E52

Key Words : forward guidance, new Keynesian model, TANK, redistribution, HANK, monetary policy

*The earlier version of the paper circulated under the title "TANK Models with Amplification and no Puzzles: the Magic of Output Stabilization and Capital". We thank participants at the 2019 CEPR conference "Heterogeneous Agents or Heterogeneous Information" in Paris, seminars at the CUNY Graduate Center and Bank of England, and the 2020 Econometric Society World Congress for useful comments.

†Department of Economics, The Graduate Center, City University of New York and CEPR

‡Department of Economics, The Graduate Center, City University of New York

1 Introduction

For decades, monetary policy was analyzed within the representative-agent new Keynesian (RANK) framework. However, many questions policymakers face today concern the distributional effects of policy. Therefore, economists are increasingly incorporating heterogeneity in asset holdings and incomes into new Keynesian models, leading to the current era of heterogeneous-agent new Keynesian (HANK) models. Key questions that HANK models address include: (i) How are the properties of solutions and equilibrium uniqueness affected by heterogeneity? (ii) How does heterogeneity affect aggregate and individual responses to monetary policy? (iii) How do monetary-policy shocks transmit through the economy when agents are heterogeneous? (iv) How does redistribution across agents affect the economy and interact with monetary policy? (v) How do different types of shocks affect inequality?

Quantitative HANK models address these questions using numerical solutions; see, e.g., Gornemann et al. (2016), Guerrieri and Lorenzoni (2017), Kaplan et al. (2018), Debortoli and Gali (2018), Auclert (2019), Auclert et al. (2020).[1] However, numerical solutions are not always conclusive. In this paper, we construct novel closed-form solutions for TANK models to derive sharp answers to the above five questions.

Specifically, the present paper takes an under-the-hatch look at two-agent new Keynesian (TANK) models and provides precise characterizations of the mechanisms at work in TANK. We build on TANK models of Bilbiie (2008) and Debortoli and Galí (2018). The model features households who differ in their ability to access asset markets. Namely, one part of population cannot save and just earns labor income, while the other part has the ability to save through bonds and shares of production firms. As a result, agents are heterogenous in their marginal propensities to consume. The government redistributes profits, and labor tax revenues between the two types of agents, which allows the study of the interaction between monetary and fiscal redistribution policies.

(i) There are few results in the literature regarding equilibrium uniqueness for HANK models. Acharya and Dogra (2018) analyze how cyclical properties of idiosyncratic shocks affect equilibrium determinacy. Auclert et al. (2019) provide a criterion for uniqueness in HANK models based on winding numbers; however, this approach does not guide on how the underlying parameters, such as those relating to fiscal redistribution, affect indeterminacy regions. The analysis in the present paper helps us inform the quantitative HANK literature on how the model's parameters, related to fiscal redistribution and the share of non-asset market participants, affect equilibrium uniqueness, as well as the properties of unique solutions.

(ii) We analyze how heterogeneity affects aggregate and individual responses to forward guidance modeled as a shock in the central bank's policy rule. Werning (2015) shows that heterogeneity, on its own, is not enough to reduce the power of a future real interest rate change; see Debortoli and Gali (2018), and Bilbiie (2019a) for examples. Bilbiie (2019a), Kaplan et al. (2016) and McKay et al. (2016) introduce additional features which diminish the power of a future real interest rate change. We consider two experiment: forward guidance as in McKay et al. (2016) and that used in conjunction with active monetary policy, as in Maliar and Taylor (2018). In the former, the sequence of real interest rates is fixed exogenously, while in the latter it is endogenous. With fixed real interest rates, we demonstrate that heterogeneity amplifies today's response of economic aggregates to forward-guidance policy. When the central bank uses forward guidance along with active policy, we find that future monetary policy shocks have small effects on today's economy. As the share of non-asset market participants increases, monetary policy shocks become even less powerful. Therefore, introducing non-asset market participants reduces the power of forward guidance in normal times.

(iii) Monetary policy can transmit through the economy much differently in HANK than it does in RANK. Auclert (2019) shows that heterogeneity in earnings, nominal rate exposure and real rate exposure affect the transmission of monetary policy when agents have different marginal propensities to consume. Kekre and Lenel (2020) emphasize the role of the marginal propensity to bear risk in monetary policy transmission and show that monetary policy shocks redistribute towards agents with a high marginal

[1] Other HANK models include Broer et al. (2020), Farhi and Werning (2016), Hagedorn et al. (2019), Melcangi and Sterk (2019), and Ravn and Sterk (2018).

propensity to bear risk. Kaplan et al. (2018) decompose the response of the economy to a contemporaneous monetary policy shock into a component that captures the direct response to changes in the interest rate and a component that captures the indirect response due to changes in income. They find that the share of indirect effects in their TANK model roughly corresponds to the share of non-asset market participants. We use closed-form solutions to characterize analytically the direct and indirect effects under a general Taylor rule. Similarly to the previous literature, when considering contemporaneous shocks, we find that the share of direct effects decreases with the share of non-asset market participants. When there are future monetary policy shocks, we find that indirect effects work to increase consumption, while direct effects work to reduce consumption when the share of non-asset market participants is high; the reverse is true when the share of non-asset market participants is low. Therefore, when asset market participation is abundant enough, the transmission of future monetary shocks in TANK works much like it does in a RANK economy, even if the transmission of contemporaneous shocks does not. In other words, the transmission of future shocks is more RANK-like than TANK-like.

(iv) Importantly, models with heterogeneous agents allow for analysis of fiscal redistribution policies; see, e.g., Oh and Reis (2012), McKay and Reis (2016). In this paper, we consider redistribution funded by either both agents (labor income taxes) or only unconstrained agents (illiquid profits). We show that the monetary authority must take into account the actions of the fiscal authority to ensure equilibrium determinacy. Moreover, we show that whether or not increases in net transfers to constrained agents allow for stronger responses by the central bank depends on how those transfers are funded. We also consider how transfers affect consumption volatility of the two agents. First, when constrained agents are provided a positive net transfer of labor-tax revenues, constrained agent consumption volatility increases. However, transfers of illiquid profits to constrained agents can significantly reduce consumption volatility for the constrained agents and may or may not increase consumption volatility of unconstrained agents. When asset-market participation is limited enough, the reduction in income volatility for the unconstrained agents more than compensates for the reduced ability to directly smooth consumption due to lower non-labor income, resulting in lower consumption volatility. Our results on indeterminacy and consumption volatility show that one cannot draw conclusions about the effects of transfers without consideration of how the transfers are funded.

(v) HANK models are used to characterize inequality over economic cycles; see, e.g., Auclert et al. (2020). Our solutions enable us to assess the cyclicality of income and consumption inequality not only for deterministic demand shocks, as is usually done in the new Keynesian literature, but also over cycles driven by either demand or supply shocks. In TANK, both income and consumption inequality are fully determined by the markup. Therefore, supply-shock-driven cycles result in procyclical inequality and procyclical markups, while demand-shock-driven cycles result in countercyclical inequality and countercyclical markups. Neither prediction is consistent with the empirical evidence that consumption and income inequality are countercyclical, while markups are procyclical. To address this, we introduce capital and adjustment costs to the model. In work occurring simultaneously, Bilbiie et al. (2019c) use a TANK model with capital to study how inequality responds to monetary-policy shocks. In their model, complementarity between capital and income inequality amplifies the aggregate response to a monetary-policy shock and results in countercyclical consumption and income inequality. Unlike their paper, we consider supply shocks in addition to monetary-policy shocks. We show that the consumption inequality depends on markups, as well as on the value of installed capital and return on capital. However, income inequality is still fully determined by the markup. We find that consumption inequality is more countercyclical than income inequality in response to demand shock cycles. Additionally, the model is able to generate countercyclical consumption inequality and procyclical markups over supply -shock-driven business cycles.

Our analysis of the mechanisms at work in TANK would not be feasible without the closed-form solutions. Such solutions allow us to analyze analytically the individual responses of constrained and unconstrained agents and decompose total effects of contemporaneous and future, one-time and persistent monetary-policy shocks into direct and indirect effects. Furthermore, the closed-form solutions allow us to understand detailed mechanisms behind the effects of fiscal redistribution policy and asset-market participation on the indeterminacy regions. The solutions we construct are applicable not only to the deterministic setting with exogenous disturbances in the natural rate of interest, desired interest rate, preferences, productivity, nominal interest rate but also to the stochastic setting.

There are few other papers in the HANK literature that derive closed-form solutions. First, a "pseudo" RANK model, PRANK, of Acharya and Dogra (2018) introduces uninsurable idiosyncratic uncertainty but allows for full asset-market participation; the assumption of exponential utility allows for analytic solutions (as there is no need to keep track of wealth distribution). While their model helps them to study the roles of precautionary savings and cyclical income risk, it is not designed to assess the impact of redistribution across agents with heterogeneous marginal propensities to consume (MPC) as all agents have the same MPC. Second, a "tractable" HANK model, THANK, of Bilbiie (2019b), which nests TANK, allows for uninsurable idiosyncratic uncertainty but obtains analytical tractability by assuming full insurance within types, steady-state equality, limited insurance across types and exogenous switches between types. Our paper abstracts from idiosyncratic uncertainty; however, it allows for steady-state inequality and more general Taylor rules with output, inflation and expected inflation targeting.

The rest of the paper is as follows: Section 2 describes both non-linear and log-linerized versions of the basic TANK model, presents the closed-form solutions and discusses when the model produces the same results as RANK. Section 3 discusses how heterogeneity affects the properties of solutions and uniqueness of equilibrium. Section 4 analyzes aggregate and individual responses to future anticipated shocks (forward-guidance, monetary-policy shocks). Section 5 decomposes the response into direct and indirect effects. Section 6 presents the solutions for the stochastic case and discusses the effects of redistribution on economic volatility. Section 7 discusses the model's predictions about consumption and income inequality. Section 8 considers the robustness of these predictions by extending the model to include capital, and finally, Section 9 concludes.

2 Closed-form solution to the TANK model

The two main features of TANK are limited asset market participation and redistribution of profits. The former assumption was first used in the context of new Keynesian models in Gali et al. (2007) and Bilbiie (2008) and the latter assumption was emphasized in Bilbiie (2008). We construct a closed-form solution to the version of the TANK model studied in Bilbiie (2008) and Debortoli and Galí (2018).[2]

2.1 TANK model

There is a continuum of agents on the unit interval. The agents are classified based on their ability to access asset markets. There is a λ share of constrained agents and a $1-\lambda$ share of unconstrained agents; the variables of such agents are denoted by superscripts K ("*Keynesian*") and U, respectively. The constrained agents receive their income from supplying labor in the market and transfers from the government, and they do not trade any assets. In addition, the economy includes a continuum of production firms, indexed by $i \in [0, 1]$, producing differentiated goods, a central bank and a government.

Unconstrained agents. An unconstraint agent solves a dynamic problem by trading one-period riskless bonds and mutual-fund shares:

$$\max_{\{C_t^U, N_t^U, B_t^U, F_t^U\}} E_0 \sum_{t=0}^{\infty} \beta^t Z_t \left[\frac{(C_t^U)^{1-\sigma} - 1}{1-\sigma} - \frac{(N_t^U)^{1+\varphi} + 1}{1+\varphi} \right] \tag{1}$$

[2] There are several important differences between the assumptions in Bilbiie (2008) and Debortoli and Galí's (2018), including complete versus incomplete markets, the logarithmic versus more general CRRA utility functions, zero versus non-zero steady state profits, the presence of preference shock and the way of modeling the redistributive policy.

subject to

$$\text{s.t. } C_t^U + \frac{B_t^U}{P_t} + Q_t F_t^U = \frac{B_{t-1}^U R_{t-1}}{P_t} + \left(1-\delta^W\right) \frac{W_t}{P_t} N_t^U + \left[Q_t + (1-\delta) D_t\right] F_{t-1}^U + T_{D,t}^U + T_{W,t}^U, \tag{2}$$

where initial condition (B_{-1}^U, F_{-1}^U) is given. Here, E_0 is the conditional expectation operator; $\beta \in (0,1)$; $\sigma > 0$ and $\varphi > 0$; C_t^U, N_t^U, B_t^U, F_t^U, $T_{D,t}^U$ and $T_{W,t}^U$are the unconstrained agent's consumption, labor, nominal bond holdings, intermediate-good producers' shares, real government transfers of illiquid profits, and real government transfers of labor tax revenue, respectively; D_t is the dividend from ownership of firms; P_t, Q_t, W_t and R_{t-1} are consumption-good price index, share price, nominal wage and (gross) nominal interest rate, respectively; Z_t is a preference shock, following an AR(1) process, $Z_{t+1} = Z_t^{\rho_z} \exp(\varepsilon_{z,t+1})$ with $\varepsilon_{z,t+1} \sim \mathcal{N}\left(0, \sigma_z^2\right)$; $\delta \in [0,1]$ is a fraction of illiquid profits allocated across agents by government and $\delta^W \in [0,1]$ is the labor-income tax rate.[3] The consumption choice C_t^U arises from the standard cost minimization over the continuum of goods produced by the firms; for the first-order conditions, see Appendix A.1, where we collected all the derivations of Section 2.

Constrained agents. A constrained agent solves a static problem

$$\max_{C_t^K, N_t^K} Z_t \left[\frac{(C_t^K)^{1-\sigma} - 1}{1-\sigma} - \frac{(N_t^K)^{1+\varphi} + 1}{1+\varphi}\right] \tag{3}$$

subject to

$$C_t^K = (1-\delta^W) \frac{W_t}{P_t} N_t^K + T_{D,t}^K + T_{W,t}^K, \tag{4}$$

where C_t^K, N_t^K, $T_{D,t}^K$ and $T_{W,t}^K$ are the constrained agent's consumption, labor, government transfers of illiquid profits and government transfers of labor-tax revenues.

Government. Government makes transfers of illiquid profits and labor-tax revenues to the unconstrained and constrained agents, $T_{D,t}^U, T_{W,t}^U$, $T_{D,t}^K$ and $T_{W,t}^K$, by re-distributing the illiquid profits of the firms δD_t and labor-tax revenues $\delta^W \frac{W_t}{P_t} N_t$, so that $(1-\lambda) T_{D,t}^U + \lambda T_{D,t}^K = \delta D_t$ and $(1-\lambda) T_{W,t}^U + \lambda T_{W,t}^K = \delta^W \frac{W_t}{P_t} N_t$. It does so according to the following rules:

$$T_{D,t}^U = \left(1 + \frac{\tau\lambda}{1-\lambda}\right) \delta D_t, \quad T_{D,t}^K = (1-\tau)\,\delta D_t, \tag{5}$$

$$T_{W,t}^U = \left(1 + \frac{\tau^W\lambda}{1-\lambda}\right) \delta^W \frac{W_t}{P_t} N_t, \quad T_{W,t}^K = \left(1-\tau^W\right) \delta^W \frac{W_t}{P_t} N_t, \tag{6}$$

where $1-\tau$ and $1-\tau^W$ are shares of profits and labor-tax revenues, respectively, distributed to the constrained agents. The redistribution of illiquid profits only directly affects unconstrained agents, and that of labor income taxes directly affects both agents. In the previous TANK models, only profits are redistributed; see Debortoli and Gali (2018) and Bilbiie (2008).

Supply side. Following Debortoli and Galí (2018), we assume that wage is a constant share of the ratio of the marginal utility of (average) consumption to the marginal utility of (average) leisure,

$$\frac{W_t}{P_t} = \mathcal{M}^w C_t^{\sigma} N_t^{\varphi}, \tag{7}$$

[3] Our definition of illiquid profits differs from that in Kaplan et al. (2018): our illiquid profits are untraded, and, therefore, not held in an account of the unconstrained agent, while their illiquid profits are distributed to an account which is subject to transaction costs.

where C_t and N_t are average consumption and labor, respectively, $C_t \equiv \int_0^1 C_t(s)$ and $N_t \equiv \int_0^1 N_t(s)$; $\mathcal{M}^w > 1$ is the average wage markup. Each agent $s \in [0,1]$ has incentives to supply labor at a market wage $\frac{W_t}{P_t}$, i.e., $\frac{W_t}{P_t} \geq \mathcal{M}^w \left(C_t^U\right)^\sigma \left(N_t^U\right)^\varphi$ and $\frac{W_t}{P_t} \geq \mathcal{M}^w \left(C_t^K\right)^\sigma \left(N_t^K\right)^\varphi$. As a consequence, both types of agents will supply the same labor.[4]

Each producer i owns a linear technology described by $Y_t(i) = A_t N_t(i)$, where A_t is the productivity level following an AR(1) process $A_{t+1} = A_t^{\rho_a} \exp(\varepsilon_{a,t+1})$ with $\varepsilon_{a,t+1} \sim \mathcal{N}\left(0, \sigma_a^2\right)$, and $N_t(i)$ is labor input. The firm's price is rigid a la Rotemberg (1982), i.e., it sets prices optimally in every period of time subject to adjustment costs, given by $\frac{\xi}{2} P_t Y_t \left(\frac{P_t(i)}{P_{t-1}(i)} - 1\right)^2$, where Y_t is aggregate output, and $\xi > 0$. The demand for its differentiated good is $Y_t(i) = Y_t \left(\frac{P_t(i)}{P_t}\right)^{-\varepsilon}$, which comes from the Dixit-Stiglitz aggregator $Y_t = \left(\int_0^1 Y_t(i)^{\frac{\varepsilon-1}{\varepsilon}} di\right)^{\frac{\varepsilon}{\varepsilon-1}}$, $\varepsilon \geq 1$.

Consumption gap. Due to different incomes, the two types of agents will differ with respect to consumption. We define a consumption gap as $\Gamma_t \equiv 1 - \frac{C_t^K}{C_t^U}$, which is given by

$$\Gamma_t = \frac{\left[\triangle_t \mathcal{M}_t - 1\right]\left[1 - \delta(1-\tau)\right] + \tau^W \delta^W}{1 + \left(\tau^W \delta^W - 1\right)\lambda + \left[\triangle_t \mathcal{M}_t - 1\right]\left[1 - \delta(1-\tau)\lambda\right]}. \tag{8}$$

Note that Γ_t is increasing in the price markup $\mathcal{M}_t$ under our assumption $\delta(1-\tau) < 1$. That is, there must be some fraction of profits that are liquid, $\delta < 1$, (so that shares are valued) or there must be less than full redistribution (i.e., unconstrained agents receive a proportionally larger share of illiquid profits than their share in the population, $(1-\lambda) T_{D,t}^U > \delta D_t (1-\lambda)$).

Monetary policy. To close the model, we assume that the central bank follows a Taylor rule which includes current inflation π_t, forward-looking inflation π_{t+1}, as well as a deviation of output from the steady state y_t,

$$i_t = i_t^* + \phi_\pi \pi_t + \phi_{E\pi} E_t\{\pi_{t+1}\} + \phi_y y_t + v_t, \tag{9}$$

where $\phi_\pi \geq 0$, $\phi_{E\pi} \geq 0$ and $\phi_y \geq 0$; i_t^* is an exogenously given desired long-run interest rate (either deterministic or stochastic), and v_t is a (log-deviation of) monetary-policy shock that can include both anticipated and unanticipated shocks. In the latter case, we assume that v_t follows an AR(1) process $v_{t+1} = \rho_v v_t + \varepsilon_{v,t+1}$ with $\varepsilon_{v,t+1} \sim \mathcal{N}\left(0, \sigma_v^2\right)$.[5] We assume the net inflation rate target is zero.

2.2 When does TANK become RANK?

There are two possible cases when the TANK model reduces to a RANK model. The first, obvious case is when the share of constrained agents is zero, $\lambda = 0$. The second case is when all profits are illiquid ($\delta = 1$) and illiquid profits and labor tax revenues are distributed uniformly across households ($\tau = 0$ and $\tau^W = 0$). While the unconstrained agents would like to save via bonds, this is not possible as there is no one to trade with. Therefore, the only way for unconstrained agents to save is through holding shares in the firms. However, because the firm will never pay out any dividends as there are no liquid profits to distribute, the share price will be zero. Therefore, the budget constraint of the unconstrained agent will reduce to that of the constrained agent. Since the unconstrained agents have no way to save for the next period, they consume their entire labor income and transfer. Because all agents have the same level of productivity and because profits are uniformly distributed, all agents consume the same amount, which produces the RANK result.

[4] We implement the whole analysis for the model in which both types choose labor optimally. We do not report the results for this case as they are close to those of the model with identical labor but the formulae become cumbersome. We comment on the differences between the two models when relevant.

[5] This Taylor rule is in the spirit of Taylor (1993); see, e.g., Gali (2008, p. 82), and Debortoli et al. (2019).

2.3 Log-linearized model

We log-linearize the model around a zero-inflation steady state following Debortoli and Gali (2018); see Appendix A.1. The resulting aggregate Euler equation (IS curve), Phillips curve and Taylor rule, respectively, are

$$x_t = E_t \{x_{t+1}\} - \frac{1}{\sigma(1-\Phi)} [i_t - E_t \{\pi_{t+1}\} - r_t^n], \tag{10}$$

$$\pi_t = \beta E_t \{\pi_{t+1}\} + \kappa x_t, \tag{11}$$

$$i_t = i_t^* + \phi_\pi \pi_t + \phi_{E\pi} E_t \{\pi_{t+1}\} + \phi_y x_t + u_t, \tag{12}$$

where x_t is the output gap (in log-deviations from the natural level of output); π_t is (net) inflation; *and* i_t is the (net) nominal interest rate. The natural rate of interest r_t^n and the shock u_t in (12) are related to the shocks in nonlinear model; see the appendix. The composite parameter Φ is defined later in (15).

The only difference between log-linearized TANK and RANK lies in the aggregate Euler equation (10) and comes from the heterogeneity parameter Φ. If $\Phi = 0$, then TANK becomes RANK (either when $\lambda = 0$ or when $\delta = 1$, $\tau = 0$ and $\tau^W = 0$). Because the effect of heterogeneity is compressed in a single parameter Φ, the TANK model provides a very tractable way to assess the role of heterogeneity on aggregate dynamics.[6]

2.4 Closed-form solutions

Our next step it to construct a closed-from solution to the model. We rewrite equations (10)–(12) as a second-order difference equation in π_t. For this purpose, we substitute i_t from the Taylor rule (9) into the Euler equation (10), use the Phillips curve (11) to get x_t and x_{t+1} and substitute them into (10) to obtain

$$E_t \{\pi_{t+2}\} + bE_t \{\pi_{t+1}\} + c\pi_t = -X_t, \tag{13}$$

$$X_t \equiv -\frac{\kappa}{\beta\sigma(1-\Phi)} (r_t^n - u_t), \tag{14}$$

where X_t summarizes all exogenous disturbances $\{r_t^n, a_t, v_t, i_t^*\}$, and constants b and c are $b \equiv -1 - \frac{1}{\beta} + \frac{1}{\beta\sigma(1-\Phi)} (\phi_{E\pi}\kappa - \beta\phi_y - \kappa)$ and $c \equiv \frac{1}{\beta} + \frac{\kappa\phi_\pi + \phi_y}{\beta\sigma(1-\Phi)}$.

Equation (13) is the same as in Maliar and Taylor (2018) who derived a closed-form solution to RANK (except that the parameter Φ is absent in their model). There is a remarkably simple way to extend their RANK solutions to our TANK model. Namely, we only need to replace σ in their analysis with $\sigma(1-\Phi)$ and all their theorems holds. In Table 1, we show the resulting characteristic roots m_1 and m_2 of (13), closed-form solutions, as well as boundaries on different types of solutions for the TANK model.[7]

The table contains the solutions for the case when shocks $\{v_t, z_t, a_t\}$ are deterministic and anticipated; the case of stochastic shocks is studied in Section 6. There are four possible types of solutions: i) an indeterminate solution characterized by an arbitrary integration constant C; ii) a unique solution with 2 distinct real roots; iii) a unique solution with 1 repeated real root; iv) a unique solution with complex roots. The formulas for the bounds on different types of solutions $\phi_{E\pi}^1 - \phi_{E\pi}^4$ are provided later in (21)–(23).

All the constructed solutions are forward stable, i.e., they satisfy the transversality condition. Finally, using the aggregate Euler equation (10), one can write a solution for x_t that is solely a function of parameters and exogenous shocks; see formula (A12) in the appendix.

[6] If we allow for heterogeneous labor across types, the slope of the Phillips curve κ will also depend on the heterogeneity and redistribution parameters $\{\lambda, \delta, \tau, \delta^W, \tau^W\}$. Therefore, TANK will differ from RANK not only in Φ but also in κ.

[7] Cochrane (2017) derives closed-form solutions in the continuous-time model for case i) under an assumption that all coefficients in the Taylor rule are set to zero.

Table 1: Closed-form solutions.

	Case i): indeterminate solution	Case ii): unique solution, 2 real roots
Roots values	$\lvert m_1\rvert \geq 1, \lvert m_2\rvert \leq 1$	$\lvert m_1\rvert > 1$, $\lvert m_2\rvert > 1$
Bounds	$\phi_{E\pi} < \phi^1_{E\pi}$ & $\phi_{E\pi} > \phi^4_{E\pi}$	$\phi^1_{E\pi} \leq \phi_{E\pi} < \phi^2_{E\pi}$ & $\phi^3_{E\pi} \leq \phi_{E\pi} < \phi^4_{E\pi}$
Solution	$Cm_2^t + \frac{1}{m_1-m_2} E_t \left[\sum_{s=t}^{\infty} m_1^{t-1-s} X_s + \sum_{s=-\infty}^{t-1} m_2^{t-1-s} X_s\right]$	$\frac{1}{m_1-m_2} E_t \left[\sum_{s=t}^{\infty} m_1^{t-1-s} X_s - \sum_{s=t}^{\infty} m_2^{t-1-s} X_s\right]$
	Case iii): unique solution, 1 repeated root	Case iv); unique solution, complex roots
Roots values	$m_1 = m_2 = m$, $\lvert m_1\rvert > 1$	$m_{1,2} = \mu \pm \eta i$, $r \equiv \sqrt{\mu^2+\eta^2} > 1$
Bounds	$\phi_{E\pi} = \phi^2_{E\pi}$ & $\phi_{E\pi} = \phi^3_{E\pi}$	$\phi^2_{E\pi} < \phi_{E\pi} < \phi^3_{E\pi}$
Solution	$\frac{1}{m} E_t \left[(t-1)\sum_{s=t}^{\infty} m^{t-1-s} X_s - \sum_{s=t}^{\infty} s m^{t-1-s} X_s\right]$	$\frac{1}{\eta} E_t \left[\sum_{s=t}^{\infty} r^{t-1-s} \sin\left(\theta\left(t-1-s\right)\right) X_s\right]$

Notes: C is an arbitrary constant of integration; for case iv), $\eta = Im[m_1]$, $r \equiv \sqrt{\mu^2+\eta^2} \geq 1$ and $\theta \equiv \arctan\left(\frac{\eta}{\mu}\right)$.

3 How heterogeneity affects the properties of solutions and equilibrium uniqueness

In this section, we study how heterogeneity affects the solutions via the composite parameter Φ. From the aggregate Euler equation (10), we see that the aggregate predictions of the model are the same as those of a representative agent model with an inverse elasticity of intertemporal substitution coefficient $\sigma(1-\Phi)$. Therefore, we only need to understand how Φ depends on heterogeneity.

3.1 Effects of heterogeneity on the composite parameter Φ

The composite parameter Φ appearing in the aggregate Euler equation is defined as

$$\Phi = \frac{\lambda\left(\sigma+\varphi\right)\Psi}{1-\lambda\Gamma}, \tag{15}$$

where constants Ψ and Γ are given by

$$\Gamma = \frac{(\mathcal{M}-1)(1-\delta(1-\tau))+\tau^W\delta^W}{1+\left(\tau^W\delta^W-1\right)\lambda+(\mathcal{M}-1)(1-\delta(1-\tau)\lambda)}, \tag{16}$$

$$\Psi = \frac{(1-\lambda)(1-\delta(1-\tau)-\delta^W\tau^W)}{\left[1+\left(\delta^W\tau^W-1\right)\lambda+(\mathcal{M}-1)(1-\delta(1-\tau)\lambda)\right]^2}, \tag{17}$$

where $\mathcal{M} \equiv \frac{\varepsilon}{\varepsilon-1}$ is the steady-state price markup; Γ is the steady-state consumption gap. We first analyze how Φ depends on each of its arguments $\left(\lambda, \tau, \delta, \tau^W, \delta^W\right)$. We focus on the more empirically relevant case of $\Phi < 1$ in which the aggregate Euler equation is downward sloping although the case $\Phi > 1$ was also documented in history; see Bilbiie (2008) for a discussion.[8]

Proposition 1. *Assume $\lambda < \frac{1}{2}$ and $\Phi < 1$. Then,*

[8] Bilbiie (2008) argues that during the pre-Volker period, the Federal Reserve followed an inverted Taylor principle (i.e. responded less than one-to-one to inflation). This policy led to determinacy because the economy was characterized by an upward sloping aggregate Euler equation due to very limited asset-market participation (i.e., high λ). He argues that in the 1980's, the United States expanded asset market participation, returning the economy to standard downward demand and requiring active monetary policy to insure determinacy.

(i) If $\mathcal{M} < 1 + \frac{1}{\delta(1-\tau)}$, *we have* $\frac{\partial\Phi}{\partial\lambda} > 0$.
(ii) If $\tau < 1$, *we have* $\frac{\partial\Phi}{\partial\tau} > 0$ *and* $\frac{\partial\Phi}{\partial\delta} < 0$.
(iii) $\frac{\partial\Phi}{\partial\tau^W} < 0$ *and* $\frac{\partial\Phi}{\partial\delta^W} < 0$.

Proof. See Appendix A.2. ■

3.2 Relation between individual and aggregate variables

We now discuss why changes in $\{\lambda, \tau, \delta, \tau^W, \delta^W\}$ affect Φ in the ways described in Proposition 1. To do so, as an example, we show what happens at the individual level when there is a cut in the real interest rate.

The following relations between individual and aggregate variables hold in the model:

$$n_t = \frac{1}{\varphi}w_t - \frac{1}{1-\Phi}\left(\frac{\sigma}{\varphi}\right)c_t^U - \frac{\Phi}{1-\Phi}\left(\frac{\sigma}{\varphi}\right)\left(\frac{1+\varphi}{\sigma+\varphi}\right)a_t, \tag{18}$$

$$c_t^U = (1-\Phi)\,c_t + \Phi\left(\frac{1+\varphi}{\sigma+\varphi}\right)a_t, \tag{19}$$

$$c_t^K = \chi \cdot c_t + (1-\chi)\left(\frac{1+\varphi}{\sigma+\varphi}\right)a_t, \tag{20}$$

where a constant χ is

$$\chi \equiv 1 + \frac{(1-\lambda)\Phi}{\lambda(1-\Gamma)} \geq 1,$$

and n_t is labor of both constrained and unconstrained agents; see Appendix A.2 for derivations. In the following discussion, we assume $a_t = 0$.

Effects of an interest rate cut. Consider a 1% one-time, contemporaneous real interest-rate cut, with the real interest rate unaffected in all other periods – this result will resurface several times in future paragraphs. From the next period on, the economy returns to the steady state, so that all the variables will be equal to zero (here, all changes in variables are relative to the steady state).

The aggregate Euler equation (10) implies that, in response to a decline in the real interest rate, $i_t - E_t\{\pi_{t+1}\}$, aggregate consumption increases more in TANK than in RANK. According to (19) and (20), consumption of unconstrained agents increases by less $(1-\Phi<1)$ and that of constrained agents increases by more $(\chi > 1)$ than aggregate consumption.

To understand the individual responses, note the following: first, only unconstrained agents are forward looking, so any changes in the economy are driven by their response to changes in the real interest rate, and second, absent transfers, constrained income depends only on labor income. A real interest rate cut reduces the price of consuming today, which, in turn, generates an increase in unconstrained demand (as seen form their Euler equation (A7)). To meet demand, firms hire more labor. Also, wages go up (as $w_t = (\sigma+\varphi)\,c_t$). Since prices are sticky, dividend income falls. As labor income increases, constrained agents' demand shifts outward, and the resulting increase in their consumption is equal to the change in labor income. For unconstrained agents, higher labor income shifts their demand out, while lower dividend income shifts their demand back in.[9] Since unconstrained agents are always on their Euler equation, in equilibrium, their consumption increases. In sum, since the constrained agent does not experience the decrease in dividend income, a real interest rate cut generates a larger increase in constrained consumption than unconstrained consumption. This explains why aggregate demand is higher in TANK than in RANK.

[9] As discussed in Bilbiie (2008), with a larger share of constrained agents, the effect of constrained agents on dividend income and, in turn, on a shift of unconstrained demand, is larger.

Sensitivity of aggregate consumption to individual consumption. In the case of no productivity shocks, equation (19) implies $c_t = \frac{1}{1-\Phi} c_t^U$, where $\frac{1}{1-\Phi}$ is the elasticity of aggregate consumption with respect to unconstrained consumption.

According to part (i) of Proposition 1, an increase in λ raises the elasticity $\frac{1}{1-\Phi}$. What is the mechanism behind this outcome? Increasing the share of constrained agents in the population amplifies the aggregate consumption response to an interest rate cut because constrained agents consume their entire labor income each period. As a result, aggregate consumption becomes more sensitive to unconstrained consumption.

For part (ii), reducing τ decreases the elasticity $\frac{1}{1-\Phi}$. To see why, compare two cases: $\tau = 1$ (no redistribution) and $\tau < 1$. In the latter case, when dividends fall in response to the real interest rate cut, constrained agents experience the fall in transfer income. Therefore, part of their increase in demand, due to higher labor income, is offset by lower transfer income. Hence, compared to the case of $\tau = 1$, constrained consumption increases by less with $\tau < 1$, while the response of the unconstrained agent is the same in both cases. Therefore, aggregate consumption is less sensitive to a unit change in unconstrained consumption with $\tau < 1$ than $\tau = 1$. In other words, the elasticity $\frac{1}{1-\Phi}$ increases with τ. For part (ii) about δ, an increase in δ is equivalent to a reduction in τ in terms of the amount of dividend income available to constrained agents, which is why the elasticity decreases with δ.

From part (iii), decreasing τ^W increases the elasticity $\frac{1}{1-\Phi}$. Compare two cases: $\tau^W = 0$ (no redistribution) and $\tau^W < 0$. Decreasing τ^W below zero results in a positive net transfer of labor taxes to constrained agents and a negative net transfer to unconstrained agents. Unlike the case of profit transfers, labor-tax transfers work in the same direction as the earned income of constrained agents, reinforcing the increase in their demand due to the interest rate cut. Consequently, constrained income increases relative to the case of $\tau^W = 0$. Aggregate consumption becomes more sensitive to increases in unconstrained consumption.

3.3 Real- and complex-root solutions and indeterminacy: TANK versus RANK

Bounds on four types of solutions. Above, we characterize the effect of Φ on the slope of the aggregate demand curve (10). Additionally, Φ affects the bounds on $\phi_{E\pi}$ leading to different types of solutions presented in Table 1,

$$\phi_{E\pi}^1 \equiv \frac{\phi_y}{\kappa}(\beta - 1) + 1 - \phi_\pi, \tag{21}$$

$$\phi_{E\pi}^{2,3} = \sigma(1-\Phi)\left[\frac{\beta+1}{\kappa} \pm 2\frac{\beta}{\kappa}\sqrt{\frac{1}{\beta} + \frac{\kappa}{\beta}\frac{1}{\sigma(1-\Phi)}\left(\phi_\pi + \frac{\phi_y}{\kappa}\right)}\right] + \frac{\beta}{\kappa}\phi_y + 1, \tag{22}$$

$$\phi_{E\pi}^4 \equiv 2\frac{\beta}{\kappa}\sigma(1-\Phi)[\beta+1] + \phi_\pi + \left(\frac{1+\beta}{\kappa}\right)\phi_y + 1. \tag{23}$$

The solution is indeterminate when either the central bank is "too hawkish" in targeting expected inflation, $\phi_{E\pi} > \phi_{E\pi}^4$, or when it is "too dovish" in doing so, $\phi_{E\pi} < \phi_{E\pi}^1$. Determinate complex roots arise when the central bank responds moderately to expected inflation, $\phi_{E\pi}^2 < \phi_{E\pi} < \phi_{E\pi}^3$, with $\phi_{E\pi}^2 > \phi_{E\pi}^1$ and $\phi_{E\pi}^3 < \phi_{E\pi}^4$. Note that the lower indeterminacy bound $\phi_{E\pi}^1$ does not depend on Φ and hence, is the same in RANK and TANK. The other three bounds, however, do include the term $1 - \Phi$ and thus, are affected by the presence of heterogeneity.[10]

Calibrated example. We further investigate how the degree of asset market participation, λ, impacts the regions of different types of solutions in a calibrated example. Figure 1 considers three parameterizations of asset market participation: RANK ($\lambda = 0$), TANK with $\lambda = 0.21$, and TANK with $\lambda = 0.38$. The value

[10] When the central bank does not target expected inflation ($\phi_{E\pi} = 0$) but does target current inflation, the bounds for determinacy can be found by setting $\phi_{E\pi}^1 = 0$ and $\phi_{E\pi}^4 = 0$ and by solving for ϕ_π. In this case, the solution is unique whenever $\phi_\pi > \frac{\phi_y}{\kappa}(\beta-1)+1$ and $\phi_\pi > -2\frac{\beta}{\kappa}\sigma(1-\Phi)[\beta+1] - \left(\frac{1+\beta}{\kappa}\right)\phi_y - 1$. Since we focus on the case of $\Phi < 1$, the second restriction is satisfied for all combinations of heterogeneity parameters we consider. In this case, the determinacy regions do not depend on heterogeneity, so we omit this case from further discussion in the following section.

$\lambda = 0.21$ is assumed in Debortoli and Gali (2018), while the value $\lambda = 0.38$ is close to the empirical estimate in Campbell and Mankiw (1989) and to the one used in Kaplan et al. (2018).

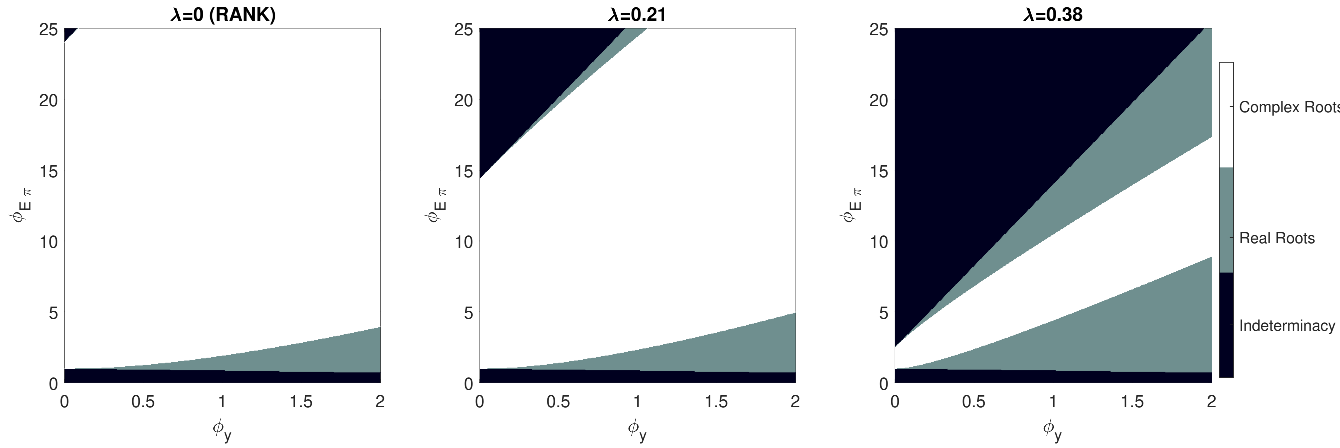

Figure 1: Types of solutions.

We distinguish three areas: a unique solution with real roots (in gray), a unique solution with complex roots (in white), and indeterminacy (in black). The figure presents comparisons in $(\phi_y, \phi_{E\pi})$ space, and we set $\phi_\pi = 0$. The tendencies are as follows: First, TANK is characterized by substantially larger areas of indeterminacy than RANK due to a negative relation between Φ and $\phi^4_{E\pi}$; see (23). For example, $\phi_{E\pi} = 3$ leads to indeterminacy if $\phi_y = 0$, $\phi_\pi = 0$ and 38% of agents are excluded from asset markets. Second, a larger λ leads to more extensive unique-solution areas with real roots. Third, the complex-root region is smaller in TANK than in RANK, but the complex root region is still relevant as it corresponds to the most empirically relevant values of $\phi_{E\pi}$ and ϕ_y; e.g., under $\lambda = 0.21$, values of $\phi_{E\pi}$ above unity imply complex roots.[11] However, the aggregate predictions of the model are similar when we consider $\phi_{E\pi}$ near the boundary that distinguishes between real and complex roots.

Next, we consider how transfers affect the upper bound on $\phi_{E\pi}$ that leads to determinacy. Figure 2 shows how $\phi^4_{E\pi}$, on the left axis, and $1 - \Phi$, on the right axis, change as the transfer parameters τ (first panel) and τ^W (second panel) are varied. In each panel, we consider the cases of $\lambda = 0.21$ (orange line) and $\lambda = 0.38$ (blue line); regions of the parameter space where $\Phi > 1$ (i.e., where the aggregate Euler equation slopes upward) are in gray.

[11] Complex-root solutions can generate large oscillations which should not necessarily be ignored. Beaudry et al. (2019) argue that a new Keynesian model, augmented to include complementarities through financial frictions, is capable of generating low-frequency business cycles consistent with the data where models with real roots do not. In their model, complex-root solutions can generate stronger internal propagation of exogenous shocks, while not leading to explosive oscillations. Therefore, whether complex roots solutions are relevant depends on what features of the economy one hopes to explain.

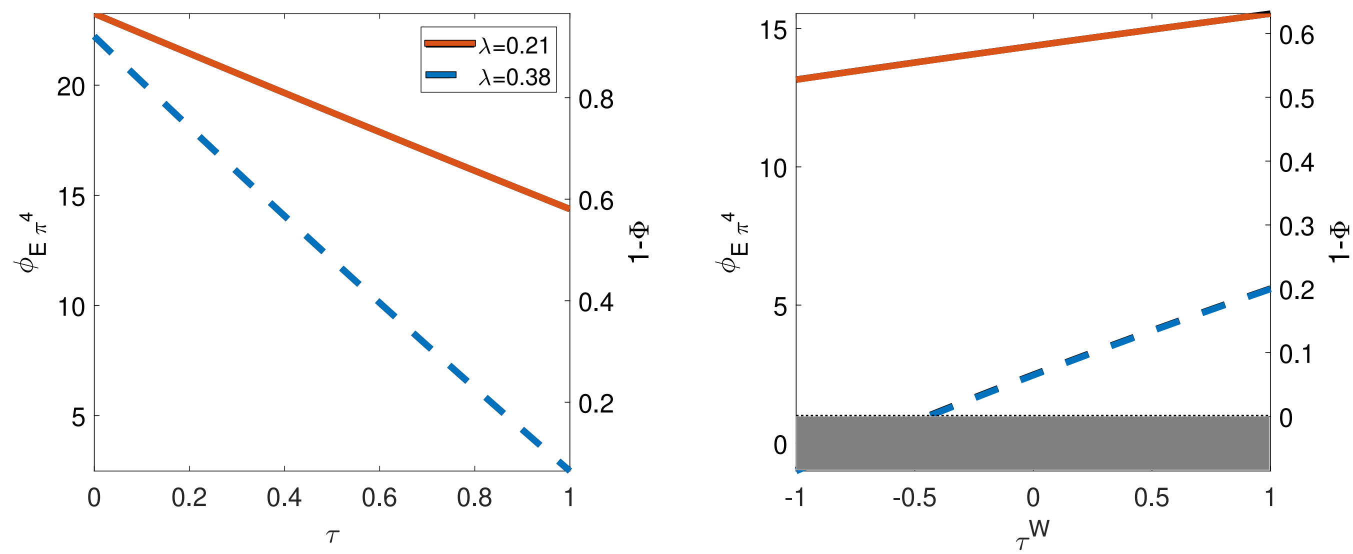

Figure 2: The effect of transfers on $\phi^4_{E\pi}$ and $1-\Phi$.

From the first panel we see that, with $\lambda = 0.38$, $\phi^4_{E\pi}$ is above 10 so long as $\tau < 0.6$. Therefore, even with low levels of asset market participation, the upper bound is empirically irrelevant so long as there is enough redistribution of illiquid profits to constrained agents. Additionally, the slopes of the lines differ for different levels of asset market participation. Therefore, the effect of a marginal increase τ on the elasticity of aggregate consumption to unconstrained consumption depends on the share of asset market participation. Specifically, the effect is larger when λ is larger.

In the second panel, we see that the blue line crosses the boundary of the gray region when $\tau^W \approx -0.5$. In other words, when λ is large and there is a high degree of redistribution of labor tax revenues towards constrained agents, the aggregate Euler equation will be upward sloping. Moreover, there is much less of a difference in the slopes of two lines in the case of labor tax revenue transfers than illiquid profit transfers. This is due to the fact that labor tax revenue transfers work in the same direction of earned income of constrained agents, while illiquid profit transfers work in the opposite direction.

According to Figures 1 and 2, the solution is indeterminate in TANK with $\lambda = 0.38$ for any $\phi_{E\pi} > 2$. Such a low upper bound is problematic for the model, implying tight restrictions on the central bank's response to expected inflation. The value of $\phi^4_{E\pi} = 2$ corresponds to $1 - \Phi = 0.14$. That is, the indeterminacy regions in TANK where agents have logarithmic utility over consumption and where heterogeneity parameters result in $1 - \Phi = 0.14$ are the same as the regions in RANK where the agent has an inverse elasticity of intertemporal substitution coefficient of 0.14.

Why do RANK and TANK differ in values of $\phi_{E\pi}$ leading to indeterminacy? Smaller values of $\phi_{E\pi}$ lead to indeterminacy in TANK relative to RANK. The key feature of TANK that explains this result is that, whenever dividends fall relative to steady state, their fall has an amplified effect on labor supply in TANK relative to RANK. In RANK, we have $\Phi = 0$, so the income effect on labor supply – the second term in (18) – is determined by σ. In TANK, however, there are two additional channels. First, when $\Phi > 0$, the income effect is multiplied by an additional term $\frac{1}{1-\Phi}$, which appears due to the fact that dividends are distributed unequally and asset market participation is limited in TANK. The second channel works through c^U_t. In equilibrium, one unconstrained agent holds $\frac{1}{1-\lambda}$ shares. If dividends change by $\triangle d_t$, each unconstrained agent experiences a change of $\frac{\triangle d_t}{1-\lambda}$ in non-labor income. Therefore, holding wages fixed, the shift in labor supply from a change in dividend income is greater under TANK. Moreover, the size of the shift increases with Φ.

Formula (18) implies that labor supply shifts outward when dividends fall. If the shift in labor supply is large enough, equilibrium wages will fall. Since wages are equal to marginal costs, and firms reduce prices when marginal costs fall, profits must be procyclical when falling dividends generate large shifts in labor supply. Note, as well, that $\phi_{E\pi}$ determines the equilibrium relationship between inflation and dividends.[12] Also, with larger values of $\phi_{E\pi}$, greater changes in dividends are needed for the returns on bonds and shares to be equal. Since dividends have an amplified effect on labor supply in TANK, a smaller $\phi_{E\pi}$ is needed in TANK than in RANK for falling dividends to generate a shift in labor supply such that wages would fall in equilibrium. In other words, a smaller $\phi_{E\pi}$ leads to procyclical profits in TANK than in RANK.

In response to a real rate cut, unconstrained agents increase demand. Moreover, with procyclical profits, the increase in dividend income generates additional rounds of consumption. In other words, unconstrained agents face increasing returns in dividend income. The increasing returns on unconstrained consumption results in an indeterminate solution as the process of increasing demand and dividends continues to infinitum.

4 Forward guidance

In this section, we study the effects of heterogeneity on aggregate and individual responses to forward guidance. We model forward guidance as a one-time shock in the central bank's policy rule, and we allow for active policy. With active policy, the equilibrium sequence of real interest rates, $\{(i_t - E_t\{\pi_{t+1}\})\}_{t=0}^{\infty}$, is determined endogenously, and the real rate changes in every period leading up to the shock.

4.1 Aggregate responses to future shocks

Forward guidance regarding real rates. To illustrate why forward guidance is powerful at the lower bound in the baseline new Keynesian model, McKay et al. (2016) consider an experiment where the central bank directly sets the real interest rate. The real rate is set to zero in all periods except period T, when the rate is cut by 1%. McKay et al. (2016) use the experiment to show why RANK suffers from the *forward guidance puzzle* predicting that output growth and inflation respond excessively to future real rate cuts and that cumulative responses of macroeconomic variables increase with how far in the future rate cuts take place; see also Del Negro et al. (2015), Carlstrom (2015), Maliar and Taylor (2018), Campbell et al. (2019).[13]

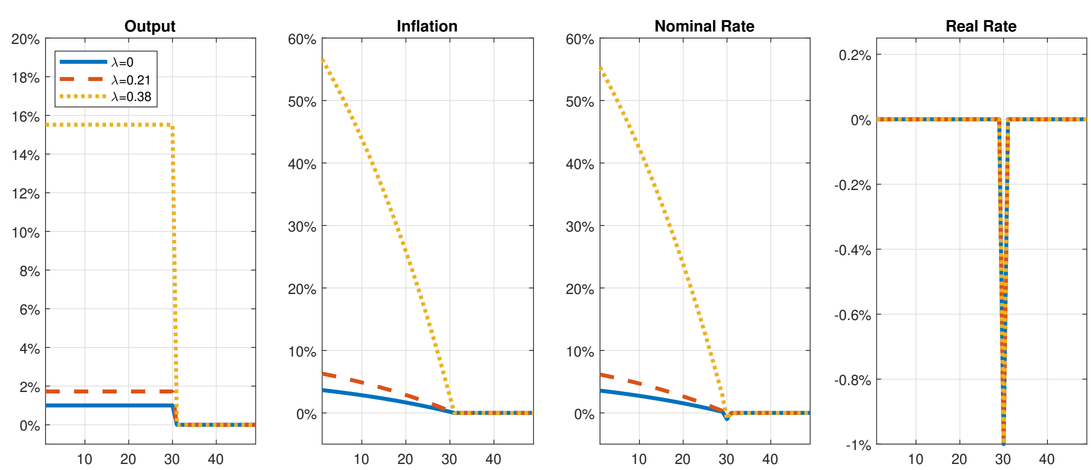

Figure 3: Aggregate impulse response for the Taylor rule with $\phi_{E\pi} \searrow 1$, $\phi_y = 0$ and $\phi_\pi = 0$.

[12] To see it, assume, for simplicity, that the central bank only targets expected inflation with $\phi_{E\pi} > 1$, which implies that the real return on bonds is $(\phi_{E\pi} - 1)\, E_t\{\pi_{t+1}\}$. The return on bonds and shares must be equal, and hence, the value of $\phi_{E\pi}$ determines the relationship between inflation and dividends.

[13] Gabaix (2019) and Husted et al. (2017) consider this experiement as well and show that bounded rationallity and monetary policy uncertainty can resolve the forward guidance puzzle.

We consider McKay's et al. (2016) experiment in the context of TANK; see Figure 3. We do so by using the Taylor rule $i_t = i_t^* + E_t\{\pi_{t+1}\} + v_t$. The figure displays the response of aggregate variables for $T = 30$ and for different values of the share of constrained agents, $\lambda \in \{0, 0.21, 0.38\}$. Our results show that heterogeneity substantially amplifies the effect of the future monetary-policy shock and worsens the puzzle even further; see the case of $\lambda = 0.38$. The response of consumption at $t < T$ is the same as the response to a contemporaneous real interest rate cut, which occurs in period T. From the discussion of Proposition 1, increasing λ raises the elasticity of aggregate consumption to unconstrained consumption and results in a larger aggregate response to a contemporaneous real interest rate cut. Therefore, the responses of economic aggregates increase with λ. An increase in τ or a decrease in $\left\{\delta, \delta^W, \tau^W\right\}$ raises the elasticity and has the same directional effect as an increase in λ.

Forward guidance with active policy. In normal times, monetary policy is active, and forward guidance policy does not result in the central bank directly setting the path for the real interest rate. Instead the central bank affects the real interest rate in all periods leading up to period T. Using forward guidance in normal times has increased since the Financial Crisis of 2007–2009. Inevitably, using forward guidance in formal policy statements while the economy is at the effective lower bound requires forward guidance away from the effective lower bound.[14] Statements about how monetary policy "plans to remain accommodative", which were put in place at the lower bound, will no longer be applicable as the economy recovers and require removal from formal policy statements. Removing statements provides guidance on how the central bank plans to conduct future policy while current policy is conducted in a conventional manner. Empirical studies of forward guidance in normal times, e.g., include, Gürkaynak et al. (2005) and Campbell et al. (2012). Maliar and Taylor (2018) study forward guidance along with active monetary policy rules in normal times in the context of RANK.

In Figure 5, we consider the Taylor rule parameterized by $\phi_{E\pi} = 0$, $\phi_y = \frac{.25}{4}$ and $\phi_\pi = 1.5$. As discussed in Maliar and Taylor (2018), when forward guidance is used along with active policy, RANK's dynamics have common-sense properties: the future shock does not have immediate effects, and the effects only become strong when approaching to the date of the shock. As is seen from Figure 5, the implications in TANK are similar. Additionally, heterogeneity increases the peaks of positive and negative output gaps and delays the period at which the first noticeable responses of output and inflation occur.

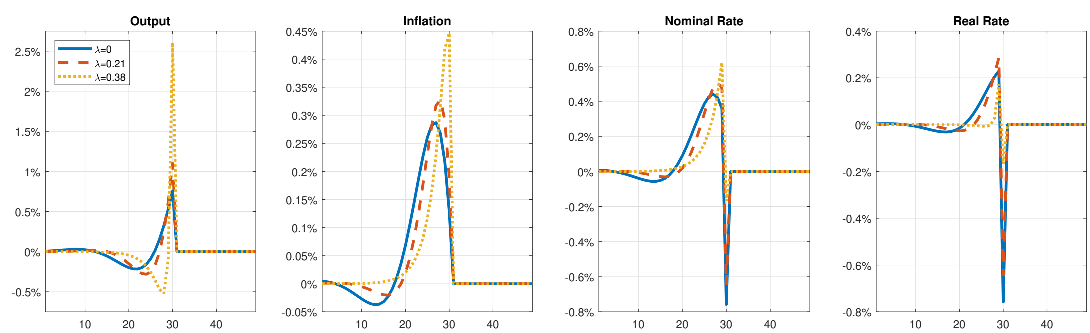

Figure 4: Aggregate impulse response for the Taylor rule with $\phi_{E\pi} = 0$, $\phi_y = \frac{.25}{4}$ and $\phi_\pi = 1.5$

What is the intuition for the results in Figure 4? If unconstrained agents were to increase consumption by a large amount in period zero in response to a shock in period $T > 0$, there would be a large increase in the real interest rate due to the fact that monetary policy is active. The rise in the real interest rate would induce unconstrained agents to reduce consumption, pushing back against the initial increase in unconstrained demand. Therefore, unconstrained agents increase consumption by a small amount which

[14] The Federal Reserve used forward guidance four times between 2018 and early 2020; *see https://www.federalreserve.gov/monetarypolicy/timeline-forward-guidance-about-the-federal-funds-rate.htm.*

results in a small increase in output today. When λ is large, there is an ample response in period T because it is a response to contemporaneous shock, and limiting asset market participation amplifies the response of the economy to a contemporaneous shock. Therefore, there is only a small increase in output in period zero but a large increase in period T.

Interaction between heterogeneity and active policy. To better illustrate the role of heterogeneity, we consider a Taylor rule that differs from the one in Figure 3 in the value of ϕ_y. Specifically, Figure 5 considers the rule with $\phi_{E\pi} \searrow 1$ and $\phi_y = 0.01$ and $\phi_\pi = 0$.

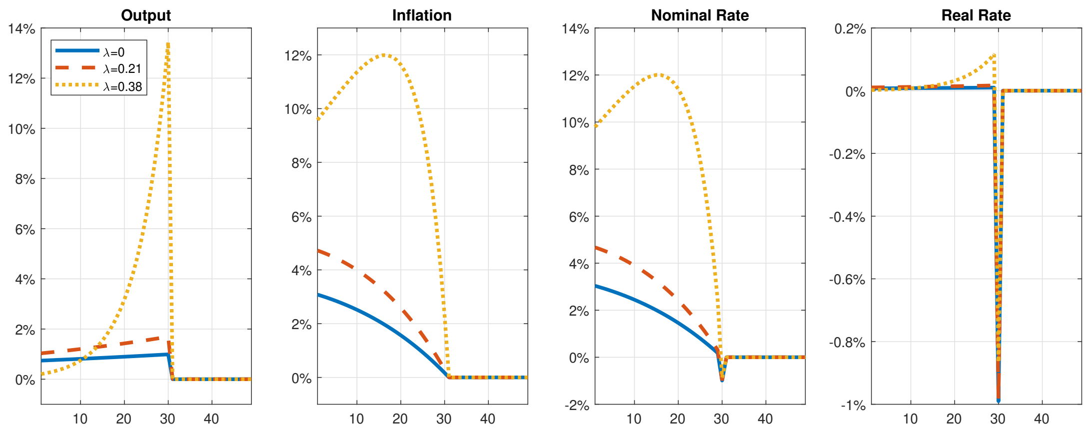

Figure 5: Aggregate impulse response for the Taylor rule with $\phi_{E\pi} \searrow 1$, $\phi_y = 0.01$ and $\phi_\pi = 0$.

As is seen, even a small value of ϕ_y helps get rid of constant jump in output. Heterogeneity brings two effects here: first, it reduces the effect of the future shock on today's output, and second, it increases its effect at the peak. When 38% of agents are excluded from asset markets, the response of consumption today to a shock in period 30 is only 1.5% of the peak response. In RANK, the response today is 75% of the peak. Therefore, future nominal interest rate shocks have a discounted impact on today's economy. The response of output differs from the case with $\phi_y = 0$ because the real interest rate is not constant prior to the shock. The fact that the central bank actively responds to changes in the economy through the policy rule leads to small today's responses to future policy shocks.

The reason future monetary policy shocks have a discounted impact on today's economy when monetary policy is active is easily seen by considering the formula for the consumption response displayed in panel 1 of Figure 6. After substituting in the Taylor rule and solving the aggregate Euler equation forward, we obtain

$$c_t = -\sum_{k=0}^{\infty} \widetilde{\beta}^k \frac{1}{\sigma(1-\Phi)} \left[i^*_{t+k} + v_{t+k} \right] - \frac{\sigma(1-\Phi)}{\phi_y + \sigma(1-\Phi)} E_t\{z_t\}, \tag{24}$$

where

$$\widetilde{\beta} = \frac{\sigma(1-\Phi)}{\phi_y + \sigma(1-\Phi)}.$$

In equation (24), c_t is written as a function of only exogenous shocks. The term $\widetilde{\beta}$ is an effective discount factor applied to the future monetary policy shock.[15] If $\phi_y \neq 0$, then $\widetilde{\beta}$ is less than unity so long as $\Phi < 1$ (downward sloping demand). In this case, any amount of output-gap targeting generates discounting of

[15] This differs from the case of Euler equation discounting where the discount factor is applied to the real interest rate; see, e.g., McKay et al. (2017).

future monetary-policy shocks.[16] The term $\widetilde{\beta}$ captures the fact that the central bank's policy affects the real interest rate in every period when the bank targets the output gap.

The composed parameter Φ influences the effect of the monetary-policy shock in (24) through $\widetilde{\beta}$ and $\frac{1}{\sigma(1-\Phi)}$. Note that $\widetilde{\beta}$ decreases in Φ and $\frac{1}{\sigma(1-\Phi)}$ increases with Φ. First, when $k > 0$, an increase in Φ results in smaller effects of future monetary-policy shocks on today's economy as the $\widetilde{\beta}^k$ term dominates the $\frac{1}{\sigma(1-\Phi)}$ term. Second, when $k = 0$, we are left with $\frac{1}{\sigma(1-\Phi)}$, so that the peak effect increases in Φ. Therefore, reducing asset market participation or reducing transfers of illiquid profits to constrained agents results in a smaller initial response and larger peak response to a future monetary policy shock.

In sum, we show that the TANK model amplifies the aggregate economy's response in the period of the monetary policy shock as the share of constrained agents increases. This finding is similar to that of Auclert et al. (2018), Debortoli and Gali (2018), and Bilbiie (2019a). Our novel finding is that heterogeneity, in conjunction with an active Taylor rule, helps generate sensible time-zero output responses to forward guidance.

4.2 Individual responses to future shocks

In this section, we explore individual responses to forward guidance policy.

First, assume the general case of the Taylor rule (9). As before, consider a forward guidance shock announced at $t = 0$ and realized at $t = T$. For $t > T$, consumption of both types of agents is zero, $c_t^U = 0$ and $c_t^K = 0$. For $t \leq T$, the impulse response functions of unconstrained and constrained consumption for the case of two distinct real roots are

$$c_t^U = -\frac{1}{\sigma}\left[\mathcal{H}_{t,T} + \upsilon_T\right], \tag{25}$$

$$c_t^K = \chi\left\{-\frac{1}{\sigma(1-\Phi)}\left[\mathcal{H}_{t,T} + \upsilon_T\right]\right\}, \tag{26}$$

where $\mathcal{H}_{t,T}$ is defined as

$$\mathcal{H}_{t,T} \equiv \frac{X_T}{m_1 - m_2}\sum_{k=0}^{T-t}\left[\left(\frac{\phi_y}{\kappa}+\phi_\pi\right)\left(m_1^{t+k-1-T} - m_2^{t+k-1-T}\right) + \left(\phi_{E\pi} - 1 - \frac{\beta\phi_y}{\kappa}\right)\left(m_1^{t+k-T} - m_2^{t+k-T}\right)\right],$$

and it captures the endogenous feedback between the central bank's policy rule and the economy. The term $\mathcal{H}_{t,T}$ makes the consumption responses of the two agents dependent on the period T when the future shock occurs. Furthermore, $\mathcal{H}_{t,T}$ makes the unconstrained agent's response dependent on the heterogeneity parameters through m_1, m_2 and Φ. Finally, $\mathcal{H}_{t,T}$ makes the response of both agents explicitly depend on the parameterization of the monetary policy rule.

Second, consider a central bank that uses the Taylor rule with $\phi_{E\pi} \searrow 1$, $\phi_y = 0$ and $\phi_\pi = 0$. As previously noted, this specification results in the real interest rate changing in only a single period. Substituting $\phi_{E\pi}$, ϕ_y, and ϕ_π into the above general formulas for c_t^U and c_t^K, the term $\mathcal{H}_{t,T}$ reduces to zero so the responses of unconstrained and constrained consumption, respectively, become

$$c_t^U = -\frac{1}{\sigma}\upsilon_T, \tag{27}$$

$$c_t^K = -\chi\cdot\frac{1}{\sigma(1-\Phi)}\upsilon_T. \tag{28}$$

The above equations imply that t-period consumption of both constrained and unconstrained agents are proportional to the future shock and independent of the period of the future shock T. Therefore, a monetary policy shock 30 periods in the future produces the same response as a shock that were to occur

[16] When $\widetilde{\beta} > 1$, there is compounding – the opposite of discounting. Note that positively sloped aggregate demand, $\Phi > 1$, is not a sufficient condition for compounding.

tomorrow. In this case, the real interest rate follows an exogenously set path. As a result, there is no feedback between the real interest rate and economic activity, which is why $\mathcal{H}_{t,T}$ reduces to zero. Hence, consumption of unconstrained agents does not depend on heterogeneity as heterogeneity does not affect the real interest rate. When there is an expansionary monetary-policy shock, values of heterogeneity parameters $\{\lambda, \tau, \delta, \delta^W, \tau^W\}$ that lead to a larger Φ lead to a greater response by the constrained agent. Therefore, changes in the underlying structure of the heterogeneity can either amplify or dampen the initial response depending on whether the changes increase or decrease Φ.

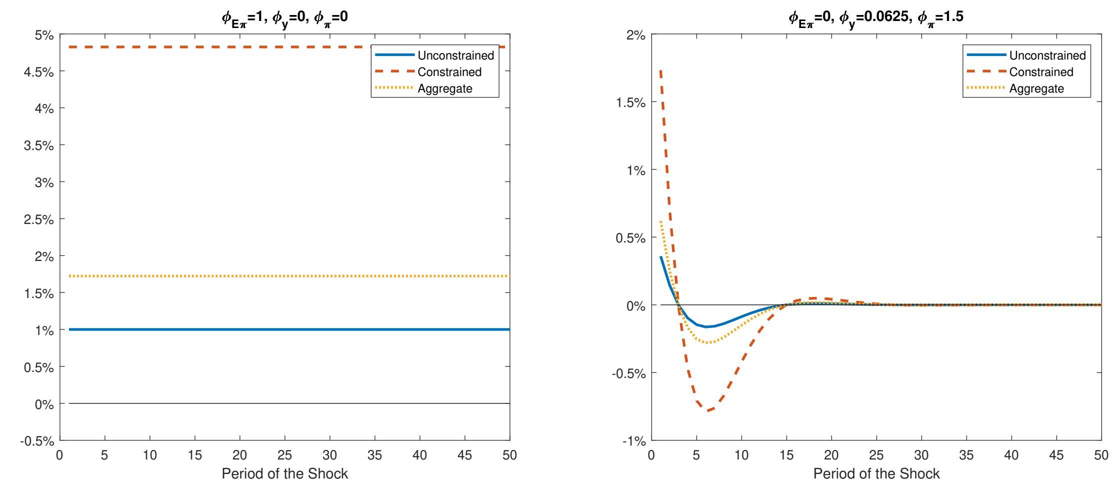

Figure 6: Time-zero consumption responses to the future 1% monetary policy shock for different horizons of the future shocks.

Figure 6 plots time-zero individual and aggregate consumption responses to a -1% monetary policy shock in period T for $T \in \{0, ..., 50\}$. The figure considers two parameterizations of the Taylor rule: $\phi_{E\pi} \searrow 1$, $\phi_y = 0$ and $\phi_\pi = 0$ in the first panel and $\phi_{E\pi} = 0$, $\phi_y = \frac{.25}{4}$ and $\phi_\pi = 1.5$, in the second panel. In the first panel, the response of consumption is the same regardless of the period of the shock. This is due to the fact that the response in any period prior to the shock is the response to a contemporaneous shock as discussed in the previous section. Additionally, the response of the constrained agent exceeds that of the unconstrained agent as the constrained agent does not experience the decline in dividend income. In the second panel, the same idea holds in that the constrained agent reacts more to the shock than the unconstrained agent. Moreover, a future nominal interest rate shock can lead to an economic contraction today. Therefore, if the shock has a contractionary effect on today's economy, consumption of the constrained agent decreases by more than that of the unconstrained agent. To check the robustness of our results, we perform the same experiments for the case of differentiated labor and obtain similar patterns of time-zero responses. Overall, the results suggest that policymakers should consider multiple horizons when using models to evaluate the potential effects of forward guidance in normal times.

5 Transmission of monetary policy

In this section, we study how monetary policy transmits through the economy by decomposing the total effects of monetary-policy shocks into direct and indirect effects. Direct effects capture the response of consumption when the real interest rate changes, keeping income fixed, while indirect responses reflect the change in consumption due to a change in income keeping the interest rate fixed.

The decomposition has received much attention in the HANK literature. First, Kaplan et al. (2018) demonstrate that RANK is characterized by large direct effects, which suggests a lack of general equilibrium feedback in response to monetary-policy shocks. In contrast, in a two-asset quantitative HANK model, indirect effects make up the majority of the response, which returns the general equilibrium feedback mechanisms to new Keynesian models. Second, Bilbiie (2019a) studies the decomposition of responses to persistent contemporaneous shocks in TANK for the case of the Taylor rule $i_t = E_t\{\pi_{t+1}\} + v_t$. The previous literature focuses only on contemporaneous monetary-policy shocks. We extend the previous results by considering both contemporaneous and future shocks for a more general family of Taylor rules.

5.1 Closed-form solutions for direct and indirect effects

Using our closed-form solutions, we derive analytical formulas for the direct and indirect effects under Taylor rule (9); see Appendix A.4 for details. The effects can be computed from the planned expenditure curve, given by

$$c_t = \underbrace{\vartheta_j^1 x_t}_{\text{Indirect}} + \underbrace{\vartheta_j^2 \left[i_t - E_t\{\pi_{t+1}\}\right]}_{\text{Direct}}, \tag{29}$$

$$\vartheta_j^1 \equiv \frac{\left[1-(1-\Phi)(1-\varsigma)\beta\right]}{1-(1-\Phi)(1-\varsigma)\beta\varrho_j}, \qquad \vartheta_j^2 \equiv -\frac{1}{\sigma}\left(\frac{\beta(1-\varsigma)}{1-(1-\Phi)(1-\varsigma)\beta\varrho_j}\right), \tag{30}$$

where ϑ_j^1 and ϑ_j^2 are, respectively, multipliers of the indirect and direct effects of a shock occurring j periods from now; $\varsigma \equiv \frac{\lambda(1-\Gamma)}{(1-\lambda)+\lambda(1-\Gamma)}$ is a constant weight of constrained consumption in aggregate consumption, $c_t = (1-\varsigma)\,c_t^U + \varsigma c_t^K$; ϱ_j is a scalar governing the relationship between consumption tomorrow and consumption today, $c_{t+1} = \varrho_j c_t$, and it is a non-linear function of how far in the future the shock occurs; see Table 5 in the appendix for ϱ_j in the considered cases.

Using the Taylor rule (9) to substitute out i_t and the Phillips curve (11) to substitute out x_t, the indirect and direct effects become

$$\text{Indirect:} \quad \vartheta_j^1 \cdot \frac{1}{\kappa}\left[\pi_t - \beta E_t\{\pi_{t+1}\}\right], \tag{31}$$

$$\text{Direct:} \quad \vartheta_j^2 \cdot \left[i_t^* + \left(\frac{\phi_y}{\kappa} + \phi_\pi\right)\pi_t + \left(\phi_{E\pi} - 1 - \frac{\beta\phi_y}{\kappa}\right)E_t\{\pi_{t+1}\} + v_t\right], \tag{32}$$

where π_t and π_{t+1} are given by our closed-form solutions in Table 1. Therefore, we have indirect and direct effects written in terms of exogenous shocks and parameters only.

5.2 Quantifying direct and indirect effects

In Tables 2 and 3, we quantify direct and indirect effects in a calibrated example. We consider three cases: (1) a one-time transitory monetary-policy shock for a unique solution with two real roots; (2) a one-time transitory shock for an indeterminate solution; (3) a one-time persistent shock that reverts back to zero with persistence ρ_v for a unique solution. For all three cases, we consider both contemporaneous shocks ($j = 0$) and future shocks ($j \geq 1$). We focus on a 1% expansionary monetary-policy shock occurring in period $t + j$. The results are computed under the assumptions of $\phi_{E\pi} = 0$, $\phi_y = \frac{.25}{4}$ and $\phi_\pi = 1.05$ for the unique solutions (cases (1) and (3) in the table), and $\phi_{E\pi} = \phi_y = \phi_\pi = 0$ for the indeterminate solution (case (2) in the table).

General tendencies: TANK versus RANK. Table 2 reports the shares of the direct and indirect effects in total, as well as total response, for $\lambda \in \{0.21, 0.08, 0\}$.

The general tendencies are as follows: First, in response to a contemporaneous monetary-policy shock ($j = 0$), direct and indirect effects both work to increase current consumption. Second, in response to a future monetary policy shock ($j = 1$, 5), direct and indirect effects work in opposite directions in both RANK and TANK.[17] Finally, for a future shock, indirect effects work to decrease consumption when the

[17] At some horizons not reported in the table, direct and indirect effects might work in the same direction.

Table 2: Direct and indirect effects in the baseline model

	Types of solutions and shock								
	Unique solution, transitory shock			Indeterminacy, transitory shock			Unique solution, persistent shock		
	(1)			(2)			(3)		
	$j = 0$	$j = 1$	$j = 5$	$j = 0$	$j = 1$	$j = 5$	$j = 0$	$j = 1$	$j = 5$
$\lambda = 0.21$									
Share in total, in%									
Direct	45.9	-64.9	-1100.95	63.64	-167.89	-167.89	29.79	-39.07	251.02
Indirect	54.1	164.9	1200.95	36.36	267.89	267.89	70.21	139.07	-151.02
Aggregate response, in %									
Total	1.2	0.82	0.09	0.61	0.35	0.04	1.81	2.24	0.08
$\lambda = 0.08$									
Share in total, in%									
Direct	77.84	8626.07	652.26	87.93	195.88	195.88	63.73	-394.72	509.01
Indirect	22.16	-8526.07	-552.26	12.07	-95.88	-95.88	36.27	494.72	-409.01
Aggregate response, in %									
Total	0.91	0.70	0.24	0.44	0.28	0.04	1.48	-0.925	0.37
$\lambda = 0$									
Share in total, in%									
Direct	97.50	111.55	109.93	98.78	105.06	105.06	95.12	114.58	109.47
Indirect	2.50	-11.55	-9.93	1.22	-5.06	-5.06	4.88	-14.58	-9.47
Aggregate response, in %									
Total	0.8	0.65	0.05	0.38	0.25	0.05	1.34	1.04	0.04

Notes: Direct and indirect effects sum up to 100%. Total effect represents a percentage change in today's consumption due to a decrease in the interest rate at $t + j$.

share of constrained agents is low ($\lambda = 0$ or $\lambda = 0.08$) and to increase consumption when the share is high ($\lambda = 0.21$).

Let us explain the last result for $j = 1$. In response to a one-period ahead expansionary monetary-policy shock, expected period-1 consumption rises. First consider RANK. Since income rises in period 1 and the current real interest rate is fixed, the present value of period-1 income increases. With higher consumption in period 1 and higher present value of period-1 income, the unconstrained agent would like to borrow. However, with bonds in zero net supply, the agent is unable to. Therefore, current income must change for the agent to choose to hold zero bonds. This requires a decrease in current income, to compensate for the increase in the present value of future income. The decline in their current income comes about due to a reduction in demand. In this case, labor income falls and dividend income rises, with the net effect being a decline in income. Therefore, the indirect effect is negative.

Now consider TANK. The key feature in TANK is that consumption of constrained agents impacts income of unconstrained agents. Since constrained agents do not experience a fall in dividend income, there is a smaller increase in the present value of period-1 unconstrained income in TANK. To generate the necessary change in current income for the unconstrained agent to choose zero bonds, unconstrained agents must increase demand. The reason unconstrained agents increase demand is that decreasing demand would generate too large of a fall in unconstrained period-0 income to be consistent with unconstrained agents choosing zero bonds. When unconstrained agents reduce demand, labor income falls for both unconstrained and constrained agents. The reduction in constrained demand reinforces the initial decline in labor income. Therefore, due to the spillovers from constrained demand to unconstrained income, a reduction in unconstrained demand would not be consistent with the agent choosing zero bonds as the fall in current income would be too large. Therefore, unconstrained demand, and, in turn, consumption, rises, which is why the indirect effect is positive.

When there are not enough constrained agents, the spillover from the constrained agents to unconstrained income is not sufficient for the unconstrained agent to choose to increase consumption. There is a threshold level of constrained agents such that the indirect effect is negative for all levels of constrained agents below that threshold. When all other parameters are kept at their baseline levels, the threshold is around $\lambda = 0.08$. Above the threshold, the effect constrained agents have on aggregate demand is large enough to generate a positive indirect effect. Below the threshold, future monetary-policy shocks work much like they do in RANK, even though contemporaneous shocks do not. Therefore, only considering how contemporaneous monetary policy shocks transmit through the economy is not sufficient for understanding how future shocks transmit.

Aggregate MPC. Table 3 contains the multiplier ϑ_j^1 of the indirect effect under the baseline calibration of $\lambda = 0.21$.

Table 3: Aggregate MPC

Shock horizon	Types of solutions and shocks		
	Unique solution, transitory shock	Indeterminacy, transitory shock	Unique solution, persistent shock
	(1)	(2)	(3)
$j = 0$	0.5410	0.3636	0.7021
$j = 1$	1.6490	2.6789	1.7241
$j = 5$	12.9546	2.6789	-1.5102

Notes: ϑ_j^1 is the multiplier of the indirect effect.

Note that ϑ_j^1 can be interpreted as the aggregate marginal propensity to consume (MPC) out of a change in current income induced by the monetary-policy shock. Therefore, the multiplier cannot be considered as an aggregate MPC out of a general change in income. Rather, it is specific to the timing and type of the shock generating the change in income.

In all cases, we find that the aggregate MPC's are non-monotonic in t and can be greater than unity. The latter occurs when ϱ_j is also greater than unity; see (30). Thus, future consumption, c_{t+1}, is expected to be greater than current consumption, c_t. When unconstrained agents expect future consumption to exceed current consumption, they will try to bring some of that consumption to the present. As unconstrained consumption rises, constrained income and therefore, constrained consumption also rises. Therefore, absent any adjustments in the real interest rate, current consumption rises by more than current income. This leads to a multiplier greater than unity.

We now consider the three cases (1)–(3) appearing in Tables 2 and 3 in detail.

A transitory shock in case of a unique solution. We first consider the case of a unique solution. For a contemporaneous monetary-policy shock ($j = 0$), the output-gap persistence ϱ_j is zero since variables are only functions of current and future shocks; see equations (A12), (A13); hence, $c_{t+1} = 0$. Consequently, in the case of future shocks ($j \geq 1$), ϱ_j does depend on the horizon of the shock.

Consider TANK with $\lambda = 0.21$ in Table 2. First, at nearly all horizons, the indirect effects in TANK exceed the direct effects and are much larger than those in RANK. Second, while the relative sizes of the direct and indirect effects deviate from each other over time, the total effect decreases with the horizon of the shock; very large positive and very large negative effects generate relatively small total effects. Finally, the increase in consumption in response to future shocks in TANK is due to large indirect effects, while the increase in RANK is due to large direct effects. Therefore, just as in the case of contemporaneous shocks, future shocks can transmit through the economy differently in TANK than in RANK.

We analyze now the special case of the Taylor rule with $\phi_{E\pi} \searrow 1$, $\phi_y = 0$ and $\phi_\pi = 0$ (this case does not appear in Table 2) which leads to the central bank directly setting the real interest rate. Because $v_t = 0$ and $i_t^* = 0$, the direct effects of future shocks are absent. Therefore, the share of direct effects in the total response jumps discretely from some positive value for $j = 0$ to 0% of the total effect for $j > 0$. Since the aggregate MPC is constant (one will get $\varrho_j = 1$ after substituting in for $m_1 = \frac{1}{\beta}$ and

$m_2 = 1$), the change in income is the same regardless of the period of the shock, and the current-period consumption response is the same for all horizons. The result suggests that the resolutions to the forward guidance puzzle (proposed by other papers) must generate indirect effects that decrease with the horizon of the shock.

A transitory shock in case of an indeterminate solution. To fix one equilibrium, we set the constant in the inflation equation in Table 1 to zero ($C = 0$). With an indeterminate solution, inflation is a function of all past and future shocks. As a result, both x_{t+1} and ϱ_0 will also be non-zero. Additionally, for all $j \geq 1$, the output-gap persistence ϱ_j's (and hence, ϑ_j^1's) does not depend on how far in the future the shock occurs: ϱ_j is equal to m_1 – the larger eigenvalue from the difference equation for inflation (13).

As is seen in Table 2, direct effects make up the majority of the response to a contemporaneous monetary-policy shock ($j = 0$). However, the sizes of the direct and indirect effects are the same across different j for $j \geq 1$ due to ϱ_j being the same and $\pi_{t+1} = m_1\pi_t$. However, the sizes of the direct and indirect effects are the same for all $j \geq 1$. The sizes are the same as ϱ_j is the same for all $j \geq 1$, and $\pi_{t+1} = m_1\pi_t$, which follows from case i) of Table 1. Additionally, the indirect effect dominates the direct effect of a future monetary-policy shock. Moreover, similar to Kaplan et al. (2018), we find that the share of direct effects decreases with λ, and in the limit, as $\lambda \to 1$, direct effects disappear: when no one is forward looking, consumption does not respond to interest rates.

A persistent shock in case of a unique solution. Finally, we assume that a deterministic monetary-policy shock occurs at time $t + j$ and reverts back to zero with persistence ρ_v. Kaplan et al. (2018) and Bilbiie (2019a) consider this case under the Taylor rule with $\phi_{E\pi} \searrow 1$, $\phi_y = 0$ and $\phi_\pi = 0$. In contrast to the indeterminacy case, $\varrho_j\ \acute{}s$ and ϑ_j^1's do depend on the horizon of the future shock.

The tendencies regarding direct and indirect effects are similar to those for a shock without persistence. One key difference though is that the aggregate MPC ϑ_j^1 is negative for $j = 5$. Formula (30) implies that ϑ_j^1 is negative when ϱ_j is sufficiently large. Also, a larger ϱ_j is needed in TANK than in RANK for a negative aggregate MPC. Note that $\vartheta_j^1 < 0$ means that an increase in income today would generate a decrease in consumption when interest rates are held fixed. Therefore, aggregate consumption behaves as if it were an inferior good.

Recall that negative aggregate MPCs arise when consumption tomorrow is expected to be much higher than consumption today. If the aggregate MPC is negative, an increase in aggregate income results in a decrease in aggregate consumption. How can this be? With a negative aggregate MPC, the change in income of the unconstrained agents needed for them to choose zero bonds comes about due to a rise in dividend income and fall in labor income, with the net effect being an increase in unconstrained income. In contrast, if the change in income was due to rising labor income and falling dividend income, constrained consumption would also increase. Therefore, the aggregate MPC would be positive. For the MPC to be negative and for both agents to increase consumption when income rises, it must be that the change in income for unconstrained agents works through rising dividends and falling labor income. The fall in constrained consumption due to lower labor income exceeds the increase in unconstrained consumption because of the unitary MPC of constrained agents (even though constrained agents constitute a smaller share of the population) so that aggregate consumption falls. Therefore, the aggregate MPC is negative.

6 Redistribution and economic volatility

In this section, we study the effects of heterogeneity on a stochastic version of the economy.

Closed-form solutions. We begin by constructing closed-form solutions for the stochastic version of the economy using the same strategy as in Maliar and Taylor (2018). The solutions for the case of productivity shocks under the assumption that the productivity level, a_t, follows an AR(1) processes with persistence

ρ_a and variance σ_a^2 are

$$\text{distinct real roots: } \pi_t = \frac{\Upsilon_a}{m_1 - m_2}\left(\frac{a_t}{\rho_a}\right)\left[\sum_{s=t}^{\infty}\left(\frac{m_1}{\rho_a}\right)^{t-1-s} - \sum_{s=t}^{\infty}\left(\frac{m_2}{\rho_a}\right)^{t-1-s}\right],$$

$$\text{complex roots: } \pi_t = \Upsilon_a\left(\frac{a_t}{\eta\rho_a}\right)\left[\sum_{s=t}^{\infty}\left(\frac{r}{\rho_a}\right)^{t-1-s}\sin\left(\theta\left(t-1-s\right)\right)\right],$$

where $\Upsilon_a = \frac{\kappa}{\beta\sigma(1-\Phi)}\frac{1+\varphi}{\sigma+\varphi}\left[\sigma(1-\rho_a)+\phi_y\right]$; see Appendix A.5 for the corresponding solutions for the other shocks, as well as for the solutions for the output gap, consumption of constrained agents and unconstrained agents.

Effect of Φ on aggregate economic volatility. Using our closed-form solutions, we can investigate how the composite parameter affects the fluctuations of economic aggregates. It turns out that output and inflation are more volatile in the economy with a larger value of the composite parameter Φ. The proposition below formalizes the claim.

Proposition 2. Assume $m_1,\ m_2 \in \mathbb{R}$, $|m_1|, |m_2| > 1$ and $\phi_{E\pi} = 0$.[18] Let $\sigma_x(\Phi_i)$ denote the standard deviation of the output gap in an economy i with composite heterogeneity parameter Φ_i. Suppose $\Phi_1 > \Phi_2$ due to differences in one of the heterogeneity parameters, $\left\{\lambda, \tau, \delta, \tau^W, \delta^W\right\}$. If $\phi_\pi > \rho_a$, then $\sigma_x(\Phi_1) > \sigma_x(\Phi_2)$.

Proof. See Appendix A.5. ■

From Proposition 2, we see that changes in the heterogeneity parameters that increase Φ result in greater output-gap volatility. Note that $1-\Phi$ provides us the value for the coefficient of relative risk aversion needed in RANK to generate the same outcome as in TANK when both agents have logarithmic utility over consumption. Therefore, changes in the heterogeneity parameters that increase Φ require a reduction in the relative risk aversion coefficient in a RANK model for TANK and RANK to generate the same results. Since a less risk averse agent will accept higher volatility, we see that volatility increases as $1-\Phi$ decreases. Since $\frac{\partial\Phi}{\partial\tau} > 0$ and $\frac{\partial\Phi}{\partial\tau^W} < 0$, Proposition 2 implies that economies with less redistribution of illiquid profits (larger τ) or more redistribution of labor-income taxes (smaller τ^W) are more volatile. We now provide a quantitative assessment of these implications.

Redistribution and heterogeneous consumption volatility. We now address how fiscal redistribution affects cyclical behavior of the economy. Transfers of either illiquid profits or labor tax revenues have little effect on the cyclicality of consumption inequality. However, transfers do affect economic volatility at the individual and aggregate levels.

In Table 4, we compare the ratios of standard deviations of individual and aggregate consumption under alternative transfer schemes to the benchmark case of no redistribution ($\tau = 1$ and $\tau^W = 0$). In particular, our alternative transfer schemes vary either τ to $\tau = 0.9$ or τ^W to $\tau^W = -0.1\left(\frac{1-\lambda}{\lambda}\right)$ – both cases represent a 10% deviation from no redistribution towards full redistribution to constrained agents. We compare the ratios under two values of the share of constrained agents, $\lambda \in \{0.21,\ 0.38\}$.

We find that transferring a larger share of illiquid profits to constrained agents decreases the volatility of constrained consumption. When illiquid profits are transferred, the income of the constrained agent is diversified, depending on both wages and profits, and thus, consumption volatility for the constrained agent decreases. Transferring a larger share of labor-tax revenues to the constrained agent has the opposite effect. Unlike illiquid-profit transfers, labor-tax revenue transfers provide no diversification benefit. When more

[18] The assumption about m_1, m_2 corresponds to case ii) in Table 1.

Table 4: Volatility of constrained and unconstrained consumption relative to benchmark

	Relative standard deviation of consumption		
	Unconstrained	Constrained	Aggregate
$\lambda = .21$, $\tau = .9$ & $\tau^W = 0$	1.004	0.9380	0.9987
$\lambda = .38$, $\tau = .9$ & $\tau^W = 0$	0.9811	0.9316	0.9965
$\lambda = .21$, $\tau = 1$ & $\tau^W = -.1\left(\frac{1-\lambda}{\lambda}\right)$	0.9987	1.0182	1.0004
$\lambda = .38$, $\tau = 1$ & $\tau^W = -.1\left(\frac{1-\lambda}{\lambda}\right)$	1.0133	1.0342	1.0020

Notes: The results are averages over 100 simlations, each of which has 50 periods.

income is transferred to constrained agents, demand is more volatile, which results in more volatile labor income. Hence, constrained consumption is more volatile with both small and large shares of constrained agents.

Consumption volatility of unconstrained agents increases as more illiquid profits are transferred to constrained agents when the share of constrained agents is low ($\lambda = 0.21$) and decreases when the share of constrained agents is high ($\lambda = 0.38$). While the resources that the unconstrained agent has available for smoothing consumption decrease, smoother consumption for the constrained agent, and hence, more steady profits, can result in smoother consumption for the unconstrained agent. This is because when only a small share of the population does not have access to assets, the ability for unconstrained agents to smooth consumption is only slightly affected by the constrained agent's decisions. As the share of constrained agents increases, and aggregate demand becomes more volatile, the consumption decisions of constrained agents have a greater impact on the unconstrained agent. Consequently, when the share of constrained agents is high, the indirect benefit of smoother income outweighs the direct decrease in ability to smooth consumption for unconstrained agents and results in a decrease in consumption volatility. Moreover, increasing transfers of labor-tax revenues decreases consumption volatility of the unconstrained agent. For the unconstrained agent, the decrease in income, which is tied to wages, reduces the relatively more volatile income source. Therefore, while the unconstrained agent has fewer resources available to smooth consumption, the reduction in income volatility results in their lower consumption volatility.

As is seen from the table, volatility of aggregate consumption increases in τ and decreases in τ^W, in line with Proposition 2. However, quantitative changes in volatility at the aggregate level are relatively small under all transfer schemes. The small effect at the aggregate level is due to the fact that the majority of agents in the economy are unconstrained and actively try to smooth consumption. Therefore, while their ability to smooth consumption is reduced by transferring illiquid profits, the relative change in volatility at the aggregate level is small.

7 Consumption and income inequality

In this section, we discuss the predictions about consumption and income inequality. In the model, consumption is equal to income, $c_t = y_t$, and hence, we can talk about consumption and income inequality interchangeably. Coibion et al. (2017) study how inequality responds to monetary policy shocks and find that both consumption and income inequality decline in response to expansionary monetary policy shocks. Krueger et al. (2010) find that consumption and income inequality are countercyclical, with income inequality being more countercyclical. While the current model can not reconcile both facts, we can assess whether the current model can predict the countercyclical nature of consumption and income inequality.

Consumption gap. The consumption gap is given by

$$\gamma_t = -(\sigma + \varphi)\Psi x_t \tag{33}$$
$$= \Psi \mu_t, \tag{34}$$

where $\Psi > 0$ is defined in (17); $\mu_t \equiv \ln \frac{\mathcal{M}_t}{\mathcal{M}}$ is the markup. According to (33) and (34), when the output gap is above steady state or markups below steady state, the consumption gap is below steady state.

Inequality responses to demand shocks. To investigate how demand shocks affect inequality, we follow the HANK literature by considering a one-time contemporaneous real interest rate cut. From the aggregate Euler equation (10), the interest rate cut increases the output gap x_t, which reduces consumption inequality, as is seen from (33). This result is known in the literature as *countercyclical consumption inequality*; see Bilbiie (2008, 2019a). Note, however, that this name is not related to the economy's business cycle fluctuations (supply shock driven fluctuations).

What drives countercyclical consumption inequality in the model? As was argued in Section 3.2, in the absence of redistribution, constrained consumption depends solely on labor income, while unconstrained consumption depends both on labor income and profits. When there is a real interest rate cut, demand increases, which drives labor income above steady state and profits below steady state. Therefore, while the incomes of both agents increase, constrained income increases by more since it is unaffected by the fall in profits. The greater increase in income for constrained agents results in a greater increase in their consumption, producing countercyclical consumption inequality.

Equation (34) provides another way to see why consumption inequality is countercyclical; namely, the model generates countercyclical markups μ_t in response to demand shocks. Indeed, in this model $w_t = \sigma c_t + \varphi n_t$ and $y_t = a_t + n_t$ (see the formulas in Appendix A.1), which imply that markups are

$$\mu_t = a_t - w_t = (1 + \varphi)\, a_t - (\sigma + \varphi)\, y_t,$$

so that μ_t falls whenever y_t goes up (holding a_t constant). Countercyclicality of markups does not depend on the deterministic nature of shocks, and markups μ_t are countercyclical when demand shocks follow the stochastic processes described in Section 2. However, empirically, markups do not appear to be countercyclical; see Ramey and Nekarda (2013). Therefore, one limitation of the current model is the inability to jointly match the empirical evidence on the cyclicality of income inequality and markups.

Inequality responses to productivity shocks. We now extend our analysis to include aggregate productivity shocks. To the best of our knowledge, the existing literature has not studied whether TANK models generate increases or decreases in income and consumption inequality in response to productivity improvements (supply shocks). As mentioned above, empirical evidence suggests that consumption and income inequality are countercyclical over the business cycle. It is easy to see that the model predicts the opposite, predicting procyclical inequality. Indeed, according to (34), the consumption gap is a constant positive share of the markup. Since markups are increasing in the level of productivity, a positive productivity shock increases consumption inequality. This counterfactual implication remains true in a version of the model with differentiated labor supply across types, which also predicts that the consumption gap is a constant share of the markup. In Section 8, we show that this model's drawback is overcome by modifying asset-market structure.

To better understand why consumption inequality is procyclical in the model, we investigate how productivity shocks affect individual consumption. Without productivity shocks, an increase in aggregate income y_t (equivalently, c_t) raises consumption of constrained agents by more than one-to-one as $\chi > 1$, while with productivity shocks, the consumption response of constrained agents is smaller. The smaller response implies that the increase in productivity has a partial negative affect on consumption of constrained agents captured by the second term in (20) (although an increase in productivity raises aggregate income y_t). The negative effect is due to the fact that the benefits of higher productivity are distributed unequally between the two agents. Constrained agents do not benefit from the positive effect productivity has on dividend income. When all profits are illiquid and distributed equally between the two agents, $\Phi \to 0$ and $\chi \to 1$. With $\chi = 1$ the second term in (20) drops out. With less than equal distribution, constrained agents experience a negative partial effect from productivity increases, while unconstrained agents experience a positive effect; see (19).

Note that, in response to a one-time productivity shock with zero persistence, we have

$$\frac{dc_t^U}{da_t} - \frac{dc_t^K}{da_t} = \underbrace{[\chi - (1-\Phi)]}_{>0} \underbrace{\left[\frac{1+\varphi}{\sigma+\varphi} - \frac{\partial y_t}{\partial a_t}\right]}_{>0} > 0,$$

since $\chi \geq 1$, $\Phi < 1$, and $\frac{\partial y_t}{\partial a_t} < \frac{1+\varphi}{\sigma+\varphi}$; see the appendix for a verification.

The term in the first set of brackets is the difference between the elasticities of unconstrained and constrained consumption to aggregate income in the case of no productivity shocks; see (19) and (20). It is positive because constrained consumption is more sensitive to aggregate income than unconstrained consumption when there are no productivity shocks. The second term is the negative of the rate at which the markup changes as productivity changes, normalized by the rate at which the markup changes as output increases (i.e.,$-\left[\frac{d\mu_t}{da_t}\right]\left[\frac{d\mu_t}{dy_t}\right]^{-1}$). In the absence of productivity changes, markups are countercyclical, $\frac{d\mu_t}{dy_t} < 0$, and they increase with productivity, $\frac{d\mu_t}{da_t} > 0$, so that the second term in brackets is positive. Therefore, while higher productivity raises output, the consumption gap also increases.

We use our numerical solutions to verify that consumption inequality is procyclical when productivity is stochastic. We consider an economy subject to exogenous productivity shocks with persistence $\rho_a = .9$ and variance $\sigma_a^2 = .016$. We conduct $1,000$ simulations of $10,000$ periods and find an average correlation between the consumption gap and productivity of 0.33. Therefore, the above results from the one time transitory shock carry over to the case where productivity is stochastic. In contrast to the empirical evidence, the model implies that consumption inequality is procyclical over the business cycle.

Summary. First, the consumption gap coincides with the income gap. Second, the behavior of inequality in the baseline model is entirely determined by markups. Finally, consumption inequality is procyclical when there are productivity shocks, and it is countercyclical when there are demand shocks.

inequality	consumption	$\gtreqless$	income
determinants	markups μ		markups μ
demand shock	*countercyclical*	=	*countercyclical*
supply shock	*procyclical*	=	*procyclical*

8 Sensitivity results: TANK with capital

We now address whether capital and adjustment costs can reverse the counterfactual predictions about inequality generated by the model without capital. We do not focus on the effects of forward guidance policy here because we find that the model with investment adjustment costs has similar predictions to the model with no capital; we present these results in Appendix B.4. Galí et al. (2004) consider a version of TANK with capital but without redistribution among agents of different types.

8.1 The model

The economy consists of a continuum of agents of two types, intermediate-good producers, final-good producers, a central bank and a government. A fraction λ of agents are constrained and cannot hold shares in the firms, bonds or capital, while the remaining agents are unconstrained, and they have full access to all three asset markets.

Consumer side. The problem of the constrained agent remains the same as in the baseline TANK model and is given by (3), (4). Unconstrained agents can save via a mutual fund which holds the intermediate-good firms' dividends, bonds and physical capital. Physical capital is subject to investment adjustment costs. The unconstrained agents rent capital and receive a rental rate R_t^k. The tax revenue from capital is distributed in the same fashion as illiquid profits. Specifically, an unconstrained agent chooses $\left\{C_t^U, N_t^U, B_t^U, F_t^U, K_t^U, I_t^U\right\}$ to maximize (1) subject to the following constraints:

$$C_t^U + \frac{B_t^U}{P_t} + Q_t F_t^U + I_t^U$$
$$= \frac{B_{t-1}^U R_{t-1}}{P_t} + \left(1-\delta^W\right)\frac{W_t}{P_t}N_t^U + (1-\tau_k)R_t^k K_{t-1}^U + (Q_t + (1-\delta)D_t)F_{t-1}^U + T_{D,t}^U + T_{W,t}^U, \quad (35a)$$

$$K_t^U = (1-\delta^k)K_{t-1}^U - \frac{\zeta}{2}\left(\frac{I_t^U}{K_{t-1}^U} - \delta^k\right)^2 K_{t-1}^U + I_t^U, \quad (36)$$

where initial condition $(B_{-1}^U, F_{-1}^U, K_{-1}^U)$ is given. Here, C_t^U, N_t^U, B_t^U, F_t^U, K_t^U, I_t^U, $T_{D,t}^U, T_{W,t}^U$ are the unconstrained agent's consumption, labor, nominal bond holdings, shares of intermediate-good firms, capital stock, investment, transfers of illiquid profits and capital tax revenues, and transfers of labor tax revenues, respectively; D_t is the dividend from ownership of intermediate-good firms; P_t, Q_t, W_t, R_t^k and R_{t-1} are the final-good price, share price, nominal wage, rate of return on physical capital and (gross) nominal interest rate, respectively; Z_t is a preference shock, following the standard AR(1) process in logs, $Z_{t+1} = Z_t^{\rho_Z}\exp(\varepsilon_{Z,t+1})$ with $\varepsilon_{Z,t+1} \sim \mathcal{N}\left(0, \sigma_Z^2\right)$; τ_k is a tax on returns from capital; a parameter $\zeta \geq 0$ captures the cost of adjusting investment; $\beta \in (0,1)$ is the subjective discount factor; $\delta^k \in [0,1]$ is the depreciation rate of capital; $\delta \in [0,1]$ is the share of illiquid profits allocated across agents by government; $\delta^W \in [0,1]$ is the labor-income tax rate; $\sigma > 0$ and $\varphi > 0$. The FOCs of the unconstrained agent's problem (1), (35a), (36) are derived in Appendix B.1.

Supply side. The supply side in this TANK economy is similar to the one in TANK without capital. The only difference now is that a firm i produces an intermediate good $Y_t(i)$ using two inputs, capital $K_t(i)$ and labor $N_t(i)$, according to

$$Y_t(i) = A_t K_t(i)^\alpha N_t(i)^{1-\alpha},$$

where $\alpha \in (0,1)$; A_t is a productivity level following the standard AR(1) process in logs, $A_{t+1} = A_t^{\rho_A}\exp(\varepsilon_{A,t+1})$ with $\varepsilon_{A,t+1} \sim \mathcal{N}\left(0, \sigma_A^2\right)$. As before, a firm's price is rigid a la Rotemberg (1982) by setting prices optimally in every period of time subject to the same adjustment-cost function.

Redistribution. Government makes transfers $\left\{T_{D,t}^U, T_{W,t}^U, T_{D,t}^K, T_{W,t}^K\right\}$ by re-distributing the illiquid profits of the intermediate-good producers δD_t, capital-tax revenues $\tau_k R_t^k K_t$ and labor-tax revenues $\delta^W \frac{W_t}{P_t}N_t$, so that $(1-\lambda)T_{D,t}^U + \lambda T_{D,t}^K = \delta D_t + \tau_k R_t^k K_t$ and $(1-\lambda)T_{W,t}^U + \lambda T_{W,t}^K = \delta^W \frac{W_t}{P_t}N_t$, respectively. It does so according to the following rules:

$$T_{D,t}^U = \left(1 + \frac{\tau\lambda}{1-\lambda}\right)\left(\delta D_t + \tau_k R_t^k K_t\right), \quad T_{D,t}^K = (1-\tau)\left(\delta D_t + \tau_k R_t^k K_t\right), \quad (37)$$

$$T_{W,t}^U = \left(1 + \frac{\tau^W\lambda}{1-\lambda}\right)\delta^W \frac{W_t}{P_t}N_t, \quad T_{W,t}^K = \left(1-\tau^W\right)\delta^W \frac{W_t}{P_t}N_t, \quad (38)$$

where $1-\tau$ and $1-\tau^W$ are, respectively, the share of profits plus capital-tax revenues and the share of labor-tax revenues, distributed to the constrained agents, with $\tau \in [0,1]$ and $\tau^W \in \left[-\frac{1-\lambda}{\lambda}, 1\right]$. All the log-linearized equations of the model are summarized in the appendix.

8.2 Consumption and income inequality

To the best of our knowledge, the new Keynesian model with capital does not allow for closed-form solutions. To solve the model numerically, we use Fair and Taylor's (1983) extended path method. The log-linearized equations are presented at the end of Appendix B.3.

Consumption gap. As before, we define Γ_t as a consumption gap, $\Gamma_t = 1 - \frac{C_t^K}{C_t^U}$. Since consumption and income no longer coincide, we also define the income gap as, $\Gamma_t^Y = 1 - \frac{Y_t^K}{Y_t^U}$, where Y_t^K and Y_t^U are income of constrained and unconstrained agents, respectively. In Appendix B.2, we derive the nonlinear formulas for the consumption and income gaps.

By log-linearizing the gaps around a zero inflation steady state, we obtain

$$\gamma_t = \theta_1 \mu_t + \theta_2 \hat{\imath}_t + \theta_3 r_t^k + \theta_4 k_t, \tag{39}$$

$$\gamma_t^Y = \theta^Y \mu_t, \tag{40}$$

where $\hat{\imath}_t$ is the deviation of investment from steady state, γ_t^Y is the (log-linear) income gap, and the θ's are constants, defined in Appendix B.3. These income and consumption gaps apply to any standard adjustment cost function which results in a steady state investment to capital ratio of δ.[19]

The cyclicality of income inequality is fully determined by the cyclicality of the markup; see (40). With countercyclical markups, the income gap will decrease in expansions. Therefore, if the model generates countercyclical income inequality, it also has the undesirable feature of countercyclical profits just as in the case of the model without capital. This should be kept in mind when analyzing richer HANK models. It may be that the reason some HANK models can match the countercyclicality of income inequality, consistent with the empirical evidence, is that the model produces countercyclical markups in response to demand shocks, inconsistent with the empirical evidence.

Unlike the model without capital, consumption inequality depends on factors other than the markup; see (39). Specifically, investment, the rental rate of capital and the level of capital all affect consumption inequality. Recall that the θ's are applicable to adjustment cost functions that result in a steady state investment to capital ratio of δ. Thus, if two adjustment costs generate different responses of consumption inequality to shocks, the difference is due to the dynamics of markups, investment, the rental rate of capital and the capital stock, not the steady state levels.

Under the adjustment cost function in (35a), the consumption gap can be rewritten as

$$\gamma_t = \varkappa_1 \mu_t + \varkappa_2 \hat{q}_t + \varkappa_3 r_t^k, \tag{41}$$

where $\varkappa_1$,$\varkappa_2$ and $\varkappa_3$ are constants defined in Appendix B.3; $\hat{q}_t$ is the value of installed capital in terms of consumption, and r_t^k is the return on physical capital (both in log deviations). When $\hat{q}_t$ is high, having an additional unit of installed capital is more valuable in terms of consumption than purchasing an additional unit of uninstalled capital. In this case, we would therefore expect to see consumption of the unconstrained agent to fall and investment to increase. The deviation of the value of installed capital from the price of the investment good is key in the cyclicality of consumption inequality in our model.

In the baseline TANK model with no capital, the consumption gap is a constant fraction of the markup. As a result, a positive demand shock that reduces the markup decreases the consumption gap, while a positive supply shock that raises the markup increases the consumption gap. For the TANK model with capital, the prediction is not so clear. The sign of $\varkappa_2$ is opposite to those of $\varkappa_1$ and $\varkappa_3$; see Appendix B.2. Therefore, the model with capital and adjustment costs can have either increasing or decreasing consumption inequality. For example, a productivity shock increases markups μ_t, the value of an additional unit of installed capital $\hat{q}_t$ and the rental rate of capital r_t^k. Therefore, without knowing the magnitudes of deviations, we cannot say whether the consumption gap increases or decreases. Higher returns on capital, which only accrue to unconstrained agents increase the consumption gap, while higher values of installed capital call for increases in investment, reducing the consumption gap.

[19] The derivation also requires sticky prices. In a model with wage instead of price rigidity, the consumption and income gaps would differ from those presented here as the markup is always constant when prices are flexible and hence, would not enter the log-linear income and consumption gaps.

Inequality responses to demand shocks. As discussed in the model without capital, the empirical studies find that consumption inequality is more countercyclical than income inequality. We now address whether capital and adjustment costs allow the model to replicate this fact.

To determine the cyclicality of consumption and income inequality in response to demand shocks, we consider a 1% one-time, contemporaneous real interest rate cut. We find that consumption inequality is more countercyclical than income inequality. The income gap falls by almost 8%, while the consumption gap falls by about 12%. While consumption and income inequality both decline due to the fall in markups, consumption inequality is also affected by changes in investment and the return on capital. The net effect of these two works to reduce consumption inequality. However, countercyclical income inequality is due to countercyclical markups. Therefore, the present TANK model with capital and adjustment costs is unable to reconcile the two empirical findings regarding income inequality and markups.

Inequality over the business cycle. In this section, we address consumption and income inequality over the business cycle (i.e., when there are exogenous stochastic supply shocks). We assume that the productivity shock is parameterized by an AR(1) process with persistence $\rho_a = .9$ and variance $\sigma_a^2 = .016$. From equation (41), we see that the magnitude of the investment adjustment costs plays a key role in whether positive productivity shocks increase or decrease the consumption gap between households with access to savings and those without. In the simulation, we consider a tax rate of 20% on capital and a Taylor rule with $\phi_\pi = 1.5$ and $\phi_y = \frac{.25}{4}$. We keep $\lambda = 0.21$, $\delta = 0.92$, and $\delta^W = 0.15$ as in the model without capital. We vary $\left(\tau, \tau^W\right) \in \left\{(1,0), (.9,0), \left(1, -.1\left(\frac{1-\lambda}{\lambda}\right)\right)\right\}$. The steady state consumption gaps range from $33\% - 37\%$ and the steady state income gaps range from $42\% - 46\%$. The steady state gaps are largest when there are no transfers and smallest when there are labor-tax revenue transfers.

We set the adjustment-cost parameter $\zeta = 6$. It implies the ratio of standard deviations of investment to output is 2.75, which is in line with U.S. data. Figure 8 presents the consumption gap, income gap and productivity level for a 50 period simulation of the economy under this parameterization of the adjustment cost for the three different transfer schemes; all variables are measured in deviations from steady state. We find that the model can generate countercyclical consumption inequality even without engaging in direct redistribution. The countercyclical nature of the consumption gap is driven by the savings decision of the unconstrained household.

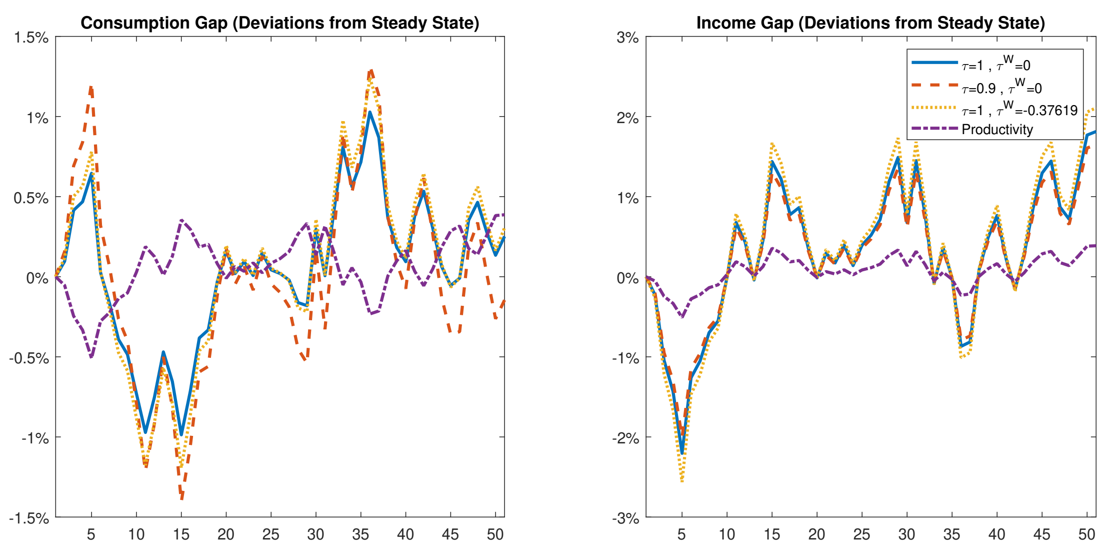

Figure 7: Consumption and income inequality over the business cycle in TANK with capital.

In our sensitivity experiments, we find that larger adjustment costs result in procyclical consumption inequality (these experiments are not reported). Intuitively, with a larger adjustment cost parameter, marginal costs of investment are larger, so the value of installed capital needs to greatly deviate from steady state for investment to be optimal. If the deviation is small, the unconstrained agent will consume instead of investing. In that case, consumption inequality would be procyclical.

While consumption inequality can be procyclical or countercyclical, we find that income inequality is always highly procyclical. Since markups are procyclical over the business cycle in this model, income inequality increases when the economy experiences a positive productivity innovation. Indeed, the correlation of the income gap with productivity is above 0.99 in all transfer schemes (see Table 6 in Appendix B.4).

Consumption inequality, however, is affected by the transfer schemes. Under no asset and labor-tax revenue transfers ($\tau = 1$ and $\tau^W = 0$), the correlation between the consumption gap and productivity is -0.3612. Therefore, the model is able to generate countercyclical consumption inequality. When we only allow for transfers of illiquid profits and capital-tax revenues (i.e., asset transfers) to constrained agents, we find that consumption inequality becomes even more countercyclical. Indeed, the correlation between the consumption gap and productivity increases to -0.5940 when each constrained agent is given one-tenth of the illiquid profits and capital tax revenues ($\tau = .9$ and $\tau^W = 0$).

As argued above, the model generates highly procyclical income inequality due to the fact that the income gap can be written solely as a function of markups when both agents are assumed to supply the same amount of labor. Nevertheless, if we allow for differentiated labor, income inequality will depend positively on markups and negatively on the difference in labor supplied by the two types of agents. However, this further increases the procyclicality of income inequality.

Summary. First, in response to demand shocks, consumption and income inequality continues to be countercyclical, as in the baseline model with no capital. Second, in response to supply shocks, consumption inequality can be either procyclical, acyclical or countercyclical, but income inequality is strongly procyclical. Finally, the behavior of income inequality continues to be completely determined by markups, while that of consumption inequality is governed by a joint effect of markups, value of installed capital and interest rate. Such a joint effect depends on the adjustment cost function assumed.

inequality	consumption	$\substack{> \\ = \\ <}$	income
determinants	μ, $\hat{q}$, r^k		μ
demand shock	*countercyclical*	$>$	*countercyclical*
supply shock	*pro*$-$, *a*$-$, *countercyclical*		*procyclical*

9 Conclusion

Our results provide insight into what mechanisms are at work in quantitative HANK models. First, the effect transfers have on individual and aggregate outcomes depend crucially on the degree of asset market participation and how transfers are funded. Transfers that diversify income of constrained agents have effects opposite to transfers that result in greater concentration. When designing transfers, one must keep in mind how the redistribution policy affects agents directly, through their tax paid and transfer received, and indirectly, through changes in income due to changes in the behavior of other agents.

Second, we show forward guidance away from the lower bound becomes less effective as more agents are excluded from asset markets. This suggests the effects of expansionary and contractionary forward guidance may not be symmetric in HANK models since the share of constrained agents is endogenous. Contractionary forward guidance may increase the share of constrained agents while expansionary policy may have opposite effects.

Furthermore, our results on direct and indirect effects show that even if the transmission of current shocks in TANK differs from that in RANK, as was shown in the previous literature, transmission of future shocks may be similar to RANK. Therefore, when thinking about the transmission of forward guidance in HANK, it is not enough to just consider the transmission of contemporaneous shocks.

Finally, our results on inequality over economic cycles provide insight into what drives the gaps in consumption and income between those agents at the very top of the wealth distribution, whose borrowing constraint never binds, and very bottom of the wealth distribution, whose borrowing constraint always binds, in HANK models. Our results show that the income gap between these two types of agents is due solely to the markup. Consequently, if a simulated HANK model correctly predicts that the income ratio of top to bottom income groups is countercyclical, we should keep in mind that it may be due to the model generating countercyclical markups, in contrast with the empirical evidence. Unlike the income gap, the consumption gap depends on the markup, value of installed capital and return on capital. We show that policies that allow for greater capital mobility (i.e., lower adjustment costs) reduce consumption inequality between those at the very top and very bottom. Therefore, in HANK models, lowering adjustment costs may have an even larger effect than the one presented here as fewer agents will fall into a zero-wealth category as adjustment costs decline.

While the model with capital is able to generate countercyclical consumption inequality, income inequality remains highly procyclical in response to supply shocks, which stands at odds with the data. We leave for future research finding specifications of TANK that would lead to countercyclical income inequality. For example, one may consider an extension where agents differ in labor productivity or face separate labor markets. In the latter case, unemployment in each market might respond differently to business cycles, which may allow the model to generate countercyclical income inequality. Adding these assumptions in an analytically tractable way would offer better understanding of mechanisms at work under the hatch in HANK models.

References

[1] Acharya, S. and K. Dogra, (2018). Understanding HANK: insights from PRANK. FRBNY Staff Report 835.

[2] Auclert, A., (2019). Monetary policy and the redistribution channel. American Economic Review 109(6), 2333–2367.

[3] Auclert, A., B. Bardóczy, M. Rognile, and L. Straub, (2019). Using the sequence-space Jacobian to solve and estimate heterogeneous-agent models. Manuscript.

[4] Auclert, A., M. Rognile, and L. Straub, (2018). The intertemporal Keynesian cross. Manuscript.

[5] Auclert, A., M. Rognile, and L. Straub, (2020). Micro jumps, macro humps:monetary policy and business cycles in an estimated HANK model. Manuscript.

[6] Beaudry, P., D. Galizia and F. Portier, (2019). Putting the cycle back into business cycle analysis. American Economic Review 2020, 110(1), 1-47.

[7] Bilbiie, F., (2008). Limited asset markets participation, monetary policy and (inverted) aggregate demand logic. Journal of Economic Theory 140, 162-196.

[8] Bilbiie, F., (2019a). The new Keynesian cross. Journal of Monetary Economics, forthcoming.

[9] Bilbiie, F., (2019b). Monetary policy and heterogeneity: an analytical framework. Journal of Monetary Economics, forthcoming.

[10] Bilbiie. F D. Känzig and P. Surico, (2019c). Capital, income inequality, and consumption: the Missing Link. CEPR working paper DP 14118.

[11] Broer T., N. Hansen, P. Krusell, and E. Öberg, (2020). The new Keynesian transmission mechanism: a heterogeneous-agent perspective. Review of Economic Studies 87 (1), 77–101.

[12] Campbell, J. and N. Mankiw, (1989). Consumption, income, and interest rates: reinterpreting the time series evidence, in: O. Blanchard, S. Fisher (Eds.), NBER Macroeconomics Annual, MIT Press, Cambridge, MA, pp. 185–216.

[13] Campbell, J., C. Evans, J. Fisher and F. Justiniano, (2012). Macroeconomic effects of Federal Reserve forward guidance. Brookings Papers on Economic Activity, Spring, 1-54.

[14] Campbell, J., F. Ferroni, J. Fisher and L. Melosi, (2019). The limits of forward guidance. Journal of Monetary Economics 108, 118-134.

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

[16] Cochrane, J. H., (2017). The new-Keynesian liquidity trap. Journal of Monetary Economics, 92, 47–63.

[17] Coibion, O, Y. Gorodnichenko, L. Kueng, and J. Silvia. (2017). Innocent Bystanders? Monetary Policy and Inequality. Journal of Monetary Economics 88: 70–89.

[18] Debortoli, D. and J. Gali, (2018). Monetary policy with heterogeneous agents: insights from TANK models. Manuscript.

[19] Debortoli, D., J. Gali and L. Gambetti, (2019). On the empirical (ir)relevance of the zero lower bound constraint. NBER working paper 25820.

[20] Del Negro, M., M. Giannoni, and C. Patterson, (2015). The forward guidance puzzle. Working Paper, Federal Reserve Bank of New York.

[21] Fair, R. and J. Taylor, (1983). Solution and maximum likelihood estimation of dynamic nonlinear rational expectations models. Econometrica 51, 1169-1185.

[22] Farhi, E. and I. Werning, (2016). A theory of macroprudential policies in the presence of nominal rigidities. Econometrica 84, 1645-1704.

[23] Gabaix, X. (2019). A behavioral new Keynesian model. Harvard University, working paper.

[24] Galí, J., D. López-Salido, J. Vallés, (2007). Understanding the effects of government spending on consumption. Journal of the European Economic Association 5(1), 227-270.

[25] Guerrieri, V. and Lorenzoni, G. (2017). Credit crises, precautionary savings, and the liquidity trap. Quarterly Journal of Economics, 132, 1427–1467.

[26] Gürkaynak, R., B. Sack, and E. Swanson, (2005). Do actions speak louder than words? The response of asset prices to monetary policy actions and statements. International Journal of Central Banking 1 (1), 55-93.

[27] Hagedorn, M., Luo, J., I. Manovski and K. Mitman, (2019). Forward guidance. Journal of Monetary Economics, 102, 1-23.

[28] Kaplan, G., B. Moll, and G. Violante, (2016). A note on unconventional monetary policy in HANK. Working Paper, Princeton University.

[29] Kaplan, G., B. Moll, and G. Violante, (2018). Monetary policy according to HANK. American Economic Review 108(3), 697-743.

[30] Kekre, R., and M. Lenel, (2020). Monetary policy, redistribution, and risk premia. Working Papers 2020-02, Becker Friedman Institute for Research In Economics.

[31] Krueger, D. F. Perri, L. Pistaferri and G. Violante, (2010). Cross sectional facts for macroeconomists. Review of Economic Dynamics, Elsevier for the Society for Economic Dynamics, vol. 13(1), pages 1-14, January.

[32] Maliar L. and J. B. Taylor, (2018). Forward guidance: is it useful after the crisis? CEPR working paper DP 13383.

[33] McKay, A., E. Nakamura, and J. Steinsson, (2016). The power of forward guidance revisited. American Economic Review, 106(10), 3133-58.

[34] McKay, A., E. Nakamura, and J. Steinsson (2017). The discounted Euler equation: a note. Economica 84, 820–831.

[35] McKay, A. and R. Reis, (2016). The role of automatic stabilizers in the U.S. business cycle. Econometrica 84(1), 141-194.

[36] Melcangi. D. and V. Sterk, (2019). Stock market participation, inequality, and monetary policy. Manuscript.

[37] Oh, H. and R. Reis, (2012). Targeted transfers and the fiscal response to the great recession. Journal of Monetary Economics 59 (S), 50-64.

[38] Taylor, J., (1993). Discretion versus policy rules in practice. Carnegie-Rochester Conference Series on Public Policy, 39, 195–214.

[39] Ramey, V. A, and C. J. Nekarda, (2013). The Cyclical Behavior of the Price-Cost Markup. NBER Working Paper Series, National Bureau of Economic Research, p. 19099,doi:10.3386/w19099.

[40] Ravn, M. and V. Sterk, (2018). Macroeconomic fluctuations with HANK & SAM: an analytical approach. Manuscript.

[41] Rotemberg, J., (1982). Monopolistic price adjustment and aggregate output. Review of Economic Studies, 49 (4), 517–531.

[42] Werning, I. (2015). Incomplete markets and aggregate demand. Working Paper, MIT.

Online appendices

Appendix A. TANK model

In this section, we provide the supplementary results of Sections 2–7.

A.1 Supplementary results of Section 2

First-order conditions (FOCs). The FOCs of the unconstrained agent's problem (1), (2) are

$$\left(F_t^U\right): \qquad Q_t = E_t\left\{\Lambda_{t,t+1}^U\left[Q_{t+1} + (1-\delta)D_{t+1}\right]\right\}, \tag{A1}$$

$$\left(B_t^U\right): \qquad 1 = \beta R_t E_t\left\{\frac{Z_{t+1}}{Z_t}\cdot\left(\frac{C_{t+1}^U}{C_t^U}\right)^{-\sigma}\frac{1}{\Pi_t}\right\}, \tag{A2}$$

$$\left(N_t^U\right): \qquad \left(1-\delta^W\right)\frac{W_t}{P_t} = (C_t^U)^\sigma (N_t^U)^\varphi, \tag{A3}$$

where $\Lambda_{t,t+1}^U \equiv \beta\frac{Z_{t+1}}{Z_t}\cdot\left(\frac{C_{t+1}^U}{C_t^U}\right)^{-\sigma}$; $\Pi_t \equiv \frac{P_{t+1}}{P_t}$ is t-period gross price inflation.

Under the assumption of a symmetric equilibrium, $P_t\left(i\right) = P_t$, $Y_t\left(i\right) = Y_t$, the profit maximization problem of the firm implies the following Phillips curve:

$$\Pi_t\left(\Pi_t - 1\right) = E_t\left\{\Lambda_{t,t+1}^U\left(\frac{Y_{t+1}}{Y_t}\right)\Pi_{t+1}\left(\Pi_{t+1}-1\right)\right\} + \frac{\varepsilon}{\xi}\left(\frac{1}{\mathcal{M}_t} - \frac{1}{\mathcal{M}}\right), \tag{A4}$$

where $\Pi_t \equiv \frac{P_t}{P_{t-1}}$ is gross price inflation; $\Lambda_{t,t+1}^U$ is a stochastic discount factor used by the firm (only the unconstrained agents own shares); $\mathcal{M}_t \equiv \frac{A_t}{W_t}$ is the average gross markup; $\mathcal{M}$ is the markup in the absence of adjustment costs, $\mathcal{M} \equiv \frac{\varepsilon}{\varepsilon-1}$. Price dispersion is defined as $\triangle_t \equiv 1 - \frac{\xi}{2}\left(\Pi_t - 1\right)^2$. Using the fact that aggregate profits are $D_t = Y_t\triangle_t - \frac{W_t}{P_t}N_t$, we get the formula for the consumption gap (8).

Log-linearization of the non-linear model. In the non-linear model, the resource constraint is given by $Y_t -_0^1 X_t(i)di = C_t$, where $Y_t =_0^1 Y_t(i)di$ is aggregate output, and $X_t(i)$ is the firm i's amount needed for price adjustment costs, defined by

$$X_t\left(i\right) = \left(\frac{P_t\left(i\right)}{P_t}\right)^\varepsilon \frac{\xi}{2} Y_t\left(\Pi_t - 1\right)^2.$$

By symmetric equilibrium, $P_t\left(i\right) = P_t$, we have $_0^1 X_t(i)di = \frac{\xi}{2}Y_t\left(\Pi_t-1\right)^2$ so that $Y_t\left(1 - \frac{\xi}{2}\left(\Pi_t - 1\right)^2\right) = C_t$. We define price dispersion $\triangle_t$ as $\triangle_t \equiv 1 - \frac{\xi}{2}\left(\Pi_t - 1\right)^2$. Hence, we obtain $Y_t\triangle_t = C_t$. In the first-order approximation, $\triangle_t$ is equal to unity, which implies $y_t = c_t$.

Log-linearization of the Phillips curve, obtained from the firm's problem in (A4), yields

$$\pi_t = \beta E_t\left\{\pi_{t+1}\right\} - \frac{\varepsilon - 1}{\xi}\mu_t.$$

(Recall that the lower case variables represent their log-linear deviations), Thus, the slope κ of the curve is $\frac{\varepsilon-1}{\xi}$. Log-linearizing the wage setting rule (7) yields

$$w_t = \sigma c_t + \varphi n_t. \tag{A5}$$

Next, the production function in the log-linear form is $y_t = a_t + n_t$. The log-linearized markup equation, $\mathcal{M}_t \equiv \frac{A_t}{W_t}$, is

$$\mu_t = a_t - w_t. \tag{A6}$$

Expressing n_t from the production function and substituting it into the wage equation (A5), we have

$$w_t = \sigma c_t + \varphi y_t - \varphi a_t.$$

Using this result in the markup equation (A6), along with $c_t = y_t$, we have

$$\mu_t = (1+\varphi)\, a_t - (\sigma+\varphi)\, y_t.$$

Setting $\mu_t = 0$ gives us the natural rate of output,

$$y_t^n = \frac{1+\varphi}{\sigma+\varphi} a_t.$$

Using the latter formula in the left-hand side of the markup equation (A6) gives us

$$\mu_t = -(\sigma+\varphi)\, x_t.$$

The Euler equation of the unconstrained agent is

$$c_t^U = E_t\{c_{t+1}^U\} - \frac{1}{\sigma}\left[i_t - E_t\{\pi_{t+1}\}\right] - \frac{1}{\sigma} E_t\{\Delta z_{t+1}\}. \tag{A7}$$

Using $C_t = C_t^U(1-\lambda\Gamma_t)$ and taking into account (33), we obtain

$$\begin{aligned} c_t = c_t^U - \frac{\lambda}{1-\lambda\Gamma}\gamma_t & \\ = c_t^U + \Phi x_t, \end{aligned} \tag{A8}$$

By substituting (A8) into (A7), we get the aggregate Euler equation,

$$c_t = E_t\{c_{t+1}\} - \frac{1}{\sigma}\left[i_t - E_t\{\pi_{t+1}\}\right] - \frac{1}{\sigma} E_t\{\Delta z_{t+1}\} - E_t\{\Phi \Delta x_{t+1}\}. \tag{A9}$$

Recall that $c_t = y_t$ and that $y_t = x_t + \frac{1+\varphi}{\sigma+\varphi} a_t$. For the RANK economy, the natural rate of interest is

$$\begin{aligned} r_t^n = -E_t\left\{\triangle z_{t+1}\right\} + \sigma E_t\left\{\triangle y_{t+1}^n\right\} & \\ = (1-\rho_z) z_t - \sigma(1-\rho_a)\frac{1+\varphi}{\sigma+\varphi} a_t. \end{aligned} \tag{A10}$$

Substituting $E_t\left\{\triangle z_{t+1}\right\}$ from (A10) into (A9) leads to the aggregate Euler equation (IS curve) (10). Note that the Taylor rule (9) can be rewritten as (12) because $y_t^n = \frac{1+\varphi}{\sigma+\varphi} a_t$ with $u_t \equiv \phi_y \frac{1+\varphi}{\sigma+\varphi} a_t + v_t$. In sum, our aggregate three-equation model is given by (10), (11) and (12).

Closed-form solution for the output gap. Regarding the solution for the output gap, we proceed as follows: We first iterate the IS curve forward to write x_t as

$$x_t = -\frac{1}{\sigma(1-\Phi)} E_t \sum_{k=0}^{\infty}\left[i_{t+k} - \pi_{t+1+k} - r_{t+k}^n\right]. \tag{A11}$$

We then substitute in for i_t using the Taylor rule (9), for x_{t+k} using the Phillips curve (11) and for the inflation terms using the above closed-form solutions. In particular, one can write x_t as

$$x_t = -H_t - \frac{1}{\sigma(1-\Phi)}\left[{}_{k=0}^{\infty} v_{t+k} - {}_{k=0}^{\infty} r_{t+k}^n\right], \tag{A12}$$

where $H_t \equiv f\left(m_1, m_2, \phi_{E\pi}, \phi_\pi, \phi_y, \kappa, \beta, \{X_s\}_{s=k+t}^{\infty}\right)$. For example, for the case of two distinct real roots leading to a unique equilibrium, H_t is defined as

$$H_t \equiv \frac{1}{m_1 - m_2} {}_{k=0}^{\infty}\left[\left(\frac{\phi_y}{\kappa} + \phi_\pi\right)\sum_{s=t+k}^{\infty}\left(m_1^{t+k-1-s} - m_2^{t+k-1-s}\right) X_s + \left(\phi_{E\pi} - 1 - \frac{\beta\phi_y}{\kappa}\right)\sum_{s=t+k+1}^{\infty}\left(m_1^{t+k-s} - m_2^{t+k-s}\right) X_s\right]. \tag{A13}$$

Therefore, the closed-form solution for the output gap depends solely on the exogenous variables, as well as on the model's parameters, among which the Taylor-rule parameters enter explicitly.

A.2 Supplementary results of Section 3

Proof of Proposition 1. Suppose $\lambda < \frac{1}{2}$.
(i) $\mathcal{M} < 1 + \frac{1}{\delta(1-\tau)}$. We will show that Φ is increasing in λ.

For the Taylor rule (9), the upper indeterminacy bound is defined in Theorem 1. The share of constrained agents λ enters the formula (15) for Φ, so that we have

$$\frac{\partial \Phi}{\partial \lambda} = \frac{(1-\lambda\gamma)\left((\sigma+\varphi)\Psi + \lambda(\sigma+\varphi)\frac{\partial \Psi}{\partial \lambda}\right) + \left(\Gamma + \lambda \frac{\partial \Gamma}{\partial \lambda}\right)(\lambda(\sigma+\varphi)\Psi)}{(1-\lambda\gamma)^2}.$$

The result holds if the numerator is positive. The derivatives of Ψ and Γ with respect to λ, respectively, are given by

$$\frac{\partial \Psi}{\partial \lambda} = \frac{2(1-\lambda)(1-\delta(1-\tau)-\tau^W\delta^W)\left(\delta(1-\tau)(\mathcal{M}-1)+1-\tau^W\delta^W\right)}{\mathcal{D}^3} - \frac{(1-\delta(1-\tau)-\tau^W\delta^W)}{\mathcal{D}^2},$$

$$\frac{\partial \Gamma}{\partial \lambda} = \frac{\left[(\mathcal{M}-1)\left(1-\delta\left(1-\tau\right)\right)+\tau^W\delta^W\right]\left((\mathcal{M}-1)\delta(1-\tau)-\left(\tau^W\delta^W-1\right)\right)}{\mathcal{D}^2},$$

where $\mathcal{D} \equiv 1 + \left(\delta^W\tau^W - 1\right)\lambda + (\mathcal{M}-1)(1-\delta(1-\tau)\lambda)$. Therefore, $\frac{\partial \Phi}{\partial \lambda} > 0$ if the following inequality holds:

$$\begin{aligned}
&(\sigma+\varphi)\Psi \\
&\quad + 2(1-\lambda\gamma)\lambda(\sigma+\varphi)\left(\frac{2(1-\lambda)(1-\delta(1-\tau)-\tau^W\delta^W)\left(\delta(1-\tau)(\mathcal{M}-1)+1-\tau^W\delta^W\right)}{\mathcal{D}^3}\right) \\
&\quad + \lambda^2(\sigma+\varphi)\Psi\left(\frac{\left[(\mathcal{M}-1)\left(1-\delta\left(1-\tau\right)\right)+\tau^W\delta^W\right]\left((\mathcal{M}-1)\delta(1-\tau)-\left(\tau^W\delta^W-1\right)\right)}{\mathcal{D}^2}\right) \\
&\qquad > (1-\lambda\gamma)\lambda(\sigma+\varphi)\left(\frac{(1-\delta(1-\tau)-\tau^W\delta^W)}{\mathcal{D}^2}\right). \qquad \text{(A14)}
\end{aligned}$$

Recall that $\Psi > 0$ and that $\mathcal{M} \equiv \frac{\varepsilon}{\varepsilon-1}$. Let us consider the third term on the left-hand side. It is greater than zero as long as the following inequality holds:

$$1 > \delta(1-\tau)(\mathcal{M}-1) + \tau^W\delta^W.$$

This holds trivially for $\tau = 1$. For $\tau \neq 1$, the above inequality implies the following restriction on the markup:

$$\mathcal{M} < 1 + \frac{1-\tau^W\delta^W}{\delta(1-\tau)},$$

and the following restriction on the parameter of the Dixit-Stiglitz aggregator: $\varepsilon > 1 + \frac{\delta(1-\tau)}{1-\tau^W\delta^W}$. The assumption will be satisfied under most reasonable parameterizations. Indeed, a reasonable value of τ^W would most likely correspond to a value of zero or lower, implying that constrained agents receive a labor income tax refund. Now, turning to the first, second and fourth terms of the inequality, and recalling the definition of Ψ in (17), we have

$$\begin{aligned}
&(1-\lambda)\cdot\left[(1-\delta(1-\tau)-\tau^W\delta^W)\right] \\
&\quad + 2(1-\lambda\gamma)\lambda(\sigma+\varphi)(1-\lambda)(1-\delta(1-\tau))\delta(1-\tau)(\mathcal{M}-1) \\
&\qquad > (1-\lambda\gamma)\lambda\cdot\left[(1-\delta(1-\tau)-\tau^W\delta^W)\right].
\end{aligned}$$

The second term is positive. The first and third terms just differ in $1-\lambda$ and $(1-\lambda\gamma)\lambda$. Since by assumption, $1-\lambda\gamma < 1$ and $\lambda < \frac{1}{2}$, the above inequality holds. Therefore, given that $(1-\lambda+(\mathcal{M}-1)(1-\delta(1-\tau)\lambda))^3 >$

0 and given that the third term in (A14) is strictly positive, inequality (A14) is satisfied, and $\frac{\partial \Phi}{\partial \lambda} > 0$. In sum, $\frac{\partial \Phi}{\partial \lambda} > 0$ so long as $\mathcal{M} < 1 + \frac{1}{\delta(1-\tau)}$ for $\tau \neq 1$. For $\tau = 1$, we do not need the restriction on the markup for the proof to hold.

(*ii*). We will show that Φ is increasing in τ and decreasing in δ. Equation (15) yields

$$\frac{\partial \Phi}{\partial \tau} = \frac{(1-\lambda\gamma)\lambda(\sigma+\varphi)\frac{\partial \Psi}{\partial \tau} + \lambda^2(\sigma+\varphi)\Psi\frac{\partial \Gamma}{\partial \tau}}{(1-\lambda\gamma)^2}.$$

If the numerator is positive the result holds.
First, let us consider the partial derivatives of Ψ and Γ with respect to τ,

$$\begin{aligned}\frac{\partial \Psi}{\partial \tau} &= \frac{\delta(1-\lambda)}{\mathcal{D}^2} - \frac{2\delta\lambda(1-\lambda)\left[(1-\lambda)(1-\delta(1-\tau)-\delta^W\tau^W)\right](\mathcal{M}-1)}{\mathcal{D}^3},\\ \frac{\partial \Gamma}{\partial \tau} &= \frac{\delta(\mathcal{M}-1)}{\mathcal{D}} - \frac{(\mathcal{M}-1)\delta\lambda\left[(\mathcal{M}-1)(1-\delta(1-\tau))+\tau^W\delta^W\right]}{\mathcal{D}^2},\end{aligned}$$

where $\mathcal{D}$ is defined as in part (*i*). We will have $\frac{\partial \Gamma}{\partial \tau} > 0$ if the following inequality holds:

$$\delta(\mathcal{M}-1)\left[1+\left(\tau^W\delta^W-1\right)\lambda+(\mathcal{M}-1)(1-\delta(1-\tau)\lambda)\right] > (\mathcal{M}-1)\delta\lambda\left[(\mathcal{M}-1)(1-\delta(1-\tau))+\tau^W\delta^W\right].$$

Given that $\mathcal{M}-1>0$, we have

$$\begin{aligned}\mathcal{D} &> \lambda\left[(\mathcal{M}-1)(1-\delta(1-\tau))+\tau^W\delta^W\right],\\ 1+\lambda\tau^W\delta^W-\lambda+(\mathcal{M}-1)(1-\delta(1-\tau)\lambda) &> \lambda(\mathcal{M}-1)(1-\delta(1-\tau))+\lambda\tau^W\delta^W,\\ (1-\lambda)+(\mathcal{M}-1)(1-\delta(1-\tau)\lambda) &> \lambda(\mathcal{M}-1)(1-\delta(1-\tau)).\end{aligned}$$

Since $\lambda \in [0,1)$, we obtain $(1-\delta(1-\tau)\lambda) > \lambda(1-\delta(1-\tau))$. Therefore, since the first term in the above inequality is always positive, and the second term is greater than the third one, the above inequality holds. Next, let us consider $\frac{\partial \Psi}{\partial \tau}$. This is also greater than zero so long as the following inequality holds:

$$\delta(1-\lambda)\mathcal{D} > 2\delta\lambda(1-\lambda)(1-\delta(1-\tau)-\delta^W\tau^W)(\mathcal{M}-1),$$

or equivalently,

$$\mathcal{D} > 2\lambda(\mathcal{M}-1)(1-\delta(1-\tau)) - 2\lambda(\mathcal{M}-1)\delta^W\tau^W.$$

For $\lambda < \frac{1}{2}$, the second term on the left hand side is greater than the term on the right hand side. Additionally, when $\tau^W \geq 0$, the sum of the first two terms on the left hand side will also be positive. Therefore, the above inequality holds. When $\tau^W < 0$ a more stringent restriction on λ may be necessary for the proof to hold.
Therefore, since both $\frac{\partial \Psi}{\partial \tau} > 0$ and $\frac{\partial \Gamma}{\partial \tau} > 0$ for $\lambda < \frac{1}{2}$, $\frac{\partial \Phi}{\partial \tau} > 0$, leading to $\frac{\partial \Phi}{\partial \tau} > 0$. This directly implies that Φ decreases in δ. This is because just the parameter Φ contains τ and δ and because $\frac{\partial \Phi}{\partial \tau} = -\frac{\partial \Phi}{\partial \delta}$.

(*iii*). We now show that Φ decreases in both τ^W and δ^W.

Suppose that $\lambda < \frac{1}{2}$. Formula (15) implies

$$\frac{\partial \Phi}{\partial \tau^W} = \frac{(1-\lambda\Gamma)\lambda(\sigma+\varphi)\frac{\partial \Psi}{\partial \tau^W} + \lambda^2(\sigma+\varphi)\Psi\frac{\partial \Gamma}{\partial \tau^W}}{(1-\lambda\Gamma)^2}.$$

If the numerator is negative, the result holds.
From (17) and (16), partial derivatives of Ψ and Γ with respect to τ^W are given by

$$\frac{\partial \Psi}{\partial \tau^W} = -\frac{\delta^W(1-\lambda)}{\mathcal{D}^2} - \frac{2\delta^W\lambda\left[(1-\lambda)(1-\delta(1-\tau)-\delta^W\tau^W)\right]}{\mathcal{D}^3},$$
$$\frac{\partial \Gamma}{\partial \tau^W} = \frac{\delta^W}{\mathcal{D}} - \frac{\delta^W\lambda\left[(\mathcal{M}-1)(1-\delta(1-\tau))+\tau^W\delta^W\right]}{\mathcal{D}^2},$$

where $\mathcal{D}$ is defined as in part (i).
Toward contradiction, we will now assume that $\frac{\partial \Phi}{\partial \tau^W} > 0$ under the conditions of $\lambda < \frac{1}{2}$ and $\tau^W > 0$. For $\frac{\partial \Phi}{\partial \tau^W} > 0$, we will need

$$\lambda\Psi\frac{\partial \Gamma}{\partial \tau^W} > -(1-\lambda\Gamma)\frac{\partial \Psi}{\partial \tau^W},$$

which is equivalent to

$$\lambda\Psi\frac{\delta^W}{\mathcal{D}} > \lambda\Psi\frac{\delta^W\lambda\left[(\mathcal{M}-1)(1-\delta(1-\tau))+\tau^W\delta^W\right]}{\mathcal{D}^2} + (1-\lambda\Gamma)\frac{\delta^W(1-\lambda)}{\mathcal{D}^2}$$
$$+(1-\lambda\Gamma)\frac{2\delta^W\lambda\left[(1-\lambda)(1-\delta(1-\tau)-\delta^W\tau^W)\right]}{\mathcal{D}^3}.$$

After using equation (17) for Ψ, the above inequality simplifies to

$$\begin{gathered}\lambda\delta^W\mathcal{D}(1-\lambda)\left[1-\delta(1-\tau)-\delta^W\tau^W\right] > \\ \lambda^2\delta^W\left[(\mathcal{M}-1)(1-\delta(1-\tau))+\tau^W\delta^W\right](1-\lambda)(1-\delta(1-\tau)-\delta^W\tau^W) \\ +(1-\lambda\Gamma)\delta^W(1-\lambda)\mathcal{D}^2 \\ +(1-\lambda\Gamma)2\delta^W\lambda\left[(1-\lambda)(1-\delta(1-\tau)-\delta^W\tau^W)\right]\mathcal{D}.\end{gathered}$$

The second term on the right hand side of the inequality is positive. Therefore, for the above to be true, the term on the left hand side must be greater than the first and third term on the right hand side. Dividing these terms by $\lambda\delta^W(1-\lambda)(1-\delta(1-\tau)-\delta^W\tau^W)$, for the left hand side to be greater than the first and third terms on the right hand side we need

$$\mathcal{D} > \lambda\left[(\mathcal{M}-1)(1-\delta(1-\tau))+\tau^W\delta^W\right] + (1-\lambda\Gamma)2\mathcal{D},$$

$$0 > \lambda\left[(\mathcal{M}-1)(1-\delta(1-\tau))+\tau^W\delta^W\right] + (1-2\lambda\Gamma)\mathcal{D}.$$

The first term on the right hand side will be positive. Additionally, if $\lambda < \frac{1}{2}$, since $\Gamma \in [0,1)$, the second term on the right hand side will also be positive. Therefore, the inequality cannot hold. Therefore, for $\lambda < \frac{1}{2}$ and $\tau^W > 0$, we have $\frac{\partial \Phi}{\partial \tau^W} < 0$. This directly implies that Φ decreases in δ^W as the two parameters only enter jointly. ■

Impulse responses. Using our closed-form solutions, we can write the responses of inflation (and hence, the output gap) to a one-time anticipated shock. Suppose that at $t = 0$, the agents anticipate a future shock X_T to occur at time T (due to an anticipated change in z_T, a_T or v_T). For $t \leq T$, the impulse-response functions for inflation in case of two distinct real roots producing a unique solution and that of complex roots, respectively, are

$$\text{distinct real roots:} \qquad \pi_t = \tfrac{m_1^{t-1-T} - m_2^{t-1-T}}{m_1 - m_2} \cdot X_T, \tag{A15}$$
$$\text{complex roots:} \qquad \pi_t = \tfrac{r^{t-1-T}}{\eta}\sin\left[\theta\left(t-1-T\right)\right] \cdot X_T. \tag{A16}$$

For $t > T$, inflation returns immediately to steady state, $\pi_t = 0$.

When the solution is indeterminate, the impulse-response functions are

$$\text{for } t \leq T: \quad \pi_t = \tfrac{m_1^{t-1-T}}{m_1 - m_2} \cdot X_T,$$
$$\text{for } t > T: \quad \pi_t = \tfrac{m_2^{t-1-T}}{m_1 - m_2} \cdot X_T,$$

As stated in the main text, the key difference between indeterminate and unique solutions is that indeterminate solutions are functions of all past, present and future shocks while unique solutions are functions only of present and future shocks.

Relation between individual and aggregate variables. First, the relationship between unconstrained consumption and aggregate consumption is derived using equation (A8) and $c_t = x_t + \frac{1+\varphi}{\sigma+\varphi} a_t$; this gives us equation (19). Second, using the log-linearized wage rule (A5) and the relationship between c_t and c_t^U gives us equation (18).

Finally, we derive the relationship between aggregate consumption and constrained consumption. Log-linearizing the aggregate consumption equation, $C_t = \lambda C_t^K + (1-\lambda) C_t^U$ and re-arranging the terms yields

$$c_t^U = \frac{C}{(1-\lambda)C^U} c_t - \frac{\lambda}{1-\lambda} \frac{C^K}{C^U} c_t^K,$$

which in terms of $\Gamma = 1 - \frac{C^K}{C^U}$ becomes

$$c_t^U = \left(\frac{C}{(1-\lambda)C^U} \right) c_t - \left(\frac{\lambda(1-\Gamma)}{1-\lambda} \right) c_t^K.$$

Using the aggregate consumption equation $C = \lambda C^K + (1-\lambda) C^U$, as well as $c_t = c_t^U + \Phi x_t$ in (A8) and re-arranging the terms, we have

$$c_t^K = c_t + \frac{(1-\lambda)\,\Phi}{(1-\Gamma)\,\lambda} \cdot x_t. \tag{A17}$$

Imposing market clearing $x_t = c_t + \frac{1+\varphi}{\sigma+\varphi} a_t$, we then have

$$c_t^K = \chi c_t + (1-\chi) \left(\frac{1+\varphi}{\sigma+\varphi} \right) a_t, \tag{A18}$$

which appears in (28) of the main text with χ being defined in (20).

A.3 Supplementary results of Section 4

Closed-form solution for consumption of the unconstrained agent. Solving the unconstrained agent's Euler equation (A7) forward gives us

$$c_t^U = -\frac{1}{\sigma} \sum_{k=0}^{\infty} \left[i_{t+k} - E_t\{\pi_{t+1+k}\} \right] + \frac{1}{\sigma} z_t. \tag{A19}$$

We use the Taylor rule (9) to substitute in for i_t,

$$c_t^U = -\frac{1}{\sigma} \sum_{k=0}^{\infty} \left[\phi_y E_t\{x_{t+k}\} + \phi_\pi E_t\{\pi_{t+k}\} + (\phi_{E\pi} - 1) E_t\{\pi_{t+1+k}\} + v_{t+k} \right] + \frac{1}{\sigma} z_t. \tag{A20}$$

Substituting x_t from the Phillips curve (11) into (A20) and considering the closed-form solution for inflation (see equation (A13) for the case of two distinct real roots with a unique solution), we have

$$c_t^U = -\frac{1}{\sigma} \left[H_t +_{k=0}^{\infty} v_{t+k} \right] + \frac{1}{\sigma} z_t, \tag{A21}$$

where H_t is defined in (A13). When there is a one-time shock in period T, the term H_t reduces to $\mathcal{H}_{t,T}$ presented in the main text. Setting $z_t = 0$ and $v_{t+k} = 0 \; \forall \; t+k \neq T$ gives us the response of unconstrained consumption (25) presented in the main text.

Closed-form solution for consumption of the constrained agent. Solving the aggregate Euler equation (A9) forward gives us

$$c_t = -\frac{1}{\sigma\,(1-\Phi)}\sum_{k=0}^{\infty}\left[i_{t+k} - E_t\{\pi_{t+1+k}\}\right] + \frac{1}{\sigma}z_t + \frac{1}{(1-\Phi)}\frac{1+\varphi}{\sigma+\varphi}a_t.$$

After substituting in for i_{t+k} from the Taylor rule (9), we get the closed -form solution for aggregate consumption,

$$c_t = -\frac{1}{\sigma(1-\Phi)}\left\{H_t + \sum_{k=0}^{\infty} v_{t+k}\right\} - \frac{1}{\sigma(1-\Phi)}\sum_{k=0}^{\infty} r^n_{t+k} + \frac{1}{\sigma}z_t.$$

Therefore, after substituting in for c_t in (A18) with $a_t = 0$, we have

$$c^K_t = \chi \cdot \left\{-\left(\frac{1}{\sigma(1-\Phi)}\right)[H_t +_{k=0}^{\infty} v_{t+k}] - \frac{1}{\sigma(1-\Phi)}\sum_{k=0}^{\infty} r^n_{t+k} + \frac{1}{\sigma}z_t\right\}. \tag{A22}$$

Again, when there is a one-time shock in period T, H_t reduces to $\mathcal{H}_{t,T}$. Setting $z_t = 0$ for $\forall\ t$ and $v_{t+k} = 0$ for $\forall\ t+k \neq T$ gives us the response of constrained consumption (26).

A.4 Supplementary results for Section 5

Planned expenditure curve. First, we derive constrained and unconstrained consumption in terms of income and the real interest rate. The lifetime budget constraint of the unconstrained agent is

$$E_{t}{}_{i=0}^{\infty}\Lambda_{t,t+i}Y^U_{t+i} = E_{t}{}_{i=0}^{\infty}\Lambda_{t,t+i}C^U_{t+i},$$

where Y^U_t is the income of the unconstrained agent. Log-linerizing the budget constraint gives us

$$E_{t}{}_{i=0}^{\infty}\beta^i\left(\hat{\Lambda}_{t,t+i} + y^U_t\right) = E_{t}{}_{i=0}^{\infty}\beta^i\left(\hat{\Lambda}_{t,t+i} + c^U_t\right),$$

where $\hat{\Lambda}_{t,t+i}$ and y^U_t are the stochastic discount factor used by the firm and the income of the unconstrained agent, respectively (in log deviations). Using the fact that $\hat{\Lambda}_{t,t+i} = -\sigma\left(c^U_{t+1} - c^U_t\right)$, after adding $\left(\frac{1}{\sigma} - 1\right)E_{t}{}_{i=0}^{\infty}\beta^i\hat{\Lambda}_{t,t+i}$ to both sides we have

$$E_{t}{}_{i=0}^{\infty}\beta^i\left(\frac{1}{\sigma}\hat{\Lambda}_{t,t+i} + y^U_t\right) = 0 + E_{t}{}_{i=0}^{\infty}\beta^i\left(\frac{1}{\sigma}\hat{\Lambda}_{t,t+i} + c^U_t\right).$$

Using the Euler equation for the unconstrained agent, the right hand side becomes $\frac{1}{1-\beta}c^U_t$. Given the $\hat{\Lambda}_{t,t} = 0$, we can rewrite the left hand side as

$$E_{t}{}_{i=0}^{\infty}\beta^i\left(\hat{\Lambda}_{t,t+i} + y^U_t\right) = 0 + E_{t}{}_{i=1}^{\infty}\beta^i\left(\hat{\Lambda}_{t,t+i} + c^U_t\right).$$

From the FOC for bonds (A2), $i_t - E_t\{\pi_{t+1}\} = -E_t\left\{\hat{\Lambda}_{t,t+i}\right\}$. Therefore,

$$E_t\left\{\hat{\Lambda}_{t,t+i}\right\} = E_{t}{}_{k=0}^{i-1}\left(i_{t+k} - \pi_{t+1+k}\right).$$

Using the latter result, we have

$${}_{i=0}^{\infty}E_t\left\{\hat{\Lambda}_{t,t+i}\right\} = -{}_{i=1}^{\infty}\beta^i E_{t}{}_{k=0}^{i-1}\left(i_{t+k} - \pi_{t+1+k}\right) = -\frac{\beta}{1-\beta}E_{t}{}_{i=0}^{\infty}\beta^i\left[i_{t+i} - \pi_{t+1+i}\right].$$

Combining the latter two results gives us

$$\frac{1}{1-\beta}c_t^U = -\frac{1}{\sigma}\frac{\beta}{1-\beta}E_{t i=0}^{\infty}\beta^i\left(i_{t+i} - E_t\left\{\pi_{t+1+i}\right\}\right) + E_{t i=0}^{\infty}\beta^i y_{t+i}^U.$$

Multiplying both sides by $1-\beta$ and removing the $i=0$ variables from the summations, we have

$$c_t^U = (1-\beta)\,y_t^U - \frac{1}{\sigma}\beta\left(i_{t+i} - \pi_{t+1+i}\right) - \beta\frac{1}{\sigma}E_{t i=0}^{\infty}\beta^i\left(i_{t+i} - \pi_{t+1+i}\right) + (1-\beta)\,E_{t i=0}^{\infty}\beta^i y_{t+i}^U.$$

Finally, noting that

$$\beta c_{t+1}^U = -\beta\frac{1}{\sigma}\sum_{i=0}^{\infty}\beta^i\left(i_{t+i} - \pi_{t+1+i}\right) + (1-\beta)\sum_{i=0}^{\infty}\beta^i y_{t+i}^U,$$

we get unconstrained consumption depending on income and the real interest rate,

$$c_t^U = (1-\beta)\,y_t^U - \frac{1}{\sigma}\beta\left[i_t - E_t\left\{\pi_{t+1}\right\}\right] + \beta E_t\left\{c_{t+1}^U\right\}.$$

Taking into account that the constrained agent consumes all of his or her income each period, unconstrained consumption is

$$c_t^K = y_t^K,$$

where y_t^K denotes income of the constrained agents.

Second, aggregation over consumption gives us $c_t = (1-\varsigma)\,c_t^U + \varsigma c_t^K$, where $\varsigma \equiv \frac{\lambda(1-\Gamma)}{(1-\lambda)+\lambda(1-\Gamma)}$ is a constant weight. Given that $c_t = y_t$ and $c_t^K = \chi c_t$, $y_t^K = \chi y_t$ and $y_t^U = \left(\frac{1-\varsigma\chi}{1-\varsigma}\right)y_t$, we can write aggregate consumption as

$$c_t = \left[1 - (1-\Phi)(1-\varsigma)\beta\right]y_t - \frac{1}{\sigma}\beta(1-\varsigma)\left[i_t - E_t\left\{\pi_{t+1}\right\}\right] + (1-\Phi)(1-\varsigma)\beta E_t\left\{c_{t+1}\right\}.$$

Third, we derive a relationship between c_t and c_{t+1}. Our closed-form solutions allow us to consider sequences of future monetary policy shocks with the following structure: $\{v_t\} = \{\phi_t v\}$ where ϕ_t's are scalars.[20] We confine our attention to the case when $\phi_t = 0$ for all but period $t+j$ or when $\phi_t = \rho^t$ (i.e., a one-time transitory or persistent future shock). We generalize the previous equation relating c_t to c_{t+1} (equivalently, y_t to y_{t+1}) as $c_{t+1} = \varrho_j c_t$, where j indicates how many periods in the future the shock occurs.[21] The value of the output-gap persistence ϱ_j depends on whether the shock has persistence or not, as well as on the type of roots (i.e., real or complex). We report ϱ_j for different cases in Table 5.

We present the formulas for a one-time transitory shock with a unique solution, one-time transitory shock with an indeterminate solution, and one-time persistent shock with a unique solution.

Finally, using the derived relationship $c_{t+1} = \varrho_j c_t$, we can write the equation for aggregate consumption and thus, the planned expenditure curve as (29).

How the weight ς changes with λ. We prove that $\frac{\partial\varsigma}{\partial\lambda} > 0$ when $\tau = 1$ and $\tau^W = 0$

$$\begin{aligned}\frac{\partial\varsigma}{\partial\lambda} &= \frac{\left[(1-\lambda)+\lambda(1-\Gamma)\right]\left[(1-\Gamma) - \lambda\frac{\partial\Gamma}{\partial\lambda}\right] - \lambda(1-\Gamma)\left[-1 + (1-\Gamma) - \lambda\frac{\partial\Gamma}{\partial\lambda}\right]}{\left[(1-\lambda)+\lambda(1-\Gamma)\right]^2}\\ &= \frac{(1-\Gamma) - \lambda(1-\lambda)\frac{\partial\Gamma}{\partial\lambda}}{\left[(1-\lambda)+\lambda(1-\Gamma)\right]^2}.\end{aligned}$$

Given $\Gamma = \frac{(\mathcal{M}-1)}{(1-\lambda)+(\mathcal{M}-1)}$ and $\frac{\partial\Gamma}{\partial\lambda} = \frac{(\mathcal{M}-1)}{(1-\lambda)+(\mathcal{M}-1)}$, for the above to be positive, we need

$$(1-\lambda) + (1-\lambda)(\mathcal{M}-1) > \lambda(1-\lambda)(\mathcal{M}-1).$$

Therefore, $\frac{\partial\varsigma}{\partial\lambda} > 0$.

---

[20] The ϕ_t's do not need to be monotonicly decreasing. The only requirement is that the infinite sums are well definied.

[21] Our result $c_{t+1} = \varrho_j c_t$ generalize that in Bilbiie (2019a) who considers the case of the Taylor rule $i_t = E_t\{\pi_{t+1}\} + v_t$, when there is a contemporaneous shock that reverts to zero with persistence ρ_v and obtains $c_{t+1} = \rho_v c_t$, where ρ_v is the persistence of the monetary policy shock. This is due to the fact that with this Taylor rule, the real interest rate is equal to the monetary policy shock, so that the real interest rate at $t+j$ is $r_{t+j} = \rho_v^j v_t$.

Table 5: Output-gap persistence

	Determinacy, transitory shock	Indeterminacy, transitory shock	Determinacy, persistent shock
$j=0$	0	$\frac{1-\beta m_2}{m_1^{-1}-\beta}$	$\frac{\sum_{s=t+1}^{\infty}\left[m_1^{t-s}-m_2^{t-s}\right]\rho_v^{s-t}-\beta\sum_{s=t+2}^{\infty}\left[m_1^{t+1-s}-m_2^{t+1-s}\right]\rho_v^{s-t}}{\sum_{s=t}^{\infty}\left[m_1^{t-s}-m_2^{t-s}\right]\rho_v^{s-t}-\beta\sum_{s=t+1}^{\infty}\left[m_1^{t-s}-m_2^{t-s}\right]\rho_v^{s-t}}$
$j=1$	$\frac{m_1^{-1}-m_2^{-1}}{m_1^{-2}-m_2^{-2}-\beta\left(m_1^{-1}-m_2^{-1}\right)}$	m_1	$\frac{\sum_{s=t+1}^{\infty}\left[m_1^{t-s}-m_2^{t-s}\right]\rho_v^{s-1-t}-\beta\sum_{s=t+2}^{\infty}\left[m_1^{t+1-s}-m_2^{t+1-s}\right]\rho_v^{s-1-t}}{\sum_{s=t+1}^{\infty}\left[m_1^{t-s}-m_2^{t-s}\right]\rho_v^{s-1-t}-\beta\sum_{s=t+1}^{\infty}\left[m_1^{t-s}-m_2^{t-s}\right]\rho_v^{s-1-t}}$
$j\geq 2$	$\frac{m_1^{-j}-m_2^{-j}-\beta\left(m_1^{1-j}-m_2^{1-j}\right)}{m_1^{-1-j}-m_2^{-1-j}-\beta\left(m_1^{-j}-m_2^{-j}\right)}$	m_1	$\frac{\sum_{s=t+j}^{\infty}\left[m_1^{t-s}-m_2^{t-s}\right]\rho_v^{s-t-j}-\beta\sum_{s=t+j}^{\infty}\left[m_1^{t+1-s}-m_2^{t+1-s}\right]\rho_v^{s-t-j}}{\sum_{s=t+j}^{\infty}\left[m_1^{t-1-s}-m_2^{t-1-s}\right]\rho_v^{s-t-j}-\beta\sum_{s=t+j}^{\infty}\left[m_1^{t-s}-m_2^{t-s}\right]\rho_v^{s-t-j}}$

Notes: Formulas in each cell correspond to the persistence in the ouput gap ϱ_j.

A.5 Supplementary results for Section 6

Closed-form solutions for inflation in the stochastic setting. The inflation solution for the case of two distinct roots leading to a unique solution, $|m_1|>1$ and $|m_2|>1$, is given by

$$\pi_t=\frac{1}{m_1-m_2}E_t\left[\sum_{s=t}^{\infty}m_1^{t-1-s}X_s-\sum_{s=t}^{\infty}m_2^{t-1-s}X_s\right]; \tag{A23}$$

see Maliar and Taylor (2018) for full derivations, as well as for solutions for the other three cases.

This solution for inflation will be different depending on the type of shock assumed. (We present solutions when either a productivity shock or preference shock happens). By definition, X_t is described by (14). For example, if the economy is subject to productivity shocks only, we have

$$X_t=\frac{\kappa}{\beta\sigma(1-\Phi)}\cdot\frac{1+\varphi}{\sigma+\varphi}\left[\sigma(1-\rho_a)+\phi_y\right]\cdot a_t.$$

For $s\geq t$, in the case of a productivity shock, we have

$$E_t\left\{X_s\right\}=\Upsilon_a\rho_a^{s-t}a_t,\qquad \Upsilon_a\equiv\frac{\kappa}{\beta\sigma(1-\Phi)}\frac{1+\varphi}{\sigma+\varphi}\left[\sigma(1-\rho_a)+\phi_y\right],$$

so that we have

$$\pi_t=\frac{\Upsilon_a}{m_1-m_2}\left(\frac{a_t}{\rho_a}\right)\left[\sum_{s=t}^{\infty}\left(\frac{m_1}{\rho_a}\right)^{t-1-s}-\sum_{s=t}^{\infty}\left(\frac{m_2}{\rho_a}\right)^{t-1-s}\right]. \tag{A24}$$

In the case of preference shocks only,

$$E_t\left\{X_s\right\}=\Upsilon_z\rho_z^{s-t}z_t,\qquad \Upsilon_z\equiv-\frac{\kappa}{\beta\sigma(1-\Phi)}(1-\rho_z),$$

so that we obtain

$$\pi_t=\frac{\Upsilon_z}{m_1-m_2}\left(\frac{z_t}{\rho_z}\right)\left[\sum_{s=t}^{\infty}\left(\frac{m_1}{\rho_j}\right)^{t-1-s}-\sum_{s=t}^{\infty}\left(\frac{m_2}{\rho_j}\right)^{t-1-s}\right]. \tag{A25}$$

Equations (A24) and (A25) can be written as the closed-form solution for inflation presented in the main text. The case of complex roots is obtained similarly.

Closed-form solutions for the output gap in the stochastic setting. We can use equation (A24) for inflation, along with the Taylor rule (9) and the Phillips curve (11), to construct our series for the output gap,

$$x_t = -E_t \left\{ \widetilde{H}_t + \frac{1}{1-\rho_v} v_t - \frac{1}{1-\rho_{r^n}} r_t^n \right\}, \tag{A26}$$

with $\widetilde{H}_t$ being defined by

$$\begin{aligned} \widetilde{H}_t \quad = \quad & \sum_{k=0}^{\infty} \left\{ \left(\frac{\phi_y}{\kappa} + \phi_\pi\right) \left(\frac{\Upsilon_j}{m_1 - m_2}\right) \left(\frac{h_{t,j}}{\rho_j}\right) \left[\sum_{s=t+k}^{\infty} \left(\frac{m_1}{\rho_j}\right)^{t+k-1-s} - \sum_{s=t+k}^{\infty} \left(\frac{m_2}{\rho_j}\right)^{t+k-1-s} \right] \right. \\ & \left. + \left(\phi_{E\pi} - 1 - \frac{\beta\phi_y}{\kappa}\right) \left(\frac{\Upsilon_j}{m_1 - m_2}\right) \left(\frac{h_{t,j}}{\rho_j}\right) \left[\sum_{s=t+k+1}^{\infty} \left(\frac{m_2}{\rho_j}\right)^{t+k-s} - \sum_{s=t+k+1}^{\infty} \left(\frac{m_2}{\rho_j}\right)^{t+k-s} \right] \right\}, \end{aligned}$$

where $j \in \{a, z\}$, $h_{t,a} = a_t$, $h_{t,z} = z_t$, $\Upsilon_j \in \{\Upsilon_a, \Upsilon_z\}$.

Proof of Proposition 2. In the proof, we use the example of an increase in τ as $\frac{\partial \Phi}{\partial \tau} > 0$. Replacing $\partial\tau$ with $\partial\lambda$, $\partial(-\tau^W)$, $\partial(-\delta)$ or $\partial(-\delta^W)$ provides the proof for the other cases.

Consider a realization of $\{a_t\}$. Then the sample variance of inflation is

$$\sigma_\pi^2 = \frac{1}{N-1} \sum_{t=1}^{N} (\pi_t - \bar{\pi})^2,$$

where $\bar{\pi}$ is the average of inflation under $\{a_t\}$. We use the closed-form solution for π_i to get

$$\begin{aligned} \pi_t \quad &= \quad -a_t (\Upsilon_a) \left[\frac{1}{(m_1 - \rho_a)(m_2 - \rho_a)} \right], \\ \bar{\pi} \quad &= \quad -\bar{a} (\Upsilon_a) \left[\frac{1}{(m_1 - \rho_a)(m_2 - \rho_a)} \right]. \end{aligned}$$

where $\bar{\pi}$ is average inflation. Then, we have

$$\sigma_\pi^2 = \left(-\Upsilon_a \left[\frac{1}{(m_1 - \rho_a)(m_2 - \rho_a)} \right] \right)^2 \sigma_a^2,$$

where σ_a^2 is the sample variance of a which is exogenous. If the term inside the parentheses multiplying σ_a^2 is decreasing in τ (since the term is less than zero), we are done. Substituting in for Υ_a we obtain

$$-\Upsilon_a \left[\frac{1}{(m_1 - \rho_a)(m_2 - \rho_a)} \right] = -\frac{\kappa}{\beta\sigma(1-\Phi)} \frac{1+\varphi}{\sigma+\varphi} [\sigma(1-\rho_a) + \phi_y] \left[\frac{1}{(m_1 - \rho_a)(m_2 - \rho_a)} \right].$$

Only Φ, m_1 and m_2 depend on τ. Therefore, we only need to know how the following changes with τ

$$-\frac{1}{(1-\Phi)} \left[\frac{1}{(m_1 - \rho_a)(m_2 - \rho_a)} \right].$$

Note that

$$(m_1 - \rho_a)(m_2 - \rho_a) = c + b\rho_a + \rho_a^2.$$

where b and c are the coefficients from the difference equation for inflation (13) and defined as $b \equiv -1 - \frac{1}{\beta} + \frac{1}{\beta\sigma(1-\Phi)} (\phi_{E\pi}\kappa - \beta\phi_y - \kappa)$ and $c \equiv \frac{1}{\beta} + \frac{\kappa\phi_\pi + \phi_y}{\beta\sigma(1-\Phi)}$. Therefore, omitting the leading negative sign, we have,

$$\frac{\partial}{\partial \tau}\left\{\frac{1}{(1-\Phi)}\left[c+b\rho_a+\rho_a^2\right]^{-1}\right\} = $$

$$-\frac{1}{(1-\Phi)}\left[c+b\rho_a+\rho_a^2\right]^{-2}\left[\frac{\partial c}{\partial \tau}+\frac{\partial b}{\partial \tau}\rho_a\right]$$
$$-\left[c+b\rho_a+\rho_a^2\right]^{-1}\frac{1}{(1-\Phi)^2}\frac{\partial \Phi}{\partial \tau}.$$

Using the definitions of b, c, $\frac{\partial c}{\partial \tau}$, $\frac{\partial b}{\partial \tau}$, we get

$$\begin{aligned}\frac{\partial}{\partial \tau}\left\{\frac{1}{(1-\Phi)}\left[c+b\rho_a+\rho_a^2\right]^{-1}\right\} &= \frac{1}{(1-\Phi)}\left[c+b\rho_a+\rho_a^2\right]^{-2}\frac{\partial \Phi}{\partial \tau}\\ &\left\{-\frac{1}{(1-\Phi)}\left[\frac{(\phi_\pi\kappa+\phi_y)}{\beta\sigma(1-\Phi)^2}+\frac{(\phi_{E\pi}\kappa-\beta\phi_y-\kappa)}{\beta\sigma(1-\Phi)^2}\rho_a\right]\right.\\ &\left.-\left[\frac{1}{\beta}+\frac{\kappa\phi_\pi+\phi_y}{\beta\sigma(1-\Phi)}+\left[-1-\frac{1}{\beta}+\frac{(\phi_{E\pi}\kappa-\beta\phi_y-\kappa)}{\beta\sigma(1-\Phi)}\right]\rho_a+\rho_a^2\right]\right\}.\end{aligned}$$

Simplifying, we obtain

$$\begin{aligned}\frac{\partial}{\partial \tau}\left\{\frac{1}{(1-\Phi)}\left[c+b\rho_a+\rho_a^2\right]^{-1}\right\} &= \frac{1}{(1-\Phi)}\left[c+b\rho_a+\rho_a^2\right]^{-2}\frac{\partial \Phi}{\partial \tau}\\ &\left\{\left(\rho_a-\frac{1}{\beta}\right)(1-\rho_a)-2\frac{(\phi_\pi\kappa+\phi_y)}{\beta\sigma(1-\Phi)}-2\frac{(\phi_{E\pi}\kappa-\beta\phi_y-\kappa)}{\beta\sigma(1-\Phi)}\rho_a\right\}.\end{aligned}$$

Since we previously ignore the negative sign, we need the term in brackets to be positive as $\frac{\partial \Phi}{\partial \tau}>0$. The term is positive so long as

$$\phi_\pi > \frac{1}{\kappa}\left\{-\left(\frac{1}{\beta}-\rho_a\right)(1-\rho_a)\frac{\beta\sigma(1-\Phi)}{2}-(1-\rho_a\beta)\phi_y+\rho_a\kappa\right\}.$$

Therefore $\phi_\pi>\rho_a$ is sufficient for $\frac{\partial \sigma_\pi^2}{\partial \tau}>0$. Taking the square root of the variance, σ_π^2, we obtain the definition of volatility. Therefore, inflation volatility is increasing in τ.

Using the result regarding inflation volatility and the Phillips curve, we can establish that the output gap volatility increases with τ. Note that $E\pi_{t+1}=\rho_a\pi_t$; see (A24). Therefore, using the Phillips curve we have

$$x_t=\frac{1}{\kappa}(1-\rho_a\beta)\pi_t.$$

The term multiplying π_t does not depend on τ. Therefore, since $\frac{\partial \sigma_\pi}{\partial \tau}>0$, $\frac{\partial \sigma_x}{\partial \tau}>0$, which yields the desired result that $\sigma_x(\Phi_1)>\sigma_x(\Phi_2)$ whenever $\Phi_1>\Phi_1$ due to differences in heterogeneity parameters.

Closed-form solutions for individual consumption in the stochastic setting. Following the same procedure, as described in Appendix A.5 for the deterministic setting, we obtain the formulas for consumption of unconstrained and constrained agents. For the unconstrained agent, we have

$$c_t^U=-\frac{1}{\sigma}\left\{\widetilde{H}_t+\frac{1}{1-\rho_v}v_t\right\}-\frac{1}{\sigma}\left(\frac{1}{1-\rho_{r^n}}\right)r_t^n+\frac{1}{\sigma}z_t, \tag{A27}$$

and for the constrained agent, we have

$$c_t^K=\chi\left[-\frac{1}{\sigma(1-\Phi)}\left\{\widetilde{H}_t+\frac{1}{1-\rho_v}v_t\right\}-\frac{1}{\sigma(1-\Phi)}\left(\frac{1}{1-\rho_{r^n}}\right)r_t^n+\frac{1}{\sigma}z_t,\right] \tag{A28}$$

A.6 Supplementary results for Section 7

Responses of income to contemporaneous productivity shocks. In Section 7, we claim that consumption inequality increases in response to productivity shock so long as $\frac{\partial y_t}{\partial a_t} < \frac{1+\varphi}{\sigma+\varphi}$. To verify this claim, we start from the aggregate Euler equation

$$\frac{\partial y_t}{\partial a_t} = -\left(1+\frac{\phi_y}{\sigma(1-\Phi)}\right)^{-1}\frac{1}{\sigma(1-\Phi)}\left(\phi_\pi\frac{\partial \pi_t}{\partial a_t}\right) - \left(1+\frac{\phi_y}{\sigma(1-\Phi)}\right)^{-1}\frac{1}{\sigma(1-\Phi)}\left(\frac{1+\varphi}{\sigma+\varphi}\right)\sigma\Phi.$$

Using the Phillips curve (11), we have

$$\frac{\partial \pi_t}{\partial a_t} = -\kappa\frac{\partial \mu_t}{\partial a_t}.$$

Combining these equations gives us

$$\frac{dy_t}{da_t} = \left(1+\frac{\phi_y}{\sigma(1-\Phi)}\right)^{-1}\frac{1}{\sigma(1-\Phi)}\left(\phi_\pi\kappa\frac{d\mu_t}{da_t}\right) - \left(1+\frac{\phi_y}{\sigma(1-\Phi)}\right)^{-1}\frac{1}{\sigma(1-\Phi)}\left(\frac{1+\varphi}{\sigma+\varphi}\right)\sigma\Phi.$$

Next, we need to substitute in for $\frac{d\mu_t}{da_t}$. Totally differentiating the markup with respect to a_t and rearranging, we get

$$\frac{d\mu_t}{da_t} = (1+\varphi) - (\sigma+\varphi)\frac{\partial y_t}{\partial a_t}.$$

Substituting in for $\frac{d\mu_t}{da_t}$, we have

$$\frac{\partial y_t}{\partial a_t} = \left[1+\left(1+\frac{\phi_y}{\sigma(1-\Phi)}\right)^{-1}\frac{\phi_\pi\kappa\,(\sigma+\varphi)}{\sigma(1-\Phi)}\right]^{-1}\left(1+\frac{\phi_y}{\sigma(1-\Phi)}\right)^{-1}\frac{1}{\sigma(1-\Phi)}\left(\phi_\pi\kappa\,(1+\varphi) - \left(\frac{1+\varphi}{\sigma+\varphi}\right)\sigma\Phi\right).$$

Simplifying the terms, we have

$$\begin{aligned}\left[1+\left(1+\frac{\phi_y}{\sigma(1-\Phi)}\right)^{-1}\frac{\phi_\pi\kappa\,(\sigma+\varphi)}{\sigma(1-\Phi)}\right]^{-1} &= \frac{\sigma(1-\Phi)+\phi_y}{\sigma(1-\Phi)+\phi_y+\phi_\pi\kappa\,(\sigma+\varphi)},\\ \left(1+\frac{\phi_y}{\sigma(1-\Phi)}\right)^{-1} &= \frac{\sigma(1-\Phi)}{\sigma(1-\Phi)+\phi_y}.\end{aligned}$$

This leaves us with

$$\frac{\partial y_t}{\partial a_t} = \frac{\phi_\pi\kappa\,(1+\varphi) - \left(\frac{1+\varphi}{\sigma+\varphi}\right)\sigma\Phi}{\sigma(1-\Phi)+\phi_y+\phi_\pi\kappa\,(\sigma+\varphi)}.$$

We need this to be less than $\frac{1+\varphi}{\sigma+\varphi}$:

$$\frac{\phi_\pi\kappa\,(1+\varphi) - \left(\frac{1+\varphi}{\sigma+\varphi}\right)\sigma\Phi}{\sigma(1-\Phi)+\phi_y+\phi_\pi\kappa\,(\sigma+\varphi)} < \frac{1+\varphi}{\sigma+\varphi},$$

$$\frac{\phi_\pi\kappa\,(1+\varphi)\,(\sigma+\varphi) - (1+\varphi)\,\sigma\Phi}{\phi_\pi\kappa\,(1+\varphi)\,(\sigma+\varphi) + (1+\varphi)\,\phi_y + (1+\varphi)\,\sigma(1-\Phi)} < 1.$$

The inequality holds as $0 < \Phi < 1$ and $\phi_y \geq 0$. Therefore, $\frac{dy_t}{da_t} < \frac{1+\varphi}{\sigma+\varphi}$, and $\frac{dy_t}{da_t} > 0$.

Claim about the response of markups to productivity shocks and changes in output. Differentiation the markup with respect to y_t yields

$$d\mu_t = -\left(\sigma + \varphi\right) dy_t.$$

Therefore, $\left[\frac{d\mu_t}{da_t}\right]\left[\frac{d\mu_t}{dy_t}\right]^{-1}$ is

$$\left[\frac{d\mu_t}{da_t}\right]\left[\frac{d\mu_t}{dy_t}\right]^{-1} = \frac{(1+\varphi) - (\sigma+\varphi)\frac{\partial y_t}{\partial a_t}}{-\left(\sigma+\varphi\right)} = -\left[\frac{1+\varphi}{\sigma+\varphi} - \frac{\partial y_t}{\partial a_t}\right].$$

Multiplying both sides by (-1) provides the desired result.

Appendix B. TANK with capital

In this section, we present the derivations for the TANK model with capital.

B.1 Non-linear model

Household's FOC's. The FOCs of the unconstrained agent's problem (1), (35a), (36) are given by

$$\left(I_t^U\right): \qquad q_t = \left(1 - \zeta\left(\frac{I_t}{K_{t-1}^U} - \delta^k\right)\right)^{-1}, \tag{B1}$$

$$\begin{aligned}\left(K_t^U\right): \qquad q_t = \beta E_t \Bigg[\frac{Z_{t+1}}{Z_t}\left(\frac{C_{t+1}^U}{C_t^U}\right)^{-\sigma} \Big\{ & (1-\tau^k)R_{t+1}^K \\ & + q_{t+1}\left(1 - \delta^k - \frac{\zeta}{2}\left(\frac{I_{t+1}^U}{K_t} - \delta^k\right)^2 + \zeta\left(\frac{I_{t+1}^U}{K_t^U} - \delta^k\right)\frac{I_{t+1}^U}{K_t^U}\right)\Big\}\Bigg],\end{aligned} \tag{B2}$$

$$\left(F_t^U\right): \qquad Q_t = \beta E_t\left\{\frac{Z_{t+1}}{Z_t}\left(\frac{C_{t+1}^U}{C_t^U}\right)^{-\sigma}\left[Q_{t+1} + (1-\delta)D_{t+1}\right]\right\}, \tag{B3}$$

$$\left(B_t^U\right): \qquad 1 = \beta R_t E_t\left\{\frac{Z_{t+1}}{Z_t}\left(\frac{C_{t+1}^U}{C_t^U}\right)^{-\sigma}\frac{1}{\pi_{t+1}}\right\}, \tag{B4}$$

$$\left(N_t^U\right): \qquad \left(1-\delta^W\right)\frac{W_t}{P_t} = (C_t^U)^{\sigma}(N_t^U)^{\varphi}, \tag{B5}$$

where $q_t \equiv \frac{\vartheta_t}{\eta_t}$, with ϑ_t being the multiplier on the capital-accumulation equation (36), and η_t being the multiplier on the budget constraint (35a).

General consumption and income gaps. We first present our derivations for the consumption gap, which can be written as

$$\Gamma_t = 1 - \frac{\left(1-\delta^W\right)\frac{W_t}{P_t}N_t + T_{D,t}^K + T_{W,t}^K}{\left(1-\delta^W\right)\frac{W_t}{P_t}N_t + \frac{1-\delta}{1-\lambda}D_t + T_{D,t}^U + T_{W,t}^U + (1-\tau^k)R_t^k K_{t-1}^U - I_t^U}.$$

It leads to

$$\Gamma_t = \frac{\frac{1-\delta}{1-\lambda}D_t + (T_{D,t}^U - T_{D,t}^K) + (T_{W,t}^U - T_{W,t}^K) + (1-\tau^k)R_t^k K_{t-1}^U - I_t^U}{W_t N_t + \frac{1-\delta}{1-\lambda}D_t + T_{D,t}^U + T_{W,t}^U + (1-\tau^k)R_t^k K_{t-1}^U - I_t^U}.$$

Aggregate capital follows from the market clearing condition $K_t^d = (1-\lambda)K_{t-1}^U \equiv K_t$; aggregate investment is given by $I_t = (1-\lambda)I_t^U$. Equations (37) and (38) imply

$$\begin{aligned} T_{D,t}^U - T_{D,t}^K &= \left(\frac{\tau\lambda}{1-\lambda} + \tau\right)\left(\delta D_t + \tau^k R_t^k K_t\right), \\ T_{W,t}^U - T_{W,t}^K &= \left(\frac{\tau\lambda}{1-\lambda} + \tau\right)\left(\delta^W \frac{W_t}{P_t} N_t\right), \end{aligned}$$

which we substitute back into the consumption-gap formula to obtain

$$\Gamma_t = \frac{\frac{1-\delta}{1-\lambda} D_t + \left(\frac{\tau\lambda}{1-\lambda} + \tau\right)\left(\delta D_t + \tau^k R_t^k K_t\right) + \left(\frac{\tau^W\lambda}{1-\lambda} + \tau^W\right)\delta^W \frac{W_t}{P_t} N_t + \frac{1-\tau^k}{1-\lambda} R_t^k K_t - I_t^U}{\frac{W_t}{P_t} N_t + \frac{1-\delta}{1-\lambda} D_t + \left(1 + \frac{\tau\lambda}{1-\lambda}\right)\left(\delta D_t + \tau^k R_t^k K_t\right) + \left(1 + \frac{\tau^W\lambda}{1-\lambda}\right)\delta^W \frac{W_t}{P_t} N_t^U + \frac{1-\tau^k}{1-\lambda} R_t^k K_t - I_t^U}.$$

We multiply top and bottom by $1-\lambda$ and rearrange the terms to get

$$\Gamma_t = \frac{\left(1 + (\tau-1)\delta\right) D_t + \tau^W \delta^W \frac{W_t}{P_t} N_t + \left(1 + \left(\tau^k - 1\right)\tau\right) R_t^k K_t - (1-\lambda) I_t^U}{\left(1 + (\tau-1)\lambda\delta\right) D_t + \left[(1-\lambda) + \left(1 + \left(\tau^W - 1\right)\lambda\right)\delta^W\right] \frac{W_t}{P_t} N_t + \left(1 + (\tau-1)\lambda\tau^k\right) R_t^k K_t - (1-\lambda) I_t^U}. \tag{B6}$$

Dividends are a difference between aggregate output and costs of inputs,

$$D_t = A_t K_t^\alpha N_t^{1-\alpha} \triangle_t - \frac{W_t}{P_t} N_t - R_t^k K,$$

where $\triangle_t \equiv 1 - \frac{\varepsilon}{\zeta}\left(\Pi_t - 1\right)^2$ is the price dispersion. Dividing the previous equation for D_t by $R_t^k K_t$ yields

$$\frac{D_t}{R_t^k K_t} = \frac{A_t K_t^{\alpha-1} N_t^{1-\alpha} \triangle_t}{R_t^k} - \frac{\frac{W_t}{P_t}}{R_t^k} \frac{N_t}{K_t} - 1.$$

From cost minimization, we have $\frac{\frac{W_t}{P_t}}{R_t^k} \frac{N_t}{K_t} = \frac{1-\alpha}{\alpha}$, which implies

$$\frac{D_t}{R_t^k K_t} = \frac{1}{\alpha} \mathcal{M} \triangle_t - \frac{1-\alpha}{\alpha} - 1.$$

We multiply the numerator and denominator of Γ_t in (B6) by $\frac{1}{R_t^k K_t}$ to get

$$\Gamma_t = \frac{\left(1 + (\tau-1)\delta\right)\left(\frac{1}{\alpha}\mathcal{M}\triangle_t - \frac{1-\alpha}{\alpha} - 1\right) + \tau\delta^W \frac{1-\alpha}{\alpha} + \left(1 + \left(\tau^k - 1\right)\tau\right) - (1-\lambda)\frac{I_t^U}{R_t^k K_t}}{\left[(1-\lambda) + \left(1 + (\tau-1)\lambda\right)\delta^W\right]\frac{1-\alpha}{\alpha} + \left(1 + (\tau-1)\lambda\delta\right)\left(\frac{1}{\alpha}\mathcal{M}\triangle_t - \frac{1-\alpha}{\alpha} - 1\right) + \left(1 + (\tau-1)\lambda\tau^k\right) - (1-\lambda)\frac{I_t^U}{R_t^k K_t}}.$$

Let us denote the steady state ratio of investment to $R^k K$ as $\Sigma \equiv \frac{I}{R^k K}$. The steady state consumption gap becomes

$$\Gamma = \frac{\left(1 + (\tau-1)\delta\right)\left(\frac{1}{\alpha}\mathcal{M} - \frac{1-\alpha}{\alpha} - 1\right) + \tau\delta^W \frac{1-\alpha}{\alpha} + \left(1 + \left(\tau^k - 1\right)\tau\right) - \Sigma}{\left[(1-\lambda) + \left(1 + \left(\tau^W - 1\right)\lambda\right)\delta^W\right]\frac{1-\alpha}{\alpha} + \left(1 + (\tau-1)\lambda\delta\right)\left(\frac{1}{\alpha}\mathcal{M} - \frac{1-\alpha}{\alpha} - 1\right) + \left(1 + (\tau-1)\lambda\tau^k\right) - \Sigma} \equiv \frac{\mathcal{N}_0^\theta}{\mathcal{D}_0^\theta}.$$

We proceed similarly to derive the income gap, Γ_t^Y:

$$\Gamma_t^Y = \frac{\left(1 + (\tau-1)\delta\right)\left(\frac{1}{\alpha}\mathcal{M}_t\triangle_t - \frac{1-\alpha}{\alpha} - 1\right) + \tau\delta^W \frac{1-\alpha}{\alpha} + \left(1 + \left(\tau^k - 1\right)\tau\right)}{\left[(1-\lambda) + \left(1 + \left(\tau^W - 1\right)\lambda\right)\delta^W\right]\frac{1-\alpha}{\alpha} + \left(1 + (\tau-1)\lambda\delta\right)\left(\frac{1}{\alpha}\mathcal{M}_t\triangle_t - \frac{1-\alpha}{\alpha} - 1\right) + \left(1 + (\tau-1)\lambda\tau^k\right)}.$$

In steady state, we have

$$\Gamma^Y = \frac{\left(1 + (\tau-1)\delta\right)\left(\frac{1}{\alpha}\mathcal{M} - \frac{1-\alpha}{\alpha} - 1\right) + \tau\delta^W \frac{1-\alpha}{\alpha} + \left(1 + \left(\tau^k - 1\right)\tau\right)}{\left[(1-\lambda) + \left(1 + \left(\tau^W - 1\right)\lambda\right)\delta^W\right]\frac{1-\alpha}{\alpha} + \left(1 + (\tau-1)\lambda\delta\right)\left(\frac{1}{\alpha}\mathcal{M} - \frac{1-\alpha}{\alpha} - 1\right) + \left(1 + (\tau-1)\lambda\tau^k\right)} \equiv \frac{\mathcal{N}_0^Y}{\mathcal{D}_0^Y}.$$

Specific adjustment costs. We now present an additional derivation of the consumption gap in the case of the adjustment cost function in (36). From the FOC (B1) for q_t, we obtain

$$\frac{I_t^U}{K_t} = \frac{q_t - 1 + \Gamma q_t \delta^k}{(1-\lambda)\,\Gamma q_t}.$$

Therefore, the consumption gap becomes

$$\Gamma_t = \frac{(1+(\tau-1)\,\delta)\left(\frac{1}{\alpha}\mathcal{M}_t\triangle_t - \frac{1-\alpha}{\alpha} - 1\right) + \tau^W\delta^W\frac{1-\alpha}{\alpha} + \left(1+\left(\tau^k-1\right)\tau\right) - \frac{q_t-1+\Gamma q_t\delta^k}{\Gamma q_t R_t^k}}{\left[1+\left(\tau^W\delta^W-1\right)\lambda\right]\frac{1-\alpha}{\alpha} + (1+(\tau-1)\,\lambda\delta)\left(\frac{1}{\alpha}\mathcal{M}_t\triangle_t - \frac{1-\alpha}{\alpha} - 1\right) + \left(1+(\tau-1)\,\lambda\tau^k\right) - \frac{q_t-1+\Gamma q_t\delta^k}{\Gamma q_t R_t^k}}, \tag{B7}$$

where $\triangle_t$ is price dispersion, defined as in the model with no capital (see Appendix A.1).

In steady state, $\triangle_t = 1$, $q_t = 1$, $\mathcal{M}_t = \mathcal{M}$, and $\frac{1}{R_t^k} = \frac{1-\tau^k}{\frac{1}{\beta}-\left(1-\delta^k\right)}$. Therefore, the consumption gap in a zero-inflation steady state is

$$\Gamma = \frac{(1+(\tau-1)\,\delta)\left(\frac{1}{\alpha}\mathcal{M} - \frac{1-\alpha}{\alpha} - 1\right) + \tau^W\delta^W\frac{1-\alpha}{\alpha} + \left(1+\left(\tau^k-1\right)\tau\right) - \frac{\left(1-\tau^k\right)\delta^k}{\frac{1}{\beta}-\left(1-\delta^k\right)}}{\left[1-\lambda+\left(1+\left(\tau^W-1\right)\lambda\right)\delta^W\right]\frac{1-\alpha}{\alpha} + (1+(\tau-1)\,\lambda\delta)\left(\frac{1}{\alpha}\mathcal{M} - \frac{1-\alpha}{\alpha} - 1\right) + \left(1+(\tau-1)\,\lambda\tau^k\right) - \frac{\left(1-\tau^k\right)\delta^k}{\frac{1}{\beta}-\left(1-\delta^k\right)}} \equiv \frac{\mathcal{N}_0}{\mathcal{D}_0}.$$

Supply side. Cost minimization of an individual intermediate-good producer i implies that the real marginal costs MC_t are the same for all producers,

$$\mathrm{MC}_t = \frac{1}{A_t}\alpha^{-\alpha}(1-\alpha)^{-(1-\alpha)}\left(R_t^k\right)^{\alpha}\left(\frac{W_t}{P_t}\right)^{1-\alpha}.$$

The Phillips curve following from the firm's profit-maximization is the same as in the model with no capital and is given by (A4).

Market clearing. We assume that the real wage is determined by the unconstrained agent's FOC and hours are identical across agents $N_t = N_t^U = N_t^K$. Since all firms set the same price, we can write the economy's resource constraint as

$$C_t + I_t = Y_t\triangle_t.$$

B.2 Log-linearization of the non-linear model

First, we present the log-linearized consumption and income gaps for the general specification of the adjustment cost function. Log-linearizing, we have

$$\begin{aligned}
\gamma_t \quad = \quad & \left(\frac{(1+(\tau-1)\,\delta)}{\mathcal{N}_0^\theta} - \frac{(1+(\tau-1)\,\lambda\delta)}{\mathcal{D}_0^\theta}\right)\left(\frac{\mathcal{M}}{\alpha}\right)\mu_t + \\
& \left(\frac{1}{\mathcal{D}_0^\theta} - \frac{1}{\mathcal{N}_0^\theta}\right)\Sigma\hat{\imath}_t + \\
& \left(\frac{1}{\mathcal{N}_0^\theta} - \frac{1}{\mathcal{D}_0^\theta}\right)\Sigma\hat{r}_t^k + \\
& \left(\frac{1}{\mathcal{N}_0^\theta} - \frac{1}{\mathcal{D}_0^\theta}\right)\Sigma k_t.
\end{aligned}$$

This gives the following expressions for the θ_i's:

$$
\begin{aligned}
\theta_1 &= \left(\frac{(1+(\tau-1)\,\delta)}{\mathcal{N}_0^\theta} - \frac{(1+(\tau-1)\,\lambda\delta)}{\mathcal{D}_0^\theta}\right)\frac{\mathcal{M}}{\alpha}, \\
\theta_2 &= \left(\frac{1}{\mathcal{D}_0^\theta} - \frac{1}{\mathcal{N}_0^\theta}\right)\Sigma, \\
\theta_3 &= \left(\frac{1}{\mathcal{N}_0^\theta} - \frac{1}{\mathcal{D}_0^\theta}\right)\Sigma = -\theta_2, \\
\theta_4 &= \left(\frac{1}{\mathcal{N}_0^\theta} - \frac{1}{\mathcal{D}_0^\theta}\right)\Sigma = -\theta_2.
\end{aligned}
$$

Log linearizing, the income gap, we arrive at

$$
\gamma_t^Y = \left(\frac{(1+(\tau-1)\,\delta)}{\mathcal{N}_0^Y} - \frac{(1+(\tau-1)\,\lambda\delta)}{\mathcal{D}_0^Y}\right)\left(\frac{\mathcal{M}}{\alpha}\right)\mu_t,
$$

$$
\theta^Y = \left(\frac{(1+(\tau-1)\,\delta)}{\mathcal{N}_0^Y} - \frac{(1+(\tau-1)\,\lambda\delta)}{\mathcal{D}_0^Y}\right)\frac{\mathcal{M}}{\alpha}.
$$

Now we present the derivation for the adjustment cost function used in (36). Taking logs and linearizing (B7) gives us the following first-order approximation:

$$
\begin{aligned}
\gamma_t = & \left(\frac{(1+(\tau-1)\,\delta)}{\mathcal{N}_0} - \frac{(1+(\tau-1)\,\lambda\delta)}{\mathcal{D}_0}\right)\left(\frac{1}{\alpha}\triangle\right)(\mathcal{M}_t - \mathcal{M}) + \\
& \left(\frac{(1+(\tau-1)\,\delta)}{\mathcal{N}_0} - \frac{(1+(\tau-1)\,\lambda\delta)}{\mathcal{D}_0}\right)\mathcal{M}\,(-\zeta)\,(\Pi-1)\,(\Pi_t - \Pi) + \\
& \left(\frac{1}{\mathcal{D}_0} - \frac{1}{\mathcal{N}_0}\right)\left[\frac{1}{q^2\zeta}\left(\frac{1-\tau^k}{\frac{1}{\beta}-(1-\delta^k)}\right)\right](q_t - q) + \\
& \left(\frac{1}{\mathcal{N}_0} - \frac{1}{\mathcal{D}_0}\right)\left(\frac{q-1+\Gamma q\delta^k}{\Gamma q}\right)\left(\frac{1-\tau^k}{\frac{1}{\beta}-(1-\delta^k)}\right)\frac{1}{R^k}\left(R_t^k - R^k\right),
\end{aligned}
$$

where R^k is the previously defined steady-state rate of return on physical capital. Since $\Pi_t = 1$ in steady state, the above reduces to (41) in the main text, where the constants are defined $\varkappa_1$,$\varkappa_2$, and $\varkappa_3$ by

$$
\begin{aligned}
\varkappa_1 &\equiv \left(\frac{(1+(\tau-1)\,\delta)}{\mathcal{N}_0} - \frac{(1+(\tau-1)\,\lambda\delta)}{\mathcal{D}_0}\right)\cdot\frac{\mathcal{M}}{\alpha}, \\
\varkappa_2 &\equiv \left(\frac{1}{\mathcal{D}_0} - \frac{1}{\mathcal{N}_0}\right)\left[\frac{1}{\zeta}\left(\frac{1-\tau^k}{\frac{1}{\beta}-(1-\delta^k)}\right)\right], \\
\varkappa_3 &\equiv \left(\frac{1}{\mathcal{N}_0} - \frac{1}{\mathcal{D}_0}\right)\left(\frac{\left(1-\tau^k\right)\delta^k}{\frac{1}{\beta}-(1-\delta^k)}\right).
\end{aligned}
$$

Therefore, the consumption gap is affected through three channels, related to the markup, investment and capital. Under the assumption that $\Gamma > 0$ so that consumption of unconstrained agents is greater than that of constrained agents in steady state and $\Gamma < 1$, we must have $\mathcal{D}_0 > \mathcal{N}_0$.

The other log-linearized equations include

$$
\begin{array}{ll}
(1) & \sigma c_t + \varphi n_t = w_t, \\
(2) & y_t = a_t + \alpha k_t + (1-\alpha) n_t, \\
(3) & k_{t+1} = \delta^k \iota_t + \left(1-\delta^k\right) k_t, \\
(4) & \hat{q}_t = \zeta \delta^k (\iota_t - k_t), \\
(5) & c_t^U = E_t\left\{c_{t+1}^U\right\} - \frac{1}{\sigma}(i_t - E_t\{\pi_{t+1}\}) - \frac{1}{\sigma} E_t\{\triangle z_{t+1}\}, \\
(6) & i_t = i_t^* + \phi_y y_t + \phi_\pi \pi_t + \phi_{E\pi}\pi_{t+1} + v_t, \\
(7) & \mu_t = -r_t^k - k_t + y_t, \\
(8) & \mu_t = -w_t - n_t + y_t, \\
(9) & \pi_t = \beta E_t\{\pi_{t+1}\} - \mu_t \frac{\varepsilon-1}{\zeta}, \\
(10) & \widehat{q}_t = \frac{\left(1-\tau^k\right)}{\frac{1}{\beta}-\left(1-\delta^k\right)} E_t\{\widehat{q}_{t+1}\} + \sigma \hat{c}_t^U - \sigma E_t\left\{c_{t+1}^U\right\} + E_t\{\triangle z_{t+1}\} + \frac{\left[1-\beta\left(1-\delta^k\right)\right]}{1-\tau^k} E_t\left\{r_{t+1}^k\right\}, \\
(11) & c_t = c_t^U - \frac{\lambda}{1-\lambda\gamma}\gamma_t, \\
(12) & \varpi c_t + (1-\varpi)\iota_t = y_t, \\
(13) & \mu_t = -\left(\frac{\alpha+\varphi}{1-\alpha} + \sigma\frac{1}{\varpi}\right) x_t,
\end{array}
$$

where all the variables are in log deviations; w_t is the real wage; r_t^k is the return on physical capital; ι_t is investment; $\widehat{q}_t$ is the value of installed capital in terms of consumption; μ_t is the markup; is the $\triangle z_{t+1} \equiv \ln Z_{t+1} - \ln Z_t$; ϖ is a steady state share of consumption in output, $\varpi \equiv 1 - \frac{\alpha\delta^k\left(\frac{\varepsilon-1}{\varepsilon}\right)}{R^k}$. Therefore, the TANK model with capital can be represented as a system of 14 equations (13 above equations plus formula (41) for γ_t) in 14 unknowns $\left\{c_t, c_t^U, n_t, y_t, x_t, k_t, \iota_t, \hat{q}_t, \mu_t, \gamma_t, \pi_t, w_t, i_t, r_t^k\right\}$. There are 3 shocks $\{v_t, a_t, z_t\}$, which can be either anticipated or unanticipated; in the latter case, they are defined by AR(1) processes given in the description of the model with no capital.

B.3 Forward guidance

We now investigate how heterogeneity impacts the effectiveness of forward guidance. We consider forward guidance along with active policy as in the model without capital. Specifically, we consider a Taylor rule with $\phi_{E\pi} = 1$, $\phi_y = 0.1$ and $\phi_\pi = 0$ and three values for the share of constrained agents, $\{0, 0.21, 0.38\}$.

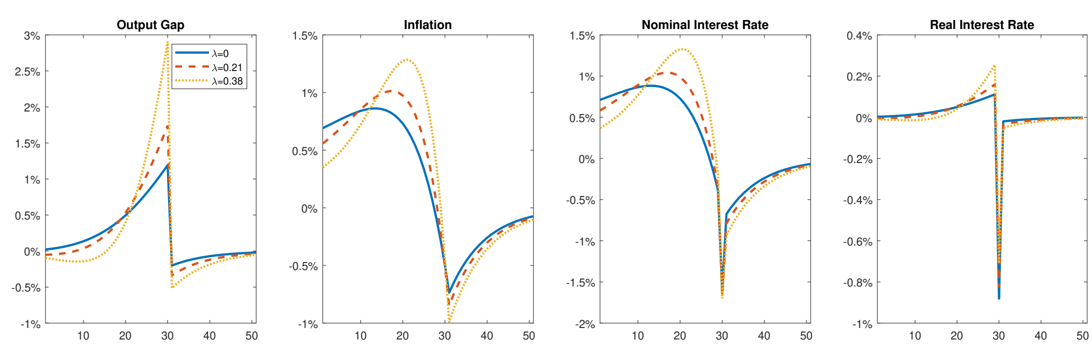

Figure 8: Response to the monetary-policy shock at $T = 30$ with $\phi_{E\pi} = 1$, $\phi_y = 0.1$ and $\phi_\pi = 0$.

In Figure 8, we plot the impulse responses of the output gap, inflation, nominal interest rate and real interest rate to a monetary policy shock at $T = 30$. From the first panel, we see that larger λ leads to greater peak responses, just as in the case of the model without capital. Additionally, from the panel presenting the responses for inflation, we see a larger λ leads to a smaller initial response of inflation. Therefore, with $\phi_y = 0.1$, we see that the impact output stabilization has on aggregate responses is similar to the effect in the model without capital.

B.4 Consumption and income inequality

Table 6 presents the results about volatility of consumption and income inequality. We consider three pairs of $\left(\tau, \tau^W\right)$: no transfer ($\tau = 1$, $\tau^W = 0$), asset transfer ($\tau = .9$, $\tau^W = 0$), and labor transfer ($\tau = 1$, $\tau^W = -.1\left(\frac{1-\lambda}{\lambda}\right)$).

Table 6: Consumption and income inequality under different transfers

	No Transfers	Asset Transfers	Labor Transfers
Correlation of consumption gap with technology	-0.3612	-0.5940	-0.3606
Correlation of income gap with technology	0.9931	0.9931	0.9931

.