Optimal Policy for Behavioral Financial Crises[*]

Paul Fontanier[†]

January 30, 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 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, as well as 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*.[1] 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).[2] In response, economists have developed a number of behavioral models of financial instability.[3] 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 social planner and the market share the same beliefs, or when agents are sophisticated? And third, how should regulators incorporate incomplete information about behavioral biases when

[1] This view has been articulated by, e.g. Gorton (2012) or Geithner (2014).

[2] Specifically, forecast errors made by market participants are reliably predictable ex ante, using for example forecast revisions as pioneered by Coibion and Gorodnichenko (2012).

[3] See Bordalo, Gennaioli, and Shleifer (2018), Greenwood, Hanson, and Jin (2019), Maxted (2020) and Krishnamurthy and Li (2020). 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.

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 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 precise extent of behavioral biases in financial markets increases the incentives for the planner to act early rather than wait.

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 which can be thought of, e.g. as real estate or mortgage loans. At the heart of the model lies a financial friction: in the intermediate period, borrowing by intermediaries needs to be secured by posting these risky assets as collateral. The amount of borrowing available depends on the quantity of collateral available, 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. It is general enough to represent many psychological phenomena, while keeping the welfare analysis tractable. Crucially, being over-optimistic in booms regarding the prospects of the collateral asset is by implication being over-optimistic regarding the capacity of the financial sector to refinance itself. Behavioral biases in the asset market thus spread over the entire economy and distort all allocations. In order to study the robustness of my results, I allow agents to be potentially sophisticated: a parameter controls by how much agents realize that the market's future expectations will be biased.

I present the welfare analysis in Section 3, where a paternalistic social planner evaluates welfare using his own (rational) expectations.[4] 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 bind in the future. Furthermore, it highlights how the predictable components of *future* behavioral biases formed inside a financial crisis also create losses and should be monitored when agents are not fully sophisticated. 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 current and past asset prices, new externalities arise. By borrowing and investing, agents influence the realization of current and future 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 or return extrapolation, this fall in asset prices can trigger irrational pessimism, which tightens collateral constraints and deepens financial crises. Belief amplification thus creates an externality that calls for reducing leverage ex-ante: by increasing the net worth of inter-

[4] The welfare analysis is set in the context of *constrained efficiency*, as is common in the literature on optimal policy with financial frictions (Geanakoplos and Polemarchakis 1985 ; Dávila and Korinek 2018). The planner is thus unable to act during crises and undo the binding financial constraints. Appendix E allows for the simultaneous choice of ex-ante and ex-post policies.

mediaries in a crisis, this policy supports asset prices, which in itself supports sentiment and thus relaxes the future collateral constraint.

An additional effect, termed a *reversal externality*, works through prices *during* the boom. When agents invest in risky assets in good times they 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 in a crisis.

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.

How can one interpret these results of the model 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.[5] 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 financial 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

[5]Counter-cyclical capital buffers are at the center of the Basel III regulatory framework. My model shows how to optimally vary the levels of buffers when sentiment is fluctuating.

do the counter-cyclical capital buffers. The presence of the reversal externality however calls for the use of a third instrument in order to control prices. Monetary policy can be used as a complementary tool: even when counter-cyclical capital buffers and LTV ratios can be flexibly adapted, an increase in the interest rate can be beneficial. By lowering contemporaneous asset prices, monetary policy influences the future equilibrium determination when sentiment is endogenous. The future price crash inside a financial crisis will be less severe, mitigating the excess pessimism and relaxing collateral constraints. 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.[6,7]

It is however undeniable that identifying a bubble, or anticipating future pessimism, is intrinsically difficult since corresponding fundamentals are not observable. Indeed, the challenge for financial authorities of detecting contemporaneous irrationality in financial markets is 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. In Section 5, 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. Similarly, I show that this intuition also goes through regarding price regulation in order to counter the reversal externality. Imagine that the regulator

[6] An interesting example is the housing boom of the 2000s: while initial price increases in 2001-2003 may have been supported by fundamentals and low interest rates, it might have been the trigger for further irrational extrapolation down the road, resulting in adverse welfare consequences. If that is the case, my model suggests that an interest rate hike is warranted.

[7] Note that the central bank needs to act ex-ante to prevent *future* extrapolation because, in the study of constrained efficiency, it is not permitted to relax financial constraints in a crisis. Would the social planner be able to change allocations freely also during a financial crisis, the need for leaning against the wind would disappear. Similarly, in Farhi and Werning (2020) the central bank cannot lower the interest rate in a crisis because of the zero lower bound, which would be beneficial to stimulate asset prices.

fears that high prices today could translate into over-pessimism in a future crisis, but is unsure of the strength of this mechanism. In that situation, the more uncertainty there is around this extrapolation force, the more the regulator wants to force lower prices in the boom. Investment regulation, however, must be *relaxed* in the face of uncertainty, in order to encourage shifting resources to the crisis state.

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.[8] In a recent survey, Sufi and Taylor (2021) 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 the Baa bond – Treasury credit spread.[9]

[8] Recent work refined our understanding of this predictability, and identified many other predictive factors. Baron and Xiong (2017) examine bank equity returns. Greenwood and Hanson (2013) focus on a measure of credit quality, and find that credit booms are accompanied by a deterioration of the average quality of corporate issuers, and that a high share of risky loans forecasts negative corporate bond returns. López-Salido, Stein, and Zakrajšek (2017) demonstrate predictable mean-reversion in credit spreads, and that elevated credit-market sentiment predicts a decline in economic activity in the following years. Kirti (2018) and Krishnamurthy and Muir (2020) use interactive regression specifications by combining credit growth with a proxy for sentiment, and find results consistent with the idea that the interplay between leverage and mispricing is central. Mian, Sufi, and Verner (2017) show that household debt is also a good predictor of future economic slowdowns, an indication that systematic extrapolation errors are not specific to the finance sector.

[9] Recent theoretical work introduced extrapolative expectations into financial frictions models, and in particular showed how behavioral biases allow standard models to match the observed behavior of credit spreads before crises (Bordalo et al. 2018; Greenwood et al. 2019; Maxted 2020; Krishnamurthy and Li 2020; Bordalo, Gennaioli, Shleifer, and Terry 2021; Camous and Van der Ghote 2021). Other papers started integrating behavioral distortions into business cycle analysis, eg. L'Huillier, Singh, and Yoo (2021) and Bianchi, Ilut, and Saijo (2021). Chodorow-Reich, Guren, and McQuade (2021) study housing, where improvement in fundamentals triggers a boom-bust-rebound driven by over-optimism. All of these papers use the diagnostic expectations mechanism of Bordalo et al. (2018). Agents learn about the fundamentals by observing "dividends" but become over-optimistic. They thus do not feature endogenous sentiment, a feature that has different implications for policy as I show in this paper.

My paper integrates these lessons into the literature on normative macro-finance.[10] My framework follows from 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.

There is also a vast literature showing that adaptive learning improve the fit of business cycles models to the data (see Gaspar, Smets, and Vestin (2010) for a survey), but such models usually deliver under-reaction. A notable exception is the work of Adam and Marcet (2011) and Adam, Marcet, and Beutel (2017a), where subjective price dynamics lead to boom-bust dynamics in asset prices. Using this mechanism, Winkler (2020) builds a model where firms face financial constraints (leading to an amplification mechanism similar to the one in my paper) but does not study financial crises or optimal macroprudential and monetary policies. Caines and Winkler (2021) and Adam, Pfäuti, and Reinelt (2022) study monetary policy in the presence of subjective asset price dynamics also using learning models in standard New-Keynesian setups.

I end this section by focusing on the most closely related papers. First, Farhi and Werning (2020) analyze an environment with aggregate demand – rather than pecuniary – 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.[11] Third, Caballero and Simsek

[10] To provide a rigorous welfare analysis under behavioral distortions, I rely on a recent literature of "behavioral public finance," with Gruber and Köszegi (2001), O'Donoghue and Rabin (2006) and Mullainathan, Schwartzstein, and Congdon (2012). Farhi and Gabaix (2020) provide a general treatment of optimal taxation with behavioral agents, and I use their result and their concept of a "behavioral wedge" to characterize uninternalized welfare effects.

[11] I also contribute to this line of research by providing an alternative way of modeling general belief distor-

(2020) study monetary policy when macroprudential policy is constrained, and agents have heterogenous beliefs.[12] I build on the results of these three papers, 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.[13] I start by laying out the different ingredients of the model, before describing in details the class of belief distortions I will consider.

tions. 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.

[12] 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."

[13] All my results go through with the same intuition when I perform the same analysis with a price-dependent collateral constraint that creates standard pecuniary externalities. See Appendix C and the discussion below.

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 expectation 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 $t = 2$.

Preferences: Bankers have log-utility in period $t = 1$ and $t = 2$, and linear utility in the last period:[14]

$$U^b = \mathbb{E}_1 \left[\ln(c_1) + \beta \ln(c_2) + \beta^2 c_3 \right] \tag{1}$$

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

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

Financial Assets: There are two financial assets in the economy: deposits and the risky asset. Financial intermediaries issue deposits $d_t \geq 0$ to households at time t, to finance their consumption and their investment in the risky asset. The price of the risky asset at time t is denoted by q_t. At time $t = 1$, financial intermediaries can create H units of the asset by paying a convex cost $c(H)$.[15] 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 independent cumulative probability distributions F_2 and F_3, with

[14] This functional form is adopted for simplicity. It allows for tractability in the equilibrium expressions during a financial crisis. Appendix H presents the analysis without linear utility in the ultimate period and with a general IES throughout. The results of the analysis and the intuitions are entirely similar, at the cost of unnecessary complexity.

[15] Each intermediary has an indivisible real option to do business. Hence even faced with convex costs, an intermediary cannot multiply to operate at the infinitesimal scale.

support on $\mathbb{R}_+^*$. Only intermediaries have the expertise required to hold risky assets.[16]

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:

$$d_2 \leq \phi H \mathbb{E}_2[z_3] \tag{3}$$

where the parameter ϕ depends on the legal environment. The lower ϕ is, the less the bank is able to issue deposits to households in the intermediate period.

Here and in what follows, the expectation operator $\mathbb{E}_t$ must not be interpreted as en objective (rational) expectation, but rather as agent's own subjective expectations. I specify belief distortions after laying out the rest of the model, in Section 2.2, but equation (3) makes clear that beliefs will 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:*

$$\phi < \beta. \tag{4}$$

Constraints: Financial intermediaries' constraints are then as follows:

$$c_1 + c(H) + q_1 h_1 \leq e_1 + d_1 + q_1 H \tag{5}$$

$$c_2 + d_1(1 + r_1) + q_2 h_2 \leq d_2 + (z_2 + q_2) h_1 \tag{6}$$

[16] 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.

$$c_3 + d_2(1 + r_2) \le z_3 h_2 \tag{7}$$

$$d_2 \le \phi h_2 \mathbb{E}_2[z_3] \tag{8}$$

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.

In order for financial intermediaries to always be able to repay their debt, I make the following parametric assumption, assumed to hold throughout the analysis.

Assumption 2. *The financial friction parameter is small enough such that:*

$$\phi \mathbb{E}_2[z_3] < \min z_3. \tag{9}$$

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:

$$n_2 = z_2 H - d_1(1 + r_1). \tag{10}$$

Finally, the budget constraints of intermediaries are simply given by:

$$c_1^h - d_1 \le e_1^h \tag{11}$$

$$c_2^h - d_2 \le e_2^h + d_1(1 + r_1) \tag{12}$$

$$c_3^h \le e_3^h + d_2(1 + r_2) \tag{13}$$

$$\tag{14}$$

with the usual non-negativity constraints:

$$c_t^h \geq 0 \text{ and } c_t^h \geq 0, \text{ for } t \in \{1, 2, 3\}. \tag{15}$$

We assume that households endowments e_t^h are large enough to guarantee positive consumption in every period.

Interpretation of the Environment: Financial intermediaries should be interpreted as levered financial institutions that are using short-term debt: commercial and investment banks, insurance companies, hedge funds, brokers, etc.

The risky asset is used as collateral for short-term debt by financial intermediaries. A favored interpretation is that H represents Mortgage-Backed Securities (MBS), complex products widely used in repo markets. In this case, the costs $c(H)$ can be interpreted as securitization costs (e.g. search costs, legal fees, or the wages of structured traders).[17] 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). In the baseline model, banks against their collateral constraints can only reduce their consumption. Appendix B extends the model to allow for real production: intermediaries lend money to firms subject to pay a wage-in-advance constraint. Naturally, banks against their collateral constraint also have to reduce their lending activity to the real sector, which creates a real recession. The welfare results are identical in this case, only more involved, which is why I focus on the simpler case. But a "financial crisis" is meant to capture moments where intermediaries are under stress and cannot fulfill their role.

Remark 1 (Microfoundations of the Collateral Constraint). The specification of the collateral constraint in equation (3) can be obtained from the following microfoundations:

[17] Alternatively, one can picture intermediaries as firm/bank coalitions, where H represent C&I loans or projects funded by the intermediaries. H may also represent real estate held by the financial sector: the dividends are then simply rents coming from these operations.

i) Financial intermediaries lack commitment to repay in the final period ;

ii) Financial intermediaries must take the decision of whether to default before observing the realization of z_3 ;

iii) In the event of default, lenders can seize a fraction ϕ of the asset held by intermediaries.

These frictions lead lenders to only be willing to lend up to a fraction of the average future payoffs of the risky asset.[18] While also realistic, this form of the collateral constraint 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.

A large part of the normative macro-finance literature, for this reason, uses an alternative formulation for financial frictions to obtain pecuniary externalities. Dávila and Korinek (2018) show that a *collateral externality* arises when the collateral constraint depends on the *current* price of the asset, as in:

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

This type of collateral constraint is used for example by Farhi and Werning (2016), Bianchi and Mendoza (2018), and Jeanne and Korinek (2019). It can be microfounded by assuming that avoiding repayment requires diverting resources in the current period, and this is perfectly observed by lenders. Ottonello, Perez, and Varraso (2022) show that the quantitative predictions of both types of constraint are similar. Without taking a stance on which one is more realistic, I focus on the future payoff constraint in equation (8) since it cleanly isolates the effects of behavioral biases, and show in Appendix C the robustness of the results to this alternative formulation.[19]

[18] The model could perfectly be written with a collateral constraint of the form $d_2 < \phi H \min z_3$. This would relax the second assumption made for the micro-foundations: borrowers could default after observing the realization of z_3. My conclusions would be unchanged because the behavioral bias in my model shifts $\min z_3$ exactly as it shifts $\mathbb{E}_2[z_3]$. I keep the $\mathbb{E}_2[z_3]$ formulation as expressions are cleaner this way.

[19] Results are also similar if intermediaries are subject to a constraint in the initial period.

2.2 Beliefs

Households and Intermediaries have homogenous beliefs. I allow for a class of deviations from rationality that strikes a balance between generality and tractability, as well as providing several different insights.

Recall that dividends are drawn from (objective) 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, the subjective expectations of agents regarding period-$t+1$ dividends are encoded in a scalar Ω_{t+1}.

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:*

$$\Omega_{t+1} : \mathcal{I}_t = \{z_{t-i}, q_{t-i}\}_{i \geq 0} \to \mathbb{R} \tag{17}$$

At time t, agents believe that dividends paid in the next period are drawn from the cumulative distribution $F_{t+1}(z - \Omega_{t+1}(\mathcal{I}_t))$. They thus believe that the dividend paid in each state of the world will be $z_{t+1} + \Omega_{t+1}(\mathcal{I}_t)$ rather than z_{t+1}.

In other words, the behavioral bias Ω_{t+1} is a location shifter on expected dividends. In that respect, Ω_{t+1} exactly represents the *predictable component* at t of forecast errors realized at $t+1$.[20] 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. In this case, 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 (z_{t-i} or q_{t-i}). This approach is particularly flexible for the subsequent welfare analysis, since it summarizes all possible distortions in a single quantity. The functional

[20] See Cieslak (2018), Bordalo et al. (2018), or Ma, Ropele, Sraer, and Thesmar (2020) for empirical evidence on this predictability.

dependence on $\mathcal{I}_t$ is, however, assumed to be independent of policy.[21] 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.[22]

Fundamental Extrapolation: This case captures models where investors extrapolate fundamentals, which here are $\{z_t\}$. Several influential papers use this class of models to explain a wide range of facts about asset prices, starting with Barberis, Shleifer, and Vishny (1998). This extrapolation can come from a variety of psychologically founded biases.[23] I model fundamental extrapolation in reduced-form as:

$$\Omega_{t+1} = \alpha_z(z_t - z_{t-1}) \tag{18}$$

where α_z is a positive number. 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 $z_2 < z_1$.

Price/Return Extrapolation: 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 as this paper will show. Early examples include mod-

[21] I could also allow Ω to depend on other equilibrium variables in particular quantities like debt or investment. This does not bring additional insight. This is because debt and investment are already targeted by the usual instruments of leverage limits and LTV regulations (see below in Section 3). The case where biases depend on policy could yield further insight and is left for future research.

[22] 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. I present and discuss several cases in Appendix F.

[23] Constraints on memory and cognition can make it difficult for agents to work with complicated models, as in Fuster, Hebert, and Laibson (2012), leading agents to excessively use recent data points. Bordalo et al. (2018) link extrapolative beliefs about fundamentals to the representativeness heuristic of Kahneman and Tversky (1972). In Rabin and Vayanos (2010), extrapolative beliefs stem from believing in the law of small numbers.

els by DeLong, Shleifer, Summers, and Waldmann (1990), Hong and Stein (1999) and Barberis and Shleifer (2003).[24] Close to this paper, Farhi and Werning (2020) use return extrapolation in a model with aggregate demand externalities to study macroprudential and monetary policy. In the present paper, price extrapolation is modeled in reduced-form as:

$$\Omega_{t+1} = \alpha_q (q_t - q_{t-1}) \tag{19}$$

where, in period $t = 1$, we will postulate the existence of a hypothetic price q_0 that prevailed in the past and anchors agents' expectations. Crucially, agents' present and *future* beliefs now move with policies that influence asset prices (a potential channel for monetary policy, as studied in Section 4.3).

Remark 2 (Agents' Beliefs and Collateral Constraint). I assumed above that all agents (borrowers and lenders) share the same behavioral biases Ω throughout the paper. In the micro-foundations of the collateral constraint proposed above, households are lending up to a limit where they think intermediaries would default. Hence, what matters in equation (3) are the beliefs of lenders, rather than borrowers. But since beliefs are homogenous, the tightness of the collateral constraint will nevertheless depend on $\mathbb{E}_2[z_3 + \Omega_3]$.[25]

2.3 Sophistication

Crucially for my 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 second parameter that controls

[24] Recent research leverages the use of survey data to motivate price or return extrapolation: McCarthy and Hillenbrand (2021) estimate that return extrapolation accounts for 23% of movements in the S&P500 index. Price and return extrapolation have been used by Barberis, Greenwood, Jin, and Shleifer (2018) to present a model of financial bubbles, while DeFusco, Nathanson, and Zwick (2017) apply it to the housing market.

[25] It would differ, however, in a model where the collateral constraint depends on the equilibrium price q_2. Since intermediaries are the marginal pricers of the asset, the beliefs of borrowers would determine the tightness of the constraint in a crisis (see Simsek (2013) for an in-depth analysis). For simplicity beliefs are assumed to be homogenous in my model.

the level of sophistication of agents, ζ:

Definition 2 (Sophistication Parametrization)**.** *The parameter $\zeta \in \mathbb{R}$ captures the level of sophistication of agents. At time $t = 1$, agents have expectations regarding their future expectations such that:*

$$\mathbb{E}_1\left[\,\mathbb{E}_2[z_3]\,\right] = \mathbb{E}_1[z_3 + \zeta\Omega_3]. \tag{20}$$

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.[26] Agents are partially sophisticated when $\zeta \in]0,1[$.

Remark 3 (Subjective Prices). Notice that the bias in the baseline version of my model is always modeled as a shift in the subjective distribution of *dividends*, not *prices*. This is a reduced-form assumption, and such a bias appears for example when agents are learning from prices as in Chahrour and Gaballo (2021) or Bastianello and Fontanier (2022). Subjective prices are then also distorted since agents form expectations in a consistent manner. I now expand on this point before formally defining the non-rational equilibrium.[27]

2.4 Equilibrium

Before formally defining the equilibrium concept used in the paper, it is informative to understand in greater details how agents form expectations about future variables given the specific form of biases I assumed in Definitions 1 and 2. One way to gain intuition is to inspect the usual pricing equation of the risky asset at t:

$$q_t = \beta\mathbb{E}_t\left[\frac{\lambda_{t+1}(z_{t+1}+\Omega_{t+1},\zeta\Omega_{t+2})}{\lambda_t}\Big(z_{t+1}+\Omega_{t+1}+q_{t+1}\big(z_{t+1}+\Omega_{t+1},\zeta\Omega_{t+2}\big)\Big)\right] \tag{21}$$

where $q_{t+1}(z_{t+1}+\Omega_{t+1},\zeta\Omega_{t+2})$ is the price that would prevail, at $t+1$, in a environment where the state of the world realizes at $z_{t+1}+\Omega_{t+1}$, and future agents will price the as-

[26] ζ thus measures the degree of inconsistency in the Law of Iterated Expectations. Implicitly, this formulation assumes that sophisticated agents understand how Ω_3 will be determined according to information available at $t = 2$.

[27] Appendix F.3 also discusses the case where the behavioral bias is distorting expected *prices*, while expectations of fundamentals stay rational (as in, e.g. Adam and Marcet 2011). In that case, some externalities survive only when the collateral constraint depends on prices, rather than expected payoffs.

set believing that at $t+2$ it will pay $z_{t+2} + \zeta\Omega_{t+2}$ rather than z_{t+2}. Similarly, agents can potentially realize that their marginal utility λ_{t+1} will be determined by biased agents: agents form marginal utility expectations taking into account that their future selves will receive dividends $z_{t+1} + \Omega_{t+1}$. But they also realize that, if they are constrained next period according to (3), their borrowing will be limited by $\mathbb{E}_{t+1}[z_{t+2} + \zeta\Omega_{t+2}]$. For instance, picture a fully sophisticated agent ($\zeta = 1$) that is over-optimistic today ($\Omega_{t+1} > 0$) but realize that everyone will be over-pessimistic tomorrow ($\Omega_{t+2} < 0$ some states of the world). This agent think that the asset is going to pay more than in reality. But the agent realize that his consumption will be more severely impacted in a crisis because of a tight collateral constraint caused by excessive pessimism.[28]

Throughout the paper, I use a streamlined notation:

$$q_t = \beta\mathbb{E}_t\left[\frac{\lambda_{t+1}}{\lambda_t}\left(z_{t+1} + \Omega_{t+1} + q_{t+1}\right)\right] \tag{22}$$

where the dependence of the stochastic discount factor λ_{t+1}/λ_t and of the price on the behavioral biases are kept implicit, for conciseness. We now proceed to formally define the equilibrium, and analyze it in more details in the rest of this section.

Definition 3 (Equilibrium Definition)**.** *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 $\{c_1, c_2(z_2), c_2(z_2, z_3), c_1^h, c_2^h(z_2), c_2^h(z_2, z_3), H\}$, 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* (1) *and* (2) *subject to* (5)-(8) *and* (11)-(13)*; and (iii) biases are consistent with definitions 1 and 2.*

2.5 Equilibrium Analysis

I solve for the equilibrium by backward induction, starting from the intermediate period.

[28] Because what matters for agents is how tight the collateral constraint will be, and how the asset will be priced, their sophistication ultimately needs to be about whether the *market* will be biased, rather than just themselves.

Households: Households are passive throughout the three periods, and they pin down the rate of interest through their Euler equation:

$$\beta(1+r_t) = 1 \tag{23}$$

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:

$$\lambda_2 = \frac{1}{c_2} = \mathbb{E}_2[\lambda_3] = 1 \tag{24}$$

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. Finally the price of the collateral asset is simply given by:

$$q_2 = \beta\mathbb{E}_2[z_3 + \Omega_3] \tag{25}$$

Financial Crisis: When the collateral constraint is binding, the associated Lagrange multiplier on the constraint, κ_2, is given by:

$$\kappa_2 = \lambda_2 - 1 > 0. \tag{26}$$

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:

$$q_2 = \beta c_2\mathbb{E}_2[z_3 + \Omega_3] + \phi(1 - c_2)\mathbb{E}_2[z_3 + \Omega_3] \tag{27}$$

where the last term is a *collateral premium* (holding marginally more of the asset is valuable since it relaxes financial constraints). Consumption, on the other hand, is directly coming from the budget constraint of financial intermediaries (6), since agents are against the collateral constraint:

$$c_2 = z_2 H - d_1(1 + r_1) + \phi H \mathbb{E}_2[z_3 + \Omega_3]. \tag{28}$$

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, c_2, and the expectation of future dividends. But this drop in asset prices does not spill back to consumption. The left panel of Figure 1 illustrates this equilibrium determination when $\Omega_3 < 0$ with the solid line, while the rational case is depicted by the dotted line. Exogenous pessimism makes the pricing condition steeper, but consumption 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 process is represented on the right panel of Figure 1. It illustrates the presence of a new feature that I call *belief amplification*.[29] 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. The

[29]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 C.

arrow on Figure 1 illustrates the further contraction in consumption c_2 due to this belief amplification.[30]

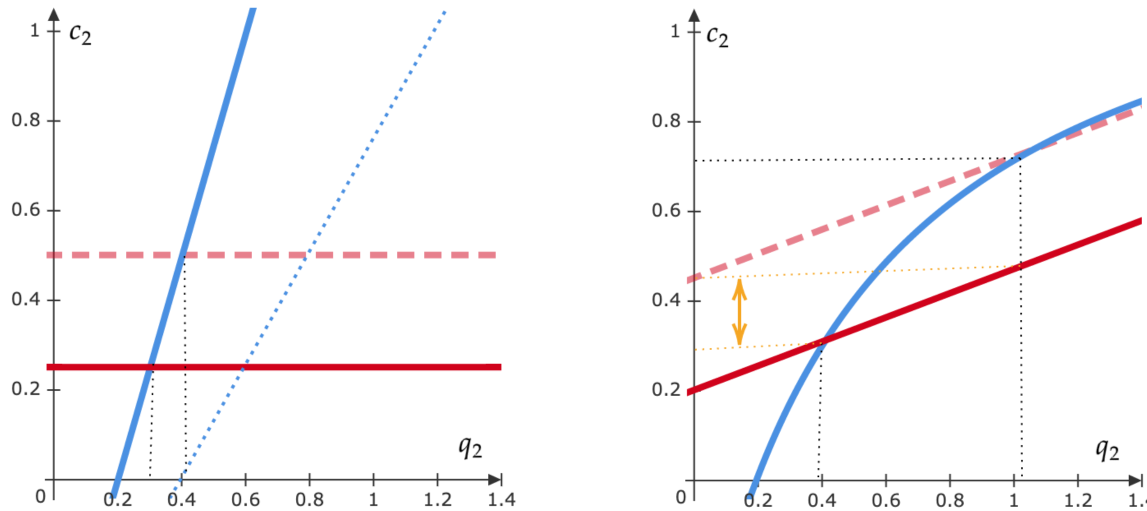

Figure 1: Graphical Illustration of Equilibrium Determination with exogenous sentiment at $t = 2$. The red line represents the budget constraint equation (28), and the blue line represents the pricing equation (27). The dotted line represents the rational pricing equation. The left panel illustrates how the equilibrium shifts after an exogenous shock to net worth n_2 (dashed line to solid line) when agents have an exogenous bias $\Omega_3 < 0$, and the blue dotted line depicts the case where $\Omega_3 = 0$. The right panel illustrates the same experiment but with an endogenous $\Omega_3(q_2) = \alpha(q_2 - q_1)$.

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

$$1 = \mathbb{E}_1 \left[\frac{\lambda_2}{\lambda_1} \right] \tag{29}$$

since financial intermediaries and households have the same time-preference parameter β. Collateral creation is driven by the pricing equation of intermediaries:

$$q_1 = c'(H) = \beta \mathbb{E}_1 \left[\frac{\lambda_2}{\lambda_1} (z_2 + \Omega_2 + q_2) \right]. \tag{30}$$

[30] As can be seen in Figure 1, the equilibrium is ensured to be unique in the exogenous sentiment case. This is not immediate anymore for endogenous sentiment. Appendix G 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.

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 consumption and the realized one is driven, for the case of unsophisticated agents, by *future* sentiment, since an $\Omega_3 < 0$ at $t = 2$ leads to a tighter collateral constraint.

3 Welfare Decomposition

This section describes the constrained planning problem of the planner, and how it differs from the decentralized allocation. The constrained problem is the following: at $t = 1$, a social planner realizes that agents are subject to behavioral biases. Taking these biases into account, the planner seeks to finds the allocation that maximizes welfare, knowing that the future allocation (at $t = 2$) will be determined by private agents.[31]. More formally, the paternalistic social planner evaluates welfare with two key distinctions relative to atomistic behavioral agents:

1. The social planner takes general equilibrium effects into account ;
2. The social planner uses its own expectations and knows that private agents are subject to a bias Ω_2, and that they can be subject to a bias Ω_3 in the future.

I adopt the notation $\mathbb{E}^{SP}$ to denote expectations formed according to this process, and use $\mathbb{E}$ for the expectations of private agents.

[31] This implies that the planner will not intervene at $t = 2$, a common assumption in this literature (Dávila and Korinek 2018). Appendix E studies the case where the planner can also influence allocations inside a crisis. This case brings additional forces that complicates the analysis, but all the uninternalized welfare effects that I highlight here are still present in this extension. This is the case as long as the planner cannot fully undone financial frictions. See Benigno, Chen, Otrok, Rebucci, and Young (2023) for a study of optimal policy in a rational setup, where a Ramsey planner can replicate the unconstrained allocation with subsidies during crises.

3.1 Decomposition

One contribution of this paper is to precisely identify how behavioral biases impact welfare. I present a general decomposition in the spirit of Dávila and Korinek (2018), that fleshes out how a marginal increase in leverage or in investment 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.

I start by analyzing how changes in debt d_1, investment H, and prices q_1, each fixing all others variables at $t = 1$, affect the welfare of individual agents.

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:

$$\mathcal{W}_d = \underbrace{\left(\mathbb{E}_1[\lambda_2] - \mathbb{E}_1^{SP}[\lambda_2]\right)}_{\mathcal{B}_d} - \underbrace{\mathbb{E}_1^{SP}\left[\kappa_2 \phi H \frac{\boldsymbol{d\Omega_3}}{\boldsymbol{dq_2}} \frac{dq_2}{dn_2}\right]}_{\mathcal{C}_d}; \tag{31}$$

ii) For investment in collateral assets H:

$$\mathcal{W}_H = \underbrace{\left(\beta \mathbb{E}_1^{SP}\left[\lambda_2(z_2 + q_2)\right] - \lambda_1 q_1\right)}_{\mathcal{B}_H} + \underbrace{\beta \mathbb{E}_1^{SP}\left[\kappa_2 \phi H \frac{\boldsymbol{d\Omega_3}}{\boldsymbol{dq_2}} \left(\frac{dq_2}{dn_2} z_2 + \frac{dq_2}{dH}\right)\right]}_{\mathcal{C}_H}; \tag{32}$$

iii) And for prices q_1:

$$\mathcal{W}_q = \beta \mathbb{E}_1^{SP}\left[\kappa_2 \phi H \frac{\boldsymbol{d\Omega_3}}{\boldsymbol{dq_1}}\right]. \tag{33}$$

Proof. All proofs are provided in Appendix A. □

Two classes of effects appear in this Proposition, which I now explore in turns.

3.2 Behavioral Wedges

The first type of effects (the terms $\mathcal{B}_d$ and $\mathcal{B}_H$ of equations 31 and 32) 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. The strength of the behavioral wedge is driven by the difference in the expected severity of crisis. Indeed, because of the linearity of utility in the last period, the marginal utility of financial intermediaries is constant outside of a crisis, while even 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:*[32]

$$\mathcal{B}_d \simeq \underbrace{-\boldsymbol{\Omega_2} H \mathbb{E}_1^{SP}\left[\boldsymbol{\lambda_2^2} \mathbb{1}_{\kappa_2>0}\right]}_{(i)} + \underbrace{\phi H(\mathbf{1}-\boldsymbol{\zeta})\mathbb{E}_1^{SP}\left[\boldsymbol{\Omega_3}\boldsymbol{\lambda_2^2}\mathbb{1}_{\kappa_2>0}\right]}_{(ii)}. \tag{34}$$

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 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. An interesting case in point of equation (34) is that even if

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

$\Omega_2 = 0$, welfare losses are possible because of the predictable behavior of *future* irrationality. Even if, on average, there is no deviation from rationality (i.e. $\mathbb{E}_1^{SP}[\Omega_3] = 0$), 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 (34) 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 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:

$$\mathcal{C}_d = -\mathbb{E}_1^{SP}\left[\kappa_2 \phi H \boldsymbol{\frac{d\Omega_3}{dq_2}} \frac{dq_2}{dn_2}\right]. \tag{35}$$

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. The fraction $d\Omega_3/dq_2$ 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 aggravating 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 equation (18) will not feature such a market failure. In the natural benchmark of price extrapolation, as in equation (19), this derivative is simply $d\Omega_3/dq_2 = \alpha_q > 0$. This externality is then pushing the private solution towards excessive borrowing.[33]

More surprisingly, the collateral externality for investment in Proposition 1 can 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 prices and thus consequently reduce pessimism. In turn, this ameliorates the borrowing capacity of the whole economy, thus improving welfare.[34] Irrational exuberance thus helps overcome the under-investment problem coming from financial frictions.[35] 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

[33]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:

$$\frac{dq_2}{dn_2} = \frac{(\beta - \phi)\mathbb{E}_2[z_3 + \Omega_3]}{1 - (\phi + (\beta - \phi)(2c_2 - n_2))\frac{d\Omega_3}{dq_2}}. \tag{36}$$

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 $d\Omega_3/dq_2 > 0$.

[34]In models where assets change hands in equilibrium, as in 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.

[35]This is reminiscent of Martin and Ventura (2016), where the presence of bubbles alleviates financing frictions.

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:

$$\mathcal{W}_q = \beta \mathbb{E}_1^{SP}\left[\kappa_2 \phi H \frac{\boldsymbol{d\Omega_3}}{\boldsymbol{dq_1}}\right]. \tag{37}$$

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 our 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 reversal 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.[36]

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

[36]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.

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 4 (Sentiment during crises). A majority of the effects presented in Proposition 1 operates through the interaction of irrationality with the health of financial intermediaries during crises. Suggestive evidence supports the assumption that the two objects Ω_3 and $\lambda_2 \mathbb{1}_{\kappa_2 > 0}$ negatively comove. Figure 2 uses two proxies to construct time series for Ω_3 and for λ_2. For the marginal utility of intermediaries λ_2, I rely on He, Kelly, and Manela (2017) who compute an intermediary capital ratio (inversely proportional to λ_2 when agents have log-utility). For Ω_3, I use the forecast errors made by stock market analysts on the long-run growth of stocks, a measure from Bordalo, Gennaioli, La Porta, and Shleifer (2020) which is directly constructed from survey data. Figure 2 shows how Ω_3 is consistently negative in crises.[37,38]

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

[37] In 2008, forecast errors Ω_3 crashed while the marginal utility of intermediaries λ_2 spiked, suggesting sizable welfare losses. But in events such as the dot-com bubble burst, pessimism was not accompanied by declines in the financial health of intermediaries. The theory I am developing suggests that these events are less of a concern for welfare.

[38] See also Bordalo et al. (2018) and Maxted (2020) for other examples of over-pessimism. An earlier version (Fontanier 2022) also presents other alternatives to measure sentiment, leading to the same result.

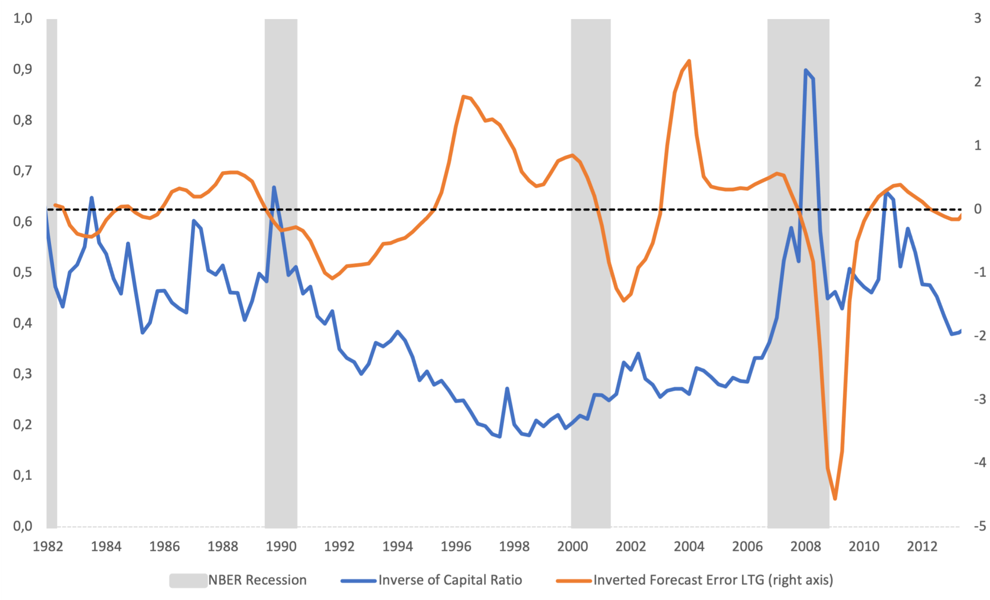

Figure 2: Time-series variation of proxies for λ_2 and Ω_3. For the financial health of intermediaries λ_2, I rely on He et al. (2017) which computes an intermediary capital ratio. The inverse of this capital ration is proportional to λ_2 when agents have log-utility, as in this model. For Ω_3, I use the inverted forecast errors made by stock market analysts on the long-term growth of stocks, a measure of Bordalo et al. (2020)

to the same constraints as agents, with prices determined by market-clearing as in the decentralized case, evaluates welfare using its own expectations and thus takes the previous uninternalized effects into account. These objects are crucial to characterize optimal policy in this setting.[39]

I first start with a natural proposition: 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 ;*

[39]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? In particular, any direct intervention at $t = 2$ is proscribed. Appendix E allows for the simultaneous choice of ex-ante and ex-post policies. In particular, it shows that in my framework the possibility of intervention at $t = 2$ does not change the desirability of macroprudential interventions at $t = 1$ (Jeanne and Korinek 2020).

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 driving the uninternalized effects from the previous part, $\mathcal{W}_d$ and $\mathcal{W}_H$: (i) the current extent of sentiment Ω_2 ; (ii) the future covariance of $(1-\zeta)\Omega_3$ with λ_2, conditional on Ω_2 ; (iii) the sensitivity of sentiment with respect to current and future prices. Second, when current asset prices impact future sentiment, three instruments are needed to achieve the second-best, and not only two.[40]

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. 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 (see Figure 2), the optimal value of this macroprudential leverage tax is also time-varying. But

[40] 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.

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.

Most macroprudential regulations on capital structure are nonetheless set in terms of leverage limits rather than leverage taxes. Are a leverage tax and a leverage limit equivalent in this model? Weitzman (1974) showed that whether price or quantity regulation is more desirable depends on which one is more robust to changes in parameters. Here, when financial intermediaries are against the regulatory leverage limit, an increase in sentiment Ω_2 increases their incentives to take on more debt, but agents simply cannot change their positions. Their leverage thus stays at the exact allocation desired by the social planner. This is not the case for debt taxes, as can be seen from equation (34): the tax needs to be calibrated at the exact level of Ω_2 to achieve the second-best. This intuition leads to the following proposition.

Proposition 4 (Leverage Limits Robustness)**.** *Leverage limits are 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.*

The intuition can also be seen when sentiment moves downward, towards less exuberance. For small departures from an equilibrium with $\Omega_2 \geq 0$, movements in Ω_2 on the downside do not call for changing the allocation desired by the planner, because the pecuniary externality still needs to be corrected. A leverage limit thus stays binding for agents, while a leverage tax would force financial intermediaries to decrease their leverage below the socially desirable outcome.[41,42]

This insight, however, does not imply that counter-cyclical restrictions are not desirable when a flat leverage limit is imposed. This is because, as explicit from Proposition

[41]Note that quantity *caps* are enough only because Ω_2 is assumed to be positive. In the event that agents are too pessimistic initially, the regulator would need to subsidize borrowing and thus caps fail to achieve this objective.

[42]The previous working paper Fontanier (2022) additionally shows that a leverage limit is robust to the introduction of belief heterogeneity in the model, while a leverage tax becomes less efficient. See Clayton and Schaab (2020) and Jeanne and Korinek (2020) for the disparities between price and quantity regulations in macroprudential policy.

1, the behavior of future sentiment matters as much as the extent of contemporaneous irrational exuberance from the perspective of period $t = 1$. As long as the planner's estimate of Ω_3 given the information available is time-varying, and agents are not fully sophisticated, the leverage limit needs to be tightened or relaxed accordingly.

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 counter-cyclical 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 is because my model is the first to highlight the need for an additional instrument that complements traditional macroprudential tools like counter-cyclical capital buffers and LTV ratios. From an abstract perspective, this instrument can be modeled as a tax on asset holdings. But the concrete goal is to directly manipulate asset prices through the demand for these assets. A direct tax on asset holdings, however, seems rather unrealistic to implement. A more natural candidate for this instrument is to use monetary policy, which is the object of the next section.

4.3 Monetary Policy as Price Regulation

A large part of the "leaning vs. cleaning" policy debate revolves around the possible use of a monetary tightening to tame asset prices in the face of irrational exuberance. The conventional view holds that "monetary policy is not a useful tool for achieving this objective" (Bernanke 2002). Recent work challenged this perspective: Caballero and Simsek (2020) show that when traditional macroprudential policy is constrained, leaning against the wind with a monetary tightening is valuable. This occurs because the gains from a preventive tightening are first-order, while the losses from deviating from perfect inflation targeting are only second-order (assuming that the output gap can be perfectly closed). As anticipated, my model features a different channel through which monetary policy can affect welfare: the reversal externality.

I start by introducing rigidities in order for monetary policy deviations to have potential costs. 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.[43] Households now have the following utility function:

$$U^h = \mathbb{E}_1 \left[\left(\ln(c_1^h) - \nu \frac{l_1^{1+\eta}}{(1+\eta)} \right) + \beta c_2^h + \beta^2 c_3^h \right] \tag{38}$$

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

[43] 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 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.

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 5 (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:*

$$\begin{aligned}\frac{d\mathcal{W}_1}{dr_1} = \; & \underbrace{\frac{dY_1}{dr_1}\mu_1}_{(i)} + \underbrace{\frac{dd_1}{dr_1}\mathcal{W}_d}_{(ii)} + \underbrace{\frac{dH}{dr_1}\mathcal{W}_H}_{(iii)} \\ & + \underbrace{\frac{d\Omega_2}{dq_1}\frac{dq_1}{dr_1}\left(\frac{dd_1}{d\Omega_2}\mathcal{W}_d + \frac{dH}{d\Omega_2}\mathcal{W}_H\right)}_{(iv)} + \underbrace{\frac{dq_1}{dr_1}\beta\mathbb{E}_1\left[\kappa_2\phi H\frac{d\Omega_3}{dq_1}\right]}_{(v)} \qquad (39)\end{aligned}$$

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 economy 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:

$$\frac{d\mathcal{W}_1}{dr_1} = \frac{dH}{dr_1}\mathcal{W}_H \tag{40}$$

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.[44]

Notice from equation (39) 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 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 $t = 1$.[45]

[44] My framework also abstracts from other considerations that could argue against tightening in such a situation. For example, Martinez-Miera and Repullo (2019) show that tightening can increase risk-taking by financial institutions by shifting investment towards riskier firms. Allen, Barlevy, and Gale (2022) present a model where risk-shifting raises asset prices above fundamentals, but tighter monetary policy further decreases investment that is already underfunded. Caines and Winkler (2021) and Adam and Woodford (2021) are other examples where the central bank leans against the wind of high house prices to avoid excessive investment in housing. In my framework such welfare costs are primarily addressed using LTV/LTI tools that directly target investment inefficiencies. Drechsler, Savov, and Schnabl (2022) also argues that the monetary tightening of 2003-2006 contributed to the increase housing prices. In that case, leaning against the wind backfires (see Proposition 6).

[45] 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.

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.[46] Inspecting Proposition 5, 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.[47]

Proposition 6 (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:*

$$\frac{d\mathcal{W}_1}{dr_1} = \frac{dY_1}{dr_1}\mu_1 + \beta\mathbb{E}_1\left[\kappa_2\phi H\frac{d\Omega_3}{dq_1}\frac{dq_1}{dr_1}\right]. \tag{41}$$

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.[48] 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 this into account leads the optimal macroprudential limit to be less strict, which raises welfare.[49]

Finally, notice how the sophistication parameter ζ is naturally absent of Proposition

[46]See e.g. Dudley (2015) ; Caballero and Simsek (2020) and Stein (2021).

[47]This is similar to Farhi and Werning (2020) when they consider the presence of biases during crises. The main difference is that my results do not depend on the inability of the central bank to change interest rates during the bust.

[48]Barlevy (2022) notes too that in his model monetary policy is effective only if it can cure optimism or pessimism.

[49]Evidently, this policy problem is also plagued with uncertainty. Section 5.2 shows that the incentives to tighten monetary policy are increasing in the uncertainty around the strength of the reversal force.

6. 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.

4.4 Small Deviations from Rationality

Suppose that we place ourselves at the REE constrained-efficient allocation. Agents are fully rational, so 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 (31), (32), and (33). At the rational expectations constrained-efficient equilibrium, behavioral wedges are zero, so the only effects left are the collateral externalities and the reversal externality:

$$\mathcal{C}_d = -\beta \mathbb{E}_1^{SP} \left[\kappa_2 \phi H \frac{\boldsymbol{d\Omega_3}}{\boldsymbol{dq_2}} \frac{dq_2}{dn_2} \right] \tag{42}$$

$$\mathcal{C}_H = \beta \mathbb{E}_1^{SP} \left[\kappa_2 \phi H \frac{\boldsymbol{d\Omega_3}}{\boldsymbol{dq_2}} \left(\frac{dq_2}{dn_2} z_2 + \frac{dq_2}{dH} \right) \right] \tag{43}$$

$$\mathcal{W}_q = \beta \mathbb{E}_1^{SP} \left[\kappa_2 \phi H \frac{\boldsymbol{d\Omega_3}}{\boldsymbol{dq_1}} \right] \tag{44}$$

which, as explicated earlier, are only present when *future* sentiment is impacted by *current* and *past* asset prices, 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. This result draws attention to irrational distress during financial crises, while the literature has mostly focused on irrational exuberance during the build-up leading to the crash.[50]

[50] Of course irrational exuberance is 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).

5 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.

5.1 Ω-Uncertainty

I start with 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.[51] 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 assumptions make the analysis more convenient:

Assumption 3. *All parameters of the model and of the probability density function $f_2(z_2)$ are common knowledge to private agents and the social planner, except possibly for its mean, $\bar{z}_2$.*

Assumption 4. *Equilibrium prices at time $t = 1$ are strictly increasing in $\bar{z}_2$.*

These assumptions imply that, in the absence of sentiment, the social planner could simply infer the value of $\bar{z}_2$ by looking at equilibrium prices in period 1, q_1. Let me now assume that the social planner's prior over sentiment is given by a uniform distribution:

$$w \sim \mathcal{U}\left[\bar{\Omega}_2 - \sigma_\Omega, \bar{\Omega}_2 + \sigma_\Omega\right] \tag{45}$$

[51] The analysis for Ω_3-uncertainty is presented in Appendix D. The results are identical.

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 regarding future dividends becomes:

$$\bar{z}_2 = g_q^{-1}(q_1) - \bar{\Omega}_2 \tag{46}$$

Taking the uncertainty in its posterior into account, the planner first-order condition for short-term debt is now given by:

$$u'(c_1) = \frac{1}{2\sigma_\Omega} \int_0^\infty \left[\int_{-\sigma_\Omega}^{\sigma_\Omega} \frac{\partial \mathcal{W}_2}{\partial n_2} \left(d_1, H; q_2, z_2 - \bar{\Omega}_2 - \omega_2 \right) d\omega_2 \right] f_2(z_2) dz_2 \tag{47}$$

This expression contains all of the intuition for how sentiment uncertainty can reinforce or weaken the need for preventive leverage tightening.[52] 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 sentiment uncertainty increases expected marginal utility. This, in turn, 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. Using the analysis of the equilibrium presented in Section 2.5, uncertainty about behavioral biases unambiguously calls for precautionary restrictions, as expressed in the following Proposition.

Proposition 7 (Ω_2-Uncertainty and Leverage Restrictions)**.** *If the social planner believes that*

[52] Note that $\bar{\Omega}_2$ is an argument of $\mathcal{W}_2$ here, because the planner is using the same distribution agents are using. In the previous Section, I was using an equivalent notation where private agents use the same distribution as the planner but add their bias $\bar{\Omega}_2$. The two are of course equivalent, but in the present section this clarifies that the planner simply extracts the distribution used by private agents.

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.8), but the intuition can be understood succinctly. The key is to notice that marginal welfare is a *convex* function, as shown in Figure 3. Intuitively, sentiment uncertainty adds terms to the expectation computed by the planner relative to the private solution, but the parts coming from intermediaries' optimism are more costly than the ones coming from pessimism.[53] The strong non-linearities 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.

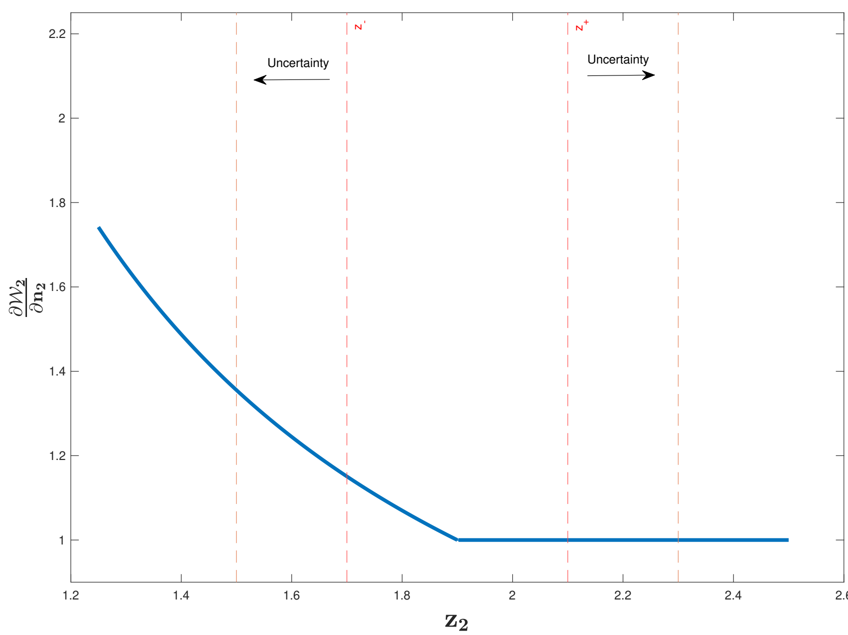

Figure 3: Non-linearities and Ω-uncertainty. This figure plots $\partial \mathcal{W}_2 / \partial n_2$ against the fundamental realization z_2. The thick dotted lines correspond to the range of values of z_2 where the expectations is taken. The thin dotted lines represent the widening of the range where the social planner computes expectations caused by the uncertainty on Ω_2. The break arises at the crisis cutoff.

Proposition 7 also highlights that investment regulation behaves in a different way:

[53] The same insights can be obtained if we were to consider uncertainty about the extent of sentiment *inside* a financial crisis. Furthermore, endogenous sentiment, for example in the form of price extrapolation, amplifies this effect by adding *more* curvature. These results are presented in Appendix D.

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 that is going to pay 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.

5.2 Reversal Uncertainty

The previous analysis 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 last proposition answers this interrogation unambiguously.

Proposition 8 (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.

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. Importantly, 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. Finally, I showed that adding uncertainty about the extent of behavioral biases in financial markets reinforced incentives for leverage regulation, as well as for the use of monetary policy to lean against the wind.

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 empirical and theoretical research is needed to fully grasp the complex timing interactions between policy, crises, and behavioral biases.

References

Adam, K. and A. Marcet (2011). Internal rationality, imperfect market knowledge and asset prices. *Journal of Economic Theory 146*(3), 1224–1252.

Adam, K., A. Marcet, and J. Beutel (2017a). Stock price booms and expected capital gains. *American Economic Review 107*(8), 2352–2408.

Adam, K., A. Marcet, and J. Beutel (2017b). Stock price booms and expected capital gains. *American Economic Review 107*(8), 2352–2408.

Adam, K., O. Pfäuti, and T. Reinelt (2022). Subjective housing price expectations, falling natural rates and the optimal inflation target.

Adam, K. and M. Woodford (2021). Robustly optimal monetary policy in a new keynesian model with housing. *Journal of Economic Theory 198*, 105352.

Allen, F., G. Barlevy, and D. Gale (2022). Asset price booms and macroeconomic policy: A risk-shifting approach. *American Economic Journal: Macroeconomics*.

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., R. Greenwood, L. Jin, and A. Shleifer (2018). Extrapolation and bubbles. *Journal of Financial Economics 129*(2), 203–227.

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.

Baron, M. and W. Xiong (2017). Credit expansion and neglected crash risk. *The Quarterly Journal of Economics 132*(2), 713–764.

Bastianello, F. and P. Fontanier (2022). Partial Equilibrium Thinking in General Equilibrium. *Working Paper*.

Benigno, G., H. Chen, C. Otrok, A. Rebucci, and E. R. Young (2023). Optimal policy for macrofinancial stability. *American Economic Journal: Macroeconomics 15*(4), 401–428.

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 (2021). Diagnostic business cycles. *NBER Working Paper*.

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 (2020). Expectations of fundamentals and stock market puzzles. *NBER Working Paper*.

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

Bordalo, P., N. Gennaioli, A. Shleifer, and S. J. Terry (2021). Real credit cycles. *NBER Working Paper*.

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

Bouchaud, J.-P., P. Krueger, A. Landier, and D. Thesmar (2019). Sticky expectations and the profitability anomaly. *The Journal of Finance 74*(2), 639–674.

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 (2020). Prudential monetary policy. *MIT Working Paper*.

Caines, C. and F. Winkler (2021). Asset price learning and optimal monetary policy. *Journal of Monetary Economics 123*, 53–67.

Camous, A. and A. Van der Ghote (2021). Financial cycles under diagnostic beliefs. *Working Paper*.

Chahrour, R. and G. Gaballo (2021). Learning from house prices: Amplification and business fluctuations. *The Review of Economic Studies 88*(4), 1720–1759.

Chodorow-Reich, G., A. M. Guren, and T. J. McQuade (2021). The 2000s housing cycle with 2020 hindsight: A neo-kindlebergerian view. *NBER Working Paper*.

Cieslak, A. (2018). Short-rate expectations and unexpected returns in treasury bonds. *The Review of Financial Studies 31*(9), 3265–3306.

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

Clayton, C. and A. Schaab (2020). Multinational banks and financial stability. *Working Paper*.

Coibion, O. and Y. Gorodnichenko (2012). What can survey forecasts tell us about information rigidities? *Journal of Political Economy 120*(1), 116–159.

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*.

DeFusco, A. A., C. G. Nathanson, and E. Zwick (2017). Speculative dynamics of prices and volume. Technical report, National Bureau of Economic Research.

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.

Dudley, W. (2015). Is the active use of macroprudential tools institutionally realistic? *Federal Reserve Bank of New York Working Paper*.

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 J. Tirole (2012). Collective moral hazard, maturity mismatch, and systemic bailouts. *American Economic Review 102*(1), 60–93.

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*.

Fuster, A., B. Hebert, and D. Laibson (2012). Natural expectations, macroeconomic dynamics, and asset pricing. *NBER Macroeconomics Annual 26*(1), 1–48.

Gaspar, V., F. Smets, and D. Vestin (2010). Inflation expectations, adaptive learning and optimal monetary policy. In *Handbook of monetary economics*, Volume 3, pp. 1055–1095. Elsevier.

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. and S. G. Hanson (2013). Issuer quality and corporate bond returns. *The Review of Financial Studies 26*(6), 1483–1525.

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.

Gruber, J. and B. Köszegi (2001). Is addiction "rational"? theory and evidence. *The Quarterly Journal of Economics 116*(4), 1261–1303.

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.

Jeanne, O. and A. Korinek (2020). Macroprudential regulation versus mopping up after the crash. *The Review of Economic Studies 87*(3), 1470–1497.

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

Kahneman, D. and A. Tversky (1972). Subjective probability: A judgment of representativeness. *Cognitive psychology 3*(3), 430–454.

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

Kirti, D. (2018). Lending standards and output growth. *IMF Working paper*.

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

Krishnamurthy, A. and W. Li (2020). Dissecting mechanisms of financial crises: Intermediation and sentiment. *NBER Working Paper*.

Krishnamurthy, A. and T. Muir (2020). How credit cycles across a financial crisis. *NBER Working Paper*.

L'Huillier, J.-P., S. R. Singh, and D. Yoo (2021). Incorporating diagnostic expectations into the new keynesian framework. *Working Paper*.

López-Salido, D., J. C. Stein, and E. Zakrajšek (2017). Credit-market sentiment and the business cycle. *The Quarterly Journal of Economics 132*(3), 1373–1426.

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

Ma, Y., T. Ropele, D. Sraer, and D. Thesmar (2020). A quantitative analysis of distortions in managerial forecasts. *NBER Working Paper*.

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. (2020). A Macro-Finance Model with Sentiment. *Working Paper*.

McCarthy, O. and S. Hillenbrand (2021). 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.

Mian, A., A. Sufi, and E. Verner (2017). Household debt and business cycles worldwide. *The Quarterly Journal of Economics 132*(4), 1755–1817.

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

Mullainathan, S., J. Schwartzstein, and W. J. Congdon (2012). A reduced-form approach to behavioral public finance. *Annual Review of Economics 4*(1), 511–540.

Odean, T. (1998). Volume, volatility, price, and profit when all traders are above average. *The journal of finance 53*(6), 1887–1934.

O'Donoghue, T. and M. Rabin (2006). Optimal sin taxes. *Journal of Public Economics 90*(10-11), 1825–1849.

Ottonello, P., D. J. Perez, and P. Varraso (2022). Are collateral-constraint models ready for macroprudential policy design? *Journal of International Economics*, 103650.

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.

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 (2021). Financial crises: A survey. *NBER Working Paper*.

Weitzman, M. L. (1974). Prices vs. quantities. *The Review of Economic Studies 41*(4), 477–491.

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

Winkler, F. (2020). The role of learning for asset prices and business cycles. *Journal of Monetary Economics 114*, 42–58.

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 can be written as:

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

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:

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

the first-order condition on borrowing gives:

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

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:

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

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

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

which proves the first part of Proposition 1.

Investment: At time $t = 2$, the welfare of borrowers can be written as:

$$\mathcal{W}_2 = \begin{cases} \beta \ln\left(n_2 + \phi H \mathbb{E}_2[z_3 + \Omega_3(q_2, q_1)]\right) + \beta^2 \left(\mathbb{E}_2[z_3] H - \phi H \mathbb{E}_2[z_3 + \Omega_3(q_2, q_1)]/\beta\right) & \text{if } z_2 \geq z^* \\ \beta\left(\beta \mathbb{E}[z_3] H + n_2\right) & \text{otherwise} \end{cases} \tag{A.6}$$

while the Lagrangian corresponding to bankers' problem in period $t = 1$ is given by:

$$\mathcal{L}_{b,1} = \left[u(c_1) + \mathbb{E}_1[\mathcal{W}_2(n_2, H; q_2, z_2\Omega_3(q_2, q_1))]\right] - \lambda_1\left[c_1 + c(H) - d_1 - e_1\right] \tag{A.7}$$

the first-order condition on investment yields:

$$\frac{\partial \mathcal{L}_{b,1}}{\partial H} = -\lambda_1 c'(H) + \beta \mathbb{E}_1\left[\lambda_2 (z_2 + \Omega_2)(z_2 + \Omega_2 + q_2(z_2 + \Omega_2, \Omega_3(q_2, q_1)))\right] \tag{A.8}$$

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:

$$\frac{\partial \mathcal{L}_{b,1}^{SP}}{\partial H} = \beta \mathbb{E}_1^{SP}\left[\lambda_2(z_2 + q_2)\right] - \lambda_1 c'(H) + \beta \mathbb{E}_1^{SP}\left[\kappa_2 \phi H \frac{\partial \Omega_3}{\partial q_2}\left(\frac{\partial q_2}{\partial n_2} z_2 + \frac{\partial q_2}{\partial H}\right)\right] \tag{A.9}$$

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

$$\frac{\partial \mathcal{L}^{SP}_{b,1}}{\partial H} = \underbrace{\beta \mathbb{E}_1\left[\lambda_2(z_2+q_2)\right] - \lambda_1 q_1}_{\text{Internalized}} + $$

$$\underbrace{\beta \mathbb{E}^{SP}_1\left[\lambda_2(z_2+q_2)\right] - \beta \mathbb{E}_1\left[\lambda_2(z_2+q_2)\right] + \beta \mathbb{E}^{SP}_1\left[\kappa_2 \phi H \frac{\partial \Omega_3}{\partial q_2}\left(\frac{\partial q_2}{\partial n_2} z_2 + \frac{\partial q_2}{\partial H}\right)\right]}_{\text{Uninternalized}}. \quad \text{(A.10)}$$

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:

$$\frac{d\mathcal{W}_1}{q_1} = \beta \mathbb{E}^{SP}_1[\lambda_2 \phi H \frac{d\Omega_3}{dq_1} - \beta \phi H \frac{d\Omega_3}{dq_1}(1+r_2)] \quad \text{(A.11)}$$

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:

$$\mathcal{W}_q = \beta \mathbb{E}^{SP}_1[\kappa_2 \phi H \frac{d\Omega_3}{dq_1}] \quad \text{(A.12)}$$

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:

$$\frac{1}{c_2(\Omega_2, \zeta\Omega_3)} - \frac{1}{c_2(0, \Omega_3)} =$$
$$\frac{1}{(\Omega_2)H - d_1(1+r_1) + \phi H \mathbb{E}_2[z_3 + \zeta\Omega_3]} - \frac{1}{z_2 H - d_1(1+r_1) + \phi H \mathbb{E}_2[z_3 + \Omega_3])} \quad \text{(A.13)}$$

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:

$$\frac{1}{(z_2+\Omega_2)H - d_1(1+r_1) + \phi H \mathbb{E}_2[z_3 + \zeta\Omega_3]} = \frac{1}{c_2(0,0)} \frac{1}{1 + \frac{\Omega_2 H}{c_2(0,0)} + \frac{\phi H \zeta \Omega_3}{c_2(0,0)}}$$
$$= \lambda_2 \left(1 - \frac{\Omega_2 H + \phi H \zeta \Omega_3}{c_2(0,0)}\right) \quad \text{(A.14)}$$

While the same algebra for the second part of equation (A.13) yields similarly:

$$\frac{1}{z_2 H - d_1(1+r_1) + \phi H \mathbb{E}_2[z_3 + \Omega_3]} = \frac{1}{c_2(0,0)} \frac{1}{1 + \frac{\phi \Omega_3 H}{c_2(0,0)}} = \lambda_2 \left(1 + \frac{\phi H \Omega_3}{c_2(0,0)}\right) \quad \text{(A.15)}$$

Taking the difference gives:

$$\frac{1}{c_2(\Omega_2, \zeta\Omega_3)} - \frac{1}{c_2(0, \Omega_3)} = \lambda_2^2 \left(H\Omega_2 - (1-\zeta)\phi H \Omega_3\right) \quad \text{(A.16)}$$

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:

$$\mathcal{B}_d \simeq -\boldsymbol{\Omega_2} H \mathbb{E}^{SP}\left[\boldsymbol{\lambda_2^2} \mathbb{1}_{\kappa_2>0}\right] + \phi H \boldsymbol{(1-\zeta)} \mathbb{E}^{SP}\left[\boldsymbol{\Omega_3 \lambda_2^2} \mathbb{1}_{\kappa_2>0}\right] \quad \text{(A.17)}$$

A.3 Derivation of Equation (36)

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

$$dq_2 = \beta dc_2 \mathbb{E}_2[z_3 + \Omega_3] + \beta c_2 d\Omega_3 - \phi dc_2 \mathbb{E}_2[z_3 + \Omega_3] + \phi(1 - c_2) d\Omega_3. \quad \text{(A.18)}$$

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

$$dq_2 = \beta(dn_2 + \phi H d\Omega_3)\mathbb{E}_2[z_3 + \Omega_3] + \beta c_2 d\Omega_3 \\ - \phi(dn_2 + \phi H d\Omega_3)\mathbb{E}_2[z_3 + \Omega_3] + \phi(1 - c_2)d\Omega_3. \quad \text{(A.19)}$$

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

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

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:

$$\frac{dq_2}{dn_2} = \frac{(\beta - \phi)\mathbb{E}_2[z_3 + \Omega_3]}{1 - (\phi + (\beta - \phi)(2c_2 - n_2)\,\frac{d\Omega_3}{dq_2}}. \quad \text{(A.21)}$$

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:

$$\mathcal{B}_H(z_2) = [\lambda_2(0;\Omega_3)(z_2 + q_2(0;\Omega_3))] - [\lambda_2(\Omega_2;\zeta\Omega_3)(z_2 + \Omega_2 + q_2(\Omega_2;\zeta\Omega_3)] \quad \text{(A.22)}$$

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:

$$\mathcal{B}_H(z_2) = (\lambda_2(0;\Omega_3) - \lambda_2(\Omega_2;\zeta\Omega_3)(z_2 + q_2^r)) \\ + \lambda_2^r \left(\Omega_3 \frac{dq_2}{d\Omega_3} - \Omega_2\left(1 + \frac{dq_2}{d\Omega_2}\right) - \zeta\Omega_3\frac{dq_2}{d\Omega_3}\right) \quad \text{(A.23)}$$

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:

$$\mathcal{B}_H \approx \beta \mathbb{E}_1^{SP}[\mathcal{B}_d(z_2)(z_2 + q_2^r)\mathbb{1}_{\kappa_2>0}]$$
$$- \beta\Omega_2 \mathbb{E}_1^{SP}[\lambda_2(1 + (\beta - \phi)Hz_3)\mathbb{1}_{\kappa_2>0}] + \beta(1-\zeta)\mathbb{E}_1^{SP}\left[\Omega_3\lambda_2 \frac{dq_2}{dz_3}\mathbb{1}_{\kappa_2>0}\right] \quad \text{(A.24)}$$

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:

$$c_1 + c(H) + \tau_h H + q_1 h \leq e_1 + d_1(1-\tau_d) + q_1 H + \tau_q h \quad \text{(A.25)}$$

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:

$$\frac{\partial \mathcal{L}_{b,0}}{\partial d_1} = \lambda_1(1-\tau_d) - \mathbb{E}_1[\lambda_2] = 0 \quad \text{(A.26)}$$
$$\frac{\partial \mathcal{L}_{b,0}}{\partial H} = c'(H) + \tau_H - q_1 = 0 \quad \text{(A.27)}$$
$$\frac{\partial \mathcal{L}_{b,0}}{\partial h} = \lambda_1 q_1 + \lambda_1 \tau_q - \mathbb{E}_1[\lambda_2(z_2 + \Omega_2 + q_2^r)] = 0 \quad \text{(A.28)}$$

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

$$\tau_d = -\frac{\mathcal{W}_d}{\lambda_1} \quad \text{(A.29)}$$

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

$$c'(H) = \beta \mathbb{E}_1^{SP}[\lambda_2(z_2 + q_2)] + \beta \mathbb{E}_1^{SP}\left[\kappa_2 \phi H \frac{\partial \Omega_3}{\partial q_2}\left(\frac{\partial q_2}{\partial n_2} z_2 + \frac{\partial q_2}{\partial H}\right)\right] \tag{A.30}$$

and because

$$\beta \mathbb{E}_1^{SP}[\lambda_2(z_2 + q_2)] + \beta \mathbb{E}_1^{SP}\left[\kappa_2 \phi H \frac{\partial \Omega_3}{\partial q_2}\left(\frac{\partial q_2}{\partial n_2} z_2 + \frac{\partial q_2}{\partial H}\right)\right] = \beta \mathbb{E}_1[\lambda_2(z_2 + \Omega_2 + q_2)] - \mathcal{W}_H \tag{A.31}$$

the tax must simply be set equal to:

$$\tau_H = -\frac{\mathcal{W}_H}{\lambda_1} \tag{A.32}$$

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 $\lambda_1 q_1^* + \lambda_1 \tau_q - \mathbb{E}_1\left[\lambda_2(z_2 + \Omega_2 + q_2)\right] = 0$, so the tax should be set at:

$$\tau_q = \frac{\mathbb{E}_1\left[\lambda_2(z_2 + \Omega_2 + q_2)\right] - \lambda_1 q_1^*}{\lambda_1} \tag{A.33}$$

A.6 Proof of Proposition 4

I keep using the notation from the previous proof. Agents' private Euler equation when a tax is imposed on leverage is: $\lambda_1(1 - \tau_d) = \mathbb{E}_1[\lambda_2]$. Since, in a crisis, $\lambda_2(d_1, z_2 + \Omega_2, H, \zeta\Omega_3)$ is unambiguously decreasing in Ω_2, and because λ_1 is decreasing in d_1, leverage is increasing with Ω_2. As long as $\Omega_2 > 0$, and there is a positive probability of a crisis, we have $\mathcal{W}_d < 0$. It directly implies, from equation (A.5), that this decreases welfare as evaluated from the planner.

However if the policy is put in place through a leverage limit, the allocation satisfies:

$$\lambda_1 = \max\left(\lambda_1^*, \mathbb{E}_1[\lambda_2]\right) \tag{A.34}$$

Since we assumed that $\mathcal{W}_d < 0$, this necessarily implies that $\lambda_1^* > \mathbb{E}_1[\lambda_2]$. In turn, this means that for a perturbation $d\omega < 0$ to initial exuberance:

$$\lambda_1^* > \mathbb{E}_1[\lambda_2(z_2 + \Omega_2)] > \mathbb{E}_1[\lambda_2(z_2 + \Omega_2 + d\omega)] \tag{A.35}$$

so leverage stays at the optimal level desired by the planner. Finally, regarding a downward movement to Ω_2, the assumption that $\mathcal{W}_d < 0$ implies that there is a non-zero gap between λ_1^* and $\mathbb{E}_1[\lambda_2]$, such that for a small enough $d\omega > 0$, it is also guaranteed that:

$$\lambda_1^* > \mathbb{E}_1[\lambda_2(z_2 + \Omega_2 - d\omega)] > \mathbb{E}_1[\lambda_2(z_2 + \Omega_2)] \tag{A.36}$$

hence guaranteed that allocations stay at the second-best.

A.7 Proof of Proposition 5

The welfare function that the planner considers is given by:

$$\mathcal{W}_1 = \Phi^h \mathbb{E}_1^{SP}\big(\ln\left[c_1^h - \nu\frac{l_1^{1+\eta}}{1+\eta}\right] + \beta c_2^h + \beta^2 c_3^h\big) + \Phi^b \mathbb{E}_1^{SP}\big(\ln(c_1) + \beta\ln(c_2) + \beta^2 c_3\big) \tag{A.37}$$

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:

$$\mathcal{W}_1^h = \ln\left[c_1^h - \nu\frac{Y_1^{1+\eta}}{1+\eta}\right] + \beta c_2^h + \beta^2 c_3^h. \tag{A.38}$$

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:

$$\frac{d\mathcal{W}_1^h}{dr_1} = \frac{Y_1}{dr_1}\lambda_1^h - \nu Y_1^\eta \frac{Y_1}{dr_1}\lambda_1^h + \underbrace{\frac{dc_1^h}{dr_1}\lambda_1^h - \beta\mathbb{E}_1\frac{dc_1^h}{dr_1}}_{\text{Euler}=0}. \tag{A.39}$$

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

$$\begin{aligned} \frac{d\mathcal{W}_1}{dr_1} &= \frac{dY_1}{dr_1}\mu_1 + \frac{dd_1}{dr_1}\mathcal{W}_d + \frac{dH}{dr_1}\mathcal{W}_H \\ &+ \frac{d\Omega_2}{dq_1}\frac{dq_1}{dr_1}\left(\frac{dd_1}{d\Omega_2}\mathcal{W}_d + \frac{dH}{d\Omega_2}\mathcal{W}_H\right) + \mathbb{E}_1\left[\kappa_2\phi H \frac{d\Omega_3}{dq_1}\frac{dq_1}{dr_1}\right] \end{aligned} \tag{A.40}$$

A.8 Proof of Proposition 7

As explained in the main text, the social planner's optimality condition under the premises of Proposition 7 can be expressed as:

$$u'(c_1) = \frac{1}{2\sigma_\Omega}\int_0^\infty \left[\int_{-\sigma_\Omega}^{\sigma_\Omega} \frac{\partial \mathcal{W}_2}{\partial n_2}\left(d_1, H; z_2 - \bar{\Omega}_2 - \omega_2\right) d\omega_2\right] f_2(z_2) dz_2. \tag{A.41}$$

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

$$\mathcal{W}_2 = \begin{cases} \beta \ln\left(n_2 + \phi H \mathbb{E}_2[z_3]\right) + \beta^2\left(\mathbb{E}^{SP}[z_3]H - \phi H\mathbb{E}_2[z_3]/\beta\right) & \text{if } z_2 \geq z^* \\ \beta\left(\beta\mathbb{E}^{SP}[z_3]H + n_2\right) & \text{otherwise} \end{cases} \tag{A.42}$$

so that the first derivative is equal to:

$$\frac{\partial \mathcal{W}_2}{\partial n_2} = \begin{cases} \beta\lambda_2 & \text{if } z_2 \geq z^* \\ \beta & \text{otherwise} \end{cases} \tag{A.43}$$

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:

$$g(w_2) = \int_0^{+\infty} \frac{\partial \mathcal{W}_2}{\partial n_2}(d_1, H; q_2, z_2 - \bar{\Omega}_2 - \omega_2) f_2(z_2) dz_2 \tag{A.44}$$

while the integral taken over the uncertainty band is:

$$G(\sigma_\Omega) = \int_{-\sigma_\Omega}^{+\sigma_\Omega} \frac{g(w_2)}{2\sigma_\Omega} dw_2. \tag{A.45}$$

Given the continuity of $\partial\mathcal{W}_2/\partial n_2$ (see equation A.42) we can differentiate with respect to σ_Ω:

$$G'(\sigma_\Omega) = -\frac{1}{2\sigma_\Omega^2}\int_{-\sigma_\Omega}^{+\sigma_\Omega}\int_0^{+\infty} \frac{\partial\mathcal{W}_2}{\partial n_2}(d_1, H; z_2 - \bar{\Omega}_2 - \omega_2) f_2(z_2) dz_2 dw_2 +$$
$$\int_0^{+\infty} \frac{\partial\mathcal{W}_2}{\partial n_2}(d_1, H; z_2 - \bar{\Omega}_2 - \sigma_\Omega) f_2(z_2) dz_2 - \int_0^{+\infty} \frac{\partial\mathcal{W}_2}{\partial n_2}(d_1, H; z_2 - \bar{\Omega}_2 + \sigma_\Omega) f_2(z_2) dz_2 \tag{A.46}$$

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

$$G'(\sigma_\Omega) = -\frac{G(\sigma_\Omega)}{\sigma_\Omega} + \frac{1}{2\sigma_\Omega}(g(\sigma_\Omega) - g(-\sigma_\Omega)) \tag{A.47}$$

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:

$$\frac{g(\sigma_\Omega) - g(-\sigma_\Omega)}{2} > \int_{-\sigma_\Omega}^{+\sigma_\Omega} \frac{g(w_2)}{2\sigma_\Omega} dw_2. \tag{A.48}$$

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:[54]

$$\frac{1}{2}\left(\frac{\partial\mathcal{W}_2}{\partial n_2}(z_2 + \sigma_\Omega) - \frac{\partial\mathcal{W}_2}{\partial n_2}(z_2 - \sigma_\Omega)\right) > \int_{-\sigma_\Omega}^{+\sigma_\Omega} \frac{\partial\mathcal{W}_2}{\partial n_2}(z_2 + \omega_2)\frac{d\omega_2}{2\sigma_\Omega} \quad \forall z_2 \in supp(f_2). \tag{A.49}$$

[54] $\bar{\Omega}_2$ 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.

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.

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

$$\varphi(t\sigma_\Omega - (1-t)\sigma_\Omega) \leq t\varphi(\sigma_\Omega) + (1-t)\varphi(-\sigma_\Omega) \quad \forall t \in [0,1]. \tag{A.50}$$

Now integrate this inequality over t to get:

$$\int_0^1 \varphi(t\sigma_\Omega - (1-t)\sigma_\Omega)dt \leq \int_0^1 t\varphi(\sigma_\Omega)dt + \int_0^1 (1-t)\varphi(-\sigma_\Omega)dt. \tag{A.51}$$

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

$$\int_{-\sigma_\Omega}^{+\sigma_\Omega} \frac{\varphi(x)}{2\sigma_\Omega}dx \leq \frac{\varphi(\sigma_\Omega) - \varphi(-\sigma_\Omega)}{2} \tag{A.52}$$

which is exactly the relationship in equation (A.49).

We now have to prove that $\partial\mathcal{W}_2/\partial n_2$ is convex to end the proof of Proposition 7. Going back to equation (A.42), denote $\partial\mathcal{W}_2/\partial n_2$ by $\mathcal{W}_{2,n}$. Given Equation (A.43), 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.[55] For investment, the first order condition becomes:

$$\lambda_1 c'(H) = \frac{1}{2\sigma_\Omega}\int_0^\infty \left[\int_{-\sigma_\Omega}^{\sigma_\Omega} \lambda_2(z_2 - \bar{\Omega}_2 - \omega_2)(z_2 - \bar{\Omega}_2 - \omega_2 + q_2(z_2 - \bar{\Omega}_2 - \omega_2))d\omega_2\right] f_2(z_2)dz_2. \tag{A.53}$$

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:

$$q_2 = \beta(n_2 + \phi M\mathbb{E}_2[z_3])\mathbb{E}_2[z_3] + \phi(1 - n_2 - \phi M\mathbb{E}_2[z_3])\mathbb{E}_2[z_3] \tag{A.54}$$

[55] 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 $\mathbb{E}_1[z_3]$ and that has no impact on the sign of these derivatives as long as $\mathbb{E}_1[z_3] + \Omega_3 > 0$, which we always assume to be the case.

Which is clearly linear in ω_2 since net worth is linear in ω_2: $n_2 = (z_2 - \bar{\Omega}_2 - \omega_2)H - d_1(1 + r_1)$. 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.9 Proof of Proposition 8

Using the premises of Proposition 8 and the expressions in Proposition 1, we are interested in the behavior of:

$$\frac{1}{2\sigma_\alpha}\int_{\bar{\alpha}_q-\sigma_\alpha}^{\bar{\alpha}_q+\sigma_\alpha} \kappa_2(\alpha_q)\alpha_q d\alpha_q = \int_{-\sigma_\alpha}^{\sigma_\alpha} (\lambda_2(\alpha_q) - 1)\alpha_q d\alpha_q \tag{A.55}$$

when σ_α varies. As we showed in the proof of Proposition 7 in Section A.8, 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:

$$\frac{d}{d\alpha}\big((\lambda_2(\alpha_q) - 1)\alpha_q\big) = (\lambda_2(\alpha_q) - 1) + \frac{\alpha_q \phi H q_1}{c_2^2} \tag{A.56}$$

differentiating once again:

$$\frac{d^2}{d\alpha_q^2}\big((\lambda_2(\alpha_q) - 1)\alpha_q\big) = \frac{\phi H q_1}{c_2^2} + \frac{\phi H q_1}{c_2^2} + \alpha_q \frac{(\phi H q_1)^2}{2c_2^3} \tag{A.57}$$

and all those terms are unambiguously positive. By convexity the strength of the reversal externality is increasing in σ_α. By proposition 6, this immediately requires a higher interest rate to satisfy Equation (41).

B Real Production

B.1 Adding Real Production

To incorporate a real side to the model, I allow households to supply labor at $t = 2$. Households have a convex disutility for supplying labor:

$$U^h = \mathbb{E}_1 \left[c_1^h + \beta \left(c_2^h - \nu \frac{l_2^{1+\eta}}{(1+\eta)} \right) + \beta^2 c_3^h \right] \tag{B.1}$$

where l_2 is the amount of labor supplied by households at time $t = 2$.

There is a fringe of competitive firms of measure one, producing from the labor of households. Firms use a decreasing returns to scale technology from labor, with productivity A:

$$Y_2 = A l_2^{\alpha} \tag{B.2}$$

To bridge the gap between Main street and Wall street, I add another financial friction. Firms need to pay a fraction γ of wage bills in advance to workers, which requires them to borrow $f_2 = \gamma w_2 l_2$ from financial intermediaries. We assume that the interest rate required by financial intermediaries to advance such funds depends on the size of the loan according to:

$$1 + r_f = \frac{\delta}{f_2}. \tag{B.3}$$

This innocuous trick allows the model to say away from corner solutions.[56] The budget constraints at $t = 2$ are now given by:

$$c_2^h + d_2 \leq e_2^h + w_2 l_2 + d_1(1 + r_1) + \pi_2 \tag{B.4}$$

$$c_2 + d_1(1 + r_1) + f_2 + q_2 h \leq d_2 + (z_2 + q_2) H \tag{B.5}$$

respectively for households and financial intermediaries. Household optimization then

[56] This also allows for belief amplification to survive. Remember that belief amplification comes from the two-way feedback effect between the stochastic discount factor and the price of the risky asset. A corner solution with respect to the borrowing of real firms would break this link.

simply yields $w_2 = \nu l_2^\eta$. It is also assumed for simplicity that loans made to firms cannot be used as collateral.[57] The specific form assumed in (B.3) simplifies matter since funds allocated to firms verify the following identity:

$$\frac{f_2}{\delta} = \beta c_2 \tag{B.7}$$

so that bankers' consumption and funds allowed to firms are proportional. Intuitively, when collateral constraints are extremely tight, this forces financial intermediaries to cut back on consumption *and* on their lending activities in the same way.[58] Thus inside a crisis the amount of labor used for production verifies:

$$l_2 = \left(\frac{z_2 H - d_1(1+r_1) + \phi H \mathbb{E}_2[z_3 + \Omega_3]}{\gamma \nu \left(1 + \frac{1}{\beta\delta}\right)} \right)^{\frac{1}{1+\eta}} \tag{B.8}$$

which translates into a production level at time $t = 2$ of:

$$Y_2 = A \left(\frac{z_2 H - d_1(1+r_1) + \phi H \mathbb{E}_2[z_3 + \Omega_3]}{\gamma \nu \left(1 + \frac{1}{\beta\delta}\right)} \right)^{\frac{\alpha}{1+\eta}} \tag{B.9}$$

A drop in expectations directly impacts output, as well as a fall in financial intermediaries' net worth $z_2 H - d_1(1+r_1)$. Hence, looking at $\mathbb{E}_2[z_3 + \Omega_3]$ inside a crisis is a sufficient statistics even in this extended model with real production. A liquidity drought spills over the real sector and propagates to employment and output.

[57] A more complete formulation of the collateral constraint could be:

$$d_2 \leq \phi H \mathbb{E}_2[z_3] + \psi f_2 \tag{B.6}$$

whereby assuming that a fraction of the amount lent to firms can be recovered by depositors in the (non-equilibrium) possibility of default. I am here analyzing the limiting case where $\psi \to 0$. The case where corporate loans are pledgeable complexifies matters without bringing any new intuition. The analysis can be found in a previous version, Fontanier (2022).

[58] Consumption is needed for the SDF to generate amplification and welfare losses: a risk-neutral valuation pricing kernel breaks the feedback loop between the price of the asset and marginal utility. But one can think of c_2 as dividends or compensation.

B.2 Welfare Analysis with Real Production

The planner maximizes:

$$\mathcal{W}_1 = \Phi^h \mathbb{E}_1^{SP}\big(c_1^h + \beta\left[c_2^h - \nu\frac{l_2^{1+\eta}}{1+\eta}\right] + \beta^2 c_3^h\big) + \Phi^b \mathbb{E}_1^{SP}\big(\ln(c_1) + \beta\ln(c_2) + \beta^2 c_3\big) \quad \text{(B.10)}$$

where Φ^h and Φ^b are the Pareto weights attached to each group by the planner. I denote by V_2^h and V_2^h the value functions of each group at time $t = 2$.

Leverage: We are interested in the derivatives of these value functions at time $t = 2$ with respect to the amount of short-term debt (or savings) chosen at time $t = 1$. Because funds allocated to firms (f_2) are chosen optimally without a constraint (see equation B.7), an infinitesimal change in f_2 does not have a first-order impact on the welfare of bankers:

$$\frac{dV_2^b}{dd_1} = \phi H(\lambda_2 - 1)\frac{d\Omega_3}{dq_2}\frac{dq_2}{dd_1} + \underbrace{\beta\frac{\delta}{f_2} - \lambda_2}_{=0}. \quad \text{(B.11)}$$

For households, however, there is a new term coming from the expansion of bank lending to firms in the real sector:

$$\frac{dV_2^h}{dd_1} = \phi H\underbrace{(\lambda_3^h - \lambda_2^h)}_{=0}\frac{d\Omega_3}{dq_2}\frac{dq_2}{dd_1} + \max\left(\underbrace{A\alpha\left(\frac{z_2 H - d_1(1+r_1) + \phi H\mathbb{E}[z_3 + \Omega_3]}{\gamma\nu\left(1+\frac{1}{\delta}\right)}\right)^{\frac{\alpha}{1+\eta}-1} - \nu}_{\to 0 \text{ when unconstrained}}, 0\right)\frac{dc_2}{dd_1}. \quad \text{(B.12)}$$

To understand why this second term is 0 when firms are unconstrained, notice that when firms are able to perfectly maximize profits they hire an amount of labor corresponding to: $\alpha A l_2^{\alpha-1} = w$, which itself implies, when combined with households first-order condition for labor/leisure: $\alpha A l^{\alpha-1-\eta} = \nu$. Similarly, the derivative dc_2/dd_1 is also 0 when financial intermediaries are unconstrained. To conclude, the planner's optimality condition for short-term debt is given by:

$$0 = \Phi^h \mathbb{E}_1^{SP} \left[\kappa_f \mathbb{1}_{\kappa_2 > 0} \left(\phi H \frac{d\Omega_3}{dq_2} \frac{dq_2}{dd_1} - (1 + r_1) \right) \right] +$$
$$\Phi^b \left\{ \mathbb{E}_1 [\lambda_2] - \mathbb{E}_1^{SP} [\lambda_2] + \mathbb{E}_1^{SP} [\phi H \kappa_2 \frac{d\Omega_3}{dq_2} \frac{\partial q_2}{\partial d_1}] \right\} \quad \text{(B.13)}$$

where $\kappa_f = \alpha A l_2^{\alpha-1} - \nu l_2^{\nu}$ plays the role of a "capacity wedge:" it measures how far firms are from their first-best production level (equivalent to the κ of intermediaries). When this wedge is positive (i.e. underemployment) a reduction in the leverage of financial intermediaries is beneficial for households, since it increases the production of real goods in a crisis. The $(1 + r_1)$ is a distributive externality: because of incomplete markets, funds are more valuable to firms than to intermediaries so reducing the leverage of intermediary is beneficial in a crisis. The first part of this expression is a collateral externality: relaxing the collateral constraint of intermediaries if beneficial to everyone when the economy in a crisis.

Current Prices: The reversal externality, similar to the collateral externality, also enters in production. The welfare effects of changing marginally equilibrium prices q_1 are given by:

$$\mathcal{W}_q = \Phi^h \mathbb{E}_1^{SP} \left[\kappa_f \mathbb{1}_{\kappa_2 > 0} \left(\beta \phi H \frac{\partial \Omega_3}{\partial q_1} \right) \right] + \Phi^b \left\{ \beta \mathbb{E}_1^{SP} \left[\kappa_2 \phi H \frac{\partial \Omega_3}{\partial q_1} \right] \right\} \quad \text{(B.14)}$$

Summary: The welfare analysis is very similar to the case without production studied in the main paper. In particular, the forces at play are exactly the same. Production simply reinforces the need for the planner to intervene in financial markets. Indeed, the worsening of pessimism during crises has repercussions on the level of employment and output, inflating the size of welfare losses. The important lesson of this extension is that the features of behavioral biases that matter for welfare are entirely identical to what was identified in the baseline welfare analysis.

C Alternative Collateral Constraint with Current Prices

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

$$d_2 \leq \phi H q_2. \tag{C.1}$$

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):

$$c_2 = z_2 H - d_1(1 + r_1) + \phi H q_2. \tag{C.2}$$

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:

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

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

Proposition 9 (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:

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

ii) For investment in collateral assets H:

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

iii) And for prices q_1:

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

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

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

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

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

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

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 $t=2$.[59]

Proposition 7 also still holds exactly in the same way, although the proof is more involved.[60]

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

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

[60] 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).

D Ω_3-Uncertainty

This section extends the insights of Section 5 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:

$$w_3 \sim \mathcal{U}\left[\bar{\Omega}_3 - \sigma_\Omega, \bar{\Omega}_3 + \sigma_\Omega\right] \tag{D.1}$$

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

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

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

$$L = \frac{1}{2\sigma_\Omega} \int_{-\sigma_\Omega}^{\sigma_\Omega} \frac{1}{n_2 + \phi H(\bar{z}_3 - \bar{\Omega}_3 - \omega_3)} d\omega_3 \tag{D.3}$$

$$\implies L = -\frac{1}{(2\sigma_\Omega)\phi H}\left[\ln(n_2 + \phi H(\bar{z}_3 - \bar{\Omega}_3 - \omega_3))\right]_{-\sigma_\Omega}^{\sigma_\Omega} \tag{D.4}$$

$$\implies L = \frac{1}{(2\sigma_\Omega)\phi H} \ln\left(\frac{n_2 + \phi H(\bar{z}_3 - \bar{\Omega}_3 + \sigma_\Omega))}{n_2 + \phi H(\bar{z}_3 - \bar{\Omega}_3 - \sigma_\Omega))}\right) \tag{D.5}$$

This is a functions of the type:

$$f(x) = \frac{1}{x} \ln\left(\frac{K + x}{K - x}\right) \tag{D.6}$$

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

$$f'(x) = \frac{(K^2 - x^2)\ln\left(\frac{K+x}{K-x}\right) + 2Kx}{x^2(K - x)(K + x)} \tag{D.7}$$

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

$$\frac{d}{dx}(K^2 - x^2)\ln\left(\frac{K+x}{K-x}\right) + 2Kx = 2x\ln\left(\frac{K+x}{K-x}\right) > 0 \tag{D.8}$$

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:

$$g(x) = \frac{1}{x}\ln\left(\frac{1}{K-x}\right) \tag{D.9}$$

Where the derivative is now:

$$g'(x) = \frac{\frac{x}{a-x} - \ln\left(\frac{1}{K-x}\right)}{x^2} \tag{D.10}$$

And the derivative of the numerator is:

$$\frac{d}{dx}\frac{x}{a-x} - \ln\left(\frac{1}{K-x}\right) = \frac{x}{(a-x)^2} > 0 \tag{D.11}$$

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.

E Bailouts

While the main analysis was made under the restriction of constrained efficiency, in reality financial crises are often addressed using direct liquidity injections. This section investigates how bailouts, and their anticipation by agents, interact with irrational exuberance and distress concerns. The first question that naturally arises is whether the presence of behavioral biases changes the optimal policy mix between ex-ante and ex-post interventions. I then explore whether irrationality mitigates or amplify moral hazard problems.

E.1 A Stylized Model of Bailouts

The social planner can now directly inject liquidity into the financial system, by providing loans to financial institutions. Concretely, it transfers an amount b from households to financial intermediaries at time $t = 2$, and financial intermediaries reimburse households at $t = 3$ at the prevailing market risk-free rate. I assume that this transfer entails a quadratic cost $g(b)$, representing distortions arising from taxation or political economy concerns:

$$g(b) = \frac{b^2}{2\xi} \tag{E.1}$$

Outside of a financial crisis, there is no point in providing liquidity to financial intermediaries. Inside a crisis, welfare at $t = 2$ becomes:[61]

$$\begin{aligned}\mathcal{W}_2 = \ln\left(z_2 H - d_1(1+r_1) + b + \phi H \mathbb{E}_2[z_3 + \Omega_3]\right) \\ + \beta\left(\mathbb{E}^{SP}[z_3]H - \phi H \mathbb{E}_2[z_3 + \Omega_3]/\beta - b/\beta\right) - g(b).\end{aligned} \tag{E.2}$$

This leads to the following expression for the optimal bailout size:

$$b^*(d_1, H, z_2, \Omega_3) = \xi\left(\frac{\partial \mathcal{W}_2}{dn_2}(d_1, H, z_2, \Omega_3) - 1\right) \tag{E.3}$$

where the partial derivative with respect to net worth can once again be expressed as:

$$\frac{\partial \mathcal{W}_2}{dn_2} = \kappa_2\left(1 + \phi H \frac{d\Omega_3}{dq_2}\frac{da_2}{dn_2}\right). \tag{E.4}$$

[61] The welfare of households is irrelevant since the loan is made at the market rate, hence households stay on their Euler equation. Alternatively, g could represent the welfare costs borne by households if the loan make them deviate from their optimality conditions.

Intuitively, a bailout is only desirable when $\kappa_2 > 0$: otherwise, there is no need to intervene to circumvent financial frictions that are not currently biting. The optimal bailout size is also increasing with κ_2: the more stringent frictions are, the more incentives to intervene and relax them. In particular, in the presence of excess pessimism $\Omega_3 < 0$, the financial crisis will be more severe and thus calling for stronger intervention. Furthermore, belief amplification also creates a new motive for ex-post intervention. By providing liquidity to distressed financial intermediaries, the social planner is indirectly supporting asset prices. This in turn can lessen pessimism and thus alleviate collateral constraints.

E.2 Optimal Policy Mix

Does the possibility of bailouts in the future change the financial authority's incentive to impose leverage restrictions? Jeanne and Korinek (2020) show, in a somewhat related setup, that macroprudential policy is still desirable and can resolve any time-consistency problems that may arise from the use of ex-post liquidity provision. In this section I confirm that their results are still valid in the presence of behavioral factors. In other words, I show that the possible existence of irrational exuberance is not an argument in favor of the ex-post "cleaning" paradigm.

Proposition 11 (Uninternalized Welfare Effects with Bailouts)**.** *Under the presence of bailouts, the decomposition developed in Section 3 holds. The the uninternalized welfare effects of a marginal increase in leverage can be expressed as:*

$$\mathcal{W}_d = \underbrace{\beta\left(\mathbb{E}_1[\lambda_2(b^*)] - \mathbb{E}_1^{SP}[\lambda_2(b^*)]\right)}_{\mathcal{B}_d} - \underbrace{\beta\mathbb{E}_1^{SP}\left[\kappa_2(b^*)\phi H \frac{\boldsymbol{d\Omega_3}}{\boldsymbol{dq_2}}\frac{dq_2}{dn_2}\right]}_{\mathcal{C}_d}. \tag{E.5}$$

Proof. At time $t = 2$, the welfare of financial intermediaries can now be written as:

$$\mathcal{W}_2 = \begin{cases} \beta \ln\left(n_2 + b^* + \phi H \mathbb{E}_2[z_3 + \Omega_3(q_2)]\right) + \beta^2 \left(\mathbb{E}[z_3]H - \phi H \mathbb{E}_2[z_3 + \Omega_3(q_2)]/\beta - b^*/\beta\right) & \text{if } z_2 \geq z^* \\ \beta\left(\beta \mathbb{E}[z_3]H + n_2\right) & \text{otherwise} \end{cases} \tag{E.6}$$

with the level of bailouts determined optimally in equilibrium. The private first-order condition on borrowing is unchanged since agents to not internalize their impact on b^* (atomistic agents):

$$\frac{\partial \mathcal{L}_{b,1}}{\partial d_1} = \lambda_1 - \mathbb{E}_1\left[\lambda_2(b^*)\right] \tag{E.7}$$

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 2 *and* the level of bailouts. This leads to the following first-order condition:

$$\frac{\partial \mathcal{L}^{SP}_{b,1}}{\partial d_1} = \lambda_1 - \mathbb{E}^{SP}_1\left[\lambda_2(b^*)\right] - \mathbb{E}^{SP}_1\left[\kappa_2 \phi H \frac{\partial \Omega_3}{\partial q_2}\frac{\partial q_2}{\partial n_2}\right] + \frac{db^*}{dn_2}\lambda_2 - \frac{db^*}{dn_2}g'(b^*) - \frac{db^*}{dn_2} \tag{E.8}$$

And the last part is equal to zero since bailouts are chosen optimally:

$$g'(b^*) = \lambda_2 - 1 \tag{E.9}$$

which proves Proposition 11. □

Equation E.5 makes the dependence of λ_2, marginal utility of financial intermediaries inside a financial crisis, on the level of bailouts explicit.[62] One can readily see that, even in the presence of bailouts, the collateral externality is still present and uninternalized, thus calling for leverage restrictions.[63] Therefore, and naturally as in Jeanne and Korinek

[62] One might seem surprising that the uninternalized welfare effect does not include a term $\partial b^*/\partial d_1$, that represents how increasing aggregate leverage changes the future size of bailouts. This is because bailouts are determined optimally in period $t = 2$, so the envelope theorem applies.

[63] This is assuming that bailouts are not effective enough to entirely prevent the occurrence of a financial crisis in the future. If bailouts are not costly at all, for example, the social planner will be able to provide enough liquidity in all states of the world such to achieve $\kappa_2 = 0$. Only under this extreme, and unrealistic case, are ex-ante restrictions undesirable.

(2020), bailouts still need to be accompanied by ex-ante leverage restrictions to compensate for uninternalized welfare effects. The next part explores how the size of the optimal intervention changes because of moral hazard, following Farhi and Tirole (2012).

E.3 Moral Hazard and Exuberance

The behavioral biases of agents furthermore interact with moral hazard concerns in a novel way.[64] This can be seen from inspecting the behavioral wedge of equation (E.5):

$$\mathcal{B}_{d,b^*} = \beta \mathbb{E}_1[\lambda_2(b^*)] - \beta \mathbb{E}_1^{SP}[\lambda_2(b^*)] \tag{E.10}$$

Marginal utility during a crisis depends on the level of bailouts b^*. But if agents recognize that bailouts will be determined optimally, according to equation (E.3), their expected bailout size state-by-state differs from the planner's. Indeed, b^* depends on the net worth of agents, but financial intermediaries believe that the asset will pay off $z_2 + \Omega_2$ instead of z_2 in each state. In other words, when agents are over-optimistic, they expect bailouts to be *smaller* than in reality, intuitively because they expect crises to be less severe than in reality. Hence, for a fixed $\Omega_2 > 0$, agents expect less aggressive bailouts than in reality: this directly reduces the behavioral wedge, which is the difference between expected marginal utilities between agents and the planner. Indeed, agents expect $\lambda_2(z_2 + \Omega_2, b^*(z_2 + \Omega_2, 0), 0)$, while the planner expected $\lambda_2(z_2, b^*(z_2, \Omega_3), \Omega_3)$. Moral hazard concerns are then attenuated by irrational optimism since $b^*(z_2 + \Omega_2, 0) < b^*(z_2, \Omega_3)$ and λ_2 is decreasing in b. This effect is further amplified by the fact that agents neglect the fact that the optimal bailout might be even larger since agents can be over-pessimistic in the future. This is summarized in the following proposition.

Proposition 12 (Moral hazard and Exogenous Biases)**.** *For a fixed $\Omega_2 > 0$ and fixed state-by-state $\Omega_3 < 0$, the behavioral wedge is negative and increasing in ξ.*

Proof. The behavioral wedge is given by:

[64] Dávila and Walther (2021) is the only work, to the best of my knowledge, that analyzes bailouts in an environment with distorted beliefs. They do not consider the moral hazard problems that arise from agents anticipating government intervention, neither do they study endogenous belief distortions, however.

$$\mathcal{B}_d = \mathbb{E}_1\left[\lambda_2(b^*)\right] - \mathbb{E}_1^{SP}\left[\lambda_2(b^*)\right] \tag{E.11}$$

We can simply compare the two marginal utilities state-by-state. Agents believe that:

$$\lambda_2(b^*) = \left((z_2+\Omega_2)H + b^*(z_2+\Omega_2, 0) - d_1(1+r_1) + \phi H\mathbb{E}[z_3+\zeta\Omega_3]\right)^{-1} \tag{E.12}$$

While the planner believes that:

$$\lambda_2^{SP}(b^*) = \left(z_2 H + b^*(z_2, \Omega_3) - d_1(1+r_1) + \phi H\mathbb{E}[z_3+\Omega_3]\right)^{-1} \tag{E.13}$$

Since bailouts are proportional to the severity of the crisis. Using equation (E.3) yields:

$$\frac{d\mathcal{B}_d}{d\xi} = -\mathbb{E}_1\left[\frac{db^*}{d\xi}\lambda_2^2(b^*)\right] + \mathbb{E}_1^{SP}\left[\frac{db^*}{d\xi}\lambda_2^2(b^*)\right] \tag{E.14}$$

$$\Longrightarrow \frac{d\mathcal{B}_d}{d\xi} = -\mathbb{E}_1\left[(\lambda_2(b^*)-1)\lambda_2^2(b^*)\right] + \mathbb{E}_1^{SP}\left[(\lambda_2(b^*)-1)\lambda_2^2(b^*)\right] \tag{E.15}$$

Which is positive since the λ_2 are always greater or equal to 1. □

Matters are however more complicated when behavioral biases are endogenous, for example when Ω_2 depends on $(q_1 - q_0)$. In this case the expectation of future bailouts also raises the attractiveness of creating financial assets: their price will be supported by government's action in the intermediate period, lowering the risk premium. This pushes up the initial price of collateral assets, thereby fueling irrational exuberance. This increase in Ω_2 leads financial intermediaries to augment their leverage. The initial equilibrium is then determined by multiple fixed-points between the value of bailouts, leverage, and sentiment, as can be seen from the following system:

$$u'(c_1) = -\mathbb{E}_1\left[\frac{\partial \mathcal{W}_2}{\partial d_1}\left(d_1, b^*(d_1, H, z_2+\Omega_2(\boldsymbol{q}_1 - q_0)), H, z_2+\Omega_2(\boldsymbol{q}_1-q_0)\right)\right] \tag{E.16}$$

$$\boldsymbol{q}_1 = \mathbb{E}_1\left[\frac{\partial \mathcal{W}_2}{\partial H}\left(d_1, b^*(d_1, H, z_2+\Omega_2(\boldsymbol{q}_1 - q_0)), H, z_2+\Omega_2(\boldsymbol{q}_1-q_0)\right)\right] \tag{E.17}$$

where bailouts b^* feed in the equilibrium price q_1 which then feeds into the Euler equation, and in turn changes the equilibrium value of bailouts, and so on. I illustrate how this interaction between bailouts and sentiment depends on the determinants of Ω_2. Figure 4 presents the optimal leverage restriction that the planner needs to impose (in percentage of the decentralized equilibrium short-term debt) to attain the second-best, with and without bailouts, for different levels of initial sentiment. The left panel presents the case where Ω_2 is set exogenously. There, when optimism increases this reduces the value of the behavioral wedge and thus diminishes the size of optimal leverage reductions. The left panel then looks at the case where $\Omega_2 = \alpha(q_1 - q_0)$, and varies α. This time, even though optimism still weakens moral hazard concerns, it is compensated by the feedback effect that functions through asset prices.[65]

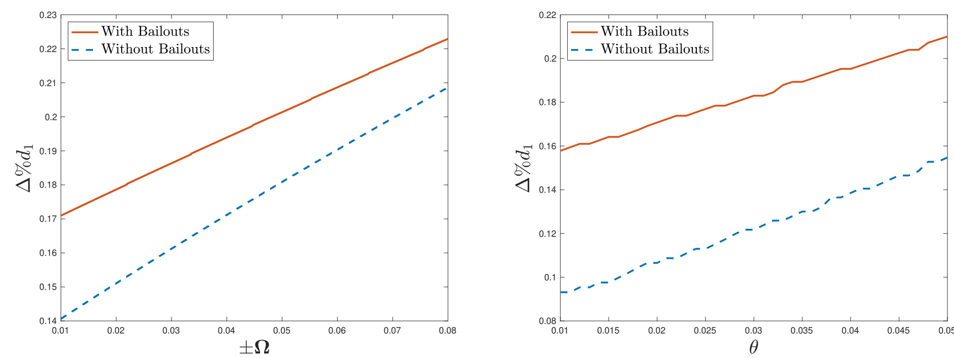

Figure 4: Supplementary Leverage Restrictions Required in the Exogenous Biases and and Endogenous Biases. In each panel the dotted lines plot the required decrease in leverage from the decentralized equilibrium to achieve the second-best in the absence of bailouts. Solid lines perform the same exercise but in the presence of bailouts in period $t = 2$. The behavioral bias in the left panel is of the fundamental extrapolation form, defined as $\Omega_{t+1} = \alpha(z_t - z_{t-1})$. z_0 and z_1 are chosen such that $\Omega_2 > 0$ to feature initial exuberance. The behavioral bias in the right panel is of the price extrapolation form, defined as $\Omega_{t+1} = \alpha(q_t - q_{t-1})$. q_0 is chosen such that $q_0 < q_1$ to feature initial exuberance.

[65]A corollary of this analysis is that announcing that bailouts will happen in case of a crisis must be done as early as possible if beliefs depend on price changes. Announcing bailouts at the last moment creates additional optimism in this case, right when the financial system is the most vulnerable.

F Various Psychological Models and Ω-Correspondence

F.1 Overconfidence

Overconfidence has been widely used to explain large trading volume, by generating substantial disagreement between investors (Odean 1998). Because this paper is about *aggregate* over-optimism or over-pessimism, I will focus in this section on the features of overconfidence that can generate momentum and reversals.

Financial institutions have a prior over the distribution of dividends in period $t = 2$:

$$z_2 \sim \mathcal{N}(\mu_0, \sigma_0^2) \tag{F.1}$$

and receive a signal $s = z_2 + \epsilon$ with:

$$\epsilon \sim \mathcal{N}(0, \sigma_s^2). \tag{F.2}$$

Overconfident financial intermediaries have a posterior of:

$$z_2 \sim \mathcal{N}\left(\mu_0 + \frac{\sigma_0^2}{\sigma_0^2 + \tilde{\sigma}_s^2}(s - \mu_0), \frac{\sigma_0^2}{1 + \frac{\sigma_0^2}{\tilde{\sigma}_s^2}}\right) \tag{F.3}$$

where $\tilde{\sigma}_s^2 < \sigma_s^2$, which means that overconfident agents believe that the signal has a higher precision than in reality. This directly implies that the bias, relative to the social planner valuation, is given by:

$$\Omega_2 = \frac{\sigma_s^2 - \tilde{\sigma}_s^2}{(\sigma_0^2 + \tilde{\sigma}_s^2)(\sigma_0^2 + \sigma_s^2)}\sigma_0(s - \mu_0) \tag{F.4}$$

so that agents become exuberant after positive news ($s > \mu_0$): $\Omega_2 > 0$.

Notice how the variance of the two distributions are different with overconfidence. As such, the results in Proposition 2 (and its equivalent for investment in Appendix A.4) are not directly applicable. But a higher $\tilde{\sigma}_s^2$ means that agents are using a narrower distribution than the social planner. This is reminiscent of the results presented in Section 5:

intuitively, this will create an even larger gap between the two solutions since agents will neglect left-tail and right-tail events. As shown in Proposition 7, this is calling for tighter macroprudential regulation ex-ante.

F.2 Sticky Beliefs

While this paper is mostly concerned with investors that adjust their views too much in response to information, there is also widespread evidence of investors adjusting their beliefs *too little*. A recent example is the work of Bouchaud, Krueger, Landier, and Thesmar (2019), where investors form expectations according to:

$$\tilde{\mathbb{E}}_1[z_2] = (1-\lambda)\mathbb{E}_1^r[z_2] + \lambda\tilde{\mathbb{E}}_0[z_2] \tag{F.5}$$

where $\mathbb{E}_1^r$ is the rational time 1 expectations about the future dividend. When $\lambda = 0$, expectations are fully rational. When $\lambda > 0$, expectations depend on past expectations. In terms of the notation of my paper, the bias can be expressed as:

$$\tilde{\mathbb{E}}_1[z_2] = \mathbb{E}_1^{SP}[z_2] + \lambda\left(\tilde{\mathbb{E}}_0[z_2] - \mathbb{E}_1^r[z_2]\right) \implies \Omega_2 = \lambda\left(\tilde{\mathbb{E}}_0[z_2] - \mathbb{E}_1^r[z_2]\right). \tag{F.6}$$

Agents are thus over-optimistic in period t when the objective expected dividend is less than the expectation agents held in period $t-1$. Expanding this expression recursively yields:

$$\Omega_2 = \lambda\left(\mathbb{E}_0^r[z_2] - \mathbb{E}_1^r[z_2]\right) + \lambda\Omega_1. \tag{F.7}$$

F.3 Internal Rationality

Adam, Marcet, and Beutel (2017b) present a model where agents are not "externally rational:" they do not know the true stochastic process for payoff relevant variables beyond their control, i.e. prices in my setup. Here I adapt their idea to my setup with some simplifying assumptions, and show in which circumstances the results change.

Agents are rational regarding the distribution of z_t, but they believe prices evolve according to:

$$q_{t+1} = q_t + \beta_{t+1} + \epsilon_{t+1} \tag{F.8}$$

with ϵ_{t+1} is a transitory shock and β_{t+1} is a persistent component evolving as:

$$\beta_{t+1} = \beta_t + \nu_{t+1}. \tag{F.9}$$

Furthermore, all innovations are jointly normal. Adam et al. (2017b) show that under some conditions, and when agents are using a steady-state precision, the filtering problem boils down to expectations evolving as:

$$\tilde{E}_t[q_{t+1}] = (1+g)(q_t - q_{t-1}) + (1-g)\tilde{E}_{t-1}[q_t] \tag{F.10}$$

where g is the equivalent of a Kalman gain, function of the variances of the noise terms. To make progress, I further assume that agents place a low conditional variance on this estimate, such that I can study the limiting case where this point estimate is believed to be certain (i.e. there is no risk for the price next period in agents' mind). I denote by $\tilde{q}_2$ this point estimate, such that agents' optimization yields:

$$q_1 = \beta \mathbb{E}_1 \left[\frac{\lambda_2}{\lambda_1} (z_2 + \tilde{q}_2) \right]. \tag{F.11}$$

Equation (F.11) can be rewritten using the correct price used by the planner q_2:

$$q_1 = \beta \mathbb{E}_1 \left[\frac{\lambda_2}{\lambda_1} (z_2 + q_2 + (\tilde{q}_2 - q_2)) \right]. \tag{F.12}$$

so an equivalent to the Ω_2 used throughout this paper is $\Omega_2^q = \tilde{q}_2 - q_2$: a bias on expected prices that is positive (exuberance) when the forecasted value if above the realized value, and vice-versa.

How does this impact the welfare analysis? It crucially depends on the form of the collateral constraint. If we stay in the benchmark case where the collateral constraint takes the form $d_2 \leq \phi H \mathbb{E}_2[z_3]$, then it is clear that since agents are correct about the distribution of fundamentals, they will make no mistake regarding their future net worth or the future borrowing capacity of the economy. Consequently, the only margin that is distorted if the investment margin: agents are too optimistic (pessimistic) regarding the payoffs of their investment, since they are too optimistic (pessimistic) regarding the resale value of the asset they are creating. Thus, only the behavioral wedge for investment is non-zero in this case.

Importantly, there are no externalities anymore. Indeed, decisions during the boom will impact time expectations of prices made at $t = 2$ but these expectations will not affect the tightness of collateral constraints.

This discussion makes clear that for externalities to survive in this case, it is necessary to have a collateral constraint that depends on prices (either current prices, or expected prices), whereas biases on fundamentals impact welfare in a "robust" way. Only then can biases impact its tightness: when agents in a crisis are over-pessimistic regarding future prices, that impacts the equilibrium value of q_2. In this case externalities survive. But notice that the sign of the key derivative for the reversal externality is clearly ambiguous:

$$\frac{d\Omega_3^q}{dq_1} = \frac{d\tilde{q}_3}{dq_1} = (1-g)\left(\frac{d\tilde{q}_2}{dq_1} - 1\right). \tag{F.13}$$

This is because sentiment is "sticky" with learning. If by reducing asset prices at $t = 1$, the planner makes future agents more pessimistic in a financial crisis, that hurts welfare.

G Multiple Equilibria

The analysis in the main paper as made under the assumption that the equilibrium was unique at $t = 2$ (see footnote 30). 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:

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

$$c_2 = z_2 H - d_1(1 + r_1) + \phi H \mathbb{E}_2[z_3 + \Omega_3(q_2)] \tag{G.2}$$

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:

$$\begin{aligned} q_2 = {} & \beta \left(n_2 + \phi H \mathbb{E}_2[z_3 + \Omega_3(q_2)]\right) \mathbb{E}_2[z_3 + \Omega_3(q_2)] \\ & + \phi(1 - (n_2 + \phi H \mathbb{E}_2[z_3 + \Omega_3(q_2)]))\mathbb{E}_2[z_3 + \Omega_3(q_2)] \end{aligned} \tag{G.3}$$

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

$$\Omega_3(q_2) = \alpha q_2 + \chi \tag{G.4}$$

The price condition is now:

$$\begin{aligned} q_2 = {} & \beta \left(n_2 + \phi H \mathbb{E}_2[z_3 + \alpha q_2 + \chi]\right) \mathbb{E}_2[z_3 + \alpha q_2 + \chi] \\ & + \phi(1 - (n_2 + \phi H \mathbb{E}_2[z_3 + \alpha q_2 + \chi]))\mathbb{E}_2[z_3 + \alpha q_2 + \chi] \end{aligned} \tag{G.5}$$

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. Figure 5 illustrates this. We can thus consider the case of unique equilibrium when sentiment is linear in prices. How more complicated forms of biases interact with frictions to create multiple equilibria is left for future work.

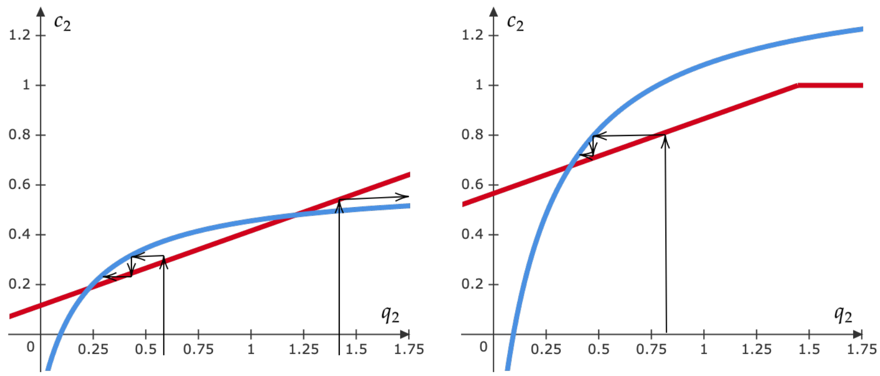

Figure 5: Graphical Illustration of Equilibrium Determination at $t = 2$ with sentiment linear in asset prices. The red line represents the budget constraint, and the blue line represents the pricing condition. The black arrows represent a tâtonnement process that starts at a given price. This price yields a certain level of sentiment and thus of consumption, which then gives rise to a different price, and so on. The left panel shows that if two equilibria are possible, only one is stable.

H General Intertemporal Elasticity of Substitution

The paper made two assumptions on the utility function form of financial intermediaries: (i) log-utility in the first two periods, and (ii) linear utility in the last period. These assumptions were made for tractability, and to avoid over-complicating expressions without bringing any new intuition. In this section, I show that a model with a more general intertemporal elasticity of substitution (henceforth IES) delivers the exact same insights.

The utility function of financial intermediaries is now given by:

$$U^b = \mathbb{E}_1 \left[\frac{c_1^{1-\sigma}}{1-\sigma} + \beta \frac{c_2^{1-\sigma}}{1-\sigma} + \beta^2 \frac{c_3^{1-\sigma}}{1-\sigma} \right] \tag{H.1}$$

where σ is the inverse of the IES. The equilibrium is now characterized by the Lagrange multiplier on the collateral constraint, κ, expressed as $\kappa = \lambda_2 - \mathbb{E}_2[\lambda_3]$, where the marginal utility is now given by $\lambda_t = c_t^{-\sigma}$. The pricing equation at $t = 2$ is thus now slightly more complicated than before:

$$q_2 = \beta \mathbb{E}_2 \left[\frac{\lambda_3}{\lambda_2} (z_3 + \Omega_3) \right] + \phi \left(1 - \mathbb{E}_2 \left[\frac{\lambda_3}{\lambda_2} \right] \right) \mathbb{E}_2 \left[(z_3 + \Omega_3) \right] \tag{H.2}$$

However, it should be clear by now that the uninternalized welfare effects take exactly the same form I presented in Proposition 1. Why? The welfare of intermediaries at time $t = 2$ during a crisis can be written as:

$$\mathcal{W}_2 = \beta u \left(n_2 + \phi H \mathbb{E}_2 [z_3 + \Omega_3(q_2, q_1)] \right) + \beta^2 u \left(\mathbb{E}_2 [z_3] H - \phi H \mathbb{E}_2 [z_3 + \Omega_3(q_2, q_1)] / \beta \right) \tag{H.3}$$

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

$$\mathcal{L}_{b,1} = \left[u(c_1) + \mathbb{E}_1 [\mathcal{W}_2(n_2, H; q_2, z_2)] \right] - \lambda_1 \left[c_1 + c(H) - d_1 - e_1 \right] \tag{H.4}$$

The first-order condition on borrowing then still gives $\partial \mathcal{L}_{b,1} / \partial d_1 = \lambda_1 - \mathbb{E}_1 \left[\lambda_2 \right]$ 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:

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

where the only difference is now that $\kappa_2 = \lambda_2 - \mathbb{E}_2[\lambda_3]$ instead of $\lambda_2 - 1$. Obviously, the same algebra ensures that Proposition 1 is in the same way still valid.

It is less obvious to sign the derivative $\partial q_2 / \partial n_2$ in this general case. But inside a financial crisis, this sensitivity is unambiguously positive. Indeed, we have:

$$\frac{d c_2}{d n_2} = 1 + \phi H \frac{d \Omega_3}{d q_2} \frac{d q_2}{d n_2} \implies d\lambda_2 = -\sigma \left(1 + \phi H \frac{d \Omega_3}{d q_2} d q_2 \right) \lambda_2^{\frac{\sigma+1}{\sigma}} \tag{H.6}$$

and

$$\frac{dc_3}{dn_2} = -\phi H \frac{d\Omega_3}{dq_2}\frac{dq_2}{dn_2}(1+r_2) \implies d\lambda_3 = \sigma\left(\phi H \frac{d\Omega_3}{dq_2} dq_2(1+r_2)\right)\lambda_3^{\frac{\sigma+1}{\sigma}} \tag{H.7}$$

which implies that $d\lambda_3$ is of the sign as dq_2. In other words, the IES value (whether it is above or below 1) is irrelevant inside a crisis, because the amount of borrowing is fixed by the collateral constraint. In the case of an exogenous behavioral bias, the price sensitivity can be written:

$$dq_2 = \beta\sigma dc_2 c_2^{\sigma-1}\Big((\beta-\phi)\mathbb{E}_2[\lambda_3(z_3+\Omega_3)] + \phi Cov(\lambda_3, z_3)\Big). \tag{H.8}$$

dc_2 is obviously positive when the change is in net worth. Because of Assumption 1, the first term in the parentheses is positive. The second term, however, is negative.[66] While we could entertain the possibility that the covariance is strongly negative, this is not robust to changes in the micro-foundations of the collateral constraint.[67] I thus only study the natural case where $dq_2/dn_2 > 0$.[68]

This calculation was made with a fixed Ω_3, but is still valid with an endogenous bias. Indeed, movements in Ω_3 only amplify this price sensitivity:

$$\begin{aligned} dq_2 = \beta\sigma dc_2 c_2^{\sigma-1}\Big((\beta-\phi)\mathbb{E}_2[\lambda_3(z_3+\Omega_3)] + \phi Cov(\lambda_3, z_3)\Big) \\ + \Big(\mathbb{E}_2[\frac{d\lambda_3}{\lambda_2}(z_3+\Omega_3)] - \phi\mathbb{E}_2[\frac{d\lambda_3}{\lambda_2}[\mathbb{E}_2[(z_3+\Omega_3)]]\Big) \\ + d\Omega_3\Big(\beta\mathbb{E}_2[\frac{\lambda_3}{\lambda_2}] + \phi(1-\mathbb{E}_2[\frac{\lambda_3}{\lambda_2}])\Big) \end{aligned} \tag{H.10}$$

where dc_2 also incorporates how the price in q_2 impact Ω_3 and thus the borrowing capacity. There is also a term (the second line) expressing how a change in sentiment brought

[66] This could not happen in the linear utility at time $t = 3$, since then λ_3 was a constant.

[67] Indeed, if we assume that agents can default after observing the realization in z_3, the collateral constraint becomes of the form $d_2 \leq \phi H \min z_3$ and in this case the price sensitivity is:

$$dq_2 = \beta\sigma dc_2 c_2^{\sigma-1}\Big(\beta\mathbb{E}_2[\lambda_3(z_3+\Omega_3)] - \phi\mathbb{E}_2[\lambda_3](\min z_3 + \Omega_3)\Big) \tag{H.9}$$

which is unambiguously positive with Assumption 1.

[68] Dávila and Korinek (2018) also assume that the price of capital assets is increasing in the net worth of the financial sector.

by a change in asset prices impact future marginal utility, λ_3. Under the same condition as before, this term is also positive (similarly, it is only needed that ϕ is small enough, and this condition disappears under the alternative collateral formulation involving the minimum payoff).

Using the same welfare function, the general formulation in the second part of Proposition 1 is also still valid:

$$\mathcal{W}_H = \left(\beta \mathbb{E}_1^{SP}\left[\lambda_2(z_2+q_2)\right] - \lambda_1 q_1\right) + \beta \mathbb{E}_1^{SP}\left[\kappa_2 \phi H \frac{d\Omega_3}{dq_2}\left(\frac{dq_2}{dn_2} z_2 + \frac{dq_2}{dH}\right)\right] \tag{H.11}$$

Here again, however, the sign of dq_2/dH is harder to determine without the linearity of utility in the ultimate period, since movements in H have effects on the future marginal utility. A first thing to notice is that even if for some levels of IES, dq_2/dH becomes negative, that is still unlikely to overturn the result that the collateral externality pushes towards under-investment. Indeed, as I just showed the first term of the collateral externality dq_2/dn_2 is positive. So dq_2/dH needs to be strongly negative to compensate for this effect. In other words, the linearity of utility at $t = 3$ or the log-utility at $t = 2$ are not directly responsible for this result: it is the assumption that $z_2 > 0$ (see Dávila and Korinek 2018 for examples where over-investment arises because dividends are negative in bad states of the world).

But in general, for the same reason dq_2/dn_2 is positive, this derivative will also be positive. Intuitively, dq_2/dH measures how an expansion of the borrowing capacity of financial intermediaries impact the equilibrium asset price. If dq_2/dn_2 is positive, we should expect the same thing for dq_2/dH: an increase in the borrowing capacity is similar to an increase in net worth during a financial crisis. Indeed, consider the case with an exogenous bias for intuition first (remember that these derivatives are keeping the net worth constant):

$$\frac{d\lambda_2}{dH} = -\sigma\phi \mathbb{E}_2[z_3+\Omega_3]\lambda_2^{\frac{\sigma+1}{\sigma}} > 0 \tag{H.12}$$

$$\frac{d\lambda_3}{dH} = \sigma\phi \mathbb{E}_2[z_3+\Omega_3](1+r_2)\lambda_3^{\frac{\sigma+1}{\sigma}} < 0 \tag{H.13}$$

so that it is clear that the stochastic discount factor (λ_3/λ_2) is increasing in H. The price sensitivity can be expressed as always as:

$$dq_2 = \beta \mathbb{E}_2 \left[d\frac{\lambda_3}{\lambda_2}(z_3 + \Omega_3) \right] - \phi \mathbb{E}_2 \left[d\frac{\lambda_3}{\lambda_2} \right] \mathbb{E}_2 \left[(z_3 + \Omega_3) \right] \tag{H.14}$$

which again will be negative only in the case where the covariance is strongly negative:

$$dq_2 = (\beta - \phi) \mathbb{E}_2 \left[d\frac{\lambda_3}{\lambda_2}(z_3 + \Omega_3) \right] + \phi Cov(d\frac{\lambda_3}{\lambda_2}, z_3) \tag{H.15}$$

And, once again, this is not robust to alternative collateral constraints like $d_2 \leq \phi H \min[z_3 + \Omega_3]$. Lastly, this goes through with endogenous sentiment (as previously for net worth):

$$dq_2 = (\beta - \phi) \mathbb{E}_2 \left[d\frac{\lambda_3}{\lambda_2}(z_3 + \Omega_3) \right] + \phi Cov(d\frac{\lambda_3}{\lambda_2}, z_3) + d\Omega_3 \left((\beta - \phi) \mathbb{E}_2 \left[\frac{\lambda_3}{\lambda_2} \right] + \phi \right) \tag{H.16}$$

where $d\frac{\lambda_3}{\lambda_2}$ now also incorporates how changes in sentiment affect the SDF. Using $d\Omega_3 = \frac{d\Omega_3}{dq_2} dq_2$, we see that the sign of dq_2 is unchanged, movements in sentiment are simply amplifying the previous sensitivity. To conclude, the model with a general CRRA utility function across all three periods deliver the same uninternalized welfare effects as in the baseline case. This generality comes at the cost of greater complexity, without bringing anymore intuition.