Unregulated Finance and Optimal Regulation*

Christopher Clayton[†] Andreas Schaab[‡]

February 2022

Abstract

We study financial regulation when some financial institutions or capital flows are unregulated and there are pecuniary externalities. Optimal financial regulation is scaled by a "regulatory arbitrage multiplier." The contribution of an unregulated actor to regulatory arbitrage can be determined using microeconomic price elasticities and aggregate financial flows. Regulatory arbitrage can motivate tighter, rather than looser, optimal regulation due to collateral price effects. We provide a classification scheme for unregulated finance based on the welfare gains from extending regulation to certain institutions or activities. The institutions and markets driving regulatory arbitrage are not necessarily the most valuable targets for regulation. We apply our theory to shadow bank institution regulation and to capital flow management in a small open economy.

JEL codes: F38, G28, D62

Key words: Unregulated finance, capital flows, regulatory arbitrage, macroprudential regulation, capital controls, pecuniary externalities, fire sales

*We are grateful to Antonio Coppola, Eduardo Dávila, Gary Gorton, Sam Hanson, Matteo Maggiori, Aakriti Mathur, Alp Simsek, Jean Tirole, Aleh Tsyvinski, and seminar participants at Yale SOM, AFFI, and 2021 SED for helpful comments and discussions. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement n°669217 - ERC MARKLIM). This paper was previously circulated under the title "Shadow Banks and Optimal Regulation."

[†]Clayton: Yale School of Management. Email: christopher.clayton@yale.edu

[‡]Schaab: Columbia Business School. Email: ajs2428@columbia.edu. Website: `https://andreasschaab.com`

First Circulated: December 2020

1 Introduction

In the wake of the 2008 financial crisis, a new regulatory regime for financial stability has emerged. A distinctive feature of this new regime is that bank holding companies face tighter regulatory requirements, and capital control measures are increasingly employed as part of the regulatory toolkit.[1] At the same time, many financial institutions that conduct similar activities remain unregulated, raising questions about the efficiency and efficacy of current regulatory policies.[2] As conventional banks curb regulated activities, unregulated institutions could step in, thus diminishing the intended effect of regulation in the first place. However, currently unregulated institutions differ from conventional banks not only in their regulatory status but also in many other fundamental characteristics and activities. It is therefore not obvious what implications their presence has for bank regulation. An active debate has ensued about whether and how to start regulating the unregulated financial sector.

Our paper develops a new framework to study financial regulation when there are pecuniary externalities that warrant regulatory intervention, but some financial insitutions or agents, whom we term "financial actors," are not subject to financial regulation. Our framework allows for rich flexibility in the types of and relationships between different financial and nonfinancial actors in the economy. Using this framework, we make two main contributions. First, we show that the impact of unregulated finance on optimal regulation is captured by the *regulatory arbitrage multiplier,* which scales optimal financial taxes. We characterize unregulated actors' contribution to this multiplier in terms of estimable sufficient statistics, and show that regulatory arbitrage may amplify or dampen optimal regulation. Our second contribution is to develop a classification scheme for unregulated finance. Our classification scheme identifies the most valuable targets for new regulation based on the welfare gains a planner can attain by extending regulation. We show that an actor's contribution to regulatory arbitrage is the key determinant of the welfare gains from new regulation. However, we show that the appropriate definition of regulatory arbitrage differs subtly between our two main policy questions: collateral price effects on demand are important drivers of regulatory arbitrage to determine optimal financial taxes, but do not directly govern the potential welfare gains from extending regulation to the unregulated financial sector.

In our model, heterogeneous financial actors—such as banks, mutual funds, and international portfolio investors—interact on a set of markets with one another and with nonfinancial actors—such as nonfinancial firms and households. Equilibrium market prices appear in the constraint sets of

[1] In the domestic U.S. context, the Dodd-Frank Wall Street Reform and Consumer Protection Act of 2010 strengthened regulation of bank holding companies and established an orderly resolution regime. At an international level, the Basel III accords strengthened international regulatory standards. In the capital controls context, see IMF (2012).

[2] See Financial Stability Board (2013) for a policy perspective, and Financial Stability Board (2020) for an annual empirical overview of the shadow banking system.

these economic actors, giving rise to pecuniary externalities that warrant regulation. For example, occasionally binding collateral constraints that depend on the market value of assets may lead to a classical fire sale externality.[3] Our model remains agnostic to the fundamental differences between regulated and unregulated financial actors, regarding not only their fundamental characteristics but also the extent to which they are affected by externalites. Our framework allows us to determine the implications of regulatory arbitrage by unregulated actors for optimal regulation. Our framework also offers guidance to policymakers on how to best extend regulation to previously unregulated financial actors. The fact that our model remains agnostic to the differences between regulated and unregulated finance allows us to identify the set of sufficient statistics relevant for each of these policy questions.

We begin by characterizing financial regulation when all financial actors can be regulated. Financial regulation in our model takes the form of identity-based (i.e., actor-specific) revenue-neutral taxes on the transactions of financial actors in each market. For example, actors that issue debt on a bond market could be subject to a tax on debt issuance. We show that the product of two terms governs optimal financial regulation: the equilibrium price impact of a change in aggregate financial sector demand, and the pecuniary externalities generated by price changes in each market. Optimal regulation in our model features equal treatment in the sense that the same taxes are applied to all regulated financial actors. This happens because all financial actors generate the same externality by changing demand in a market.[4]

Our first main result is that optimal financial taxes in the presence of unregulated financial actors are scaled by a *regulatory arbitrage multiplier*. The regulatory arbitrage multiplier reflects how the effectiveness of regulation changes in the presence of equilibrium arbitrage by unregulated actors. As the planner adjusts financial taxes on a regulated financial actor, that actor's demand changes, leading to changes in market prices. Unregulated financial actors respond to changing market prices by adjusting their demand, which results in regulatory arbitrage. A feedback loop emerges: Regulatory arbitrage induces further equilibrium prices changes that in turn prompt more regulatory arbitrage. The regulatory arbitrage multiplier is the fixed point of this process, representing the total arbitrage response by unregulated actors that arises in equilibrium.

Our agnostic framework allows us to identify the relevant and empirically estimable sufficient statistics that determine an actor's contribution to regulatory arbitrage. These comprise a set of microeconomic price elasticities and flow volumes. Unregulated financial actors with larger micro elasticities or larger flow volumes contribute more to the regulatory arbitrage process, and so have a larger impact on optimal regulation.

[3] As in Kiyotaki and Moore (1997) for example.

[4] Equal treatment is recognized as an important objective in the policy debate, for example ECB (2018) and IMF (2012).

Intuitively, one might expect that regulatory arbitrage would undermine the effectiveness of financial regulation due to the classical substitution effect: As a planner tightens regulation, unregulated actors respond to the fall in prices with increased demand, making regulation less effective. A surprising result of our framework is that regulatory arbitrage can actually increase the effectiveness of financial regulation and lead to higher taxes. In models with pecuniary externalities, a counterveiling collateral price effect on unregulated demand emerges that counteracts the classical substitution effect. The intuition is that as the fire sale price of an asset rises, demand for the asset by unregulated financial actors actually increases because (collateral) constraints are relaxed. This offsets the classical substitution effect. Indeed, we show in a simple model that when unregulated actors are forced sellers due to fixed liquidity needs (Holmström and Tirole 1998), their demand for retaining assets increases in the price. Higher prices allow them to sell less to meet their liquidity need. This makes financial regulation even more effective and implies larger optimal taxes. If, on the other hand, unregulated actors are asset purchasers (i.e., arbitrageurs) with fixed liquidity surpluses (Allen and Gale 1994), higher prices reduce their loss absorbing capacity. This makes financial regulation less effective and leads to smaller optimal taxes.

The flexibility of our framework allows us to study not only the regulation of certain financial actors and institutions, which we classify as *identity-based* regulation, but also regulation of specific markets, which we classify as *activity-based* regulation. Identity-based regulation involves imposing a full set of taxes on a specific institution, whereas activity-based regulation involves placing a uniform tax across actors for trades in a specific market.[5] In contrast to identity-based regulation, activity regulation tends to be more effective for markets associated with high degrees of regulatory arbitrage. The intuition is that the regulator is able to calibrate a larger tax that also corrects for regulatory arbitrage.

The post-crisis regulatory framework has not been extended to unregulated finance more broadly in part because of the complexity of the unregulated financial sector. There are many different types of unregulated financial actors—such as mutual funds, insurance companies, hedge funds, and international portfolio investors—with differing business models. It is therefore not simple to determine whether and how to extend financial regulation to the unregulated financial system. Prominent policy proposals to extend financial regulation have advocated both regulating specific insitutions—such as targeted regulation of mutual funds—and regulating specific activities—such as a uniform tax on leverage.[6]

Our second main contribution is to propose a regulatory classification scheme for unregulated finance that can be used to evaluate targets for identity- and activity-based regulation. Our classifi-

[5] For example, proposals for activity-based regulation have included taxes on leverage (Feldman and Heincecke 2018).

[6] See for example Gorton et al. (2010) and Feldman and Heincecke (2018).

cation is tied to estimable sufficient statistics. Formally, we evaluate the first-order welfare gains achievable from extending identity- and activity-based regulation to previously unregulated actors or activities. We use these welfare gain metrics to classify the attributes of actors and activities that are the most valuable targets for regulation. The welfare-based approach to classification is desirable because it directly identifies and ranks actors and activities based on the welfare gains that can be attained by extending regulation. A key strength of our framework is that we do not take a stance on the structure of unregulated finance. Our sufficient statistics approach identifies ex post which types of institutions and markets should be classified as unregulated finance from the perspective of a regulator designing new regulation.

Intuitively, one might expect that the actors and activities contributing most to regulatory arbitrage are also the most valuable targets for regulation. We show that this intuition is correct under two sufficient conditions: (i) there are no wealth effects; and, (ii) the proposed regulation taxes markets whose prices do not appear in the constraints. These assumptions are economically plausible when regulation is imposed during times when actors have high net worth (i.e., wealth effects are small) and constraints are not binding (i.e., prices do not appear in constraints). However, the simple intuition breaks down more generally because regulatory arbitrage in response to price changes involves both income effects and collateral price effects. In contrast, revenue-neutral financial taxes do not generate income effects or change collateral prices. This means that the relevant definition of regulatory arbitrage from the classification perspective is the classical substitution effect. This highlights an important difference between how to think about regulatory arbitrage from the perspective of designing optimal regulation versus for extension of new regulation. The former requires thinking about the entire (Marshallian) effects of price changes on demand, whereas the latter requires thinking about the classical (Hicksian) substitution effect. The above-mentioned model of liquidity needs and surpluses represents an illustrative limiting case where the classical substitution effect is zero: Even though unregulated finance matters in this model for optimal financial taxes due to collateral price effects, the planner cannot increase welfare by extending regulation to unregulated actors.

In general, regulatory arbitrage occurs across many markets simultaneously. It is therefore not obvious how to rank the regulatory arbitrage responses of different actors or activities. An important contribution of our classification scheme is to show that the vector of optimal financial taxes is also the vector of weights used to compare regulatory arbitrage responses across different markets. Intuitively, markets with large optimal taxes are associated with large pecuniary externalities, so that regulatory arbitrage in these markets is assigned a particularly large weight in our classification scheme.

Our framework allows us to compare the relative benefits of new identity- and activity-based regulation. One key insight is that identity-based regulation targets regulatory arbitrage across markets for a given institution, whereas activity-based regulation targets regulatory arbitrage across institutions for a given market. Thus activity-based regulation is particularly important when a market is central to many unregulated financial actors, whereas identity-based regulation is particularly important when an institution is central in many markets. A surprising insight is that activity regulation allows for implicit discrimination against institutions that rely particularly on that market, an idea we develop further in our applications.

Finally, in Section 5 we apply our theory to two leading applications. Our first application studies the regulation of shadow banking institutions. We consider a simple model, in which shadow banks issue debt to finance initial investment, but then face a binding rollover constraint during the crisis that forces them to fire sell assets. We show that the regulatory classification of shadow banking institutions depends on their ex-ante illiquid investment elasticities and on their total illiquid investment. In the case of Cobb-Douglas productivity, we show more concretely that shadow banks with high levels of illiquid investment and high illiquid investment factor shares are the most valuable targets for new regulation.

Our second application is to capital flow regulation by a small open economy (SOE). The small open economy faces inflows and outflows by international investors, who may be flighty or may value capital retrenchment during crises. We show that greater investor flight dampens the efficacy of initial investment (inflow) taxes because flighty investors arbitrage the lower investment price and then generate costly outflows. Similarly, greater investor retrenchment can amplify or dampen the efficacy of ex post (outflow) taxes when capital inflows are valuable but outflows are costly. This is because the outflow tax boosts the liquidation price, which simultaneously encourages valuable inflows by investors who value retrenchment but also encourages outflows. Interestingly, we show that outflow taxes tend to be more valuable than inflow taxes, since outflow taxes implicitly discriminate against flight and retrenching investors.[7]

Related literature. The paper closest to ours is Dávila and Walther (2021) who, in independent and contemporaneous work, develop an elegant theory of second-best corrective policy with imperfect instruments. The two papers share a similar focus on regulatory arbitrage ("leakage" in their language) but are otherwise complementary in their approach and focus. They study the role of leakage elasticities—capturing whether one agent or activity is a gross complement or substitute for another agent or activity—in determining second-best regulation and the value of relaxing constraints on regulatory instruments. Our paper focuses on the interaction between regulatory arbitrage and pecuniary externalities. We introduce the regulatory arbitrage multiplier and develop a sufficient statistics representation in terms of micro price elasticities and flow volumes. We highlight

[7]This mechanism is similar to the idea that financial transaction taxes discriminate against speculative short term traders (e.g., Davila 2021).

the importance of collateral prices, and a key focus of our paper is regulatory classification. Our applications focus on the regulatory classification of shadow banks and on international capital flow regulation.

A closely related normative literature studies the impacts of unregulated shadow banks on different forms of regulation. Begenau and Landvoigt (2020) study quantitatively the impacts of capital requirements when banks face competition from shadow banks. Bengui and Bianchi (2021) show that the optimal debt tax on banks may be larger in the presence of otherwise identical unregulated shadow banks because the magnitude of their pecuniary externality increases, even though shadow banks increase risk due to safer banks. Several papers study the impacts of shadow banks on capital requirements/debt taxes (Plantin 2014, Huang 2018, Martinez-Miera and Repullo (2019)), liquidity regulation (Grochulski and Zhang 2019), and reserve requirements (He et al. 2018). Ordoñez (2018) and Farhi and Tirole (2020) study incentives for a bank to choose to become regulated. Ordoñez (2018) proposes explicit subsidies (transfers) to banks who opt to be regulated, whereas Farhi and Tirole (2020) propose that the promise of public support, such as lender-of-last-resort and deposit insurance, incentivizes banks to choose to be regulated. Our contribution is to provide a general framework with pecuniary externalities in which unregulated financial actors may differ from regulated financial actors in regulatory status, fundamentals, and in their endogenous activities, and in which the regulator has a rich set of regulatory instruments. We use this framework not only to characterize the impact of regulatory arbitrage on optimal regulation, but also to propose a classification scheme for unregulated finance. This literature builds an a substantial literature on pecuniary externalities and optimal regulation when all financial actors are regulated (e.g., Caballero and Krishnamurthy 2001, Bianchi 2011, Bianchi and Mendoza 2018, Dávila and Korinek 2018, Farhi et al. 2009, and Lorenzoni 2008).[8] Our study begins with a benchmark environment in which all financial actors are regulated, and then relaxes this assumption to both study the impacts on optimal regulation and to propose a classification scheme for unregulated finance.

The normative literature is guided by the positive literature studying shadow banking and capital flows. The positive literature on shadow banking studies differences between banks and shadow banks, including regulatory arbitrage (Acharya et al. 2013, Claessens et al. 2012, Gorton et al. 2010), combinations of regulatory arbitrage and technological differences (Buchak et al. 2018), implications of deposit insurance access for asset holdings (Hanson et al. 2015) or side of the asset resale market (Chretien and Lyonnet 2020), different forms of liquidity creation (Moreira and Savov 2017), and differences between safe hand insurance companies and weak hand mutual funds (Coppola 2021). The positive literature on capital flows studies the sectoral drivers of capital

[8]See also related work on optimal regulation with aggregate demand externalities (Schmitt-Grohé and Uribe 2016, Farhi and Werning 2016, Korinek and Simsek 2016) and fiscal externalities (Farhi and Tirole 2012, Bianchi 2016, Chari and Kehoe 2016). See Bianchi and Lorenzoni (2021) and Erten et al. (2021) for recent syntheses and surveys of the capital controls literature.

flows (Avdjiev et al. 2018), determinants of international banking (Niepmann 2015, Shen 2019), sudden stops, flight and retrenchment (Calvo (1988), Mendoza 2010, Milesi-Ferretti and Tille 2011, Forbes and Warnock 2012, Broner et al. 2013, Davis and Van Wincoop 2018, Caballero and Simsek 2020), global imbalances (Caballero et al. 2008, Mendoza et al. 2009, Maggiori 2017), currency denomination (Maggiori et al. 2019), information acquisition (De Marco et al. 2019), and tax havens (Coppola et al. 2021). Our framework identifies estimable sufficient statistics that in conjunction with empirical work can guide policymakers in determining the impact of regulatory arbitrage and the classification of unregulated finance.

2 Model

The economy is populated by financial actors—such as banks, mutual funds, and international portfolio investors—and nonfinancial actors—such as firms and households. Financial actors are indexed by their type $i \in \mathcal{I}$, where $\mathcal{I}$ is a finite set, and μ_i denotes the mass of type i financial actors in the economy. Nonfinancial actors are indexed by their type $h \in \mathcal{H}$ and their mass is denoted μ_h. The total measure across both types of agents sums to one. While we introduce financial and nonfinancial actors in a similar and general manner, the latter will not be subject to financial regulation. As a result, it is expositionally convenient to present and discuss the two groups of agents separately.

The economy has $M+1$ goods available for trade: a numeraire, denoted by $c \in \mathbb{R}$, as well as a vector of M goods, denoted by $I \in \mathbb{R}^M$, which are traded at a price vector p. Separating out the numeraire will be notationally convenient. Since c and I denote flows of goods between agents, they will be in zero net supply. Consistent with our main applications, our notation is suggestive of a numeraire consumption good c and a vector of investment goods I. Our framework is sufficiently general, however, that traded goods can correspond to trades of different types of goods – such as consumption goods, investment goods, state contingent securities, and labor supply – both within and between dates. For example, the set of markets may enable agents to trade a consumption good and a capital good over a set of histories over a set of periods.

2.1 Financial Actors

Financial actor i can engage in trades c_i of the numeraire and I_i of other goods, with $I_i(m) > 0$ denoting the purchase of good m. The decision problem of financial actor i is given by

$$\max_{c_i, I_i} \omega_i U_i(c_i, I_i)$$

subject to

$$c_i + pI_i \leq w_i \tag{1}$$

$$\Gamma_i(c_i, I_i, p) \geq 0 \tag{2}$$

where $\Gamma_i : \mathbb{R}^{M+1} \times \mathbb{R}^M_+ \to \mathbb{R}^{O_i}$ is vector-valued. We scale i's lifetime utility, U_i, by a social welfare weight, $\omega_i > 0$, without loss of generality. This facilitates the characterization of the regulatory planning problem below. Equation (1) is the budget constraint of financial actor i over all trades of goods. Since this budget constraint is defined over trades, i.e., flows of goods between actors, the aggregate tradeable wealth across all actors in the economy must be zero, that is $\sum_i \mu_i w_i = 0$. Equation (2) represents a general constraint set that may include technological constraints (e.g., production functions and capital evolution equations), financing constraints (e.g., collateral constraints), incomplete markets constraints (e.g., restrictions on goods trades), and so on.

From here, we define the private Lagrangian of financial actor i as

$$\mathcal{L}_i = \omega_i U_i(c_i, I_i) + \lambda_i(w_i - c_i - pI_i) + \Lambda_i \Gamma_i(c_i, I_i, p), \tag{3}$$

which also corresponds to i's indirect utility function, $\mathcal{L}_i(p, w_i)$, when substituting in the Marshallian demand functions $c_i(p, w_i)$ and $I_i(p, w_i)$. The Lagrange multiplier $\lambda_i > 0$ is the marginal value of tradeable wealth to financial actor i, and the multiplier $\Lambda_i \in \mathbb{R}^{O_i}$ on the constraint set is vector-valued.

Purchase price and collateral price. In our model, the price vector p appears in both the budget constraint (1) and the constraint set (2). It will at times be expositionally helpful to the "purchase price" of goods to be the price P appearing in the budget constraint, and the "collateral price" of goods to be the price q appearing in the constraint set. Absent planner intervention, the purchase price and collateral price are both equal to the equilibrium price, that is $P = q = p$. This notation is helpful because we will allow the planner to impose regulatory taxes that affect behavior by changing the purchase price. However, these taxes will not affect the collateral price, that is the planner cannot directly relax constraints such as collateral constraints.

Regulated and unregulated financial actors. The economy may be populated by both regulated financial actors, denoted $i \in \mathcal{B} \subset \mathcal{I}$, and unregulated financial actors, denoted $i \in \mathcal{S} \subset \mathcal{I}$. If $\mathcal{S} = \emptyset$, then all financial actors in the economy are regulated by the planner. We have adopted the suggestive notation of regulated financial actors, $i \in \mathcal{B}$, corresponding to conventional "banks" and unregulated actors, $i \in \mathcal{S}$, to "shadow banks." Our first main application in Section 5.1 develops a model of shadow banking.

Example 1: capital stocks. Our framework naturally nests models of capital and investment. Let $I_i(1),...,I_i(T)$ denote the capital investment decision of financial actor i in periods $t = 1,\dots,T$. Capital evolution then implies that the capital stock at date t is given by $k_{i,t} = \sum_{m=1}^{t}(1-\delta)^{t-m}I_i(m)$, where δ denotes the rate of depreciation. Financial actor i's outstanding stock of capital in period t has market value $p(t)k_{i,t}$.

2.2 Nonfinancial Actors

Nonfinancial actor h also engages in trades of goods, denoted c_h and I_h. Because in aggregate nonfinancial actors must be on the opposite side of each market as financial actors, it is notationally convenient to denote $I_h(m) > 0$ to be the *supply* of good m by nonfinancial actor h, rather than demand, and, similarly, $c_h > 0$ to be the supply of the numeraire. Adopting this notational convention is without loss of generality. As with financial actors in equation (3), we define the private Lagrangian of nonfinancial actor h as

$$\mathcal{L}_h = \omega_h U_h(c_h, I_h) + \lambda_h(c_h + pI_h - w_h) + \Lambda_h \Gamma_h(c_h, I_h, p), \tag{4}$$

where h's lifetime utility, U_h, is once again scaled by a social welfare weight, $\omega_h > 0$. $\lambda_h > 0$ again denotes the marginal value of wealth and $\Lambda_h \in \mathbb{R}^{O_h}$ is vector-valued. Under our convention that nonfinancial actors supply goods, h's tradeable wealth is now given by $-w_h$, that is $w_h < 0$ corresponds to positive wealth.

Example 2: labor supply. Let tradeable good m' denote the supply of labor. Then nonfinancial actor h supplies (sells) labor, $I_h(m') > 0$, and is compensated at wage rate $p(m')$.

Example 3: second-best capital users. In models with collateral constraints, second-best users of capital are one simple way to model fire sales (see, e.g., Kiyotaki and Moore, 1997). In our framework, let h be an arbitrageur that can purchase the capital good, $-I_h(m) > 0$, and transform it into the consumption good using a technology $\mathcal{F}_h$. The net income earned by arbitrageur h is then given by $\mathcal{F}_h(-I_h(m)) + p(m)I_h(m)$, so that the demand of arbitrageur h for the capital good solves $\frac{\partial \mathcal{F}_h(-I_h(m))}{\partial(-I_h(m))} = p(m)$.

2.3 Market Clearing and Competitive Equilibrium

In general equilibrium, the markets for all $M+1$ traded goods must clear. We adopt the notational convention that $I_{\mathcal{N}} \in \mathbb{R}^M$ denotes the aggregate traded goods positions summed over a set $\mathcal{N}$ of agents. That is, $I_{\mathcal{I}}(m) = \sum_{i\in\mathcal{I}} \mu_i I_i(m)$ is the aggregate demand of all financial actors for good m and $I_{\mathcal{H}}(m) = \sum_{h\in\mathcal{H}} \mu_h I_h(m)$ is the aggregate supply of good m by all nonfinancial actors.

Market clearing then requires that

$$I_{\mathcal{I}} = I_{\mathcal{H}}. \tag{5}$$

There is also a market clearing condition for the numeraire, $c_{\mathcal{I}} = c_{\mathcal{H}}$, which we can drop due to Walras' law. From here, a competitive equilibrium is an allocation, $\{c_i, I_i\}$ and $\{c_h, I_h\}$, and a vector of prices, p, such that financial and nonfinancial actors optimize given prices according to (3) and (4), and markets clear for all tradeable goods according to (5).

2.4 Sufficient Statistics

We now characterize two sufficient statistics that will play an important role in the results we present below.

Definition 1. Let $\Xi_i = \nabla_p I_i$ denote a Jacobian, ordered such that $\Xi_i(m,m') = \frac{\partial I_i(m')}{\partial p(m)}$.

1. The *aggregate supply response* is defined by

$$\Xi_{\mathcal{H}} \equiv \nabla_p I_{\mathcal{H}} = \sum_{h\in\mathcal{H}} \mu_h \Xi_h \tag{6}$$

   which measures the response of aggregate nonfinancial actor supply to price changes.

2. The *aggregate unregulated demand response* is defined by

$$\Xi_{\mathcal{S}} \equiv \nabla_p I_{\mathcal{S}} = \sum_{i\in\mathcal{S}} \mu_i \Xi_i \tag{7}$$

   which measures the response of aggregate *unregulated* financial actor demand to price changes.

Definition 1 provides two important mappings for results to come. The first is the aggregate supply response, which measures how aggregate supply of nonfinancial actors responds to a change in the equilibrium price vector p. Formally, we have that $dI'_{\mathcal{H}} = dp'\Xi_{\mathcal{H}}$, where x' indicates the transpose of x. Inverting $\Xi_{\mathcal{H}}$, this means that $dp' = dI'_{\mathcal{H}}\Xi_{\mathcal{H}}^{-1}$, which is to say that $\Xi_{\mathcal{H}}^{-1}$ provides the inverse mapping from changes in prices, dp, into changes in the aggregate supply of nonfinancial actors, $dI_{\mathcal{H}}$. From market clearing, $I_{\mathcal{H}} = I_{\mathcal{I}}$, this then also provides the mapping from a change in aggregate demand of financial actors into the price changes required for nonfinancial actors to be willing to accommodate that change in demand. Similarly, the matrix $\Xi_{\mathcal{S}}$ provides the mapping from changes in prices to changes in the aggregate demand of unregulated financial actors, that is $dI'_{\mathcal{S}} = dp'\Xi_{\mathcal{S}}$.

3 Optimal Regulation

In this section, we characterize optimal regulation of financial actors in this economy. We begin in Section 3.1 by studying the case where all financial actors can be regulated, which serves as an important benchmark. In Section 3.2, we then derive optimal regulation in an environment with both regulated and unregulated financial actors. In both cases, the behavior of nonfinancial actors, who are not subject to financial regulation, must be taken as given by the social planner when designing regulation for financial actors. When there are also unregulated financial actors, the planner has to take their behavior as given as well when designing regulation for regulated financial actors. We show that optimal regulation in the presence of unregulated financial actors is characterized by a regulatory arbitrage multiplier, which is the main result of this section.[9]

3.1 Optimal Regulation Without Unregulated Financial Actors

Consider first the case where all financial actors can be subjected to financial regulation, that is $\mathcal{B} = \mathcal{I}$ and $\mathcal{S} = \emptyset$. Regulation of financial actor i takes the form of a complete vector of taxes, $\tau_i \in \mathbb{R}^M$, where $\tau_i(m)$ denotes the tax rate i must pay on trades of good m. The optimal tax on the numeraire c_i will be zero, and we omit it in our exposition. Therefore, the modified budget constraint of financial actor i can be written as $c_i + (p + \tau_i)I_i \leq w_i + T^*$, where $T^* = \sum_m \tau_i(m)I_i(m)$ denotes equilibrium remitted revenues, which i takes as given. This means that the purchase price $P_i = p + \tau_i$ is affected by regulation, while the collateral price $q = p$ is not. With a complete set of taxes for all regulated financial actors, the planner is able to select any feasible allocation $\{c_i, I_i\}_{i \in \mathcal{B}}$ through the appropriate choice of taxes. Thus, we can equivalently think of the problem as the social planner directly choosing the allocations of regulated financial actors and then back out the taxes τ_i that implement these allocations in equilibrium.

The social welfare function of the planner is

$$\sum_{i \in \mathcal{I}} \mu_i \mathcal{L}_i + \sum_{h \in \mathcal{H}} \mu_h \mathcal{L}_h, \tag{8}$$

where the Pareto weights ω_i and ω_h are already incorporated in the respective Lagrangians. The problem of the planner is therefore to choose financial regulation $\{\tau_i\}_{i \in \mathcal{I}}$ to maximize social welfare (8), taking as given the behavior of nonfinancial actors, $c_h(p, w_h)$ and $I_h(p, w_h)$, and that the equilibrium price is determined through market clearing (5). The following proposition characterizes optimal financial regulation.

[9] In Appendix C, we analyze optimal regulation in an extended environment that allows for ingredients such as non-pecuniary externalities.

Proposition 2. *Suppose that all financial actors are regulated ($\mathcal{S} = \emptyset$). Then, optimal financial regulation satisfies $\tau_i = \tau$ for all $i \in \mathcal{I}$, with*

$$\tau = -\frac{1}{\lambda}\Xi_{\mathcal{H}}^{-1}\mathcal{E} \tag{9}$$

where $\lambda_i = \lambda$ for all i, and where $\mathcal{E} \in \mathbb{R}^M$ is the vector of pecuniary externalities given by

$$\mathcal{E} = \underbrace{\sum_{h\in\mathcal{H}} \mu_h \lambda_h I_h - \sum_{i\in\mathcal{I}} \mu_i \lambda_i I_i}_{\text{Distributive Externalities}} + \underbrace{\sum_{h\in\mathcal{H}} \mu_h \nabla_p \Lambda_h \Gamma_h + \sum_{i\in\mathcal{I}} \mu_i \nabla_p \Lambda_i \Gamma_i}_{\text{Collateral Externalities}}. \tag{10}$$

Proposition 2 tells us that optimal regulation is determined by the interaction of two objects: the aggregate supply response introduced in Definition 1, Ξ_h, and the vector $\mathcal{E}$ of pecuniary externalities induced by price changes. Taken together, these objects map changes in flows in the financial sector into the externalities that result from them. Consider a change in the demand for traded goods by financial actor i, denoted dI_i. This change in demand generates a change $dI_{\mathcal{I}} = \mu_i dI_i$ in the aggregate flows of the financial sector, which must then be accommodated by the nonfinancial sector, that is $dI_{\mathcal{H}} = \mu_i dI_i$ by market clearing. The aggregate supply response determines the change in prices necessary to induce the nonfinancial sector to absorb the change in demand, that is $dp' = \mu_i dI_i' \Xi_{\mathcal{H}}^{-1}$. Finally, $\mathcal{E}$ maps the resulting change in prices into its effect on the planner's welfare function, $dp'\mathcal{E}$. Put together, this provides the mapping $\mu_i d_i' \Xi_{\mathcal{H}}^{-1}\mathcal{E}$ from changes in flows of a regulated financial actor into changes in social welfare induced through changes in equilibrium prices. Scaling by the private marginal value of wealth λ_i puts the regulation in money-metric terms, where we note that a constrained efficient allocation features $\lambda_i = \lambda$ across regulated financial actors so that the private marginal values of wealth are equalized.[10] Finally because the measure of financial actor i is μ_i, this same measure also scales the total private benefit to i from a welfare perspective, so that μ_i drops out of the optimal tax formula.

In this model, non-zero-sum pecuniary externalities arise because prices appear in the constraint sets of financial and nonfinancial actors, for example due to collateral constraints.[11] A source of Pareto inefficiency, these externalities provide a motivation for regulation. The total externality of a change in the price $p(m)$ is captured by $\mathcal{E}$, that is to say $\mathcal{E}(m) = \sum_{j\in\mathcal{I}\cup\mathcal{H}} \mu_j \frac{\mathcal{L}_j}{\partial p(m)}$. In a Pareto efficient economy, we would have $\mathcal{E}(m) = 0$ for all m.

Equation (10) decomposes the total externality induced by a change in price $p(m)$ into two types of externalities: distributive externalities and collateral externalities (see Dávila and Korinek,

[10]Formally, equating of private marginal values of wealth can happen either through the required choice of Pareto weights ω_i, or through the use of lump sum transfers required to achieve Pareto efficiency.

[11]Pareto inefficiency can also arise due to incomplete markets (see, e.g., Geanakoplos and Polemarchakis, 1985).

2018). Distributive externalities arise because an increase in price redistributes money away from buyers of good m towards sellers of good m. In a Pareto efficient economy absent prices in constraints, distributive externalities sum to zero: An increase in the price of good m benefits the sellers but reduces the welfare of the buyers of good m. When prices appear in constraints, however, distributive externalities need not be zero. In particular, a constrained efficient allocation equalizes the *social* marginal value of wealth across agents, but the distributive externality depends on the private marginal value of wealth. In general, the private and social marginal values of wealth in this model differ for nonfinancial actors. This is because a change in nonfinancial actor wealth dw_h not only induces a change in private value but also elicits a response in Marshallian demand, $I_h(p, w_h)$, which in turn leads to price changes and thus pecuniary externalities.

The collateral externality term in $\mathcal{E}(m)$ reflects the total welfare impact of a change in price $p(m)$ through changes in the constraint set. For example, an increase in the collateral price relaxes the collateral constraint, which generates a positive externality. In welfare terms, the shadow price Λ_i of relaxing the constraint through price changes reflects this collateral externality.

Finally, Proposition 2 features an "equal-treatment" property, whereby optimal regulation does not depend on the identity of agents, $\tau_i = \tau$. The pecuniary externality that results from changes in prices depends on aggregate demand by financial actors, which must be absorbed by nonfinancial actors. It is only the change in aggregate demand itself, therefore, that matters from an externality perspective, and not the identity of the financial actor that generates the change in demand. As a result, the externality generated by different financial actors (normalizing by their measure) for the same change in demand is the same, justifying the same tax treatment across all regulated financial actors. Equal regulatory treatment of financial actors has been emphasized in policy debates on both bank regulation and capital flow management.[12]

3.2 Optimal Regulation With Unregulated Financial Actors

We now study the design of financial regulation when some financial actors cannot be regulated, that is $\mathcal{S} \neq \emptyset$. The social planner now only possesses instruments τ_i for agents $i \in \mathcal{B}$. As before, the planner optimally chooses financial regulation $\{\tau_i\}_{i \in \mathcal{B}}$ for those actors that can be regulated to maximize social welfare (8), now taking as given the behavior of both nonfinancial actors and unregulated financial actors, and internalizing the price impact through market clearing (5). The following proposition characterizes optimal regulation in the presence of unregulated financial actors and is the main result of this section.

[12]The ECB lists "ensuring a level playing field and equal treatment of all supervised institutions" as an objective of the Single Supervisory Mechanism (ECB, 2018), while the IMF states that "[i]t is generally preferable that CFMs not discriminate between residents and non-residents" (IMF, 2012). Equal treatment has also been emphasized in the prior academic literature (see, e.g., Clayton and Schaab, 2022).

Proposition 3. *Suppose there are unregulated financial actors ($\mathcal{S} \neq \emptyset$). Then, optimal financial regulation is given by $\tau_i^* = \tau^*$ for all $i \in \mathcal{B}$, with*

$$\tau^* = \mathcal{M}\tau \tag{11}$$

where $\mathcal{M} = (\mathbb{I} - \Xi_{\mathcal{H}}^{-1}\Xi_{\mathcal{S}})^{-1}$ is the regulatory arbitrage multiplier*, where $\mathbb{I}$ is the identity matrix, and where τ is given by the formula in equation (9).*

Optimal regulation in the presence of unregulated financial actors is characterized by the same formula as optimal regulation with only regulated financial actors, up to one additional term: the *regulatory arbitrage multiplier* $\mathcal{M}$. Notice that when $\Xi_{\mathcal{S}} = 0$, as is the case without unregulated agents, then $\mathcal{M} = \mathbb{I}$ and we recover the result of Proposition 2.

To build intuition for the regulatory arbitrage multiplier, consider a first-order approximation around $\Xi_{\mathcal{H}}^{-1}\Xi_{\mathcal{S}} = 0$. We have

$$\mathcal{M} \approx \underbrace{\mathbb{I}}_{\text{Direct Impact}} + \underbrace{\Xi_{\mathcal{H}}^{-1}\Xi_{\mathcal{S}}}_{\text{First Round Regulatory Arbitrage}}$$

To illustrate the economic mechanism, consider again a perturbation dI_i in the demand for traded goods by regulated financial actor $i \in \mathcal{B}$. To first order, $\mathcal{M}$ reflects two effects of such a perturbation. First, dI_i has a direct effect on the aggregate demand for traded goods by financial actors, which was the effect underlying Proposition 2. Second, there is an indirect effect in general equilibrium, whereby unregulated financial actors respond to the change in prices that results from the direct effect. In particular, the direct effect of perturbation dI_i results in (first-round) price changes $dp' = dI_i'\Xi_{\mathcal{H}}^{-1}$, as the nonfinancial sector must be induced to change its supply to accommodate the change in financial sector demand. These price changes in turn elicit a demand response by unregulated financial actors, given by $dX_{\mathcal{S}}' = dp'\Xi_{\mathcal{S}} = dI_i'\Xi_{\mathcal{H}}^{-1}\Xi_{\mathcal{S}}$. As a result, to first order the total change in aggregate financial actor demand that must be absorbed by nonfinancials is $dI_i' + dI_{\mathcal{S}}' = dI_i'(\mathbb{I} + \Xi_{\mathcal{H}}^{-1}\Xi_{\mathcal{S}})$, with $\Xi_{\mathcal{H}}^{-1}\Xi_{\mathcal{S}}$ capturing regulatory arbitrage to first order.

Beyond first order, the process of regulatory arbitrage iterates, with changes in unregulated demand fueling price changes. This feedback loop generates a fixed point,

$$\mathcal{M} = \underbrace{\mathbb{I}}_{\text{Direct Impact}} + \underbrace{\sum_{t=1}^{\infty}(\Xi_{\mathcal{H}}^{-1}\Xi_{\mathcal{S}})^t}_{\text{Total Regulatory Arbitrage}} = (\mathbb{I} - \Xi_{\mathcal{H}}^{-1}\Xi_{\mathcal{S}})^{-1},$$

which captures the final price change following the iterated process of regulatory arbitrage by unregulated financial actors.[13]

To further illuminate the impact of regulatory arbitrage on optimal regulation, we study the special case where $\Xi_{\mathcal{H}}$ is diagonal, that is, the cross-price elasticities of aggregate nonfinancial actor supply are zero. For example, this assumption is satisfied in a model with investment (capital) goods over multiple periods, where a different capital producing firm produces or deconstructs the capital good in each period.[14] In this limit case, our first-order characterization of $\mathcal{M}$ becomes even sharper and yields additional insights.

Corollary 4. *Suppose that $\Xi_{\mathcal{H}}$ is diagonal. Under a first-order approximation of the regulatory arbitrage multiplier, optimal policy is then given by*

$$\tau^*(m) \approx \underbrace{\left(1+\alpha(m)\frac{\xi_{\mathcal{S}}(m,m)}{\xi_{\mathcal{H}}(m,m)}\right)\tau(m)}_{\text{Own Price}} + \underbrace{\sum_{m'\neq m}\alpha(m')\frac{\xi_{\mathcal{S}}(m,m')}{\xi_{\mathcal{H}}(m,m)}\frac{I_{\mathcal{I}}(m')}{I_{\mathcal{I}}(m)}\tau(m')}_{\text{Cross Price}} \tag{12}$$

where $\xi_{\mathcal{S}}(m,m') = \frac{p(m)}{I_{\mathcal{S}}(m')}\frac{\partial I_{\mathcal{S}}(m')}{\partial p(m)}$ is the unregulated financial sector demand elasticity, $\alpha(m) = \frac{I_{\mathcal{S}}(m)}{I_{\mathcal{I}}(m)}$ is the unregulated share of total financial sector flows, and $\frac{I_{\mathcal{I}}(m')}{I_{\mathcal{I}}(m)}$ is the relative market size measured by flow volume.

Corollary 4 reveals that, to first order, there are three objects that are important for determining the implications of regulatory arbitrage for optimal regulation: (i) the market shares of unregulated agents, α, (ii) the relative price elasticities of nonfinancial actor supply and unregulated financial actor demand, $\frac{\xi_{\mathcal{S}}}{\xi_{\mathcal{H}}}$, and (iii) the relative sizes of different markets in terms of total financial sector flow volume, $I_{\mathcal{I}}$. In particular, Corollary 4 highlights two channels through which regulatory arbitrage affects optimal financial regulation, reflecting how changes in regulation of one market affect unregulated financial actor demand across all markets. In the special case where $\Xi_{\mathcal{H}}$ is diagonal, the direct effect that results from a reduction in demand for good m is an isolated change in the price of good m, which is captured by the supply elasticity $\xi_{\mathcal{H}}(m,m)$. In addition to this direct effect from the regulation, there are also two sets of indirect effects that arise due to regulatory arbitrage.

First, the change in the price of good m induces an own-price effect on unregulated demand for good m. This channel is important when the own-price elasticity of unregulated financial actors is

[13] The regulatory arbitrage multiplier $\mathcal{M}$ is parallel to the Leontief inverse that characterizes input-output multipliers in the network literature. See, e.g., Baqaee and Farhi (2019).

[14] Concretely, we could model the capital producing firm at date t as having a cost $\Phi_t(I_h(t))$ of producing $I_h(t)$ of the date t capital good. Thus, we can write the objective function as maximizing $U_h = c_h - \Phi_t(I_h(t))$ subject to $c_h + p(t)I_h(t) \leq w_h$. To see the limitations of this assumption, note that it rules out the possibility that adjustment costs Φ_t depend on the size of the existing capital stock, which depends on $I_h(t-1)$. In such models, the zero cross-price elasticity assumption does not generally hold.

high relative to that of nonfinancial actors. In this case, small price changes induce large behavioral demand responses by unregulated actors, which in turn require large price adjustments to encourage relatively inelastic nonfinancials to meet that demand. The own-price effect is also large when unregulated actors constitute a large share of total flows in market m, which amplifies the price impact for given elasticities. To build intuition, suppose that the planner tries to increase the price of good m. To increase the price of m, the planner needs to increase the amount that must be supplied by nonfinancial actors, which is done by increasing demand by regulated financial actors. When the demand elasticity of unregulated financial actors and the supply elasticity of nonfinancial actors have opposite signs, unregulated financial actor decrease their demand for good m in response to the price increase. Their reduction in demand means that the overall increase in demand is lower, leading to a smaller price increase. This makes the regulation less effective and promotes an optimal tax rate $\tau^*(m)$ that is lower in magnitude that $\tau(m)$. By contrast, if unregulated demand and nonfinancial supply elasticities have the same sign, then the price increase actually *increases* demand by unregulated financial. This increase in demand amplifies the original demand increase by regulated financial actors, leading to further price increases. This makes the impact of regulation stronger, rather than weaker, and promotes a larger optimal tax rate $\tau^*(m)$. We will show in Section 3.3 that this latter case can be economically relevant in models with pecuniary externalities.

The second channel works through cross-price effects. A change in the price of m also induces cross-price substitution by unregulated financial actors into or out of other markets, $m' \neq m$. As demand changes, equilibrium prices in these markets adjust, which leads to pecuniary externalities. Even if an activity m is not directly associated with an externality, that is $\tau(m) = 0$, it may become regulated in equilibrium in the presence of regulatory arbitrage. As with the own-price effect, cross-price substitution becomes important when unregulated financial actors constitute a large share of market m', or when the cross-price demand elasticity of unregulated actors is large relative to the own-price demand elasticity of the nonfinancial sector. Lastly, the significance of cross-price substitution also depends on the relative size of the two markets: When market m' is large relative to market m, then—all else equal—the total flow response is larger in market m', resulting in larger externalities.

3.3 Magnitude of the Regulatory Arbitrage Multiplier

According to Proposition 3, optimal financial regulation becomes either tighter or looser in the presence of unregulated financial actors depending on whether the regulatory arbitrage multiplier is "larger" or "smaller" than 1.[15] In this section, we analyze the magnitude of the regulatory arbitrage

[15]The simplest, although incomplete, way of thinking of the idea of $\mathcal{M}$ being "larger" or "smaller" than one is whether the diagonal entries $\mathcal{M}(m,m)$ are larger than or smaller than 1. More generally, it can be understand from the

multiplier and study its determinants.

An intuitve line of reasoning may suggest that, in the presence of regulatory arbitrage, optimal financial regulation should become looser. Consider a standard model of collateral constraints and forced deleveraging, where both regulated and unregulated financial actors contribute to fire sales.[16] Tighter regulation of regulated financial actors increases financial stability by raising the fire sale price and relaxing the collateral constraint. Higher liquidation prices, however, encourage unregulated financial actors to take on more risk. By this logic, regulatory arbitrage reduces the effectiveness of financial regulation and calls for smaller taxes τ^*. This argument emphasizes a classical substitution effect, whereby unregulated financial actors' *supply* of capital liquidations increases in the fire sale price. In the language of Corollary 5, the argument corresponds to the case $\alpha(m)\xi_{\mathcal{S}}(m,m) < 0$ for assets m that are fire sold in a crisis.

We now show that models of pecuniary externalities naturally give rise to a counterveiling force that motivates *tighter* regulation in the presence of unregulated financial actors: the income effect. Intuitively, when financial regulation raises fire sale prices, unregulated financial actors are less likely to face binding collateral constraints, earn more income from liquidations, and benefit from greater collateralizability of their assets. These positive pecuniary externalities loosen the binding collateral constraints of unregulated financial actors, and so lead to fewer forced liquidations. When this income effect dominates the classical substitution effect, higher taxes on regulated financial actors have net positive stability spillovers on the liquidation decisions of unregulated financial actors. Financial regulation becomes more rather than less effective in this case, with tighter optimal regulation in the presence of regulatory arbitrage.

3.3.1 An Example: the Liquidity Shock Model

We begin with an illustrative case to convey the intuition, followed by a formal characterization of income effects in the presence of regulatory arbitrage. Suppose that there are three dates, $t = 1,2,3$. There are two markets, $M = 2$, one for investment at date 1 and one for investment at date 2. We assume that all investment fully depreciates at date 3. Unregulated financial actors are endowed with a fixed scale A_i in the investment project at date 1 but no other tradeable wealth, so they do not participate in date 1 markets. At the intermediate date 2, i's project experiences a "liquidity event" ρ_i.

case that $M = 2$. In this case $\mathcal{M}$ is two-by-two, and we have

$$\tau^*(1) = \mathcal{M}(1,1)\tau(1) + \mathcal{M}(1,2)\tau(2) = \left(\mathcal{M}(1,1) + \mathcal{M}(1,2)\frac{\tau(2)}{\tau(1)}\right)\tau(1).$$

Thus, formally what we mean is that $\mathcal{M}(1,1) + \mathcal{M}(1,2)\frac{\tau(2)}{\tau(1)}$ is larger or smaller than 1, which can be generated either by a sufficiently large diagonal entry, $\mathcal{M}(1,1) > 1$, or by an appropriate off-diagonal entry $\mathcal{M}(1,2)\frac{\tau(2)}{\tau(1)} > 1$.

[16] Many such models exist with only regulated financial actors, for example the canonical Kiyotaki and Moore (1997).

The liquidity event ρ_i is not stochastic, that is its value for unregulated financial actor i is known with certainty at date 1. If $\rho_i > 0$, then i has a liquidity need of size ρ_i, which is the amount it must pay to maintain the project. And since project cashflows are not pledgeable to investors, i must liquidate part of the project at date 2 to meet this liquidity need, that is $p(2)I_i(2) = -\rho_i$, where we recall that $I_i < 0$ indicates supply. If instead $-\rho_i > 0$, then i has a liquidity surplus and can use this surplus to buy projects. As cashflows are not pledgeable, i's purchases are constrained to $p(2)I_i(2) = -\rho_i$. Therefore, unregulated financial actors with liquidity needs are forced sellers at date 2 (Holmström and Tirole 1998), whereas unregulated financial actors with liquidity surpluses are arbitrageurs with limited wealth (Allen and Gale 1994). Finally, there are two different nonfinancial actors, one at date 1 and one at date 2, who live for one period and convert projects into consumption goods (as in Example 3). The cross-price elasticities of the date 1 and date 2 markets are therefore zero. We assume they have a constant elasticity of 1 at both dates.[17]

We colloquially refer to this model as the "liquidity shock model." We obtain the following result.

Corollary 5. *In the liquidity shock model, optimal regulation of market $m = 2$ in the presence of unregulated financial actors is given by*

$$\tau^*(2) = \left(1 + \frac{I_{\mathcal{S}}(2)}{I_{\mathcal{B}}(2)}\right)\tau(2) \tag{13}$$

Corollary 5 shows that the impact of regulatory arbitrage on optimal financial regulation in the liquidity shock model depends on whether or not regulated and unregulated financial actors are in aggregate on the same side of the market for date 2 liquidations, i.e., whether the ratio $\frac{I_{\mathcal{S}}(2)}{I_{\mathcal{B}}(2)}$ is positive or negative. When on the same (opposite) side of the liquidation market, the optimal tax on regulated financial actors is *larger* (smaller) than in the case where all financial actors are regulated. The underlying economic force is a dominating income effect from financial stability in the liquidity shock model. When regulated financial actors are selling assets at date 2 and unregulated financial actors experience liquidity shocks, then all financial actors are forced to engage in sales at date 2 to pay their liquidity needs. Since unregulated financial actors must raise a fixed amount ρ_i by selling assets, they need to liquidate less of the project to meet liquidity needs when the liquidation price rises. In other words, $-I_i(2)$ falls as $p(2)$ increases. When tighter financial regulation reduces

[17] For a more general elasticity, we have $\tau^*(2) = \left(1 + \frac{1}{1+\frac{\xi_{\mathcal{H}}(2)-1}{\xi_{\mathcal{H}}(2)}\frac{I_{\mathcal{S}}(2)}{I_{\mathcal{B}}(2)}}\frac{I_{\mathcal{S}}(2)}{I_{\mathcal{B}}(2)}\right)\tau(2)$. Thus the same side of the market result generalizes immediately when $\xi_{\mathcal{H}} \geq 1$ while the opposite side of the market result generalizes immediately for $\xi_{\mathcal{H}} \leq 1$. Both results generalize more broadly when the market share of regulated financial actors is sufficiently large relative to the market share of unregulated financial actors. For example if market shares are fixed but nonfinancial actor supply is highlight inelastic, $\xi_{\mathcal{H}}(2) \to 0$, then $\mathcal{M}(2,2) \to 0$.

forced asset sales by regulated financial actors, thus raising the equilibrium price, it also *reduces* forced asset sales by unregulated financial actors. Thus further promoting financial stability, the effectiveness of financial regulation increases. Optimal regulation therefore becomes tighter, with larger taxes $\tau^*(2)$ (in magnitude) on regulated financial intermediaries.

The above logic rests on $\rho_i > 0$, so that regulated and unregulated financial actors sell assets at date 2 in tandem. Suppose that instead $-\rho_i > 0$, so that unregulated financial actors are buying assets at date 2 from regulated financial actors. In this case, unregulated financial actors are serving as arbitrageurs with a fixed amount of arbitrage capital $-\rho_i$ to engage in purchases. Now when the price $p(2)$ rises, the loss-absorbing capacity $I_{\mathbb{S}}(2)$ of unregulated financial actors falls due to their decreasing purchasing power. This means that the price rise forces more fire sold assets to be purchased by nonfinancial actors, rather than by unregulated financial actors, leading to an offsetting decrease in the price $p(2)$. This undermines the effectiveness of financial regulation, and leads the optimal tax $\tau^*(2)$ to be smaller in magnitude than in the case without unregulated financial actors.

In sum, Corollary 5 tells us that the income effect can amplify the effectiveness of financial regulation when higher fire sale prices relax constraints of unregulated financial actors, and so reduce their forced asset sales. By contrast, the income effect undermines the effectiveness of financial regulation when unregulated financial actors play role of arbitrageurs in asset resale markets, by reducing their loss absorbing capacity.

3.3.2 Income Effects in the Presence of Regulatory Arbitrage

Having introduced in a simple example the economic forces that determine the magnitude of the regulatory arbitrage multiplier, we now show formally that regulatory arbitrage is governed by a set of competing income and substitution effects. In particular, we focus on unpacking the own-price response and determining its direction. To do so, we employ tools from price theory to give a decomposition of own-price demand, $\frac{\partial I_i(m)}{\partial p(m)}$, which underlies the own-price elasticity, into income and substitution effects. Our decomposition differs from the canonical one because prices appear not only in the budget constraint (purchase price), but also in the constraint set (collateral price).[18] In particular we define the Marshallian demand function $x_i(P,q,w_i)$ to be Marshallian demand that explicitly separates out the purchase price P and the collateral price q, where in equilibrium $P = q = p$. This allows us to denote separate derivatives in these prices. We similarly define Hicksian demand $h_i(P,q,\overline{U}) \in \mathbb{R}^M$ and the expenditure function $e_i(P,q,\overline{U})$ from the expenditure minimization problem.[19]

[18] The price theory logic that follows is closely related to Farhi and Gabaix (2020), who characterize price theory decompositions when behavioral agents have demand functions that exhaust budget constraints but do not maximize utility.

[19] The expenditure minimization problem is $\min h_i^c + ph_i$ subject to $U(h_i^c, h_i) \geq \overline{U}$ and $\Gamma(h_i^z, h_i, q) \geq 0$, Hicksian demand is the solution $h_i(P,q,\overline{U})$ to this expenditure minimization problem, and the expenditure function is $e_i(P,q,\overline{U}) =$

From here, we obtain the formal decomposition of $\Xi_i(m,m)$.

Proposition 6. *A price-theory decomposition of unregulated financial actor i's demand in market m yields*

$$\Xi_i(m,m) = \underbrace{\frac{\partial h_i(m)}{\partial P(m)}}_{\substack{\text{Standard Substitution Effect} \\ <0}} + \underbrace{\frac{\partial h_i(m)}{\partial q(m)}}_{\substack{\text{Constraint Set Substitution Effect} \\ \lesseqgtr 0}} + \underbrace{\frac{\partial I_i(m)}{\partial w_i}\frac{1}{\lambda_i}\mathcal{E}_i(m)}_{\substack{\text{Total Income Effect} \\ \lesseqgtr 0}} \tag{14}$$

Proposition 6 tells us that the sign of $\Xi_i(m,m)$ generally depends on a combination of income and substitution effects. When $\Xi_i(m,m) > 0$, i.e., the own-price demand for good m increases in the price, Corollary 5 revealed that this pushes for larger taxes on regulated financial actors because unregulated financial actors amplified, rather than dampened, the demand response of regulated financial actors. The right-hand side of equation (14) consists of three terms. The first term, $\frac{\partial h_i(m)}{\partial P(m)}$, is the standard substitution effect in the purchase price. This term is negative, reflecting the usual intuitive effect that a compensated increase in a good's own price shifts demand away from that good. This term leads regulatory arbitrage to undermine the effects of tighter regulation, as reductions in demand by regulated financial actors reduce prices and fuel increases in demand by unregulated financial actors. This term thus reflects the logic underlying the claim that regulatory arbitrage might loosen optimal regulation of financial actors.

However, there are two additional forces. The first is the substitution effect in the collateral price. Unlike the purchase price substitution effect, the collateral price substitution effect cannot be signed in general. However, it can be positive in natural models of fire sales. For example, an increase in the collateral price may encourage unregulated financial actors to sell less of an asset if debt-back rollover is tied to the market value of collateral, resulting in $\frac{\partial h_i(m)}{\partial q(m)} > 0$. In this case, the collateral price substitution effect enhances the effectiveness of financial regulation: as regulation increases the price $p(m)$, unregulated financial actors are incentivized to hold, rather than sell, asset m due to the higher collateral price. From Corollary 5, this increases the magnitude of taxes on regulated financial actors.

Finally, there is also an income effect. Absent prices in constraints, the traditional income effect is based on demand $I_i(m)$: an increase in demand reduces the effective wealth of unregulated financial actor i because it costs more to purchase the same amount $I_i(m)$. With prices in constraints, the income effect is instead given by $\mathcal{E}_i(m)$, which is the total wealth-equivalent value to unregulated financial actor i of the price change. $\mathcal{E}_i(m)$ incorporates not only the classical income effect, but

$h_i^c + ph_i$ substituting in Hicksian demand.

also an additional income effect that arises from changing tightness in the constraint set. This changing constraint set tightness is equivalent to an increase in wealth, and hence incorporated into the wealth equivalent measure $\mathcal{E}_i(m)$. Hence if m is a normal good, that is $\frac{\partial I_i(m)}{\partial w_i}$ and there is a positive pecuniary externality from increasing the price $p(m)$, that is $\mathcal{E}_i(m) > 0$ as in a fire sale model, then total income effect increases demand (reduces supply) of unregulated financial actor i for good m. When pecuniary externalities are particularly large, the income effect is also large, and pushes for an upward sloping demand curve for good m in the price of good m.

Taken together, Proposition 6 tells us that although the classical substitution effect from the purchase price undermines the effectiveness of financial regulation, both the substitution effect from the collateral price and the income effect from the pecuniary externality can enhance the effectiveness of financial regulation. When the positive spillover effects of the higher collateral price dominate the negative effects of the purchase price, regulatory arbitrage enhances rather than dampens the effectiveness of financial regulation.

We started this subsection with an intuitive account of why regulatory arbitrage might make financial regulation less effective, thus implying optimal regulation should become looser in the presence of unregulated financial actors. This mechanism corresponds to a standard substitution effect that is captured by the first term in Proposition 6. In summary, we have demonstrated in this subsection that regulatory arbitrage also gives rise to two counterveiling economic forces that potentially make financial regulation more rather than less effective.

3.4 Taking the Regulatory Arbitrage Multiplier to the Data

An important question is how to quantify the regulatory arbitrage multiplier and take it to the data. A key contribution of this paper is that we can represent the regulatory arbitrage multiplier in terms of sufficient statistics that are in principle estimable in the data.

Our results can be used to inform regulators on how to adjust financial regulation in response to regulatory arbitrage by different types of unregulated financial actors. We suppose that the financial regulator already knows how to determine optimal financial regulation *without* unregulated financial actors. This requires knowledge of the nonfinancial aggregate supply response $\Xi_{\mathcal{H}}^{-1}$ and the vector of pecuniary externalities $\mathcal{E}$. We think of these as being the objects the regulator has already evaluated in order to be able to calibrate optimal financial regulation, absent concerns about regulatory arbitrage. Empirically, these are the objects that regulators have evaluated when designing post-crisis financial regulatory regimes.

Given the regulator has knowledge of the price response and the vector of externalities, the last remaining step to evaluating the regulatory arbitrage multiplier is to evaluate empirically the aggregate unregulated demand response $\Xi_{\mathcal{S}}$. We now show that we can decompose the regulatory arbitrage multiplier into a combination of aggregate financial flows and micro estimable price elasticities. In particular, element (m,m') of the regulatory arbitrage multiplier can be represented as

$$\Xi_{\mathcal{S}}(m,m') = \underbrace{\frac{1}{p(m)}}_{\substack{\text{Market}\\ \text{Price}}} \sum_{i\in\mathcal{S}} \underbrace{\xi_i(m,m')}_{\substack{\text{Price}\\ \text{Elasticities}}} \quad \underbrace{\mu_i I_i(m')}_{\substack{\text{Aggregate}\\ \text{Flows}}} \tag{15}$$

This decomposition depends on three objects that are in principle empirically observable. The first is the price $p(m)$ in the market. The second are the micro price elasticities $\xi_i(m,m')$ of flows of i. The third are the aggregate flow positions $\mu_i I_i(m')$, which could be drawn for example from the Flow of Funds.

3.5 Regulating Markets

Our analysis has so far focused on the regulation of certain financial actors or institutions. We classify such regulation as *identity-based* regulation. For example, this can correspond to regulation of a bank holding company. Financial regulation, however, can also apply directly to specific markets. In that case, we classify it as *activity-based* regulation. For example, the planner may directly regulate markets for specific forms of debt, such as the repo market. Such interventions would involve applying taxes uniformly across all financial actors that participate in that market, rather than applying specific (identity-based) taxes to different institutions across their activities. The distinction between identity- and activity-based regulation is not only of interest from a theoretical point of view but also highly relevant in practice.[20]

In this subsection, we characterize optimal regulation of markets, i.e., optimal activity-based regulation, and contrast it with our results on optimal identity-based regulation in Sections 3.1 through 3.3. To sharpen the analysis, we consider regulation of a single market m^* in the main text and defer a general characterization to Appendix B. Formally, we assume that the planner imposes uniform regulation $\hat{\tau}(m^*)$ on a single market, m^*, across all unregulated financial actors. In addition, we make two simplifying assumptions that are also relaxed in Appendix B: (i) there are no wealth effects for unregulated financial actors, that is $\nabla_{w_i} I_i = 0$ for $i \in \mathcal{S}$;[21] and, (ii) the price $p(m^*)$ does not appear in the constraint sets of unregulated financial actors. In this case, we obtain the following sharp characterization of optimal activity regulation.

[20]For example, Feldman and Heincecke (2018) emphasizes combining strengthened equity capital requirements for systemically important financial institutions (identity-based regulation) with a tax on the leverage of unregulated financial intermediaries (activity-based regulation).

[21]For example, absence of wealth effects can arise if unregulated financial actors have quasilinear utility $A_0^i + c_0^i + U_i(I_i)$, where A_0^i is a sufficiently large endowment of the numeraire consumption good so that nonnegativity $A_0^i + c_0^i \geq 0$ is slack. In this case, $\frac{\partial c_0^i}{\partial w_i} = 1$ and $\nabla_{w_i} I_i = 0$.

Proposition 7. *Suppose that wealth effects are zero for unregulated financial actors and that $p(m^*)$ does not appear in constraint sets of unregualted financial actors. Optimal regulation $\hat{\tau}^*(m^*)$ of market m^* is then given by*

$$\hat{\tau}^*(m^*) = \frac{\overbrace{\Xi_{\mathcal{S}}(m^*,\cdot)\tau^*}^{\text{Pecuniary Externalities}}}{\underbrace{\Xi_{\mathcal{S}}(m^*,m^*) + \Xi_{\mathcal{S}}(m^*,\cdot)\mathcal{M}\Xi_{\mathcal{H}^{-1}}\Xi_{\mathcal{S}}(\cdot,m^*)}_{\text{Unregulated Financial Actor Private Value}}} \tag{16}$$

where $\Xi_{\mathcal{S}}(m^*,\cdot)$ is the m^*-th row of $\Xi_{\mathcal{S}}$.

Proposition 7 states that optimal financial regulation of market m^* under *activity-based* regulation is determined by the ratio of two effects. The numerator of equation (16) captures the welfare gain that arises because regulation of activity m^* mitigates pecuniary externalities. This welfare gain results from the fact that all unregulated financial actors adjust their demand for good m^* in response to the tax $\hat{\tau}^*(m^*)$. This change in demand is measured by the m^*-th row of the aggregate unregulated demand response matrix, $\Xi_{\mathcal{S}}(m^*,\cdot)$. As discussed in Sections 3.1 and 3.2, the change in aggregate unregulated demand must be absorbed by nonfinancial actors, resulting in price changes. These price changes lead to the same regulatory arbitrage dynamic as in Proposition 3. Therefore, the externality consequences of these shifts in aggregate demand are measured by τ^*, i.e., the vector of taxes applied to regulated financial actors.

The denominator of equation (16) summarizes the private loss to unregulated financial actors that results from regulation of activity m^*. Since unregulated actors are optimizing, the private marginal benefit to any unregulated financial actor of higher demand in market m^* is equal to the tax rate $\hat{\tau}^*(m^*)$ charged for that activity. Across all other markets, the private marginal benefit is 0, as those markets are not taxed. The loss in equity value is therefore proportional to the total change in demand in market m^*. This total change in demand has two components. First, there is the direct change in demand $\Xi_{\mathcal{S}}(m^*,m^*)$ induced by the tax. Second, there are the indirect effects on demand that arise through price changes. These are given by the second term in the denominator, which states that the change in demand across all markets in response to the price, $\Xi_{\mathcal{S}}(m^*,\cdot)$, induces changes in prices (accounting for regulatory arbitrage) $dp = \Xi_{\mathcal{S}}(m^*,\cdot)\mathcal{M}\Xi_{\mathcal{H}}^{-1}$. These changes in prices induce changes in demand $dp\Xi_{\mathcal{S}}(\cdot,m^*)$, resulting in further changes in private equity value.

Optimal activity-based regulation—summarized in equation (16)—differs from optimal identity-based regulation—summarized in equation (11)—in important ways. Under identity-based regulation, the planner can tax all decisions made by the regulated financial actor i and consequently does not need to take into account that equilibrium price changes affect i's behavior. Instead, the planner can simply adjust the vector of taxes. By contrast, regulation of market m^* allows the planner to influence the behavior of all participants in that market, including all unregulated financial actors. However, the planner's influence is limited to a tax on that market. Thus, the planner must account for how unregulated financial actors shift among different activities in direct response to the tax, that is regulatory arbitrage between regulated and unregulated activities.

A particularly sharp and illustrative characterization emerges when there are zero cross-price elasticities in good and price m^*. In this case, the optimal tax formula becomes

$$\tau(m^*) = \left(1 - \alpha(m^*)\frac{\xi_{\mathcal{S}}(m^*, m^*)}{\xi_{\mathcal{H}}(m^*, m^*)}\right)\tau^*(m^*).$$

When the demand curve is downward sloping and the supply curve is upward sloping, we generally have $-\alpha(m^*)\frac{\xi_{\mathcal{S}}(m^*,m^*)}{\xi_{\mathcal{H}}(m^*,m^*)} > 0$. The optimal tax on activity m^* therefore increases in both the market share $\alpha(m^*)$ of unregulated financial actors and in the elasticity of unregulated financial actors relative to regulated financial actors, $\frac{\xi_{\mathcal{S}}(m^*,m^*)}{\xi_{\mathcal{H}}(m^*,m^*)}$. When the market share of unregulated financial actors is small, the optimal tax on market m^* is approximately $\tau^*(m^*)$, reflecting equal treatment. However, there is an additional advantage of activity-based regulation: It counteracts regulatory arbitrage in market m^*. Recall that $\alpha(m^*)\frac{\xi_{\mathcal{S}}(m^*,m^*)}{\xi_{\mathcal{H}}(m^*,m^*)}$ represents the regulatory arbitrage response of unregulated financial actors, which emerged as a key determinant in the optimal tax formula for regulated financial actors, $\tau^*(m)$, in Corollary 4—there, it appeared with opposite sign. When negative, this term therefore *dampens* optimal identity-based regulation but *amplifies* optimal activity-based regulation. The difference arises because under identity-based regulation, the regulator must take as given the regulatory arbitrage response of unregulated actors—which counteracts price changes induced by the regulator when $-\alpha(m^*)\frac{\xi_{\mathcal{S}}(m^*,m^*)}{\xi_{\mathcal{H}}(m^*,m^*)} > 0$—and hence a price increase reduces demand. By contrast under activity-based regulation, applying a tax to that activity reduces demand precisely in proportion to regulatory arbitrage. In other words, the activity-based tax is more effective when $-\alpha(m^*)\frac{\xi_{\mathcal{S}}(m^*,m^*)}{\xi_{\mathcal{H}}(m^*,m^*)}$ is large, and hence unregulated financial actors are more responsive to the tax.

4 A Classification Scheme for Unregulated Finance

Section 3 characterizes optimal financial regulation in the presence of unregulated financial actors, distinguishing between identity- and activity-based regulation. The unregulated financial sector contributes to pecuniary externalities through regulatory arbitrage. A natural question therefore is whether the planner should extend regulation to previously unregulated institutions and markets. In practice, policymakers and regulators have long debated the merits of, for instance, regulating shadow banks (see Section 5.1) or imposing controls on cross-border capital flows (see Section 5.2).

In this section, we develop a regulatory classification scheme for unregulated finance. The framework of Sections 2 and 3 allows us to sharply characterize the welfare gains from extending regulation to the unregulated financial sector. In particular, we derive a welfare-based metric that identifies the most valuable targets for regulation amongst previously unregulated actors (Section 4.1) and activities (Section 4.2).

The optimal tax formula of Proposition 3 clarifies that it is the effect of regulatory arbitrage on the efficacy of regulation that determines the implications of unregulated finance for the optimal taxation of regulated actors. It seems intuitive, therefore, that the value of extending regulation to unregulated actors should also depend on their contribution to regulatory arbitrage. The main contribution of this section, however, is to show these two policy questions meaningfully differ: The notion of regulatory arbitrage that we show governs the gains from extending regulation is different from that underlying the regulatory arbitrage multiplier. This is because our model assumes that the planner cannot *directly* relax constraints of financial actors by imposing revenue-neutral taxes. This is formally reflected in the fact that taxes affect the purchase price of goods, but not the collateral price. Intuitively, a tax on asset sales affects how much the bank can buy/sell assets for, but not how much the bank's creditor can get by seizing the asset and then selling it, which is what matters for the collateral constraint. This means that because regulation affects only the purchase price, the correct notion of regulatory arbitrage for the policy question of this section is simply the classical substitution effect, i.e. the response of an unregulated actor to a compensated increase in the purchase price, holding fixed the collateral price. In contrast, the relevant notion of regulatory arbitrage for Section 3 was the total derivative of Marshallian demand in both the purchase price and collateral price. These two measures align in general only when the extended regulation only applies to activities whose prices do not appear in constraints, and when there are no wealth effects. We discuss the empirical relevance of these two assumptions, and argue that they can be relevant in important settings.

Finally, the contribution of financial actors and markets to regulatory arbitrage cannot always be ordered uniformly. For example, two banks may engage in regulatory arbitrage in two different markets. From the regulator's perspective, it is not straightforward to identify which of the two banks is a more valuable target for regulation in this case. An important result of this section is that the vector τ^* of optimal financial taxes provides the relevant weights to add up regulatory arbitrage contributions across different markets. Our regulatory classification scheme therefore uses τ^* as weights to add up the welfare gains from extending regulation to different actors or markets.

4.1 Regulatory Classification of Unregulated Financial Actors

We first characterize the potential welfare gains from regulating a previously unregulated financial actor. When these potential welfare gains are large, we think of the financial actor in question as a valuable target for regulation. Formally, we consider the following exercise. Consider the equilibrium of Proposition 3 and suppose that the planner can impose a vector of taxes $d\tau$ on unregulated actor $i \in \mathcal{S}$, while holding fixed the activities of regulated financial actors. We first build intuition in a simple environment that we use as a benchmark to discuss the general case below.

4.1.1 Welfare Gains and Regulatory Classification in a Simple Benchmark

To gain intuition, we first discuss a special case. We assume that:

(A1) There are no wealth effects for unregulated financial actor i, that is $\nabla_{w_i} I_i = 0$.[22]

(A2) The proposed regulation $d\tau$ only taxes markets m whose prices do not appear in the constraint set of unregulated financial actor i, that is $d\tau(m) \neq 0$ only if $p(m)$ does not appear in Γ_i.

The special case of assumptions (A1) and (A2) is both empirically relevant and analytically insightful, and therefore deserves attention. Economically, (A1) states that an increase in i's wealth level does not affect i's trading behavior, I_i. Instead, changes in wealth are fully absorbed in a change of numeraire consumption, that is $\frac{\partial c_i}{\partial w_i} = 1$.[23] Under assumption (A2), we restrict the planner to extend new regulation only to activities whose prices are not directly associated with welfare-reducing pecuniary externalities. Assumptions (A1) and (A2) are empirically plausible when we think of the model as starting during an ex-ante investment period with deep credit markets. During the initial investment stage, financial actors have ample net worth and credit market access, and so net worth effects will be relatively muted (A1). Regulation is applied to ex-ante investment and capital structure decisions, rather than to ex-post asset sales and liquidations, whose prices do not appear in constraints (A2).[24]

The following result identifies the potential welfare gains from extending regulation to an unregulated financial actor i under Assumptions (A1) and (A2). We use this result as an important and instructive benchmark to discuss the general case below.

Proposition 8. *Under Assumptions (A1) and (A2), the wealth-equivalent welfare gain from introducing regulation $d\tau$ on financial actor $i \in \mathcal{S}$ is to first order given by*

$$\Delta_i^*(d\tau) = -\mu_i d\tau' \Xi_i \tau^* = -\mu_i \sum_{m'} \sum_m d\tau(m) \Xi_i(m, m') \tau^*(m'). \tag{17}$$

[22] It is useful to note that (A1) is needed because we assume the regulatory taxes $d\tau$ are revenue-neutral. If taxes were not revenue-neutral, e.g. residency based transaction taxes applied to foreign capital flows, they would induce wealth effects and we would not need (A1).

[23] A useful interpretation is that of a financial actor that has already achieved its optimal investment scale and simply pays out additional increases in net worth as dividends to shareholders.

[24] In the capital controls context, we might expect inflow taxes to satisfy (A2) but outflow taxes might not satisfy (A2).

Under (A1) and (A2), the welfare gains from extending financial regulation to previously unregulated intermediary i depend critically on the demand response matrix, Ξ_i. Reflecting unregulated actor i's behavioral response to price changes, Ξ_i determined i's relative contribution to the regulatory arbitrage multiplier in Proposition 3. If i has uniformly larger demand responses to price changes than i', then i is a bigger driver of regulatory arbitrage than i'. In this case, Proposition 8 also tells us that the welfare gains from extending regulation to i are larger than those of extending regulation to i'.[25]

Economically, the reason why the regulatory arbitrage measure Ξ_i is relevant is that it precisely reflects the behavioral responses that the planner is able to induce by extennding regulation. Regulation of financial actors induces price changes, which in turn generate regulatory arbitrage, reflected in the Marshallian demand responses Ξ_i. Similarly by extending regulatory taxes to i, under (A1) and (A2) these taxes produce precisely the same Marshallian demand responses as the price changes did in Section 3. In other words, Ξ_i reflects both regulatory arbitrage and responsiveness to new regulation. Proposition 8 therefore tells us that the greatest drivers of regulatory arbitrage are also the most responsive to new regulation, and therefore the most valuable targets for new regulation.[26]

Aggregation weights. When two financial actors cannot be uniformly ranked by regulatory arbitrage, Proposition 8 provides the welfare-relevant weights to sum up the two actors' contributions to regulatory arbitrage across different markets. In particular, $\Xi_i \tau^*$ captures i's welfare-weighted contribution to the regulatory arbitrage process. Consequently, a tax on good m changes welfare by $-\sum_{m'} \Xi_i(m,m')\tau^*(m)$, reflecting that i responds to the tax change by substituting activities across all markets. In particular, it states that the arbitrage into market m' is $\Xi_i(m,m')$, and that the welfare cost of that arbitrage is $-\tau^*(m)$, that is the negative of the optimal tax rate applied to regulated financial actors in market m. This means that the optimal taxes τ^* are also the relevant weights used to translate the contribution of i to regulatory arbitrage, Ξ_i, into the welfare gains from extending regulation, $\Delta_i^*(d\tau)$.

[25] Formally, uniformly larger responses means that $\Xi_i = (1+\kappa)\Xi_{i'}$ for $\kappa > 0$, in which case we have $\Delta_i^*(d\tau) = (1+\kappa)\Delta_{i'}^*(d\tau)$ (holding equal the measure of both types). This statement can be extended to differential measures by redefining Ξ_i to be scaled by its measure μ_i.

[26] It is important to note that Proposition 8 does not feature a welfare impact of the regulation on the private surplus of i. Since unregulated financial actor i was previously at a privately optimal allocation, the private cost to agent i under the new financial regulation $d\tau$ is only second order from a welfare perspective. This follows from a standard Envelope Theorem intuition: At an optimal allocation in which the private first-order conditions hold, the private welfare loss from a marginal change in activities is zero to first order. By contrast, the social welfare gain from mitigating pecuniary externalities is first order. This means that the first order welfare gains from extending regulation depend only on the social benefit of mitigating pecuniary externalities.

Estimable sufficient statistics. As in Section 3.4, we can decompose the overall welfare gain into elasticities and flows. In particular, we have

$$\Delta_i^*(d\tau) = -\mu_i \sum_{m'm} \frac{d\tau(m)}{p(m)} \xi_i(m,m')I_i(m')\tau^*(m').$$

For example, the gain from extending a 1% ad valorem tax to previously unregulated financial actor i, that is $d\tau(m) = (1+0.01)p(m)$, is $-\frac{1}{100}\sum_{m'} \xi_i(m,m')I_i(m')\tau^*(m)$ to first order. These gains depend on the same micro elasticities ξ_i and aggregate flows I_i that determined the contribution of unregulated financial actors to the regulatory arbitrage multiplier in Section 3.2. This is another way to see that under (A1) and (A2), the unregulated financial actors that are the main drivers of regulatory arbitrage thus also tend to be the most valuable targets for new regulation.

4.1.2 Welfare Gains and Regulatory Classification in the General Framework

Proposition 8 establishes assumptions (A1) and (A2) as sufficient conditions to ensure that the same measure Ξ_i of regulatory arbitrage emphasized in Section 3 remains the key determinant of the potential welfare gains from extending regulation to the unregulated financial sector. We now show, however, that the notion of regulatory arbitrage that determines the potential gains from extending regulation in the general environment is a slightly different one. This insight, summarized in the following proposition, is the main result of this section; importantly, it clarifies how the policy question of extending regulation differs from the question of setting optimal taxes discussed in Section 3.

Proposition 9. *To first order, the wealth-equivalent welfare gain from introducing regulation $d\tau$ on financial actor $i \in \mathcal{S}$ is*

$$\Delta_i^*(d\tau) = -\mu_i d\tau' H_i \tau^* = -\mu_i \sum_{m'} \sum_{m} \tau(m) H_i(m,m')\tau^*(m'), \tag{18}$$

where $H_i(m,m') = \frac{\partial I_i(m')}{\partial P(m)} + \frac{\partial I_i(m')}{\partial w_i} I_i(m)$ is the matrix of (classical) substitution effects in the purchase price, holding fixed collateral prices.

Proposition 9 generalizes the result of Proposition 8 by relaxing (A1) and (A2). The only difference is that the regulatory arbitrage measure Ξ_i used in Proposition 8 is now replaced by the measure H_i. Whereas Ξ_i measures regulatory arbitrage according to the entire price effect on demand, H_i measures regulatory arbitrage according to the classical substitution effect in the purchase price of a good.

The classical substitution effect is therefore the appropriate measure of regulatory arbitrage when assessing the gains from extending regulation. Measuring the behavioral demand responses to a wealth-compensated change in the purchase price holding fixed the collateral price, the matrix of classical substitution effects H_i is the relevant measure of regulatory arbitrage in this context: It is the purchase price—rather than the collateral price—that changes when the planner introduces a new tax on financial transactions.

Introducing regulation on i imposes a set of compensated (i.e. revenue neutral) taxes on the purchase price paid for goods by i. Whereas price increases are uncompensated and introduce wealth effects, compensated taxes remit the revenue to the same agent and do not introduce wealth effects. This highlights the importance of (A1) in Proposition 8: assuming that wealth effects on demand are zero means that compensated and uncompensated price increases have the same impact on demand.[27] In addition, introducing taxes on i alters the purchase price faced by i but not the collateral price, since the planner cannot directly relax the constraint set. In contrast, a price increase affects both the purchase price and the collateral price. This highlights the importance of (A2) in Proposition 8: assuming that the regulated activity prices do not appear in constraint sets means that collateral price effects are zero for regulated activities, and so the purchase price effect is the sole determinant of demand responses.

The combination of (A1) and (A2) gives rise to the role of the classical substitution effect as the measure of regulatory arbitrage for determining the value of extending regulation. Formally, we can see this through the relationship between H_i and Ξ_i. For simplicity we focus on this difference for the own-price terms (i.e. the diagonal entries), which is given by which is given by[28]

$$\Xi_i(m,m) = H_i(m,m) + \underbrace{\frac{\partial I_i(m)}{\partial q(m)}}_{\text{Collateral Price}} - \underbrace{\frac{\partial I_i(m)}{\partial w_i} I_i(m)}_{\text{Wealth Compensation}} .$$

Under assumptions (A1) and (A2), the latter two terms are zero. This means that $H_i = \Xi_i$, yielding Proposition 8. Without (A1) and (A2), in general we no longer have $H_i = \Xi_i$. Recall that the classical substitution effect is negative, that is $H_i(m,m) < 0$. First, consider the collateral price effect. It is often natural for the collateral price effect to be positive, that is $\frac{\partial I_i(m)}{\partial q(m)} > 0$. For example, in fire sale models with collateral constraints, the collateral price effect is positive because an increase in the collateral value of assets allows the bank to retain more assets on its balance sheet.

[27]Formally, regulatory arbitrage in response to a price increase includes the set of wealth effects that result because buyers have become relatively poorer and sellers have become relatively richer. Note that if instead taxes were not revenue neutral, this difference would disappear. In this sense, the key robust difference is (A2).

[28]Observe that the difference from equation (14) in Proposition 6 is that equation (14) expresses the collateral price effect decomposed into a substitution effect and income effect, whereas this equation combines the two into the total (Marshallian) collateral price effect.

Because the classical substitution effect is negative whereas the collateral price effect is plausibly positive, the collateral price effect tends to push $\Xi_i(m,m)$ to be above $H_i(m,m)$. Similarly, when m is a normal good the wealth effect is positive, that is $\frac{\partial I_i(m)}{\partial w_i} > 0$. This pushes $Xi_i(m,m)$ to be above $H_i(m,m)$ when i is a seller of good m, and below $H_i(m,m)$ when i is a buyer of good m. This is intuitive: a seller experiences a positive wealth effect, increasing demand, from the higher price, whereas a buyer experiences a negative wealth effect, reducing demand, from the higher price. These competing forces can therefore push in the same or opposite directions as the classical substitution effect. This gives rise to the difference between the classical substitution effect, H_i, and the total Marshallian demand response, Ξ_i.

Finally, the optimal taxes τ^* again serve as the welfare-relevant aggregation weights as in Proposition 8. This is because although the relevant notion of regulatory arbitrage has changed, the relevant mapping from regulatory arbitrage into welfare is precisely the same as before. In this case, the previous total mapping $\Xi_i\tau^*$ has now been replaced with the mapping $H_i\tau^*$. This reflects that Ξ_i captured the behavioral response of i to price changes, whereas H_i captured the behavior response of i to compensated changes in the purchase price alone.

Liquidity shock model revisited. The contrast between the two notions of regulatory arbitrage manifests starkly in the liquidity shock model of Section 3.3.1. While an extreme case, this simple model helps further clarify the difference between the two policy questions addressed in Sections 3 and 4.

Proposition 10. *In the liquidity shock model of Section 3.3.1, $\Delta_i^*(d\tau) = 0$ for all $i \in \mathcal{S}$.*

In Section 3, we emphasized that unregulated financial actors are important drivers of regulatory arbitrage in the liquidity shock model. And yet, there are no welfare gains from extending regulation to any unregulated financial actor in this model according to Proposition 10.

The fact that the classical substitution effect is the relevant measure of regulatory arbitrage for determining the value of extending regulation to unregulated financial actors, whereas the full effect is what matters for determining optimal regulation, is at the heart of this somewhat paradoxical result. In the liquidity shock model, unregulated financial actors had a fixed liquidity need or liquidity surplus, $-\rho_i$, to use on buying or selling assets at the interim date 2. This meant that they faced a constraint on flows given by $p(2)I_i(2) = -\rho_i$, where they had a liquidity need if $-\rho_i < 0$ and hence they were forced sellers, whereas they had a liquidity surplus if $-\rho_i > 0$ and hence they were buyers. This represented a constraint on these actors, that is it belongs to the set Γ_i of constraints on their activities. As a result, their position $I_i(2)$ depended entirely on the price in constraints, that is the collateral price, and not at all on the price in the budget constraint, that is the purchase price. As a result, the classical substitution effect of demand for $I_i(2)$ in the purchase price for 2 is in fact zero. As a result, there is no welfare gain from extending regulation to unregulated actors in the liquidity shock model—as such welfare gains would be determined by the classical substitution effect—even though these actors contribute to the regulatory arbitrage multiplier. By contrast, there was an effect of the collateral price on the position $I_i(2)$ of i through the constraint. As a result, the price change that drove regulatory arbitrage, which included changes the collateral price, induced a change in demand. In this case, the regulatory arbitrage response that impacts optimal regulation was driven entirely by changes in the collateral price.

4.1.3 Key Takeaways for Identity-Based Regulation

Sections 3 and 4 ask related but meaningfully different policy questions. Proposition 8 reveals that under (A1) and (A2), the same notion of regulatory arbitrage identified as being important for determining optimal regulation is also essential to determining the best targets for regulation. For example, this may be empirically plausible when thinking about extending ex ante regulatory measures, such as leverage requirements or inflow controls, in times when net worth effects are small. Proposition 8 also provides the way of comparing regulatory arbitrage responses across prices and markets for two different unregulated financial actors from a welfare perspective. In particular, the weights used are the optimal taxes τ^* applied to regulated financial actors. This weighted regulatory arbitrage metric classifies which are the most value targets for identity-based regulation.

Proposition 9 reveals that when considering applying regulations when wealth effects are large or to activities whose prices do appear in constraints, for example capital controls on outflows, it is important to define the measure of regulatory arbitrage in terms of only the classical substitution effect. Because financial regulation can only affect the purchase price and not the collateral price, the relevant notion of regulatory arbitrage that can be impacted by regulation is the classical substitution effect. Interestingly, this qualifies the notion that unregulated financial actors that are important for regulatory arbitrage are also important for extending these forms of regulation. In this case, they remain important targets for regulation to the extent that their regulatory arbitrage is driven by the classical substitution effect, and not by responses driven by changes in the collateral price. For example, an unregulated financial actor that voluntarily increases asset sales to capitalize on the resale price would be an important driver of regulatory arbitrage from both an optimal regulation perspective and from the perspective of being a valuable target for regulation of asset sales. By contrast, an unregulated financial actor that involuntarily increases asset sales in response to a collateral price decrease may not be a valuable target for *ex post* regulation, since their arbitrage response is not driven by the classical substitution effect. Naturally, they may still be a valuable target for ex ante regulation. Thus, this also tells us more not only about what institutions are most valuable targets for regulation, but also about what forms of regulation are most likely to be successful for what type of institutions.

4.2 Regulatory Classification of Unregulated Markets

We now present a regulatory classification scheme to identify unregulated markets (activities) that are valuable targets for regulation. We study the same exercise as in Section 4.1, except that the planner instead considers imposing a uniform tax $d\tau(m^*)$ on market m^* across all unregulated financial actors. The relevant notion of regulatory arbtirage again differs from that of Section 3 because regulation cannot affect collateral prices. At the same time, assumptions (A1) and (A2) again serve as sufficient conditions for the Marshallian responses Ξ to capture the correct notion of regulatory arbitrage. In the more general case, the relevant notion of regulatory arbitrage is again the classical substitution effect.

Proposition 11. *To first order, the wealth-equivalent welfare gain from introducing a uniform tax $d\tau(m^*)$ on all unregulated financial actors is,*

1. *Under Assumptions (A1) and (A2),*

$$\Delta^*(d\tau(m^*)) = -d\tau(m^*)\Xi_{\mathcal{S}}(m^*,\cdot)\tau^* = -\sum_{m'} d\tau(m^*)\Xi_{\mathcal{S}}(m^*,m')\tau^*(m'). \tag{19}$$

2. *In general,*

$$\Delta^*(d\tau(m^*)) = -d\tau(m^*)H_{\mathcal{S}}(m^*,\cdot)\tau^* = -\sum_{m'} d\tau(m^*)H_{\mathcal{S}}(m^*,m')\tau^*(m') \tag{20}$$

where $H_{\mathcal{S}} = \sum_{i\in\mathcal{S}} \mu_i H_i$

The intuition of Proposition 11 is close to that of Propositions 8 and 9. The welfare gain from inducing changes in aggregate demand of unregulated financial actors is $-\tau^*$ in wealth equivalent, so that τ^* is the relevant weighting used to measure the welfare implications of regulatory arbitrage. The key difference is how activities are changed by the regulation. In the case where a specific unregulated financial actor was targeted for regulation, regulatory arbitrage was measured by Ξ_i in the case of (A1) and (A2), and by H_i in general. By contrast, under market (activity) regulation, a single tax $d\tau(m^*)$ is applied to the same activity across all agents. As a result, the equilibrium response is the isolated effect of the change in that purchase price, but aggregated across all agents. This is therefore captured by $\Xi_{\mathcal{S}}(m^*,\cdot)$ when (A1) and (A2) hold, that is by regulatory arbitrage across all unregulated financial actors in response to the change in price m^*. Likewise absent (A1) and (A2), regulatory arbitrage is captured by the average classical substitution effect across unregulated financial actors, $H_{\mathcal{S}}$. Market regulation thus focuses around inducing changes in demand through a single purchase price across all agents, rather than inducing a change in demand in all prices for a single agent.

Estimable sufficient statistics. It is instructive to characterize the effect of a 1% ad valorem tax on market m^*, normalizing regulation by the price so that $d\tau(m^*) = \frac{1}{100}p(m^*)$. Under (A1) and (A2), the welfare gain is given by

$$\Delta^*(m^*) = -\frac{1}{100}\sum_{m'}\left[\sum_i \mu_i \xi_i(m^*, m')I_i(m')\right]\tau^*(m'). \tag{21}$$

The key determinants of the associated welfare gain are again the average elasticity and flow volume $\xi_i(m^*, m')I_i(m')$ across unregulated financial actors. A market is thus particularly valuable to regulate when changes in its price induces regulatory arbitrage into activities that regulated financial actors are regulated for, that is $\tau^*(m)$ is large, and when unregulated financial actors on average have large flow elasticities or flow volumes in those markets.

Activity Regulation as Implicit Discrimination. An interesting observation is that activity-based regulation act as a discriminatory tax against certain actors and business models, even though it in principle applies equally to all agents. The intuition is seen most starkly when a subset $\mathcal{S}(m^*)$ of unregulated actors do not participate in market m^*. In this case, the welfare gains from a uniform tax on market m^* is

$$\Delta^*(m^*) = -\frac{1}{100}\sum_{m'}\left[\sum_{i \notin \mathcal{S}(m^*)} \mu_i \xi_i(m^*, m')I_i(m')\right]\tau^*(m').$$

This tells us that the welfare consequences of regulation of market m^* are determined by the behavioral responses induced in the set participants in that market. This means that the uniform tax nevertheless implicitly discriminates against participants, who are affected by the tax, in favor of nonparticipants, who are not affected by it. We develop this idea further in our capital control application in Section 5.2.

4.3 Tradeoffs Between Different Instruments

The regulatory classification of Sections 4.1 and 4.2 identifies the potential welfare gains from extending regulation to previously unregulated institutions and markets. An important question that our classification framework helps address is what the relative trade-offs of identity- and activity-based regulation are, and when a planner should consider employing one over the other.

Formally, we can compare the value of extending identity-based regulation τ^* to unregulated financial actor $i \in \mathcal{S}$ to the value of extending activity-based regulation $\tau^*(m)$ to all unregulated financial actors. Under Assumptions (A1) and (A2), this difference is given by

$$\Delta_i^*(\tau^*) - \Delta^*(\tau^*(m^*)) = \underbrace{-\mu_i \sum_{m'} \sum_{m \neq m^*} \tau^*(m) \Xi_i(m,m') \tau^*(m')}_{\text{Multiple Activities Within Actor}} + \underbrace{\sum_{m'} \tau^*(m^*) \left[\sum_{j \neq i} \mu_j \Xi_j(m^*,m') \right] \tau^*(m')}_{\text{Single Activity Actross Actors}}.$$

The trade-off revolves around the ability to target many activities within a single actor versus the ability to target a single activity across actors. The first term reflects that the vector of taxes τ^* applied to i induces demand responses from changes in all prices in i. For example, this means it induces own price responses in every activity. By contrast, isolating to activity m^*, activity-based regulation induces demand responses across all other unregulated financial actors, and not only on i.

It is particularly instructive to write out this equation under two assumptions: there are only two goods, $M = 2$, and cross-price elasticities are zero. Under these two assumptions, denoting $m^* = 1$ without loss of generality, we have

$$\Delta_i^*(\tau^*) - \Delta^*(\tau^*(1)) = -\mu_i \Xi_i(2,2) \tau^*(2)^2 + \left[\sum_{j \neq i} \mu_j \Xi_j(1,1) \right] \tau^*(1)^2.$$

Concretely, define $\omega_1 = \frac{\tau^*(1)^2}{\tau^*(1)^2 + \tau^*(2)^2}$, which captures the importance of market 1 relative to the importance of market 2 from the perspective of the optimal tax. This tells us that identity-based regulation is valuable relative to activity-based regulation when

$$-\mu_i \Xi_i(2,2) \geq -\left[\sum_{j \neq i} \mu_j \Xi_j(1,1) \right] \left(\frac{\tau^*(1)}{\tau^*(2)} \right)^2.$$

This equation tells us that the behavioral response of i of demand for good 2 to the price of good 2 should be sufficiently high relative to the total behavioral response of all other actors of demand for good 1 to the price of good 1, when appropriately weighted by the relative taxes applied on the two goods. This gives a notion of importance or "centrality" of an actor i versus the importance of centrality of market 1. When a large behavioral elasticity can be induced in i across both markets 1 and 2, the value of identity-based regulation tends to rise, and we might think of i as central in the sense of being a large driver of flows. By contrast, when large behavioral elasticities can be induced across agents in market 1, the value of activity-based regulation tends to rise. In this case, we might think of market 1 as being central in the sense that it is a large driver of regulatory arbitrage. The relationship between these behavioral responses is driven by the relative taxes $\tau^*(1)/\tau^*(2)$ on the two activities, reflecting the social costs of regulatory arbitrage across the two different markets.

Non-regulatory instruments. In practice, proposals for interventions in unregulated finance include both regulatory interventions and fiscal interventions, such as access to the lender of last resort (LOLR). Such access looks to bolster rollover by allowing financial actors to borrow at rates consistent with "fundamental" value of assets, rather than temporarily low fire sale prices. In our language, this could be viewed as an intervention that boosts the collateral price. Following the logic of this section, the (Marshallian) measure $\frac{\partial I_i}{\partial q}$ of demand response to the collateral price is the right measure of "arbitrage" with regards to the collateral price, and the same vector $-\tau^*$ of weights is required to weight that arbitrage response. In other words, for example the value of a collateral price intervention for actor i would be $-dq'\frac{\partial I_i}{\partial q}\tau^*$. Notice that this accounts for the possiblity of moral hazard, that is the change in collateral price of m induces changes in demand for m', but has not accounted for the costs of the intervention that induces dq. In a world in which wealth effects are small (A1), a large regulatory arbitrage response Ξ_i indicates that there is either a large (classical) substitution effect or a large (Marshallian) collateral price effect (or both). In the former case, i is a valuable target for extension of regulation, whereas in the latter case i is a valuable target for extension of support programs that boost collateral prices. This suggests that more generally the weighted regulatory arbitrage response $\Xi_i\tau^*$ tells a regulator that there is value to some for of intervention for i. The results of this section can help provide additional guidance on whether that intervention should take the form of extension of regulation, extension of fiscal backstops, or both.

5 Applications

In this section, we apply our theory to two primary applications. In our first application in Section 5.1, we study how a planner should identify shadow banking institutions as targets for regulation. In particular, we identify characteristics of unregulated financial institutions, such as mutual funds or hedge funds, that make these institutions desirable targets for financial regulation. In our second application in Section 5.2, we study how a planner should target capital control measures to manage capital flows. We use this to evaluate what types of capital flows are most desirable to regulate.

5.1 Shadow Bank Institution Regulation

Our first application studies extending financial regulation to unregulated "shadow banking" institutions, such as mutual funds or hedge funds. We present a simple model in which shadow banks issue debt at date 0, but suffer a binding debt rollover constraint and forced deleveraging when the economy is in a recession. We study what properties make a shadow banking institution a particularly desirable target for financial regulation.

There are three periods, $t = 0, 1, 2$. An aggregate state $s \in \{s_H, s_L\}$ is realized at date 1, with the probability of the high state being π_H. There is one capital good which can be purchased and sold at dates 0 and 1, and we term purchases and sales of capital to be "investment." The economy features forced deleveraging and fire sales in the low state, s_L, but not in the high state, s_H, where the price is constant. The date 0 price of capital is also endogenous. We therefore refer to prices at date 0 as p_0, and we denote as p_1 the price vector at date 1 in the low state.

At date 0, shadow banks (unregulated financial actors) can finance a project by purchasing the capital good, I_0^i, at price p_0, where $I_0^i > 0$ denotes a purchase of the capital good. At date 0, shadow bank i can use the capital good to create $R(s)\Phi_i(I_0^i)$ units of the capital good at date 1, which then pay out 1 unit of the consumption good per unit of scale if held to maturity at date 2. $R(s)$ is a capital quality shock, with $R_H > R_L$. We normalize $\mathbb{E}[R] = 1$ for simplicity, since $\mathbb{E}[R] > 1$ can be folded into the technology Φ_i. Shadow banks can sell the capital good at date 1, denoted by I_1^i, where $I_1^i < 0$ denotes selling the capital good. The resale price in the low state is $p_1 \leq 1$, while the resale price in the high state s_H is constant at 1.

Shadow banks can also issue debt, D_0^i and D_1^i, and consume C_t^i. Shadow banks can trade the consumption good at date 0, c_0^i, to purchase the investment good. Debt is short-term and is traded with deep-pocketed risk-neutral households, and so has a fixed price of 1. Given that debt is short-term, the required debt level at date 0 is

$$D_0^i = c_0^i + p_0 I_0^i - w_i,$$

where w_i is the tradeable wealth level. This debt must be repaid at date 1 either by issuing new debt or liquidating assets. In the high state s_H, there is no constraint to debt rollover, and hence $D_i^1 = D_i^0$ and $C_2^i(s_H) = R_H \Phi_i(I_0^i) - D_i^0$ is final shadow bank consumption in the high state. However, in the low state shadow banks are not able to roll over debt, that is $D_1^i \leq 0$. As a result, in the low state debt repayment must be done using asset liquidations, $p_1 I_1^i = -D_0^i$. Hence, consumption in the low state is $C_2^i(s_L) = R_L \Phi_i(I_0^i) + I_1^i$, since the entire debt level is repaid at date 1 through asset liquidations. Substituting $p_1 I_1^i = -D_0^i$ in to preferences and the budget constraint, we obtain that the bank's object is to maximize

$$U_i = c_0^i + \Phi_i(I_0^i) + \left[\pi_H p_1 + \pi_L\right] I_1^i$$

subject to the budget constraint

$$c_0^i + p_0 I_0^i + p_1 I_1^i = w_i$$

and the non-negativity constraint $c_0^i \geq 0$. In the general notation, the two traded goods apart from the numeraire are date 0 capital I_0^i and date 1 low-state capital I_1^i. Date 1 capital sales are beneficial in that they relax the budget constraint, but are costly when sold at a price lower than 1. Notice that the interesting case arises when the non-negativity constraint binds, that is $c_0^i = 0$. We will assume this is the case throughout the remainder of this section.

From here, we obtain the following result.

Proposition 12. *The regulatory classification of shadow bank i is*

$$\Delta_i^* = -\xi_i I_0^i \Delta^*$$

where $\xi_i = \frac{p_0}{I_0^i}\frac{\partial I_0^i}{\partial p_0} = \frac{\Phi_i'(I_0^i)}{\Phi_i''(I_0^i)I_0^i}$ *is the elasticity of shadow bank investment at date* 0 *to the date* 0 *price, and* Δ^* *is a constant that does not depend on i and is defined in the proof.*

Proposition 12 provides an unambigous (relative) regulatory classification of shadow banks, which requires only minimal knowledge of a shadow bank's characteristics. According to this classification, shadow banks are institutions whose ex-ante investment has a high elasticity to the ex-ante price of investment, or that are associated with large aggregate flows I_0^i. Because shadow banks are debt-financed and face binding collateral constraints, large positive investment flows at date 0 are also associated with large negative flows at date 1, generating large externalities. Unregulated institutions with a large investment elasticity or large initial flows produce large demand responses to regulation, and hence also in equilibrium produce large responses in forced sales at date 1. This suggests that, from a regulatory perspective, shadow banking institutions can be classified based on their investment price elasticity and aggregate flows.

The special case of Cobb-Douglas production yields a particularly sharp classification formula for the welfare benefits of regulating shadow bank i. Let $\Phi_i(I_0^i) = A_i(I_0^i)^{\alpha_i}$, where we can interpret each bank as having a fixed factor of "bank labor" with supply 1 and factor share $1-\alpha_i$. In the Cobb-Douglas case, we have $-\xi_i = \frac{1}{1-\alpha_i}$. Therefore, our results suggest that extending regulation to shadow banks can generate particularly large welfare gains when these previously unregulated institutions have (i) a high level of illiquid investment, I_0^i, and (ii) a large illiquid investment factor share, α_i.

5.2 International Capital Flow Regulation

Our second application studies regulation of international capital flows to a small open economy (SOE), such as an emerging market. We present a simple model of capital inflows by international investors. The SOE can experience a crisis at an intermediate date, which may result in a sudden stop or capital flight from different investors. We study the impact of unregulated capital flows on optimal regulation by the SOE planner, as well as the potential welfare gains the SOE can realize when imposing restrictions on initial inflows or on outflows during the crisis.

There are three periods, $t = 0, 1, 2$. The SOE faces aggregate uncertainty that is realized at date 1, with $s \in \{s_H, s_L\}$. The probability of the crisis state, s_L, is denoted $1 - \pi_H$. The economy has N domestic capital goods which can be purchased and sold at both date 0 and date 1. We denote prices in period 0 as $p_0(n)$. At date 1, the price of capital good n is denoted $p_1(n)$ if the economy is in the crisis state, $s = s_L$, and it is constant and normalized to 1 if the SOE does not experience a crisis, $s = s_H$. We denote p_0 and p_1 to be the vectors of date 0 and date 1 prices.[29]

At date 0, international investor i (i.e., unregulated financial actor i) can purchase a vector I_0^i of domestic capital goods, with $I_0^i(n)$ denoting purchases of good n. If the high state is realized at date 1, international investor i earns a high payoff $F_H^i(I_0^i)$ from their investment in units of the consumption good, at which point their project ends. If instead the crisis state is realized at date 1, the project yields nothing at date 1 and a lower final value $F_L^i(I_0^i, I_1^i)$ at date 2. This fall in project value can be interpreted as arising due to a negative fundamental shock in the SOE or from stochastic movements in real exchange rates.[30] I_1^i is the endogenous vector of date 1 flows into or out of domestic capital goods during the crisis state, with $I_1^i(n)$ denoting flows from good n.

International investors are deep-pocketed at date 0 and can therefore finance investment by setting $c_0^i < 0$, which makes it convenient to fold consumption at date 0 into the wealth level. The utility of international investors from their investments in the SOE can then be written as

$$\pi_H F_H^i(I_0^i) + (1 - \pi_H)\left(\lambda_1^i c_1^i + F_L^i(I_0^i, I_1^i)\right).$$

We denote by λ_1^i the marginal value of repatriated wealth at date 1. It may be larger than 1, for example if international investors experience a binding collateral constraint in their home country or if there is a movement in the real exchange rate. International investors may find it desirable to sell domestic capital goods in the low state if they have a high marginal value of wealth or if they can earn higher returns by investing abroad rather than by continuing the project in the SOE.

The budget constraint of international investor i in the SOE is

$$c_1^i + \hat{p}_0 I_0^i + p_1 I_1^i \leq \hat{w}^i$$

[29]Note that in the general notation, the index m corresponds to pairs (t, n). Thus we can index $(0, 1)$ by $m = 1$, $(0, n)$ by $m = n$, $(1, 1)$ by $m = n + 1$, and so on. It is expositionally clearer in this example to maintain pair dependence (t, n) as opposed to the index m.

[30]For example, we could assume that the domestic projects pay off in the domestic consumption good, and that foreign investors sell the domestic consumption good to purchase the foreign consumption good. We can capture this by premultiplying the project payoff F_L^i by the real exchange rate ε_L in the low state.

where $\hat{p}_0 = \frac{p_0}{1-\pi_H}$ denotes probability-normalized prices, and similarly for wealth, and where we have adopted inner product notation $\hat{p}_0 I_0^i = \sum_n \hat{p}_0(n) I_0^i(n)$. Given deep pockets, we can interpret $\hat{w}^i$ as the amount of wealth that international investor i allocates for investment in the SOE.

To obtain sharp results, we assume that investment technologies are separable across goods for an investor. Formally, this means that $F_H^i(I_0^i) = \sum_n F_H(I_0^i(n), n)$ and $F_L^i(I_0^i, I_1^i) = \sum_n F_L^i(I_0^i(n), I_1^i(n), n)$. In this case, cross-price elasticities between different goods are zero. But the cross-price elasticities between inflows and outflows of the same good are not zero.

We define two useful concepts. The first is investor i's tendency for "flight" from capital good n, which we define as

$$\omega^i(n) \equiv -\frac{\partial I_1^i(n)/\partial p_0(n)}{\partial I_0^i(n)/\partial p_0(n)}. \tag{22}$$

Intuitively, $\omega^i(n)$ measures the fraction of a new inflow $dI_0^i(n)$ at date 1 that ends up withdrawing from the SOE as an outflow at date 1. It is natural for $\omega^i(n) \geq 0$ when an increase in inflows is associated with an increase in outflows. Note that the negative sign on the right hand side of equation (22) appears because an increase in inflows in our model is a more positive value of $I_0^i(n)$, whereas an increase in outflows is a more negative value of $I_1^i(n)$.

We also define investor i's tendency for "retrenchment" from capital good n as

$$\zeta^i(n) \equiv -\frac{\partial I_0^i(n)/\partial p_1(n)}{\partial I_1^i(n)/\partial p_1(n)}. \tag{23}$$

Intuitively, $\zeta^i(n)$ measures how much an increase $dI_1^i(n)$ in outflows at date 1 results in an increase in inflows $dI_0^i(n)$ at date 0. When $\zeta^i(n)$ is large, investor i increases initial invest in the SOE when an increase in the date 1 price also leads her to withdraw more capital at date 1.

Investor flight $\omega^i(n)$ and investor retrenchment $\zeta^i(n)$ are closely related notions but capture distinct ideas. An investor who can realize a large project payoff in the high state but is almost indifferent between maintaining the project and fleeing in the low state might have a large tendency for flight, $\omega^i(n)$. By contrast, that same investor would likely have a low tendency for retrenchment, $\zeta^i(n)$. This is because her near indifference between maintaining and fleeing in the low state suggests a large outflow elasticity to the price of outflows, that is $\frac{\partial I_1^i(n)}{\partial p_1(n)}$ is large. However, near indifference also means that switching from maintaining investment to retrenching at date 1 likely has little impact on the value she gets from investment, that is $\frac{\partial I_0^i(n)}{\partial p_1(n)}$ is low. Put together, this means that retrenchment $\zeta^i(n)$ is low.

Aggregate unregulated demand response. Given that we have written the model without cross-price elasticities across goods, the demand response matrix Ξ_i of international investor i can be written as the block diagonal matrix of demand responses $\Xi_i(n)$. That is, we have

$$\Xi_i(n) = \begin{pmatrix} \frac{\partial I_0^i(n)}{\partial p_0(n)} & -\omega^i(n)\frac{\partial I_0^i(n)}{\partial p_0(n)} \\ -\zeta^i(n)\frac{\partial I_1^i(n)}{\partial p_1(n)} & \frac{\partial I_1^i(n)}{\partial p_1(n)} \end{pmatrix}$$

This represents off-diagonal elements of $\Xi_i(n)$ as the product between the own-price responses of inflows and outflows and the measures of flight and retrenchment identified. Summing over i, we obtain the aggregate unregulated demand response matrix given by

$$\Xi_{\mathcal{S}}(n) = \begin{pmatrix} \frac{\partial I_0^{\mathcal{S}}(n)}{\partial p_0(n)} & -\omega^{\mathcal{S}}(n)\frac{\partial I_0^{\mathcal{S}}(n)}{\partial p_0(n)} \\ -\zeta^{\mathcal{S}}(n)\frac{\partial I_1^{\mathcal{S}}(n)}{\partial p_1(n)} & \frac{\partial I_1^{\mathcal{S}}(n)}{\partial p_1(n)} \end{pmatrix}$$

where $\omega^{\mathcal{S}}(n) = \sum_i \beta_0^i(n)\omega^i(n)$ is the average flight across international investors and $\zeta^{\mathcal{S}}(n) = \sum_i \beta_1^i(n)\zeta^i(n)$ the average retrenchment. The weight $\beta_0^i(n) = \frac{\mu_i \xi_0^i(n)\alpha_0^i(n)}{\sum_j \mu_j \xi_0^j(n)\alpha_0^j(n)}$ reflects the relative inflow elasticity of i in good n weighted by i's share of inflows, and similarly where $\beta_1^i(n) = \frac{\mu_i \xi_1^i(n)\alpha_1^i(n)}{\sum_j \mu_j \xi_1^j(n)\alpha_1^j(n)}$ is the analogous weighting measure for outflows.[31] Therefore, aggregate flight $\omega^{\mathcal{S}}(n)$ is high when flighty investors with high $\omega^i(n)$ also have high inflow elasticities and high market shares of inflows. Similarly, aggregate retrenchment $\omega^{\mathcal{S}}(n)$ is high when retrenching investors with high $\zeta^i(n)$ have high outflow elasticities and high market shares of outflows.

Aggregate supply response and taxes. We assume that a representative capital producing firm operates a separable technology to produce capital goods goods at date 0, with constant elasticity $\xi_0^{\mathcal{H}}(n)$ for good n.[32] Similarly, a representative capital deconstructing firm operates a separable technology to convert capital goods into the consumption goods at date 1, with constant elasticity $\xi_1^{\mathcal{H}}(n)$. Because technologies are separable, all cross price elasticities are zero, and $\Xi_{\mathcal{H}}$ is the block diagonal matrix with

$$\Xi_{\mathcal{H}}(n) = \begin{pmatrix} \frac{1}{p_0(n)}\xi_0^{\mathcal{H}}(n)I_0^{\mathcal{H}}(n) & 0 \\ 0 & \frac{1}{p_1(n)}\xi_1^{\mathcal{H}}(n)I_1^{\mathcal{H}}(n) \end{pmatrix}$$

Finally, we take as a primitive the vector τ of taxes on goods, which is a stacking of $\tau(n) = \begin{pmatrix} \tau_0(n) \\ \tau_1(n) \end{pmatrix}$. It is, however, natural to have $\tau_1(n) < 0$ in models of pecuniary externalities, since this

[31] Observe the parallels to domar aggregation (e.g., Baqaee and Farhi 2019).

[32] Formally, this arises from the maximization problem with objective function $\sum_n \left[p_0(n)I_0^{\mathcal{H}}(n) - \Phi_i(I_0^{\mathcal{H}}(n), n)\right]$ where the cost function $\Phi_i(I_0^{\mathcal{H}}(n), n))$ yields constant elasticity.

corresponds to subsidizing regulated financial actors to maintain their investments at date 1.

Example: safe and flighty investors. Suppose that there are only two types of investors: fully safe and fully flighty. Fully safe investors inelastically set $I_1^s = 0$, while fully flighty investors inelastically set $I_1^f = -I_0^i$. Their measures are μ^s and μ^f, respectively. In this case, we have $\omega^s = \zeta^s = 0$ and $\omega^f = \zeta^f = 1$. Moreover, we have $\frac{\partial I_1^s}{\partial p_1} = 0$ and $\frac{\partial I_1^s}{\partial p_1} = -\frac{\partial I_0^s}{\partial p_1} = \frac{\partial I_0^s}{\partial \hat{p}_0}$.[33] Therefore, we have $\omega^{\mathcal{S}}(n) = \zeta^{\mathcal{S}}(n) = \frac{\mu^f \xi_0^f(n)\alpha_0^f(n)}{\mu^f \xi_0^f(n)\alpha_0^f(n)+\mu^s\xi_0^s(n)\alpha_0^s(n)}$, which is the elasticity-weighted share of capital flows of flighty investors. In the limiting case where both types of investors have the same inflow elasticities $\xi_0^s = \xi_0^f = \xi_0^{\mathcal{S}}$, then we have $\omega^{\mathcal{S}} = \zeta^{\mathcal{S}} = \frac{\mu^f \alpha^f(n)}{\mu^f \alpha^f(n)+\mu^s\alpha^s(n)}$ is the share of flows of flighty investors relative to safe investors. Finally, observe that we have $\frac{\partial I_0^{\mathcal{S}}}{\partial p_0(n)} = \mu^s \frac{\partial I_0^s}{\partial p_0(n)} + \mu^f \frac{\partial I_0^f}{\partial p_0(n)}$ but $\frac{\partial I_1^{\mathcal{S}}}{\partial p_1(n)} = \mu^f \frac{\partial I_0^f}{\partial p_0(n)}$, reflecting that only flighty investors engage in outflows.

5.2.1 Optimal Financial Regulation with Unregulated Capital Flows

From here, we can characterize the impact of unregulated capital flows on optimal financial regulation. Because both $\Xi_{\mathcal{S}}$ and $\Xi_{\mathcal{H}}$ are block diagonal, $\mathcal{M}$ is also block diagonal. We then obtain the following result.

Proposition 13. *Optimal regulation with unregulated capital flows is*

$$\tau_0^*(n) = \overbrace{\frac{1}{1+\delta_0(n)}\frac{1}{1-\delta_{01}(n)}}^{\text{Inflow Regulatory Arbitrage}}\left[\tau_0(n) + \overbrace{\delta_0(n)\omega^{\mathcal{S}}\frac{1}{1+\delta_1(n)}}^{\text{Flight}}\tau_1(n)\right]$$
$$\tau_1^*(n) = \underbrace{\frac{1}{1+\delta_1(n)}\frac{1}{1-\delta_{01}(n)}}_{\text{Outflow Regulatory Arbitrage}}\left[\underbrace{\delta_1(n)\zeta^{\mathcal{S}}(n)\frac{1}{1+\delta_0(n)}}_{\text{Retrenchment}}\tau_0(n)+\tau_1(n)\right]$$

where $\delta_t(n) = -\alpha_t^{\mathcal{S}}(n)\frac{\xi_t^{\mathcal{S}}(n)}{\xi_t^{\mathcal{H}}(n)} \geq 0$ *reflects own price regulatory arbitrage, and where* $\delta_{01}(n) = \omega^{\mathcal{S}}(n)\zeta^{\mathcal{S}}(n)\frac{\delta_0(n)\delta_1(n)}{(1+\delta_0(n))(1+\delta_1(n))}$ *reflects cross-price substitution.*

Although Proposition 13 appears complicated, it is in fact intuitive. We begin by discussing the optimal tax at date 0, $\tau_0^*(n)$. Applying a positive tax at date 0 reduces demand and increases the price $p_0(n)$. This price leads to regulatory arbitrage in the form of inflows, which is captured by $\frac{1}{1+\delta_0(n)}$. This regulatory arbitrage is large when $\delta_0(n)$ is large, that is when the elasticity of

[33] This means that we have $\Xi_s(n) = \begin{pmatrix} \frac{\partial I_0^i(n)}{\partial p_0(n)} & 0 \\ 0 & 0 \end{pmatrix}$ and $\Xi_f(n) = \begin{pmatrix} \frac{\partial I_0^i(n)}{\partial p_0(n)} & -\frac{\partial I_0^i(n)}{\partial p_0(n)} \\ \frac{\partial I_0^i(n)}{\partial p_0(n)} & \frac{\partial I_0^i(n)}{\partial p_0(n)} \end{pmatrix}$.

unregulated inflows is high relative to the supply elasticity, or when the unregulated inflow share of the date 0 market is large. This direct regulatory arbitrage dampens the magnitude of the optimal tax, as the increase in inflows counteracts the decrease in demand from regulated actors.

When $\omega^{\mathbb{S}} = \zeta^{\mathbb{S}} = 0$, that is there is no flight or retrenchment, this direct effect is the only additional component of the optimal tax formula. In this case, we have $\tau_t^*(n) = \frac{1}{1+\delta_t(n)}\tau_t(n)$ and so the magnitude of both taxes are dampened. However, when there is flight and retrenchment, $\omega^{\mathbb{S}}\zeta^{\mathbb{S}} > 0$, this is not the entire effect. First, there is a counterveiling effect, measured by $\frac{1}{1-\delta_{01}(n)} > 1$, which partially offsets the direct effect. This counterveiling effect arises because regulatory arbitrage in the form of inflows leads to outflows due to investor flight. The increase in outflows requires a drop in the date 1 price to equilibrate markets (since the capital deconstructing firm is on the demand side at date 1), which in turn fuels a decrease in inflows due to retrenchment. This partially offsets regulatory arbitrage at date 0. In the limit where In general it does not fully offset the initial effect as long as $\omega^{\mathbb{S}}\zeta^{\mathbb{S}} < 1$, so that the regulatory arbitrage process on net dampens the magnitude of tax rates. Thus surprisingly, flight and retrenchment indirectly counteract the dampening effect of regulatory arbitrage on optimal tax rates.

The regulatory arbitrage term $\frac{1}{1+\delta_t(n)}\frac{1}{1-\delta_{01}}$ reflects the total change in inflows (outflows) that arises in equilibrium due to the regulatory arbitrage process. Thus, this term scales the original tax rate $\tau_t(n)$ applied. In addition, the equilibrium changes in inflows from regulatory arbitrage also generates outflows via flight, while the equilibrium change in outflows from regulatory arbitrage generates inflows via retrenchment.

The effect of inflow regulatory arbitrage on outflows is captured by the "Flight" term in the optimal tax formula for $\tau_0^*(n)$. Intuitively, an increase in the date 0 tax increases outflows via flight. This is captured by the product of the inflow effect, $\delta_0(n)$, multiplied by the average flight among international investors, $\omega^{\mathbb{S}}(n)$. Outflows are themselves associated with a fixed point regulatory arbitrage process, captured by the rescaling $\frac{1}{1+\delta_1(n)}$ of the initial change, that is more elastic outflows mean that flight can only generate small equilibrium price changes. Finally, this total change in outflows is multiplied by the $\tau_1(n)$. If $\tau_1(n) < 0$ and so outflows are taxed (i.e. retained assets are subsidized), then this total effect promotes a *lower* date 0 tax. Intuitively, this is because the higher tax promotes regulatory arbitrage by flighty international investors, which then results in flight and destructive outflows. Therefore, an inflow tax becomes less desirable.

Conversely, there is an equivalent "Retrenchment" term in the optimal date 1 tax formula. A higher date 1 subsidy (i.e. tax on liquidations) promotes a higher date 1 price, which leads to retrenchment by international investors and higher inflows. This retrenchment is measured by $\delta_1(n)\zeta^{\mathbb{S}}(n)$, and produces through regulatory arbitrage a change in price proportional to $\frac{1}{1+\delta_0(n)}$, which is dampened the more elastic the inflow regulatory arbitrage response is. If inflows are taxed, i.e., $\tau_0(n) > 0$, this this reduces the effectiveness of the outflow tax and calls for a lower tax on outflows. If inflows are subsidized, i.e., $\tau_0(n) < 0$, on the other hand, then this effect calls for a larger outflow tax.

Interestingly, if $\tau_0(n)$ and $\tau_1(n)$ share the same sign, then both the flight and retrenchment terms enhance the effectiveness of the original regulation. For example, suppose that both are negative. Then, the inflow subsidy raises the date 0 price, which discourages foreign inflows and reduces flight. The reduction in flight reduces outflows and so raises the date 1 price. Similarly, the outflow tax raises the date 1 price, which increases retrenchment and promotes beneficial foreign inflows.

5.2.2 Value of Imposing Capital Controls

In the following, we characterize the first order welfare gains that the SOE planner can achieve by imposing uniform taxes on capital inflows and outflows. We then leverage these results in two leading examples to study the benefits of taxing different forms of capital inflows, and to study the benefits of taxes on inflows versus taxes on outflows. In particular, let $\tau_0(n)$ denote a uniform tax on inflows into good n at date 0 and $\tau_1(n)$ a uniform tax on outflows at date 1, so that the after-tax prices faced by international investors are $p_0 + \tau_0$ and $p_1 + \tau_1$. Following Section 4, we define $\Delta_0^*(n)$ as the first-order money-metric welfare gain to the SOE from imposing a uniform 1% ad valorem inflow tax on good n across international investors, that is $\tau_0(n) = \frac{1}{100}p_0(n)$. The money-metric measure is the date 0 change in wealth that would yield the same welfare as imposing the inflow tax. Similarly, we define $\Delta_1^*(n)$ as the money-metric welfare gain from imposing a uniform 1% ad valorem outflow tax on good n across investors. We characterize these welfare gains in the following proposition.

Proposition 14. *The money-metric first-order welfare gains for capital controls are:*

1. *For a 1% tax on inflows into good n,*

$$\Delta_0^*(n) = -\frac{1}{100}\xi_0^{\mathcal{S}}(n)I_0^{\mathcal{S}}(n)\left[\tau_0^*(n) - \omega^{\mathcal{S}}(n)\tau_1^*(n)\right] \tag{24}$$

2. *For a 1% tax on outflows from good n,*

$$\Delta_1^*(n) = -\frac{1}{100}\xi_1^{\mathcal{S}}(n)I_1^{\mathcal{S}}(n)\left[\zeta^{\mathcal{S}}(n)\tau_0^*(n) - \tau_1^*(n)\right] \tag{25}$$

Proposition 14 shows that the value of applying an inflow control depends on the flightiness of investors, $\omega^{\mathcal{S}}(n)$. Given $-\xi_0^{\mathcal{S}}(n)I_0^{\mathcal{S}}(n) > 0$, that is the capital inflow demand curve is downward sloping, then a tax on inflows generates welfare gains both when date 0 investment is taxed for regulated financial actors, $\tau_0^*(n) > 0$, and when date 1 asset sales are taxed (i.e. retaining assets is subsidized) for regulated financial actors, that is $\tau_0^*(n) < 0$. The strength of the latter effect is proportional to the flightiness of capital inflows.

On the other hand, it shows that the value of an outflow tax depends on the retrenchment of investors. Given $-\xi_0^{\mathcal{S}}(n)I_0^{\mathcal{S}}(n) > 0$, that is supply curves are upwards sloping, then an outflow tax is directly valuable when $-\tau_1^*(n) > 0$, that is asset sales are taxed among regulated financial actors. Moreover if $\tau_0^*(n) > 0$, then there is a second indirect benefit that arises due to retrenchment. When retrenchment $\zeta^{\mathcal{S}}(n)$ is high, the tax on outflows also discourages capital inflows. Conversely, if inflows are valuable, $\tau_0^*(n) < 0$, then the value of outflow regulation is dampened by the retrenchment, since restricting outflows reduces the desirability of initial inflows.

Inflow versus outflow taxes. Proposition 14 highlights an important difference between the efficacies of inflow and outflow regulation. Consider the extreme case where $\xi_0^{\mathcal{S}}(n)I_0^{\mathcal{S}}(n) = \xi_1^{\mathcal{S}}(n)I_1^{\mathcal{S}}(n)$, then we have

$$\Delta_1^*(n) - \Delta_0^*(n) = \frac{-\xi_0^{\mathcal{S}}(n)I_0^{\mathcal{S}}(n)}{100}\left[(1-\zeta^{\mathcal{S}}(n))\tau_0^*(n) + (1-\omega^{\mathcal{S}}(n))\tau_1^*(n)\right]$$

Suppose first that $\tau_0^*(n) \leq 0$, that is capital inflows are valuable to the SOE. In this case, the relative value of the inflow control increases in both flightiness and retrenchment. This is because the outflow control discourages retrenching investors from investing in the begin with, leading to a sharper reduction in beneficial inflows. By contrast if $\tau_0^*(n) \geq 0$, then the relative value of the inflow control increases in flightiness but decreases in retrenchment.

Notably, this also makes clear that outflow regulation can be valuable even when inflow regulation is not. The reason is due to differential targeting of the two instruments. Suppose that both flightiness and retrenchment are low, and that $\tau_0^*(n) \leq 0$. In this case, the inflow regulation is negative value while the outflow regulation has positive value. The difference arises due to differential targeting of the two instruments. The inflow tax targets all investors, including ones that are not flighty, meaning that the SOE must suffer a large contraction in inflows to generate a small reduction in outflows when flightiness is low. By contrast, the outflow tax only harms retrenching investors. When retrenching investors are a small fraction of the population, the outflow tax allows effectively discriminates against retrenching investors in favor of nonretrenching investors. This highlights an important function of activity regulation: it can be used to implicitly discriminate between different types of unregulated actors without applying explicitly discriminatory taxes. If an SOE cannot identify what investors will be flighty/retrenching ex ante, the outflow tax provides an incentive compatible method of differentially screening out risky investors in favor of safe investors.[34]

Safe and flighty investors revisited. Suppose we revisit the case of only fully safe or fully flighty investors with the same elasticities. It is easy to see here that outflow regulation only affects flighty investors, since they are the sole drivers of date 1 flow elasticities. By contrast, inflow regulation discourages outflows by flighty investors in exactly the same proportion, since $I_0^f = -I_1^f$, but also discourages inflows by safe investors. This is the limiting case of the differential highlighted above, and highlights the advantage of outflow regulation as a method of screening out flighty investors in favor of safe investors.

6 Conclusion

We study optimal regulation when there are pecuniary externalities, but there is regulatory arbitrage by the unregulated financial sector. Optimal regulation is scaled by a regulatory arbitrage multiplier. The contribution of an unregulated financial institution or actor can be determined by a combination of microeconomic price elasticities and aggregate flow volumes, which are in principle estimable in the data. Interestingly, we show that regulatory arbitrage can naturally amplify the effectiveness of financial regulation in economies with pecuniary externalities.

Our framework contributes a regulatory classification of the institutions, activities, and capital flows that should be identified by regulators and policymakers as the most valuable targets for regulation. This classification is based on in principle estimable sufficient statistics, and did not rely on knowledge of the structure of unregulated finance. Our results can help provide guidance to policymakers for thinking about how to identify targets for regulation in a complex and heterogeneous financial system.

References

Acharya, V. V., P. Schnabl, and G. Suarez (2013, mar). Securitization without risk transfer. *Journal of Financial Economics 107*(3), 515–536.

Allen, F. and D. Gale (1994). Limited market participation and volatility of asset prices. *The American Economic Review 84*(4), 933–955.

Avdjiev, S., B. Hardy, S. Kalemli-Ozcan, and L. Servén (2018, jan). Gross Capital Flows by Banks, Corporates and Sovereigns.

[34] This idea is related to Davila (2021), who argues that short-term financial transaction taxes discriminate against speculative investors.

Baqaee, D. R. and E. Farhi (2019). The Macroeconomic Impact of Microeconomic Shocks: Beyond Hulten's Theorem. *Econometrica 87*(4), 1155–1203.

Begenau, J. and T. Landvoigt (2020). Financial Regulation in a Quantitative Model of the Modern Banking System.

Bengui, J. and J. Bianchi (2021). Macroprudential Policy with Leakages.

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

Bianchi, J. (2016, dec). Efficient bailouts? *American Economic Review 106*(12), 3607–3659.

Bianchi, J. and G. Lorenzoni (2021). The Prudential Use of Capital Controls and Foreign Currency Reserves. *SSRN Electronic Journal*.

Bianchi, J. and E. G. Mendoza (2018, apr). Optimal Time-Consistent Macroprudential Policy. *Journal of Political Economy 126*(2), 588–634.

Broner, F., T. Didier, A. Erce, and S. L. Schmukler (2013, jan). Gross capital flows: Dynamics and crises. *Journal of Monetary Economics 60*(1), 113–133.

Buchak, G., G. Matvos, T. Piskorski, and A. Seru (2018, dec). Fintech, regulatory arbitrage, and the rise of shadow banks. *Journal of Financial Economics 130*(3), 453–483.

Caballero, R. J., E. Farhi, and P. O. Gourinchas (2008, feb). An equilibrium model of global imbalances and low interest rates. *American Economic Review 98*(1), 358–393.

Caballero, R. J. and A. Krishnamurthy (2001, dec). International and domestic collateral constraints in a model of emerging market crises. *Journal of Monetary Economics 48*(3), 513–548.

Caballero, R. J. and A. Simsek (2020, dec). A Model of Fickle Capital Flows and Retrenchment. *Journal of Political Economy 128*(6), 2288–2328.

Calvo, G. A. (1988). Servicing the Public Debt: The Role of Expectations. *American Economic Review 78*(4), 647–661.

Chari, V. V. and P. J. Kehoe (2016, sep). Bailouts, Time Inconsistency, and Optimal Regulation: A Macroeconomic View. *American Economic Review 106*(9), 2458–2493.

Chretien, E. and V. Lyonnet (2020, may). Are Traditional and Shadow Banks Symbiotic?

Claessens, S., Z. Pozsar, L. Ratnovski, and M. Singh (2012). Shadow Banking: Economics and Policy. *IMF Staff Discussion Note*.

Clayton, C. and A. Schaab (2022). Multinational Banks and Financial Stability. *Quarterly Journal of Economics. Forthcoming*.

Coppola, A. (2021). In Safe Hands: The Financial and Real Impact of Investor Composition Over the Credit Cycle.

Coppola, A., M. Maggiori, B. Neiman, and J. Schreger (2021). Redrawing the Map of Global Capital Flows: The Role of Cross-Border Financing and Tax Havens. *Quarterly Journal of Economics 136*(3), 1499–1556.

Davila, E. (2021). Optimal Financial Transaction Taxes.

Dávila, E. and A. Korinek (2018, jan). Pecuniary Externalities in Economies with Financial Frictions. *The Review of Economic Studies 85*(1), 352–395.

Dávila, E. and A. Walther (2021). Corrective Regulation with Imperfect Instruments.

Davis, J. S. and E. Van Wincoop (2018, dec). Globalization and the increasing correlation between capital inflows and outflows. *Journal of Monetary Economics 100*, 83–100.

De Marco, F., M. Macchiavelli, and R. Valchev (2019). International Diversification and Information Heterogeneity.

ECB (2018). SSM Supervisory Manual.

Erten, B., A. Korinek, and J.-a. Ocampo (2021). Capital Controls: Theory and Evidence. *Journal of Economic Literature 59*(1), 45–89.

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

Farhi, E., M. Golosov, and A. Tsyvinski (2009, jul). A theory of liquidity and regulation of financial intermediation. *Review of Economic Studies 76*(3), 973–992.

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

Farhi, E. and J. Tirole (2020). Shadow Banking and the Four Pillars of Financial Intermediation.

Farhi, E. and I. Werning (2016, sep). A Theory of Macroprudential Policies in the Presence of Nominal Rigidities. *Econometrica 84*(5), 1645–1704.

Feldman, R. J. and K. Heincecke (2018). The Minneapolis Plan to End Too Big to Fail: The Right Plan at the Right Time.

Financial Stability Board (2013). Strengthening Oversight and Regulation of Shadow Banking: Policy Framework for Strengthening Oversight and Regulation of Shadow Banking Entities.

Financial Stability Board (2020). Global Monitoring Report on Non-Bank Financial Intermediation 2019.

Forbes, K. J. and F. E. Warnock (2012, nov). Capital flow waves: Surges, stops, flight, and retrenchment. *Journal of International Economics 88*(2), 235–251.

Geanakoplos, J. and H. M. Polemarchakis (1985). Existence, Regularity, and Constrained Suboptimality of Competitive Allocations When the Asset Market Is Incomplete. *Cowles Foundation*

Discussion Papers.

Gorton, G., A. Metrick, A. Shleifer, and D. K. Tarullo (2010). Regulating the shadow banking system. *Brookings Papers on Economic Activity* (2), 261–312.

Grochulski, B. and Y. Zhang (2019). Optimal Liquidity Policy with Shadow Banking. *Economic Theory 68*(4), 967–1015.

Hanson, S. G., A. Shleifer, J. C. Stein, and R. W. Vishny (2015, sep). Banks as patient fixed-income investors. *Journal of Financial Economics 117*(3), 449–469.

He, P., Z. Liu, and C. Xie (2018, sep). In the Shadow of Shadow Banking.

Holmström, B. and J. Tirole (1998). Private and public supply of liquidity. *Journal of Political Economy 106*(1), 1–40.

Huang, J. (2018, nov). Banking and shadow banking. *Journal of Economic Theory 178*, 124–152.

IMF (2012). The Liberalization and Management of Capital Flows - An Institutional View.

Kiyotaki, N. and J. Moore (1997, apr). Credit Cycles. *Journal of Political Economy 105*(2), 211–248.

Korinek, A. and A. Simsek (2016, mar). Liquidity trap and excessive leverage. *American Economic Review 106*(3), 699–738.

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

Maggiori, M. (2017, oct). Financial intermediation, international risk sharing, and reserve currencies. *American Economic Review 107*(10), 3038–3071.

Maggiori, M., B. Neiman, and J. Schreger (2019). International Currencies and Capital Allocation. *Journal of Political Economy. Forthcoming*.

Martinez-Miera, D. and R. Repullo (2019). Markets, Banks, and Shadow Banks.

Mendoza, E. G. (2010). Sudden Stops, Financial Crises, and Leverage. *American Economic Review 100*(5), 1941–1966.

Mendoza, E. G., V. Quadrini, and J. V. Ríos-Rull (2009, jun). Financial integration, financial development, and global imbalances. *Journal of Political Economy 117*(3), 371–416.

Milesi-Ferretti, G. M. and C. Tille (2011, apr). The great retrenchment: International capital flows during the global financial crisis. *Economic Policy 26*(66), 285–342.

Moreira, A. and A. Savov (2017, dec). The Macroeconomics of Shadow Banking. *Journal of Finance 72*(6), 2381–2432.

Niepmann, F. (2015, jul). Banking across borders. *Journal of International Economics 96*(2), 244–265.

Ordoñez, G. (2018, jan). Sustainable shadow banking. *American Economic Journal: Macroeconomics 10*(1), 33–56.

Plantin, G. (2014, jan). Shadow Banking and Bank Capital Regulation.

Schmitt-Grohé, S. and M. Uribe (2016, oct). Downward Nominal Wage Rigidity, Currency Pegs, and Involuntary Unemployment. *Journal of Political Economy 124*(5), 1466–1514.

Shen, L. S. (2019). Global vs. Local Banking: Firm Financing in a Globalized Financial System.

A Proofs

A.1 Proof of Proposition 2

Because all financial actors are regulated, the social optimality condition for $I_i(m)$ is

$$0 = \frac{\partial \mathcal{L}_i}{\partial I_i(m)} - \nu(m)$$

where ν is the Lagrange multiplier on market clearing, that is $\nu(m)$ is the Lagrange multiplier on market clearing for m. The private optimality condition of an optimizing agent facing tax rate m is

$$0 = \frac{\partial \mathcal{L}_i}{\partial I_i(m)} - \lambda_i \tau(m),$$

so that matching terms we obtain the optimal tax

$$\tau_i(m) = \frac{1}{\lambda_i} \nu(m).$$

Next, by Envelope Theorem the optimal external price $p(m)$ is

$$0 = \sum_{h \in \mathcal{H}} \mu_h \frac{\partial \mathcal{L}_h}{\partial p(m)} + \sum_{i \in \mathcal{I}} \mu_i \frac{\partial \mathcal{L}_i}{\partial p(m)} + \nabla_{p(m)} X_{\mathcal{H}} \nu$$

Given the definition of $\mathcal{E}(m)$ and $\Xi_{\mathcal{H}}(m, \cdot)$, that is the m-th row of $\Xi_{\mathcal{H}}$, this gives

$$\Xi_{\mathcal{H}}(m, \cdot) \nu = -\mathcal{E}(m).$$

Putting in vector form and inverting $\Xi_{\mathcal{H}}$, we have $\nu = -\Xi_{\mathcal{H}}^{-1} \mathcal{E}$. Finally, substituting back in to the tax formula in vector form, we obtain

$$\tau_i(m) = -\frac{1}{\lambda_i} \Xi_{\mathcal{H}}^{-1} \mathcal{E}.$$

Finally, lump sum transfers mean that $\lambda_i = \lambda$ is constant across regulated financial actors, giving the result.

A.2 Proof of Proposition 3

The proof follows the same steps as Proposition 2. The optimal tax vector is given by

$$\tau_i = \frac{1}{\lambda}\nu^*,$$

where ν^* is the new Lagrange multiplier on market clearing. By Envelope Theorem, the optimal price vector p is now given by

$$0 = \mathcal{E} + (\Xi_{\mathcal{H}} - \Xi_{\mathcal{S}})\nu^*$$

From here, we obtain

$$\tau_i = -\frac{1}{\lambda}(\Xi_{\mathcal{H}} - \Xi_{\mathcal{S}})^{-1}\mathcal{E} = -\frac{1}{\lambda}(\mathbb{I} - \Xi_{\mathcal{H}}^{-1}\Xi_{\mathcal{S}})^{-1}\Xi_{\mathcal{H}}^{-1}\mathcal{E}$$

which gives the result.

A.3 Proof of Corollary 4

The proof follows immediately from the first order approximation $\mathcal{M} \approx \mathbb{I} + \Xi_{\mathcal{H}}^{-1}\Xi_{\mathcal{S}}$. Given that $\Xi_{\mathcal{H}}^{-1}$ is diagonal, then the m-th row of the matrix $\Xi_{\mathcal{H}}^{-1}\Xi_{\mathcal{S}}$ is $\mathbb{I}(m,\cdot) + \frac{1}{\partial I_{\mathcal{H}}(m)/\partial p(m)}\Xi_{\mathcal{S}}(m,\cdot)$. Thus, we obtain

$$\tau^*(m) = \left(1 + \frac{\partial I_{\mathcal{S}}(m)/\partial p(m)}{\partial I_{\mathcal{H}}(m)/\partial p(m)}\right)\tau(m) + \sum_{m' \neq m} \frac{\partial I_{\mathcal{S}}(m')/\partial p(m)}{\partial I_{\mathcal{H}}(m)/\partial p(m)}\tau(m').$$

Finally, the last step follows since

$$\frac{\partial I_{\mathcal{S}}(m')/\partial p(m)}{\partial I_{\mathcal{H}}(m)/\partial p(m)} = \frac{\partial I_{\mathcal{S}}(m')/\partial p(m)}{\partial I_{\mathcal{H}}(m)/\partial p(m)} \cdot \frac{p(m)/I_{\mathcal{S}}(m')}{p(m)/I_{\mathcal{H}}(m)} \cdot \frac{I_{\mathcal{S}}(m')}{I_{\mathcal{H}}(m)} = \alpha(m')\frac{\xi_{\mathcal{S}}(m,m')}{\xi_{\mathcal{H}}(m,m)}\frac{I_{\mathcal{I}}(m')}{I_{\mathcal{I}}(m)}$$

where we have $\frac{I_{\mathcal{I}}(m')}{I_{\mathcal{I}}(m)} = 1$ when $m' = m$.

A.4 Proof of Corollary 5

Given household cross-price elasticities are zero and they have a constant elasticity of 1, we have

$$\Xi_{\mathcal{H}} = \begin{pmatrix} I_{\mathcal{H}}(1) & 0 \\ 0 & -I_{\mathcal{H}}(2) \end{pmatrix}$$

where the negative sign at date 2 reflects that they are buyers, and so buy less (supply more) as the price rises. For unregulated financial actor i, we have

$$\frac{\partial I_i(2)}{\partial p(2)} = -I_i(2)$$

and so we have $\Xi_{\mathcal{S}}(2,2) = -I_{\mathcal{S}}(2)$. Given cross price elasticity of zero, then we have

$$\mathcal{M}(2,2)^{-1} = 1 - \frac{-I_{\mathcal{S}}(2)}{-I_{\mathcal{H}}(2)} = \frac{I_{\mathcal{B}}(2)}{I_{\mathcal{H}}(2)}.$$

Inverting, we obtain

$$\mathcal{M}(2,2) = \frac{I_{\mathcal{B}}(2) + I_{\mathcal{S}}(2)}{I_{\mathcal{B}}(2)} = 1 + \frac{I_{\mathcal{S}}(2)}{I_{\mathcal{B}}(2)}$$

which gives the result.

A.5 Proof of Proposition 6

Define the expenditure function $e_i(P,q,\overline{U})$ and Hicksian demand $h_i(P,q,\overline{U})$ as the solutions to the expenditure minimization problem, and define Marshallian demand $I_i(P,q,w_i)$ as the solution to the utility maximization problem. As usual, Marshallian and Hicksian demand are related by

$$I_i(P,q,e_i(P,q,\overline{U})) = h_i(P,q,\overline{U}).$$

From here, we set $P = q = p$ and totally differentiate demand for m in $p(m)$, obtaining

$$\frac{\partial I_i(m)}{\partial P(m)} + \frac{\partial I_i(m)}{\partial q(m)} + \frac{\partial I_i(m)}{\partial w_i}\left(\frac{\partial e_i}{\partial P(m)} + \frac{\partial e_i}{\partial q(m)}\right) = \frac{\partial h_i(m)}{\partial P(m)} + \frac{\partial h_i(m)}{\partial q(m)}.$$

Observe that $\frac{\partial I_i(m)}{\partial P(m)} + \frac{\partial I_i(m)}{\partial q(m)} = \Xi_i(m,m)$ by definition. It remains only to characterize the derivatives of the expenditure function. The Lagrangian of the expenditure minimization problem is

$$\mathcal{L}^H = h_i^c + ph_i - \frac{1}{\lambda_i}(U(h_i^c,h_i) - \overline{U}) - \frac{1}{\lambda_i}\Lambda_i\Gamma_i(h_i^c,h_i,q).$$

From here we obtain the first order condition

$$0 = p(m)h_i(m) - \frac{1}{\lambda_i}\frac{\partial U}{\partial h_i(m)} - \frac{1}{\lambda_i}\Lambda_i\frac{\partial \Gamma_i}{\partial h_i(m)}$$

Observe that this is the same first order condition as from the utility maximization problem. Thus when we use $\overline{U} = V_i(P,q,w_i)$ so that Marshallian and Hicksian demand are equal, we obtain the same Lagrange multipliers. From here, Envelope Theorem immediately yields

$$\frac{de_i}{\partial p(m)} = h_i(m) - \frac{1}{\lambda_i}\frac{\partial \Lambda_i}{\partial p(m)} = -\mathcal{E}_i(m).$$

Finally, substituting back in we obtain the result.

A.6 Proof of Proposition 7

Consider the social planner's Lagrangian, given as before.

$$\mathcal{L} = \sum_{h\in\mathcal{H}} \mu_h \mathcal{L}_h + \sum_{i\in\mathcal{I}} \mu_i \mathcal{L}_i + (I_{\mathcal{H}} - I_{\mathcal{I}})'\nu^*.$$

By Envelope Theorem, for unregulated actor i we have

$$\frac{\partial \mathcal{L}_i}{\partial \tau(m^*)} = \lambda_i \left(I_i^*(m^*) + \frac{\partial I_i^*(m^*)}{\partial \tau(m^*)}\tau(m^*) - I_i(m) \right) = \lambda_i \frac{\partial I_i(m^*)}{\partial \tau(m^*)}\tau(m^*).$$

Next, by Envelope Theorem for any price $p(m)$

$$\frac{\partial \mathcal{L}_i}{\partial p(m)} = \lambda_i \frac{\partial I_i^*(m^*)}{\partial p(m)}\tau(m^*) + \mathcal{E}_i(m)$$

where the first term reflects the remitted revenues. Intuitively, the first term reflects that the regulation of m^* means that i is no longer on her private first order condition in market m^* due to the activity tax. Thus from here we have

$$0 = \frac{\partial \mathcal{L}}{\partial p(m)} = \sum_{i\in\mathcal{S}} \mu_i \lambda_i \frac{\partial I_i^*(m^*)}{\partial p(m)}\tau(m^*) + \mathcal{E}(m) + \left(\Xi_{\mathcal{H}} - \Xi_{\mathcal{S}} \right)\nu^*.$$

Re-expressing in vector form,

$$0 = \sum_{i\in\mathcal{S}} \mu_i \lambda_i \nabla_p I_i^*(m^*)\tau(m^*) + \mathcal{E} + \left(\Xi_{\mathcal{H}} - \Xi_{\mathcal{S}} \right)\nu^*,$$

which inverting yields

$$\nu^* = -\mathcal{M}\Xi_{\mathcal{H}}^{-1} \sum_{i\in\mathcal{S}} \mu_i \lambda_i \nabla_p I_i^*(m^*)\tau(m^*) - \mathcal{M}\Xi_{\mathcal{H}}^{-1}\mathcal{E}.$$

Thus putting these together in the FOC for $\tau(m^*)$, we have

$$0 = \sum_i \mu_i \lambda_i \frac{\partial I_i(m^*)}{\partial \tau(m^*)} \tau(m^*) - \nabla_{\tau^*} I_{\mathcal{S}} \nu^*$$

From here, we can use absence of wealth effects and $p(m^*)$ not appearing in constraints to write

$$\left[\sum_i \mu_i \lambda_i \frac{\partial I_i(m^*)}{\partial p(m^*)} + \Xi_{\mathcal{S}}(m^*,\cdot)\mathcal{M}\Xi_{\mathcal{H}}^{-1} \sum_{i\in\mathcal{S}} \mu_i \lambda_i \nabla_p I_i^*(m^*)\right] \tau(m^*) = \Xi_{\mathcal{S}}(m^*,\cdot)\nu^*.$$

Finally, given absence of wealth effects we have $\lambda_i = \lambda$ for all $i \in \mathcal{S}$. From this, we obtain

$$\tau(m^*) = \frac{1}{\lambda} \frac{\Xi_{\mathcal{S}}(m^*,\cdot)\lambda\tau^*}{\Xi_{\mathcal{S}}(m^*,m^*) + \Xi_{\mathcal{S}}(m^*,\cdot)\mathcal{M}\Xi_{\mathcal{H}}^{-1}\Xi_{\mathcal{S}}(\cdot,m^*)}.$$

Finally, following the steps of Proposition 3, we have $\tau^* = \frac{1}{\lambda}\nu^*$, which yields

$$\tau(m^*) = \frac{\Xi_{\mathcal{S}}(m^*,\cdot)\tau^*}{\Xi_{\mathcal{S}}(m^*,m^*) + \Xi_{\mathcal{S}}(m^*,\cdot)\mathcal{M}\Xi_{\mathcal{H}}^{-1}\Xi_{\mathcal{S}}(\cdot,m^*)}.$$

A.7 Proof of Proposition 8

Recalling the social planner's Lagrangian, we have

$$\nabla_{d\tau}\mathcal{L} \approx d\tau'\nabla_{\tau_i}\mathcal{L}_i - \mu_i d\tau' \nabla_{d\tau} I_i \nu = -\mu_i d\tau' \nabla_{\tau_i} I_i \nu^*$$

where the last line follows since by Envelope Theorem,

$$\nabla_{\tau_i}\mathcal{L}_i = \nabla_{\tau_i} c_i \nabla_{c_i}\mathcal{L}_i + \nabla_{\tau_i} I_i \nabla_{I_i}\mathcal{L}_i = 0$$

given that i was at a private optimum under which the FOCs $\nabla_{c_i}\mathcal{L}_i = 0$ and $\nabla_{I_i}\mathcal{L}_i = 0$ held. Therefore, we are left with only the impact on market clearing. From the proof of Proposition 3, we have $\nu^* = \lambda\tau^*$. Therefore, we have

$$\nabla_{d\tau}\mathcal{L} = -\mu_i \lambda d\tau' \nabla_{\tau_i} I_i \tau^*.$$

The only remaining element is to charcterize $\nabla_{\tau_i} I_i$. We have the Marshallian demand function $I_i(P,q,w_i)$. Given (A1), there are no wealth effects, and so we have $I_i(P,q)$. Define $\hat{M} \subset M$ as the set of activities to which $d\tau$ does not apply, that is so that $d\tau(m) = 0$. (A2) says that we can write

$I_i(P,q(\hat{M}))$. Therefore, we have for $m \notin \hat{M}$

$$\frac{\partial I_i(m')}{\partial d\tau(m)} = \frac{\partial I_i(m')}{\partial P(m)} = \frac{\partial I_i(m')}{\partial P(m)} + \underbrace{\frac{\partial I_i(m')}{\partial q(m)}}_{=0} = \frac{\partial I_i(m')}{\partial p(m)} = \Xi_i(m,m').$$

Lastly, note that rows $m \in \hat{M}$ of Ξ_i are all multiplied by $d\tau(m) = 0$. Therefore, without loss we can write $\nabla_{\tau_i} I_i = \nabla_p I_i = \Xi_i$. From here, we obtain

$$\frac{1}{\lambda}\nabla_{d\tau}\mathcal{L} = -\mu_i d\tau' \Xi_i \tau^*$$

which re-expresses the gain normalizing by the marginal value of wealth λ. Therefore, we denote

$$\Delta_i^*(d\tau) = \frac{1}{\lambda}\nabla_{d\tau}\mathcal{L} = -\mu_i d\tau' \Xi_i \tau^*$$

giving the result.

A.8 Proof of Proposition 9

The proof follows the same steps as the proof of Proposition 8, giving for us

$$\Delta_i^*(d\tau) = -\mu_i d\tau' \nabla_{\tau_i} I_i \tau^*,$$

but we can no longer substitute in Ξ_i. Instead, we have to consider full Marshallian demand $I_i(P,q,w_i + d\tau I_i^*)$, where $d\tau + I_i^*$ reflects remitted revenues. Differentiating around $d\tau = 0$, we have

$$\frac{\partial I_i(m')}{\partial d\tau(m)} = \frac{\partial I_i(m')}{\partial P(m)} + \frac{\partial I_i(m')}{\partial w_i} I_i(m),$$

since in equilibrium we have $I_i^*(m) = I_i(m)$. Finally, from Section 3.3 and the proof of Proposition 6, the equivalence between Marshallian demand and Hicksian demand gives us

$$I_i(P,q,e_i(P,q,\overline{U})) = h_i(P,q,\overline{U}).$$

Differentiating element m' of both sides in $d\tau(m)$ for $P = P + d\tau$, we have

$$\frac{\partial I_i(m')}{\partial P(m)} + \frac{\partial I_i(m')}{\partial w_i}\frac{\partial e_i}{\partial P(m)} = \frac{\partial h_i(m')}{\partial P(m)}$$

Lastly, from the proof of Proposition 6 and Marshallian-Hicksian equalization, we have $\frac{\partial e_i}{\partial P(m)} = h_i(m) = I_i(m)$, which yields

$$\frac{\partial I_i(m')}{\partial P(m)} + \frac{\partial I_i(m')}{\partial w_i} \frac{\partial e_i}{\partial P(m)} = \frac{\partial h_i(m')}{\partial P(m)}.$$

Therefore, substituting in above we obtain

$$\frac{\partial I_i(m')}{\partial d\tau(m)} = \frac{\partial h_i(m')}{\partial P(m)}.$$

This gives in matrix form $\nabla_{d\tau} I_i = H_i$, giving the result.

A.9 Proof of Proposition 10

By definition, we have the demand functions $I_i(1) = 0$ and $I_i(2) = I_i(P, q, w_i, 2) = -\rho_i / q(2)$. Therefore, $H_i = 0$, yielding the result.

A.10 Proof of Proposition 11

The proofs follow exactly as in Propositions 8 and 9. We now have by Envelope Theorem

$$\Delta^*(d\tau(m)) = \frac{1}{\lambda} \nabla_{d\tau(m)} \mathcal{L} = -\sum_i \mu_i d\tau' \nabla_{d\tau(m)} I_i \tau^*.$$

Finally, under (A1) and (A2) we have $\nabla_{d\tau(m)} = \Xi_i(m, \cdot)$ as before. Absent (A1) and (A2), we instead have $\nabla_{d\tau(m)} = H_i(m, \cdot)$. Finally, we have defined $\Xi_{\mathcal{S}} = \sum_i \mu_i \Xi_i$ and similarly for $H_{\mathcal{S}}$, giving the result.

B Optimal Activity Regulation

Formally, suppose now that the social planner can set uniform wedges $\hat{\tau}(m)$ on a subset $\hat{M} \subset M$ of unregulated flows. As before, the social planner has a full set of controls over regulated actors. The following result characterizes optimal activity regulation.

Proposition 15. *Optimal regulation $\hat{\tau}$ for $m \in \hat{M}$ satisfies*

$$0 = I(\hat{M}) \sum_{i \in \mathcal{S}} \mu_i H_i \left(\mathbb{I} - \hat{\tau} \nabla_{w_i} I_i \right)^{-1} \left[\left(\lambda_i \mathbb{I} + \mathcal{M} \Xi_{\mathcal{H}^{-1}} \Theta_{\mathcal{S}} \right) \hat{\tau} + \tau^* \right]. \tag{26}$$

where $I(\hat{M})$ is a diagonal matrix with $I(m,m) = 1$ for $m \in \hat{M}$ and $I(m,m) = 0$ otherwise, and where $\Theta_{\mathcal{S}} = \sum_{i \in \mathcal{S}} \mu_i \lambda_i \Xi_i$.

The intuition is the same as in Proposition 7, except that we now have the term $H_i\left(I - \hat{\tau}\nabla_{w_i} x_i\right)^{-1}$. This reflects that taxes are compensated, so the activity regulation induces substitution effects H_i but also wealth effects from the remissions. These wealth effects were irrelevant under complete instruments in identity-based regulation. Note that the first-order condition is pre-multiplied by $I(\hat{M})$, which ensures that it only holds for regulatory instruments which the planner possesses.

B.0.1 Proof of Proposition 15

Consider the social planner's Lagrangian, given as before.

$$\mathcal{L} = \sum_{h \in \mathcal{H}} \mu_h \mathcal{L}_h + \sum_{i \in \mathcal{I}} \mu_i \mathcal{L}_i + (I_{\mathcal{H}} - I_{\mathcal{I}})' \lambda.$$

Using the private unregulated Lagrangian, accounting for activity regulation, by Envelope Theorem

$$\frac{\partial \mathcal{L}_i}{\partial \tau(m)} = \lambda_i \left(I_i^*(m) + (\nabla_{\tau(m)} I_i^*)\tau - I_i(m) \right) = \lambda_i (\nabla_{\tau(m)} I_i)\tau.$$

From here, the first order condition for activity regulation $\tau(m)$ is given by

$$0 = \sum_{i \in \mathcal{S}} \mu_i \lambda_i (\nabla_{\tau(m)} I_i)\tau - (\nabla_{\tau(m)} I_{\mathcal{S}})\lambda.$$

Defining $I_i = I_i(p + \tau, w_i, p)$ and differentiating, we have

$$\frac{\partial I_i(m')}{\partial \tau(m)} = \frac{\partial I_i(m')}{\partial \tilde{p}(m)} + \frac{\partial I_i(m')}{\partial w_i}\left(I_i(m) + (\nabla_{\tau(m)} I_i)\tau \right).$$

In matrix form, this yields

$$\nabla_{\tau(m)} I_i = \nabla_{\tilde{p}(m)} I_i + I_i(m) \nabla_{w_i} I_i + \nabla_{\tau(m)} I_i \tau \nabla_{w_i} I_i.$$

Rearranging and inverting, we obtain, we have

$$\nabla_{\tau(m)} I_i = \left(\nabla_{\tilde{p}(m)} I_i + I_i(m) \nabla_{w_i} I_i \right) \left(I - \tau \nabla_{w_i} I_i \right)^{-1}$$

Finally, noting that $H_i(m,\cdot) = \nabla_{\tilde{p}(m)} I_i + I_i(m)\nabla_{w_i} I_i$, we have

$$\nabla_{\tau(m)} I_i = H_i(m,\cdot)\Big(\mathbb{I} - \tau\nabla_{w_i} I_i\Big)^{-1}.$$

From here, we obtain the first order condition, we obtain

$$0 = \sum_{i\in\mathcal{S}} \mu_i H_i(m,\cdot)\Big(\mathbb{I} - \tau\nabla_{w_i} I_i\Big)^{-1}\Big(\lambda_i\tau - \lambda\Big).$$

Finally, we can take the derivative in p. Note that for unregulated financial actors, we have

$$d\mathcal{L}_i = \nabla_p \mathcal{L}_i + \lambda_i \nabla_p I_i^* \tau = \nabla_p \mathcal{L}_i + \lambda_i \nabla_p I_i \tau$$

so that we have the social FOC for p

$$0 = \mathcal{E} + \sum_{i\in\mathcal{S}} \mu_i\lambda_i \Xi_i \tau + (\Xi_{\mathcal{H}} - \Xi_{\mathcal{S}})\lambda$$

giving

$$\lambda = -\mathcal{M}\Xi_{\mathcal{H}}^{-1}\Big[\mathcal{E} + \sum_{i\in\mathcal{S}} \mu_i\lambda_i\Xi_i\tau\Big]$$

giving

$$0 = \sum_{i\in\mathcal{S}} \mu_i H_i(m,\cdot)\Big(\mathbb{I} - \tau\nabla_{w_i} I_i\Big)^{-1}\Big(\lambda_i\tau + \mathcal{M}\Xi_{\mathcal{H}}^{-1}\Big[\mathcal{E} + \sum_{i\in\mathcal{S}} \mu_i\lambda_i\Xi_i\tau\Big]\Big).$$

Notice that for complete instruments ($\hat{M} = M$), then given λ_i is constant across i in this case we recover $\tau = -\frac{1}{\lambda_i}\mathcal{M}\Xi_{\mathcal{H}}^{-1}\mathcal{E}$ as before. Finally, let us define $I(\hat{M})$ to be a diagonal matrix with $I(m,m) = 1$ for $m \in \hat{M}$ and $I(m,m) = 0$ otherwise. Then, notice that we can write this system of equations as

$$0 = I(\hat{M})\sum_{i\in\mathcal{S}} \mu_i H_i\Big(\mathbb{I} - \tau\nabla_{w_i} I_i\Big)^{-1}\Big[\Big(\lambda_i\mathbb{I} + \mathcal{M}\Xi_{\mathcal{H}^{-1}}\Theta_{\mathcal{S}}\Big)\tau + \mathcal{M}\Xi_{\mathcal{H}}^{-1}\mathcal{E}\Big].$$

where we define $\Theta_{\mathcal{S}} = \sum_{i\in\mathcal{S}} \mu_i\lambda_i\Xi_i$.

C Non-Pecuniary Externalities

In the main text sections, we have considered the case where the social planner has a complete set of regulatory instruments (wedges) for regulated actors, and moreover considers pecuniary externalities.

We now allow for non-pecuniary externalities, and characterize an analogue of Proposition 3.

Incorporating non-pecuniary externalities requires only a slight modification to the framework above. We maintain the notation of I_i being the vector of bank activities and w_i the wealth level, but incorporate the numeraire c_i into the vector I_i to simplify notation. Financial actors maximize $U_i(I_i, P)$ subject to the constraint set $\Gamma_i(I_i, P, w_i) \geq 0$, which now includes any budget constraints. In this notation, P reflects a set of equilibrium aggregates, which may include not only prices but also any other welfare-relevant aggregates. Nonfinancial actors are defined similarly, with their supply functions denoted $I_h(P, w_i)$.

The equilibrium aggregates are defined by an equilibrium relationship

$$\Phi(I, P) = 0 \tag{27}$$

where $I = \{I_j\}_{j \in \mathcal{J} \cup \mathcal{H}}$ is the activities of all actors in the economy. This notation captures the framework of Section 2 with pecuniary externalities when equation (27) is a set of market clearing conditions.[35]

Aggregate Response Matrices. In the baseline model, the aggregate response matrices $\Xi_{\mathcal{H}}$ and $\Xi_{\mathcal{S}}$ were sufficient statistics, because they captured price impacts on market clearing. For this section, we define analogous expressions capturing price impacts on the constraint set $\Phi(I, P)$. In particular, we can define the *aggregate supply impact matrix* as

$$\Theta_{\mathcal{H}} = \sum_{h \in \mathcal{H}} \nabla_P I_h \nabla_{I_h} \Phi,$$

and analogously we can define the *aggregate unregulated demand impact matrix* as

$$\Theta_{\mathcal{S}} = \sum_{i \in \mathcal{S}} \nabla_P I_i \nabla_{I_i} \Phi.$$

These are the impacts of changes in P on the constraint set through the activities of nonfinancial actors and unregulated financial actors. In the baseline model, the impact $\nabla_{I_i} \Phi$ is simply the measure μ_i, and $\Theta_{\mathcal{H}}$ and $\Theta_{\mathcal{S}}$ collapse to $\Xi_{\mathcal{H}}$ and $\Xi_{\mathcal{S}}$, respectively.

Example 16 (Aggregate Demand Externalities)**.** We give an example of regulation with aggregate demand externalities. Aggregate demand externalities arise in general because price rigidities force one or more actors to absorb residual demand in a market, even though those actors may not be on

[35] Our environment can also capture incomplete instruments. We give an example of incomplete instruments in the spirit of macroprudential regulation, where future allocations cannot be controlled at the current date. Activities $I_i(m)$ for $m \geq \tilde{M}$ are future activities, with the constraint set specifying demand functions $I_i(m) = I_i(\tilde{m}, P, I_i^0)$ for $m \geq \tilde{M}$, where $I_i^0 = (I_i(1), \ldots, I_i(\tilde{M} - 1))$ is the vector of activities that can be regulated.

their first order conditions.[36] We can impose price rigidities via a set of constraints $\Psi_i(P) = 0$, and impose a rationing rule on good m by denoting an aggregate $P(m)$ to be residual demand for that good, which then actors agents are forced by their constraint sets to absorb.

C.1 Optimal Regulation

We can now characterize optimal regulation with and without unregulated financial actors

Proposition 17. *In this environment,*

1. *Optimal regulation without unregulated financial actors is given by*

$$\tau_i = -\frac{1}{\lambda_i}\nabla_{I_i}\Phi\left(\nabla_P\Phi + \Theta_{\mathcal{H}}\right)^{-1}\mathcal{E}. \tag{28}$$

2. *Optimal regulation with unregultaed financial actors is given by*

$$\tau_i = -\frac{1}{\lambda_i}\nabla_{I_i}\Phi\left(\nabla_P\Phi + \Theta_{\mathcal{H}}\mathcal{M}^{-1}\right)^{-1}\mathcal{E} \tag{29}$$

where $\mathcal{M} = \left(\mathbb{I} + \Theta_{\mathcal{H}}^{-1}\Theta_{\mathcal{S}}\right)^{-1}$.

Optimal regulation accounts for externalities resulting from changes in aggregates P. With only regulated financial actors, a change in activities of i leads to an effect $\Phi_{I_i}\Phi$ on the constraint set. There are two ways this change can be offset by a change in aggregates P. The first is the direct effect of a change in P, given by $\nabla_P\Phi$. The second is the indirect effect through nonfinancial actors, given by $\Theta_{\mathcal{H}}$. When aggregates appear in the constraint set only through nonfinancial supply, as with prices and market clearing, we have $\nabla_P\Phi = 0$ and are left only with the indirect effect, as in previous sections. The total change in both current and future prices multiplies the vector $\mathcal{E}$ of externalities arising from changes in aggregates.

Next introducing unregulated financial actors, we obtain a regulatory arbitrage multiplier $\mathcal{M}$, which multiplies only the indirect effect but not the direct effect. This regulatory arbitrage multiplier has a similar intuition to the previous sections: a change in aggregate demand forces a change in aggregates P. However, this yields a change in unregulated demand, which forces further changes in aggregates, and so forth.

[36]For example, in the conventional New Keynesian model, firms that are unable to reset their price are forced to supply whatever quantity is demanded.

C.2 Proof of Proposition 17

The Lagrangian of the social planner is given by

$$\mathcal{L} = \sum_{j \in \mathcal{I} \cup \mathcal{H}} \mu_j \mathcal{L}_j + \Phi' \lambda.$$

We obtain by the usual steps

$$\tau_i = \frac{1}{\lambda_i} \nabla_{x_i} \Phi \lambda.$$

Without Unregulated Financial Actors. We have

$$0 = \nabla_P \sum_{j \in \mathcal{I} \cup \mathcal{H}} \mu_j \mathcal{L}_j + \left(\nabla_P \Phi + \sum_{h \in \mathcal{H}} \nabla_P I_h \nabla_{I_h} \Phi \right) \lambda.$$

Noting that $\nabla_P \Phi + \sum_{h \in \mathcal{H}} \nabla_P I_h \nabla_{I_h} \Phi$ is a square matrix, we have

$$\lambda = -\left(\nabla_P \Phi + \sum_{h \in \mathcal{H}} \nabla_P I_h \nabla_{I_h} \Phi \right)^{-1} \mathcal{E}$$

where we now have $\mathcal{E} = \nabla_P \sum_{j \in \mathcal{I} \cup \mathcal{H}} \mu_j \mathcal{L}_j$. Substituting in yields

$$\tau_i = -\frac{1}{\lambda_i} \nabla_{I_i} \Phi \left(\nabla_P \Phi + \sum_{h \in \mathcal{H}} \nabla_P I_h \nabla_{I_h} \Phi \right)^{-1} \mathcal{E}.$$

With Unregulated Financial Actors. We have

$$0 = \nabla_P \sum_{j \in \mathcal{I} \cup \mathcal{H}} \mu_j \mathcal{L}_j + \left(\nabla_P \Phi + \sum_{h \in \mathcal{H}} \nabla_P I_h \nabla_{I_h} \Phi + \sum_{j \in \mathcal{S}} \nabla_P I_j \nabla_{I_j} \Phi \right) \lambda,$$

so that we have

$$\lambda = -\left(\nabla_P \Phi + \sum_{h \in \mathcal{H}} \nabla_P I_h \nabla_{I_h} \Phi \mathcal{M}^{-1} \right)^{-1} \mathcal{E},$$

where we have defined

$$\mathcal{M} = \left(I + \left(\sum_{h \in \mathcal{H}} \nabla_P I_h \nabla_{I_h} \Phi \right)^{-1} \sum_{j \in \mathcal{S}} \nabla_P I_j \nabla_{I_j} \Phi \right)^{-1}.$$

This yields the result

$$\tau_i = -\frac{1}{\lambda_i} \nabla_{I_i} \Phi \left(\nabla_P \Phi + \sum_{h \in \mathcal{H}} \nabla_P I_h \nabla_{I_h} \Phi \mathcal{M}^{-1} \right)^{-1} \mathcal{E}.$$