Optimal Policy for Behavioral Financial Crises [*]

Paul Fontanier[†]

November 2024

Abstract

Should policymakers adapt their macroprudential and monetary policies when the financial sector is vulnerable to belief-driven boom-bust cycles? I develop a model in which financial intermediaries are subject to collateral constraints, and that features a general class of deviations from rational expectations. I show that distinguishing between the drivers of behavioral biases matters for the precise calibration of policy: when biases are a function of equilibrium asset prices, as in return extrapolation, new externalities arise, even in models that do not have any room for policy in their rational benchmark. These effects are robust to the degree of sophistication of agents regarding their future biases. I show how time-varying leverage, investment and price regulations can achieve constrained efficiency. Importantly, greater uncertainty about the extent of behavioral biases in financial markets reinforces incentives for preventive action.

JEL Classification Numbers: D62, E44, E52, E70, G28

Keywords: Financial crises, Beliefs, Extrapolation, Macroprudential Policy, Optimal Policy under Uncertainty

[*]I am extremely grateful to my advisors Emmanuel Farhi, Sam Hanson, Andrei Shleifer, Jeremy Stein and Ludwig Straub for invaluable guidance and support. I thank Philipp Schnabl (the editor) and several referees for helpful suggestions that substantially improved the paper. I also thank Nicholas Barberis, Francesca Bastianello, Jonathan Benchimol, Adrien Bilal, John Campbell, Gabriel Chodorow-Reich, Chris Clayton, Maria Fayos Herrera, Xavier Gabaix, Nicola Gennaioli, Robin Greenwood, Antoine Jacquet, Spencer Yongwook Kwon, Jean-Paul L'Huillier, Albert Marcet, Kenneth Rogoff, Alp Simsek, Larry Summers, Adi Sunderam, Jean Tirole, Luis Viceira, Jonathan Wallen, Michael Woodford, and participants at the 2019 Sloan-NOMIS Workshop on the Cognitive Foundations of Economic Behavior, the CU-RIDGE 2021 Financial Stability Workshop, the 2022 Theories & Methods in Macroeconomics Conference, the AQR Top Finance Graduate Award Conference at CBS, the Virtual Israel Macro Meeting, and numerous seminars for helpful comments. I am grateful to the Corps des Mines for supporting this research.
[†]Yale School of Management. paul.fontanier@yale.edu & paulfontanier.github.io

1 Introduction

Should policymakers be concerned about asset price booms, and should they act preemptively before they burst? Historically the dominant paradigm among policymakers has relied on the idea that financial crises are "bolts from the sky," triggered by unpredictable and large negative shocks. Because private agents implicitly understand the riskiness of the activities they engage in, rapid growth in asset prices can only be supported by sound fundamentals and is not a cause for concern *per se* (Gorton 2012, Geithner 2014). This contrasts sharply with the alternative, behavioral view of financial bubbles and crises that has been revived after the great financial crisis. Following in the footsteps of Minsky (1977) and Kindleberger (1978), researchers showed that factors such as credit growth and asset price booms successfully predict financial crises (Jordà, Schularick, and Taylor 2015). At the same time, survey data supports the idea that investors' beliefs are inconsistent with the Rational Expectations hypothesis, generally pointing towards the importance of extrapolation in financial markets (Gennaioli and Shleifer 2018). In response, economists have developed a number of behavioral models of financial instability.[1] Still, how policymakers should adapt their toolbox when financial instability is driven by systematic behavioral biases is largely an open question.

I tackle this question by constructing a model of financial crises in which agents display arbitrary deviations from rationality, and analyze optimal policy from the perspective of a social planner who recognizes that agents may have behavioral biases. I use this model to clarify three key normative questions surrounding the policy debate. First, which features of behavioral biases matter for welfare and should therefore be a concern for financial stability? Second, is there still room for intervention when the planner and the market share the same beliefs? And third, how should regulators incorporate incomplete information about behavioral biases when contemplating early action?

The contributions of this paper can be summarized in three points. First, I show that welfare losses are driven by three key features of behavioral biases: (i) irrational optimism

[1] See Bordalo, Gennaioli, and Shleifer (2018), Greenwood, Hanson, and Jin (2019), Maxted (2024) and Krishnamurthy and Li (2024). These models are able to match moments that are inconsistent with rational frameworks, such as low credit spreads during the the run-up to financial crises.

in booms if financial frictions might bind later on; (ii) future irrational pessimism during financial crises; and (iii) how equilibrium prices impact biases. Second, I show how regulators can use leverage, investment, and price regulation (such as monetary policy) to improve welfare, and highlights which welfare effects are robust to the degree of sophistication of private agents and thus survive even if the planner shares the same beliefs as the market. And third, I show that greater uncertainty about the extent of behavioral biases in financial markets increases the incentives for the planner to act early.

I present the model in Section 2. It features three periods and two types of agents: financial intermediaries and households. Financial intermediaries borrow by issuing deposits to households, and can invest in the creation of risky assets. At the heart of the model lies a financial friction: in the intermediate period, borrowing by intermediaries needs to be secured by posting risky assets as collateral. The amount of borrowing available depends on the quantity of collateral, and on the expectation of its future payoff. Such a friction, while keeping the economy away from the first-best, does not create any externality in a rational benchmark, and thus does not leave any room for policy. The central element of the model is a general class of deviations from rationality in the formation of agents' expectations, which applies in all periods. I introduce a behavioral bias that shifts agents' perceived distribution of future dividends. The behavioral bias is allowed to depend on both fundamentals and asset prices, and agents are allowed to be potentially sophisticated about future biases. It is general enough to represent many psychological phenomena, while keeping the welfare analysis tractable.

I present the welfare analysis in Section 3, where a paternalistic social planner evaluates welfare using his own (rational) expectations. I start by fleshing out how behavioral factors and financial frictions interact to create uninternalized welfare effects. This analysis clarifies that irrational over-optimism in booms creates first-order welfare losses only when there is a chance that financial frictions might later bind. Furthermore, it highlights how the predictable components of *future* behavioral biases formed inside a crisis also create losses and should be monitored. Indeed, if private agents tend to be over-pessimistic during financial crises, but neglect this future bias, they over-borrow in good times. If the social planner anticipates that future behavioral biases will be on the side of over-pessimism during an eventual financial crisis, there is a wedge between private expectations and those of the social planner. Here again, the interaction with financial frictions is crucial. Expected losses are greater when over-pessimism coincides with deeper financial crises: behavioral biases are tightening an already tight collateral constraint.

The welfare decomposition delivers a second key insight: precisely distinguishing between the drivers of these behavioral biases matters. When behavioral biases depend on asset prices, new externalities arise. By borrowing and investing, agents influence the realization of equilibrium prices, which can in turn alter the magnitude of behavioral biases. These effects, only present in the case of endogenous sentiment, are akin to pecuniary externalities but work through beliefs. For example, short-term borrowing lowers agents' net worth in a future crisis, which has a negative effect on future equilibrium prices. With endogenous sentiment such as price extrapolation, this fall in prices can trigger irrational pessimism, which tightens collateral constraints and deepens crises. Belief amplification thus creates an externality that calls for reducing leverage ex-ante. An additional effect, termed a *reversal externality*, works through prices *during* the boom. Agents' demand for risky assets in good times bid up their prices. This can feed pessimism tomorrow by impacting the magnitude of behavioral biases in the future. For instance, if agents are simply extrapolating price changes, a high price in the past is a force that pushes agents towards irrational pessimism later (Farhi and Werning 2020). Hence an increase in prices today will cause a reversal in beliefs tomorrow, which will tighten collateral constraints and prevent all intermediaries from rolling over their debt.

Notably, these externalities are present even when private agents are *fully sophisticated* about future biases, and even if the market shares the same beliefs as the regulator. Even though financial intermediaries can be fully aware that the market will be irrationally panicking in a future financial crisis, their decisions are still privately optimal. Atomistic intermediaries cannot coordinate to collectively reduce their leverage or decrease asset prices in order to alleviate the effects of future pessimism. Only an intervention from the planner can solve these externalities, showing that naïvety or belief differences between policymakers and market participants are not key for these results.

Section 4 fleshes out how to interpret these results in terms of real-world policy. The tax on short-term borrowing can naturally be interpreted as capital structure regulation. If behavioral biases fluctuate along the business cycle, the optimal level of these restrictions is time-varying. My model thus calls for the use of counter-cyclical capital buffers. It shows that the time-variation should not only track the contemporaneous extent of over-optimism in financial markets, but should also consider how it will influence the *future* realizations of behavioral biases in eventual crises, as well as the expected impact of future prices on future biases. Similarly, to regulate the quantity of investment, regulators can rely on the implementation of Loan-to-Value (LTV) ratios. The optimal LTV limit should also be time-varying, and should closely track the same behavioral biases as do the counter-cyclical capital buffers. The reversal externality however calls for the use of a third instrument in order to control prices. Monetary policy can thus be used as a complementary tool, even when counter-cyclical capital buffers and LTV ratios can be flexibly adapted. A monetary tightening lowers contemporaneous asset prices which influences future sentiment, relaxing collateral constraints in crises. Such action does not require any information about biases in the boom phase, and is robust to the level of sophistication of agents. This suggests that the concern for the central bank should not only be placed on whether prices are rational, but also on whether price booms will trigger further rounds of price extrapolation later on.

It is however undeniable that identifying a bubble, or anticipating future pessimism, is intrinsically difficult since corresponding fundamentals are not observable, a recurring argument from the advocates of the "wait-and-see" approach (Bernanke 2002). I acknowledge this issue but show that the intuition goes in the opposite direction. I allow the social planner to have an imprecise estimate of behavioral biases. The key result is that the strength of the desired ex-ante intervention on leverage is actually *increasing* in uncertainty. The more uncertainty there is about irrationality today, the more important it is to tighten leverage restrictions today. Intuitively, this is because sentiment interacts with financial frictions to create strong non-linearities: the costs of having intervened when it turns out that the price boom was entirely justified by sound fundamentals are dwarfed by the benefits of mitigating a possible sentiment-driven financial crisis.

Relation to the Literature: This paper is primarily motivated by the recent empirical evidence on credit cycles that revived the Minsky (1977) and Kindleberger (1978) narratives. This line of research started with Borio and Lowe (2002) showing that asset price growth and credit growth predict banking crises, stimulating research on the predictability of financial crises. Schularick and Taylor (2012) demonstrate that credit expansions forecast real activity slowdowns. Jordà et al. (2015) and Greenwood, Hanson, Shleifer, and Sørensen (2022) show that combining credit growth measures with asset price growth substantially increases the out-of-sample predictive power on a subsequent financial crisis. In a recent survey, Sufi and Taylor (2022) argue that "all told, the emerging historical evidence supports the existence of systematic behavioral biases in explaining credit cycles." Direct evidence of such biases comes from survey data: Bordalo et al. (2018) document the predictability of forecast errors regarding credit spreads.

My paper integrates these lessons into the literature on normative macro-finance. I follow earlier work characterizing generic inefficiencies created by incomplete markets (Geanakoplos and Polemarchakis 1985 ; Greenwald and Stiglitz 1986). In my model, the amount of borrowing is limited by the expectation of the asset's future payoffs, a friction similar to Kiyotaki and Moore (1997). By contrast, most of the recent normative literature (as in Mendoza 2010, Bianchi 2011 and Jeanne and Korinek 2019) uses a collateral constraint that features instead the current price of the asset. This creates a pecuniary externality, since agents do not internalize how their ex-ante leverage decisions impact market prices tomorrow, and hence the aggregate borrowing capacity of the financial sector in the future. Dávila and Korinek (2018) offer a sharp analysis of this market failure.

I end this section by focusing on the most closely related papers. First, Farhi and Werning (2020) analyze an environment with a zero lower bound and aggregate demand externalities, where agents extrapolate returns. Second, Dávila and Walther (2021) study an environment without financial frictions with general belief distortions during the boom, and characterize optimal leverage and monetary policies.[2] Third, Caballero and Simsek (2020a) study monetary policy when macroprudential policy is constrained, and agents have heterogenous beliefs. I build on these results, and also complement them by showing how: (i) behavioral biases create powerful welfare effects even in a model without a market failure in its rational benchmark ; (ii) different types of biases lead to different forms of optimal intervention ; (iii) the externalities created by biases are robust to the degree of sophistication of agents ; and importantly (iv) uncertainty about the precise extent of biases in financial markets reinforce the motives for preventive intervention.

2 Model

This section presents the framework that will serve as the basis for the subsequent welfare analysis. The model is stylized in the tradition of the over-borrowing literature (Lorenzoni 2008), and financial intermediaries play a crucial role (He and Krishnamurthy 2013). To isolate the effects of behavioral biases, it features a borrowing constraint that does not create externalities in a rational equilibrium.

2.1 Setup

Time is discrete, with three periods $t \in \{1, 2, 3\}$. There are two types of agents: financial intermediaries (or banks) and households. Both types are present in measure 1. There is a single good used both for consumption and for investment in the creation of a risky asset. The risky asset can only be held by financial intermediaries, and pays a stochastic dividend at times $t = 2$ and $t = 3$. The asset is also used as a collateral by financial intermediaries to issue deposits in period $t = 2$, and this constraint depends on the ex-

[2] I also contribute to this line of research by providing an alternative way of modeling general belief distortions. My proposal is simpler to use, especially for the welfare analysis, but at the cost of not being able to replicate the arbitrary distortions on the entire probability distribution used in Dávila and Walther (2021). For instance, Dávila and Walther (2021) can investigate how policy depends on whether agents are optimistic regarding left-tail or right-tail outcomes, a case my modeling choice cannot nest. However, it proves particularly convenient when I study the empirically relevant case where the social planner is uncertain about the precise extent of irrationality in financial markets.

pectation of the future payoff of the asset. I define a "financial crisis" as a moment when the borrowing constraint of financial intermediaries binds at time $t = 2$.

Throughout the paper, the aggregate state of the world corresponds to the dividend payment of the risky asset, and is denoted by z_t at time t, while the equilibrium price of the risky asset is denoted by q_t.

2.1.1 Beliefs

Beliefs about the Exogenous State: The objective distribution of the aggregate state z_t is given by two independent cumulative probability distributions $z_2 \sim F_2(z)$ and $z_3 \sim F_3(z)$, with support on $\mathbb{R}_+^*$. At time $t-1$, all agents hold the same subjective beliefs about the aggregate state of the world at t, given by distributions $\tilde{F}_t$. To keep the welfare analysis tractable while being general enough to represent several psychological phenomena, the deviations from the objective probability distribution F_t are encoded in a scalar Ω_t:

$$\tilde{F}_t(z) = F_t(z - \Omega_t) \tag{1}$$

Agents at time t thus believe that the dividend paid in each state of the world at $t+1$ will be $z_{t+1} + \Omega_{t+1}$ rather than z_{t+1}.

The behavioral bias Ω_{t+1} is thus a location shifter on expected dividends. In that respect, Ω_{t+1} exactly represents the *predictable component* at t of forecast errors realized at $t+1$. A positive bias Ω_{t+1} means that agents are over-optimistic at time t regarding the prospects of dividends in the future. In this case, sentiment will be said to be high, or equivalently that markets are displaying "irrational exuberance" (Shiller 2015). A negative bias Ω_{t+1} means that agents are over-pessimistic at t: sentiment will be said to be low, or equivalently that markets are displaying "irrational distress" (Fisher 1932). The generality of the approach comes from allowing the Ω biases to depend on several variables: fundamentals and prices.

Definition 1 (Behavioral Bias in Expectations)**.** *A behavioral bias at time t is a function* Ω_{t+1} *that can depend on fundamental or prices:*

$$\Omega_{t+1} : \mathcal{I}_t = \{z_{t-i}, q_{t-i}\}_{i \geq 0} \to \mathbb{R} \tag{2}$$

This implies that Ω can also be an *equilibrium* object. This approach is particularly flexible for the subsequent welfare analysis, since it summarizes all possible distortions in a single quantity. Throughout the paper the biases Ω are kept general, highlighting the properties of sentiment that matter for welfare. It will be useful to flesh out specific examples to build intuition, however. I will focus on two functional forms that are common in the behavioral finance literature, and have been used to explain the credit cycle facts I reviewed in the introduction.[3] The first case is *fundamental extrapolation* (Barberis, Shleifer, and Vishny 1998 ; Rabin and Vayanos 2010 ; Bordalo et al. 2018), modelled here in reduced-form as:

$$\Omega_{t+1} = \alpha_z(z_t - z_{t-1}) \tag{3}$$

where α_z is a positive number. The second case of interest is *price extrapolation* (DeLong, Shleifer, Summers, and Waldmann 1990 ; Hong and Stein 1999 ; Barberis and Shleifer 2003 ; Bastianello and Fontanier 2024), modelled as:

$$\Omega_{t+1} = \alpha_q(q_t - q_{t-1}) \tag{4}$$

with $\alpha_q > 0$.[4] While price extrapolation is aimed at explaining the same set of facts as fundamental extrapolation, it can have drastically different implications, and in particular in terms of policy. This is because agents' present and *future* beliefs now move with policies that move asset prices.

[3] A particularly clear survey of this literature can be found in Barberis (2018). While the core of the paper focuses on these two cases, other behavioral models are nested by the Ω-formulation, such as sticky beliefs, behavioral inattention, and overconfidence.

[4] Because there is no fundamental realization at $t = 1$ or before, I assume that there are hypothetic values z_1 and z_0 driving initial sentiment. The bias at $t = 1$ about next period's payoff will thus be $\Omega_2 = \alpha_z(z_1 - z_0)$, while the bias in the intermediate period will be given by $\Omega_3 = \alpha_z(z_2 - z_1)$. A boom-bust cycle in the spirit of Gennaioli and Shleifer (2018) is thus represented by fundamental realizations $z_1 > z_0$ (good news at $t = 1$) followed by $z_2 < z_1$. Similarly, we will postulate the existence of a hypothetic price q_0 that prevailed in the past and anchors agents' expectations.

Sophistication: Crucially for the results, agents are allowed to be biased in the initial period $t = 1$ as well as in the intermediate (crisis) period $t = 2$. It thus begs the question of whether agents realize that they, or the market, might be biased in the future. To flexibly show how my results change with sophistication or naïvety, I introduce a parameter that controls the level of sophistication of agents, ζ:

Definition 2 (Sophistication Parametrization)**.** *At time $t = 1$, agents believe that their future time 2-selves will form expectations about z_3 according to the following distorted probability distribution:*

$$\tilde{F}_{3,\zeta}(z) = F_3(z - \zeta \cdot \Omega_t) \tag{5}$$

The parameter $\zeta \in \mathbb{R}$ captures the level of sophistication of agents.

When $\zeta = 0$, agents expect their future selves to have unbiased expectations regarding future dividends (*naïvety*). When $\zeta = 1$, agents understand that their future selves will have expectations biased by Ω_3. Agents are partially sophisticated when $\zeta \in]0, 1[$.[5]

Throughout the paper, I will use $\tilde{\mathbb{E}}$ for the expectation of private agents (formed using $\tilde{F}$ and $\tilde{F}_\zeta$) and $\mathbb{E}$ for the Rational Expectations operator:

$$\tilde{\mathbb{E}}_t[z_{t+1}] = \mathbb{E}_t[z_{t+1}] + \Omega_{t+1} \tag{6}$$

2.1.2 Environment

Preferences: Bankers have log-utility in periods $t \in \{1, 2\}$, and linear utility at $t = 3$:

$$U^b = \tilde{\mathbb{E}}_1 \left[\ln(c_1) + \beta \ln(c_2) + \beta^2 c_3 \right] \tag{7}$$

where c_t is the consumption of bankers at t, and β is the standard time discount factor. For simplicity, households (lenders) have deep pockets and linear utility throughout:

$$U^h = \tilde{\mathbb{E}}_1 \left[c_1^h + \beta c_2^h + \beta^2 c_3^h \right]. \tag{8}$$

[5] Implicitly, this formulation assumes that sophisticated agents understand how Ω_3 will be determined according to information available at $t = 2$.

Financial Assets: There are two assets in the economy: deposits and the risky asset. Intermediaries issue deposits $d_t \geq 0$ to households at time t, to finance their consumption and their investment in the risky asset. At time $t = 1$, financial intermediaries can create H units of the asset by paying a convex cost $c(H)$. The equilibrium price of the risky asset at $t = 1$, by no-arbitrage, is thus $q_1 = c'(H)$. This asset pays stochastic dividends z_2 and z_3 in future periods, drawn from the previously introduced independent cumulative probability distributions F_2 and F_3. Only intermediaries have the expertise required to hold risky assets.[6]

Financial Friction: At time $t = 2$, financial intermediaries face a collateral constraint: the amount they can borrow by issuing deposits must be secured by the risky asset, and is thus limited by the expectation of its future payoff. The collateral constraint takes the specific form:

$$d_2 \leq \phi H \tilde{\mathbb{E}}_2[z_3]. \tag{9}$$

The lower ϕ is, the less the bank is able to issue deposits to households in the intermediate period. Equation (9) makes clear that subjective beliefs will directly alter the refinancing capacity of the financial sector. I make one parametric assumption that guarantees that the equilibrium is not trivial.

Assumption 1. *The financial friction parameter is small enough such that the collateral constraint is not always slack, and hence that financial crises are non-zero probability events:*

$$\phi < \beta. \tag{10}$$

[6]Assets thus never change hands in equilibrium. This is in contrast with the notion of "fire sales," developed first in Shleifer and Vishny (1992), where liquidation does not necessarily allocate assets to the highest value users. Dávila and Korinek (2018) call these "distributive" externalities, where redistribution of wealth between agents with different marginal rates of substitution creates an inefficiency. This includes, for instance, the models in Caballero and Krishnamurthy (2003), Lorenzoni (2008) and Fanelli and Straub (2021). Although a rather stark assumption, it is consistent with He, Khang, and Krishnamurthy (2010), documenting that toxic MBS were always on the balance sheet of financial intermediaries during the 2008 financial crisis. Haddad and Muir (2021) provide further evidence suggesting that intermediaries are responsible for a large fraction of risk premium variation in various asset classes. This also allows me to sidestep "distributive" externalities (Dávila and Korinek 2018) that can lead to under- as well as over-borrowing in the rational benchmark.

Constraints: Financial intermediaries' constraints are then as follows:

$$c_1 + c(H) + q_1 h_1 \leq e_1 + d_1 + q_1 H \tag{11}$$

$$c_2 + d_1(1 + r_1) + q_2 h_2 \leq d_2 + (z_2 + q_2) h_1 \tag{12}$$

$$c_3 + d_2(1 + r_2) \leq z_3 h_2 \tag{13}$$

$$d_2 \leq \phi h_2 \tilde{\mathbb{E}}_2[z_3] \tag{14}$$

where H is the quantity of the asset created at $t = 1$, h_1 is the quantity intermediaries keep on their balance sheet at the end of period $t = 1$, and h_2 is the quantity of the risky asset held by financial intermediaries at time $t = 2$. In equilibrium, $h_1 = h_2 = H$ since households cannot hold the asset, and all intermediaries are identical. Financial intermediaries have an endowment e_1 in the initial period.[7]

Throughout the paper, I make use of the marginal utility of consumption of financial intermediaries, $\lambda_t = 1/c_t$ in period 1 and 2, while $\lambda_3 = 1$ in the last period because of linear utility. A key object of interest, as in most models with financial frictions, is the *net worth* n_2 of financial intermediaries at $t = 2$, defined as:

$$n_2 = z_2 H - d_1(1 + r_1). \tag{15}$$

2.1.3 Discussion of the Environment

Financial intermediaries should be interpreted as levered financial institutions that are using short-term debt: banks, insurance companies, hedge funds, brokers, etc. A favored interpretation is that H represents Mortgage-Backed Securities (MBS), complex products widely used in repo markets. When concerns about the future value of these assets arise, collateral values fall, forcing the banking system to cut back on other activities in order to roll-over its short-term debt. The model thus seeks to capture a typical "run on repo" as the panic of 2007-2008 (Gorton and Metrick 2012).

[7] In order for financial intermediaries to always be able to repay their debt, I assume that ϕ is low enough such that $\phi \tilde{\mathbb{E}}_2[z_3] < \min z_3$ is satisfied.

The specific form of the collateral constraint assumes that short-term debt is collateralized by the future cash-flows of the risky asset. This allows me to fully isolate the effects of behavioral biases on welfare: despite the presence of financial frictions, the equilibrium is constrained-efficient when expectations are rational (Dávila and Korinek 2018). A large part of the normative macro-finance literature, for this reason, uses an alternative formulation for financial frictions to obtain pecuniary externalities (Farhi and Werning 2016, Bianchi and Mendoza 2018, Jeanne and Korinek 2019). A *collateral externality* indeed arises when the collateral constraint depends on the *current* price of the asset:

$$d_2 \leq \phi H q_2. \tag{16}$$

Without taking a stance on which one is more realistic, I focus on the future payoff constraint in equation (9) since it cleanly isolates the effects of behavioral biases, and show in Appendix B the robustness of the results to this alternative formulation.

2.2 Equilibrium

We now proceed to formally define the equilibrium, and analyze it in more details in the rest of this section.

Definition 3 (Equilibrium). *Given the behavioral biases functions Ω_2 and Ω_3, a sophistication parameter ζ, as well as initial values z_0, z_1, and q_0, an equilibrium consists of a real allocation $\{c_1, c_2(z_2), c_3(z_2, z_3), c_1^h, c_2^h(z_2), c_3^h(z_2, z_3), H\}$, prices $\{q_1, q_2(z_2)\}$, and biases $\{\Omega_2(\mathcal{I}_1), \Omega_3(\mathcal{I}_2(z_2))\}$, such that: (i) markets clear; (ii) agents maximize* (7) *and* (8) *subject to* (11)-(14)*; and (iii) behavioral biases are consistent with definitions 1 and 2.*

2.3 Equilibrium Analysis

Households: Households are passive throughout the three periods, and they pin down the rate of interest through their Euler equation: $\beta(1 + r_t) = 1$.

Financial Intermediaries at $t = 2$: Entering period $t = 2$ with a stock H of collateral assets, and debt d_1 to repay, financial intermediaries must decide on their borrowing and consumption levels. There are two separate cases.

No Crisis: When financial intermediaries are not constrained, their Euler equation simply sets consumption such that:

$$\lambda_2 = \frac{1}{c_2} = \tilde{\mathbb{E}}_2[\lambda_3] = 1 \tag{17}$$

because of the linearity of utility in the last period. The consumption level is thus independent of the price of the risky asset, and consequently of any behavioral bias. The price of the collateral asset is then simply given by $q_2 = \beta\tilde{\mathbb{E}}_2[z_3] = \beta\big(\mathbb{E}_2[z_3] + \Omega_3\big)$.

Financial Crisis: When the collateral constraint is binding, the associated Lagrange multiplier on the constraint, κ_2, is given by $\kappa_2 = \lambda_2 - 1 > 0$. This directly quantifies the severity of the crisis: it encodes how far we are from the unconstrained equilibrium. The asset price comes from intermediaries' maximization which yields:

$$q_2 = \beta c_2 \mathbb{E}_2[z_3 + \Omega_3] + \phi(1 - c_2)\mathbb{E}_2[z_3 + \Omega_3] \tag{18}$$

where the last term is a *collateral premium*. Consumption is coming from the budget constraint of financial intermediaries (12), since agents are against the collateral constraint:

$$c_2 = z_2 H - d_1(1 + r_1) + \phi H \mathbb{E}_2[z_3 + \Omega_3]. \tag{19}$$

This last expression makes clear that, unlike in the unconstrained case, behavioral biases have direct effects on real allocations in crises. Pessimism ($\Omega_3 < 0$) reduces the amount households are willing to lend to financial intermediaries, leading to a one-for-one fall in their consumption level c_2. I distinguish the cases when Ω is exogenous or endogenous to clarify their differences, which will be crucial for welfare.

Exogenous Bias: When Ω_3 is exogenously set, the budget constraint equation is sufficient to obtain the consumption level in a crisis (as in a rational benchmark). Sentiment simply shifts consumption by a constant relative to the REE benchmark. It also has an effect on asset prices through the stochastic discount factor and the expectation of future dividends. But this drop in asset prices does not spill back to consumption, which is pinned down independently.

Endogenous Bias: When the behavioral bias Ω_3 depends on equilibrium prices q_2, however, the budget constraint is not enough anymore to determine the consumption level of financial intermediaries in a crisis. The equilibrium now requires solving for a fixed-point between the budget constraint and the pricing equation. This is a manifestation of *belief amplification*.[8] Intuitively, a fall in net worth causes a fall in current consumption. This decreases the stochastic discount factor used by agents to price the risky asset, which in itself creates endogenous pessimism. This leads the price of the asset to fall further, which tightens the borrowing constraints of financial intermediaries by aggravating pessimism, and in turn creates a further fall in the price that leads to more pessimism.[9]

Financial Intermediaries at $t = 1$: The consumption Euler equation for financial intermediaries in the initial period is simply given by:

$$1 = \tilde{\mathbb{E}}_1 \left[\frac{\lambda_2}{\lambda_1} \right] \tag{20}$$

since financial intermediaries and households have the same time-preference parameter β. Because consumption inside a crisis, c_2, depends directly on z_2, agents with an optimistic bias $\Omega_2 > 0$ expect their future consumption to be higher than in reality. Accordingly, their Euler equation directly implies that an optimistic bias translates into overconsumption at $t = 1$ relative to the rational benchmark, financed by borrowing (so a higher d_1). Similarly, the gap between expected and realized consumption is driven, for the case of unsophisticated agents, by *future* sentiment, since an $\Omega_3 < 0$ at $t = 2$ leads to

[8] In a setup where the collateral constraint depends on current prices q_2, this belief amplification channel compounds the traditional *financial amplification* mechanism. See Appendix B.

[9] While the equilibrium is ensured to be unique in the exogenous sentiment case, this is not immediate anymore for endogenous sentiment. Appendix A.9 shows how linear forms of price extrapolation also guarantee the uniqueness of a stable equilibrium. Complex non-linear forms of endogenous sentiment can however lead to multiple equilibria. Since this is not the focus of this analysis, for the rest of the paper I assume that belief distortions are not strong enough such that equilibrium uniqueness is guaranteed.

a tighter collateral constraint.

Finally, collateral creation is driven by the pricing equation of intermediaries:

$$q_1 = c'(H) = \beta \tilde{\mathbb{E}}_1 \left[\frac{\lambda_2}{\lambda_1} (z_2 + q_2) \right]. \tag{21}$$

3 Welfare Decomposition

This section describes the constrained planning problem of the planner: at $t = 1$, a social planner holding rational expectations realizes that agents are subject to behavioral biases. Taking these biases as well as general equilibrium effects into account, the planner seeks to find the allocation that maximizes welfare, knowing that the future allocation (at $t = 2$) will be determined by private agents (subject to a bias Ω_3).

3.1 Decomposition

I present a general decomposition in the spirit of Dávila and Korinek (2018), that fleshes out how a marginal increase in leverage, investment or prices leads to uninternalized welfare consequences, and classify the different channels. A key advantage of this approach is that the decomposition naturally determines which features of behavioral biases matter for financial stability, and need to be quantified by the regulator. Two classes of effects appear in the following proposition, which I then explore in turns.

Proposition 1 (Uninternalized Effects). *The uninternalized first-order impact on welfare, when infinitesimally varying one aggregate variable while keeping the others constant, are given by:*

i) For short-term debt d_1:

$$\mathcal{W}_d = \underbrace{\left(\tilde{\mathbb{E}}_1[\lambda_2] - \mathbb{E}_1[\lambda_2] \right)}_{\mathcal{B}_d} - \underbrace{\mathbb{E}_1 \left[\kappa_2 \phi H \frac{\boldsymbol{d\Omega_3}}{\boldsymbol{dq_2}} \frac{dq_2}{dn_2} \right]}_{\mathcal{C}_d}; \tag{22}$$

ii) For investment in collateral assets H:

$$\mathcal{W}_H = \underbrace{\left(\beta \mathbb{E}_1 \left[\lambda_2 (z_2 + q_2)\right] - \lambda_1 q_1\right)}_{\mathcal{B}_H} + \underbrace{\beta \mathbb{E}_1 \left[\kappa_2 \phi H \frac{\boldsymbol{d\Omega_3}}{\boldsymbol{dq_2}} \left(\frac{dq_2}{dn_2} z_2 + \frac{dq_2}{dH}\right)\right]}_{\mathcal{C}_H}; \tag{23}$$

iii) And for prices q_1:

$$\mathcal{W}_q = \beta \mathbb{E}_1 \left[\kappa_2 \phi H \frac{\boldsymbol{d\Omega_3}}{\boldsymbol{dq_1}}\right]. \tag{24}$$

Proof. All proofs are provided in Appendix A. □

3.2 Behavioral Wedges

The first type of effects (the terms $\mathcal{B}_d$ and $\mathcal{B}_H$ of equations 22 and 23) are *behavioral wedges*, as in Farhi and Gabaix (2020). They quantify *differences in beliefs* between the planner and the market. Take the behavioral wedge for leverage for instance. Its strength is driven by the difference in the expected severity of crisis: when optimistic agents expect a crisis, they expect to withstand it with stronger capital buffers thanks to a payoff $z_2 + \Omega_2$ on their holdings of risky assets, rather than just z_2. Because of the strong non-linearity of the model, the behavioral wedge is a complex object. Nonetheless, an infinitesimal perturbation around the REE is enlightening (assuming Ω_2 and Ω_3 are small state-by-state):

Proposition 2 (Behavioral Wedge Approximation)**.** *If Ω_2 and Ω_3 are small state-by-state, the behavioral wedge for short-term debt, $\mathcal{B}_d$, can be approximated by:*[10]

$$\mathcal{B}_d \simeq \underbrace{-\boldsymbol{\Omega_2} H \mathbb{E}_1 \left[\boldsymbol{\lambda_2^2} \mathbb{1}_{\kappa_2 > 0}\right]}_{(i)} + \underbrace{\phi H (\boldsymbol{1} - \boldsymbol{\zeta}) \mathbb{E}_1 \left[\boldsymbol{\Omega_3} \boldsymbol{\lambda_2^2} \mathbb{1}_{\kappa_2 > 0}\right]}_{(ii)}. \tag{25}$$

The first term quantifies the welfare losses from contemporaneous irrationality at $t = 1$. It is negative when Ω_2 is positive, naturally implying that an additional unit of leverage

[10] The same first-order analysis for investment can be found in Appendix A.4.

is costly when agents are over-optimistic. Importantly the bias is multiplied by a measure of the expected severity of a future financial crisis, outlining that what affects welfare is not simply deviations from rationality, but their *interaction* with financial frictions.

The second term quantifies welfare changes emanating from the predictable behavior of future deviations from rationality, Ω_3. Once again, predictable pessimism in the future is not enough to generate first-order welfare losses: this term is non-zero only when the *product* of sentiment with marginal utility in a crisis is non-zero. In other words, it is the *comovement* of irrationality with the health of financial intermediaries that is a cause of concern for the planner (see Section 5.1 for suggestive evidence regarding this covariance). An interesting case in point of equation (25) is that even if $\Omega_2 = 0$, welfare losses are possible because of the predictable behavior of *future* irrationality. Even if, on average, there is no deviation from rationality (i.e. $\mathbb{E}_1[\Omega_3] = 0$), the possible covariance of Ω_3 with the health of financial intermediaries, λ_2, creates a welfare loss from increasing leverage in period $t = 1$. This implies that is it not necessary for the social planner to know the current state of irrational exuberance to be justified to act preemptively: knowing that agents will be pessimistic in bad states of the world is enough. This insight, however, heavily depends on the degree of sophistication of agents. Indeed, equation (25) makes clear that the second part of this behavioral wedge disappears when agents are sophisticated ($\zeta = 1$): in this instance, agents realize that the crisis will be worse because of over-pessimism in the future, and thus lower their leverage in the initial period accordingly.

3.3 Externalities

The second class of effects are externalities that works through *future (crises) beliefs* and *prices*. They are operative even though, as explained earlier, there is no externality in the rational benchmark. Even more surprisingly, they are effective even in the case of fully sophisticated agents as can be seen from the absence of ζ in these expressions. Let us examine in detail the terms composing this externality for leverage:

$$\mathcal{C}_d = -\mathbb{E}_1\left[\kappa_2 \phi H \frac{\boldsymbol{d\Omega_3}}{\boldsymbol{dq_2}} \frac{dq_2}{dn_2}\right]. \tag{26}$$

The first term is the Lagrange multiplier κ_2, again indicating that welfare losses are present only in the event of a binding financial friction at $t = 2$. The term ϕH then corresponds to the fact that this externality operates at the level of the friction that limits borrowing at $t = 2$. The derivative dq_2/dn_2 quantifies the change in asset prices implied by the change in short-term debt at $t = 1$: taking on more leverage mechanically lowers net worth in the future, which impacts equilibrium prices in the future (see Section 2). For now, all of these terms also exist in a rational world. The bold term, however, is specific to behavioral distortions and is thus zero in a rational counterfactual, making the expression zero in total. It measures how sentiment *inside* a financial crisis changes when equilibrium prices change.

This externality can be intuitively described as follows. Agents fail to internalize that, by increasing their leverage in good times, they lower asset prices tomorrow, which can make everyone in the economy more pessimistic. This pessimism, in turn, tightens the collateral constraint of financial intermediaries, preventing them to roll-over their debt as desired, and aggravates the financial crisis.

For this externality to exist it is necessary for the belief derivative to be non-zero. In other words, the collateral externality is operative if and only if behavioral biases at $t = 2$ are a direct function of equilibrium prices at $t = 2$. This means, for example, that any fundamental-based behavioral bias as in (3) will not feature such a market failure. In the natural benchmark of price extrapolation, as in (4), this derivative is simply $d\Omega_3/dq_2 = \alpha_q > 0$. This externality is then pushing towards excessive borrowing.[11]

More surprisingly, the collateral externality for investment in Proposition 1, $\mathcal{C}_H$, can go in the opposite direction as the leverage one. Agents are not taking into account how a supplementary unit of collateral, by raising net worth next period, can support asset

[11] Notice that when this externality exists because of endogenous sentiment, the price sensitivity dq_2/dn_2 that enters this expression is also magnified by *belief amplification*. A change in net worth in period $t = 2$ impacts equilibrium asset prices as:

$$\frac{dq_2}{dn_2} = \frac{(\beta - \phi)\mathbb{E}_2[z_3 + \Omega_3]}{1 - (\phi + (\beta - \phi)(2c_2 - n_2))\frac{d\Omega_3}{dq_2}}. \tag{27}$$

A positive change in net worth leads to a change in price through the stochastic discount factor c_2, which itself can alleviate pessimism, supporting asset prices in a feedback loop. This makes the price more sensitive to changes in net worth when $d\Omega_3/dq_2 > 0$.

prices and thus consequently reduce pessimism. In turn, this ameliorates the borrowing capacity of the whole economy, thus improving welfare.[12] Irrational exuberance thus helps overcome the under-investment problem coming from financial frictions.[13] In a model with a collateral constraint directly featuring q_2, the rational benchmark features such a positive collateral externality. Irrational exuberance thus helps to alleviate this market failure. The welfare impact of an additional unit of investment is however unambiguously negative for large enough Ω_2, since the behavioral wedge can be unboundedly negative.

The last part of Proposition 1 highlights another effect. In most models (rational models or models with exogenous sentiment) the two uninternalized effects for leverage and investment are enough to characterize efficiency. Indeed, once allocations are fixed the equilibrium level of prices has no effect on welfare. The problem is different, however, in the presence of endogenous sentiment since a new state-variable enters the optimal policy problem: the equilibrium level of asset prices today can enter the determination of *future* allocations, and thus the expected level of welfare:

$$\mathcal{W}_q \quad = \quad \beta \mathbb{E}_1 \left[\kappa_2 \phi H \frac{\boldsymbol{d\Omega_3}}{\boldsymbol{dq_1}} \right]. \tag{28}$$

This effect works through the interaction of financial frictions, *past* asset prices, and sentiment in a crisis. The intuition for this term is as follows. When private agents push up the price of the asset today, it might influence the formation of behavioral biases in the future. This is represented by the term $d\Omega_3/dq_1$. Typically, in the illustrative price extrapolation case where $\Omega_3 = \alpha(q_2 - q_1)$, this derivative is equal to $-\alpha$, a negative term. This change in sentiment at time $t = 2$ impacts the collateral limit for short-term debt d_2, in proportion to ϕH, a positive quantity. It then impacts welfare if agents are against their borrowing constraint, i.e. if $\kappa_2 > 0$, since it directly alters the amount they can borrow. Succinctly, when agents bid up prices, this can feed pessimism tomorrow by increasing the anchor agents use to form expectations: an increase in prices today will cause a rever-

[12] In models where assets can change hands (Lorenzoni 2008) the price of the asset is decreasing in the aggregate quantity, since outside investors are usually assumed to have a concave production function.
[13] This is reminiscent of Martin and Ventura (2016), where bubbles help alleviate financing frictions.

sal tomorrow. I thus call this effect a *reversal externality*. This new force is independent of the current extent of behavioral biases in the initial period, and has important implication for the conduct of policy as I show next.[14]

Finally, these externalities are robust to the degree of sophistication of agents, as can be seen from the absence of ζ in these expressions. As for regular externalities, agents do not internalize that their decisions at $t = 1$ can influence the determination of sentiment at $t = 2$ if sentiment depends on equilibrium variables like prices. Take the collateral externality for example. Market's participants can be sophisticated enough to realize that low prices in a future crises will create endogenous pessimism and tighten collateral constraints. But they cannot coordinate to reduce their aggregate leverage at $t = 1$, in order to sustain asset prices in a future bust and attenuate future pessimism. Only the regulator can internalize these effects, even though the regulator and the market share the same beliefs about how sentiment will manifest itself in the future.

Remark 1 (Intermediaries vs. Households Beliefs). For simplicity and conciseness, beliefs are assumed to be homogeneous through the paper. Appendix D derives the same welfare decomposition as in Proposition 1 when financial intermediaries and households have different beliefs. It highlights that, while intermediaries' beliefs are what matters during the boom, the beliefs of households are the ones shaping the welfare losses during the bust. This is because the tightness of the collateral constraint directly depends on households beliefs about the future payoffs of the collateral (see Simsek (2013) for an in-depth analysis of belief disagreement under collateral constraints).

[14] Similarly, Schmitt-Grohé and Uribe (2016) show that with downward wage rigidity, past wages become a relevant state variable, motivating the planner to reduce real wages in booms.This effect is also present in Farhi and Werning (2020). In their model, a high price in the initial period can translate into over-pessimism when the ZLB hits. This creates a downward pressure on prices at the ZLB, which affects aggregate demand via a wealth effect.

4 Optimal Policy

4.1 Constrained Efficiency

I can now characterize the allocation the planner would like to implement in the presence of sentiment, when solving the constrained planning problem. A planner subject to the same constraints as agents, with prices determined by market-clearing as in the decentralized case, evaluates welfare using its own expectations. In particular, any direct intervention at $t = 2$ is proscribed.[15] The planner thus takes the previous uninternalized effects into account. These objects are crucial to characterize optimal policy in this setting.

In order to achieve the second-best, the planner makes agents internalize their uninternalized welfare effects. This is done by choosing taxes or subsidies that exactly cancel out the uninternalized effects described in Proposition 1.

Proposition 3 (Second-Best Policy)**.** *The social planner achieves the constrained-efficient second-best with three instruments defined by:*

1. *A tax $\tau_d = -\mathcal{W}_d / \lambda_1$ on short-term borrowing ;*
2. *A tax $\tau_H = -\mathcal{W}_H / (\lambda_1 q_1^*)$ on the creation of collateral assets ;*
3. *A tax $\tau_q = \frac{q_1 - q_1^*}{q_1^*}$ on the holding of collateral assets*

where λ_1 is the marginal utility of financial intermediaries at time $t = 1$ evaluated at the desired allocation, q_1 is the price that would arise through market-clearing at the desired allocation without the holding tax, and q_1^* is the price such that $W_q = 0$ when evaluated at the desired allocation.

Proposition 3 is rather abstract, but makes two simple points. First, the calibration of macroprudential policy should be done by focusing on the key aspects of sentiment

[15] The concept of constrained efficiency also restricts the analysis to a planner who takes financial frictions as given, following Hart (1975), Stiglitz (1982) and Geanakoplos and Polemarchakis (1985). It can be understood as answering the following question: can a planner subject to the same constraints as private agents improve on the market outcome? An earlier version (Fontanier 2022) allows the planner to intervene in crises with bailouts.

driving the uninternalized effects from the previous part, $\mathcal{W}_d$ and $\mathcal{W}_H$. Second, when current asset prices impact future sentiment, three instruments are needed to achieve the second-best, and not only two.[16]

How can one interpret the results of Proposition 3 in terms of real-world policy? The optimal taxes on debt and investment correspond to the usual instruments in the macroprudential toolkit: capital requirements and Loan-to-Value (LTV) restrictions (Claessens 2014). This is not the case for the tax on holdings, designed to influence equilibrium price.[17] I now explore the concrete policy lessons coming out of the analysis.

4.2 Implementation

Counter-cyclical Capital Buffers: The tax on short-term borrowing can naturally be interpreted as capital structure regulation. Proposition 3 thus provides the financial regulator with the features of behavioral biases that are necessary to quantify in order to optimally calibrate leverage restrictions. Because Ω_2 is a largely volatile object (Bordalo, Gennaioli, Porta, and Shleifer 2024), the optimal value of this macroprudential leverage tax is also time-varying. But importantly, the time-variation in τ_d should not only track Ω_2, but also take into account the expected *future* realizations of Ω_3 (if agents are not fully sophisticated) as well as the expected impact of future prices on Ω_3.[18]

[16] Achieving the second-best does not imply that the benchmark rational level of welfare is attainable. This is because, while the effects of over-optimism can be fully resorbed by appropriate prudential regulation, the presence of pessimism in crises makes intermediaries worse off even with optimal policy compared to the rational benchmark, since no intervention is possible at $t = 2$ to directly increase borrowing. The numerical exercise in Section 5.3 suggests that this gap between the rational and constrained-efficient allocations can be substantial.

[17] Conceptually, another instrument is needed because biases are driven by an equilibrium variable that falls outside the scope of traditional macroprudential tools. In that sense, the role of prices is not crucial: one could imagine an alternative where biases are a function of an equilibrium *quantity* not affected by traditional tools, and another instrument would be needed to specifically target it. I focus on prices in this paper as there is extensive empirical evidence of price extrapolation (e.g. Greenwood and Shleifer 2014).

[18] Leverage limits in this framework are also more robust than leverage taxes to fluctuations in irrational exuberance: small movements in the behavioral bias around a positive Ω_2 lead to smaller welfare losses when a leverage limit is imposed rather than leverage tax. Time-variation in the leverage limit is nevertheless still required as long as the planner's estimate of Ω_3 given the information available is time-varying.

LTV Regulation: The second tax in Proposition 3 directly aims at regulating the *quantity* of risky investments. For this reason, this policy can be interpreted as loan-to-value (LTV) regulation. Importantly, the welfare analysis highlights again that the optimal LTV limit is time-varying, tracking the same behavioral biases as do leverage restrictions.

The crucial difference with capital requirements lies in the time-variation required by variation in the expected impact of prices on sentiment. When the regulator is concerned that a future crash in prices will result in a greater sensitivity of sentiment with respect to prices in a crisis (all else equal), the optimal reaction is to tighten leverage restrictions more but to *relax* LTV ratios. Indeed, as explained in Section 3.3, the collateral externality for H calls for *higher* investment than in the decentralized equilibrium, in order to alleviate pessimism during crises by strengthening the net worth of the financial sector.

Price Regulation: The third tax in Proposition 3 does not have a simple relation to the current macroprudential toolbox, however. This underlines the need for an additional instrument that complements traditional macroprudential policy. From an abstract perspective, this instrument can be modeled as a tax on asset holdings. This seems rather unrealistic to implement, however. A more natural candidate for this instrument is to use monetary policy, which is studies formally in Appendix C. Leaning against the wind lowers *current* asset prices, which will then cure *future* pessimism in a possible crisis – a new channel for monetary policy. Furthermore, such action does not necessarily require information about contemporaneous biases. A sharp increase in asset prices could be entirely due to fundamentals, but the planner can still have an incentive to make prices deviate from their rational value today to prevent irrational distress from happening. Finally, implementing such a policy allows for financial regulation to be adapted and *relaxed*. Indeed, by acting preventively the central bank makes the future realizations of pessimism less severe, thus directly reducing the size of behavioral wedges and of the collateral externality. Taking this into account leads the optimal macroprudential limit to be less strict, which raises welfare. The numerical exercise in Section 5.3 suggests that this complementarity can be substantial.

4.3 Sentiment Uncertainty

I so far assumed that the social planner had perfect information about behavioral biases. In a famous speech on asset price bubbles, Bernanke (2002) discussed the "identification problem" that naturally arises once the financial stability authority contemplates a proactive approach to bubbles. A natural question of practical importance is then whether my results are impaired in the presence of imperfect knowledge about behavioral biases.

The short answer is: no, to the contrary. Sentiment uncertainty reinforces motives for preventive action, in contrast with Brainard (1967)'s "attenuation principle." While recognizing that identifying a bubble is intrinsically difficult, this section shows that the widespread intuition that this uncertainty calls for more *laissez-faire* is actually erroneous.

I study the case where the regulator is uncertain about the level of irrational exuberance. To this end, I leverage the previous equilibrium and welfare analysis. Ω_3 is assumed to be certain and constant in the future.[19] Recall that private agents are shifting the entire distribution of future dividends by Ω_2, believing that dividends will be $z_2 + \Omega_2$ instead of z_2. The following assumption make the analysis more convenient:

Assumption 2. *All parameters of the model and of the probability distribution $F_2(z)$ are common knowledge to private agents and the social planner, except possibly for its mean, $\bar{z}_2$. Additionally, q_1 is strictly increasing in $\bar{z}_2$.*

This imply that, in the absence of sentiment, the social planner could simply infer the value of $\bar{z}_2$ by looking at equilibrium prices. We now assume that the social planner's prior over sentiment is given by a uniform distribution:

$$w \sim \mathcal{U}\left[\bar{\Omega}_2 - \sigma_\Omega, \bar{\Omega}_2 + \sigma_\Omega\right] \tag{29}$$

where $\bar{\Omega}_2$ is the average level of sentiment according to the planner's prior, and σ_Ω controls the amount of uncertainty around it. By observing asset prices the planner can infer what agents believe the average future dividend is, so that the planner's posterior mean

[19] The analysis for Ω_3-uncertainty is presented in Appendix A.7. The results are identical.

regarding future dividends becomes $\bar{z}_2 = g_q^{-1}(q_1) - \bar{\Omega}_2$. Taking the uncertainty in its posterior into account, the planner first-order condition for short-term debt is now given by:

$$u'(c_1) = \frac{1}{2\sigma_\Omega} \int_0^\infty \left[\int_{-\sigma_\Omega}^{\sigma_\Omega} \frac{\partial \mathcal{W}_2}{\partial n_2} \left(d_1, H; q_2, z_2 - \bar{\Omega}_2 - \omega_2\right) d\omega_2 \right] dF_2(z_2) \tag{30}$$

This expression contains all of the intuition for how sentiment uncertainty can reinforce or weaken the need for preventive leverage tightening. Once deducing the average behavioral error $\bar{\Omega}_2$, the planner is uncertain about the exact distribution of the state of the world next period. It thus takes the distribution that agents use, but factors in the noise it attributes to their expectations. This leads the social planner to consider, for each realization z_2, all values inside the segment $[z_2 - \sigma_\Omega, z_2 + \sigma_\Omega]$ as equally likely.

If expanding the set of possible behavioral biases, by increasing σ_Ω, increases the value of the expectations term, it means that uncertainty increases expected marginal utility. This implies that the social planner wishes to reduce the leverage of agents today to get back to the optimality condition, by increasing initial marginal utility $u'(c_1)$, and by diminishing expected future marginal utility. Conversely, if enlarging the possible values of ω_2 decreases expected marginal utility, the social planner should relax leverage constraints compared to the absolute certainty case. The following Proposition expresses that uncertainty about behavioral biases unambiguously calls for precautionary restrictions.

Proposition 4 (Ω_2-Uncertainty and Leverage Restrictions)**.** *If the social planner believes that the behavioral bias at $t = 1$ can be expressed as $\bar{\Omega}_2 + \omega$, where ω is uniformly distributed on $[-\sigma_\Omega, \sigma_\Omega]$, and Ω_3 is constant state-by-state at $t = 2$, then the optimal leverage tax is increasing in uncertainty σ_Ω, and the optimal investment tax is decreasing in σ_Ω.*

The proof is rather involved (see Appendix A.6), but the intuition can be understood succinctly. The key is to notice that marginal welfare is a *convex* function. Intuitively, sentiment uncertainty adds terms to the expectation computed by the planner relative to the private solution, but the parts coming from intermediaries' relative optimism are more costly than the ones coming from relative pessimism.[20] The strong non-linearities

[20] The same insights can be obtained if we were to consider uncertainty about the extent of sentiment *inside* a

associated with the interaction of sentiment with financial frictions make it attractive to tighten capital requirements in the face of uncertainty. Using the words of Yellen (2009), a "type 1" error is simply much less costly than a "type 2" error.

Remark 2 (Uncertainty and H-Regulation). Proposition 4 also highlights that investment regulation behaves in a different way: an increase in σ_Ω calls for more investment in H in the planner's problem relative to the private solution. This is because increasing uncertainty increases the incentive to shift consumption to the next period. Indeed, if there is a risk that agents are extremely over-optimistic and that a crisis will be extremely severe, it is even more valuable to hold an asset paying dividends, albeit low, in this state of the world. Concretely, this means that in times of heightened uncertainty, the regulator should tighten counter-cyclical capital buffers but at the same time relax LTV ratios.

5 Applications and Empirical Relevance

5.1 The Importance of Irrational Pessimism

A perhaps surprising lesson of the model is that many of the effects of behavioral biases on welfare are actually coming from the presence of Ω_3. Indeed, a majority of the effects presented in Proposition 1 operates through the interaction of irrationality with the health of financial intermediaries during crises.

One way to highlight this new result is to place ourselves at the REE constrained-efficient allocation, where the planner has no reason to intervene. If we add an infinitesimal degree of irrationality, which forces cause first-order welfare losses? The answer comes by inspecting equations (22), (23), and (24). At the rational expectations constrained-efficient equilibrium, behavioral wedges are zero, so the only effects left are the collateral externalities and the reversal externality, expressions (26) and (28). As explicated earlier, these are only present when *future* sentiment is impacted by *current* and *past* asset prices,

financial crisis. Endogenous sentiment, for example in the form of price extrapolation, amplifies this effect by adding more curvature. Additionally, uncertainty around the *strength* of the extrapolation mechanism reinforces the need for price regulation. These results are presented in Appendix A.7.

and there is a positive probability of a crisis in the future.

The fundamental intuition behind this result is that small changes in leverage due to fluctuating sentiment are not harmful to the first-order since agents are on the objective Euler equation. But anything that directly impacts the tightness of the collateral constraint in a crisis, where agents are *not* on their Euler equation, has a first-order impact on welfare by aggravating financial turmoil.[21]

5.2 Relevant Empirical Moments for Policymakers

These results draw attention to irrational distress during financial crises, while the literature has mostly focused on irrational exuberance during the build-up leading to the crash. It also provides guidance to policymakers on what precise information and moments about behavioral biases should be collected and monitored. In addition to the level of irrational exuberance, my theoretical analysis identifies four key properties of beliefs that need to be quantified by the regulator:

1. $\boldsymbol{Cov(\Omega_3, \lambda_2)}$: Proposition 2 underlines that the optimal level of macroprudential policy depends on the covariance between the predictable amount of irrational pessimism and the balance sheet health of financial intermediaries. Suggestive evidence supports the assumption that the two objects Ω_3 and $\lambda_2 \mathbb{1}_{\kappa_2>0}$ negatively comove: figure 1 uses two proxies to construct time series for Ω_3 and for λ_2, and shows how Ω_3 is consistently negative in crises.[22]

2. $\boldsymbol{\frac{d\Omega_3}{dq_2}}$: Does the magnitude of irrational pessimism depend on the level of asset prices inside a crisis? If yes, this is a first-order effect for the calibration of macroprudential policy. For instance, McCarthy and Hillenbrand (2022) estimate that between 19% and 33% of mispricing movements in the stock market can be attributed to price

[21] Irrational exuberance is of course also costly, as it triggers more frequent credit crunches. It is also possible that irrational distress is a direct function of past optimism, creating the same kind of reversal externality, but the first-order damages to welfare would not be directly attributable to irrational exuberance either. There is also the possibility that over-optimism has other effects on investment in the real sector, which can be costly as in Rognlie, Shleifer, and Simsek (2018).

[22] See also Bordalo et al. (2018) and Maxted (2024) for other examples of over-pessimism. An earlier version (Fontanier 2022) also presents other alternatives to measure sentiment, leading to the same result.

extrapolation rather than fundamental extrapolation. It remains an open question whether this also applies to assets used as collateral by financial intermediaries.

3. $\frac{d\Omega_3}{dq_1}$: Does the magnitude of irrational pessimism depend on the equilibrium level of asset prices *before* the bust? While some form of price extrapolation is a necessary ingredient here to generate a role for policy, it is not sufficient. To counter the Reversal externality fleshed out in (28), the planner must act to decrease prices in order to predictably lessen the amount of irrational pessimism in a future crisis. The planner must thus understand if agents mechanically extrapolate from equilibrium prices or from the component of prices that is invariant to policy. Whether a monetary tightening, for instance, causally impact the future magnitude of irrational pessimism is an open question. A quantification of this object is necessary to know whether monetary policy should lean against the wind.

4. σ_Ω: Given the planner's estimate of behavioral biases at a given time, the width of the confidence interval around that estimate also matters for optimal policy. Given the model of expectation formation preferred by the planner to explain behavioral biases, how much of the realized variation in forecast errors of private participant can be explained by this model? The lower this share, the more important it is to factor in this uncertainty. As shown in Section 4.3, this uncertainty is then compounded by the degree of non-linearities in financial crises.

5.3 Numerical Illustration

For conventional models of expectation formation, how large are likely to be these effects? This section proposes a numerical exercise to illustrate these results. Since the model is stylized and designed for a closed-form normative analysis, this exercise is meant as an illustration rather than a full-fledged calibration. I use standard values for the discount factor ($\beta = 0.98$) and risk aversion ($\gamma = 2$). I calibrate the financial friction parameter to 0.75: an average value for margins required by the Federal Reserve when pledging risky

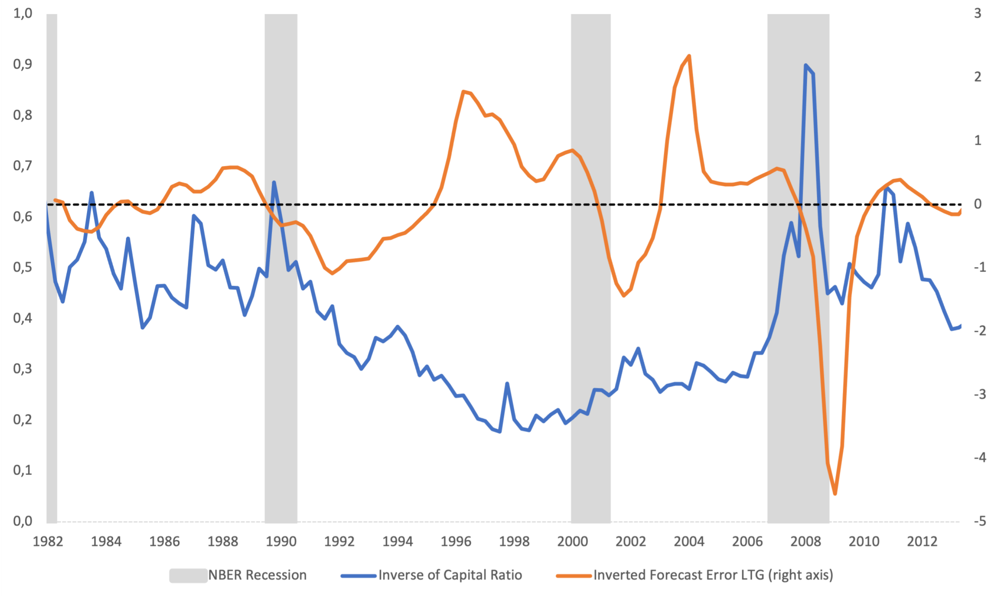

Figure 1: Time-series variation of proxies for λ_2 and Ω_3. For the financial health of intermediaries λ_2, I rely on He et al. (2017) which computes an intermediary capital ratio. The inverse of this capital ratio is proportional to λ_2 when agents have log-utility. For Ω_3, I use the inverted forecast errors made by stock market analysts on the long-term growth of stocks, a measure of Bordalo et al. (2024) which is directly constructed from survey data.

loans.[23] The dividend processes z_2 and z_3 are assumed to be following an i.i.d uniform distribution, and H is normalized to 1.

To close the calibration and match reasonable moments, I start by assuming that belief distortions are formed according to diagnostic expectations (Bordalo et al. 2018):

$$\Omega_3 = \theta \left(z_2 - \mathbb{E}[z_2] \right) \tag{31}$$

To be conservative about the role of behavioral biases in driving financial crises, I use for θ the estimate of Bordalo, Gennaioli, Ma, and Shleifer (2020), at the very-low end of values found in the literature.[24] Since the asset does not pay dividends at $t = 1$, Ω_2 is drawn randomly from the probability distribution of Ω_3 to ensure that both behavioral biases have the same magnitude on average. I then calibrate the mean and volatility of the

[23]See the margins tables at https://www.frbdiscountwindow.org/Pages/Collateral/collateral_valuation.

[24]Specifically, Bordalo et al. (2020) estimate an average diagnostic parameter of 0.5. With a persistence of fundamentals of 0.4 (Bordalo et al. 2018), this means that agents' reaction to dividend realization is roughly 20 percent larger than the REE benchmark. Other studies have found higher values for this parameter, closer to 1 (Bordalo, Gennaioli, La Porta, and Shleifer 2019 ; L'Huillier, Singh, and Yoo 2024 ; Bianchi, Ilut, and Saijo 2024). Despite this conservatism, the differences with the rational case are large.

probability distribution of dividends z to match two aggregate empirical moments. First, the frequency of financial crises in the data is between 4% and 6% (Taylor 2015). Second, the Intermediary Capital Ratio of intermediaries (proportional to λ_2^{-1}) fell by around 60% in 2008 (He et al. 2017). The model with diagnostic expectations is then calibrated to reproduce an average probability of crises close to 5%, and the fall in λ_2^{-1} to be 60% in a worst-case scenario.

Table 1 starts by showing the outcome of the simulation over all possible values for Ω_2 (so that investors are initially as likely to be optimistic as pessimistic). The unconditional probability of a financial crisis is 5.2%, comparing with only a 0.76% in the rational expectations case. Interestingly, Table 1 suggests that most of this increase in the frequency of financial crises is attributable to the presence of over-pessimism during bad times rather than over-optimism in good times, which echoes the previous discussion in Section 5.1.

Table 1: Exogenous Sentiment, Unconditional Moments

	Baseline	Rational	Only Ω_2	Only Ω_3	Sophistication
Probability of Crises (%)	5.2	0.76	1.5	5.3	4.3
Perceived Probability (%)	2.1	0.76	2.1	0.8	4.3
Severity of Crises (%)	-48.0	-37.9	-38.6	-47.6	-45.7
Optimal Change in Leverage (%)	-2.6	-	0.48	-3.0	2.0

Note: The third column estimates the moments, while imposing $\Omega_3 = 0$ in all states of the world. The fourth column imposes $\Omega_2 = 0$. The last column assumes that intermediaries are fully sophisticated ($\zeta = 1$) about the future bias Ω_3. The severity of crises is equal to the average decrease in λ_2^{-1}. The optimal change in leverage computes the required change in d_1 to implement the planner's solution.

From a normative perspective, Table 1 shows that the average change in leverage required by the rational planner is very small. This is because, as pointed out in Section 4.2, optimal policy becomes highly state-contingent. Table 2 thus presents the results of the numerical exercise when conditioning on a positive realization of the behavioral bias Ω_2 (in other words when intermediaries are over-optimistic ex-ante). As expected, crises are more frequent (probability of 6.7%) as well as more severe, but agents almost entirely neglect this risk. The optimal policy then requires a substantial decrease in the leverage of financial intermediaries, -10.1%, highlighting that macroprudential policy must be highly flexible over the cycle. It should also be noted that the optimal policy does not amount to implementing the REE allocation, as discussed in Section 4.1. The presence of eventual pessimism in bad times makes it optimal to implement a *lower* level of leverage for intermediaries than in the REE benchmark, since this directly worsens crises through the collateral constraint. Table 2 shows that this difference is not negligible: with behavioral biases, equilibrium borrowing is only 7.3% larger than in the REE benchmark.

Table 2: Exogenous Sentiment, Moments Conditional on $\Omega_2 \geq 0$

	Baseline	Rational	Only Ω_2	Sophistication	Optimal Policy
Probability of Crises (%)	6.7	0.76	2.7	5.9	4.6
Perceived Probability (%)	0.06	0.76	0.06	2.3	0.0
Severity of Crises (%)	-51.6	-37.9	-42.4	-49.8	-46.0
Optimal Change in Leverage (%)	-10.1	-	-7.3	-6.8	-
Change to REE Leverage (%)	-7.3	-	-7.3	-3.9	3.1

Note: The third column estimates the moments, while imposing $\Omega_3 = 0$ in all states of the world. The fourth column imposes $\Omega_2 = 0$. The last column assumes that intermediaries are fully sophisticated ($\zeta = 1$) about the future bias Ω_3. The severity of crises is equal to the average decrease in λ_2^{-1}. The optimal change in leverage computes the required change in d_1 to implement the planner's solution. The Change to REE leverage computes the average difference between the equilibrium d_1 and its rational counterfactual: a negative number indicates that leverage is higher than in the REE benchmark.

Finally, Table 3 performs the same exercise but instead condition on Ω_2 being two standard deviations above its zero-mean. The results are intuitive and expected: crises are more probable, more severe, and the optimal policy wants to rein in credit more. This echoes the crises predictability results surveyed in (Sufi and Taylor 2022): when credit growth rises one standard deviation above its mean, the expected crisis probability almost doubles relative to its baseline level. Overall, this numerical exercise suggests that the crises predictability indicators of this literature could be useful to serve as proxies for the optimal time-variation in macroprudential policy.

Table 3: Exogenous Sentiment, Moments Conditional on high Ω_2 (2 s.d.)

	Baseline	Rational	Only Ω_2	Sophistication	Optimal Policy
Probability of Crises (%)	8.4	0.76	4.6	7.6	4.6
Perceived Probability (%)	0.0	0.76	0.0	0.0	0.0
Severity of Crises (%)	-55.6	-37.9	-46.9	-53.3	-46.0
Optimal Change in Leverage (%)	-17.0	-	-14.4	-13.3	-
Change to REE Leverage (%)	-14.4	-	-14.4	-10.5	3.1

Note: The third column estimates the moments, while imposing $\Omega_3 = 0$ in all states of the world. The fourth column imposes $\Omega_2 = 0$. The last column assumes that intermediaries are fully sophisticated ($\zeta = 1$) about the future bias Ω_3. The severity of crises is equal to the average decrease in λ_2^{-1}. The optimal change in leverage computes the required change in d_1 to implement the planner's solution. The Change to REE leverage computes the average difference between the equilibrium d_1 and its rational counterfactual: a negative number indicates that leverage is higher than in the REE benchmark.

To place these numbers in the context of the broader macroprudential literature, it is useful to look at (rational) models that feature collateral externalities. Bianchi and Mendoza (2018) find that the social planner implements a debt level that is approximately 10% lower than the decentralized equilibrium before a financial crisis hits (see Figure 1 in Bianchi and Mendoza 2018). My numerical illustration thus suggests that the optimal policy aimed at taming behavioral financial fragility has a similar order of magnitude. A crucial difference with this literature, however, is that in my model the optimal policy fails to bring the probability of crises close to zero (still elevated at 4.6%). This is because of the presence of Ω_3 in bad times, which ex-ante policy cannot solve. This stands in sharp contrast with conventional macroprudential models, such as Bianchi (2011) and Bianchi and Mendoza (2018), where the social planner cuts the probability of a financial crisis more than tenfold.

The above analysis was conducted under the assumption that sentiment is driven by fundamentals, a case where the literature gives us empirical guidance on the size of the behavioral biases. The analytical results of Section 3.3, however, emphasized that other welfare consequences can arise when sentiment is instead driven by prices. I thus repeat the exercise, using instead price extrapolation as in (4), calibrated to match the same unconditional crisis probability from Taylor (2015). The results are presented in Table 4.

Table 4: Endogenous Sentiment

	Unconditional	Rational	$\Omega_2 > 0$	High Ω_2 (2 s.d.)
Probability of Crises (%)	4.8	0.76	9.3	12.2
Perceived Probability (%)	2.1	0.76	0.1	0.0
Severity of Crises (%)	-39.9	-37.9	-54.0	-61.4
Optimal Change in Leverage (%)	0.5	-	-7.2	-14.4

Note: The third column estimates the moments, conditional on Ω_2 being positive at $t = 1$. The last column estimates the moments, conditional on Ω_2 being two standard-deviation higher than its unconditional mean at $t = 1$. The severity of crises is equal to the average decrease in λ_2^{-1}. The optimal change in leverage computes the required change in d_1 to implement the planner's solution.

Two results stand out. First, while the unconditional probability of crises is similar, it becomes higher once conditioning on over-optimism in good times. This is due to the specific interaction that appears between sentiment in periods 1 and 2 with price extrapolation: over-optimism boosts asset prices, making the reversal sharper. Similarly, and for the same reason, crises are more pronounced. At the same time, the optimal policy becomes *less* aggressive. While perhaps surprising, this result is intuitive: by lowering leverage, macroprudential policy has a second-round positive impact with price extrapolation. Entering a crisis period with higher net worth supports asset prices q_2, which in turn reduces the magnitude of over-pessimism Ω_3. The required reduction in leverage stays nevertheless substantial, at 14.8% when Ω_2 is two standard deviations above its mean.

Lastly, this numerical illustration can help us assess whether the reversal externality is a likely important channel for policy. In the spirit of Appendix C, I explore the impact of using interest rates to lean against asset prices. Nagel and Xu (2024) estimate that a 25 bp surprise interest rate increase leads to a 156 bp fall in asset prices. Table 5 shows the results of this experiment. By indirectly relaxing the collateral constraint in future crises, this allows the planner be less aggressive in forcing the intermediaries to reduce their leverage. Tightening substantially more, by a full 50 bp, has a more pronounced effect. Interestingly, it is still not enough to bring the crisis probability close to rational levels: in the case of high irrational exuberance, this probability is still elevated at 7.6%. While this suggests that leaning against the wind might not be the right tool to *avoid* the occurrence of behavioral financial crises, it still seems to be a powerful complement to traditional macroprudential policy. Fully assessing its optimal use would need a full-fledged calibration of the benefits of relaxing leverage limits, and to compare it to the real costs of a surprise monetary tightening.

Table 5: Monetary Policy Experiment

	Ω_2: 1 s.d.			Ω_2: 2 s.d.		
Monetary Tightening	0	25 bps	50 bps	0	25 bps	50 bps
Probability of Crises (%)	10.7	10.7	7.6	12.2	12.2	12.2
Optimal Change in Leverage (%)	-7.0	-3.3	-0.53	-14.4	-10.9	-7.4

Note: The first three columns condition on Ω_2 being one standard-deviation higher than its unconditional mean at $t = 1$, and the last three columns condition on Ω_2 being two standard-deviation higher. For each, I estimate the moments for three cases: (i) no change in the policy rate, (ii) a 25 bp increase, and (iii) a 50 bps increase. The optimal change in leverage computes the required change in d_1 to implement the planner's solution.

6 Conclusion

Should financial regulators and monetary authorities try to mitigate the potential instabilities associated with irrational booms and busts? In this paper I provide a framework that allows for the rigorous analysis of this crucial policy question. I showed how leverage, investment and price regulations can achieve constrained efficiency in the presence of behavioral biases, even in an environment that does not feature any externality in its rational benchmark. Some of the effects uncovered depend directly on beliefs being a function of equilibrium prices, and are robust to the degree of sophistication of agents.

While the model can be extended along several dimensions, the results suggest a need for research on two specific dimensions. First, while empirical research has convincingly demonstrated that overreaction is a pervasive feature of financial markets, we have less certainty about its drivers. My paper shows that understanding what drives deviations from rationality will simultaneously advance our comprehension of what policy can and should do to deal with financial bubbles. Second, in my model the small number of periods obfuscates the timing subtleties faced by regulators. But we have little understanding over the dynamic build-up of sentiment, and over which horizon it is influenced by monetary policy and asset prices. Further research is needed to fully grasp the complex timing interactions between policy, crises, and behavioral biases.

References

Barberis, N. (2018). Psychology-based models of asset prices and trading volume. In *Handbook of Behavioral Economics: Applications and Foundations 1*, Volume 1, pp. 79–175. Elsevier.

Barberis, N. and A. Shleifer (2003). Style investing. *Journal of Financial Economics 68*(2), 161–199.

Barberis, N., A. Shleifer, and R. Vishny (1998). A model of investor sentiment. *Journal of Financial Economics 49*(3), 307–343.

Barlevy, G. (2022). On speculative frenzies and stabilization policy.

Bastianello, F. and P. Fontanier (2024). Partial equilibrium thinking, extrapolation, and bubbles. *Working Paper*.

Bernanke, B. S. (2002). Asset-price "bubbles" and monetary policy. *Speech before the New York Chapter of the National Association for Business Economics*.

Bernanke, B. S. and M. Gertler (2000). Monetary policy and asset price volatility. *Working Paper*.

Bernanke, B. S. and K. N. Kuttner (2005). What explains the stock market's reaction to federal reserve policy? *The Journal of finance 60*(3), 1221–1257.

Bianchi, F., C. Ilut, and H. Saijo (2024). Diagnostic business cycles. *Review of Economic Studies 91*(1), 129–162.

Bianchi, J. (2011). Overborrowing and systemic externalities in the business cycle. *American Economic Review 101*(7), 3400–3426.

Bianchi, J. and E. G. Mendoza (2018). Optimal time-consistent macroprudential policy. *Journal of Political Economy 126*(2), 588–634.

Bordalo, P., N. Gennaioli, R. La Porta, and A. Shleifer (2019). Diagnostic expectations and stock returns. *The Journal of Finance 74*(6), 2839–2874.

Bordalo, P., N. Gennaioli, Y. Ma, and A. Shleifer (2020). Overreaction in macroeconomic expectations. *American Economic Review 110*(9), 2748–2782.

Bordalo, P., N. Gennaioli, R. L. Porta, and A. Shleifer (2024). Belief overreaction and stock market puzzles. *Journal of Political Economy 132*(5), 1450–1484.

Bordalo, P., N. Gennaioli, and A. Shleifer (2018). Diagnostic expectations and credit cycles. *The Journal of Finance 73*(1), 199–227.

Borio, C. E. and P. W. Lowe (2002). Asset prices, financial and monetary stability: exploring the nexus. *BIS working paper*.

Brainard, W. C. (1967). Uncertainty and the effectiveness of policy. *The American Economic Review 57*(2), 411–425.

Caballero, R. J. and A. Krishnamurthy (2003). Excessive dollar debt: Financial development and underinsurance. *The Journal of Finance 58*(2), 867–893.

Caballero, R. J. and A. Simsek (2020a). Prudential monetary policy. *MIT Working Paper*.

Caballero, R. J. and A. Simsek (2020b). A risk-centric model of demand recessions and speculation. *The Quarterly Journal of Economics 135*(3), 1493–1566.

Claessens, M. S. (2014). An overview of macroprudential policy tools. *IMF Working Paper*.

Dávila, E. and A. Korinek (2018). Pecuniary externalities in economies with financial frictions. *The Review of Economic Studies 85*(1), 352–395.

Dávila, E. and A. Walther (2021). Prudential policy with distorted beliefs. *Working Paper*.

DeLong, J. B., A. Shleifer, L. H. Summers, and R. J. Waldmann (1990). Positive feedback investment strategies and destabilizing rational speculation. *the Journal of Finance 45*(2), 379–395.

Drechsler, I., A. Savov, and P. Schnabl (2018a). Liquidity, risk premia, and the financial transmission of monetary policy. *Annual Review of Financial Economics 10*, 309–328.

Drechsler, I., A. Savov, and P. Schnabl (2018b). A model of monetary policy and risk premia. *The Journal of Finance 73*(1), 317–373.

Drechsler, I., A. Savov, and P. Schnabl (2022). How monetary policy shaped the housing boom. *Journal of Financial Economics 144*(3), 992–1021.

Fanelli, S. and L. Straub (2021, 03). A Theory of Foreign Exchange Interventions. *The Review of Economic Studies*.

Farhi, E. and X. Gabaix (2020). Optimal taxation with behavioral agents. *American Economic Review 110*(1), 298–336.

Farhi, E. and I. Werning (2016). A theory of macroprudential policies in the presence of nominal rigidities. *Econometrica 84*(5), 1645–1704.

Farhi, E. and I. Werning (2020). Taming a minsky cycle. *Working Paper*.

Fisher, I. (1932). *Booms and depressions: some first principles*. Adelphi Company New York.

Fontanier, P. (January 2022). Optimal Policy for Behavioral Financial Crises. *Harvard University Working Paper*.

Geanakoplos, J. (2010). The leverage cycle. *NBER macroeconomics annual 24*(1), 1–66.

Geanakoplos, J. D. and H. M. Polemarchakis (1985). Existence, regularity, and constrained suboptimality of competitive allocations when the asset market is incomplete. *Cowles Foundation Discussion Papers 764*.

Geithner, T. F. (2014). *Stress test: Reflections on financial crises*. Broadway Books.

Gennaioli, N. and A. Shleifer (2018). *A crisis of beliefs*. Princeton University Press.

Gorton, G. (2012). *Misunderstanding financial crises: Why we don't see them coming*. Oxford University Press.

Gorton, G. and A. Metrick (2012). Securitized banking and the run on repo. *Journal of Financial economics 104*(3), 425–451.

Gourio, F., A. K. Kashyap, and J. W. Sim (2018). The trade offs in leaning against the wind. *IMF Economic Review 66*(1), 70–115.

Greenspan, A. (2002). Economic volatility. *Remarks by Chairman Alan Greenspan at a symposium sponsored by the Federal Reserve Bank of Kansas City, Jackson Hole, Wyoming*.

Greenwald, B. C. and J. E. Stiglitz (1986). Externalities in economies with imperfect information and incomplete markets. *The quarterly journal of economics 101*(2), 229–264.

Greenwood, R., S. G. Hanson, and L. J. Jin (2019). Reflexivity in credit markets. *NBER Working Paper*.

Greenwood, R., S. G. Hanson, A. Shleifer, and J. A. Sørensen (2022). Predictable financial crises. *The Journal of Finance 77*(2), 863–921.

Greenwood, R. and A. Shleifer (2014). Expectations of returns and expected returns. *The Review of Financial Studies 27*(3), 714–746.

Haddad, V. and T. Muir (2021). Do intermediaries matter for aggregate asset prices? *The Journal of Finance*.

Hart, O. D. (1975). On the optimality of equilibrium when the market structure is incomplete. *Journal of economic theory 11*(3), 418–443.

He, Z., B. Kelly, and A. Manela (2017). Intermediary asset pricing: New evidence from many asset classes. *Journal of Financial Economics 126*(1), 1–35.

He, Z., I. G. Khang, and A. Krishnamurthy (2010). Balance sheet adjustments during the 2008 crisis. *IMF Economic Review 58*(1), 118–156.

He, Z. and A. Krishnamurthy (2013). Intermediary asset pricing. *American Economic Review 103*(2), 732–70.

Hong, H. and J. C. Stein (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. *The Journal of finance 54*(6), 2143–2184.

Jeanne, O. and A. Korinek (2019). Managing credit booms and busts: A pigouvian taxation approach. *Journal of Monetary Economics 107*, 2–17.

Jordà, Ò., M. Schularick, and A. M. Taylor (2015). Leveraged bubbles. *Journal of Monetary Economics 76*, S1–S20.

Kindleberger, C. P. (1978). *Manias, panics and crashes: a history of financial crises*. Palgrave Macmillan.

Kiyotaki, N. and J. Moore (1997). Credit cycles. *Journal of political economy 105*(2), 211–248.

Krishnamurthy, A. and W. Li (2024). Dissecting mechanisms of financial crises: Intermediation and sentiment. *Journal of Political Economy, Forthcoming*.

Lorenzoni, G. (2008). Inefficient credit booms. *The Review of Economic Studies 75*(3), 809–833.

L'Huillier, J.-P., S. R. Singh, and D. Yoo (2024). Incorporating diagnostic expectations into the new keynesian framework. *Review of Economic Studies 91*(5), 3013–3046.

Martin, A. and J. Ventura (2016). Managing credit bubbles. *Journal of the European Economic Association 14*(3), 753–789.

Martinez-Miera, D. and R. Repullo (2019). Monetary policy, macroprudential policy, and financial stability. *Annual Review of Economics 11*, 809–832.

Maxted, P. (2024). A macro-finance model with sentiment. *Review of Economic Studies 91*(1), 438–475.

McCarthy, O. and S. Hillenbrand (2022). Heterogeneous investors and stock market fluctuations. *Working Paper*.

Mendoza, E. G. (2010). Sudden stops, financial crises, and leverage. *American Economic Review 100*(5), 1941–66.

Minsky, H. P. (1977). The financial instability hypothesis: An interpretation of keynes and an alternative to "standard" theory. *Challenge 20*(1), 20–27.

Nagel, S. and Z. Xu (2024). Movements in yields, not the equity premium: Bernanke-kuttner redux. Technical report, Working paper.

Rabin, M. and D. Vayanos (2010). The gambler's and hot-hand fallacies: Theory and applications. *The Review of Economic Studies 77*(2), 730–778.

Rigobon, R. and B. Sack (2004). The impact of monetary policy on asset prices. *Journal of Monetary Economics 51*(8), 1553–1575.

Rognlie, M., A. Shleifer, and A. Simsek (2018). Investment hangover and the great recession. *American Economic Journal: Macroeconomics 10*(2), 113–53.

Scheinkman, J. A. and W. Xiong (2003). Overconfidence and speculative bubbles. *Journal of political Economy 111*(6), 1183–1220.

Schmitt-Grohé, S. and M. Uribe (2016). Downward nominal wage rigidity, currency pegs, and involuntary unemployment. *Journal of Political Economy 124*(5), 1466–1514.

Schmitt-Grohé, S. and M. Uribe (2021). Multiple equilibria in open economies with collateral constraints. *The Review of Economic Studies 88*(2), 969–1001.

Schularick, M. and A. M. Taylor (2012). Credit booms gone bust: Monetary policy, leverage cycles, and financial crises, 1870-2008. *American Economic Review 102*(2), 1029–61.

Shiller, R. J. (2015). *Irrational exuberance*. Princeton university press.

Shleifer, A. and R. W. Vishny (1992). Liquidation values and debt capacity: A market equilibrium approach. *The Journal of Finance 47*(4), 1343–1366.

Simsek, A. (2013). Belief disagreements and collateral constraints. *Econometrica 81*(1), 1–53.

Stein, J. C. (2021). Can policy tame the credit cycle? *IMF Economic Review 69*(1), 5–22.

Stiglitz, J. E. (1982). The inefficiency of the stock market equilibrium. *The Review of Economic Studies 49*(2), 241–261.

Sufi, A. and A. M. Taylor (2022). Financial crises: A survey. *Handbook of international economics 6*, 291–340.

Taylor, A. M. (2015). Credit, financial stability, and the macroeconomy. *Annu. Rev. Econ. 7*(1), 309–339.

Werning, I. (2015). Incomplete markets and aggregate demand. *NBER Working Paper*.

Woodford, M. (2012). Inflation targeting and financial stability. *Working Paper*.

Yellen, J. L. (2009). A minsky meltdown: Lessons for central bankers. presentation to the 18th annual hyman p. In *Minsky Conference on the State of the US and World Economies "Meeting the Challenges of the Financial Crisis", Levy Economics Institute of Bard College, New York City, April*, Volume 16.

Appendices

A Proofs and Derivations

A.1 Proof of Proposition 1

Leverage: At time $t = 2$, the welfare of financial intermediaries according to the planner's (rational) expectations can be written as:

$$\mathcal{W}_2 = \begin{cases} \beta \ln\left(n_2 + \phi H \mathbb{E}_2[z_3 + \Omega_3(q_2, q_1)]\right) + \beta^2 \left(\mathbb{E}_2[z_3]H - \phi H \mathbb{E}_2[z_3 + \Omega_3(q_2, q_1)]/\beta\right) & \text{if } z_2 \leq z^* \\ \beta\left(\beta \mathbb{E}_2[z_3]H + n_2\right) & \text{otherwise} \end{cases} \tag{A.1}$$

with $n_2 = z_2 H - d_1(1 + r_1)$, while the Lagrangian corresponding to bankers' problem in period $t = 1$ is given by:

$$\mathcal{L}_{b,1} = \left[u(c_1) + \tilde{\mathbb{E}}_1[\mathcal{W}_2(n_2, H; q_2, z_2, \zeta\Omega_3(q_2, q_1))]\right] - \lambda_1\left[c_1 + c(H) - d_1 - e_1\right] \tag{A.2}$$

the first-order condition on borrowing gives:

$$\frac{\partial \mathcal{L}_{b,1}}{\partial d_1} = \lambda_1 - \tilde{\mathbb{E}}_1\left[\lambda_2\right] \tag{A.3}$$

where λ_t is the Lagrange multiplier on the budget constraint at time t. The social planner maximizes the same function, but under its own expectations, and by also taking into account how a change in d_1 impacts asset prices in period 2. This leads to the following first-order condition:

$$\frac{\partial \mathcal{L}^{SP}_{b,1}}{\partial d_1} = \lambda_1 + \mathbb{E}_1\left[\lambda_2\right] - \beta \mathbb{E}_1\left[\kappa_2 \phi H \frac{\partial \Omega_3}{\partial q_2}\frac{\partial q_2}{\partial n_2}\right]\frac{dn_2}{dd_1} \tag{A.4}$$

where κ_2 is the Lagrange multiplier on the collateral constraint at $t = 2$. Hence simply by incorporating $\tilde{\mathbb{E}}_1[\lambda_2]$ we can express the total change in welfare as internalized plus uninternalized effects:

$$\frac{\partial \mathcal{L}_{b,1}^{SP}}{\partial d_1} = \underbrace{\lambda_1 - \tilde{\mathbb{E}}_1\left[\lambda_2\right]}_{\text{Internalized}} + \underbrace{\tilde{\mathbb{E}}_1\left[\lambda_2\right] - \beta\mathbb{E}_1\left[\lambda_2\right] - \mathbb{E}_1\left[\kappa_2 \phi H \frac{\partial \Omega_3}{\partial q_2}\frac{\partial q_2}{\partial n_2}\right]}_{\text{Uninternalized}} \tag{A.5}$$

which proves the first part of Proposition 1.

Investment: Using the same expressions as in (A.1), the Lagrangian corresponding to bankers' problem in period $t = 1$ is given by:

$$\mathcal{L}_{b,1} = \left[u(c_1) + \tilde{\mathbb{E}}_1[\mathcal{W}_2(n_2, H; q_2, z_2\Omega_3(q_2, q_1))]\right] - \lambda_1\left[c_1 + c(H) - d_1 - e_1\right] \tag{A.6}$$

the first-order condition on investment yields:

$$\frac{\partial \mathcal{L}_{b,1}}{\partial H} = -\lambda_1 c'(H) + \beta\tilde{\mathbb{E}}_1\left[\lambda_2(z_2 + \Omega_2)(z_2 + \Omega_2 + q_2(z_2 + \Omega_2, \Omega_3(q_2, q_1)))\right] \tag{A.7}$$

The social planner maximizes the same function, but under its own expectations, and by also taking into account how a change in d_1 impacts asset prices in period 1 and 2. This leads to the following first-order condition:

$$\frac{\partial \mathcal{L}_{b,1}^{SP}}{\partial H} = \beta\mathbb{E}_1\left[\lambda_2(z_2 + q_2)\right] - \lambda_1 c'(H) + \beta\mathbb{E}_1\left[\kappa_2\phi H \frac{\partial \Omega_3}{\partial q_2}\left(\frac{\partial q_2}{\partial n_2}z_2 + \frac{\partial q_2}{\partial H}\right)\right] \tag{A.8}$$

the second part of Proposition 1 is then proved once we notice that:

$$\begin{aligned}\frac{\partial \mathcal{L}_{b,1}^{SP}}{\partial H} = & \underbrace{\beta\tilde{\mathbb{E}}_1\left[\lambda_2(z_2 + q_2)\right] - \lambda_1 q_1}_{\text{Internalized}} + \\ & \underbrace{\beta\mathbb{E}_1\left[\lambda_2(z_2 + q_2)\right] - \beta\tilde{\mathbb{E}}_1\left[\lambda_2(z_2 + q_2)\right] + \beta\mathbb{E}_1\left[\kappa_2\phi H \frac{\partial \Omega_3}{\partial q_2}\left(\frac{\partial q_2}{\partial n_2}z_2 + \frac{\partial q_2}{\partial H}\right)\right]}_{\text{Uninternalized}}.\end{aligned} \tag{A.9}$$

Prices: The only variable that can be changed, at $t = 2$, by a change in q_1, is Ω_3 (remember that we are keeping everything else fixed at $t = 1$). Hence the welfare change is given by:

$$\frac{d\mathcal{W}_1}{q_1} = \beta\mathbb{E}_1[\lambda_2 \phi H \frac{d\Omega_3}{dq_1} - \beta\phi H \frac{d\Omega_3}{dq_1}(1 + r_2)] \tag{A.10}$$

where once again the first part in the expectation corresponds to the change in consumption at $t = 2$ induced by the shift in the collateral limit, and the second part corresponds to the decrease in consumption at $t = 3$ since the amount that needs to be repaid is higher. That leads, using $\kappa_2 = \lambda_2 - 1$ and $\beta(1 + r_2) = 1$, to the reversal externality formulation:

$$\mathcal{W}_q = \beta\mathbb{E}_1[\kappa_2 \phi H \frac{d\Omega_3}{dq_1}] \tag{A.11}$$

A.2 Proof of Proposition 2

I compute the difference between λ_2 expected by private agents and λ_2 expected by the planner, state by state z_2. When both expect a realization z_2 not to produce a financial crisis, marginal utilities are equalized to 1, so the difference disappears. For the rest there are two cases: either both marginal utilities correspond to binding collateral constraints, either one agent expect the friction to bind and the other not. The first case yields:

$$\begin{aligned} &\frac{1}{c_2(\Omega_2, \zeta\Omega_3)} - \frac{1}{c_2(0, \Omega_3)} = \\ &\quad \frac{1}{(\Omega_2)H - d_1(1 + r_1) + \phi H\mathbb{E}_2[z_3 + \zeta\Omega_3]} - \frac{1}{z_2 H - d_1(1 + r_1) + \phi H\mathbb{E}_2[z_3 + \Omega_3])} \end{aligned} \tag{A.12}$$

I take the first-order approximation around the REE $\lambda_2 = 1/(z_2 H - d_1(1 + r_1) + \phi H\mathbb{E}_2[z_3]) = 1/c_2(0, 0)$. It gives:

$$\begin{aligned} \frac{1}{(z_2 + \Omega_2)H - d_1(1 + r_1) + \phi H\mathbb{E}_2[z_3 + \zeta\Omega_3]} &= \frac{1}{c_2(0,0)} \frac{1}{1 + \frac{\Omega_2 H}{c_2(0,0)} + \frac{\phi H \zeta \Omega_3}{c_2(0,0)}} \\ &= \lambda_2 \left(1 - \frac{\Omega_2 H + \phi H \zeta \Omega_3}{c_2(0,0)}\right) \end{aligned} \tag{A.13}$$

While the same algebra for the second part of equation (A.12) yields similarly:

$$\frac{1}{z_2H - d_1(1+r_1) + \phi H \mathbb{E}_2[z_3+\Omega_3]} = \frac{1}{c_2(0,0)} \frac{1}{1+\frac{\phi\Omega_3 H}{c_2(0,0)}} = \lambda_2 \left(1 + \frac{\phi H \Omega_3}{c_2(0,0)}\right) \quad \text{(A.14)}$$

Taking the difference gives:

$$\frac{1}{c_2(\Omega_2, \zeta\Omega_3)} - \frac{1}{c_2(0,\Omega_3)} = \lambda_2^2 \left(H\Omega_2 - (1-\zeta)\phi H \Omega_3\right) \quad \text{(A.15)}$$

Finally we do not need to study the part coming from the planner and agents disagreeing about the occurrence of a crisis for a given z_2 since that effect is second-order: an infinitesimal difference integrated over an infinitesimal band. It thus follows that, to the first order:

$$\mathcal{B}_d \simeq -\boldsymbol{\Omega_2} H \mathbb{E}\left[\boldsymbol{\lambda_2^2} \mathbb{1}_{\kappa_2>0}\right] + \phi H \boldsymbol{(1-\zeta)} \mathbb{E}\left[\boldsymbol{\Omega_3 \lambda_2^2} \mathbb{1}_{\kappa_2>0}\right] \quad \text{(A.16)}$$

A.3 Derivation of Equation (27)

Totally differentiating the equilibrium pricing equation at $t = 2$ yields:

$$dq_2 = \beta dc_2 \mathbb{E}_2[z_3+\Omega_3] + \beta c_2 d\Omega_3 - \phi dc_2 \mathbb{E}_2[z_3+\Omega_3] + \phi(1-c_2)d\Omega_3. \quad \text{(A.17)}$$

Then use the budget constraint, also totally differentiated, to get $dc_2 = dn_2 + \phi H d\Omega_3$. Combining these two conditions gives:

$$dq_2 = \beta(dn_2 + \phi H d\Omega_3)\mathbb{E}_2[z_3+\Omega_3] + \beta c_2 d\Omega_3 \\ - \phi(dn_2 + \phi H d\Omega_3)\mathbb{E}_2[z_3+\Omega_3] + \phi(1-c_2)d\Omega_3. \quad \text{(A.18)}$$

then notice that by assumption, $d\Omega_3 = {}^{d\Omega_3}/_{dq_2} dq_2$. Thus rearranging yields:

$$dq_2 \left(1 - \beta\phi H \mathbb{E}_2[z_3+\Omega_3]\frac{d\Omega_3}{dq_2} - \beta c_2 \frac{d\Omega_3}{dq_2} + \phi^2 H \mathbb{E}_2[z_3+\Omega_3]\frac{d\Omega_3}{dq_2} - \phi(1-c_2)\frac{d\Omega_3}{dq_2}\right)$$

$$= \left(\beta \mathbb{E}_2[z_3 + \Omega_3] - \phi \mathbb{E}_2[z_3 + \Omega_3]\right) dn_2 \quad \text{(A.19)}$$

and using $c_2 + \phi H \mathbb{E}_2[z_3 + \Omega_3] = 2c_2 - n_2$ through the budget constraint, we end up with:

$$\frac{dq_2}{dn_2} = \frac{(\beta - \phi)\mathbb{E}_2[z_3 + \Omega_3]}{1 - (\phi + (\beta - \phi)(2c_2 - n_2)\frac{d\Omega_3}{dq_2}}. \quad \text{(A.20)}$$

A.4 Behavioral Wedge for Investment

I use the same notation as for the proof of Proposition 2, presented in Appendix A.2. The behavioral wedge for investment can consequently be expressed state-by-state as:

$$\mathcal{B}_H(z_2) = [\lambda_2(0;\Omega_3)(z_2 + q_2(0;\Omega_3))] - [\lambda_2(\Omega_2;\zeta\Omega_3)(z_2 + \Omega_2 + q_2(\Omega_2;\zeta\Omega_3)] \quad \text{(A.21)}$$

As for leverage, it is sufficient to only look at states where the borrowing constraint binds both in the expectation of the social planner and of private agents. To the first-order, we can write:

$$\mathcal{B}_H(z_2) = (\lambda_2(0;\Omega_3) - \lambda_2(\Omega_2;\zeta\Omega_3)(z_2 + q_2)) + \lambda_2^r \left(\Omega_3 \frac{dq_2}{d\Omega_3} - \Omega_2\left(1 + \frac{dq_2}{d\Omega_2}\right) - \zeta\Omega_3 \frac{dq_2}{d\Omega_3}\right) \quad \text{(A.22)}$$

The part $\lambda_2(0;\Omega_3) - \lambda_2(\Omega_2;\zeta\Omega_3)$ exactly corresponds to the behavioral wedge for leverage state-by-state, that we will denote by $\mathcal{B}_d(z_2)$ for conciseness. The behavioral wedge for investment can thus be expressed as:

$$\mathcal{B}_H \approx \beta \mathbb{E}_1[\mathcal{B}_d(z_2)(z_2 + q_2)\mathbb{1}_{\kappa_2 > 0}] - \beta\Omega_2 \mathbb{E}_1[\lambda_2(1 + (\beta - \phi)Hz_3)\mathbb{1}_{\kappa_2 > 0}] + \beta(1 - \zeta)\mathbb{E}_1\left[\Omega_3 \lambda_2 \frac{dq_2}{dz_3}\mathbb{1}_{\kappa_2 > 0}\right] \quad \text{(A.23)}$$

where $\mathcal{B}_d(z_2) = ((1 - \zeta)\Omega_3 - \Omega_2)\lambda_2^2$.

A.5 Proof of Proposition 3

The proof of Proposition 3 is straightforward once the uninternalized effects of leverage and investment have been derived. By assumption, the planner can impose taxes or subsidies on leverage, on the creation of collaterals assets, and on the holdings of collateral assets, which are rebated or funded lump-sum. Denote these taxes/subsidies respectively by τ_d, τ_H and τ_q. The budget constraint can be written:

$$c_1 + c(H) + \tau_h H + q_1 h \leq e_1 + d_1(1-\tau_d) + q_1 H + \tau_q h \tag{A.24}$$

where H is the amount invested and h is the amount kept on the balance sheet. Of course in equilibrium $h = H$.

The first-order conditions of private agents are given by:

$$\frac{\partial \mathcal{L}_{b,0}}{\partial d_1} = \lambda_1(1-\tau_d) - \tilde{\mathbb{E}}_1\left[\lambda_2\right] = 0 \tag{A.25}$$

$$\frac{\partial \mathcal{L}_{b,0}}{\partial H} = c'(H) + \tau_H - q_1 = 0 \tag{A.26}$$

$$\frac{\partial \mathcal{L}_{b,0}}{\partial h} = \lambda_1 q_1 + \lambda_1 \tau_q - \tilde{\mathbb{E}}_1\left[\lambda_2(z_2+q_2)\right] = 0 \tag{A.27}$$

The planner wants the agent to internalize the effects of leverage. This is simply done with a tax equal to:

$$\tau_d = -\frac{\mathcal{W}_d}{\lambda_1} \tag{A.28}$$

For investment, the planer wants to fix the level of investment at a level H such that:

$$c'(H) = \beta\mathbb{E}_1\left[\lambda_2(z_2+q_2)\right] + \beta\mathbb{E}_1\left[\kappa_2\phi H \frac{\partial \Omega_3}{\partial q_2}\left(\frac{\partial q_2}{\partial n_2} z_2 + \frac{\partial q_2}{\partial H}\right)\right] \tag{A.29}$$

and because

$$\beta\mathbb{E}_1\left[\lambda_2(z_2+q_2)\right] + \beta\mathbb{E}_1\left[\kappa_2\phi H \frac{\partial \Omega_3}{\partial q_2}\left(\frac{\partial q_2}{\partial n_2} z_2 + \frac{\partial q_2}{\partial H}\right)\right] = \beta\tilde{\mathbb{E}}_1[\lambda_2(z_2+q_2)] - \mathcal{W}_H \tag{A.30}$$

the tax must simply be set equal to:

$$\tau_H = -\frac{\mathcal{W}_H}{\lambda_1} \tag{A.31}$$

Finally, denote by q_1^* the price at $t = 1$ such that the reversal externality is equal to 0. This is the price the planner wants to set. We thus simply want $\lambda_1 q_1^* + \lambda_1 \tau_q - \tilde{\mathbb{E}}_1 \left[\lambda_2 (z_2 + q_2)\right] = 0$, so the tax should be set at:

$$\tau_q = \frac{\tilde{\mathbb{E}}_1 \left[\lambda_2 (z_2 + q_2)\right] - \lambda_1 q_1^*}{\lambda_1} \tag{A.32}$$

A.6 Proof of Proposition 4

As explained in the main text, the social planner's optimality condition under the premises of Proposition 4 can be expressed as:

$$u'(c_1) = \frac{1}{2\sigma_\Omega} \int_0^\infty \left[\int_{-\sigma_\Omega}^{\sigma_\Omega} \frac{\partial \mathcal{W}_2}{\partial n_2} \left(d_1, H; z_2 - \bar{\Omega}_2 - \omega_2\right) d\omega_2 \right] dF_2(z_2). \tag{A.33}$$

Key to this proposition is the shape of $\partial \mathcal{W}_2 / \partial n_2$ with respect to z_2. First recall that:

$$\mathcal{W}_2 = \begin{cases} \beta \ln \left(n_2 + \phi H \tilde{\mathbb{E}}_2[z_3]\right) + \beta^2 \left(\mathbb{E}[z_3] H - \phi H \tilde{\mathbb{E}}_2[z_3]/\beta\right) & \text{if } z_2 \geq z^* \\ \beta \left(\beta \mathbb{E}[z_3] H + n_2\right) & \text{otherwise} \end{cases} \tag{A.34}$$

so that the first derivative is equal to:

$$\frac{\partial \mathcal{W}_2}{\partial n_2} = \begin{cases} \beta \lambda_2 & \text{if } z_2 \geq z^* \\ \beta & \text{otherwise} \end{cases} \tag{A.35}$$

which is constant outside of a crisis, as expected. I use the following notation to simplify the exposition of the proof. First, the expectation over z_2 for a given w_2 is denoted by:

$$g(w_2) = \int_0^{+\infty} \frac{\partial \mathcal{W}_2}{\partial n_2} (d_1, H; q_2, z_2 - \bar{\Omega}_2 - \omega_2) dF_2(z_2) \tag{A.36}$$

while the integral taken over the uncertainty band is:

$$G(\sigma_\Omega) = \int_{-\sigma_\Omega}^{+\sigma_\Omega} \frac{g(w_2)}{2\sigma_\Omega} dw_2. \tag{A.37}$$

Given the continuity of $\partial \mathcal{W}_2 / \partial n_2$ (see equation A.34) we can differentiate with respect to σ_Ω:

$$G'(\sigma_\Omega) = -\frac{1}{2\sigma_\Omega^2} \int_{-\sigma_\Omega}^{+\sigma_\Omega} \int_0^{+\infty} \frac{\partial \mathcal{W}_2}{\partial n_2}(d_1, H; z_2 - \bar{\Omega}_2 - \omega_2) dF_2(z_2) dw_2 + \\ \int_0^{+\infty} \frac{\partial \mathcal{W}_2}{\partial n_2}(d_1, H; z_2 - \bar{\Omega}_2 - \sigma_\Omega) dF_2(z_2) - \int_0^{+\infty} \frac{\partial \mathcal{W}_2}{\partial n_2}(d_1, H; z_2 - \bar{\Omega}_2 + \sigma_\Omega) dF_2(z_2) \tag{A.38}$$

which can be expressed in terms of the notation just defined above as:

$$G'(\sigma_\Omega) = -\frac{G(\sigma_\Omega)}{\sigma_\Omega} + \frac{1}{2\sigma_\Omega}(g(\sigma_\Omega) - g(-\sigma_\Omega)) \tag{A.39}$$

Before proceeding further, remember that the social planner optimally sets leverage such that $u'(c_1) = G(\sigma_\Omega)$, while the decentralized equilibrium is independent of σ_Ω. Thus, leverage restrictions are increasing in σ_Ω if and only if G is increasing in σ_Ω. This condition is then equivalent, using the derivative just computed, to:

$$\frac{g(\sigma_\Omega) - g(-\sigma_\Omega)}{2} > \int_{-\sigma_\Omega}^{+\sigma_\Omega} \frac{g(w_2)}{2\sigma_\Omega} dw_2. \tag{A.40}$$

Since $\partial \mathcal{W}_2 / \partial n_2$ is continuous in z and in ω_2, and since ω_2 is defined in the compact set $[-\sigma_\Omega, \sigma_\Omega]$, g is continuous (by continuity of parametric integrals) and Fubini's theorem implies that a sufficient condition for $G'(\sigma_\Omega) > 0$ is that:[25]

$$\frac{1}{2}\left(\frac{\partial \mathcal{W}_2}{\partial n_2}(z_2 + \sigma_\Omega) - \frac{\partial \mathcal{W}_2}{\partial n_2}(z_2 - \sigma_\Omega)\right) > \int_{-\sigma_\Omega}^{+\sigma_\Omega} \frac{\partial \mathcal{W}_2}{\partial n_2}(z_2 + \omega_2) \frac{d\omega_2}{2\sigma_\Omega} \quad \forall z_2 \in supp(f_2). \tag{A.41}$$

In other words, this condition requires that the average taken over a segment is below the average of the two extreme points of this same segment.

[25] $\bar{\Omega}_2$ does not need to appear in this condition since this inequality is required to hold for all z_2 in the support of the definition, so equivalently for all $z_2 - \bar{\Omega}_2$ also in the support.

Next, notice that any convex function satisfies this requirement. For a convex function φ, Jensen's inequality yields:

$$\varphi(t\sigma_\Omega - (1-t)\sigma_\Omega) \leq t\varphi(\sigma_\Omega) + (1-t)\varphi(-\sigma_\Omega) \quad \forall t \in [0,1]. \tag{A.42}$$

Now integrate this inequality over t to get:

$$\int_0^1 \varphi(t\sigma_\Omega - (1-t)\sigma_\Omega)dt \leq \int_0^1 t\varphi(\sigma_\Omega)dt + \int_0^1 (1-t)\varphi(-\sigma_\Omega)dt. \tag{A.43}$$

A change of variable $t \to (x-\sigma_\Omega)/(2\sigma_\Omega)$ in the left-hand side thus yields:

$$\int_{-\sigma_\Omega}^{+\sigma_\Omega} \frac{\varphi(x)}{2\sigma_\Omega}dx \leq \frac{\varphi(\sigma_\Omega) - \varphi(-\sigma_\Omega)}{2} \tag{A.44}$$

which is exactly the relationship in equation (A.41).

We now have to prove that $\partial\mathcal{W}_2/\partial n_2$ is convex to end the proof of Proposition 4. Going back to equation (A.34), denote $\partial\mathcal{W}_2/\partial n_2$ by $\mathcal{W}_{2,n}$. Given Equation (A.35), start with the derivative of marginal utility ${}^{d\lambda_2}/_{dz_2} = -{}^{H}/_{c_2^2}$, and so ${}^{d^2\lambda_2}/_{dz_2^2} = {}^{2}/_{c_2^3}H > 0$, which concludes the proof for leverage.[26] For investment, the first order condition becomes:

$$\lambda_1 c'(H) = \frac{1}{2\sigma_\Omega}\int_0^\infty \left[\int_{-\sigma_\Omega}^{\sigma_\Omega} \lambda_2(z_2 - \bar{\Omega}_2 - \omega_2)(z_2 - \bar{\Omega}_2 - \omega_2 + q_2(z_2 - \bar{\Omega}_2 - \omega_2))d\omega_2\right] dF_2(z_2). \tag{A.45}$$

Fortunately, it is now straightforward to sign the derivative of this function given the previous proof for leverage. We know that $\lambda_2(z_2 - \bar{\Omega}_2 - \omega_2)$ is convex in ω_2. This is multiplied by a linear and positive function of ω_2 (the dividends), and then by the price realization at $t = 2$.The price at $t = 2$ is given by:

$$q_2 = \beta(n_2 + \phi M\mathbb{E}_2[z_3])\mathbb{E}_2[z_3] + \phi(1 - n_2 - \phi M\mathbb{E}_2[z_3])\mathbb{E}_2[z_3] \tag{A.46}$$

Which is clearly linear in ω_2 since net worth is linear in ω_2: $n_2 = (z_2 - \bar{\Omega}_2 - \omega_2)H -$

[26]For the sake of brevity, Ω_3 is left out of the expressions as, by assumption, it is a constant. It thus only shifts the value of $\tilde{\mathbb{E}}_1[z_3]$ and that has no impact on the sign of these derivatives as long as $\tilde{\mathbb{E}}_1[z_3] + \Omega_3 > 0$, which we always assume to be the case.

$d_1(1+r_1)$. Hence this function is convex in ω_2, which implies that the right-hand side of the first-order condition is increasing in uncertainty. This time, however, this means that $c'(H)$ in equilibrium needs to be higher than in the decentralized equilibrium. Hence, uncertainty calls for increasing investment. Intuitively, uncertainty increases the stochastic discount factor that prices the asset, while keeping the rest fixed, meaning that more consumption should be shifted to the future.

A.7 Ω_3-Uncertainty

This part extends the insights of Section 4.3 to the case where the uncertainty pertains to Ω_3. I start by studying the realization of only one state of the world, and complete the proof using the linearity of expectations.

I assume that for a given realization of z_2, the planner has a uniform distribution on sentiment during a crisis:

$$w_3 \sim \mathcal{U}\left[\bar{\Omega}_3 - \sigma_\Omega, \bar{\Omega}_3 + \sigma_\Omega\right] \tag{A.47}$$

The integral (denoted by L) used by the social planner to compute the marginal effect on welfare on increasing leverage becomes:

$$L = \frac{1}{2\sigma_\Omega}\int_{-\sigma_\Omega}^{\sigma_\Omega} \frac{\partial \mathcal{W}_2}{\partial n_2}\left(d_1, H; q_2, z_2, \bar{z}_3 - \bar{\Omega}_3 - \omega_3\right) d\omega_3 \tag{A.48}$$

Assume first that for all realizations of ω_3 the resulting equilibrium is a crisis one. This yields:

$$L = \frac{1}{2\sigma_\Omega}\int_{-\sigma_\Omega}^{\sigma_\Omega} \frac{1}{n_2 + \phi H(\bar{z}_3 - \bar{\Omega}_3 - \omega_3)} d\omega_3 \tag{A.49}$$

$$\implies L = -\frac{1}{(2\sigma_\Omega)\phi H}\left[\ln(n_2 + \phi H(\bar{z}_3 - \bar{\Omega}_3 - \omega_3))\right]_{-\sigma_\Omega}^{\sigma_\Omega} \tag{A.50}$$

$$\implies L = \frac{1}{(2\sigma_\Omega)\phi H}\ln\left(\frac{n_2 + \phi H(\bar{z}_3 - \bar{\Omega}_3 + \sigma_\Omega))}{n_2 + \phi H(\bar{z}_3 - \bar{\Omega}_3 - \sigma_\Omega))}\right) \tag{A.51}$$

This is a functions of the type:

$$f(x) = \frac{1}{x} \ln\left(\frac{K+x}{K-x}\right) \tag{A.52}$$

And we can show that this is increasing in x, for $x \in [0, K]$. Indeed, the derivative is given by:

$$f'(x) = \frac{(K^2 - x^2)\ln\left(\frac{K+x}{K-x}\right) + 2Kx}{x^2(K-x)(K+x)} \tag{A.53}$$

The denominator is clearly positive, but the denominator is indeterminate. Take the derivative of the denominator:

$$\frac{d}{dx}(K^2 - x^2)\ln\left(\frac{K+x}{K-x}\right) + 2Kx = 2x\ln\left(\frac{K+x}{K-x}\right) > 0 \tag{A.54}$$

The denominator is thus increasing and its limit in 0 is 0. Hence, f is increasing on $[0, K]$. Accordingly, L is increasing in σ_Ω.

Left now is the same calculation when for some parts of the uncertainty set, the economy is outside of a crisis. Following the same steps as before, this boils down to the study of:

$$g(x) = \frac{1}{x} \ln\left(\frac{1}{K-x}\right) \tag{A.55}$$

Where the derivative is now:

$$g'(x) = \frac{\frac{x}{a-x} - \ln\left(\frac{1}{K-x}\right)}{x^2} \tag{A.56}$$

And the derivative of the numerator is:

$$\frac{d}{dx}\frac{x}{a-x} - \ln\left(\frac{1}{K-x}\right) = \frac{x}{(a-x)^2} > 0 \tag{A.57}$$

Since $g'(0^+) > 0$, g is increasing. Thus the same result applies. This concludes the proof by linearity of expectations: since this integral is increasing in σ_Ω, all components of the expectations over all future states of the world are increasing, and it then follows that the overall expectation is increasing in σ_Ω.

A.8 Reversal Uncertainty

The analysis in Section 4.3 shows how the regulator should adapt leverage and investment regulations in the face of sentiment uncertainty. The last natural question is how price regulation (and thus the eventual use of interest rates) should be adapted. The following proposition answers this interrogation unambiguously.

Proposition 5 (Reversal Uncertainty and Price Regulation)**.** *Assume that inside crises, the behavioral bias takes the form $\Omega_3 = \bar{\Omega}_3 - \alpha_q q_1$ with $\bar{\Omega}_3$ a constant, and the planner believes that α is uniformly distributed on $[\bar{\alpha}_q - \sigma_\alpha, \bar{\alpha}_q + \sigma_\alpha]$, where $\bar{\alpha}_q$ and σ_α are positive constants. Then the optimal interest rate at $t = 1$ is increasing in uncertainty σ_α if the regulator has access to unconstrained leverage and investment regulations.*

This proposition formalizes the following intuition: the regulator fears that high prices today could translate into over-pessimism inside a future crisis, but is unsure of the strength of the extrapolation. In that situation, the more uncertainty there is around this extrapolation mechanism, the more the regulator wants to lower prices in the boom. This is one again coming from a similar convexity insight: cases where the extrapolation parameter is strong are more costly because of the non-linearities typically found in financial crises.

Proof: Using the premises of Proposition 5 and the expressions in Proposition 1, we are interested in the behavior of:

$$\frac{1}{2\sigma_\alpha}\int_{\bar{\alpha}_q-\sigma_\alpha}^{\bar{\alpha}_q+\sigma_\alpha} \kappa_2(\alpha_q)\alpha_q d\alpha_q = \int_{-\sigma_\alpha}^{\sigma_\alpha} (\lambda_2(\alpha_q) - 1)\alpha_q d\alpha_q \tag{A.58}$$

when σ_α varies. As we showed in the proof of Proposition 4 in Section A.6, it is sufficient to show that the function $(\lambda_2(\alpha_q) - 1)\alpha_q$ is convex in α_q to prove that this integral is increasing in σ_α. Using $\lambda_2 = {}^{1}/_{c_2}$ once again, we have:

$$\frac{d}{d\alpha}\big((\lambda_2(\alpha_q) - 1)\alpha_q\big) = (\lambda_2(\alpha_q) - 1) + \frac{\alpha_q \phi H q_1}{c_2^2} \tag{A.59}$$

differentiating once again:

$$\frac{d^2}{d\alpha_q^2}\big((\lambda_2(\alpha_q) - 1)\alpha_q\big) = \frac{\phi H q_1}{c_2^2} + \frac{\phi H q_1}{c_2^2} + \alpha_q \frac{(\phi H q_1)^2}{2c_2^3} \tag{A.60}$$

and all those terms are unambiguously positive. By convexity the strength of the reversal externality is increasing in σ_α. By proposition 9, this immediately requires a higher interest rate to satisfy Equation (C.14).

A.9 Multiple Equilibria

The analysis in the main paper as made under the assumption that the equilibrium was unique at $t = 2$ (see footnote 9). When sentiment is exogenous, the uniqueness of the equilibrium is straightforward to prove: the budget constraint directly pins down the consumption in equilibrium. This in turn directly pins down the asset price, and the equilibrium is unique.

Multiple equilibria can arise only when sentiment depends on asset prices (see Jeanne and Korinek 2019 ; Schmitt-Grohé and Uribe 2021). With endogenous biases, the system of equation becomes:

$$q_2 = \beta c_2 \mathbb{E}_2[z_3 + \Omega_3(q_2)] + \phi(1 - c_2)\mathbb{E}_2[z_3 + \Omega_3(q_2)] \tag{A.61}$$

$$c_2 = z_2 H - d_1(1 + r_1) + \phi H \mathbb{E}_2[z_3 + \Omega_3(q_2)] \tag{A.62}$$

which makes it clear that, as long as Ω_3 is strictly increasing in q_2, different equilibrium levels of asset prices result in different equilibrium levels of consumption. The asset price determination is given by:

$$q_2 = \beta \left(n_2 + \phi H \mathbb{E}_2[z_3 + \Omega_3(q_2)]\right) \mathbb{E}_2[z_3 + \Omega_3(q_2)] \\ + \phi(1 - (n_2 + \phi H \mathbb{E}_2[z_3 + \Omega_3(q_2)])) \mathbb{E}_2[z_3 + \Omega_3(q_2)] \quad \text{(A.63)}$$

Depending on the shape of $\Omega_3(q_2)$, an arbitrary number of equilibria are possible. I illustrate the problem with a linear function:

$$\Omega_3(q_2) = \alpha q_2 + \chi \quad \text{(A.64)}$$

The price condition is now:

$$q_2 = \beta \left(n_2 + \phi H \mathbb{E}_2[z_3 + \alpha q_2 + \chi]\right) \mathbb{E}_2[z_3 + \alpha q_2 + \chi] \\ + \phi(1 - (n_2 + \phi H \mathbb{E}_2[z_3 + \alpha q_2 + \chi])) \mathbb{E}_2[z_3 + \alpha q_2 + \chi] \quad \text{(A.65)}$$

This is a quadratic equation, hence will have at most two solutions (it can also only have one solution if the intersection gives a $c_2 = 1$, in which case the constraint is not binding). Only one of them will however be stable: since the consumption equation is linear in q_2, dc_2/dq_2 as computed along the pricing equation is necessarily below the slope of the budget constraint on one of the two equilibria. How more complicated forms of biases interact with frictions to create multiple equilibria is left for future work.

B Alternative Collateral Constraint with Current Prices

This section shows the robustness of my results when, instead, we consider a collateral constraint of the form:

$$d_2 \leq \phi H q_2. \quad \text{(B.1)}$$

Financial amplification comes into play because the consumption level c_2 that prices the asset directly depends on the price of the asset through the collateral constraint (with $h = H$ in equilibrium):

$$c_2 = z_2 H - d_1(1 + r_1) + \phi H q_2. \quad \text{(B.2)}$$

A fall in the price of the risky asset tightens the budget constraint even more, thus leading the price to fall further as a result of stronger discounting, and so on. Asset price and consumption are then determined in general equilibrium according to the fixed-point:

$$q_2 = \beta c_2(q_2)\mathbb{E}_2[z_3 + \Omega_3(q_2)] + \phi q_2(1 - c_2(q_2)). \tag{B.3}$$

Proposition 1 now takes the following – extremely similar – form.

Proposition 6 (Uninternalized Effects with $\phi H q_2$)**.** *The uninternalized first-order impact on welfare, when infinitesimally varying one aggregate variable while keeping the others constant, when the collateral constraint has current prices in it, are given by:*

i) For short-term debt d_1:

$$\mathcal{W}_d = \underbrace{\left(\tilde{\mathbb{E}}_1[\lambda_2] - \mathbb{E}_1[\lambda_2]\right)}_{\mathcal{B}_d} - \underbrace{\mathbb{E}_1\left[\kappa_2 \phi H \frac{dq_2}{dn_2}\right]}_{\mathcal{C}_d}. \tag{B.4}$$

ii) For investment in collateral assets H:

$$\mathcal{W}_H = \underbrace{\left(\beta\mathbb{E}_1\left[\lambda_2(z_2 + q_2)\right] - \lambda_1 q_1\right)}_{\mathcal{B}_H} + \underbrace{\beta\mathbb{E}_1\left[\kappa_2 \phi H \left(\frac{dq_2}{dn_2} z_2 + \frac{dq_2}{dH}\right)\right]}_{\mathcal{C}_H} \tag{B.5}$$

iii) And for prices q_1:

$$\mathcal{W}_q = \beta\mathbb{E}_1\left[\phi \kappa_2 H \frac{dq_2}{d\Omega_3}\frac{d\Omega_3}{dq_1}\right] \tag{B.6}$$

As in the core of the paper, an infinitesimal perturbation around the REE is enlightening (assuming Ω_2 and Ω_3 are small state-by-state):

Proposition 7 (Behavioral Wedge Approximation)**.** *If Ω_2 and Ω_3 are small state-by-state, the behavioral wedges can be expressed as:*

$$\mathcal{B}_d \simeq -\boldsymbol{\Omega_2} H \mathbb{E}\left[\boldsymbol{\lambda_2^2}\left(1 + \phi\frac{dq_2}{dn_2}\right)\mathbb{1}_{\kappa_2>0}\right] + \phi H(\boldsymbol{1 - \zeta})\mathbb{E}[\boldsymbol{\Omega_3 \lambda_2^2}\frac{dq_2}{dz_3}\mathbb{1}_{\kappa_2>0}]. \tag{B.7}$$

$$\mathcal{B}_H = \beta\mathbb{E}_1\left[\mathcal{B}_d(z_2)(z_2+q_2)\right] - \beta\Omega_2\mathbb{E}_1\left[\lambda_2^r\left(1+\frac{dq_2}{dz_2}\right)\mathbb{1}_{\kappa_2>0}\right] + \beta(1-\zeta)\mathbb{E}_1\left[\lambda_2^r\Omega_3\frac{dq_2}{dz_3}\mathbb{1}_{\kappa_2>0}\right] \tag{B.8}$$

where $\mathcal{B}_d(z_2)$ is the behavioral wedge for leverage for a realization z_2 of the dividend process at $t = 2$:

$$\mathcal{B}_d(z_2) = \Omega_2\lambda_2^2\left(H\Omega_2 + \phi\frac{dq_2}{dn_2}\right)\mathbb{1}_{\kappa_2>0} - \phi H(1-\zeta)\Omega_3\lambda_2^2\frac{dq_2}{dz_3}\mathbb{1}_{\kappa_2>0}. \tag{B.9}$$

As can readily be seen from this expression, all the intuitions are preserved with this collateral constraint: the comovement of future sentiment with the health of the financial sector, and the necessary interaction with financial frictions. The new terms are simply coming from the fact that an error in the expectation of dividends directly spills over expected consumption, through the level of asset prices at $t = 2$.[27]

Proposition 4 also still holds exactly in the same way, although the proof is more involved.[28]

C Monetary Policy as Price Regulation

I start by introducing rigidities in order for monetary policy deviations to have potential costs.[29] Because aggregate demand is not the focus on this paper, this is done by following Farhi and Werning (2020): households supply labor and output is demand-determined at $t = 1$ by assuming wages are fully rigid.[30] Households now have the following utility

[27] The price sensitivity is slightly different, because irrationality at $t = 2$, represented by Ω_3, influences equilibrium asset prices:

$$\frac{dq_2}{dn_2} = \frac{\beta\mathbb{E}_2\left[z_3 + \boldsymbol{\Omega_3}\right] - \phi q_2}{1 - \beta\phi H(\mathbb{E}_2\left[z_3+\boldsymbol{\Omega_3}\right]) + 2\phi^2 Hq_2 - c_2\beta\frac{d\boldsymbol{\Omega_3}}{dq_2}}. \tag{B.10}$$

[28] This is because the price in the collateral constraint introduces additional curvature. The proof is omitted for space constraints but can also be found in Fontanier (2022).

[29] The proposal to use interest rate hikes to act early has been central to the policy debate on asset bubbles, even though it has often been resisted by policy makers (Greenspan 2002; Bernanke 2002). Bernanke and Gertler (2000) show in a conventional model that asset prices are relevant to monetary policy only to the extent that they may signal inflationary pressures. Woodford (2012) complements this analysis by demonstrating that if the probability of a financial crisis increases with the output gap, it may be necessary to conduct tighter policy. Gourio, Kashyap, and Sim (2018) study this problem quantitatively when the probability depends on the amount of inefficient credit. Barlevy (2022) shows that this may backfire if the boom is driven by a "speculation shock."

[30] In Farhi and Werning (2020), there is an aggregate demand externality because wages are also fully rigid at $t = 2$, when the economy hits the ZLB. In my model there is no ZLB at $t = 2$, thus no aggregate demand

function:

$$U^h = \tilde{\mathbb{E}}_1 \left[\left(\ln(c_1^h) - \nu \frac{l_1^{1+\eta}}{(1+\eta)} \right) + \beta c_2^h + \beta^2 c_3^h \right] \tag{C.11}$$

which introduces curvature in consumption utility, and labor disutility in period $t = 1$. Firms produce using labor linearly, $Y_1 = l_1$. Wages are fully rigid and normalized to 1, causing workers to be potentially off their labor supply curve. This creates a role for monetary policy: the central bank can close the output gap by choosing the nominal rate of interest that brings workers back to their labor supply curve. The *labor wedge* $\mu_1 = 1 - \nu c_1^h l_1^\eta$ quantifies how far off are workers from their optimality condition. The labor wedge is positive when there is underemployment, and negative when there is overheating. Perfectly achieving natural employment means that $\mu_1 = 0$. Finally, Pareto weights are simply taken to be equal to the marginal utility of each type of agent at $t = 1$, in order to sidestep redistribution concerns.

A change in the nominal interest works through five different channels: (i) traditional aggregate demand ; (ii) credit ; (iii) investment ; (iv) current beliefs and (v) future beliefs. We can once again leverage the prior general welfare analysis.

Proposition 8 (Welfare Effects of Monetary Policy)**.** *The total welfare effects, as evaluated through the central bank's expectations, of an infinitesimal interest rate can be expressed by:*

$$\begin{aligned} \frac{d\mathcal{W}_1}{dr_1} = & \underbrace{\frac{dY_1}{dr_1}\mu_1}_{(i)} + \underbrace{\frac{dd_1}{dr_1}\mathcal{W}_d}_{(ii)} + \underbrace{\frac{dH}{dr_1}\mathcal{W}_H}_{(iii)} \\ & + \underbrace{\frac{d\Omega_2}{dq_1}\frac{dq_1}{dr_1}\left(\frac{dd_1}{d\Omega_2}\mathcal{W}_d + \frac{dH}{d\Omega_2}\mathcal{W}_H\right)}_{(iv)} + \underbrace{\frac{dq_1}{dr_1}\beta\mathbb{E}_1\left[\kappa_2\phi H\frac{d\Omega_3}{dq_1}\right]}_{(v)} \end{aligned} \tag{C.12}$$

where $\mathcal{W}_d = \mathcal{B}_d + \mathcal{C}_d$, the sum of the behavioral wedge and the collateral externality for leverage, and $\mathcal{W}_H = \mathcal{B}_H + \mathcal{C}_H$, the sum of the behavioral wedge and the collateral externality for investment. The last term is proportional to $\mathcal{W}_q$, the reversal externality (see Section 3 for details).

If the monetary authority is able to perfectly close the output gap and bring the econ-

externality. The results in this section are thus complementary to those in Farhi and Werning (2020), and do not rely on the inability of the central bank to lower rates sufficiently in crises.

omy to full employment, then it can achieve $\mu_1 = 0$ (and the perturbation is taken around the natural rate). As mentioned earlier, there is thus no first-order costs from deviating slightly from perfect inflation targeting. This expression thus embodies the idea in Stein (2021) that financial stability concerns loom large when unemployment is low (μ_1 close to zero), and should be negligible when unemployment is extremely high (μ_1 strongly positive).

This, however, does not necessarily imply that leaning against the wind is always desirable when the output gap can be closed, however. To see why, take the extreme case where the financial authority is able to adapt its leverage restrictions perfectly such that $\mathcal{W}_d = 0$, and look at the simpler case where $d\Omega_3/dq_1 = d\Omega_2/dq_1 = 0$ such that channels (iv) and (v) disappear. The welfare effects are thus now given in this special case by:

$$\frac{d\mathcal{W}_1}{dr_1} = \frac{dH}{dr_1}\mathcal{W}_H \tag{C.13}$$

because investment is unambiguously decreasing in the interest rate r_1, tightening is desirable only if $\mathcal{W}_H < 0$, i.e. if the uninternalized welfare effects of marginally increasing the creation of collateral assets is negative. As fully explained in Section 3.3, this object is actually positive for small belief deviations and becomes negative only if irrational exuberance is large enough. In other words the central bank would only pursue leaning against the wind when facing large enough behavioral distortions.[31]

Notice from equation (C.12) that the ability of the central bank to improve financial stability largely depends on the reaction of beliefs to policy. Without the belief channels (iv) and (v), the potential efficacy of leaning against the wind rests on the ability to curb leverage directly by raising rates, dd_1/dr_1. As emphasized by Werning (2015) and Farhi and Werning (2020), this is not a robust prediction of these models: it varies with the initial debt position as well as the shape of the utility function. To the contrary, the fact that increasing interest rates has a negative impact on asset prices is unambiguous in our models and is supported by robust empirical evidence (see e.g. Rigobon and Sack 2004

[31] My framework also abstracts from other considerations that could argue against tightening in such a situation. See Martinez-Miera and Repullo (2019) and Drechsler, Savov, and Schnabl (2022) for cases where leaning against the wind can backfire.

and Bernanke and Kuttner 2005). Thus if Ω_2 or Ω_3 depend directly on asset prices, leaning against the wind can have first-order benefits, by providing a supplementary instrument affecting equilibrium prices, and not only real allocations at $t = 1$.[32]

These results also directly speak to the debate about time-varying macroprudential tools. A common argument for using monetary policy to rein in financial excess is that, practically, macroprudential policy cannot be quickly adapted to be synchronized in real-time (Stein 2021). Inspecting Proposition 8, however, suggests that this is only part of the story. To focus on this question, assume: (i) fully unconstrained counter-cyclical capital regulation and (ii) fully unconstrained LTV regulation. Despite these assumptions, monetary policy still has an effect through prices and future behavioral biases.

Proposition 9 (Monetary Policy as Complement)**.** *When policymakers have access to unconstrained leverage and investment taxes, welfare changes evaluated around the equilibrium with optimal taxes are given by:*

$$\frac{d\mathcal{W}_1}{dr_1} = \frac{dY_1}{dr_1}\mu_1 + \beta\mathbb{E}_1\left[\kappa_2\phi H\frac{d\Omega_3}{dq_1}\frac{dq_1}{dr_1}\right]. \tag{C.14}$$

This particular case calls for leaning against the wind in order to lower *current* asset prices, which will then cure *future* pessimism in a possible crisis – a new channel for monetary policy. Furthermore, such action does not necessarily require information about contemporaneous biases. A sharp increase in asset prices could be entirely due to fundamentals, but the planner can still have an incentive to make prices deviate from their rational value today to prevent irrational distress from happening. Finally, implementing such a policy allows for financial regulation to be adapted and *relaxed*. Indeed, by acting preventively the central bank makes the future realizations of pessimism less severe, thus directly reducing the size of behavioral wedge and of the collateral externality. Taking

[32]In this paper I only consider biases that directly depend on asset prices or fundamentals. A more general formulation could allow for biases that are only function of the risk-premium or the risk-free rate separately. The role of monetary policy could then be different depending on its transmission mechanism (Drechsler, Savov, and Schnabl 2018a ; Drechsler, Savov, and Schnabl 2018b). In practice, the bulk of the effect of monetary policy comes from changes in the equity premium (Bernanke and Kuttner 2005). Understanding whether price extrapolation is differentially affected by different monetary policy channels is an open (and empirical) question.

this into account leads the optimal macroprudential limit to be less strict, which raises welfare.[33]

Finally, notice how the sophistication parameter ζ is naturally absent of Proposition 9. Sophisticated financial intermediaries realize very well that a high price today will translate into over-pessimism and tight collateral constraints in a future crisis, but cannot coordinate to reduce their buying pressure in order to cool off asset prices.

Proof of Propositions 8 and 9 The welfare function for the planner is given by:

$$\mathcal{W}_1 = \Phi^h \mathbb{E}_1\big(\ln \left[c_1^h - \nu \frac{l_1^{1+\eta}}{1+\eta} \right] + \beta c_2^h + \beta^2 c_3^h \big) + \Phi^b \mathbb{E}_1\big(\ln(c_1) + \beta \ln(c_2) + \beta^2 c_3 \big) \quad \text{(C.15)}$$

where Φ^h and Φ^b are the Pareto weights attached to each group by the planner. In equilibrium, we have $Y_1 = l_1$ by assumption of linear production. We thus write utility of households at $t = 1$ as:

$$\mathcal{W}_1^h = \ln \left[c_1^h - \nu \frac{Y_1^{1+\eta}}{1+\eta} \right] + \beta c_2^h + \beta^2 c_3^h. \quad \text{(C.16)}$$

Households' welfare is affected by two effects: first, a change n r_1 changes the incentives for savings, forcing agents to substitute wealth across periods. Second, it changes output and thus consumption and labor supply levels. However, since households are on their Euler equation at $t = 1$, the first effect is exactly 0:

$$\frac{d\mathcal{W}_1^h}{dr_1} = \frac{Y_1}{dr_1}\lambda_1^h - \nu Y_1^\eta \frac{Y_1}{dr_1}\lambda_1^h + \underbrace{\frac{dc_1^h}{dr_1}\lambda_1^h - \beta \tilde{\mathbb{E}}_1 \frac{dc_1^h}{dr_1}}_{\text{Euler}=0}. \quad \text{(C.17)}$$

Next, the change in the interest rate have an impact on the borrowing of financial intermediaries. This is not zero as for households, because of the uninternalized effects explored in Section 3. It also has an impact on investment, which for the same reason is not zero in general. Finally, it has an impact on prices, which can spill over on sentiment. Because

[33] Evidently, this policy problem is also plagued with uncertainty. Section A.8 shows that the incentives to tighten monetary policy are increasing in the uncertainty around the strength of the reversal force.

Pareto weights are chosen such as $\Phi^j = 1/\lambda_1^j$, we simply end up with:

$$\frac{d\mathcal{W}_1}{dr_1} = \frac{dY_1}{dr_1}\mu_1 + \frac{dd_1}{dr_1}\mathcal{W}_d + \frac{dH}{dr_1}\mathcal{W}_H + \frac{d\Omega_2}{dq_1}\frac{dq_1}{dr_1}\left(\frac{dd_1}{d\Omega_2}\mathcal{W}_d + \frac{dH}{d\Omega_2}\mathcal{W}_H\right) + \mathbb{E}_1\left[\kappa_2\phi H\frac{d\Omega_3}{dq_1}\frac{dq_1}{dr_1}\right] \tag{C.18}$$

Proposition 9 then follows directly from noticing that, when leverage and investment are set at the second-best level, terms (ii)-(iii)-(iv) in equation C.12 are zero by definition.

D Belief Heterogeneity

This section studies the case where lenders and borrowers have different beliefs about the distribution of future payoffs. This allows me to connect my results to the literature on belief disagreement as a contributing factor to boom-bust cycles (Scheinkman and Xiong 2003 ; Geanakoplos 2010 ; Simsek 2013 ; Caballero and Simsek 2020b), as well as highlighting whose beliefs actually matter for welfare.

D.1 Setup and Notation

Agents of type i, with $i \in \{h, b\}$, hold subjective beliefs about the aggregate state of the world at t, given by distributions $\tilde{F}_t^i$. As before, the deviations are encoded in scalars Ω_t^i:

$$\tilde{F}_t^i(z) = F_t(z - \Omega_t^i) \tag{D.1}$$

We also introduce two sophistication parameters: how bankers form beliefs about their future beliefs, and how they form beliefs about future households' beliefs.[34] At time $t = 1$, bankers believe that agents of type i at $t = 2$ will form expectations about z_3 according to

[34]It turns out that the sophistication level of households is irrelevant in my framework. I thus do not introduce new parameters to keep the notation as light as possible.

the following distorted probability distribution:

$$\tilde{F}^i_{3,\zeta^i}(z) = F_3(z - \zeta^i \cdot \Omega^i_t) \tag{D.2}$$

For instance, $\zeta^b = 1$ and $\zeta^h = 0$ means that bankers are fully sophisticated about their future beliefs (they understand that their expectation of z_2 in period $t = 2$ will be biased by Ω_3) but they are fully naive about future households (they think that households will form expectations about z_3 in a rational way).

Although households in my framework are passive and only extend funds to intermediaries, their beliefs are crucial at time $t = 2$. This is because the tightness of the collateral constraint directly depends on their beliefs about the collateral since they are lending up to (what they believe is) the default incentive:

$$d_2 \leq \phi H \mathbb{E}_2[z_3 + \Omega^h_3] \tag{D.3}$$

The rest of the model is fully identical. I will use $\tilde{\mathbb{E}}^i$ for the expectation operator under the beliefs of agents of type i.

D.2 Welfare Propositions

Leverage: Under this framework of heterogeneous beliefs, the uninternalized first-order impact on welfare, when varying d_1, is given by:

$$\mathcal{W}_d \quad = \quad \underbrace{\left(\tilde{\mathbb{E}}^b_1[\lambda_2] - \mathbb{E}_1[\lambda_2]\right)}_{\mathcal{B}_d} \quad - \quad \underbrace{\mathbb{E}_1\left[\kappa_2 \phi H \frac{\boldsymbol{d\Omega^h_3}}{\boldsymbol{dq_2}} \frac{dq_2}{dn_2}\right]}_{\mathcal{C}_d} \tag{D.4}$$

Relative to Proposition 1, the behavioral wedge is determined by the difference with the belief of intermediaries, but the collateral externality depends on the belief distortion of households. This is because it is the over-borrowing of intermediaries that is problematic (who are subject to a constraint, while households have linear utility so their consumption smoothing is irrelevant) but the tightness of the constraint depends on Ω^h_3.

The beliefs of households do enter the behavioral wedge indirectly, nevertheless. This can be seen in the following approximation, the counterpart of Proposition 2:

$$\mathcal{B}_d \simeq -\boldsymbol{\Omega^b_2} H\mathbb{E}\left[\boldsymbol{\lambda^2_2}\mathbb{1}_{\kappa_2>0}\right] + \phi H(\mathbf{1}-\boldsymbol{\zeta^h})\mathbb{E}\left[\boldsymbol{\Omega^h_3\lambda^2_2}\mathbb{1}_{\kappa_2>0}\right] \tag{D.5}$$

Investment: The uninternalized first-order impact on welfare, when varying H, is given by:

$$\mathcal{W}_H = \underbrace{\left(\beta\mathbb{E}_1\left[\lambda_2(z_2+q_2)\right]-\lambda_1 q_1\right)}_{\mathcal{B}_H} + \underbrace{\beta\mathbb{E}_1\left[\kappa_2\phi H\frac{\boldsymbol{d\Omega^h_3}}{\boldsymbol{dq_2}}\left(\frac{dq_2}{dn_2}z_2+\frac{dq_2}{dH}\right)\right]}_{\mathcal{C}_H} \tag{D.6}$$

Notice, however, that the equilibrium price q_1 depends on almost all behavioral factors. First, the expected λ_2 depends on Ω^b_2 and $\zeta^h\Omega^h_3$: net worth in period 2 depends on dividend payments, and consumption depends on the tightness of the collateral constraint. Second, q_2 also depends on $\zeta^b\Omega^b_3$: the beliefs of bankers in the intermediate period are the ones pricing the risky asset.

Asset Price: The uninternalized first-order impact on welfare, when varying q_1, is given by:

$$\mathcal{W}_q = \beta\mathbb{E}_1\left[\kappa_2\phi H\frac{\boldsymbol{d\Omega^h_3}}{\boldsymbol{dq_1}}\right]. \tag{D.7}$$

D.3 Discussion

Two broad lessons emerge from this analysis. First, during the boom, the beliefs of borrowers (intermediaries in this framework) are what should be monitored by the regulator. But second, during the bust, the beliefs of lenders are what matter since they determine the tightness of the collateral constraint. In particular, for the reversal and collateral externalities to appear, the belief distortions of lenders must be direct functions of equilibrium prices.

This does not imply that the beliefs of borrowers during financial crises are irrelevant, nonetheless. Remember, from (D.4), (D.6), and (D.7), that all these externalities are greater in magnitude when κ_2, i.e. when the collateral constraint is tighter. This happens in particular when Ω_3^h is very negative (excessive pessimism from households). But Ω_3^h depends on q_2, which itself reflects the beliefs of intermediaries in the bust, Ω_2^b (since they are the marginal pricers of the asset). In other words, both types of irrational pessimism contribute to the severity of the crisis, but the endogeneity of the beliefs of households are what determine the existence of these novel externalities.[35]

[35] Note that, in the alternative case where the collateral constraint depends on current equilibrium prices, only the beliefs of intermediaries would matter: the belief of households would not have any impact, since they would value the collateral at its market price. While a general analysis of this interaction between disagreement and financial frictions is outside of the scope this paper, this suggests that the choice of micro-foundations is not innocuous in such settings.