# Optimal Monetary Policy with Behavioral Consumers*

Marco Airaudo† Archil Dvalishvili‡ Mikheil Dvalishvili§

February 19, 2023

**Abstract**

Motivated by a large body of experimental and field evidence documenting the occurrence of preference reversals and present bias in intertemporal consumptions decisions, we study optimal monetary policy in a tractable New Keynesian model populated by Gul-Pesendorfer (Econometrica, 2001, 2004) *temptation-with-self-control* behavioral consumers. Through distortionary wealth effects in the Euler equation, public debt dynamics are no longer redundant for real activity and aggregate welfare: Ricardian equivalence fails. The latter, combined with a novel term related to debt volatility in the central bank's policy objective, induces a meaningfull inflation-output stabilization trade-off, making full neutralization of demand-side shocks no longer possible: divine coincidence fails. As the cognitive costs of self-control are negatively related to wealth volatility, the welfare costs of economic fluctuations are a declining function of temptation.

JEL Classification: E32, E44, E50.

Keywords: Optimal monetary policy; Temptation; Self-control; Inflation Targeting

## 1 Introduction

The baseline New Keynesian model used for policy analysis yields two strong predictions: 1) absent cost-push shocks in the Phillips curve, stabilizing inflation allows the central bank to stabilize also the welfare-relevant output gap, hence fully hedging the economy from demand-side disturbances

*We would like to thank participants to the 16th Annual Dynare Conference (Lancaster Management University) for useful feedbacks and comments.

†**Corresponding Author**. School of Economics, LeBow College of Business, Drexel University, Gerri C. LeBow Hall, 3220 Market Street, Philadelphia, PA, 19104, USA. Tel: +1-215-895-6982. Email: marco.airaudo@drexel.edu.

‡Ph.D. Student, Graduate Center, CUNY, New York.

§CUNY, New York

(*Divine Coincidence*); 2) public debt dynamics have neutral effects on households' decisions and their welfare (*Ricardian Equivalence*). The first is a direct consequence of the fact that, in the baseline model, the gap between the *natural* (flexible price) level of output and its *efficient* (first best) level is constant (possibly zero) and therefore does not respond to shocks. The second is implied by allowing the government access to non-distortionary taxation and assuming lack of liquidity/wealth effects from Treasury bonds holdings by the public.

From a theoretical standpoint, both predictions are not robust to reasonable amendments to the baseline framework (see some examples in the Related Literature section below). Moreover, the empirical evidence in their support is, at best, rather weak. On the one hand, structural VARs provide ample evidence of both inflation and real activity responding to changes in total factor productivity (see, for instance, Gali, 1999, and Gali and Rabanal, 2004), while the sizable standard deviation of residuals from Phillips curve's estimations suggests that keeping inflation stable does not necessarily stabilize output (see Blanchard, 2016).[1] On the other hand, the Ricardian equivalence hypothesis appears rather fragile looking both at econometric evidence on aggregate data (Ricciuti, 2003; Haug, 2020) and at incentivized individual responses to current and future tax changes in laboratory experiments (Meissner and Rostam-Afschar, 2017).

We show that both the divine coincidence and Ricardian equivalence break down in a New Keynesian economy populated by *behavioral consumers* à-la Gul and Pesendorfer (2001, 2004), that is, consumers characterized by *temptation with self-control* preferences (GP-preferences) whereby an internal conflict arises between the *ex ante* optimal long-run ranking of options and *ex post* short-run temptations. More specifically, in our set-up, the representative infinitely-lived household is tempted to behave like a *hand-to-mouth* consumer by using his entire financial wealth (e.g., Treasury bonds) for the purpose of immediate consumption. By exerting cognitive effort (self-control) - hence suffering some disutility - to resist the urge to consume - his optimal behavior trades off the temptation for immediate satisfaction (temptation utility) with long-run optimal consumption smoothing (commitment utility).[2]

[1] Non-zero (conditional) correlation of inflation and/or real activity with total factor productivity may also result from the central bank implementing sub-optimal policy rules (e.g. a simple Taylor rule responding to inflation and output). Chen et al. (2017) and Coroneo et al. (2018) provide econometric evidence for the U.S. in favor of optimal monetary policy under discretion, compared to alternative monetary policy regimes.

[2] Gul and Pesendorfer (2001, 2004) axiomatize temptation with self-control preferences, both in a static and dynmaic choice context, showing that individual utility must depend on the menu of available choices and not only on the actually made choice. A larger menu might then be worse than a smaller one if the former includes tempting choices which are eventually harmful. In a static context, a classic example is a meal at a resturant whose menu includes both tasty unhealthy dishes (e.g. a high sugar dessert) and healthier options (e.g. a fruit salad). Consumers

We use this behavioral New Keynesian framework to study the design of optimal monetary policy, in particular for what concerns a) the identification of central bank's objectives and its relevant constraints; b) the optimal response to shocks to total factor productivity and fiscal variables; and c) the consumption equivalent welfare costs of business cycle fluctuations. we find three key consequences of GP-preferences. **First**, with self-control costs depending on available wealth, households' anticipation of upcoming temptation-commitment trade-offs yields a generalized Euler equation featuring direct wealth effects from government bonds holdings: Ricardian equivalence fails! A second-order approximation to the representative household's welfare shows that the central bank should be less concerned about price stability, but be cognizant of the welfare consequences of public debt volatility.

**Second**, GP-preferences introduce a two-way feedback channel between public debt (responding to fiscal shocks and the real interest rate) and real activity (responding to the real interest rate, shocks affecting its natural level, and debt itself). This combined with the modified policy objective makes it impossible to fully neutralize shocks to total factor productivity and/or fiscal variables (government spending, fiscal surplus), and therefore obtain complete inflation and output gap stabilization: divine coincidence fails! The optimal targeting rule features, on the one hand, a dynamic trade-off between stabilizing current versus next period inflation and output gap, and, on the other hand, a positive response to the debt gap.

**Third**, we find the consumption-equivalent welfare costs of aggregate fluctuations to be strictly decreasing in temptation, with the possibility of welfare benefits when sufficiently strong temptation is complemented by a weaker impact of supply-side shocks. Under (concave) GP-preferences, some volatility in wealth is in fact desirable as it dampens the cognitive costs of self-control.

We provide empirical support for our modeling framework by bringing to the data two testable (steady state) implications of GP-preferences: 1) real interest rates should be, on average, higher in economies characterized by stronger temptation; and 2) conditional on the latter, they should be decreasing with respect to public debt. Using available evidence on the incidence of present bias/temptation in consumption-saving decisions across countries, we find empirical evidence in favor of both predictions.

The motivation to introduce *behavioral* elements into the baseline New Keynesian framework

would like to exclude ex ante tempting items from a choice menu - in our case, exclude the possibility of acting like hand–to-mouth - but, this not being a viable option, they exercise ex post welfare-reducing self-control to stick with the optimal intertemporal consumption/saving plan.

comes from a large body of experimental and field research documenting the occurrence of preference reversals and present bias in intertemporal consumptions decisions. More precisely, lab experiments document that subjective discounting between today and tomorrow is stronger than between any two future dates $t$ and $t+1$. This leads to time-inconsistent behavior that cannot be explained by the standard exponential discounting utility framework where, because of constant subjective discount rate between any two subsequent periods, future selves always choose according to the preferences of earlier selves (preferences are time consistent).[3] Gul and Pesendorfer's temptation with self-control framework is consistent with the observed reversals because small rewards are tempting only if immediately available (the case of a decision-maker sitting at time $t > 0$) but not if expected to occur in the future (the case of a decision maker sitting at $t = 0$).

Compared to the popular $\beta$-$\delta$ (hyperbolic/quasi-geometric discounting, HQGD) preferences introduced by Strotz (1956) and further refined by Harris and Laibson (2001), modeling temptation in consumption/saving decisions via GP-preferences presents several advantages. First, GP-preferences are time-consistent, which in turns allows to use standard recursive techniques to compute *unique* optimal decision rules. On the contrary, under HQGD, equilibrium multiplicity is an overwhelming issue, which requires to impose additional refinements (e.g. Markov Perfect Equilibrium, MPE).[4] Second, GP-preferences allow to uniquely define a welfare criterion for policy analysis. This is clearly not the case with HQGD due to the ongoing (unresolved) conflict between current and future selves. Third, GP-preferences can be tested directly by estimating the statistical significance of the wealth-to-consumption ratio in the Euler equation. This is more problematic with HQGD since wealth enters the Euler equation only implicitly via its unknown equilibrium impact on consumption. Moreover, convincing experimental evidence on the existence of Gul-Pesendorfer preferences is provided by Toussaerts (2018, 2019). Her lab and field experiments allow to distinguish between present-biased/time inconsistent agents (who value commitment as they expect to fall to temptation) and self-control types (who value commitment as it enables them to reduce/eliminate self-control costs), shows close coincidence between perceived and actual self-control (a sign of consumers' sophistication).

[3] Frederick et al. (2002) provide an extensive review of the discounted utility framework, highlight its key features and anomalies, overview the experimental and field evidence on the present bias in consumption-saving decisions, as well as introduce several alternative models of intertemporal choice to account for such evidence.

[4] However, as in shown in Maliar and Maliar (2006, 2016), even after imposing the MPE requirement, multiplicity is still not ruled out completely.

**Related Literature** Our paper is complementary to the literature on the fragility of both the divine coincidence and/or Ricardian equivalence to reasonable *structural* amendments to the baseline New Keynesian model. For what concerns the former, under standard preferences, a meaningful inflation-output gap stabilization trade-off occurs in presence of time-varying price mark-ups (Steinsson, 2003), real wage rigidities (Blanchard and Gali, 2005), a credit channel (Ravenna and Walsh, 2006), as well as in more sophisticated heterogeneous agents environments (HANK models, Acharya et al., 2022). Debt dynamics are instead non-neutral for real activity and therefore welfare with distortionary taxation (Linnemann, 2006), a Blanchard-Yaari-type stochastic OLG/perpetual youth structure (Leith and von Thadden, 2006; Rigon and Zanetti, 2018), and/or the introduction of liquidity effects from Treasury bonds holdings (Canzoneri et al., 2011; Michaillat and Saez, 2021).[5]

Our paper also contributes to the literature exploring the positive and normative implications of Gul-Pesendorfer preferences in dynamic macroeconomic models. Several contributions have assessed the effects of *temptation with self-control* for what concerns asset pricing (Krusell et al., 2002; DeJong and Ripoll, 2007; Airaudo, 2019), the design of optimal capital taxation (Krusell et al., 2010), retirement accounts and social security (Kumru and Thanopoulos, 2011), Friedman's rule (Hiraguchi, 2018), the welfare cost of business cycle fluctuations (Huang et al., 2015), the large share of privately held illiquid assets and the presence of "wealthy hand-to-mouth" consumers (Kovacs et al., 2021; Attanasio et al., 2022), housing and the mortgage market (Nakajima, 2012; Schlafmann, 2021), and the forward guidance puzzle in monetary policy (Airaudo, 2020).

Finally, our paper belongs to a fast growing literature introducing other *behavioral elements* into dynamic macro models, hence labeled *Behavioral Macroeconomics*.[6] This includes works proposing a departure from full-information rational expectations (FIRE), such as models with learning (Evans and Honkapohja, 2003), rational inattention (Mackowiak et al., 2021), dispersed information (Angeletos and Lian, 2016), level-k thinking (Angeletos et al., 2021), cognitive discounting (Gabaix, 2020), diagnostic expectations (Bianchi et al., 2022; L'Huillier et al., 2022) and several other forms of bounded-rationality in expectation formation (Garcia-Schmidt and Woodford, 2019; Woodford, 2013). Preference-based contributions to this literature include models with ambiguity aversion (Adam and Woodford, 2012; Ilut and Schneider, 2014), Epstein-Zin time non-additive util-

[5] The relevant literature is obviously much more extensive and not limited to the cited papers.

[6] The literature on Behavioral Macroeconomics is expanding extremely fast, and is therefore much wider than what cited below.

ity (Rudebusch and Swansson, 2012; Andreasen et al., 2018), and reference-dependent preferences (Barberis et al., 2001;Kőszegi and Rabin, 2009).[7]

The rest of the paper is organized as follows. Section 2 describes the model. Section 3 defines the equilibrium of the model, its efficient allocation (the social planner's problem). Section 4 defines the steady state, the baseline calibration, and provides empirical evidence in support of some key testable implications of the model. Section 5 derives the key reduced form equations which characterize the local equilibrium dynamics, as well as constitute the relevant constraints for the optimal policy problem. Section 6 states the welfare criterion for optimal monetary policy, and studies the impulse responses to key exogenous shocks under both discretion and commitment. Section 7 provides a qualitative and quantitative analysis of the welfare costs from business cycle fluctuations. Section 8 states our preliminary conclusions.

# 2 The Model

The backbone of our model economy is identical to the baseline New Keynesian model used for monetary policy analysis. It includes a continuum of identical infinitely-lived households who consume and save (demand side), a continuum of sticky price monopolistically competitive good producing firms (supply side), and a unified monetary/fiscal authority. The novel aspect of the model is the introduction of Gul-Pesendorfer-type *temptation with self-control preferences* on the households' side.

## 2.1 Households

Consider an economy populated by a continuum of identical infinitely-lived households. Borrowing is not permitted, but the representative household can save by investing in a risk-free bond issued by the government (a short-term Treasury bill). Gross returns on the latter, together with labor income and dividends from firm ownership, constitute his total resources that can be used for consumption expenditure, $P_t c_t$, as well as to finance new bond holdings, $B_t$. The relevant constraints for the

[7] See also Backus et al. (2004) for an extensive review of "exotic preferences" in macroeconomic models.

representative household, in real terms, are:

$$c_t + b_t = R_{t-1}\frac{b_t}{\pi_{t-1}} + w_t h_t + d_t - \tau_t, \tag{1}$$

$$b_t \geq 0, \tag{2}$$

where $b_t = \frac{B_t}{P_t}$, $\pi_t = \frac{P_t}{P_{t-1}}$, $w_t = \frac{W_t}{P_t}$, $d_t = \frac{D_t}{P_t}$ and $\tau_t$ denote, respectively, real bond holdings, gross inflation, the real wage, real dividends and the lump-sum taxes paid to the government.

The household has temptation with self-control preferences (henceforth, just GP-utility or temptation preferences), as formalized in Gul and Pesendorfer (2001, 2004) for general settings, and recently introduced in a New Keynesian framework by Airaudo (2020). GP preferences describe a household who, in every period, is tempted to use all his financial wealth for current consumption purposes - thus behavior like a *hand-to-mouth* consumer - but that, to resist such temptation, incurs a self-control cognitive cost (disutility). Loosely speaking, let $c^r$ denote the singleton set of optimal consumption chosen by a forward-looking *Ricardian* consumer, and $c^{htm}$ be the singleton set of consumption made by a myopic *non-Ricardian* consumer.[8] Letting $\succeq$ denote household's preferences over consumption choices (satisfying completeness, transitivity, continuity, and independence), temptation with self-control requires

$$\{c^r\} \succeq \left\{c^r, c^{htm}\right\} \succeq \left\{c^{htm}\right\} \tag{3}$$

The fact that $\{c^r\}$ is weakly preferred to the enlarged set $\left\{c^r, c^{htm}\right\}$ signifies that $c^{htm}$ is a tempting option, while having $\left\{c^r, c^{htm}\right\}$ preferred to $\left\{c^{htm}\right\}$ means that, eventually, the consumer exerts self-control and chooses $c^r$. From Gul and Pesendorfer (2004), the recursive representation for the household's intertemporal utility maximization problem is a Bellman equation:

$$\mathcal{U}_t = \max_{\{c_t, h_t, b_t\}} \left[u(c_t, h_t) + v(c_t, h_t) + \beta E_t \mathcal{U}_{t+1}\right] - \max_{\{\tilde{c}_t, \tilde{h}_t, \tilde{b}_t\}} v(\tilde{c}_t, \tilde{h}_t) \tag{4}$$

subject to (1) and (2), where $u$ and $v$ are both Von Neuman-Morgenstern utility functions. On the one hand, the term $u_t + \beta E_t \mathcal{U}_{t+1}$ represents the standard *commitment* utility: it captures the household's evaluation of the long-run best. On the other hand, $v_t$ captures the *temptation*

[8] See Axioms 1-8 in Gul and Pesendorfer (2004) for a formal definition of these preferences over a compact metric space in an infinite horizon consumption-saving decision problem.

utility, i.e. how the household values his urges. The household is sophisticated in the sense that he is cognizant of his current and future costs of self-control.[9] Letting $\tilde{c}_t$, $\tilde{h}_t$ and $\tilde{b}_t$ denote optimal levels of, respectively, consumption, hours worked and bond holdings under temptation, the term $\max v(\tilde{c}_t, \tilde{h}_t) - v(c_t, h_t)$ corresponds the temptation opportunity cost: this is the utility loss the household suffers when he exerts self-control by choosing the triple $(c_t,\ h_t,\ b_t)$ over the most tempting option $(\tilde{c}_t,\ \tilde{h}_t,\ \tilde{b}_t)$. We will refer to this as the *cost of self-control.*[10] The utility $\mathcal{U}_t$ is therefore the maximum of commitment utility net of costs of self-control. For the purpose of our analysis, we adopt the following functional forms:

$$u_t = \ln x_t, \qquad v_t = \xi \ln x_t, \qquad x_t \equiv c_t - \frac{h_t^{1+\chi}}{1+\chi} \tag{5}$$

This specification is similar to the one used in Airaudo (2020). First of all, it assumes a Greenwood-Hercovitz-Huffman-type (as in Greenwood et al., 1998, henceforth GHH) temporary utility, both for commitment $u$ and temptation $v$. By eliminating wealth effects in labor supply, a GHH specification will allow us to solve for the temptation allocation in closed form.[11] Second, by imposing a log specification on both $u$ and $v$ in (5), temptation and commitment will both feature a unitary degree of relative risk aversion. This differs from Airaudo (2020) who, for the case of linear costs of self-control, allows for higher risk aversion in temptation (with respect to commitment).[12] While a log-utility specification is without loss of generality for the purpose of our analysis, it significantly simplifies the second-order Taylor approximation to the household's welfare for the optimal monetary policy design. Third, and most importantly, the strength of temptation in the model is captured by the parameter $\xi$. For $\xi = 0$, the model reduces to a baseline New Keynesian

[9] Muraven et al. (2006) (see also references therein) and, more recently, Schilback (2019) provide experimental evidence on the existence of sophistication (foresight) in decision problems with persistent self-control. A naive household would instead neither recognize nor care about future self-control costs, as well as would not anticipate future preference reversals. This would lead to a game-theoretic set-up between current and future selves, as in models with hyperbolic discounting. See Ahn et al. (2020) for a model of naivete about temptation/self-control in a quasi-hyperbolic discounting framework.

[10] As in Gul and Pesendorfer (2001, 2004), the cost of self-control is *linear* in the opportunity cost of temptation. Noor and Takeota (2010) and Fudenberg and Levine (2006, 2011) consider the case of convex costs of self-control.

[11] GHH preferences find empirical support both at the *aggregate* macro (see Schmitt-Grohé and Uribe, 2012) and *individual* micro (see Cesarini et al., 2017) levels. These preferences are often used to generate government spending multipliers for output larger than unity. See Monacelli and Perotti (2008) and Bardóczy et al. (2021) for a recent thorough discussion.

[12] Airaudo (2020) introduces GP preferences into a New Keynesian model showing that higher risk aversion in temptation yields discounting in the linearized Euler equation, which then helps solve/tame the forward guidance puzzle of monetary policy. It is worth clarifying that a log-GHH utility yields a unitary relative risk aversion with respect to the consumption-labor bundle $x$, but not with respect to consumption alone (see Airaudo and Hajdini, 2021, for a discussion).

framework with GHH preferences.

The solution to the representative household optimal consumption-saving decision involves two stages. In the first stage, we identify the optimal tempting choice by solving a simple static optimization problem:

$$\max \xi \ln \left( \tilde{c}_t - \frac{\tilde{h}_t^{1+\chi}}{1+\chi} \right),$$

subject to the budget $\tilde{c}_t + \tilde{b}_t = R_{t-1}\frac{b_{t-1}}{\pi_t} + w_t \tilde{h}_t + d_t - \tau_t$ and the non-negativity constraint $\tilde{b}_t \geq 0$. Letting $\tilde{\lambda}_t$ and $\tilde{\lambda}_{b,t}$ denote the respective Lagrange multipliers, first order conditions with respect to $\tilde{c}_t$, $\tilde{h}_t$ and $\tilde{b}_t$ give:

$$\xi \tilde{x}_t^{-1} = \tilde{\lambda}_t, \qquad \xi \tilde{x}_t^{-1} \tilde{h}_t^{\chi} = \tilde{\lambda}_t w_t, \qquad \tilde{\lambda}_t = \tilde{\lambda}_{b,t}. \tag{6}$$

Inada conditions guarantee that $\tilde{x}_t > 0$, such that both multipliers are positive. This, together with the complementary slackness condition, $\tilde{\lambda}_{b,t}\tilde{b}_t = 0$, implies optimal zero bond holdings, $\tilde{b}_t = 0$. After simple manipulation of the equations in (6), the optimal choice under temptation is summarized by the following conditions:

$$\tilde{b}_t = 0, \qquad \tilde{h}_t = w_t^{\frac{1}{\chi}}, \tag{7}$$

$$\tilde{c}_t = \frac{b_{t-1}}{\pi_t} + w_t^{\frac{1+\chi}{\chi}} + d_t - \tau_t, \tag{8}$$

$$\tilde{x}_t = R_{t-1}\frac{b_{t-1}}{\pi_t} + \frac{\chi}{1+\chi} w_t^{\frac{1+\chi}{\chi}} + d_t - \tau_t, \tag{9}$$

i.e., no savings, a labor supply which, due to GHH preferences, solely depends on the real wage (with elasticity $\chi^{-1}$), and consumption of all available resources.

Letting $\tilde{x}(b_{t-1})$ denote the expression in (9), the representative household's Bellman equation (4) becomes

$$\mathcal{U}_t = \max_{\{c_t, h_t, b_t\}} (1+\xi)\ln x_t - \xi \ln \tilde{x}(b_{t-1}) + \beta E_t \mathcal{U}_{t+1}, \tag{10}$$

subject to (1)-(2), with respective multipliers $\lambda_t$ and $\lambda_{b,t}$. First order conditions and simple manipulation of terms yield the following relationships:

$$\lambda_t = (1+\xi)x_t^{-1}, \qquad h_t = w_t^{\frac{1}{\chi}}, \tag{11}$$

$$\lambda_t = \lambda_{b,t} + \beta R_t E_t\left(\frac{\lambda_{t+1}}{\pi_{t+1}}\Gamma_{t+1}\right), \qquad \lambda_{b,t} b_t = 0, \tag{12}$$

where

$$\Gamma_{t+j} \equiv 1 - \frac{\tilde{\lambda}_{t+j}}{\lambda_{t+j}}, \qquad \text{for } j \geq 1.^{13} \tag{13}$$

The expressions in (11) represent, respectively, the marginal utility of consumption and the household's labor supply. With respect to the baseline no-temptation case, the former is augmented by a factor which depends on self-control costs, while the GHH specification insulates labor supply from direct temptation effects. The expressions in (12) constitute the household's Euler equation, with the consumption-saving trade-off distorted by the temptation-driven factor $\Gamma_{t+1}$, as defined in (13) for $j = 1$.

## 2.2 Firms

The supply side of the economy is standard. Production is split into two sectors: retail and wholesale. The retail sector is perfectly competitive and produces a final consumption good $y_t$ out of a continuum of intermediate goods via the CRS technology $y_t = \left[\int_0^1 y_t(i)^{(\epsilon-1)/\epsilon} di\right]^{\epsilon/(\epsilon-1)}$, with $\epsilon > 1$ denoting the elasticity of substitution between any two varieties of intermediate goods. Prices in the retail sector are perfectly flexible. The optimal demand for the intermediate good $y_t(i)$ is given by $y_t(i) = \left(\frac{P_t(i)}{P_t}\right)^{-\epsilon} y_t$, where $P_t \equiv \left[\int_0^1 P_t(i)^{1-\epsilon}\, di\right]^{1/(1-\epsilon)}$ is the price of the final consumption good.

The wholesale sector is made of a continuum of firms indexed by $i$, for $i \in [0,1]$. They act under monopolistic competition and are subject to nominal rigidities in price setting. The $i$-th firm hires labor from a competitive labor market to produce the $i$-th variety of a continuum of differentiated intermediate goods which are sold to retailers. Wholesale firms operate a simple linear technology: $y_t(i) = z_t h_t(i)$, where $z_t$ denotes TFP with unconditional mean $z$. Letting $\hat{z}_t \equiv \ln(z_t/z)$, we assume $\hat{z}_t = \rho_z \hat{z}_{t-1} + \hat{\varepsilon}_{z,t}$, with $|\rho_z| < 1$ and $\hat{\varepsilon}_{z,t} \sim iid\mathcal{N}\left(0, \sigma_z^2\right)$.

We introduce nominal rigidities following Calvo's staggered price setting: each firm in the wholesale sector optimally revises its price with probability $1-\theta$ in any given period $t$. Real marginal

[13] Throughout the paper, we adopt the notational convention that $\Gamma_t = 1$.

costs are equal across firms and given by $mc_t = (1-s_t)\frac{w_t}{z_t}$, with $s_t \in [0,1)$ denoting a time-varying subsidy rate to labor, common across all firms.[14] The $i$-th firm chooses the optimal price $P_t^*(i)$ to maximize $E_t \sum_{j=0}^{\infty} \theta^j \mathcal{F}_{t,t+j} y_{t+j}(i) \left(P_t^*(i) - P_{t+j} mc_{t+j}\right)$, subject to the demand constraint $y_{t+j}(i) = \left(\frac{P_t^*(i)}{P_{t+j}}\right)^{-\epsilon} y_{t+j}$. The term $\mathcal{F}_{t,t+j}$ denotes the household's stochastic discount factor (SDF) for nominal payoffs between period $t$ and a generic $t+j$, for $j \geq 0$.[15] In particular, with the SDF for riskless nominal payoffs between period $t$ and $t+1$ defined as $R_t E_t \mathcal{F}_{t,t+1} = 1$, from the Euler equation in (12) with $\lambda_{b,t} = 0$ (as we will focus on an equilibrium with positive bond supply), we have that $\mathcal{F}_{t,t+1} = \beta \frac{P_t}{P_{t+1}} \frac{\lambda_{t+1}}{\lambda_t} \Gamma_{t+1}$,or, more generally

$$\mathcal{F}_{t,t+j} = \beta^j \frac{P_t}{P_{t+j}} \frac{\lambda_{t+j}}{\lambda_t} \prod_{k=1}^{j} \Gamma_{t+k}, \quad \text{for} \quad j \geq 0.$$

It is straightforward to solve for the optimal price $P_t^*$ (where the index $i$ has been dropped since all price-setting firms face same economic conditions and therefore choose the same price) relative to $P_t$:

$$\frac{P_t^*}{P_t} = \mu \frac{E_t \sum_{j=0}^{\infty} \theta^j Q_{t,t+j} y_{t+j} mc_{t+j} \pi_{t,t+j}^{\epsilon}}{E_t \sum_{j=0}^{\infty} \theta^j Q_{t,t+j} y_{t+j} \pi_{t,t+j}^{\epsilon-1}} \tag{14}$$

where $\mu \equiv \frac{\epsilon}{\epsilon-1}$ is the (steady-state) gross price mark-up, and $Q_{t,t+j} \equiv \mathcal{F}_{t,t+j}\pi_{t,t+j}$, with $\pi_{t,t+j} \equiv \frac{P_{t+j}}{P_t}$. Firms' dividends are distributed to households in a lump-sum fashion, with $d_t(i) \equiv \frac{D_t(i)}{P_t} = \frac{P_t(i)}{P_t} y_t(i) - \frac{W_t}{P_t} h_t(i)$ given by the following expression:

$$d_t(i) = \left(\frac{P_t(i)}{P_t}\right)^{1-\epsilon} y_t - mc_t \left(\frac{P_t(i)}{P_t}\right)^{-\epsilon} y_t \tag{15}$$

### 2.3 Government

A government/central bank conducts both fiscal and monetary policy. The government's budget constraint is given by

$$\tau_t + b_t = R_{t-1}\frac{b_{t-1}}{\pi_t} + g_t + s_t w_t h_t, \tag{16}$$

that is, the proceeds from tax revenues $\tau_t$ and newly issued treasuries $b_t$ are used to pay-off outstanding public debt (capital plus interests), public spending $g_t$ and the labor subsidy to firms.

[14] Time variation in the subsidy rate will be the source of cost-push supply shocks in the Phillips curve.

[15] As standard in the literature, we assume firms discount future profits using the household's stochastic discount factor, which, by default, incorporates his dynamic self-control problem.

For simplicity, we assume $\tau_t = \tau_t^H + \tau_t^F$, where $\tau_t^F$ is entirely used to finance the labor subsidy: $\tau_t^F = s_t w_t h_t$. Assuming public spending $g_t$ is exogenous - specifically, $\hat{g}_t \equiv \ln(g_t/g)$ follows an AR(1) process, $\hat{g}_t = \rho_g \hat{g}_{t-1} + \hat{\varepsilon}_{g,t}$, with $\left|\rho_g\right| < 1$ and $\hat{\varepsilon}_{g,t} \sim iid\mathcal{N}\left(0, \sigma_g^2\right)$ - we let $f_t \equiv \tau_t^H - g_t$ denote the fiscal surplus, and introduce the following debt-feedback fiscal rule:

$$f_t = \left(\frac{b_t}{b}\right)^{\phi_b} v_{f,t} \tag{17}$$

where $b$ is steady state debt, and $\phi_b > 0$.[16] The factor $v_{f,t}$ is a fiscal shock, with $\hat{v}_{f,t} \equiv \ln(v_{f,t}/v_f)$ following the AR(1) process $\hat{v}_{f,t} = \rho_f \hat{v}_{f,t-1} + \hat{\varepsilon}_{f,t}$, for $\left|\rho_f\right| < 1$ and $\hat{\varepsilon}_{f,t} \sim iid\mathcal{N}\left(0, \sigma_f^2\right)$.

The central bank will implement the optimal monetary policy plan by maximizing the representative household's lifetime utility. A detailed description is given in Section 6.

# 3 Equilibrium

In equilibrium, households and firms optimize, given aggregate quantities and prices, and all markets clear. Output equals private consumption plus government spending $-$ i.e. $y_t = c_t + g_t$ $-$ and labor supply equals labor demand $-$ i.e. $h_t = \int_0^1 h_t(i)\, di = y_t \Delta_t / z_t$ with $\Delta_t \equiv \int_0^1 \left(\frac{P_t(i)}{P_t}\right)^{-\epsilon} di$ denoting price dispersion due to Calvo pricing. The law of motion for $P_t$ is implicitly given by $P_t^{1-\epsilon} = (1-\theta)\left(P_t^*\right)^{1-\epsilon} + \theta P_{t-1}^{1-\epsilon}$. Aggregate real dividends, $d_t \equiv \int_0^1 d_t(i)\, di$, are obtained by integrating (15) across all firms: $d_t = y_t\left(1 - mc_t \Delta_t\right)$, where $mc_t = (1 - s_t)(w_t/z_t) = (1 - s_t)\, z_t^{-(1+\chi)} \left(y_t \Delta_t\right)^{\chi}$.

The consumption-labor composite $x_t$ is

$$x_t = y_t - g_t - \frac{\left(\frac{y_t \Delta_t}{z_t}\right)^{1+\chi}}{1+\chi}. \tag{18}$$

After combining the expression for $\tilde{x}_t$ in (9) with the household's budget (1), simple algebra gives the temptation composite,

$$\tilde{x}_t = x_t + b_t. \tag{19}$$

Focusing on an equilibrium with positive supply of government bonds in very period - namely,

[16]It is common practice in the literature to assume the fiscal rule is function of *lagged* outstanding debt. Our specification with respect to *current* (newly issued) debt will bring some analytical advantage when pursuing a second order Taylor approximation to the government's intertemporal budget constraint. As discussed later, this step will be necessary for an accurate optimal monetary policy analysis via linear-quadratic techniques.

$b_t > 0$ for $t \geq 0$ - by the complementary slackness condition in (12), we have that $\lambda_{b,t} = 0$, which, in turn, leads to the generalized Euler equation:

$$\lambda_t = \beta R_t E_t \frac{\lambda_{t+1}}{\pi_{t+1}} \Gamma_{t+1}. \tag{20}$$

The following lemma establishes that the Euler equation wedge $\Gamma_{t+1}$ is always positive, which, in turn, guarantees that the right hand side of (20) is always positive as well.

**Lemma 1** *Recall the definition of* $\Gamma_{t+j}$ *in (13). If* $b_{t+j} > 0$, *then* $\Gamma_{t+j} \in (0, 1)$, *for any* $j \geq 1$.

**Proof.** See Appendix A.4.1. ∎

We can further elaborate on (20) to show that temptation preferences introduce real wealth effects from bond accumulation in the Euler equation, thus making debt dynamics no longer neutral for business cycle fluctuations. Substituting the expressions for $\lambda_t$ in (11) and $\Gamma_{t+1}$ in (13) into (20), we obtain:

$$(1 + \xi) x_t^{-1} = \beta R_t E_t \left[ \frac{x_{t+1}^{-1} + \xi \left( x_{t+1}^{-1} - \tilde{x}_{t+1}^{-1} \right)}{\pi_{t+1}} \right]. \tag{21}$$

Wealth effects are embedded in the temptation choice term $\tilde{x}_{t+1}$. Setting $\xi = 0$, we retrieve the baseline model without temptation: in this case, at the optimum, neither the marginal utility of consuming today, $x_t^{-1}$, nor the expected marginal utility of saving for future consumption, $\beta R_t E_t \left( \pi_{t+1} x_{t+1} \right)^{-1}$, are distorted by government bonds. The household does not perceive them as net wealth, and Ricardian equivalence holds.

With $\xi > 0$, temptation generates an increase in the marginal benefits of both consumption and savings in (21). The left hand side of (21) is in fact accrued by a factor $(1 + \xi)$ as the household evaluates current consumption by both his commitment utility $u$ and temptation utility $v = \xi u$. The right hand side is instead augmented by the term $\xi \left( x_{t+1}^{-1} - \tilde{x}_{t+1}^{-1} \right)$. This is clearly positive since both utilities are strictly concave and $\tilde{x}_{t+1} = x_{t+1} + b_{t+1} > x_{t+1}$. For a deeper economic intuition, notice that the term $\xi \left( \tilde{x}_{t+1}^{-1} - x_{t+1}^{-1} \right)$ corresponds to the *marginal* disutility cost of self-control. By increasing $\tilde{x}_{t+1}$ (hence, lowering $\tilde{x}_{t+1}^{-1}$), an increase in $b_{t+1}$ lowers the future marginal costs of self-control, giving a forward-looking household an additional incentive to save. Hence, an expected increase in future government debt induces a *negative wealth effect* on current consumption.

## 4 Steady State, Calibration and Empirical Evidence

We focus on a zero inflation steady state ($\pi = 1$), and, without loss of generality, set $z = v_f = 1$, $s_t = s$, and $g = g_y y$ for $g_y \in (0, 1)$.[17] From the optimal pricing of firms, we have $mc = \mu^{-1}$, which, combined with the definition of marginal costs, gives the steady state real wage: $w = [(1-s)\mu]^{-1}$. Plugging the latter into the labor supply equation yields $h = y = [(1-s)\mu]^{-1/\chi}$, such that $c = c_y y$, for $c_y \equiv 1 - g_y$. The definition of $x_t$ in (5) and simple algebra give

$$x = \omega y, \qquad \text{for} \qquad \omega \equiv \frac{(1+\chi)\, c_y (1-s)\mu - 1}{(1+\chi)(1-s)\mu} \tag{22}$$

where $\omega \in (0, 1)$ provided the steady state share of consumption in GDP, $c_y$, is sufficiently large.[18]

From the Euler equation (20), we find the steady state nominal (and real) interest rate $R = (\beta\Gamma)^{-1}$, where, making use of $\tilde{x} = x + b$ and the relationship in (22), with $b_y \equiv \frac{b}{y}$ denoting the debt-to-GDP ratio, the temptation-driven wedge $\Gamma$ is

$$\Gamma = 1 - \frac{\xi\omega}{(1+\xi)(\omega + b_y)} \in (0, 1). \tag{23}$$

Finally, from the government's budget (16), for given $b_y$ and $R$, we find the steady state surplus-to-GDP ratio, $f_y \equiv b_y (R - 1)$.

**Proposition 1** *The steady state interest rate $R$*

1. *is strictly increasing with respect to temptation $\xi$;*

2. *conditional on positive temptation, is strictly decreasing with respect to the debt-to-GDP $b_y$.*

**Proof.** See Appendix A.4.3. ■

Figure 1 displays the steady state real interest rate as function of temptation $\xi$ and the debt-to-GDP ratio $b_y$ under the baseline calibration displayed in Table 1. The latter has the inverse Frisch elasticity of labor supply $\chi$ equal to unity, an intratemporal elasticity of substitution across goods varieties $\epsilon$ equal to 8 (hence a 15% net price mark-up), and a Calvo probability of price stickiness

[17] For each variable, we drop the time subscript $t$ to denote its steady state value.

[18] This is guaranteed for any realistic parameterization of the model. For instance, for the case of an efficient subsidy - such that $(1-s)\mu = 1$ (see below) - this requires $c_y$ to be larger than $(1+\chi)^{-1}$, where the latter is 0.5 or smaller for $\chi \geq 1$. A positive $\omega$ is even more likely for the case of no subsidy.

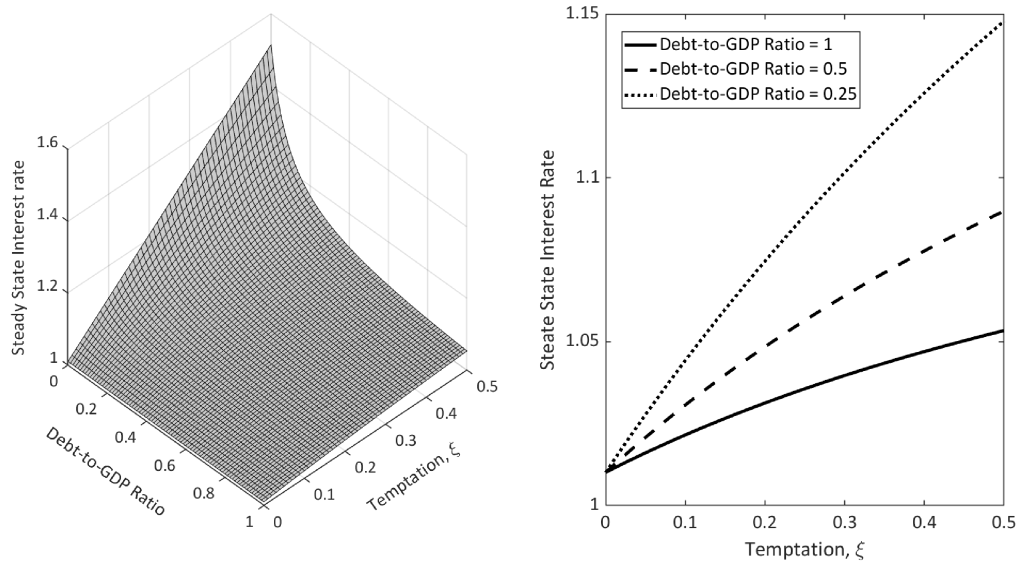


Figure 1: Steady state real interest rate.

$\theta$ equal to 0.715 (hence a price duration of approximately 3.5 quarters).[19] For what concerns fiscal quantities, we set the public spending to GDP ratio $g_y$ equal to 0.18, and assume (for the moment) no labor subsidy to firms, $s = 0$.[20] In the figure, we let the quarterly ratio $b_y$ range from 0 to 4 - hence, from 0 to 1 yearly (in the analysis to follow, we will set $b_y = 2$, that is a 50% debt-to-GDP ratio an annual frequency) - and consider $\xi \in [0, 0.5]$. This range for temptation includes the Euler equation based estimates available in the literature, going from the low-end $\xi \approx 0.05$ in Bucciol (2012), to the intermediate value $\xi \approx 0.23$ in Huang et al. (2015), to the high-end estimate $\xi \approx 0.39$ in Kovacs et al. (2021).[21] As shown in the left panel, for any $b_y \geq 0$, the real rate is strictly increasing in $\xi$, and more so when $b_y$ approaches zero. Conditional on positive temptation, the real rate is strictly decreasing in $b_y$, as also evident from the right panel.

[19]A unitary value for $\chi$ is intermediate between the *macro based* (below unity, hence higher labor elasticity) and the *micro based* (above unity, hence lower labor elasticity) evidence (see Kean and Rogerson, 2012, for a discussion).

[20]In the next section, we will set $s$ equal to its efficient level to guarantee an undistorted steady state.

[21]The Euler equation GMM estimation by Huang et al. (2015) yields $\xi = 0.1$ if using NIPA data, and $\xi = 0.24$ using CEX data. Restricting to the latter, Kovacs et al. (2021) find $\xi = 0.28$ using GMM and $\xi = 0.39$ using the method of simulated moments on the full structural model. Attanasio et al. (2022) show that a life-cycle model with GP-temptation preferences and $\xi = 0.28$ matches well the life-cycle profile of aggregate consumption and liquid asset accumulation, as well as it generates the age-dependent hand-to-mouth behavior seen in U.S. data. Airaudo et al. (2022) find similar values from a fully-fledged Bayesian DSGE model estimation on post-WWII U.S. data.

| Table 1 Calibration | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Standard Parameters | | | | | | | | | Temptation |
| $\chi$ | $\epsilon$ | $\theta$ | $b_y$ | $g_y$ | $\phi_b$ | $\rho_e$ | $\rho_i$ $(i{=}z,g,f)$ | $\sigma_i$ $(i{=}z,g,f,e)$ | $\xi$ |
| 1 | 8 | 0.715 | 2 | 0.18 | 1.5 | 0.5 | 0.9 | 0.01 | $[0, 0.3]$ |

## 4.1 Testable Implications

Statements 1. and 2. in Proposition 1 identify two testable implications of our theoretical model. As mentioned, there exist some attempts in the literature to test for the presence of GP preferences in consumption choices. However, to the best of our knowledge, direct empirical evidence of time and/or cross-section variation for the temptation parameter $\xi$ is not available. To circumvent this problem, as proxies for $\xi$, we use estimates of quasi-hyperbolic discounting by Rieger et al. (2016, 2021) from the cross-country International Test of Risk Attitudes (INTRA) survey data at the University of Zurich. Our approach is supported by the fact that, under some restrictions, consumption choice models featuring hyperbolic discounting are in reduced form isomorphic to models where consumers display temptation with self-control preferences. In particular, Lu (2016) shows that the decision maker's preferences in a dual-self model with linear costs of self-control à-la Fudenberg-Levine (2006) - which is isomorphic to the GP-preferences set-up - could be mapped into those of a hyperbolic discounter with discount factor $\beta$ (in his $\beta\delta$-preferences) equal $(1+\xi)^{-1}$.[22] We restrict our attention to a subsample of 28 countries belonging to neighboring economic areas and/or displaying similar levels of development.[23]

**Testable Implication 1: Higher Temptation leads to Higher Interest Rates** A scatter plot of the annual average real interest rate (computed on quarterly data for the period 1990:1-2007:4) and the estimated $\beta$ for hyperbolic discounting (HD) is presented in Figure 2.[24] From the

[22] Lu (2016) derives the result assuming the Fudenberg-Levine's long-run self is not concerned about *future* costs of self-control. Fudenberg and Levine (2006) show that their dual-self framework with linear self-control costs satisfies all axioms posed in Gul and Pesendorfer (2001, 2004). Dekel and Lipman (2012) provide an equivalence result between models with Strotz-type dynamically inconsistent preferences and those with GP-type temptation-with-self-control preferences.

[23] The full sample includes 53 countries. We have excluded countries from Latin America (featuring highly volatile interest rates due to recurrent episodes of macroeconomic instability and default risk), Africa (low income economies) and Asia (Japan because of its prolonged history of near-zero interest rates and deflation, Thailand and Korea because of currency crises in the late '90s, and China being a non-free market economy).

[24] Following Uribe and Yue (2006), the real interest rate is computed as the difference between the nominal return on a 3-month safe government bond and expected inflation, the latter measured as the average inflation over the past 4 quarters.

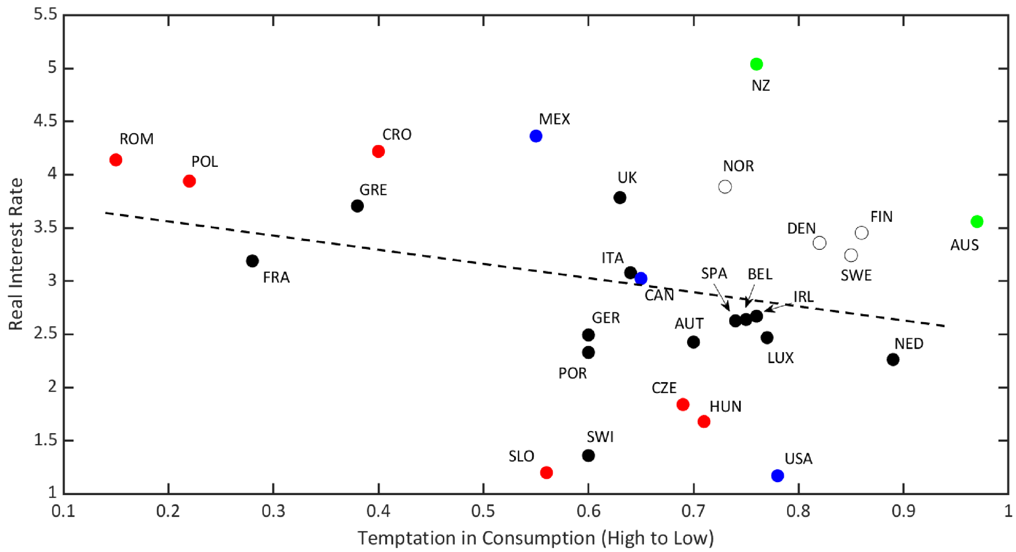


Figure 2: Real Interest Rates and Temptation

previous paragraph, a lower (respectively, higher) value of $\beta$ signifies higher (respectively, lower) present-bias in consumption choices, hence higher (respectively, lower) temptation. Different colors are assiged to different geo-economic areas: Western Europe (WE, black dots), Eastern Europe (EE, red), Northern Europe (NE, white), North America (NA, blue), and Australia-New Zealand (ANZ, green). Simple visual inspection highlights the possibility of a negative relationship between the two quantities, with the dashed line giving fitted values from a simple OLS regression (see estimates in the second column of Table 1).

For a more accurate quantitative assessment, we estimate the following linear regression by OLS:

$$RR_i = a_0 + a_1 HD_i + \sum_{j=1}^{4} b_j DUM_{j,i} + \sum_{j=1}^{4} c_j HD_i * DUM_{j,i} + f' CONT_i + \varepsilon_i \tag{24}$$

where $RR$ is the net real rate, the subscript $i$ indexed the country, $HD$ is hyperbolic discounting, $DUM_{j,i}$ is a dummy variable equal to 1 if country $i$ belongs to country group $j$ (where $j = 1$ is for EE, 2 for NE, 3 for ANZ, and 4 for NA) and zero otherwise, and $CONT_i$ is a vector of controls, including TFP growth, life expectancy and the old-age dependency ratio.[25] Results for

[25] Quarterly time series for these variables come from the FRED data set at the Federal Reserve Bank of St. Louis. See Ferrero et al. (2016) and Busetti and Caivano (2019) for an empirical and theoretical analysis of the relationship between real interest rate and demographic factors.

three alternative specifications are reported in Table 1. A baseline specification regressing the real interest rate on HD yields a poor fit and does not allow to reject the null hypothesis of no statistical relationship ($a_1 = 0$) at usual confidence levels.[26] Adding intercept dummies (I.D.) to capture geo-economic level effects produces a much better outcome: HD has a negative and statistically significant impact on interest rates - or, equivalently, temptation has a positive impact on interest rate - with positive dummies for NE and ANZ capturing the occurrence of higher rates in those areas. According to standard model evaluation criteria, allowing temptation to exert a different impact across different geo-economic areas yields an even better fit, with more significant upward pressure on rates in NA and EE (more negative slopes). Overall, these empirical results appear consistent with Statement 1 in Proposition 1.

**Table 2** Regression (24): Interest Rate and Temptation

| | 1) Baseline | 2) With I.D. | 3) With I.D. and S.D. |
|---|---|---|---|
| $a_0$ | $3.83^{**}$ (0.63, 5.47e$^{-06}$) | $4.65^{**}$ (0.50, 4.55e$^{-09}$) | $4.68^{**}$ (0.44, 1.04$e^{-9}$) |
| $a_1$ | $-1.33$ (0.94, 0.29) | $-3.21^{**}$ (0.81, 0.0014) | $-3.04^{**}$ (0.69, 0.0006) |
| $b_2$ (NE) | | $1.44^{**}$ (0.45, 0.007) | $1.28^{**}$ (0.38, 0.0066) |
| $b_3$ (ANZ) | | $2.42^{**}$ (0.58, 0.0009) | $2.25^{**}$ (0.5, 0.0004) |
| $b_4$ (NA) | | | $7.34^{**}$ (2.52, 0.019) |
| $c_1$ (EE) | | | $-1.21^{**}$ (0.57, 0.095) |
| $c_4$ (NA) | | | $-10.86^{**}$ (3.79, 0.021) |
| | $R^2_{adj} = 0.04$, F-test = 1.006 (0.236) | $R^2_{adj} = 0.46$, F-test = 6.56 (0.021) | $R^2_{adj} = 0.63$, F-test = 7.39 (0.018) |
| | AIC=0.1, SBIC=0.24 | AIC=-0.42,SBIC=-0.18 | AIC=-0.71,SBIC=-0.33 |

Notes: in parenthesis the S.E. and the 2-tailed p-value. I.D. = intercept dummy; S.D. = slope dummy

AIC = Akaike Information Creterion, SBIC = Schwarz Bayesian Information Criterion

[26]Under all specifications, none of the control variables' coefficients appears significantly different from zero. This is likely due to the limited cross-country variation as we restrict the analysis to a sub-sample of similar economies.

**Testable Implication 2: Higher Debt-to-GDP leads to Lower Interest Rates** We test the hypothesis by running the following simple regression:

$$RR_i = a_0 + a_1 DY_i + \varepsilon_i \tag{25}$$

with $DY$ denoting the (annualized) quarterly average of debt-to-GDP in country $i$. Estimating (25) over the full sample used in the previous exercise yields a poor fit and a coefficient $a_1$ not significantly different from zero. With the intent of removing economies whose levels of indebtedness and/or interest rates appear *weak* outliers, we restrict the analysis to a subset of *core* countries, whose average debt-to-GDP ratio and average real rate (for the 1990:1-2007:4 period) are within ±1.5 standard deviations from the respective full sample average.[27] The identified core economies are the black dots in the left panel of Figure 3. The red line displays the fitted values from the OLS regression (25), based on the estimates in Table 3.

**Table 3** Regression (25): Interest Rates and Debt-to-GDP

| $a_0$ | $a_1$ | |
|---|---|---|
| $4.22^{**}$ $(0.43,\ 1.72e^{-08})$ | $-0.026$ $(0.009,\ 0.029)$ | $R^2_{adj} = 0.255$, F-test$= 3.75^{**}$ $(0.047)$ |
| | | AIC $= -0.536$, SBIC $= -0.387$ |

Notes: in parenthesis the S.E. and p-values.

AIC = Akaike Information Criterion, SBIC = Schwarz Bayesian Information Criterion

Although quantitatively small, the impact of indebtedness on the real interest rate is negative and statistically significant. This can also be inferred by inspecting the relationship in four smaller groups characterized by different degrees of temptation, as in the scatter plots on the right of Figure 3: low (HD between 0.8 and 1), medium (HD between 0.7 and 0.8), high (HD between 0.6 and 0.7), and extreme (HD below 0.6). Except for the high temptation case, the remaining plots clearly suggest a negative relationship between debt and interest rates.

[27] The choice of "±1.5 standard deviation from mean" threshold is indeed arbitrary. Under assumption of normality, this would include, roughly, 80% of the sample. For larger thresholds, we do not find a statistical significant estimate for $a_1$.

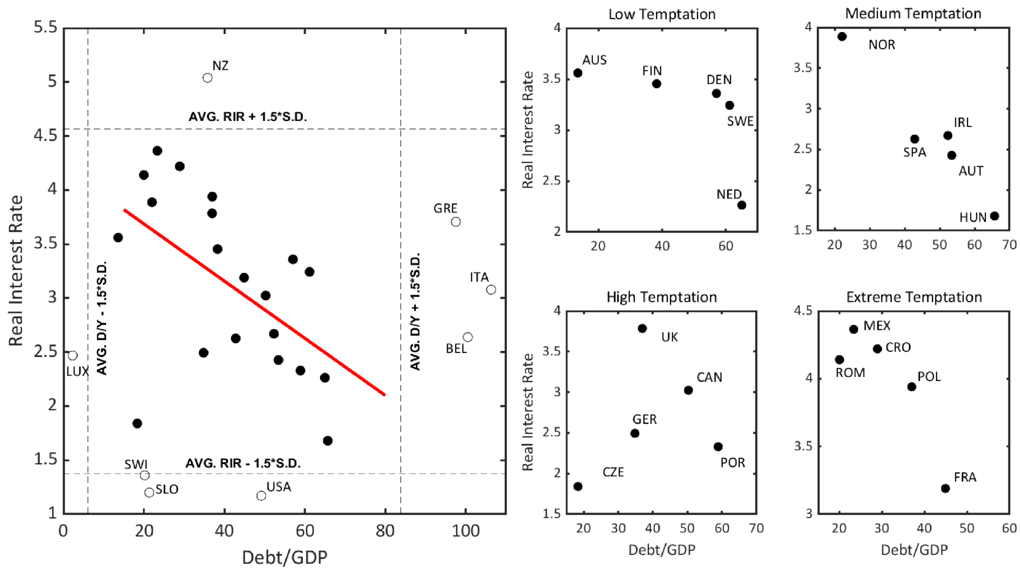


Figure 3: Real interest rates and debt-to-GDP.

# 5 Equilibrium Dynamics

To guarantee steady state efficiency by eliminating the monopolistic distortion, we introduce Assumption 1, which also normalizes both steady state hours and output to unity, $h = y = 1$.[28]

**Assumption 1:** $s = s^* = \frac{\mu-1}{\mu}$.

We approximate (to first order) all equilibrium conditions around the efficient steady state, letting "hat" on top of a variable denote its log-deviations from the respective steady state value. From the optimal price setting problem of firms, we obtain an expectational Phillips curve, $\hat{\pi}_t = \tilde{\beta} E_t \hat{\pi}_{t+1} + \kappa \hat{mc}_t$, where, as usual, the coefficient $\kappa = (1 - \tilde{\beta}\theta)(1-\theta)/\theta$ depends negatively on the extent of nominal rigidities. Making use of the expressions for real marginal costs $\hat{mc}_t = \hat{w}_t - \hat{z}_t - (\mu - 1)\,\hat{s}_t$, labor supply $\hat{w}_t = \chi\left(\hat{y}_t - \hat{z}_t\right)$, efficient output $\hat{y}_t^* = \frac{1+\chi}{\chi}\hat{z}_t$, as well as Assumption 1, simple algebra gives a standard relationship between inflation, its one-step-ahead expectation and the output gap:

$$\begin{aligned} \hat{\pi}_t &= \tilde{\beta} E_t \hat{\pi}_{t+1} + \kappa_y (\hat{y}_t - \hat{y}_t^*) + \hat{e}_t \qquad (26) \\ \kappa_y &\equiv \kappa \chi \end{aligned}$$

[28] With an undistorted/efficient steady state, we will be able to use standard linear-quadratic techniques for the evaluation of optimal policies without requiring a second order approximation to the Phillips curve. See Beningno and Woodford (2005) for a detailed discussion.

where the cost push shock $\hat{e}_t \equiv -\kappa(\mu-1)\hat{s}_t$ follows a standard AR(1) process, $\hat{e}_t = \rho_e \hat{e}_{t-1} + \hat{\varepsilon}_{e,t}$, with $|\rho_e| < 1$ and $\hat{\varepsilon}_{e,t} \sim iid\mathcal{N}\left(0, \sigma_e^2\right)$.

The government's budget constraint (16) combined with the fiscal rule (17) gives the law of motion of public debt:

$$\hat{b}_t = \rho_b(\hat{b}_{t-1} + \hat{R}_{t-1} - \hat{\pi}_t) - \rho_v \hat{v}_{f,t} \tag{27}$$

$$\rho_b \equiv \frac{R}{1+(R-1)\phi_b}, \qquad \rho_v \equiv \frac{(R-1)}{1+(R-1)\phi_b} \tag{28}$$

To guarantee that $\rho_b \in (0,1)$ - hence, stationary debt dynamics - we introduce the following assumption.

**Assumption 2:** $\phi_b > 1$.[29]

Moving to the demand side, from the Euler equation (20) and the marginal utility of consumption in (11), we obtain:

$$\hat{x}_t = E_t\hat{x}_{t+1} - \left(\hat{R}_t - E_t\hat{\pi}_{t+1}\right) - E_t\hat{\Gamma}_{t+1}, \tag{29}$$

Letting $\vartheta \equiv \frac{b_y}{\omega + b_y} \in (0,1)$, the definition $\Gamma_{t+1}$ in (13) yields

$$\hat{\Gamma}_{t+1} = -\varkappa\vartheta[\hat{x}_{t+1} - \hat{b}_{t+1}], \qquad \varkappa \equiv \frac{\xi(1-\vartheta)}{\Gamma(1+\xi)} > 0 \tag{30}$$

From the definition of $x_t$ in (5), we have instead

$$\begin{aligned} \hat{x}_t &= \frac{(1-s)\mu - 1}{\omega(1-s)\mu}\hat{y}_t - \frac{g_y}{\omega}\hat{g}_t + \frac{\hat{z}_t}{\omega(1-s)\mu} \\ &= \omega^{-1}(\hat{z}_t - g_y\hat{g}_t) \end{aligned} \tag{31}$$

where the second equality is a consequence of Assumption 1. Namely, around the efficient steady state, the consumption-labor composite $\hat{x}_t$ is completely exogenous. Letting $\gamma \equiv \varkappa\vartheta$, we plug (30)-(31) into (29), and, by simple manipulation of terms, derive the aggregate Euler equation:

$$\hat{R}_t - E_t\hat{\pi}_{t+1} = -\gamma E_t\hat{b}_{t+1} + \Psi_z\hat{z}_t - \Psi_g\hat{g}_t, \tag{32}$$

where $\Psi_z \equiv \omega^{-1}[\rho_z(1+\gamma)-1]$ and $\Psi_g \equiv \omega^{-1}g_y[\rho_g(1+\gamma)-1]$. Iterating the law of motion of debt (27) one period forward and taking expectations, we can substitute the term $E_t\hat{b}_{t+1}$ in (32), and,

[29] In the numerical analysis, we set $\phi_b = 1.5$ (see Table 1).

by simple algebra, reach the following expression:

$$\begin{aligned}\hat{R}_t &= E_t\hat{\pi}_{t+1} + \gamma_b\hat{b}_t + \gamma_z\hat{z}_t - \gamma_g\hat{g}_t + \gamma_v\hat{v}_{f,t} \\ \gamma_b &\equiv -\frac{\rho_b\gamma}{1+\rho_b\gamma}, \quad \gamma_i \equiv \frac{\Psi_i}{1+\rho_b\gamma} \text{ for } i = z, g, \quad \gamma_v \equiv \frac{\gamma\rho_v\rho_f}{1+\rho_b\gamma}\end{aligned} \tag{33}$$

With respect to the textbook New Keynesian framework, the aggregate Euler equation (33) does not display any dependence on current and expected output. This feature is the result of combining GHH preferences with steady state efficiency, and has therefore nothing to do with temptation. The Euler equation takes the form of a generalized Fisher equation whereby the *ex ante* real interest $\hat{R}_t - E_t\hat{\pi}_{t+1}$ is a function of exogenous fundamentals (TFP, government spending and fiscal shocks) but also of public debt, where the latter enters as a consequence of the temptation-driven wealth effects.

# 6 Optimal Monetary Policy

This section analyzes the consequences of temptation preferences for the design of optimal monetary policy, both under discretion and commitment. We obtain the central bank's objective by taking a second order approximation to the representative household's lifetime welfare, and expressing its arguments as squared deviations from welfare-relevant efficient targets. We start by defining the efficient equilibrium allocation, as the solution to the associated *static* social planner's utility maximization problem.

**Proposition 2** *The efficient equilibrium is characterized by the following relationships:* $h_t^* \equiv z_t^{\frac{1}{\chi}}, \quad y_t^* \equiv z_t^{\frac{1+\chi}{\chi}}, \quad c_t^* \equiv y_t^* - g_t.$

**Proof.** See Appendix A.4.2. ■

The recursive representation given in (4) can be reformulated as follows:[30]

$$\mathcal{U}_0 \equiv E_0 \sum_{t=0}^{\infty} \beta^t U_t, \qquad \text{for} \qquad U_t \equiv (1+\xi)\ln x_t - \xi \ln \tilde{x}_t \tag{34}$$

[30] See Drugeon and Wigniolle (2017) for a similar infinite summation representation of recursive GP preferences.

As shown in Appendix A.1.1, under the assumption of steady state efficiency, a second order approximation to $U_t$ yields

$$U_t - U \approx -U_y \left[\frac{\chi}{2}\left(\hat{y}_t - \hat{y}_t^*\right)^2 + \hat{\Delta}_t\right] - U_b \left[\hat{b}_t + (1-\vartheta)\left(\frac{\hat{b}_t^2}{2} - \hat{b}_t \hat{x}_t\right)\right] \tag{35}$$

$$U_y \equiv \frac{1+\xi\vartheta}{\omega}, \qquad U_b \equiv \xi\vartheta$$

The key element of differentiation with respect to the loss generally obtained for the baseline New Keynesian model is in the second bracketed term in (35). While households' welfare continues to depend (negatively) on fluctuations of the (log) output gap $\hat{y}_t - \hat{y}_t^*$ and (log) price dispersion $\hat{\Delta}_t$ around their efficient level (zero in both cases), temptation makes public debt dynamics no longer neutral for welfare evaluations, and hence for optimal policy. Public debt fluctuations distort households' intertemporal consumption decisions through the wealth effects related to the self-control costs to resist temptation. For $\xi = 0$ (no temptation), we have $U_b = 0$ and the second term drops out, leaving the output gap and price dispersion as the only welfare relevant policy objectives.

**Proposition 3** *The maximization of the representative household's welfare function $\mathcal{U}_0$ in (34) is equivalent to the minimization of the following intertemporal loss:*

$$\mathcal{L}_0 \equiv \frac{1}{2} E_0 \sum_{t=0}^{\infty} \beta^t \left[\alpha_y \left(\hat{y}_t - \hat{y}_t^*\right)^2 + \alpha_\pi \hat{\pi}_t^2 + \alpha_b (\hat{b}_t - \hat{b}_t^*)^2\right] \tag{36}$$

*where the welfare-relevant debt target $\hat{b}_t^*$ is a linear combination of shocks,*

$$\hat{b}_t^* \equiv M_z \hat{z}_t - M_g \hat{g}_t + M_v \hat{v}_{ft}, \tag{37}$$

$\alpha_y \equiv \chi > 0$, *while the welfare weights $\alpha_\pi > 0$ and $\alpha_b < 0$ are function of structural parameters of the model.*[31]

**Proof.** See Appendix A.1. ■

For what concerns the welfare weights entering (36), while $\alpha_y \equiv \chi$ is clearly positive, both $\alpha_\pi$ and $\alpha_b$ are convoluted expressions of standard and temptation-related parameters. Figure 4 displays

[31] See Appendix A.1 for analytical expressions for both the welfare weights $\alpha_\pi$ and $\alpha_b$, as well as for the coefficients $M_i$, for $i = z, g, v$.

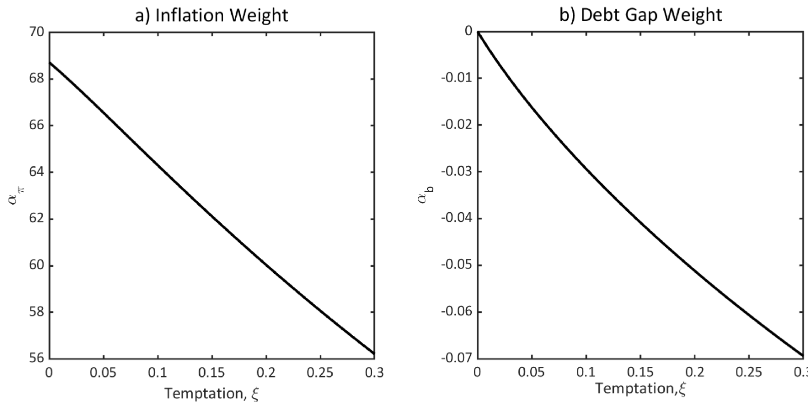


Figure 4: Welfare Weights

a numerical evaluations of both with respect to $\xi$. With the inverse Frisch elasticity parameter $\chi$ typically ranging between 0 and 5 (see Kean, 2011; Kean and Rogerson, 2012), it is evident from the left panel that price stability remains the overwhelming concern for monetary policy. For the parametric range considered, the weight on inflation $\alpha_\pi$ is strictly declining in temptation, ranging between 69 (for $\xi = 0$) and 56 (for $\xi = 0.3$). A similar pattern characterizes $\alpha_b$, although the latter is always negative, making fluctuations of debt around target always welfare-increasing. Overall, as temptation is strengthened, the central bank should be more tolerant of aggregate volatility.[32]

Optimal monetary policy is found by minimizing (36) subject to the reduced form equilibrium conditions (26), (27) and (33) written in terms of welfare-relevant gaps. Using the definition of $\hat{b}_t^*$ in (37), we have:

$$\hat{\pi}_t = \tilde{\beta} E_t \hat{\pi}_{t+1} + \kappa_y (\hat{y}_t - \hat{y}_t^*) + \hat{e}_t \quad (38)$$

$$\hat{R}_t = E_t \hat{\pi}_{t+1} + \gamma_b \left( \hat{b}_t - \hat{b}_t^* \right) + \mathcal{M}_t \quad (39)$$

$$\hat{b}_t - \hat{b}_t^* = \rho_b (\hat{b}_{t-1} - \hat{b}_{t-1}^* + \hat{R}_{t-1} - \hat{\pi}_t) + \mathcal{N}_t \quad (40)$$

[32] A more detailed discussion and some economic intuition for this result will be provided later when discussing about the welfare costs of alternative optimal policies.

where

$$\mathcal{M}_t \equiv H_z\hat{z}_t - H_g\hat{g}_t + H_v\hat{v}_{ft}, \qquad \text{with} \qquad H_i \equiv \gamma_i + \gamma_b M_i, \text{ for } i = z, g, v \tag{41}$$
$$\mathcal{N}_t \equiv (\rho_b L - 1)(M_z\hat{z}_t - M_g\hat{g}_t) + [(\rho_b L - 1) M_v - \rho_v]\,\hat{v}_{ft},$$

with $L$ denoting the lag operator. To make sure that the solution to the linear-quadratic problem indeed represents a loss minimum, we need the quadratic objective defined in (36) to be convex. This is generally the case in standard optimal monetary policy problems (with either *ad hoc* or micro-founded objectives) where all welfare weights are positive, a sufficient condition for convexity. In out context, the negativity of $\alpha_b$ rules out this argument. Following Benigno and Woodford (2005, 2006), we make use of some results in Telser and Graves (1972) and prove that our objective is indeed convex for any parameterization of the model. A detailed analysis is provided in Appendix A.1.4.[33]

## 6.1 Discretion

Under discretion, we compute the optimal time consistent monetary policy. For this purpose, we restrict the analysis to the concept of Markov-Perfect-Equilibrium (MPE), that is, an equilibrium where endogenous variables are functions only of relevant state variables, namely, the outstanding debt gap, $\hat{b}_{t-1} - \hat{b}^*_{t-1}$, the lagged nominal interest rate, $\hat{R}_{t-1}$, and all exogenous shocks (current and one-period lagged). Although a discretionary regime implies that policy announcements are not credible, current policy choices can still affect future expectations via their impact on current values for the debt gap $\hat{b}_t - \hat{b}^*_t$ and the nominal interest rate $\hat{R}_t$, both state variables in period $t+1$. The optimal monetary policy problem is therefore dynamic also under discretion, even if the policy-maker cannot strategically exploit this intertemporal linkage (which he takes as a given equilibrium relationship). This is an important element of differentiation with respect to the baseline model without temptation for which the absence of i) direct wealth effects of debt fluctuations in the Euler equation and ii) the debt gap in the policy objective allow us to seek the optimal time consistent policy by solving (analytically) a simple static loss minimization problem.

Letting $\hat{n}_t \equiv (\hat{e}_t, \hat{z}_t, \hat{g}_t, \hat{v}_{f,t})'$ denote the vector of all exogenous shocks and $V_t \equiv V(\hat{b}_{t-1} - \hat{b}^*_{t-1},$

[33] A similar issue appears in the two-country model of Groll and Monacelli (2020) where, under the baseline calibration, the micro-founded loss function features a negative weight on the squared deviation of the terms of trade from target.

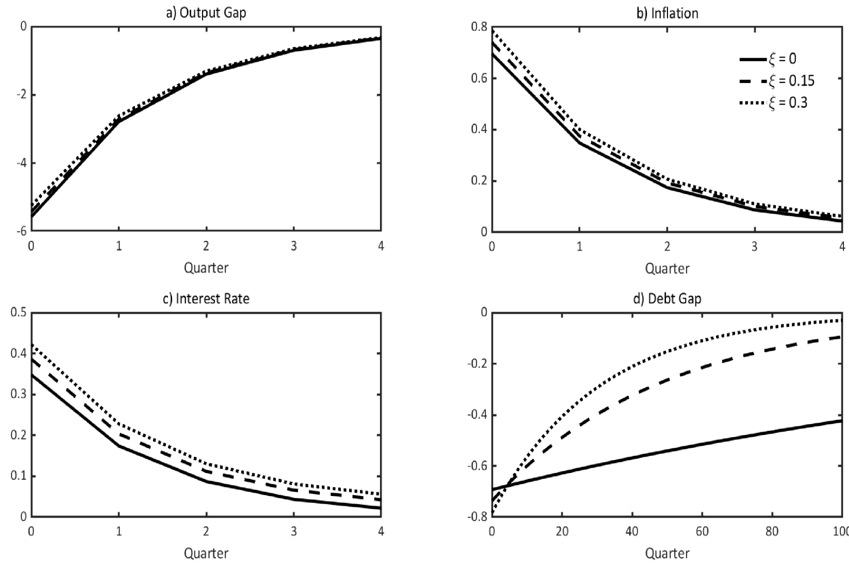


Figure 5: Impulse Responses to 1% Positive Cost Push Shock under Discretion.

$\hat{R}_{t-1}$, $\hat{n}_t$, $\hat{n}_{t-1}$) the value function, the central bank's optimization problem is described by the Bellman equation

$$V_t = \min \frac{1}{2}\left[\alpha_y \left(\hat{y}_t - \hat{y}_t^*\right)^2 + \alpha_\pi \hat{\pi}_t^2 + \alpha_b (\hat{b}_t - \hat{b}_t^*)^2\right] + \beta E_t V_{t+1}, \tag{42}$$

subject to constraints (38)-(40), and the stochastic processes for the exogenous shocks. As an analytical characterization of the solution is not possible in our case, we solve for the optimal discretionary policy following the computational procedure of Soderlind (1999), and obtain key variables' impulse responses to a 1% disturbance to the cost push shock, total factor productivity, government spending and the fiscal surplus.[34] For these, we use a textbook calibration: $\rho_i = 0.9$ for $i = z, g, f$ and $\rho_e = 0.5$, while assuming a 1% standard deviation for all of them.[35] Results are displayed in Figures 5-8. In all panels, the bold, dashed and dotted lines correspond, respectively, to a version of the model where temptation is absent ($\xi = 0$), intermediate ($\xi = 0.15$) and high ($\xi = 0.3$).

[34] The computational procedure is described in full details in Appendix A.2.

[35] The AR(1) processes for TFP, government spending and the fiscal surplus are usually assumed to be quite persistent. On the contrary, in the literature, the AR(1) coefficient $\rho_e$ for a cost push shock ranges from 0 (*iid*, as in Smets and Wouters, 2003) to 0.9 (as in Smets and Wouters, 2007). We choose the intermediate value 0.5 as our baseline.

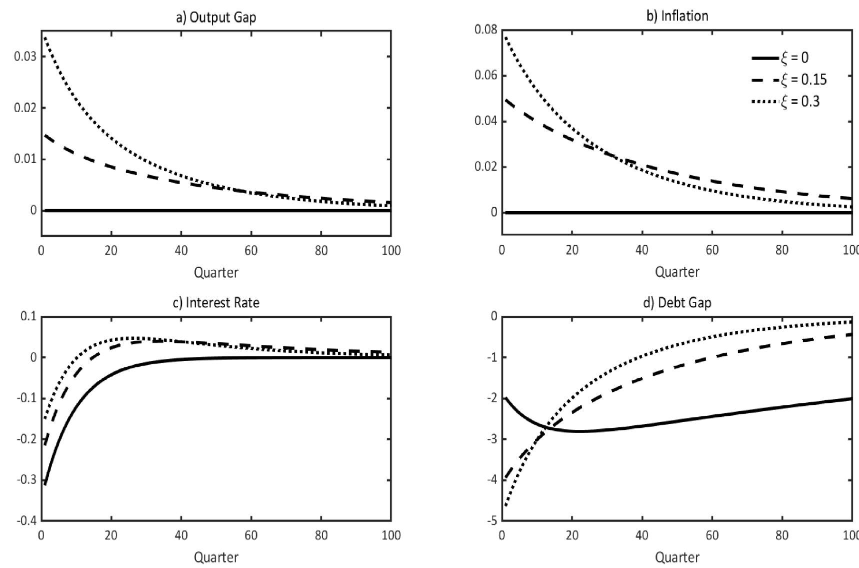


Figure 6: Impulse Responses to 1% Positive TFP Shock under Discretion.

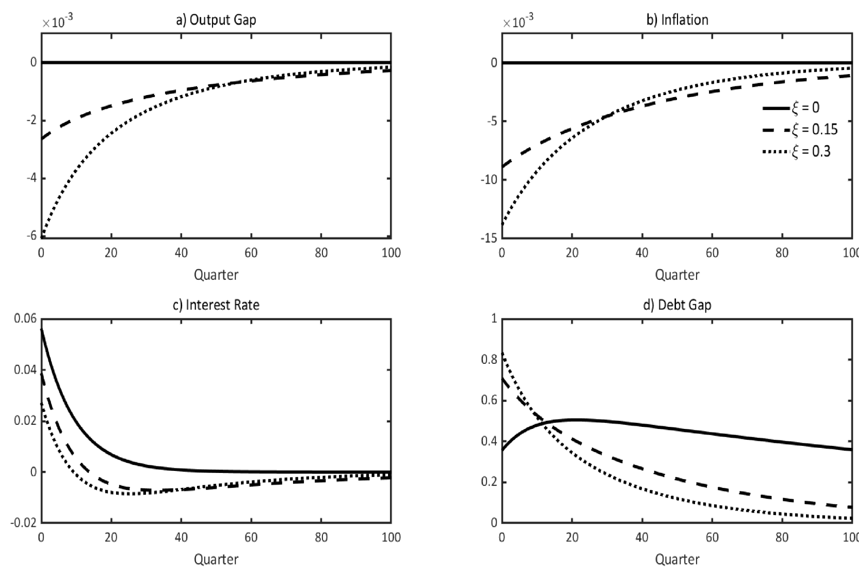


Figure 7: Impulse Responses to 1% Positive Government Spending Shock under Discretion.

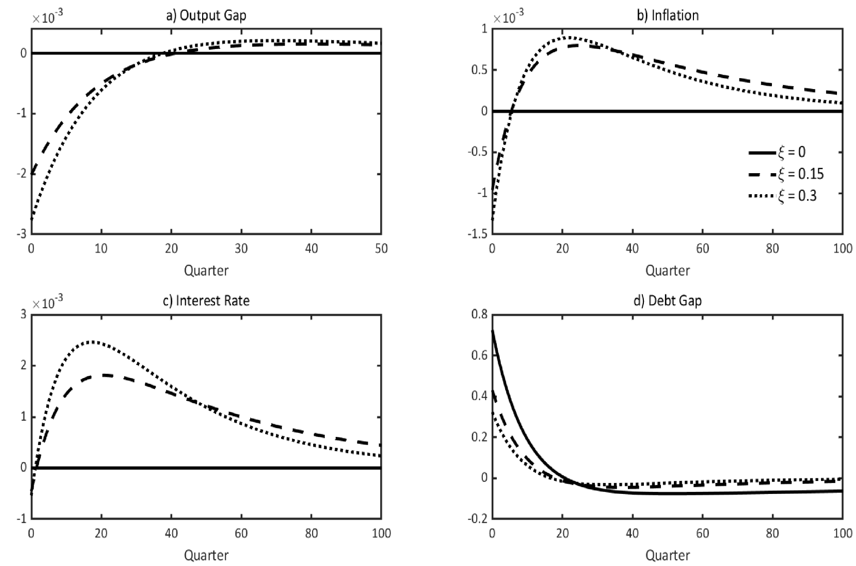


Figure 8: Impulse Responses to 1% Positive Fiscal Surplus Shock under Discretion.

In a baseline NK model without temptation, the so-called *divine coincidence* holds: the central bank can fully shield inflation and the output gap from exogenous fluctuations in TFP, government spending and the fiscal surplus (with the nominal rate and public debt playing the role of shock absorbers).[36] On the contrary, a cost push shock induces a significant jump of inflation above target and a sizable contraction in real activity.[37]

The impact of temptation is twofold. On the one hand, temptation appears to have only mild *quantitative* effects for what concerns the cost push shock: the responses of inflation, the output gap and the nominal interest rate are relatively unchanged, while the debt gap reverts back faster to steady state. On the other hand, temptation brings in interesting *qualitative* changes for the other shocks. In particular, inflation and the output gap feature a positive on-impact response to TFP (top panels in Figure 6), and negative ones to government spending and the fiscal surplus (top panels in Figures 7 and 8). The debt gap jumps significantly on impact (the more so the stronger is temptation) and takes a long time to revert back to its steady state (or welfare-relevant target) level.

To gain insight into the underlying mechanism of policy transmission, consider the minimum

[36] See Blanchard and Gali (2007) for a discussion about the divine coincidence, and Blanchard (2016) for some evidence on its empirical failure.

[37] The large responses of inflation and the output gap to the cost puch shock are due to GHH preferences. See Bardóczy et al. (2021) for a thourough discussion.

state variable (MSV) solution for inflation along the MPE:

$$\hat{\pi}_t = \pi_b \left(\hat{b}_{t-1} - \hat{b}^*_{t-1}\right) + \pi_R \hat{R}_{t-1} + \pi_n \hat{n}_t + \pi_l \hat{n}_{t-1}, \tag{43}$$

where $\pi_i$, for $i = b, R, n, l$, denotes the elasticity of inflation to variable $i$ computed by the numerical procedure.[38] Taking the first order conditions of (42) with respect to the output gap $\hat{y}_t - \hat{y}^*_t$, subject to the constraints (38)-(40) where $E_t \hat{\pi}_{t+1}$ is computed using (43), simple calculus and extensive algebra deliver a targeting rule describing the relevant trade-off for the optimal monetary policy under discretion:[39]

$$\alpha_\pi \bar{\kappa}_y \hat{\pi}_t + \alpha_y \left(\hat{y}_t - \hat{y}^*_t\right) - \alpha_b \rho_b \bar{\kappa}_y \left(\hat{b}_t - \hat{b}^*_t\right) = \beta \rho_b \left(1 + R_b\right) E_t \left[\alpha_\pi \bar{\kappa}_y \hat{\pi}_{t+1} + \alpha_y \frac{\bar{\kappa}_y}{\kappa_y} \left(\hat{y}_{t+1} - \hat{y}^*_{t+1}\right)\right] \tag{44}$$

where

$$\bar{\kappa}_y \equiv \frac{\kappa_y}{1 + \rho_b \kappa_b}, \quad \kappa_b \equiv \frac{\pi_b + \gamma_b \pi_R}{1 - \pi_R}, \quad \text{and} \quad R_b \equiv \frac{\pi_b + \gamma_b}{1 - \pi_R} \tag{45}$$

Absent temptation ($\xi = 0$), we have $\gamma_b = 0$ in the constraint (39) and $\alpha_b = 0$ in the objective (36), such that the MSV solution of inflation is a function of current shocks only, i.e. $\pi_b = \pi_R = \pi_l = 0$. In this case, equation (44) simplifies to

$$\alpha_\pi \kappa_y \hat{\pi}_t + \alpha_y \left(\hat{y}_t - \hat{y}^*_t\right) = \beta \rho_b E_t \left[\alpha_\pi \kappa_y \hat{\pi}_{t+1} + \alpha_y \left(\hat{y}_{t+1} - \hat{y}^*_{t+1}\right)\right]. \tag{46}$$

Since $\beta \rho_b \in (0, 1)$, by forward iteration, the unique stable solution to (46) is the standard *static* targeting rule:

$$\alpha_\pi \kappa_y \hat{\pi}_t + \alpha_y \left(\hat{y}_t - \hat{y}^*_t\right) = 0. \tag{47}$$

Once combined with the Phillips curve (38), the targeting rule (47) implies that, along the optimal discretionary policy path, inflation and the output gap are only driven by the cost push shock (the bold lines in Figure 5). Absent the latter, the central bank would be able to attain full inflation and output gap stabilization, $\hat{\pi}_t = \hat{y}_t - \hat{y}^*_t = 0$, i.e. the so called *divine coincidence* (the flat bold black lines in Figures 6-8).

With temptation, the targeting rule is modified along two dimensions. On the one hand, the central bank now faces a *dynamic* trade-off between stabilizing current versus expected next period

[38] Clearly, both $\pi_n$ and $\pi_l$ are vectors of coefficients.
[39] See Appendix A.3.1 for a detailed derivation.

inflation and output (the right hand side of (44)). On the other hand, it includes an additional *static* component related to deviations of current debt from target (on the left hand side of (44)). To gain a more transparent economic intuition, for analytical simplicity, consider a central bank with no explicit concern for output stabilization, $\alpha_y = 0$.[40] Letting $\alpha \equiv \alpha_b/\alpha_\pi$ be the relative welfare weight on the debt gap and $\eta \equiv \beta\rho_b (1 + R_b)$ , the targeting rule (44) reduces to

$$\hat{\pi}_t = \eta E_t \hat{\pi}_{t+1} + \alpha\rho_b \left(\hat{b}_t - \hat{b}_t^*\right). \tag{48}$$

Suppose $\pi_b = \pi_R < 0$ (which appears to be the case in the numerical solution). Since $\gamma_b \in (-1, 0)$, from the definition in (45), we have then $R_b \in (-1, 0)$, and therefore $\eta \in (0, 1)$ . Iterating equation (48) forward and imposing a standard limiting condition, we obtain

$$\hat{\pi}_t = \alpha\rho_b E_t \sum_{j=0}^{\infty} \eta^j \left(\hat{b}_{t+j} - \hat{b}_{t+j}^*\right) \tag{49}$$

According to (49), inflation should respond *negatively* (since $\alpha < 0$) to changes in the expected present discounted value of future debt deviations from target. Indeed, as displayed in panels b) and d) in Figures 6-8, the latter is *negative* for the case of TFP (hence a *positive* on-impact response of inflation), and *positive* for the case of government spending and the fiscal shock (hence a *negative* on-impact response of inflation).

Consider a positive TFP shock, and, for additional simplicity, let's suppose the shock is *iid*, such that $E_t\hat{b}_{t+j}^* = 0$ and the right hand side of (49) can be written as $\alpha\rho_b E_t \sum_{j=0}^{\infty} \eta^j \hat{b}_{t+j} - \alpha\rho_b\hat{b}_t^*$. On the one hand, with $\hat{b}_t^* = M_z\hat{z}_t$ and $M_z > 0$ (for $\xi$ within the relevant range), the debt target $\hat{b}_t^*$ increases, which, because of a negative $\alpha$, puts upward pressure on inflation. On the other hand, from the Euler equation (33) where $\gamma_z < 0$ (for any parameterization of the model), the same shock exerts a negative impact on the ex ante real interest rate, which, in turn, generates a decline in expected future debt $E_t\hat{b}_{t+j}$, for any $j \geq 1$, as the interest rate costs of servicing debt diminish. As a result of theses two channels, inflation responds positively to TFP.[41]

This outcome contrasts with the baseline NK model whereby inflation either does not respond to TFP (under optimal policy) - as the nominal rate fully absorbs any shock affecting the natural

[40] Given the overwhelmingly stronger concern for price stability, setting $\alpha_y$ to zero is, from a quantitative standpoint, without loss of generality, but involves a slight deviation from the welfare-based policy objective.

[41] A positive shock to government spending follows a similar transmission, but, with $M_g > 0$ (hence, a drop in the debt target) and $\gamma_g < 0$ (hence an increase expected future debt), it generates a persistent drop in inflation.

interest rate - or responds negatively (under a sub-optimal Taylor rule) - as real marginal costs decline with higher productivity. A similar positive response of inflation occurs instead in Heterogeneous Agents NK (HANK) models, both under optimal policy (see Davila and Schaab, 2022; Acharya et al., 2022) and instrumental Taylor rules (see Ravn and Sterk, 2021), but for a different reason. In HANK models, incomplete markets introduce an earning risk channel. When the latter is sufficiently counter-cyclical, higher TFP stimulates demand, which, in turn, puts upward pressure on price setting by firms. This mechanism - which is absent in the complete markets representative agent NK model - counter-acts the negative impact of positive TFP on marginal costs. Ravn and Sterk (2021) present empirical evidence in favor of inflationary consequences of increased TFP.

## 6.2 Commitment

Under commitment, the policy-maker announces and implements the optimal state-contingent Ramsey plan that maximizes aggregate welfare, taking into account its direct impact on individual expectations. Under the timeless perspective of Woodford (2003), the targeting rule under commitment is given by the following expressions:[42]

$$\alpha_\pi \hat{\pi}_t + \frac{\alpha_y}{\kappa_y}\left[(\hat{y}_t - \hat{y}_t^*) - \Gamma\left(\hat{y}_{t-1} - \hat{y}_{t-1}^*\right)\right] + \rho_b\left(\lambda_t^b - E_{t-1}\lambda_t^b\right) = 0, \tag{50}$$

where, letting again $\eta \equiv \beta\rho_b(1+\gamma_b) \in (0,1)$, the Lagrange multiplier on debt accumulation $\lambda_t^b$ can be written as

$$\lambda_t^b = -\alpha_b E_t \sum_{j=0}^{\infty} \eta^j \left(\hat{b}_{t+j} - \hat{b}_{t+j}^*\right) \tag{51}$$

The impulse responses under commitment displayed in Figure 9 appears *qualitatively* similar to the case of discretion, although the reversion to steady state is much faster. For instance, both inflation and the output gap absorb the initial 1% increase in TFP, government spending or the fiscal surplus in about 4 quarters, compared to 100 quarters under discretion.

Since price stability remains overwhelmingly the key policy objective for the central bank, we further scrutinize the difference between discretion and commitment for what concerns the impulse response of inflation. Figure 10 displays the results. Stronger temptation continues to amplify the on-impact response of all shocks (with a milder effect for the cost-push), while, indeed, committing to a policy yields a much faster recovery.

[42] See Appendix A.3.2 for a detailed derivation.

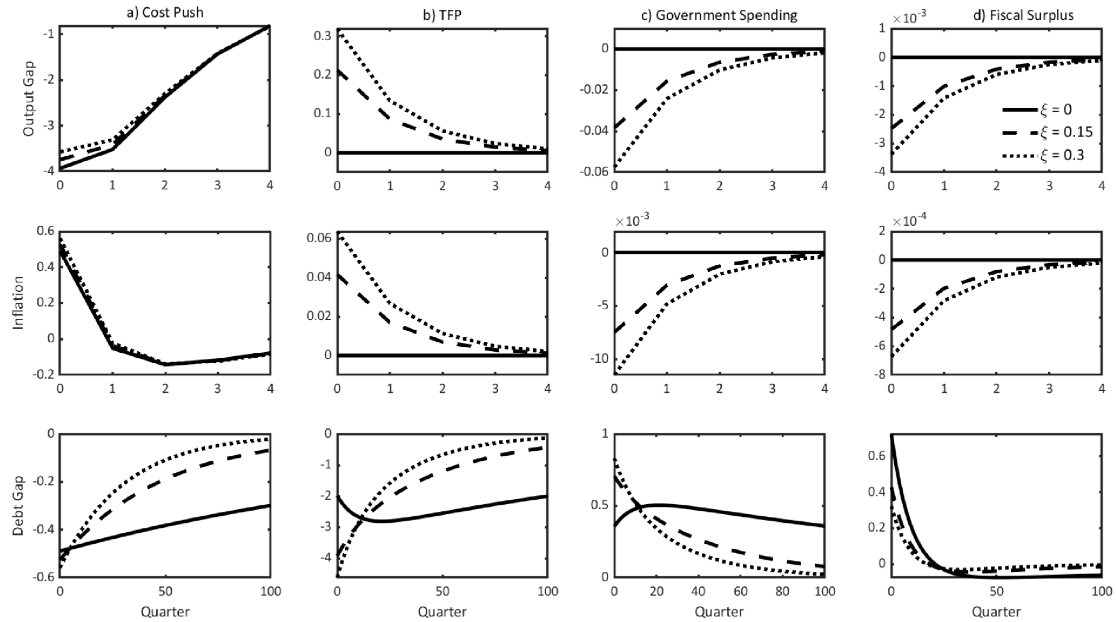


Figure 9: Impulse Responses under Commitment.

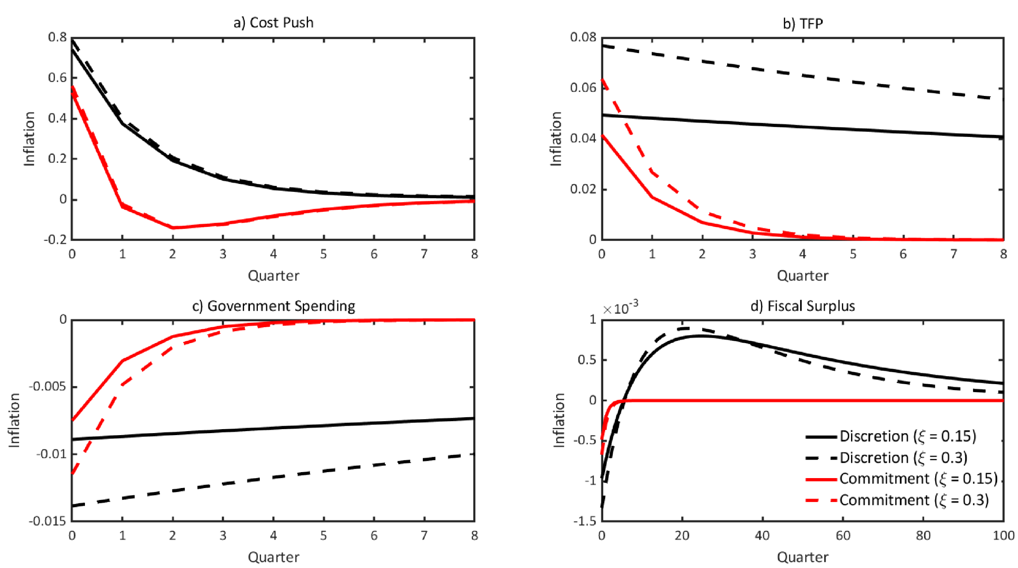


Figure 10: Inflation Responses: Discretion vs. Commitment.

Similar to the discretionary regime, the baseline case without temptation features zero responses of both inflation and the output gap to TFP, government spending and the fiscal transfer: absent the cost push shock, the divine coincidence holds. As shown in Appendix A.3.2, for $\xi = 0$ (such that $\alpha_b = \gamma_b = 0$), equation (50) collapses to the baseline targeting rule under commitment linking inflation to output gap growth:

$$\alpha_\pi \hat{\pi}_t + \frac{\alpha_y}{\kappa_y} \left[ (\hat{y}_t - \hat{y}_t^*) - \left( \hat{y}_{t-1} - \hat{y}_{t-1}^* \right) \right] = 0, \tag{52}$$

The latter combined with the Phillips curve (38) and $\hat{e}_t = 0$ indeed delivers $\hat{\pi}_t = 0$ and $\hat{y}_t = \hat{y}_t^*$ in every period.

With temptation, the targeting rule (50) requires inflation to respond to a pseudo-growth rate in the output gap, $(\hat{y}_t - \hat{y}_t^*) - \Gamma \left( \hat{y}_{t-1} - \hat{y}_{t-1}^* \right)$, but also to the debt multiplier's forecast error, $\lambda_t^b - E_{t-1}\lambda_t^b$. To compare commitment to discretion more transparently, once again, consider a central bank unconcerned about output gap fluctuations, i.e. $\alpha_y = 0$. By straightforward iteration, we can then express (50) as follows:

$$\hat{\pi}_t = \alpha \rho_b \left( E_t - E_{t-1} \right) \sum_{j=0}^{\infty} \eta^j \left( \hat{b}_{t+j} - \hat{b}_{t+j}^* \right), \tag{53}$$

With respect to the discretionary solution (49), inflation responds to the *ex post* forecast revision in the expected discounted sum of future debt deviations from target - rather then the expected discounted sum itself - and therefore displays a more contained on-impact deviation from and a faster convergence to price stability.

# 7 Welfare Analysis

In the spirit of Lucas (1987), we evaluate the *qualitative* and *quantitative* importance of temptation for the welfare cost of business cycle fluctuations under alternative monetary policy settings. Following Schmitt-Grohé and Uribe (2007), we define the consumption equivalent (CE) welfare cost of a given policy alternative $J$ as the share $\delta_J$ of steady state consumption $c$ that must be given up to make the household as well off in the stochastic equilibrium under policy $J$ as in the non-stochastic efficient steady state. Besides the optimal policies under discretion ($D$) and commitment ($C$) described in the previous section, we study two sub-optimal rules: *strict inflation targeting* ($SIT$), whereby the central bank sets $\hat{\pi}_t = 0$ at all times - and a standard *Taylor rule* ($TR$), whereby the short-term nominal rate is set according to $\hat{R}_t = \phi_\pi \hat{\pi}_t$. For this purpose, we define $\mathcal{L}_J$, the unconditional welfare-based loss function under policy $J$:[43]

$$\mathcal{L}_J \equiv \alpha_y Var\left(\hat{y}_{J,t} - \hat{y}_t^*\right) + \alpha_\pi Var\left(\hat{\pi}_{J,t}\right) + \alpha_b Var(\hat{b}_{J,t} - \hat{b}_t^*) \tag{54}$$

**Proposition 4** *Let $\mathcal{U}_J$ and $\mathcal{U}(\delta_J)$ denote the household's unconditional lifetime welfare, respectively, under monetary policy J, and at the efficient steady state with $(1-\delta_J)\,c$ replacing c. For any $\xi \geq 0$, there exists a unique $\delta_J^* < \frac{\omega}{c_y} < 1$ such that $\mathcal{U}_J = \mathcal{U}(\delta_J^*)$, with the following properties:*

**i.** $\delta_J^* \gtreqless 0$ *if and only if* $\mathcal{L}_J \gtreqless 0$;

**ii.** *if $\mathcal{L}_J$ is non-negative and strictly decreasing in $\xi$, then $\delta_J^*$ is strictly decreasing in $\xi$ as well.*

**Proof.** See Appendix A.4.4. ■

Lacking an analytical expression for the welfare loss $\mathcal{L}_J$, it is not possible to retrieve the solution $\delta_J^*$ explicitly. Moreover, both statements in Proposition 4 - i.e., that temptation may lower the welfare costs of economic fluctuations (ii.) and possibly induce some benefits (i.) - rely on assumptions about $\mathcal{L}_J$ that cannot be verified analytically. Guided by the theoretical insights of the proposition, we resort to numerical methods to find the solution $\delta_J^*$ and scrutinize its properties.

## 7.1 Temptation and Welfare: Sensitivity Analysis

Setting all parameters at baseline values, with $\phi_\pi = 1.5$ for the TR, Figures 11 plots $\delta_J^*$, for $J = D, C, SIT, TR$, as function of temptation $\xi$. As expected, for any level of temptation, commitment delivers the lowest welfare costs, followed by discretion, SIT and then a TR.[44] Since private agents' decisions depend on expectations of future quantities and prices, by announcing a credible policy plan, a committed government can strategically manipulate expectations and attain a more favorable policy trade-off than discretion when attempting to stabilizing all objectives around their targets.

However, the key takeaway from the analysis is that $\delta_J^*$ is *strictly decreasing* in $\xi$: namely, under all policies considered, temptation lowers the welfare cost of aggregate fluctuations, as predicted

[43] This expression is obtained from applying the unconditional expectation operator $E$ to $\mathcal{L}_0$ in (36). We intentionally omit the scaling factor $0.5(1-\beta)^{-1}$ multyplying $\mathcal{L}_J$.

[44] The welfare costs under a TR are strictly decreasing in the response coefficient to inflation $\phi_\pi$. Raising $\phi_\pi$ above 10 - a rather implausible scenario would generate welfare costs similar to those obtained under SIT.

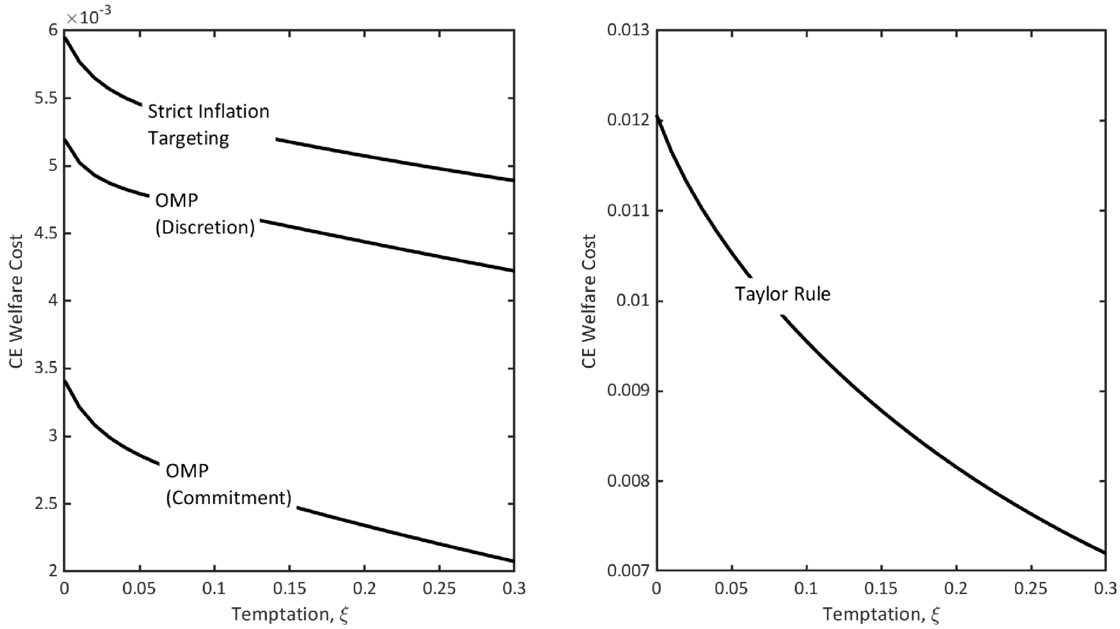


Figure 11: Consumption Equivalent (CE) Welfare Costs of Alternative Monetary Policies.

by statement ii. in Proposition 4. As it turns out, if the cost-push shock is a sufficiently weak source of uncertainty, temptation may yield welfare benefits, i.e. $\delta_J^*$ can be negative. The top left panel of Figure 12 plots $\delta_J^*$ as function of $\xi$ and the ratio of standard deviations $\sigma_e/\sigma_z$ - taking the latter as measure of the relative importance of the cost push shock volatility - for two alternative parameterization of persistence: $\rho_e = 0$ (the *iid* case) and $\rho_e = 0.5$ (baseline case). Focusing on the latter, the negatively inclined surface (dark grey) intersects the zero welfare cost flat plane (very light grey) along a concave frontier in the $\left(\xi, \frac{\sigma_e}{\sigma_z}\right)$ space. The corresponding contour plot in the right panel displays such frontier as a threshold value for $\xi$ beyond which welfare costs turn negative. A higher $\sigma_e/\sigma_z$ and/or higher persistence $\rho_e$ (both yielding stronger incidence for the cost push shock) require stronger temptation for the occurrence of welfare benefits.[45] Notice that for $\sigma_e/\sigma_z = 0$ (no cost-push shock) there are always welfare benefits from implementing the optimal discretionary policy, for any positive degree of temptation.[46]As displayed in the bottom panels of

[45] For instance, suppse $\sigma_e/\sigma_z = 0.3$. Then, the stochastic equilibrium under the optimal discretionary policy is welfare-superior to the efficient steady state for $\xi$ larger than, about, 0.1 for $\rho_e = 0.5$, but only 0.01 for $\rho_e = 0$. For $\sigma_e/\sigma_z = 0.65$ instead, fluctuations under the optimal policy are always detrimental to welfare for any $\xi \in [0, 0.3]$ and any $\rho_e \in [0, 1]$.

[46] SIT displays similar features, while, for any realistic calibration, a Taylor rule always leads to positive welfare costs. Detailed results are available from the authors upon request.

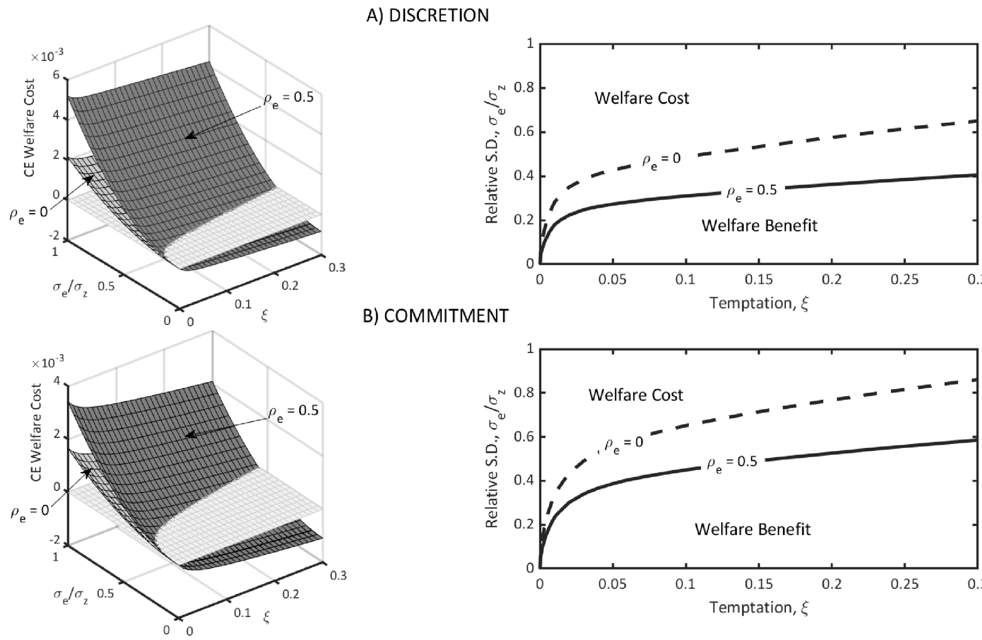


Figure 12: Cost Push Shock Volatility and the Welfare Costs of Optimal Monetary Policy.

Figure 12, the possibility of welfare benefits is even more prominent under commitment.

## 7.2 Temptation and Welfare: Inspecting the Mechanism

At the roots of these results is the fact that, with temptation, wealth volatility may increase household's welfare. We explain this apparently counter-intuitive channel in two steps. First, we show that the unconditional loss, under both discretion and commitment, is strictly decreasing in $\xi$, and attribute that to the debt gap volatility term. Then, we zoom in on the GP-utility to show that an increase in wealth volatility may generate welfare benefits by lowering the expected cognitive costs of self-control.

Figure 13 shows the effect of temptation on the welfare loss $\mathcal{L}_J$ defined in (54), for $J = D$ (discretion), $C$ (commitment).[47] Under the baseline calibration (i.e. $\sigma_e/\sigma_z = 1$), stronger temptation lowers $\mathcal{L}_J$ (second panel from the left), a result which is consistent with the sufficient conditions $\partial\mathcal{L}_J/\partial\xi < 0$ for declining welfare costs introduced in Proposition 4. It is evident from the leftmost panel that this negative pattern is mostly driven by the new term $\alpha_b Var(\hat{b}_t - \hat{b}_t^*)$: with $\alpha_b < 0$ and strictly decreasing in $\xi$ (see Figure 4 again), temptation strengthens the *negative* effect of wealth

[47] In this numerical exercise, we have set $\rho_e = 0.5$ and $\sigma_e = 0.01$, their baseline values.

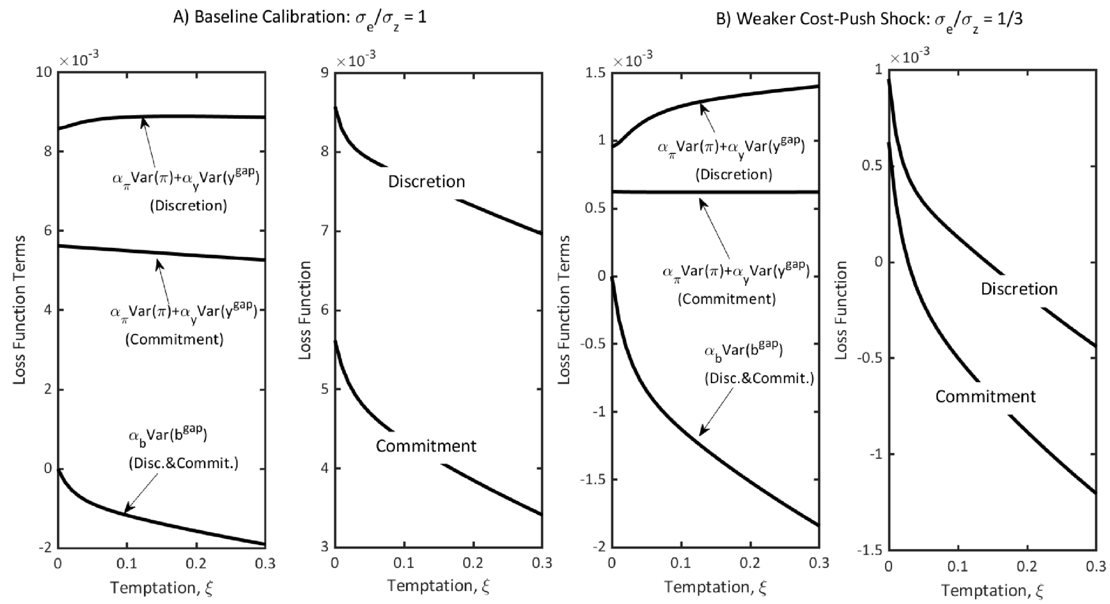


Figure 13: Loss Function and Temptation

volatility on the unconditional loss. Weakening the cost-push shock incidence (the third and fourth panels in Figure 13, with $\sigma_e/\sigma_z = 1/3$) lowers welfare losses (almost) by a factor of ten, and, most importantly, makes them negative for $\xi > 0.02$ (under commitment) and $\xi > 0.12$ (under discretion).

At the micro level, the positive relationship between temptation and welfare is a consequence of the dampening effect of wealth volatility on the cognitive costs of self-control in the GP-utility. Making use of the equilibrium condition $\tilde{x}_t = x_t + b_t$, and defining the wealth-to-income ratio $b_{x,t} \equiv b_t/x_t$, we write self-control costs $\xi(\ln \tilde{x}_t - \ln x_t)$ as function of $b_{x,t}$:[48]

$$SC(b_{x,t}) = \xi [\ln(1 + b_{x,t})] \tag{55}$$

**Proposition 5** *The expected costs of self-control $E[SC(b_{x,t})]$ decline as the wealth ratio $b_{x,t}$ becomes more volatile.*

**Proof.** See Appendix A.4.5. ■

[48] Recall that around the steady state $x = \omega y$, hence $b_x = \frac{b_y}{\omega}$ is proportional to the debt-to-GDP ratio.

## 8 Conclusions

In the baseline small-scale New Keynesian model - the backbone of larger-scale DSGE models currently used for policy analysis by central banks in developed and developing countries - absent cost-push shocks in the Phillips curve, optimal monetary policy attains full stabilization of inflation and the output gap at their respect targets (divine coincidence), while, with Ricardian equivalence holding, the central bank can pursue its dual mandate without any concerns about the dynamic of public debt. As shown in the literature, both policy outcomes are not robust to realistic *structural* modification to the baseline model, as well as the empirical evidence in their support appears, at best, rather weak.

We propose a *behavioral* modification to the baseline model which, by breaking Ricardian equivalence, yields a meaningful inflation-output stabilization trade-off for a policy-maker seeking optimal monetary policy. In our model, households are characterized by temptation with self-control preferences as formalized in Gul and Pesendorfer (2001, 2004): in every period, they are tempted to behave like a *hand-to-mouth* consumers by using their entire financial wealth for the purpose of immediate consumption. They resist the urge by exerting self-control, the latter being measured by the opportunity cost of forgoing tempting immediate consumption to pursue instead the optimal consumption smoothing plan. As self-control costs depend on the amount of wealth in hand, anticipation of such trade-off also occurring in the future yields a generalized Euler equation which is distorted by the household's holdings of Treasury bond: Ricardian equivalence fails!

To characterize optimal monetary policy, we identify the central bank's policy objective by second-order approximation of household's welfare. The implied loss function features a lower weight on inflation stabilization, and a negative weight on (squared) public debt deviations from a welfare-relevant target (a debt gap). The modified trade-off combined with the wealth distortion in the Euler equation breaks the *divine coincidence*. Similar to the baseline framework, total factor productivity and government spending affects the real interest rate (via the Euler equation), which, in turn, puts upward pressure on public debt. However, differently from the baseline, public debt, further disturbed by fiscal shocks, feeds back into the Euler equation through the above mentioned temptation-driven wealth effect.

With the Euler equation, the law of motion of public debt and the Phillips curve being now all binding policy constraints - together with the modified policy objective - the central bank can no longer fully stabilize inflation and the output gap with respect to shocks to total factor productivity and/or fiscal variables (government spending, fiscal surplus). The targeting rule under discretion features, on the one hand, a dynamic trade-off between stabilizing current versus next period inflation and output gaps, and, on the other hand, a positive response to the debt gap. Interestingly, we find the consumption-equivalent welfare costs of economic fluctuations to be strictly decreasing in temptation, potentially turning into benefits. As shown, under the assumption of concave utilities, wealth volatility dampens the cognitive costs of self-control in the GP-preference specification, hence increasing household's welfare.

# A Appendix

## A.1 Welfare Approximation

This Appendix provides a detailed derivation of the second order Taylor approximation to the representative household's welfare around the efficient steady state, and therefore a formal proof of the statement in Proposition 3.

Throughout the approximation, we will drop all those terms that are independent from policy, *t.i.p.* (i.e., terms that do not affect the optimal policy choices by the government, such as exogenous disturbances), as well as those terms that are of order of approximation higher than 2, *h.o.t.* Before getting started with the analysis, we state some useful approximation results. For generic variables $m_t$ and $n_t$ (with respective steady state $m$ and $n$), after dropping *h.o.t.*, we have the following relationships:

$$\frac{m_t - m}{m} \approx \hat{m}_t + \frac{1}{2}\hat{m}_t^2, \qquad \left(\frac{m_t - m}{m}\right)^2 \approx \hat{m}_t^2, \qquad \left(\frac{m_t - m}{m}\right)\left(\frac{n_t - n}{n}\right) \approx \hat{m}_t \hat{n}_t \tag{A.1}$$

As the analytical derivation is rather involved, we proceed in steps:

**Step 1** We derive a second-order approximation to the temporary utility $U_t$, showing that it involves the presence of linear terms in debt.

**Step 2** We pursue a second-order approximation to the government's intertemporal budget constraint from which we derive an analytical expression for the discounted sum of linear terms in debt entering the welfare objective.

**Step 3** We combine the results in to write the welfare objective as the expected present discounted value of quadratic deviations of output, inflation and debt from their respective targets.

**Step 4** We appeal to some results in Telser and Graves (1972) to prove that our loss function is convex, such that the solution to the linear-quadratic problem indeed corresponds to a loss minimum.

### A.1.1 Second-Order Approximation of Temporary Utility $U_t$ (STEP 1)

Let $U_t \equiv \ln x_t - I_t$ where $I_t \equiv \xi\left(\ln \tilde{x}_t - \ln x_t\right)$. A second-order Taylor expansion yields:[49]

$$\begin{aligned} U_t - U &\approx \frac{x_t - x}{x} - \frac{1}{2}\left(\frac{x_t - x}{x}\right)^2 - I\frac{I_t - I}{I} \\ &\approx \left(\hat{x}_t + \frac{\hat{x}_t^2}{2}\right) - \frac{\hat{x}_t^2}{2} - I\left(\hat{I}_t + \frac{\hat{I}_t^2}{2}\right) \end{aligned} \tag{A.2}$$

where the second equality makes use of the results in (A.1). Following similar steps, we obtain:

$$\frac{I_t - I}{I} \approx \frac{\xi}{I}\frac{\tilde{x}_t - \tilde{x}}{\tilde{x}} - \frac{\xi}{2I}\left(\frac{\tilde{x}_t - \tilde{x}}{\tilde{x}}\right)^2 - \frac{\xi}{I}\left(\frac{x_t - x}{x}\right) + \frac{\xi}{2I}\left(\frac{x_t - x}{x}\right)^2 \tag{A.3}$$

From the latter, we obtain:

$$\hat{I}_t + \frac{\hat{I}_t^2}{2} \approx \frac{\xi}{I}\left(\hat{\tilde{x}}_t + \frac{\hat{\tilde{x}}_t^2}{2}\right) - \frac{\xi}{I}\frac{\hat{\tilde{x}}_t^2}{2} - \frac{\xi}{I}\left(\hat{x}_t + \frac{\hat{x}_t^2}{2}\right) + \frac{\xi}{I}\frac{\hat{x}_t^2}{2} \tag{A.4}$$

$$\hat{I}_t^2 \approx \left(\frac{\xi}{I}\right)^2 \left(\hat{\tilde{x}}_t - \hat{x}_t\right)^2 \tag{A.5}$$

where (A.4) makes use of (A.1), and (A.5) comes from squaring (A.4) and dropping all *h.o.t.*

Next, consider the consumption-labor bundle $x_t \equiv c_t - \frac{h_t^{1+\chi}}{1+\chi}$. Its approximation yields

$$\frac{x_t - x}{x} \approx \frac{c}{x}\frac{c_t - c}{c} - \frac{h^{1+\chi}}{x}\frac{h_t - h}{h} - \chi\frac{h^{1+\chi}}{2x}\left(\frac{h_t - h}{h}\right)^2 \tag{A.6}$$

Making use of (A.1) again, we can rewrite the latter as

$$\left(\hat{x}_t + \frac{\hat{x}_t^2}{2}\right) \approx \frac{c}{x}\left(\hat{c}_t + \frac{\hat{c}_t^2}{2}\right) - \frac{h^{1+\chi}}{x}\left(\hat{h}_t + \frac{\hat{h}_t^2}{2}\right) - \chi\frac{h^{1+\chi}}{2x}\hat{h}_t^2 \tag{A.7}$$

[49] To shorten the notation, in the expressions to follow, we drop all *t.i.p.* and *h.o.t.*, unless required by the exposition of results.

Following similar steps, the market clearing condition $c_t = y_t - g_t$ becomes

$$\left(\hat{c}_t + \frac{\hat{c}_t^2}{2}\right) = \frac{y}{c}\left(\hat{y}_t + \frac{\hat{y}_t^2}{2}\right) - \frac{g}{c}\left(\hat{g}_t + \frac{\hat{g}_t^2}{2}\right), \tag{A.8}$$

while the aggregate technology $z_t h_t = y_t \Delta_t$ gives

$$\hat{h}_t + \frac{\hat{h}_t^2}{2} \approx \hat{\Delta}_t + \hat{y}_t + \frac{\hat{y}_t^2}{2} - \left(\hat{z}_t + \frac{\hat{z}_t^2}{2}\right) - \hat{z}_t\left(\hat{y}_t - \hat{z}_t\right), \tag{A.9}$$

$$\hat{h}_t^2 \approx \left(\hat{y}_t - \hat{z}_t\right)^2, \tag{A.10}$$

where (A.10) is obtained by squaring (A.9), and then dropping all *h.o.t.* (recalling that $\hat{\Delta}_t$ is already second-order). Plugging (A.8)-(A.10) into (A.7), while dropping *t.i.p.* and *h.o.t.*, simple algebra yields

$$\begin{aligned}\hat{x}_t + \frac{\hat{x}_t^2}{2} &\approx \frac{y}{x}\left[1 - h^\chi\right]\hat{y}_t + \frac{y}{x}\left[(1 - h^\chi) - \chi h^\chi\right]\frac{\hat{y}_t^2}{2} - \frac{y}{x}h^\chi\hat{\Delta}_t + \frac{y}{x}h^\chi\left[1+\chi\right]\hat{z}_t\hat{y}_t \\ &\approx -\frac{\chi}{\omega}\frac{\hat{y}_t^2}{2} - \frac{1}{\omega}\hat{\Delta}_t + \frac{1+\chi}{\omega}\hat{z}_t\hat{y}_t \\ &\approx -\omega^{-1}\left[\frac{\chi}{2}\left(\hat{y}_t - \hat{y}_t^*\right)^2 + \hat{\Delta}_t\right]\end{aligned} \tag{A.11}$$

where the second equality follows from $\frac{y}{x} = \frac{1}{\omega}$ (see equation (22)) and the fact that $y = h = 1$ at the efficient steady state (see Section 4), and the third from the definition of efficient output, $\hat{y}_t^* = \frac{1+\chi}{\chi}\hat{z}_t$. Making use of (A.1), we have

$$\begin{aligned}\hat{x}_t^2 &\approx \left(\frac{x_t - x}{x}\right)^2 \\ &\approx \left(\frac{c}{x}\frac{c_t - c}{c} - \frac{h^\chi h}{x}\frac{h_t - h}{h}\right)^2 \\ &\approx \left[\frac{c}{x}\left(\frac{y}{c}\hat{y}_t - \frac{g}{c}\hat{g}_t\right) - \frac{h^{1+\chi}}{x}\left(\hat{y}_t - \hat{z}_t\right)\right]^2 \\ &\approx \left[\frac{(1-s)\mu - 1}{\omega(1-s)\mu}\hat{y}_t - \frac{g_y}{\omega}\hat{g}_t + \frac{1}{\omega(1-s)\mu}\hat{z}_t^2\right]^2 \\ &= \left(-\frac{g_y}{\omega}\hat{g}_t + \frac{1}{\omega}\hat{z}_t\right)^2\end{aligned}$$

where the second line follows from (A.6) (where only first order terms are considered since the expression is then squared), the third from first order approximations of market clearing and the aggregate technology, the fourth from the definition of $\omega$ and $\frac{h}{x}h^{\chi} = \frac{y}{x}w = \frac{1}{\omega}\frac{1}{\mu(1-s)}$, and the fifth from steady state efficiency. Notice that this makes $\hat{x}_t^2$ only a functions of exogenous shocks, and therefore a $t.i.p.$

We are left with the temptation term $\tilde{x}_t = x_t + b_t$. Given its linearity, and recalling the definition $\vartheta \equiv \frac{b}{\tilde{x}}$, its approximation gives

$$\frac{\tilde{x}_t - \tilde{x}}{\tilde{x}} = (1-\vartheta)\,\frac{x_t - x}{x} + \vartheta\frac{b_t - b}{b}$$

From the latter, we find

$$\begin{aligned} \hat{\tilde{x}}_t + \frac{\hat{\tilde{x}}_t^2}{2} &= (1-\vartheta)\left(\hat{x}_t + \frac{\hat{x}_t^2}{2}\right) + \vartheta\left(\hat{b}_t + \frac{\hat{b}_t^2}{2}\right) && \text{(A.12)} \\ \hat{\tilde{x}}_t^2 &= [(1-\vartheta)\hat{x}_t + \vartheta\hat{b}_t]^2 && \text{(A.13)} \end{aligned}$$

where (A.13) comes from squaring (A.12) and dropping all $t.i.p.$

Moving back to utility (A.2), we have:

$$\begin{aligned} U_t - U &\approx \left(\hat{x}_t + \frac{\hat{x}_t^2}{2}\right) - I\frac{\xi}{I}\left[\left(\hat{\tilde{x}}_t + \frac{\hat{\tilde{x}}_t^2}{2}\right) - \frac{\hat{\tilde{x}}_t^2}{2} - \left(\hat{x}_t + \frac{\hat{x}_t^2}{2}\right)\right] \\ &\approx \left(\hat{x}_t + \frac{\hat{x}_t^2}{2}\right) - \xi I\left\{(1-\vartheta)\left(\hat{x}_t + \frac{\hat{x}_t^2}{2}\right) + \vartheta\left(\hat{b}_t + \frac{\hat{b}_t^2}{2}\right)\right. \\ &\quad \left. -\frac{1}{2}[\vartheta\hat{b}_t + (1-\vartheta)\hat{x}_t]^2 - \left(\hat{x}_t + \frac{\hat{x}_t^2}{2}\right)\right\} && \text{(A.14)} \\ &\approx (1+\xi\vartheta)\left(\hat{x}_t + \frac{\hat{x}_t^2}{2}\right) - \xi\vartheta\left(\hat{b}_t + \frac{\hat{b}_t^2}{2}\right) + \frac{\xi}{2}\left[\vartheta^2\hat{b}_t + 2\vartheta(1-\vartheta)\hat{x}\hat{b}_t\right] \\ &\approx -\frac{(1+\xi\vartheta)}{\omega}\left[\frac{\chi}{2}\left(\hat{y}_t - \hat{y}_t^*\right)^2 + \hat{\Delta}_t\right] - \xi\vartheta\,(1-\vartheta)\left(\frac{\hat{b}_t^2}{2} - \hat{x}\hat{b}_t\right) - \xi\vartheta\hat{b}_t \end{aligned}$$

where the first equality makes use of (A.4)-(A.5), the second of (A.12)-(A.13), the third regroups similar terms and makes use of the fact that $\hat{x}_t^2 \approx 0$ (being a function of shocks only), and the fourth is a consequence of (A.11) and a simple rearrangement of terms. Notice that for $\xi = 0$, the utility approximation reduces to $-\omega^{-1}\left[\frac{\chi}{2}\left(\hat{y}_t - \hat{y}_t^*\right)^2 + \hat{\Delta}_t\right]$: i.e., absent temptation, temporary utility depends negatively on price dispersion and the squared deviation of output from its efficient level. Letting $U_y \equiv \frac{1+\xi\vartheta}{\omega}$ and $U_b \equiv \xi\vartheta$, we can write (A.14) more compactly as follows:

$$U_t - U \approx -U_y\left[\frac{\chi}{2}\left(\hat{y}_t - \hat{y}_t^*\right)^2 + \hat{\Delta}_t\right] - U_b\left[\hat{b}_t + (1-\vartheta)\left(\frac{\hat{b}_t^2}{2} - \hat{b}_t\hat{x}_t\right)\right], \tag{A.15}$$

which corresponds to equation (35) in Section 6. Hence, up to a second order approximation, the maximization of the representative household's welfare is:

$$\max \mathcal{U}_0 \approx -U_y E_0 \sum_{t=0}^{\infty} \beta^t \left[\frac{\chi}{2}\left(\hat{y}_t - \hat{y}_t^*\right)^2 + \hat{\Delta}_t\right] - U_b E_0 \sum_{t=0}^{\infty} \beta^t \left[\hat{b}_t + (1-\vartheta)\left(\frac{\hat{b}_t^2}{2} - \hat{b}_t\hat{x}_t\right)\right] \tag{A.16}$$

A technical issue with the formulation in (A.16) is the presence of the sum of linear terms in $\hat{b}_t$. Following the discussion in Benigno and Woodford(2005), this undermines the evaluation accuracy of alternative policy specifications (including the truly optimal one) using a linear-quadratic framework. While a linear term in the output gap (due to the monopolistic competition distortion) has been eliminated by the efficient labor subsidy, we take care of $\hat{b}_t$ by resorting to a second order approximation to the government's intertemporal budget constraint.

### A.1.2 Second-Order Approximation of Government's Intertemporal Budget (STEP 2)

The following Lemma states the key result of this section.

**Lemma 2** *Let* $\mathcal{W}_t \equiv \lambda_t \frac{R_{t-1}b_{t-1}}{\pi_t}$, *for* $t \geq 0$. *Then*

$$\frac{\mathcal{W}_0 - \mathcal{W}}{\mathcal{W}} \approx \Upsilon_b E_0 \sum_{t=0}^{\infty} \tilde{\beta}^t \hat{b}_t + E_0 \sum_{t=0}^{\infty} \tilde{\beta}^t \left[\Upsilon_{bb}\frac{\hat{b}_t^2}{2} - \Upsilon_\pi \hat{\Delta}_t + \hat{b}_t\left(\Upsilon_{bz}\hat{z}_t + \Upsilon_{bv}\hat{v}_{ft} - \Upsilon_{bg}\hat{g}_t\right)\right] \tag{A.17}$$

*where* $\Upsilon_i$ *for* $i = b, \pi,$ *and* $\Upsilon_{bj}$ *for* $j = b, z, v, g$ *are composite coefficients.*

To prove the result in (A.17), we start from the government budget constraint (16), which, after setting $\tau_t^F = s_t w_t h_t$, becomes:

$$\frac{R_{t-1}b_{t-1}}{\pi_t} = f_t + b_t, \tag{A.18}$$

where $f_t = \tau_t^H - g_t$ denotes its primary surplus. Its forward iteration gives:

$$\begin{aligned}
\frac{R_{t-1}b_{t-1}}{\pi_t} &= f_t + \beta E_t\left(\frac{\lambda_{t+1}}{\lambda_t}\Gamma_{t+1}\frac{R_t}{\pi_{t+1}}b_t\right) \\
&= f_t + \beta E_t\left(\frac{\lambda_{t+1}}{\lambda_t}\Gamma_{t+1}f_{t+1}\right) + \beta E_t\left(\frac{\lambda_{t+1}}{\lambda_t}\Gamma_{t+1}b_{t+1}\right) \\
&= f_t + \beta E_t\left(\frac{\lambda_{t+1}}{\lambda_t}\Gamma_{t+1}f_{t+1}\right) + \beta E_t\left\{\frac{\lambda_{t+1}}{\lambda_t}\Gamma_{t+1}E_{t+1}\left[\beta\frac{\lambda_{t+2}}{\lambda_{t+1}}\Gamma_{t+2}\left(f_{t+2}+b_{t+2}\right)\right]\right\} \\
&= f_t + \beta E_t\left(\frac{\lambda_{t+1}}{\lambda_t}f_{t+1}\Gamma_{t+1}\right) + \beta^2 E_t\left(\frac{\lambda_{t+2}}{\lambda_t}f_{t+2}\Gamma_{t+1}\Gamma_{t+2}\right) + \beta^2 E_t\left(\frac{\lambda_{t+2}}{\lambda_t}\Gamma_{t+1}\Gamma_{t+2}b_{t+2}\right)
\end{aligned}$$

where the first equality follows from the fact that $b_t = R_t E_t\left(\beta\frac{\lambda_{t+1}}{\lambda_t}\frac{\Gamma_{t,t+1}}{\pi_{t+1}}b_t\right)$ by the Euler equation (20); the second is obtained by leading (A.18) one period forward to find the expression for $\frac{R_t}{\pi_{t+1}}$; the third by using (20) again (as for the first line) to substitute for $b_{t+1}$ and then by making use of fact that $\frac{R_{t+1}b_{t+1}}{\pi_{t+2}} = f_{t+2} + b_{t+2}$; and the fourth by a simple rearrangement of terms. Continuing with the forward iteration and imposing a standard transversality condition gives the household's intertemporal budget constraint (or intertemporal solvency condition):

$$\frac{R_{t-1}b_{t-1}}{\pi_t} = E_t\sum_{j=0}^{\infty}\beta^j\frac{\lambda_{t+j}}{\lambda_t}f_{t+j}\Gamma_{t,t+j} \tag{A.19}$$

where $\Gamma_{t,t+j} \equiv \Pi_{i=1}^{j}\Gamma_{t+i}$, with $\Gamma_t = 1$. Multiplying both sides of (A.19) by $\lambda_t$, while denoting $\mathcal{W}_t \equiv \lambda_t\frac{R_{t-1}b_{t-1}}{\pi_t}$ and $q_{t,t+j} \equiv \lambda_{t+j}f_{t+j}\Gamma_{t,t+j}$ for $j \geq 0$, we rewrite (A.19) more compactly as follows:

$$\mathcal{W}_t = E_t\sum_{j=0}^{\infty}\beta^j q_{t,t+j} \tag{A.20}$$

We introduce the following definition for later use:

$$Q_{t+j} \equiv E_{t+j} \sum_{i=0}^{\infty} \beta^i q_{t+j,t+j+i}, \quad \text{for} \quad j \geq 0, \tag{A.21}$$

such that $\mathcal{W}_{t+j} = Q_{t+j}$ for any $j \geq 0$.

Recalling the definition of $\tilde{\beta} \equiv \beta\Gamma$ and noticing that, at the steady state, $\Gamma_{t,t+j} = \Gamma^j$, a second-order approximation to (A.20) yields

$$\mathcal{W} + \mathcal{W}\frac{\mathcal{W}_t - \mathcal{W}}{\mathcal{W}} \approx \frac{\lambda f}{1-\tilde{\beta}} + \lambda f E_t \sum_{j=0}^{\infty} \tilde{\beta}^j \left( \frac{q_{t,t+j} - q}{q} \right)$$

where the first term on the right hand side corresponds to the steady state of the right hand side of (A.20). Since $\mathcal{W} = \frac{\lambda f}{1-\tilde{\beta}}$, the first approximation result in (A.1) implies that

$$\frac{\mathcal{W}_t - \mathcal{W}}{\mathcal{W}} \approx \left(1-\tilde{\beta}\right) E_t \sum_{j=0}^{\infty} \tilde{\beta}^j \left( \hat{q}_{t,t+j} + \frac{\hat{q}_{t,t+j}^2}{2} \right). \tag{A.22}$$

We are going to evaluate the infinite summations of linear and quadratic terms in (A.22) separately. To do that, define $q_{t+j} \equiv \lambda_{t+j} f_{t+j}$, such that $q_{t,t+j} = q_{t+j}\Gamma_{t,t+j}$, and therefore $\hat{q}_{t,t+j} = \hat{q}_{t+j} + \hat{\Gamma}_{t,t+j}$ where $\hat{\Gamma}_{t,t+j} = \sum_{i=t+1}^{t+j} \hat{\Gamma}_i$. Recalling that $\hat{\Gamma}_t = 0$ (since $\Gamma_t = 1$), after simple algebra, the sum of linear terms in (A.22) is therefore equivalent to

$$E_t \sum_{j=0}^{\infty} \tilde{\beta}^j \hat{q}_{t,t+j} = E_t \sum_{j=0}^{\infty} \tilde{\beta}^j \hat{q}_{t+j} + \frac{1}{1-\tilde{\beta}} E_t \sum_{j=1}^{\infty} \tilde{\beta}^j \hat{\Gamma}_{t+j}. \tag{A.23}$$

Using the expression for $\hat{q}_{t,t+j}$ again, we obtain the quadratic term $\hat{q}^2_{t,t+j} = \hat{q}^2_{t+j} + \hat{\Gamma}^2_{t,t+j} + 2\hat{q}_{t+j}\hat{\Gamma}_{t,t+j}$ entering (A.22). Let's evaluate first the summation $E_t \sum_{j=0}^{\infty} \tilde{\beta}^J \hat{\Gamma}^2_{t,t+j}$:

$$
\begin{aligned}
& E_t \sum_{j=0}^{\infty} \tilde{\beta}^J \hat{\Gamma}^2_{t,t+j} = E_t \sum_{j=0}^{\infty} \tilde{\beta}^j \left( \sum_{i=t+1}^{t+j} \hat{\Gamma}_i \right)^2 \\
& = E_t \Big[ \tilde{\beta}\hat{\Gamma}^2_{t+1} + \tilde{\beta}^2 \left( \hat{\Gamma}^2_{t+1} + \hat{\Gamma}^2_{t+2} + 2\hat{\Gamma}_{t+1}\hat{\Gamma}_{t+2} \right) \\
& + \tilde{\beta}^3 \left( \hat{\Gamma}^2_{t+1} + \hat{\Gamma}^2_{t+2} + \hat{\Gamma}^2_{t+3} + 2\hat{\Gamma}^2_{t+1}\hat{\Gamma}^2_{t+2} + 2\hat{\Gamma}^2_{t+2}\hat{\Gamma}^2_{t+3} + 2\hat{\Gamma}^2_{t+1}\hat{\Gamma}^2_{t+3} \right) + ... \Big] \\
& = E_t \Big\{ \tilde{\beta}\hat{\Gamma}_{t+1} \Big[ \hat{\Gamma}_{t+1}(1 + \tilde{\beta} + \tilde{\beta}^2 + ...) + 2\tilde{\beta}\hat{\Gamma}_{t+2}(1 + \tilde{\beta} + \tilde{\beta}^2 + ...) + 2\tilde{\beta}^2\hat{\Gamma}_{t+3}(1 + \tilde{\beta} + \tilde{\beta}^2 + ...) + ... \Big] \\
& + \tilde{\beta}^2\hat{\Gamma}_{t+2} \Big[ \hat{\Gamma}_{t+2}(1 + \tilde{\beta} + \tilde{\beta}^2 + ...) + 2\tilde{\beta}\hat{\Gamma}_{t+3}(1 + \tilde{\beta} + \tilde{\beta}^2 + ...) + 2\tilde{\beta}^2\hat{\Gamma}_{t+4}(1 + \tilde{\beta} + \tilde{\beta}^2 + ...) ... \Big] + ... \Big\} \\
& = \frac{1}{1-\tilde{\beta}} E_t \Big\{ \tilde{\beta}\hat{\Gamma}_{t+1} \left( \hat{\Gamma}_{t+1} + 2\tilde{\beta}\hat{\Gamma}_{t+2} + 2\tilde{\beta}^2\hat{\Gamma}_{t+3} + ... \right) + \tilde{\beta}^2\hat{\Gamma}_{t+2} \left( \hat{\Gamma}_{t+2} + 2\tilde{\beta}\hat{\Gamma}_{t+3} + 2\tilde{\beta}^2\hat{\Gamma}_{t+4} + ... \right) + ... \Big\} \\
& = \frac{1}{1-\tilde{\beta}} E_t \sum_{j=1}^{\infty} \tilde{\beta}^j \hat{\Gamma}_{t+j} \left[ \hat{\Gamma}_{t+j} + 2\hat{\mathcal{P}}_{t+j} \right] \qquad \text{(A.24)}
\end{aligned}
$$

where we have defined $\hat{\mathcal{P}}_{t+j} \equiv \sum_{i=1}^{\infty} \tilde{\beta}^i \hat{\Gamma}_{t+j+i}$. Next, we consider

$$
\begin{aligned}
E_t \sum_{j=0}^{\infty} \tilde{\beta}^j \hat{q}_{t+j}\hat{\Gamma}_{t,t+j} &= E_t \left[ \tilde{\beta}\hat{q}_{t+1}\hat{\Gamma}_{t+1} + \tilde{\beta}^2 q_{t+2}(\hat{\Gamma}_{t+1} + \hat{\Gamma}_{t+2}) + \tilde{\beta}^3 q_{t+3}(\hat{\Gamma}_{t+1} + \hat{\Gamma}_{t+2} + \hat{\Gamma}_{t+3}) + ... \right] \\
&= E_t \left[ \tilde{\beta}\hat{\Gamma}_{t+1}(\hat{q}_{t+1} + \tilde{\beta}\hat{q}_{t+2}) + \tilde{\beta}^2\hat{\Gamma}_{t+2}(\hat{q}_{t+2} + \tilde{\beta}\hat{q}_{t+3} + ...) + ... \right] \\
&= E_t \sum_{j=1}^{\infty} \tilde{\beta}^j \hat{\Gamma}_{t+j}\hat{\Theta}_{t+j} \qquad \text{(A.25)}
\end{aligned}
$$

where we have defined $\hat{\Theta}_{t+j} \equiv \left( \sum_{i=t+j}^{\infty} \tilde{\beta}^{i-(t+j)} \hat{q}_i \right)$, such that $\hat{\Theta}_t = \sum_{j=0}^{\infty} \tilde{\beta}^j \hat{q}_{t+j}$ (to be used later). We substitute (A.23)-(A.25) back into (A.22):

$$
\begin{aligned}
\frac{\mathcal{W}_t - \mathcal{W}}{\mathcal{W}} \approx & \left(1 - \tilde{\beta}\right) E_t \sum_{j=0}^{\infty} \tilde{\beta}^j \left( \hat{q}_{t+j} + \frac{\hat{q}^2_{t+j}}{2} \right) + E_t \sum_{j=1}^{\infty} \tilde{\beta}^j \hat{\Gamma}_{t+j} \\
& + \frac{1}{2} E_t \sum_{j=1}^{\infty} \tilde{\beta}^j \hat{\Gamma}^2_{t+j} + \underbrace{E_t \sum_{j=1}^{\infty} \tilde{\beta}^j \hat{\Gamma}_{t+j}\hat{\mathcal{P}}_{t+j}}_{A_2} + \underbrace{\left(1 - \tilde{\beta}\right) E_t \sum_{j=1}^{\infty} \tilde{\beta}^j \hat{\Gamma}_{t+j}\hat{\Theta}_{t+j}}_{\equiv A_1} \qquad \text{(A.26)}
\end{aligned}
$$

Next, we are going to simplify (A.26) as much as possible. Let's start with term $A_1$ in (A.26):

$$\begin{aligned} A_1 &= (1-\tilde{\beta})E_t\left[\sum_{j=1}^{\infty}\tilde{\beta}^j\hat{\Gamma}_{t+j}E_{t+j}\left(\hat{q}_{t+j}+\tilde{\beta}\hat{q}_{t+j+1}+\tilde{\beta}^2\hat{q}_{t+j+2}+...\right)\right] \\ &= (1-\tilde{\beta})E_t\left[\sum_{j=1}^{\infty}\tilde{\beta}^j\hat{\Gamma}_{t+j}E_{t+j}\left(\sum_{i=0}^{\infty}\tilde{\beta}^i\hat{q}_{t+j,t+j+i}-\frac{1}{1-\tilde{\beta}}\sum_{i=1}^{\infty}\tilde{\beta}^i\hat{\Gamma}_{t+j+i}\right)\right] \end{aligned}$$

where the second equality follows from (A.23). Notice that, from the definition of $\hat{\mathcal{P}}_{t+j}$, the sum of terms $A_1$ and $A_2$ in (A.26) equals

$$\begin{aligned} A_1+A_2 &= (1-\tilde{\beta})E_t\left[\sum_{j=1}^{\infty}\tilde{\beta}^j\hat{\Gamma}_{t+j}E_{t+j}\left(\sum_{i=0}^{\infty}\tilde{\beta}^i\hat{q}_{t+j,t+j+i}\right)\right] \\ &= E_t\sum_{j=1}^{\infty}\tilde{\beta}^j\hat{\Gamma}_{t+j}\hat{Q}_{t+j} \qquad \text{(A.27)} \end{aligned}$$

where the second equality follows from the definition of $Q_{t+j}$ in (A.21) and the fact that, to first order approximation,

$$\hat{Q}_{t+j} \equiv \left(1-\tilde{\beta}\right)E_{t+j}\sum_{i=0}^{\infty}\tilde{\beta}^i\hat{q}_{t+j,t+j+i}, \quad \text{for} \quad j \geq 0, \qquad \text{(A.28)}$$

Plugging (A.27) back into (A.26), together with $\hat{\Gamma}_t = 0$, after a simple rearrangement of terms, we obtain:

$$\frac{\mathcal{W}_t-\mathcal{W}}{\mathcal{W}} \approx E_t\sum_{j=0}^{\infty}\tilde{\beta}^j\left[\left(1-\tilde{\beta}\right)\left(\hat{q}_{t+j}+\frac{\hat{q}_{t+j}^2}{2}\right)+\left(\hat{\Gamma}_{t+j}+\frac{\hat{\Gamma}_{t+j}^2}{2}\right)+\hat{\Gamma}_{t+j}\hat{Q}_{t+j}\right] \qquad \text{(A.29)}$$

Consider the term $\hat{q}_{t+j}$. We can write it as follows we have

$$\begin{aligned} \hat{q}_{t+j} &= \hat{\lambda}_{t+j}+\hat{f}_{t+j} \\ &= -\hat{x}_{t+j}+\phi_b\hat{b}_{t+j}+\hat{v}_{f,t+j} \qquad \text{(A.30)} \end{aligned}$$

where the second equality makes use $\hat{\lambda}_{t+j} = -\hat{x}_{t+j}$ and the linearized version of the fiscal rule (17). Furthermore, recalling the result in (30), we have

$$\hat{\Gamma}_{t+j} = \Gamma_x \hat{x}_{t+j} + \Gamma_b \hat{b}_{t+j}, \quad \text{for} \quad \Gamma_b \equiv \varkappa\vartheta, \quad \Gamma_x = -\Gamma_b \tag{A.31}$$

Finally, since $\mathcal{W}_{t+j} \equiv \frac{R_{t+j-1}b_{t+j-1}}{\pi_{t+j}}\lambda_{t+j}$ and $\mathcal{W}_{t+j} = Q_{t+j}$ for any $j \geq 0$, from the fiscal budget (A.18) we have $Q_{t+j} = \lambda_{t+j}\left(f_{t+j} + b_{t+j}\right)$, whose first order approximation gives

$$\begin{aligned} \hat{Q}_{t+j} &= \hat{\lambda}_{t+j} + \left(1-\tilde{\beta}\right)\hat{f}_{t+j} + \tilde{\beta}\hat{b}_{t+j} \\ &= -\hat{x}_{t+j} + Q_b\hat{b}_{t+j} + \left(1-\tilde{\beta}\right)\hat{v}_{f,t+j}, \quad \text{for} \quad Q_b \equiv \left(1-\tilde{\beta}\right)\phi_b + \tilde{\beta} \end{aligned} \tag{A.32}$$

Making use of equations (A.30)-(A.31) (as well as their square, while dropping *t.i.p.*) and equation (A.32), we rewrite (A.29) as follows:

$$\begin{aligned} \frac{\mathcal{W}_t - \mathcal{W}}{\mathcal{W}} &= E_t \sum_{j=0}^{\infty} \tilde{\beta}^j \left[ \Upsilon_x \hat{x}_{t+j} + \Upsilon_b \hat{b}_{t+j} + \Upsilon_{xx} \frac{\hat{x}_{t+j}^2}{2} + \Upsilon_{bb} \frac{\hat{b}_{t+j}^2}{2} \right. \\ &\quad \left. + \Upsilon_{bx} \hat{b}_{t+j} \hat{x}_{t+j} + \Upsilon_{bv} \hat{b}_{t+j} \hat{v}_{f,t+j} + \Upsilon_{xv} \hat{x}_{t+j} \hat{v}_{f,t+j} \right] \end{aligned} \tag{A.33}$$

where

$$\Upsilon_x \equiv \Gamma_x - (1-\tilde{\beta}) \tag{A.34}$$

$$\Upsilon_b \equiv \Gamma_b + (1-\tilde{\beta})\phi_b \tag{A.35}$$

$$\Upsilon_{xx} \equiv (1-\tilde{\beta}) + \Gamma_x(\Gamma_x - 1) \tag{A.36}$$

$$\Upsilon_{bb} \equiv \Gamma_b(\Gamma_b + Q_b) + (1-\tilde{\beta})\phi_b^2 \tag{A.37}$$

$$\Upsilon_{bx} \equiv \Gamma_x\Gamma_b - (1-\tilde{\beta})\phi_b + \Gamma_x Q_b - \Gamma_b \tag{A.38}$$

$$\Upsilon_{bv} \equiv (1-\tilde{\beta})\phi_b + \Gamma_b(1-\tilde{\beta}) \tag{A.39}$$

$$\Upsilon_{xv} \equiv \Gamma_x(1-\tilde{\beta}) \tag{A.40}$$

To conclude the derivation, we recall that, around the efficient steady state, a second-order approximation to the consumption-labor composite $x_t$ yields

$$\hat{x}_t = \omega^{-1}\left(\hat{z}_t - g_y\hat{g}_t - \hat{\Delta}_t\right), \tag{A.41}$$

After plugging the latter, as well as its square, back into (A.33) and dropping *t.i.p.* and *h.o.t.*, a simple rearrangement of terms gives

$$\frac{\mathcal{W}_t - \mathcal{W}}{\mathcal{W}} = \Upsilon_b E_t \sum_{j=0}^{\infty} \tilde{\beta}^j \hat{b}_{t+j} + E_t \sum_{j=0}^{\infty} \tilde{\beta}^j \left[\Upsilon_{bb}\frac{\hat{b}_{t+j}^2}{2} - \Upsilon_\pi \hat{\Delta}_{t+j} + \left(\Upsilon_{bz}\hat{z}_{t+j} + \Upsilon_{bv}\hat{v}_{f,t+j} - \Upsilon_{bg}\hat{g}_{t+j}\right)\hat{b}_{t+j}\right]$$

where $\Upsilon_\pi \equiv \omega^{-1}\Upsilon_x$. Setting $t = 0$ and relabeling the time subscript $j$ by $t$, we obtain the expression in (A.17)

### A.1.3 Elimination of Linear Terms from the Welfare Objective (STEP 3)

Equation (A.17) implies that

$$E_0 \sum_{t=0}^{\infty} \tilde{\beta}^t \hat{b}_t \approx \Upsilon_b^{-1}\frac{\mathcal{W}_0 - \mathcal{W}}{\mathcal{W}} + \Upsilon_b^{-1} E_0 \sum_{t=0}^{\infty} \tilde{\beta}^t \left[\Upsilon_\pi \hat{\Delta}_t - \Upsilon_{bb}\frac{\hat{b}_t^2}{2} - \hat{b}_t\left(\Upsilon_{bz}\hat{z}_t + \Upsilon_{bv}\hat{v}_{ft} - \Upsilon_{bg}\hat{g}_t\right)\right] \tag{A.42}$$

Notice that the discount factor in (A.42) is $\tilde{\beta}$ while it is $\beta > \tilde{\beta}$ in (A.16). To substitute (A.42) into (A.16), we assume $E_0 \sum_{t=0}^{\infty} \beta^t \hat{b}_t \approx E_0 \sum_{t=0}^{\infty} \tilde{\beta}^t \hat{b}_t + h.o.t.$, that is, a discrepancy of order higher than two. This is the case if $\tilde{\beta}$ is sufficiently close to $\beta$.[50]

Based on this argument, we replace $E_0 \sum_{t=0}^{\infty} \tilde{\beta}^t \hat{m}_t$ with $E_0 \sum_{t=0}^{\infty} \beta^t \hat{m}_t$ in (A.42) for any stochastic process $\hat{m}_t$, including, obviously, $\hat{b}_t$:

$$E_0 \sum_{t=0}^{\infty} \beta^t \hat{b}_t \approx \Upsilon_b^{-1}\frac{\mathcal{W}_0 - \mathcal{W}}{\mathcal{W}} - \mathcal{A}_{\mathcal{W}} \tag{A.43}$$

[50] Using previous definitions and steady state relationships, simple algebra gives

$$\beta - \tilde{\beta} = \beta \frac{\omega}{\omega + b_y} \frac{\xi I^\varsigma}{1 + \xi I^\varsigma}.$$

In the quantitative exercise, we restrict to parameterizations for which $\beta \approx \tilde{\beta} + h.o.t.$

where we have defined

$$\mathcal{A}_{\mathcal{W}} \equiv \Upsilon_b^{-1} E_0 \sum_{t=0}^{\infty} \beta^t \left[ -\Upsilon_\pi \hat{\Delta}_t + \Upsilon_{bb} \frac{\hat{b}_t^2}{2} + \hat{b}_t \left( \Upsilon_{bz} \hat{z}_t + \Upsilon_{bv} \hat{v}_{ft} - \Upsilon_{bg} \hat{g}_t \right) \right] \tag{A.44}$$

Similarly, we let

$$\mathcal{A}_{\mathcal{U}} \equiv -U_y E_0 \sum_{t=0}^{\infty} \beta^t \left[ \frac{\chi}{2} \left( \hat{y}_t - \hat{y}_t^* \right)^2 + \hat{\Delta}_t \right] - U_b E_0 \sum_{t=0}^{\infty} \beta^t \left( 1 - \vartheta \right) \left( \frac{\hat{b}_t^2}{2} - \hat{b}_t \hat{x}_t \right), \tag{A.45}$$

and write the welfare objective $\mathcal{U}_0$ in (A.16) as

$$\mathcal{U}_0 \approx -U_b E_0 \sum_{t=0}^{\infty} \beta^t \hat{b}_t + \mathcal{A}_{\mathcal{U}} \tag{A.46}$$

We plug (A.43) in the latter to obtain

$$\begin{aligned}
\mathcal{U}_0 &\approx \mathcal{A}_{\mathcal{U}} + U_b \mathcal{A}_{\mathcal{W}} + t.i.p. \\
&\approx -U_y E_0 \sum_{t=0}^{\infty} \beta^t \left[ \frac{\chi}{2} \left( \hat{y}_t - \hat{y}_t^* \right)^2 + \frac{\epsilon}{2\kappa} \hat{\pi}_t^2 \right] - U_b E_0 \sum_{t=0}^{\infty} \beta^t \left( 1 - \vartheta \right) \left( \frac{\hat{b}_t^2}{2} - \hat{b}_t \hat{x}_t \right) \\
&\quad + \Upsilon_b^{-1} E_0 \sum_{t=0}^{\infty} \beta^t \left[ -\Theta_\pi \hat{\Delta}_t + \Theta_{bb} \frac{\hat{b}_t^2}{2} + \hat{b}_t \left( \Theta_{bz} \hat{z}_t + \Theta_{bv} \hat{v}_{ft} - \Theta_{bg} \hat{g}_t \right) \right] \\
&\approx -\frac{U_y}{2} E_0 \sum_{t=0}^{\infty} \beta^t \left[ \chi \left( \hat{y}_t - \hat{y}_t^* \right)^2 \right] - \frac{U_y}{2} E_0 \sum_{t=0}^{\infty} \beta^t \left[ \frac{\epsilon}{\kappa} \left( 1 + \frac{U_b}{U_y} \Theta_\pi \right) \hat{\pi}_t^2 \right] \\
&\quad - U_b E_0 \sum_{t=0}^{\infty} \beta^t \left( B_{bb} \frac{\hat{b}_t^2}{2} - B_{bz} \hat{b}_t \hat{z}_t - B_{bv} \hat{b}_t \hat{v}_{ft} + B_{bg} \hat{b}_t \hat{g}_t \right) \\
&\approx -\frac{U_y}{2} E_0 \sum_{t=0}^{\infty} \beta^t \left[ \chi \left( \hat{y}_t - \hat{y}_t^* \right)^2 \right] - \frac{U_y}{2} E_0 \sum_{t=0}^{\infty} \beta^t \left[ \frac{\epsilon}{\kappa} \left( 1 + \frac{U_b}{U_y} \Theta_\pi \right) \hat{\pi}_t^2 \right] \\
&\quad - \frac{U_b}{2} E_0 \sum_{t=0}^{\infty} \beta^t B_{bb} \left( \hat{b}_t - \hat{b}_t^* \right)
\end{aligned} \tag{A.47}$$

where the first equality follows from the fact that $U_b \Upsilon_b^{-1} \frac{\mathcal{W}_0 - \mathcal{W}}{\mathcal{W}}$ does not depend on policy; the second from the standard result $E_0 \sum_{t=0}^{\infty} \beta^t \hat{\Delta}_t \approx \frac{\epsilon}{2\kappa} E_0 \sum_{t=0}^{\infty} \beta^t \hat{\pi}_t^2 + t.i.p. + h.o.t.$ (see Woodford, 2003), the first order approximation $\hat{x}_t = \omega^{-1} \left( \hat{z}_t - g_y \hat{g}_t \right)$ (second order terms, like price dispersion, can be dropped since $\hat{x}_t$ enters only while multiplied by $\hat{b}_t$) and simple substitution of terms, with

$\Theta_\pi \equiv \Upsilon_b^{-1}\Upsilon_\pi$ and $\Theta_{bi} \equiv \Upsilon_b^{-1}\Upsilon_{bi}$ for $i = b, z, v, g$; the third from grouping similar terms and letting

$$B_{bb} \equiv (1-\vartheta-\Theta_{bb}), \qquad B_{bv} \equiv \Theta_{bv}, \tag{A.48}$$

$$B_{bz} \equiv \left[(1-\vartheta)\,\omega^{-1} + \Theta_{bz}\right], \quad B_{bg} \equiv \left[(1-\vartheta)\,\omega^{-1} g_y + \Theta_{bg}\right]; \tag{A.49}$$

and the fourth from defining the welfare-relevant debt target

$$\hat{b}_t^* = M_z \hat{z}_t - M_g \hat{g}_t + M_v \hat{v}_{f,t}, \tag{A.50}$$

with

$$M_z = \frac{B_{bz}}{B_{bb}}, \qquad M_g = \frac{B_{bg}}{B_{bb}}, \qquad M_v = \frac{B_{bv}}{B_{bb}}. \tag{A.51}$$

Finally, starting from the approximation in (A.47) and defining the welfare weights

$$\alpha_y \equiv \chi, \qquad \alpha_\pi \equiv \left(1 + \frac{U_b}{U_y}\Theta_\pi\right)\frac{\epsilon}{\kappa}, \qquad \alpha_b \equiv \frac{U_b B_{bb}}{U_y}, \tag{A.52}$$

the maximization of the representative household's lifetime utility $\mathcal{U}_0$ is equivalent to the minimization of the following intertemporal loss function

$$\mathcal{L}_0 \equiv \frac{1}{2} E_0 \sum_{t=0}^{\infty} \beta^t \left[\alpha_y\left(\hat{y}_t - \hat{y}_t^*\right)^2 + \alpha_\pi \hat{\pi}_t^2 + \alpha_b(\hat{b}_t - \hat{b}_t^*)^2\right] \tag{A.53}$$

Consider the definition of $\alpha_\pi$ in (A.52):

$$\alpha_\pi \equiv \left(1 + \frac{U_b}{U_y}\Theta_\pi\right)\frac{\epsilon}{\kappa} = \left(1 + \frac{\xi\vartheta}{1+\xi\vartheta}\frac{\Upsilon_x}{\Upsilon_b}\right)\frac{\epsilon}{\kappa}$$

where the second equality follows from $U_y \equiv \frac{1+\xi\vartheta}{\omega}$, $U_b \equiv \xi\vartheta$ and $\Upsilon_\pi \equiv \omega^{-1}\Upsilon_x$. Let's focus on the ratio $\frac{\Upsilon_x}{\Upsilon_b}$:

$$\frac{\Upsilon_x}{\Upsilon_b} = \frac{\Gamma_x - (1-\tilde{\beta})}{\Gamma_b + (1-\tilde{\beta})\phi_b} = \frac{-\varkappa\vartheta - (1-\tilde{\beta})}{\varkappa\vartheta + (1-\tilde{\beta})\phi_b} < 0 \tag{A.54}$$

where the first equality makes use of the definitions of $\Upsilon_x$ and $\Upsilon_b$ in (A.34) and (A.35), and the second follows from $\Gamma_x = -\Gamma_b = -\varkappa\vartheta$. $\frac{\Upsilon_x}{\Upsilon_b} < 0$. With $\frac{\xi\vartheta}{1+\xi\vartheta} \in (0,1)$, a sufficient condition for $\alpha_\pi$ to be positive is that $\frac{\Upsilon_x}{\Upsilon_b} > -1$. Using the expression in (A.54), simple algebra shows that this last inequality holds if and only if $\phi_b > 1$ (which is guaranteed by Assumption 2).

Next, consider the weight $\alpha_b$, also defined in (A.52):

$$\alpha_b \equiv \frac{U_b B_{bb}}{U_y} = \omega \frac{\xi\vartheta}{1+\xi\vartheta} B_{bb},$$

where the second equality makes use of the definitions of $U_y$ and $U_b$ again. It then follows that $\text{sign}(\alpha_b) = \text{sign}(B_{bb})$, where $B_{bb} \equiv 1 - \vartheta - \Theta_{bb}$, as defined in (A.48). Consider the term $\Theta_{bb}$, as defined in Appendix A.1.3:

$$\Theta_{bb} \equiv \frac{\Upsilon_{bb}}{\Upsilon_b} = \frac{\Gamma_b(\Gamma_b + Q_b) + (1-\tilde{\beta})\phi_b^2}{\Gamma_b + (1-\tilde{\beta})\phi_b}$$

where the second equality follows from the definitions of $\Upsilon_{bb}$ and $\Upsilon_b$ in, respectively, (A.37) and (A.35). This term is always larger than unity since $\Gamma_b > 0$ (see definition in (A.31)), $Q_b > 1$ (see definition in (A.32)) and $\phi_b > 1$ by Assumption 2. With $1 - \vartheta \in (0,1)$, we can then conclude that $B_{bb} < 0$, which in turn gives $\alpha_b < 0$.

### A.1.4 Second Order Conditions for Loss Minimum (STEP 4)

To guarantee that the solution to the first order conditions associated with the minimization of (A.53) indeed attain a loss minimum, we need to verify that the temporary loss in (A.53) is strictly convex. This is always the case in the baseline New Keynesian model without temptation where all welfare weights are positive - i.e. $\alpha_y, \alpha_\pi > 0$ - but might not apply here since $\alpha_b$ is negative. To deal with this issue, we follow Benigno and Woodford (2005, 2006) and make use of results by Telser and Graves (1972) about necessary and sufficient conditions for constrained maximization in (deterministic) dynamic linear-quadratic problems. We prove that our objective is indeed strictly convex for any parameterization of the model.

From Proposition 3 in Appendix A of Benigno and Woodford (2005), we know that this is equivalent to verifying the second order conditions for optimality in the associated deterministic problem:

$$\min_{\{\hat{y}_t-\hat{y}_t^*,\hat{\pi}_t,\hat{b}_t-\hat{b}_t^*\}} \frac{1}{2}\sum_{t=0}^{\infty}\beta^t\left[\alpha_y\left(\hat{y}_t-\hat{y}_t^*\right)^2+\alpha_\pi\hat{\pi}_t^2+\alpha_b(\hat{b}_t-\hat{b}_t^*)^2\right] \tag{A.55}$$

for bounded deterministic sequences $\left\{\hat{y}_t-\hat{y}_t^*,\hat{\pi}_t,\hat{b}_t-\hat{b}_t^*\right\}$ - in the sense that $\sum_{t=0}^{\infty}\beta^t v_t^2<\infty$ for $v=\hat{y}_t-\hat{y}_t^*$, $\hat{\pi}_t$, $\hat{b}_t-\hat{b}_t^*$ - satisfying

$$\tilde{\beta}\hat{\pi}_t = \hat{\pi}_{t-1}-\kappa\chi\left(\hat{y}_{t-1}-\hat{y}_{t-1}^*\right), \tag{A.56}$$

$$\hat{b}_t-\hat{b}_t^* = \rho_b\left(1+\gamma_b\right)\left(\hat{b}_{t-1}-\hat{b}_{t-1}^*\right), \tag{A.57}$$

for given initial conditions of predetermined variables.[51] The linear-quadratic problem in (A.55)-(A.57) is in the same form of the general problem studied by Telser and Graves (1972). Using a notation similar to theirs, let $\hat{x}_t\equiv\left[\hat{y}_t-\hat{y}_t^*,\hat{\pi}_t,\hat{b}_t-\hat{b}_t^*,\right]'$, such that the problem (A.55)-(A.57) can be written as follows:[52]

$$\min_{\{\hat{y}_t-\hat{y}_t^*,\hat{\pi}_t,\hat{b}_t-\hat{b}_t^*\}} \frac{1}{2}\sum_{t=0}^{\infty}\beta^t\hat{x}_t'B\hat{x}_t, \qquad \text{for} \qquad B\equiv\operatorname{diag}\left(\alpha_y,\alpha_\pi,\alpha_b\right) \tag{A.58}$$

subject to

$$A\left(L\right)\hat{x}_t=0, \qquad \text{for} \qquad A\left(L\right)\equiv\begin{bmatrix}-\kappa\chi L & L-\tilde{\beta} & 0\\ 0 & 0 & 1-\rho_b\left(1+\gamma_b\right)L\end{bmatrix}. \tag{A.59}$$

[51] Equation (A.56) is a one-period lagged deterministic version of the Phillips curve. Equation (A.57) is derived from substituting a deterministic version of the Euler equation, lagged by one period, into the law of motion of public debt.

[52] There is a slight abuse of notation here since $x$ has been previously used to defined the consumption-labor composite entering the GHH utility specification in the main text, and is also used in Appendix A.2 to denote the vector of predetermined and non-predetermined variables in the computational algorithm for optimal policy. The definition $\hat{x}_t\equiv\left[\hat{y}_t-\hat{y}_t^*,\hat{\pi}_t,\hat{b}_t-\hat{b}_t^*,\right]'$ pertains only to this section.

with $L$ denoting the lag operator. Second order conditions hold if and only if the quadratic form $\hat{x}_t' B \hat{x}_t$ in (A.58) is positive definite for all bounded sequences $\{\hat{x}_t\}$ satisfying constraint (A.59). Define the bordered Hessian matrix

$$H(\theta) \equiv \begin{bmatrix} 0 & A(\tilde{\beta}^{\frac{1}{2}} e^{-i\theta}) \\ A'(\tilde{\beta}^{\frac{1}{2}} e^{-i\theta}) & B \end{bmatrix},$$

where

$$\begin{aligned} A(\tilde{\beta}^{\frac{1}{2}} e^{-i\theta}) &= \begin{bmatrix} -\kappa\chi\tilde{\beta}^{\frac{1}{2}} e^{-i\theta} & \tilde{\beta}^{\frac{1}{2}} e^{-i\theta} - \tilde{\beta} & 0 \\ 0 & 0 & 1 - \rho_b (1+\gamma_b) \tilde{\beta}^{\frac{1}{2}} e^{-i\theta} \end{bmatrix}, \\ A'(\tilde{\beta}^{\frac{1}{2}} e^{i\theta}) &= \begin{bmatrix} -\kappa\chi\tilde{\beta}^{\frac{1}{2}} e^{i\theta} & 0 \\ \tilde{\beta}^{\frac{1}{2}} e^{i\theta} - \tilde{\beta} & 0 \\ 0 & 1 - \rho_b (1+\gamma_b) \tilde{\beta}^{\frac{1}{2}} e^{i\theta} \end{bmatrix}. \end{aligned}$$

Since $B$ is a diagonal non-singular matrix (provided all welfare weights are different from zero) and $A'(\tilde{\beta}^{\frac{1}{2}} e^{i\theta})$ has rank 2 (hence equal to the number of linear constraints), Theorems 5.1-5.3 in Ch. 2 of Telser and Graves (1972) imply that the positive definiteness requirement is satisfied (hence, second order conditions hold) if and only if all northwest principal minors of $H(\theta)$ of order $p > 4$ (where 4 is the product of 2 and the number of linear constraints) has the same sign of $(-1)^{\#\text{constraints}}$ for all $|\theta| \leq \pi$.[53] It follows that we only need to verify that the determinant of the bordered Hessian $H(\theta)$ is strictly positive for all $|\theta| \leq \pi$. Writing $H(\theta)$ more extensively as

$$H(\theta) = \begin{bmatrix} 0 & 0 & -\kappa\chi\tilde{\beta}^{\frac{1}{2}} e^{-i\theta} & \tilde{\beta}^{\frac{1}{2}} e^{-i\theta} - \tilde{\beta} & 0 \\ 0 & 0 & 0 & 0 & 1 - \rho_b (1+\gamma_b) \tilde{\beta}^{\frac{1}{2}} e^{-i\theta} \\ -\kappa\chi\tilde{\beta}^{\frac{1}{2}} e^{i\theta} & 0 & \alpha_y & 0 & 0 \\ \tilde{\beta}^{\frac{1}{2}} e^{i\theta} - \tilde{\beta} & 0 & 0 & \alpha_\pi & 0 \\ 0 & 1 - \rho_b (1+\gamma_b) \tilde{\beta}^{\frac{1}{2}} e^{i\theta} & 0 & 0 & \alpha_b \end{bmatrix},$$

[53] Obvsiouly $\pi$ here stands for the numerical "greek pi" and not inflation.

the Laplace expansion gives the following result:

$$
\begin{aligned}
& \det\left(H(\theta)\right) \\
= \; & -(1-\rho_b\left(1+\gamma_b\right)\tilde{\beta}^{\frac{1}{2}}e^{i\theta}) * \\
& \det\left(\begin{bmatrix} 0 & -\kappa\chi\tilde{\beta}^{\frac{1}{2}}e^{-i\theta} & \tilde{\beta}^{\frac{1}{2}}e^{-i\theta}-\tilde{\beta} & 0 \\ 0 & 0 & 0 & 1-\rho_b\left(1+\gamma_b\right)\tilde{\beta}^{\frac{1}{2}}e^{-i\theta} \\ -\kappa\chi\tilde{\beta}^{\frac{1}{2}}e^{i\theta} & \alpha_y & 0 & 0 \\ \tilde{\beta}^{\frac{1}{2}}e^{i\theta}-\tilde{\beta} & 0 & \alpha_\pi & 0 \end{bmatrix}\right) \\
= \; & -\left[1-\rho_b\left(1+\gamma_b\right)\tilde{\beta}^{\frac{1}{2}}e^{i\theta}\right]\left[1-\rho_b\left(1+\gamma_b\right)\tilde{\beta}^{\frac{1}{2}}e^{-i\theta}\right]\det\left(\begin{bmatrix} 0 & -\kappa\chi\tilde{\beta}^{\frac{1}{2}}e^{-i\theta} & \tilde{\beta}^{\frac{1}{2}}e^{-i\theta}-\tilde{\beta} \\ -\kappa\chi\tilde{\beta}^{\frac{1}{2}}e^{i\theta} & \alpha_y & 0 \\ \tilde{\beta}^{\frac{1}{2}}e^{i\theta}-\tilde{\beta} & 0 & \alpha_\pi \end{bmatrix}\right) \\
= \; & \left[1-\rho_b\left(1+\gamma_b\right)\tilde{\beta}^{\frac{1}{2}}e^{i\theta}\right]\left[1-\rho_b\left(1+\gamma_b\right)\tilde{\beta}^{\frac{1}{2}}e^{-i\theta}\right]\left[\kappa^2\chi^2\tilde{\beta}\alpha_\pi+\alpha_y(\tilde{\beta}^{\frac{1}{2}}e^{i\theta}-\tilde{\beta})(\tilde{\beta}^{\frac{1}{2}}e^{-i\theta}-\tilde{\beta})\right] \\
= \; & \underbrace{\left\{1+\left[\rho_b\left(1+\gamma_b\right)\right]^2\tilde{\beta}-2\rho_b\left(1+\gamma_b\right)\tilde{\beta}^{\frac{1}{2}}\cos\theta\right\}}_{H_1(\theta)}\underbrace{\left(\kappa^2\chi^2\tilde{\beta}\alpha_\pi+\alpha_y\tilde{\beta}+\alpha_y\tilde{\beta}^2-2\alpha_y\tilde{\beta}^{\frac{3}{2}}\cos\theta\right)}_{H_2(\theta)} \qquad \text{(A.60)}
\end{aligned}
$$

where (A.60) relies on the equality $\cos\theta = \frac{e^{i\theta}+e^{-i\theta}}{2}$.

First of all, notice that $\det(H(\theta))$ is not affected by the sign of the welfare weight $\alpha_b$. Hence, the latter does not play any role in whether the bordered Hessian $H(\theta)$ is positive definite or negative definite, and hence in establishing the validity of second order conditions for optimality. Then, consider factor $H_1(\theta)$ in (A.60). This expression is clearly strictly decreasing in $\cos\theta$, whose highest value is $\cos\theta = 1$ for $\theta = 0$. It is therefore sufficient to show that $H_1(0) > 0$. Simple algebra gives

$$
\begin{aligned}
H_1(0) &= 1+\left[\rho_b\left(1+\gamma_b\right)\right]^2\tilde{\beta}-2\rho_b\left(1+\gamma_b\right)\tilde{\beta}^{\frac{1}{2}} \\
&= \left[1-\rho_b\left(1+\gamma_b\right)\tilde{\beta}^{\frac{1}{2}}\right]^2 > 0
\end{aligned}
$$

since $\gamma_b \in (-1, 0)$. Next, consider factor $H_2(\theta)$:

$$H_2(\theta) = \kappa^2 \chi^2 \tilde{\beta} \alpha_\pi + \alpha_y \tilde{\beta} \left( 1 + \tilde{\beta} - 2\tilde{\beta}^{\frac{1}{2}} \cos\theta \right) \tag{A.61}$$

Clearly, the expression within brackets in (A.61) is also strictly decreasing in $\cos\theta$. Hence, by the same logic used for $H_1(\theta)$ and given that $\kappa^2 \chi^2 \tilde{\beta} \alpha_\pi > 0$, we just need to verify that $1 + \tilde{\beta} - 2\tilde{\beta}^{\frac{1}{2}} > 0$. This is indeed the case for any $\tilde{\beta} \in (0, 1)$. We can therefore conclude that $\det(H(\theta)) > 0$ for all $-\pi \leq \theta \leq \pi$. As a result, the second order conditions always hold for any parameterization of the model.

## A.2 Computation of Optimal Monetary Policy

The computation of optimal monetary policy - both under discretion and commitment - closely follows the linear-quadratic approach of Soderlind (1999). The first step requires writing the equilibrium system (38)-(40) in the appropriate state-space form. Starting with equation (40), the law of motion of debt (in deviation from its target) is:

$$\begin{aligned} \hat{b}_t - \hat{b}_t^* &= \rho_b \left( \hat{b}_{t-1} - \hat{b}_{t-1}^* + \hat{R}_{t-1} - \hat{\pi}_t \right) - (M_v + \rho_v)\hat{v}_{f,t} - M_z \hat{z}_t + M_g \hat{g}_t \\ &\quad + \rho_b M_z \hat{z}_{t-1} - \rho_b M_g \hat{g}_{t-1} + \rho_b M_v \hat{v}_{f,t-1} \end{aligned} \tag{A.62}$$

From the Phillips curve (38), we obtain the law of motion of expected inflation:

$$E_t \hat{\pi}_{t+1} = \frac{1}{\tilde{\beta}} \hat{\pi}_t - \frac{\kappa\chi}{\tilde{\beta}} (\hat{y}_t - \hat{y}_t^*) - \frac{1}{\tilde{\beta}} \hat{e}_t \tag{A.63}$$

From the Euler equation (39), after substituting in the expressions for the debt gap in (A.62) and for expected inflation in (A.63), and grouping similar terms, we obtain:

$$\begin{aligned} \hat{R}_t &= \left( \frac{1}{\tilde{\beta}} - \gamma_b \rho_b \right) \hat{\pi}_t - \frac{\kappa\chi}{\tilde{\beta}} (\hat{y}_t - \hat{y}_t^*) + \gamma_b \rho_b \left( \hat{b}_{t-1} - \hat{b}_{t-1}^* + \hat{R}_{t-1} \right) \\ &\quad + H^z \hat{z}_t - H^g \hat{g}_t + H^v \hat{v}_{f,t} + H_1^z \hat{z}_{t-1} - H_1^g \hat{g}_{t-1} + H_1^v \hat{v}_{f,t-1} - \frac{1}{\tilde{\beta}} \hat{e}_t \end{aligned} \tag{A.64}$$

where $H^i \equiv H_i - \gamma M_i$, for $i = z, g$, $H^v \equiv H_v - \gamma\left(M_v + \rho_v\right)$, and $H_1^i \equiv \gamma \rho_b M_i$ for $i = z, g, v$. To the newly defined state-space system, we append the stochastic laws of motion for the exogenous shocks:

$$\hat{e}_{t+1} = \rho_e \hat{e}_t + \hat{\varepsilon}_{et+1}, \tag{A.65}$$

$$\hat{z}_{t+1} = \rho_z \hat{z}_t + \hat{\varepsilon}_{zt+1}, \tag{A.66}$$

$$\hat{g}_{t+1} = \rho_g \hat{g}_t + \hat{\varepsilon}_{gt+1}, \tag{A.67}$$

$$\hat{v}_{ft+1} = \rho_f \hat{v}_{ft} + \hat{\varepsilon}_{ft+1}. \tag{A.68}$$

Letting $\hat{n}_t \equiv \left(\hat{e}_t, \hat{z}_t, \hat{g}_t, \hat{v}_{f,t}\right)'$, we define the following vectors:

$$\hat{x}_t \equiv \begin{bmatrix} \hat{x}_{1,t} \\ \hat{x}_{2,t} \end{bmatrix}, \qquad \hat{x}_{1,t} \equiv \left[\hat{e}_t, \hat{z}_t, \hat{g}_t, \hat{v}_{ft}, \hat{z}_{t-1}, \hat{g}_{t-1}, \hat{v}_{ft-1,}\hat{b}_{t-1} - \hat{b}^*_{t-1}, \hat{R}_{t-1}\right]', \tag{A.69}$$

$$\hat{x}_{2,t} \equiv \hat{\pi}_t, \qquad \hat{u}_t \equiv \hat{y}_t - \hat{y}^*_t, \qquad \hat{\varepsilon}_{t+1} \equiv \left[\hat{\varepsilon}_{et+1}, \hat{\varepsilon}_{z,t+1}, \hat{\varepsilon}_{g,t+1}, \hat{\varepsilon}_{v,t+1}, 0_{1\text{x}6}\right]', \tag{A.70}$$

where $\hat{x}_{1t}$ includes all exogenous shocks and predetermined variables, $\hat{x}_{2t}$ includes all non-predetermined variables (in our case, only inflation), $\hat{u}_t$ includes the instrument or control variables (in our case, the the output gap), and $\hat{\varepsilon}_{t+1}$ gathers the *iid* disturbances to the exogenous shocks (with $0_{m\text{x}n}$ denoting a $m$-row/$n$-column matrix of zeros).[54] Using this notation, the state-space system is

$$\begin{bmatrix} \hat{x}_{1,t+1} \\ E_t \hat{x}_{2,t+1} \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} \hat{x}_{1,t} \\ \hat{x}_{2,t} \end{bmatrix} + B\hat{u}_t + \hat{\varepsilon}_{t+1}, \tag{A.71}$$

[54] The lack of output gap expectations in the equilibrium system is the joint consequence of a GHH utlity and an efficient labor subsidy. As it only enters as a period $t$ dated variable, the output gap is the natural choice as control variable for the linear-quadratic optimization problem.

where

$$A_{11} \equiv \begin{bmatrix} \rho_e & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \rho_z & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \rho_g & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \rho_f & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -M_z & M_g & -(M_v+\rho_v) & \rho_b M_z & -\rho_b M_g & \rho_b M_v & \rho_b & \rho_b \\ -\frac{1}{\tilde{\beta}} & H^z & -H^g & H^v & H_1^z & -H_1^g & H_1^v & \gamma_b\rho_b & \gamma_b\rho_b \end{bmatrix},$$

$$A_{12} \equiv \left[0_{7\text{x}1}, -\rho_b, \frac{1}{\tilde{\beta}} - \gamma_b\rho_b\right]', \quad A_{21} \equiv \left[-\frac{1}{\tilde{\beta}}, 0_{1\text{x}8}\right], \quad A_{22} \equiv \left[\frac{1}{\tilde{\beta}}\right]$$

$$B \equiv \left[0_{1\text{x}8}, -\frac{\kappa\chi}{\tilde{\beta}}, -\frac{\kappa\chi}{\tilde{\beta}}\right]'.$$

Looking back at the system (A.71), the first four equations correspond to the exogenous processes (A.65)-(A.68); the next three are simple identities $\hat{z}_t = \hat{z}_t, \hat{g}_t = \hat{g}_t$ and $\hat{v}_{f,t} = \hat{v}_{f,t}$ to allow the recursive inclusion of lagged shocks in the system; the next three are, respectively, the law of motion of debt (A.62), the Euler equation (A.64), and the Phillips curve (A.63).

Next, consider the temporary welfare-based loss function $L_t$ derived in Proposition 3:

$$\begin{aligned} L_t &\equiv \frac{1}{2}\left[\alpha_y\left(\hat{y}_t - \hat{y}_t^*\right)^2 + \alpha_\pi\hat{\pi}_t^2 + \alpha_b(\hat{b}_t - \hat{b}_t^*)^2\right] \\ &= \frac{1}{2}\begin{bmatrix}\hat{b}_t - \hat{b}_t^* & \hat{\pi}_t & \hat{y}_t - \hat{y}_t^*\end{bmatrix}\begin{bmatrix}\alpha_b & 0 & 0 \\ 0 & \alpha_\pi & 0 \\ 0 & 0 & \alpha_y\end{bmatrix}\begin{bmatrix}\hat{b}_t - \hat{b}_t^* \\ \hat{\pi}_t \\ \hat{y}_t - \hat{y}_t^*\end{bmatrix} \\ &= \frac{1}{2}\begin{bmatrix}\hat{x}_t' & \hat{u}_t'\end{bmatrix} K' \begin{bmatrix}\alpha_b & 0 & 0 \\ 0 & \alpha_\pi & 0 \\ 0 & 0 & \alpha_y\end{bmatrix} K \begin{bmatrix}\hat{x}_t \\ \hat{u}_t\end{bmatrix} \end{aligned} \quad \text{(A.72)}$$

where the third equality follows from

$$\begin{bmatrix} \hat{b}_t - \hat{b}_t^* \\ \hat{\pi}_t \\ \hat{y}_t - \hat{y}_t^* \end{bmatrix} = K \begin{bmatrix} \hat{x}_t \\ \hat{u}_t \end{bmatrix}, \tag{A.73}$$

for

$$K \equiv \begin{bmatrix} 0 & -M_z & M_g & -(M_v + \rho_v) & \rho_b M_z & -\rho_b M_g & \rho_b M_v & \rho_b & \rho_b & -\rho_b & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{A.74}$$

Define the 11x11 matrix $W$,

$$W \equiv K^{'} \begin{bmatrix} \alpha_b & 0 & 0 \\ 0 & \alpha_\pi & 0 \\ 0 & 0 & \alpha_y \end{bmatrix} K = \begin{bmatrix} Q & U \\ U^{'} & R \end{bmatrix} \tag{A.75}$$

for $U \equiv [0_{1\text{x}10}]'$, $R \equiv \begin{bmatrix} \alpha_y \end{bmatrix}$ and an appropriately defined 10x10 matrix $Q$. After combining the expressions in (A.73)-(A.75) with (A.72), we rewrite the temporary loss function $L_t$ in matrix form:

$$\begin{aligned} L_t &= \frac{1}{2} \begin{bmatrix} \hat{x}_t^{'} & \hat{u}_t^{'} \end{bmatrix} \begin{bmatrix} Q & U \\ U^{'} & R \end{bmatrix} \begin{bmatrix} \hat{x}_t \\ \hat{u}_t \end{bmatrix} \\ &= \frac{1}{2} \begin{bmatrix} \hat{x}_{1,t}^{'} & \hat{x}_{2,t}^{'} \end{bmatrix} Q \begin{bmatrix} \hat{x}_{1,t} \\ \hat{x}_{2,t} \end{bmatrix} + \frac{1}{2} \hat{u}_t^{'} R \hat{u}_t \end{aligned} \tag{A.76}$$

Given (A.76), the Bellman equation associated with the optimal monetary policy problem is:

$$J_t = \frac{1}{2} \hat{x}_t^{'} Q \hat{x}_t + \frac{1}{2} \hat{u}_t^{'} R \hat{u}_t + \beta \frac{1}{2} E_t J_{t+1} \tag{A.77}$$

Since the model is linear-quadratic, we conjecture a value function $J_t$ which is quadratic in the predetermined variables:

$$J_t = \frac{1}{2} \hat{x}_{1,t}^{'} V_t \hat{x}_{1,t} + \frac{1}{2} \nu_t \tag{A.78}$$

where $V_t$ is a matrix (whose entries are unknown) and $\nu_t$ is a scalar (also unknown). We also conjecture a linear relationship between non-predetermined and predetermined variables in $t+1$:

$$\hat{x}_{2,t+1} = C_{t+1}\hat{x}_{1,t+1} \tag{A.79}$$

with $C_{t+1}$ also an unknown matrix.[55] In our specific case, the guessed solution (A.79) includes only inflation and takes the following form:

$$\begin{aligned} \hat{\pi}_{t+1} \;\; = \;\; & C_{e,t+1}\hat{e}_{t+1} + C_{z,t+1}\hat{z}_{t+1} + C_{g,t+1}\hat{g}_{t+1} + C_{v,t+1}\hat{v}_{f,t+1} + C^1_{z,t+1}\hat{z}_t + C^1_{g,t+1}\hat{g}_t \\ & + C^1_{v,t+1}\hat{v}_{f,t} + C^1_{b,t+1}\left(\hat{b}_t - \hat{b}^*_t\right) + C^1_{R,t+1}\hat{R}_t \end{aligned} \tag{A.80}$$

After inserting (A.79) into (A.71) and taking expectations, we obtain:

$$\begin{aligned} \begin{bmatrix} I \\ C_{t+1} \end{bmatrix} E_t\hat{x}_{1,t+1} \;\; &= \;\; \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} \hat{x}_{1,t} \\ \hat{x}_{2,t} \end{bmatrix} + \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \hat{u}_t \\ &\Longrightarrow \begin{bmatrix} I & -A_{12} \\ C_{t+1} & -A_{22} \end{bmatrix} \begin{bmatrix} E_t\hat{x}_{1,t+1} \\ \hat{x}_{2,t} \end{bmatrix} = \begin{bmatrix} A_{11} \\ A_{21} \end{bmatrix} \hat{x}_{1,t} + \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \hat{u}_t \\ &\Longrightarrow \begin{bmatrix} E_t\hat{x}_{1,t+1} \\ \hat{x}_{2,t} \end{bmatrix} = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} \left[ \begin{bmatrix} A_{11} \\ A_{21} \end{bmatrix} \hat{x}_{1,t} + \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \hat{u}_t \right] \end{aligned} \tag{A.81}$$

where the second line follows from moving $\hat{x}_{2,t}$ to the left-hand-side, and the third line from a simple matrix inversion with

$$\begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} = \begin{bmatrix} I & -A_{12} \\ C_{t+1} & -A_{22} \end{bmatrix}^{-1}$$

$$P_{21} = (A_{22} - C_{t+1}A_{12})^{-1}C_{t+1} \tag{A.82}$$

$$P_{22} = -(A_{22} - C_{t+1}A_{12})^{-1} \tag{A.83}$$

[55] The time subscript $t$ for matrices $V_t$ and $C_t$ is due to the iterative recursive algorithm used to compute the time-invariant matrices $V$ and $C$ as their fixed points.

Making use of (A.82)-(A.83), from (A.81) we obtain

$$\hat{x}_{2,t} = D_t \hat{x}_{1,t} + G_t \hat{u}_t \tag{A.84}$$

where

$$D_t \equiv (A_{22} - C_{t+1}A_{12})^{-1} C_{t+1} A_{11} - (A_{22} - C_{t+1}A_{12})^{-1} A_{21} \tag{A.85}$$

$$G_t \equiv (A_{22} - C_{t+1}A_{12})^{-1} C_{t+1} B_1 - (A_{22} - C_{t+1}A_{12})^{-1} B_2 \tag{A.86}$$

Plugging (A.84) back into the upper-block of the state space system (A.71) yields

$$\hat{x}_{1,t+1} = A_t^* \hat{x}_{1,t} + B_t^* \hat{u}_t + \varepsilon_{t+1} \tag{A.87}$$

where

$$A_t^* \equiv A_{11} + A_{12} D_t \tag{A.88}$$

$$B_t^* \equiv B_1 + A_{12} G_t \tag{A.89}$$

After partitioning

$$Q = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix},$$

we substitute (A.84) into the temporary loss (A.76):

$$\begin{aligned} L_t &= \frac{1}{2} \begin{bmatrix} \hat{x}_{1,t} \\ D_t \hat{x}_{1,t} + G_t \hat{u}_t \end{bmatrix}' \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} \begin{bmatrix} \hat{x}_{1,t} \\ D_t \hat{x}_{1,t} + G_t \hat{u}_t \end{bmatrix} + \frac{1}{2} \hat{u}_t' R \hat{u}_t \\ &= \frac{1}{2} \hat{x}_{1,t}' Q_t^* \hat{x}_{1,t} + \frac{1}{2} \hat{u}_t' R_t^* \hat{u}_t + \hat{x}_{1,t}' U_t^* \hat{u}_t \end{aligned} \tag{A.90}$$

where the second equality follows from simple matrix algebra and the following definitions

$$Q_t^* \equiv Q_{11} + Q_{12}D_t + D_t^{'}Q_{21} + D_t^{'}Q_{22}D_t \tag{A.91}$$

$$R_t^* \equiv R + G_t^{'}Q_{22}G_t \tag{A.92}$$

$$U_t^* \equiv Q_{12}G_t + D_t^{'}Q_{22}G_t \tag{A.93}$$

After plugging the loss function equation (A.90), the value function guess (A.78) updated to $t+1$, and the law of motion for $\hat{x}_{1,t+1}^{'}$ (A.87) into the Bellman equation (A.77), the optimal monetary policy problem is

$$\begin{aligned}\min_{\hat{u}_t} J_t = & \frac{1}{2}\hat{x}_{1,t}^{'}Q_t^*\hat{x}_{1,t} + \frac{1}{2}\hat{u}_t^{'}R_t^*\hat{u}_t + \hat{x}_{1,t}^{'}U_t^*\hat{u}_t \\ & + \frac{\beta}{2}E_t\left[(A_t^*\hat{x}_{1,t} + B_t^*\hat{u}_t + \varepsilon_{t+1})^{'}V_{t+1}(A_t^*\hat{x}_{1,t} + B_t^*\hat{u}_t + \varepsilon_{t+1}) + \nu_{t+1}\right]\end{aligned} \tag{A.94}$$

Since the vector $\hat{x}_{1,t}$ includes only exogenous terms and predetermined endogenous variables, taking expectations, the problem in (A.94) is equivalent to the following:

$$\min_{\hat{u}_t} \check{J}_t = \frac{1}{2}\hat{u}_t^{'}R_t^*\hat{u}_t + \hat{x}_{1,t}^{'}U_t^*\hat{u}_t + \frac{\beta}{2}\hat{u}_t^{\prime}\left(B_t^*\right)^{\prime}V_{t+1}B_t^*\hat{u}_t + \beta\hat{x}_{1,t}^{\prime}\left(A_t^*\right)^{\prime}V_{t+1}B_t^*\hat{u}_t \tag{A.95}$$

Provided both $R_t^*$ and $V_{t+1}$ are symmetric, the first order condition with respect to $\hat{u}_t$ yields:

$$\hat{u}_t = -F_t\hat{x}_{1,t} \tag{A.96}$$

for

$$F_t \equiv \left[R_t^* + \beta(B_t^*)^{'}V_{t+1}B_t^*\right]^{-1}\left[(U_t^*)^{'} + \beta(B_t^*)^{'}V_{t+1}A_t^*\right] \tag{A.97}$$

Equation (A.96) corresponds to the optimal instrument $\hat{u}_t$ (the output gap) as linear function of the vector of state variables $\hat{x}_{1,t}$. Substituting the expression in (A.96) into (A.84), we find the linear relationship between the endogenous non-predetermined (e.g. inflation) and the predetermined variables under the optimal plan:

$$\hat{x}_{2,t} = C_t \hat{x}_{1,t} \tag{A.98}$$

where

$$C_t \equiv D_t - G_t F_t \tag{A.99}$$

We substitute the optimal instrument rule (A.96) back into the value function (A.94) so that we can rewrite it in terms of endogenous predetermined variables and exogenous shocks only:

$$\begin{aligned} J_t^{opt} &= \frac{1}{2}\hat{x}_{1,t}^{'} Q_t^* \hat{x}_{1,t} + \frac{1}{2}\hat{x}_{1,t}^{'} F_t^{'} R_t^* F_t \hat{x}_{1,t} - \hat{x}_{1,t}^{'} U_t^* F_t \hat{x}_{1,t} \\ &+ \frac{\beta}{2} E_t \left\{ \left[ (A_t^* - B_t^* F_t \hat{x}_{1,t}) + \varepsilon_{t+1} \right]^{'} V_{t+1} \left[ (A_t^* - B_t^* F_t) \hat{x}_{1,t} + \varepsilon_{t+1} \right] + \nu_{t+1} \right\} && \text{(A.100)} \\ &= \frac{1}{2}\hat{x}_{1,t}^{'} \left[ Q_t^* + F_t^{'} R_t^* F_t - 2U_t^* F_t + \beta (A_t^* - B_t^* F_t)^{'} V_{t+1} (A_t^* - B_t^* F_t) \right] \hat{x}_{1,t} && \text{(A.101)} \\ &+ \frac{\beta}{2} E_t \varepsilon_{t+1}^{'} V_{t+1} \varepsilon_{t+1} + \frac{\beta}{2} E_t \nu_{t+1} \end{aligned}$$

where the second equality follows from collecting similar terms. Noticing that

$$E_t \varepsilon_{t+1}^{'} \beta V_{t+1} \varepsilon_{t+1} = \beta \text{tr}(V_{t+1} \Sigma),$$

our initial conjecture $J_t = \frac{1}{2}\hat{x}_{1,t}^{'} V_t \hat{x}_{1,t} + \frac{1}{2}\nu_t$ implies the following recursive equations:

$$V_t = \left[ Q_t^* + F_t^{'} R_t^* F_t - 2U_t^* F_t + \beta (A_t^* - B_t^* F_t)^{'} V_{t+1} (A_t^* - B_t^* F_t) \right] \tag{A.102}$$

$$\nu_t = \beta \text{tr}(V_{t+1} \Sigma) + \beta E_t \nu_{t+1} \tag{A.103}$$

Starting with any initial guess for the scalar $v_{t+1}$, matrix $C_{t+1}$ and a symmetric and positive definite matrix guess for $V_{t+1}$, we use (A.99) and (A.102)-(A.103) to find the updated matrices $C_t$ and $V_t$ (and scalar $v_t$). We keep iterating until we reach convergence: $C_t \to C$, $V_t \to V$, as well as $F_t \to F$. Once convergence is attained, the rational expectations solution under the optimal policy is described by the following linear relationships:

$$\hat{u}_t = -F\hat{x}_{1,t}, \qquad \hat{x}_{2t} = C\hat{x}_{1,t} \tag{A.104}$$

$$\hat{x}_{1,t} = (A^* - B^*F)\,\hat{x}_{1,t-1} + \varepsilon_t \tag{A.105}$$

The optimized value function is then

$$J_t^{opt} = \hat{x}_{1,t}^{'} V \hat{x}_{1,t} + \frac{\beta}{1-\beta}\text{tr}(V\Sigma)$$

## A.3 Additional Derivations of Results

### A.3.1 Targeting Rule under Discretion

We start from the minimum state variable solution for inflation along the MPE:

$$\hat{\pi}_t = \pi_b\left(\hat{b}_{t-1} - \hat{b}^*_{t-1}\right) + \pi_R\hat{R}_{t-1} + \pi_n\hat{n}_t + \pi_l\hat{n}_{t-1}, \tag{A.106}$$

where $\hat{n}_t \equiv (\hat{e}_t, \hat{z}_t, \hat{g}_t, \hat{v}_{f,t})'$ follows the AR(1) process

$$\hat{n}_t = \Xi\hat{n}_{t-1} + \hat{\varepsilon}_t \tag{A.107}$$

with *iid* disturbances included in vector $\hat{\varepsilon}_t$ and diagonal matrix $\Xi$ containing autoregressive coefficients. We use (A.106)-(A.107) to compute the conditional expectation of one-period-ahead inflation, $E_t\hat{\pi}_{t+1}$. Leading (A.106) one period forward and applying the operator $E_t$, we find that

$$E_t\hat{\pi}_{t+1} = \pi_b\left(\hat{b}_t - \hat{b}^*_t\right) + \pi_R\hat{R}_t + (\pi_n\Xi + \pi_l)\,\hat{n}_t. \tag{A.108}$$

Substituting the latter into the Euler/Fisher equation (39), simple algebra gives:

$$\begin{aligned} \hat{R}_t &= R_b\left(\hat{b}_t - \hat{b}_t^*\right) + R_n\hat{n}_t \\ R_b &\equiv \frac{\pi_b + \gamma_b}{1-\pi_R}, \qquad R_n \equiv \frac{\pi_n\Xi + \pi_l + H}{1-\pi_R}, \qquad H \equiv [0, H_z, -H_g, H_v] \end{aligned} \tag{A.109}$$

After inserting (A.108)-(A.109) into the Phillips curve (38) - writing $\hat{e}_t = [1,0,0,0]\hat{n}_t$ - and rearranging terms, we obtain:

$$\begin{aligned} \hat{\pi}_t &= \kappa_b\left(\hat{b}_t - \hat{b}_t^*\right) + \kappa_y(\hat{y}_t - \hat{y}_t^*) + \kappa_n\hat{n}_t \\ \kappa_b &\equiv \tilde{\beta}\frac{\pi_b + \gamma_b\pi_R}{1-\pi_R}, \qquad \kappa_y \equiv \kappa\chi, \qquad \kappa_n \equiv \tilde{\beta}\frac{\pi_n\Xi + \pi_l + \pi_R H}{1-\pi_R} + [1,0,0,0] \end{aligned} \tag{A.110}$$

Next, we recall the law of motion for government debt (40), and write it as follows:

$$\begin{aligned} \hat{b}_t - \hat{b}_t^* &= \rho_b(\hat{b}_{t-1} - \hat{b}_{t-1}^* + \hat{R}_{t-1} - \hat{\pi}_t) + N_n\hat{n}_t + N_l\hat{n}_{t-1} \\ N_n &\equiv \left[-M_z, M_g, -\left(M_v + \rho_v\right)\right], \qquad N_l \equiv [\rho_b M_z, -\rho_b M_g, \rho_b M_v] \end{aligned} \tag{A.111}$$

Plugging the latter back into (A.110), we obtain a Phillips curve as function of the output gap and state variables (endogenous and exogenous):

$$\begin{aligned} \hat{\pi}_t &= \bar{\kappa}_b\left(\hat{b}_{t-1} - \hat{b}_{t-1}^* + \hat{R}_{t-1}\right) + \bar{\kappa}_y(\hat{y}_t - \hat{y}_t^*) + \bar{\kappa}_n\hat{n}_t + \bar{\kappa}_l\hat{n}_{t-1} \\ \bar{\kappa}_b &\equiv \frac{\kappa_b\rho_b}{1+\kappa_b\rho_b}, \qquad \bar{\kappa}_y \equiv \frac{\kappa_y}{1+\kappa_b\rho_b}, \qquad \bar{\kappa}_n \equiv \frac{\kappa_n + \kappa_b N_n}{1+\kappa_b\rho_b}, \qquad \bar{\kappa}_l \equiv \frac{\kappa_b N_l}{1+\kappa_b\rho_b} \end{aligned} \tag{A.112}$$

Then, we use the latter to eliminate inflation $\hat{\pi}_t$ from (A.111):

$$\begin{aligned} \hat{b}_t - \hat{b}_t^* &= \bar{\rho}_b(\hat{b}_{t-1} - \hat{b}_{t-1}^* + \hat{R}_{t-1}) - \rho_b\bar{\kappa}_y(\hat{y}_t - \hat{y}_t^*) + \bar{N}_n\hat{n}_t + \bar{N}_l\hat{n}_{t-1} \\ \bar{\rho}_b &\equiv \frac{\rho_b}{1+\kappa_b\rho_b}, \qquad \bar{N}_n \equiv N_n - \rho_b\bar{\kappa}_n, \qquad \bar{N}_l \equiv N_l - \rho_b\bar{\kappa}_l \end{aligned} \tag{A.113}$$

The dynamic programming problem the benevolent government solves under discretion consists then in the Bellman equation

$$V_t = \min \frac{1}{2}\left[\alpha_y\left(\hat{y}_t - \hat{y}_t^*\right)^2 + \alpha_\pi \hat{\pi}_t^2 + \alpha_b(\hat{b}_t - \hat{b}_t^*)^2\right] + \beta E_t V_{t+1}, \tag{A.114}$$

for $V_t \equiv V(\hat{b}_{t-1} - \hat{b}_{t-1}^*, \hat{R}_{t-1})$, subject to constraints (A.112)-(A.113). Making use of partial derivatives $\frac{\partial \hat{\pi}_t}{\partial(\hat{y}_t - \hat{y}_t^*)} = \bar{\kappa}_y$ (from (A.112)) and $\frac{\partial(\hat{b}_t - \hat{b}_t^*)}{\partial(\hat{y}_t - \hat{y}_t^*)} = -\rho_b \bar{\kappa}_y$ (from (A.113)), the first order condition with respect to the output gap $\hat{y}_t - \hat{y}_t^*$ yields:

$$\alpha_y\left(\hat{y}_t - \hat{y}_t^*\right) + \alpha_\pi \bar{\kappa}_y \hat{\pi}_t - \alpha_b \rho_b \bar{\kappa}_y (\hat{b}_t - \hat{b}_t^*) + \beta E_t \frac{\partial V_{t+1}}{\partial(\hat{y}_t - \hat{y}_t^*)} = 0, \tag{A.115}$$

where

$$\begin{aligned} \frac{\partial V_{t+1}}{\partial(\hat{y}_t - \hat{y}_t^*)} &= \left[\frac{\partial V_{t+1}}{\partial \hat{R}_t}\frac{\partial \hat{R}_t}{\partial\left(\hat{b}_t - \hat{b}_t^*\right)} + \frac{\partial V_{t+1}}{\partial\left(\hat{b}_t - \hat{b}_t^*\right)}\right]\frac{\partial\left(\hat{b}_t - \hat{b}_t^*\right)}{\partial(\hat{y}_t - \hat{y}_t^*)} \\ &= -\rho_b \bar{\kappa}_y \left[R_b \frac{\partial V_{t+1}}{\partial \hat{R}_t} + \frac{\partial V_{t+1}}{\partial\left(\hat{b}_t - \hat{b}_t^*\right)}\right] \end{aligned} \tag{A.116}$$

with the second equality following from partial derivatives $\frac{\partial \hat{R}_t}{\partial\left(\hat{b}_t - \hat{b}_t^*\right)} = R_b$ (from (A.109)) and again $\frac{\partial\left(\hat{b}_t - \hat{b}_t^*\right)}{\partial(\hat{y}_t - \hat{y}_t^*)} - \rho_b \bar{\kappa}_y$. By the Envelope Theorem, simple calculus gives

$$\frac{\partial V_t}{\partial\left(\hat{b}_{t-1} - \hat{b}_{t-1}^*\right)} = \alpha_\pi \bar{\kappa}_b \hat{\pi}_t + \alpha_b \bar{\rho}_b (\hat{b}_t - \hat{b}_t^*) + \beta \bar{\rho}_b E_t \left[\frac{\partial V_{t+1}}{\partial\left(\hat{b}_t - \hat{b}_t^*\right)} + \frac{\partial V_{t+1}}{\partial \hat{R}_t} R_b\right], \tag{A.117}$$

where we have made use of partial derivatives $\frac{\partial \hat{\pi}_t}{\partial\left(\hat{b}_{t-1} - \hat{b}_{t-1}^*\right)} = \bar{\kappa}_b$ (from (A.112)) and $\frac{\partial\left(\hat{b}_t - \hat{b}_t^*\right)}{\partial\left(\hat{b}_{t-1} - \hat{b}_{t-1}^*\right)} = \bar{\rho}_b$ (from (A.113)). Doing the same with respect to the endogenous state $\hat{R}_{t-1}$, it is easy to verify that $\frac{\partial V_{t+j}}{\partial\left(\hat{b}_{t+j-1} - \hat{b}_{t+j-1}^*\right)} = \frac{\partial V_{t+j}}{\partial \hat{R}_{t+j-1}}$ for $j = 0, 1$. This allows us to rewrite (A.117) as

$$\frac{\partial V_t}{\partial\left(\hat{b}_{t-1} - \hat{b}_{t-1}^*\right)} = \alpha_\pi \bar{\kappa}_b \hat{\pi}_t + \alpha_b \bar{\rho}_b (\hat{b}_t - \hat{b}_t^*) + \beta \bar{\rho}_b \left(1 + R_b\right) E_t \frac{\partial V_{t+1}}{\partial\left(\hat{b}_t - \hat{b}_t^*\right)} \tag{A.118}$$

and (A.116) as

$$\frac{\partial V_{t+1}}{\partial(\hat{y}_t - \hat{y}_t^*)} = -\rho_b \bar{\kappa}_y \left(1 + R_b\right) E_t \frac{\partial V_{t+1}}{\partial \left(\hat{b}_t - \hat{b}_t^*\right)}.$$

Noticing that $\rho_b \bar{\kappa}_y = \bar{\rho}_b \kappa_y$, we plug the latter into the optimal policy first order condition (A.115), yielding

$$\alpha_y \left(\hat{y}_t - \hat{y}_t^*\right) + \alpha_\pi \bar{\kappa}_y \hat{\pi}_t - \alpha_b \rho_b \bar{\kappa}_y (\hat{b}_t - \hat{b}_t^*) = \rho_b \bar{\kappa}_y \left(1 + R_b\right) \beta E_t \frac{\partial V_{t+1}}{\partial \left(\hat{b}_t - \hat{b}_t^*\right)}, \tag{A.119}$$

and then use it to eliminated $E_t \frac{\partial V_{t+1}}{\partial \left(\hat{b}_t - \hat{b}_t^*\right)}$ from (A.118). After a couple of steps of algebraic simplification, we find

$$\frac{\partial V_t}{\partial \left(\hat{b}_{t-1} - \hat{b}_{t-1}^*\right)} = \alpha_\pi \hat{\pi}_t + \frac{\alpha_y}{\kappa_y} (\hat{y}_t - \hat{y}_t^*). \tag{A.120}$$

Leading the latter one period forward, taking conditional expectations, and substituting the result back into (A.119), we obtain the targeting rule under the discretionary monetary policy:

$$\alpha_\pi \bar{\kappa}_y \hat{\pi}_t + \alpha_y \left(\hat{y}_t - \hat{y}_t^*\right) - \alpha_b \rho_b \bar{\kappa}_y \left(\hat{b}_t - \hat{b}_t^*\right) = \beta \rho_b \left(1 + R_b\right) E_t \left[\alpha_\pi \bar{\kappa}_y \hat{\pi}_{t+1} + \alpha_y \frac{\bar{\kappa}_y}{\kappa_y} \left(\hat{y}_{t+1} - \hat{y}_{t+1}^*\right)\right]$$

This corresponds to equation (44) in Section 6.1.

### A.3.2 Targeting Rule under Commitment

Under the timeless perspective of Woodford (2003), the Lagrangian problem for the optimal monetary policy under commitment is:

$$\begin{aligned} &\min E_0 \sum_{t=0}^{\infty} \beta^t \left\{ \frac{1}{2} \left[\alpha_y \left(\hat{y}_t - \hat{y}_t^*\right)^2 + \alpha_\pi \hat{\pi}_t^2 + \alpha_b (\hat{b}_t - \hat{b}_t^*)^2\right] \right. \\ &+ \lambda_t^\pi \left[\hat{\pi}_t - \tilde{\beta} \hat{\pi}_{t+1} - \kappa_y (\hat{y}_t - \hat{y}_t^*) - \hat{e}_t\right] + \lambda_t^R \left[\hat{R}_t - \hat{\pi}_{t+1} - \gamma_b \left(\hat{b}_t - \hat{b}_t^*\right) - \mathcal{M}_t\right] \\ &\left. + \lambda_t^b \left[\left(\hat{b}_t - \hat{b}_t^*\right) - \rho_b (\hat{b}_{t-1} - \hat{b}_{t-1}^* + \hat{R}_{t-1} - \hat{\pi}_t) - \mathcal{N}_t\right] \right\} \end{aligned} \tag{A.121}$$

where $\lambda_t^i$ for $i = \pi, R, b$ are Lagrange multipliers. First order conditions with respect to choice variables and simple algebra deliver the following set of optimality conditions:

$$\alpha_y \left(\hat{y}_t - \hat{y}_t^*\right) = \kappa_y \lambda_t^\pi \tag{A.122}$$

$$\beta \left(\alpha_\pi \hat{\pi}_t + \lambda_t^\pi + \rho_b \lambda_t^b\right) = \tilde{\beta} \lambda_{t-1}^\pi + \lambda_{t-1}^R \tag{A.123}$$

$$\lambda_t^b = \beta \rho_b E_t \lambda_{t+1}^b + \gamma_b \lambda_t^R - \alpha_b (\hat{b}_t - \hat{b}_t^*) \tag{A.124}$$

$$\lambda_t^R = \beta \rho_b E_t \lambda_{t+1}^b \tag{A.125}$$

Substituting the expressions for $\lambda_t^\pi$ in (A.122) and for $\lambda_t^R$ (lagged one period) in (A.125) into (A.123), and recalling that $\tilde{\beta} \equiv \beta\Gamma$, simple algebra yields the targeting rule under commitment:

$$\alpha_\pi \hat{\pi}_t + \frac{\alpha_y}{\kappa_y} \left[\left(\hat{y}_t - \hat{y}_t^*\right) - \Gamma \left(\hat{y}_{t-1} - \hat{y}_{t-1}^*\right)\right] + \rho_b \left(\lambda_t^b - E_{t-1} \lambda_t^b\right) = 0, \tag{A.126}$$

where

$$\lambda_t^b = -\alpha_b E_t \sum_{j=0}^{\infty} \eta^j \left(\hat{b}_{t+j} - \hat{b}_{t+j}^*\right), \quad \text{for} \quad \eta \equiv \beta \rho_b \left(1 + \gamma_b\right) \in (0, 1) \tag{A.127}$$

is obtained by combining (A.124)-(A.125) and simple forward iteration since $\gamma_b \in (-1, 0)$.

Absent temptation ($\xi = 0$), we would have $\alpha_b = \gamma_b = 0$, and the unique stationary solution to equations (A.124) and (A.125) would be $\lambda_t^b = \lambda_t^R = 0$ in every period. With $\Gamma = 1$, equation (A.126) would collapse to the baseline targeting rule under commitment linking inflation to output gap growth:

$$\alpha_\pi \hat{\pi}_t + \frac{\alpha_y}{\kappa_y} \left[\left(\hat{y}_t - \hat{y}_t^*\right) - \left(\hat{y}_{t-1} - \hat{y}_{t-1}^*\right)\right] = 0.$$

This corresponds to the no-temptation targeting rule under commitment in equation (52).

## A.4 Proofs

### A.4.1 Proof of Lemma 1

We start by proving that the self-control costs $\xi\left(\ln\tilde{x}_{t+j}-\ln x_{t+j}\right)$ are strictly positive, for any $j\geq 1$. Since equations (1) and (9) hold in every period, including a generic $t+j$, we can substitute the expression for $d_{t+j}-\tau_{t+j}$ from the budget constraint (1) into the expression for $\tilde{x}_{t+j}$ in (9). Simple algebra gives $\tilde{x}_{t+j}=x_{t+j}+b_{t+j}$, such that $\tilde{x}_{t+j}>x_{t+j}$ if $b_{t+j}>0$. It immediately follows that $\xi\left(\ln\tilde{x}_{t+j}-\ln x_{t+j}\right)>0$. The latter, together with the fact that both Lagrange multipliers $\tilde{\lambda}_{t+j}$ and $\lambda_{t+j}$ are strictly positive, implies that $\Gamma_{t+j}<1$. It remains to show that $\Gamma_{t+j}>0$. Making use of the equalities $\tilde{\lambda}_{t+j}=\xi\tilde{x}_{t+j}^{-1}$ and $\lambda_{t+j}=(1+\xi)x_{t+j}^{-1}$, in (6) and (11) respectively, we have $\Gamma_{t+j}>0$ if and only if $\xi\tilde{x}_{t+j}^{-1}<(1+\xi)x_{t+j}^{-1}$. The latter is equivalent to $\xi\tilde{x}_{t+j}^{-1}\left(1-\frac{\tilde{x}_{t+j}}{x_{t+j}}\right)<x_{t+j}^{-1}$, which always holds since the term in brackets on the left hand side of the inequality is always negative. Hence, as long as $b_{t+j}>0$ we can conclude that $\Gamma_{t+j}\in(0,1)$.

### A.4.2 Proof of Proposition 2

As for the decentralized equilibrium, it involves two stages. First, we identify the efficient allocation under temptation by solving $\max\xi\ln\left(\tilde{c}_t-\frac{\tilde{h}_t^{1+\chi}}{1+\chi}\right)$, subject to $\tilde{c}_t+g_t=y_t=z_t\tilde{h}_t$. Simple calculus gives $\tilde{h}_t=z_t^{\frac{1}{\chi}}$, and $\tilde{x}_t=z_t^{\frac{1+\chi}{\chi}}-g_t$. Plugging the latter into the temporary utility of the representative agent, the second stage problem involves $\max\left(1+\xi\right)\ln x_t-\xi\ln\tilde{x}_t$ subject to $x_t=z_th_t-g_t-\frac{h_t^{1+\chi}}{1+\chi}$ and $\tilde{x}_t=z_t^{\frac{1+\chi}{\chi}}-g_t$. The first order condition with respect to $h_t$ yields $x_t^{-1}\left(z_t-h_t^{\chi}\right)=0$, hence $h_t=h_t^*\equiv z_t^{\frac{1}{\chi}}$. The latter yields the efficient levels of output, $y_t^*\equiv z_t^{\frac{1+\chi}{\chi}}$, and consumption, $c_t^*\equiv y_t^*-g_t$.

### A.4.3 Proof of Proposition 1

At the steady state, the nominal (and real) interest rate is

$$R=\frac{\left(1+\xi\right)\left(\omega+b_y\right)}{\beta\left[\left(1+\xi\right)\left(\omega+b_y\right)-\xi\omega\right]}\tag{A.128}$$

Assuming $\omega > 0$, differentiating (A.128) with respect to respect $\xi$ gives

$$\frac{\partial R}{\partial \xi} = \frac{\omega (\omega + b_y)}{\beta \left[\omega + b_y (1+\xi)\right]^2} > 0, \qquad \text{and} \qquad \frac{\partial^2 R}{\partial \xi^2} < 0$$

Differentiating (A.128) with respect to respect $b_y$ gives

$$\frac{\partial R}{\partial b_y} = -\frac{\xi\omega (1+\xi)}{\beta \left[\omega + b_y (1+\xi)\right]^2} < 0, \qquad \text{and} \qquad \frac{\partial^2 R}{\partial b_y^2} > 0$$

### A.4.4 Proof of Proposition 4

Let $U(x_{J,t}, \tilde{x}_{J,t})$ denote the temptation-augmented GHH utility under monetary policy $J$. We define the consumption equivalent (CE) welfare cost as the unique solution $\delta_J$ to the following equality:

$$\underbrace{E\sum_{t=0}^{\infty} \beta^t U\left[x\left(c_{J,t}, h_{J,t}\right), \tilde{x}_{J,t}\right]}_{\mathcal{U}_J} = \underbrace{E\sum_{t=0}^{\infty} \beta^t U\left[x\left(c\left(1-\delta_J\right), h\right), \tilde{x}\right]}_{\mathcal{U}(\delta_J)}, \tag{A.129}$$

where the right hand side is the unconditional household's lifetime welfare at the undistorted steady state, with consumption $c$ diminished by a share $\delta_J$. Under the utility specification in (5), the right hand side of (A.129) gives

$$\begin{aligned}(1-\beta)\,\mathcal{U}(\delta_J) &= \ln\left[(1-\delta_J)\,c_y y - \frac{y^{1+\chi}}{1+\chi}\right] \\ &\quad - \left\{\xi \ln\left[(1-\delta_J)\,c_y y - \frac{y^{1+\chi}}{1+\chi} + b\right] - \xi \ln\left[(1-\delta_J)\,c_y y - \frac{y^{1+\chi}}{1+\chi}\right]\right\} \\ &= \ln\left[\omega - \delta_J c_y\right] - \left\{\xi \ln\left[\frac{\omega - \delta_J c_y + b}{\omega - \delta_J c_y}\right]\right\}, \end{aligned} \tag{A.130}$$

where the first equality makes use of market clearing conditions ($c = c_y y$ and $h = y$) and $\tilde{x}_t = x_t + b_t$, and the second one follows from the fact that, at the efficient steady state, $\omega = \frac{(1+\chi)c_y - 1}{(1+\chi)}$ and $y = 1$.

From the second-order welfare approximation, the left hand side of (A.129) yields

$$
\begin{aligned}
(1-\beta)\mathcal{U}_J &= U-(1-\beta)\frac{U_y}{2}E\sum_{t=0}^{\infty}\beta^t\left[\alpha_y\left(\hat{y}_{a,t}-\hat{y}_t^*\right)^2+\alpha_\pi\hat{\pi}_{a,t}^2+\alpha_b(\hat{b}_{a,t}-\hat{b}_t^*)^2\right] \\
&= \ln(\omega)-\left[\xi\ln\left(1+\frac{b_y}{\omega}\right)\right] \qquad (A.131)\\
&\quad \underbrace{+\frac{1}{2}\left[\Lambda_y Var\left(\hat{y}_{J,t}-\hat{y}_t^*\right)+\Lambda_y Var\left(\hat{\pi}_{J,t}\right)+\Lambda_y Var(\hat{b}_{J,t}-\hat{b}_t^*)\right]}_{SOT_J}
\end{aligned}
$$

where the second equality follows from the steady state expression for $U$ and the fact that $E\left(\hat{v}_t^2\right)=Var\left(\hat{v}_t\right)$ for generic variable $v$, with $\Lambda_i=-U_y\alpha_i$ for $i=y,\pi,b$. Notice that the second order term $SOT_J$ is a just linear transformation of the unconditional loss $\mathcal{L}_J$ defined in (54): $SOT_J=-U_y\mathcal{L}_J$. Equating the expressions in (A.130) and (A.131), we obtain an implicit equation for $\delta_J$:

$$
\ln(\omega)-\left[\xi\ln\left(1+\frac{b_y}{\omega}\right)\right]+SOT_J=\ln\left[\omega-\delta_J c_y\right]-\left\{\xi\ln\left[\frac{\omega-\delta_J c_y+b}{\omega-\delta_J c_y}\right]\right\} \qquad (A.132)
$$

Notice that for the right hand side of (A.132) to be well-defined we need to impose the restriction $\delta_J<\frac{\omega}{c_y}\in(0,1)$.

Let's start with the case of **no temptation**. Setting $\xi=0$ in (A.132), we find simple algebra gives a unique solution $\delta_J^{NK}$ (with the superscript $NK$ standing for "baseline New Keynesian"):

$$
\delta_J^{NK}=\frac{\omega}{c_y}\left[1-\exp\left(SOT_J\right)\right] \qquad (A.133)
$$

Since in this case

$$
SOT_J=-U_y\mathcal{L}_J=-\omega^{-1}\left[\chi Var\left(\hat{y}_{J,t}-\hat{y}_t^*\right)+\frac{\epsilon}{\kappa}Var\left(\hat{\pi}_{J,t}\right)\right]<0,
$$

we have $\delta_J^{NK}\in\left(0,\frac{\omega}{c_y}\right)$: absent temptation, welfare costs are always positive.

Moving to the case of **temptation**, for analytical convenience, define $m\equiv\omega-\delta_a c_y$ and rewrite

(A.132) as follows:

$$SOT_J = -\ln(\omega) + \left[\xi \ln\left(1 + \frac{b_y}{\omega}\right)\right] + \ln m - \frac{1}{1+\zeta}\left[\xi \ln\left(1 + \frac{b_y}{m}\right)\right]^{1+\zeta} \tag{A.134}$$

Let $G(m)$ denote the right hand side of (A.134), which, again, is well-defined only for $m > 0$. Simple calculus and algebra identify the following properties: i) $G(m) \gtreqless 0$ for $m \gtreqless \omega$, with $\lim_{m\to 0} G(m) = -\infty$ and $\lim_{m\to+\infty} G(m) = +\infty$; ii) $G'(m) > 0$ and $G''(m) < 0$, with $\lim_{m\to+\infty} G'(m) = 0$. By these properties, it is straightforward to infer that, for any $\xi > 0$, there exists a unique $m^* > 0$ such that $G(m^*) = SOT_J$, where $m^* \gtreqless \omega$ if and only if $SOT_J \lesseqgtr 0$. Equivalently, by the definition of $m$ and $SOT_J$, we can conclude there exists a unique $\delta_J^* < \frac{\omega}{c_y}$ such that $\mathcal{U}_J = \mathcal{U}(\delta_J^*)$, where $\delta_J^* \lesseqgtr 0$ if and only if $\mathcal{L}_J \gtreqless 0$.

To study the relationship between welfare costs and temptation, let $m^*(\xi)$ denote the solution $m^*$ defined above as function of temptation $\xi$. By a simple application of the Implicit Function Theorem, we have that

$$\frac{\partial m^*(\xi)}{\partial \xi} = \left.\frac{\frac{\partial SOT_J}{\partial \xi} - \frac{\partial G(m)}{\partial \xi}}{G'(m)}\right|_{m=m^*} \tag{A.135}$$

Simple calculus gives $\frac{\partial G(m)}{\partial \xi} \gtreqless 0$ for $m \gtreqless \omega$. Suppose that $\frac{\partial SOT_J}{\partial \xi} > 0$: namely, the unconditional loss is strictly decreasing in $\xi$.[56] Since $G'(m) > 0$, from (A.135) and the fact that $\delta_J^* = c_y^{-1}(\omega - m^*)$ (from the definition of $m$), we can conclude that $\delta_J^*$ is strictly decreasing in $\xi$ for any $\delta_J^* \geq 0$.[57]

### A.4.5 Proof of Proposition 5

Define $a_t \equiv 1 + b_{x,t}$. Using standard notation, its log-linearization around the steady state yields $\hat{a}_t = \frac{b_x}{a}\left(\hat{b}_t - \hat{x}_t\right)$. Since $\hat{n}_t \sim N\left(0, \sigma_n^2\right)$ for $n = z, g, e, v_f$, along the equilibrium path, all (log-linearized) endogenous variables are mean-zero normally distributed variables. In particular, we have that $\hat{a}_t \sim N\left(0, \sigma_{\hat{a}_t}^2\right)$, such that $w_t \equiv \ln a_t \sim N\left(\bar{w}, \sigma_w^2\right)$ with $\bar{w} = \ln a$ and $\sigma_w^2 = \sigma_{\hat{a}_t}^2$. Hence,

[56] Although this cannot be proved analytically, it appears to be the case in all our numerical simulations.
[57] Our numerical analysis will show that $\delta_J^*$ is strictly decreasing in $\xi$ also for negative values of $\delta_J^*$ .

$a_t$ is log-normally distributed with mean $\bar{a}$ and variance $\sigma_a^2$ given by, respectively,

$$\bar{a} = e^{\bar{w}+\frac{\sigma_w^2}{2}}, \qquad \sigma_a^2 = \left(e^{\sigma_w^2} - 1\right) e^{2\left(\bar{w}+\frac{\sigma_w^2}{2}\right)} = \bar{a}^2 \left(e^{\sigma_w^2} - 1\right)$$

Using there expressions, simple algebra gives $\bar{w} = \ln \bar{a} - \ln \sqrt{1 + \frac{\sigma_a^2}{\bar{a}^2}}$, such that the unconditional expected costs of self-control are $E\left[SC\left(a_t\right)\right] = \xi E\left(w_t\right) = \xi \left(\ln \bar{a} - \ln \sqrt{1 + \frac{\sigma_a^2}{\bar{a}^2}}\right)$. Keeping the unconditional mean $\bar{a}$ constant, an increase in wealth volatility $\sigma_a^2$ always lowers $E\left[SC\left(a_t\right)\right]$.

## References


[1] Acharya, S., Challe, E., Dogra, K. (2022). Optimal monetary policy according to HANK, *American Economic Review*, forthcoming

[2] Adam, K., Woodford, M. (2012). Robustly optimal monetary policy in a microfounded New Keynesian model. *Journal of Monetary Economics*, 59, 468-487.

[3] Ahn, D.S., Iijima, R., Sarver, T. (2020). Naivete about temptation and self-control: foundations for recursive naive quasi-hyperbolic discounting. *Journal of Economic Theory*, 189, 1-40.

[4] Airaudo, M. (2020). Temptation and forward guidance. *Journal of Economic Theory*, 186, 1-29.

[5] Airaudo, M., Hajdini, I. (2023). Wealth effects, price mark-ups, and the neo-Fisherian hypothesis. *European Economic Review*, forthcoming.

[6] Andreasen, M.M., Fernandez-Villaverde, J., Rubio-Ramirez, J.F. (2018). The pruned state-space system for non-linear models. *Review of Economic Studies*, 85, 1-49.

[7] Angeletos, G.M., Huo, Z., Sastry, K. (2020). Imperfect macroeconomic expectations: evidence and theory. *NBER Macroeconomic Annual 2020*, 35, 1-86.

[8] Angeletos, G.M., Laibson, D., Repetto, A., Tobacman, J., Weinberg, S. (2001). The hyperbolic consumption model: calibration, simulation, and empirical evaluation. *Journal of Economic Perspectives*, 15, 47-68.

[9] Angeletos, G.M., Lian, C. (2016). Incomplete information in macroeconomics: accommodating frictions in coordination. *Handbook of Macroeconomics*, Volume 2, Ch. 14, 1065-1240.

[10] Attanasio, O., Kovacs, A., Moran, P. (2020). Temptation and commitment: understanding hand-to-mouth behavior. NBER Working Paper 27944.

[11] Attanasio, O., Kovacs, A., Moran, P. (2021). Temptation and incentives to wealth accumulation. NBER Working Paper 28938.

[12] Attanasio, O., Low, H. (2004). Estimating Euler equations. *Review of Economic Dynamics*, 7, 405-435.

[13] Auclert, A., Bardóczy, B., Ronglie, M. (2021). MPCs, MPEs, and multipliers: a trilemma for New Keynesian models. *The Review of Economics and Statistics*, forthcoming.

[14] Backus, D.K., Routledge, B.R., Zin, S.E. (2004). Exotic preferences for macroeconomists. *NBER Macroeconomics Annual 2004*, 19, 319-390.

[15] Banerjee, A., Mullainathan, S. (2010). The shape of temptation: implications for the economic lives of the poor. NBER, Working Paper 15973.

[16] Barberis, N., Huang, M., Santos, T. (2001). Prospect theory and asset prices. *Quarterly Journal of Economics*, 116, 1-53.

[17] Benigno, P., Woodford, M. (2005). Inflation stabilization and welfare: the case of a distorted steady state. *Journal of the European Economic Association*, 3, 1185-1236.

[18] Benigno, P., Woodford, M. (2006). Optimal taxation in an RBC model: a linear-quadratic approach. *Journal of Economic Dynamics and Control*, 30, 1445-1489.

[19] Benigno, P., Woodford, M. (2012). Linear-quadratic approximation of optimal policy problems. *Journal of Economic Theory*, 147, 1-42.

[20] Bianchi, F., Ilut, C., Saijo, H. (2022). Diagnostic business cycles. *Review of Economic Studies*, forthcoming.

[21] Blanchard, O. (2016). The Phillips curve: back to the '60s? *American Economic Review P&P*, 105 (5), 31-34.

[22] Blanchard, O., Gali, J. (2007). Real wage rigidities in the New Keynesian model. *Journal of Money, Credit and Banking*, 39, 37-65.

[23] Bucciol, A. (2012). Measuring self-control problems: a structural estimation. *Journal of the European Economic Association*, 10, 1084-1115.

[24] Busetti, F., Caivano, M. (2018). Low-frequency drivers of the real interest rate: empirical evidence from advanced economies. *International Finance*, 22, 171-185.

[25] Canzoneri, M., Cumby, R., Diba, B., López-Salido, D. (2011). The role of liquid government bonds in the great transformation of the American monetary policy. *Journal of Economic Dynamics and Control*, 35, 282-294.

[26] Carvalho, C. Ferrero, A., Nechio, F. (2016). Demographics and real interest rates: inspecting the mechanism. *European Economic Review*, 88, 208-226.

[27] Cesarini, D., Lindqvist, E., Notowidigdo, M.J., Östling, R. (2017). The effect of wealth on individual and household labor supply: evidence from Swedish lotteries. *American Economic Review*, 107, 3917-3946.

[28] Chen, X., Kirsanova, T., Leith, C. (2017). How optimal is U.S. monetary policy? *Journal of Monetary Economics*, 92, 96-111.

[29] Coroneo, L., Corradi, V., Santos Monteiro, P. (2018). Testing for optimal monetary policy via moment inequalities. *Journal of Applied Econometrics*, 33, 780-796.

[30] Davila, E., Schaab, A. (2022). Optimal monetary policy with heterogeneous agents: a timeless primal-dual approach. Yale University, manuscript.

[31] DeJong, D. N., Ripoll, M. (2007). Do self–control preferences help explain the puzzling behavior of asset prices. *Journal of Monetary Economics*, 54, 1035-1050.

[32] Dekel, E., Lipman, B.L. (2012). Costly self-control and random self-indulgence. *Econometrica*, 80, 1271-1303.

[33] Drugeon, J.P., Wigniolle, B. (2017). On impatience, temptation, and Ramsey conjecture. *Economic Theory*, 63, 73-98.

[34] Frederick, S., Loewenstein, G., O'Donoughe, T. (2002). Time discounting and time preference: a critical review. *Journal of Economic Literature*, 40, 351-401.

[35] Evans, G.W., Honkapohja, S. (2001). *Learning and Expectations in Macroeconomics.* Princeton University Press.

[36] Fudenberg D., Levine, D.K. (2006). A dual-self model of impulse control. *American Economic Review*, 96, 1449-1476.

[37] Fudenberg D., Levine, D.K. (2011). Risk, delay, and convex self-control costs. *American Economic Journal: Microeconomics*, 3, 34-68.

[38] Gabaix, X. (2020). A behavioral New Keynesian model. *American Economic Review*, 110, 2271-2327.

[39] Gali, J. (1999). Technology, employment, and the business cycle: do technology shocks explain aggregate fluctuations? *American Economic Review*, 89, 249-271

[40] Gali, J., Rabanal, P. (2004) Technology shocks and aggregate fluctuations: how well does the real business cycle model fit postwar U.S. data? *NBER Macroeconomics Annual*, 225-288.

[41] Garcia-Schmidt, M., Woodford, M. (2019). Are low interest rates deflationary? A paradox of perfect-foresight analysis. *American Economic Review*, 19, 86-120.

[42] Greenwood, J., Hercovitz, Z., Huffman, G.W. (1988). Investment, capacity utilization and the real business cycle. *American Economic Review*, 78, 402-17.

[43] Groll, D., Monacelli, T. (2020). The inherent benefit of monetary unions. *Journal of Monetary Economics*, 111, 63-79.

[44] Gul, F., Pesendorfer, W. (2001). Temptation and self-control. *Econometrica*, 69, 1403-1435.

[45] Gul, F., Pesendorfer, W. (2004). Self-control and the theory of consumption. *Econometrica*, 72, 119-158.

[46] Harris, C., Laibson, D. (2001). Dynamic choices of hyperbolic consumers. *Econometrica*, 69, 935-957

[47] Haug, A.A. (2020). Testing Ricardian equivalence with the narrative record of tax changes. *Oxford Bulletin of Economics and Statistics*, 82, 387-404.

[48] Hiraguchi, R. (2018). Temptation and self-control in a monetary economy. *Macroeconomic Dynamics*, 22, 1076-1095.

[49] Huang, K.X.D., Liu, Z., Zhu, J.Q. (2015). Temptation and self-control: some evidence and applications. *Journal of Money, Credit and Banking*, 47, 481-615.

[50] Ilut, C., Schneider, M. (2014). Ambiguous business cycle. *American Economic Review*, 104, 2368-2399

[51] Kean, M., Rogerson, R. (2012). Micro and macro labor supply elasticities: a reassessment of conventional wisdom. *Journal of Economic Literature*, 50, 464-476.

[52] Kőszegi, B., Rabin, M. (2009). Reference-dependent consumption plans. *American Economic Review*, 99, 909-936.

[53] Kovacs, A., Low, H., Moran, P. (2021). Estimating temptation and commitment over the life cycle. *International Economic Review*, 62, 101-139.

[54] Krusell, P., Kuruscu, B., Smith, A.A. (2002). Time orientation and asset prices. *Journal of Monetary Economics*, 49, 107-135.

[55] Krusell, P., Kuruscu, B., Smith, A.A. (2010). Temptation and taxation. *Econometrica*, 78, 2063-2084.

[56] Kumru, C.S., Thanopoulos, A.C. (2011). Social security reform with self-control preferences. *Journal of Public Economics*, 95, 886-899.

[57] L'Huillier, J.P., Singh, S.S., Yoo, D. (2021). Incorporating diagnostic expectations into the New Keynesian framework. Federal Reserve Bank of Cleveland, manuscript.

[58] Leith, C., von Thadden, L. (2008). Monetary and fiscal policy interactions in a New Keynesian model with capital accumulation and non-Ricardian consumers. *Journal of Economic Theory*, 140, 279-313-

[59] Linnemann, R. (2006). Interest rate policy, debt, and indeterminacy with distortionary taxation. *Journal of Economic Dynamics and Control*, 30, 487-510.

[60] Lu, S.E. (2016). Models of limited self-control: comparison and implications for bargaining. *Economics Letters*, 145, 186-191

[61] Lucas, R. (1987). *Models of Business Cycles,* Oxford, Blackwell.

[62] Mackowiak, B., Matejka, F., Wiederholt, M. (2021). Rational inattention: a review. *Journal of Economic Literature*, forthcoming.

[63] Maliar, L., Maliar, S. (2006). Indeterminacy in a log-linearized neo-classical growth model with quasi-geometric discounting. *Journal of Economic Dynamics and Control*, 23, 492-505.

[64] Maliar, L., Maliar, S. (2016). Ruling out multiplicity of smooth equilibria in dynamic games: a hyperbolic discounting example. *Dynamic Games and Applications*, 6, 243-261.

[65] Meissner, T., Rostam-Afschar, D. (2017). Learning Ricardian equivalence. *Journal of Economic Dynamics and Control*, 82, 273-288.

[66] Michaillat, P., Saez, E. (2021). Resolving New Keynesian anomalies with wealth in the utility function. *Review of Economics and Statistics*, 103, 197-215.

[67] Monacelli, T., Perotti R. (2008). Fiscal policy, wealth effects, and markups. NBER Working Paper N. 14584.

[68] Muraven, M., Shmueli D., Burkley, E. (2006). Conserving self-control strength. *Journal of Personality and Social Psychology*, 91, 524-537.

[69] Nakajima, M. (2012). Rising indebtedness and temptation: a welfare analysis. *Quantitative Economics*, 3, 257-288.

[70] Noor, J., Takeoka, N. (2010). Uphill self-control. *Theoretical Economics*, 5, 127-158.

[71] Ravenna, F., Walsh, C. (2006). Optimal monetary policy with the cost channel. *Journal of Monetary Economics*, 53, 199-216.

[72] Ravn, M.O., Sterk, V. (2021). Macroeconomic fluctuations with HANK and SAM: an analytical approach. *Journal of the European Economic Association*, 19, 1162-1202.

[73] Rieger, M.O., Wang, M., Hens, Thorsten (2016). How time preferences differ: evidence from 53 countries. *Journal of Economic Psychology*, 52, 115-135.

[74] Rieger, M.O., Wang, M., Hens, Thorsten (2021). Universal time preference. *PLoS One*, 16, 1-15.

[75] Ricciuti, R. (2003). Assessing Ricardian equivalence. *Journal of Economic Surveys*, 17, 55-78.

[76] Rigon, M., Zanetti, F. (2018). Optimal monetary policy and fiscal policy interaction in a non-Ricardian economy. *International Journal of Central Banking*, June, 389-436

[77] Rudebusch, G.D., Swanson, E.T. (2012). The bond premium in a DSGE model with long-run real and nominal risks. *American Economic Journal: Macroeconomics*, 4, 105-143.

[78] Rustichini, A. (2008). A dual or a unitary system? Two alternative models of decision making. *Cognitive, Affective, and Behavioral Neuroscience*, 8, 355-362.

[79] Schilback F. (2019). Alcohol and self-control: a field experiment in India. *American Economic Review*, 109, 1290-1322.

[80] Schmitt-Grohé, S., Uribe, M. (2012). Optimal simple and implementable monetary and fiscal rules. *Journal of Monetary Economics*, 54, 1702-1725.

[81] Schmitt-Grohé, S., Uribe, M. (2012). What's news in business cycles. *Econometrica*, 80, 2733-2764.

[82] Schlafmann, K. (2021). Housing, mortgages, and self-control. *Review of Financial Studies*, 34, 2648-2687.

[83] Smets, F., Wouters, R. (2003). An estimated dynamic stochastic general equilibrium model of the euro area. *Journal of the European Economic Association*, 1, 1123-1175.

[84] Smets, F., Wouters, R. (2007). Shocks and frictions in U.S. business cycles: a Bayesian DSGE approach. *American Economic Review*, 97, 586-606.

[85] Soderlind, P. (1999). Solution and estimation of RE macromodels with optimal policy. *European Economic Review*, 43, 813-823.

[86] Steinsson, J. (2003). Optimal monetary policy in an economy with inflation persistence. *Journal of Monetary Economics*, 50, 1425-1456

[87] Strotz, R. (1956). Myopia and inconsistency in dynamic utility maximization. *Review of Economic Studies*, 23, 165-180.

[88] Telser, L.G., Graves, R.L. (1972). *Functional Analysis in Mathematical Economics: Optimization over Infinite Horizons*, University of Chicago Press.

[89] Toussaert, S. (2018). Eliciting temptation and self-control through menu choices: a lab experiment. *Econometrica*, 86, 859-889.

[90] Toussaert, S. (2019). Revealing temptation through menu choice: field evidence. Manuscript, University of Oxford.

[91] Uribe, M., Yue, V.Z. (2006). Country spreads and emerging countries: who drives whom? *Journal of International Economics*, 69, 6-36.

[92] Woodford, M. (2003). *Interest and Prices.* Princeton University Press.

[93] Woodford, M. (2013). Macroeconomic analysis without the rational expectations hypothesis. *Annual Review of Economics*, 5, 303-346.

[94] Woodford, M. (2019). Monetary polilcy analysis when planning horizons are finite. Columbia University, manuscript.