WORKING PAPERS

N° 1445

June 2023

# "Regulation with Externalities and Misallocation in General Equilibrium"

Christopher Clayton and Andreas Schaab

Toulouse School of Economics

# Regulation with Externalities and Misallocation in General Equilibrium

Christopher Clayton[*] Andreas Schaab[†]

**First version:** December, 2020

**This version:** November, 2022

**Abstract**

We study allocative efficiency and optimal regulation in inefficient economies with misallocation and pecuniary externalities. We characterize the allocative value of a market based on its price, cross-sectional misallocation among participants, and pecuniary externalities. With both complete and incomplete regulation, a social planner equalizes prices faced by fully regulated agents with the allocative value of markets. With incomplete regulation, the planner uses regulation of fully regulated agents to trade off correcting externalities against misallocation from regulatory arbitrage by unregulated agents. The planner uses partial regulation of unregulated agents to reduce misallocation from regulatory arbitrage. We leverage our framework to answer relevant policy questions, including: (i) the social value of a new unregulated agent is its profits plus a simple measure of social value of its activities; (ii) the social value of new regulation is summarized by its reduction in misallocation. We apply our theory to shadow bank institution regulation and capital flow management in a small open economy. We extend our theory to environments with multiple regulators and common agency.

**JEL codes:** F38, G28, D62

**Keywords:** Misallocation, regulatory arbitrage, unregulated finance, macroprudential regulation, capital flows, capital controls, pecuniary externalities

[*]Yale School of Management. Email: christopher.clayton@yale.edu. http://www.christopherdclayton.com

[†]Toulouse School of Economics. Email: andreas.schaab@tse-fr.eu. https://andreasschaab.com

We are grateful to David Baqaee, Antonio Coppola, Eduardo Dávila, Gary Gorton, Sam Hanson, Ernest Liu, Matteo Maggiori, Aakriti Mathur, Alp Simsek, Jean Tirole, Aleh Tsyvinski, and seminar and conference participants 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 subsumes a working paper previously circulated under the title "Shadow Banks and Optimal Regulation."

# 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 framework to study allocative efficiency and optimal regulation in inefficient economies with distortions arising from misallocation and pecuniary externalities. We then leverage this framework to speak to policy questions about unregulated finance, in particular shadow banking and capital flows.

We study an exchange economy in which heterogeneous agents interact in markets for goods. We abstract from direct bilateral relationships between any two agents, such as customer-supplier relationships. Nonetheless, markets in our economy are interconnected through agents' trading patterns: if an agent retrenches out of one market, prompting a fall in the market price, other agents' trading responses may lead to spillovers to other markets. We allow for two sources of inefficiency. First, agents are subject to constraints that depend on market prices, giving rise to endogenous pecuniary externalities. Second, we allow for distortions in the transaction prices agents face. When two agents face different transaction prices for the same good, *constrained misallocation* arises: even accounting for agents' binding constraints, a trade between two agents with different transaction prices can be welfare improving if done at an alternate transaction price. Misallocation manifests as differences in *constrained marginal rates of substitution* (CMRS), that is a marginal rate of substitution that accounts for the implied penalties associated with binding constraints that limit trades for agents.

Our first set of results characterizes allocative inefficiency in exchange economies with general distortions from misallocation and pecuniary externalities. We begin our analysis in Section 3 by defining the *allocative value of market $m$* as the marginal social value of a new producer entering the economy with a marginal unit of good $m$. We show that the allocative value of markets decomposes into three terms. The first is the average transaction price of the good, capturing the average value to

[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.

[2] See Financial Stability Board (2013) for a policy perspective.

agents in the economy of an extra unit of that good. The second is cross-sectional misallocation that arises when different agents have different constrained marginal rates of substitution. Intuitively, an extra unit of the good is most valuable when endogenous demand responses allocate it to agents with high CMRS. The third is pecuniary externalities that arise from the adjustment in equilibrium prices. In economies without pecuniary externalities and with symmetric distortions across agents, both misallocation and externalities are zero, and the allocative value of a market is its transaction price. This reflects a first welfare theorem logic: when distortions are zero and there are no externalities, markets allocate resources efficiently according to their price.

We next turn to the question of optimal regulation in this environment. A social planner maximizes social welfare by choosing wedges in transaction prices of agents. We begin in Section 4 with the benchmark of complete regulation: the planner is able to freely choose the wedge applied to every agent in every market spearately. In this environment, we show that optimal policy is to set the CMRS of every agent equal to the allocative value of a market. Doing so provides equal regulatory treatment across agents, and eliminates cross-sectional misallocation. At the same time, it also allows the planner to separate out transaction prices from market prices through appropriate scale of regulation. In doing so, the planner is able to set average pecuniary externalities across agents to zero. As a result, optimal complete regulation allows the planner to face no trade-off between misallocation and externalities.

We then turn to our next main set of results in Section 5: characterizing optimal policy with incomplete regulation. We begin by studying the case where the planner can only freely choose wedges for a subset of regulated agents, while taking the (possibly nonzero) distortions on other agents as exogenously specified. We show that as in the case of complete regulation, optimal regulation again ensures that the CMRS of regulated agents is equal to the allocative value of a market. However, incomplete regulation results in misallocation: there is both average misallocation between regulated and unregulated agents, and also misallocation within unregulated agents. Optimal policy encodes a targeting rule that trades off misallocation against pecuniary externalities.

Our results give a key role to *regulatory arbitrage* by unregulated agents in determining the *direction* of misallocation. Intuitively, misallocation in the targeting rule arises because a price change in a market that mitigates externalities, also induces changes in demand by unregulated agents. Changes in unregulated agent demand induce costs in proportion to misallocation. Intuitively, if a price increase induces a negative externality, under optimal policy that must be counteracted by a reduction in misallocation. In the targeting rule, reduction in misallocation can happen either because unregulated agents with low CMRS reduce demand, or unregulated agents with high CMRS increase demand in response to the price change.

Our model is rich enough to allow us to study partial regulation of unregulated agents, that is when the planner can freely choose taxes on a subset of their activities. We show that optimal policy can be divided into two blocks. In the first block, the planner takes as given the CMRS of regulated agents, and chooses partial regulation in order to minimize misallocation between partially and fully regulated agents. Then, the planner feeds these wedges into the targeting rule from before in order to derive optimal regulation of regulated agents, which trades off misallocation against externalities. Interestingly, this illustrates a division of the two ideas. Partial regulation targets misallocation relative to the completely regulated agents, whereas complete regulation targets the trade-off between misallocation and externalities.

Our theory is rich enough to shed light on a number of important policy questions. An important concern in practice is that the post-crisis financial 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.[3]

We leverage our framework to propose classification schemes for agents (such as unregulated finance). In our first exercise, we ask what is the social welfare impact of entry by a new unregulated player, such as a FinTech company. We show that the welfare impact of a new player in our model can be characterized by two components. The first component is the profits of the new company. The second is the social value of its activities in each market, which are simply the optimal tax applied to regulated agents in that market times its activities in the market. This provides a simple way for a regulator to evaluate the social value of a new entrant, by using the regulation of regulated agents as weights to evaluate the indirect social welfare benefits of its activities above and beyond its direct profits. This characterization is useful because it only requires knowledge of an agent's profits and its total activities in each market, and not knowledge of deeper objects such as its cost function, technology, and so on.

In our next exercises, we propose classification schemes for evaluating targets identity- and activity-based regulation. We evaluate the first-order welfare gains achievable from extending identity- and activity-based regulation to previously unregulated agents or activities. Surprisingly, the welfare benefits of new regulation is summarized by its impact on misallocation. We show in both cases that the welfare gains from new regulation are simply the product of the direct change in demand induced by the new regulation, times the cost of misallocation in agents in which demand is changed. New regulation is therefore most beneficial if it can induce increases (decreases) in demand of agents with high (low) CMRS relative to that of regulated agents.

Intuitively, one might expect that the agents and activities contributing most to regulatory arbitrage are also the most valuable targets for regulation. We show that this intuition is not necessarily correct. The reason is that regulatory arbitrage in the targeting rule for optimal policy

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

accounts for the full effects of a change in the market price, including the fact that the market price affects agents' binding constraints (pecuniary externalities). By contrast, new regulation only targets the purchase price of agents, and not directly binding constraints. Interestingly, this means that if an unregulated agent has strong demand responses due to binding constraints, that agent may not be a valuable target for new regulation.

Our final exercise evaluates the gains from extending new support programs to unregulated agents, such as LOLR, that boost their market price and so directly relax constraints. We show that the first order gains from a support program combine the direct effect of managing pecuniary externalities, with indirect effects that arise because changes in demand induced by the support program increase or reduce misallocation. Interestingly, this latter effect can be negative even when the pecuniary externality is positive, meaning the indirect effect can counter the direct effect. The intuition is that if a positive externality relaxes constraints and increases demand, then the targeting rule for optimal policy implies that the direction of misallocation has a higher CMRS for *regulated* agents than unregulated agents. Thus the indirect consequence of the support program is in fact to *increase* misallocation. This suggests a novel potential synergy between regulation and support programs: a support program targeting a regulate agent generates no misallocation precisely because the regulated agent's CMRS is equal to the allocative value of a market.

In Section 7 we apply our theory to two leading applications. Our first application studies regulatory classification 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 leading 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.

Finally in Section 8, we extend our framework to study models with multiple regulators and common agency. Environments with multiple regulators often feature incomplete regulation – for example, international regulatory environments typically feature countries having partial regulatory jurisdiction over foreign agents operating domestically. We show in this environment that many of the core ideas of our framework carry over, with modifications, to this setting. The allocative value of a market to an individual regulator is the same allocative value, but netting out taxes applied by other regulators. Intuitively, each regulator perceives the taxes of other regulators as a part of allocative value that does not accrue to them. We then characterize optimal incomplete regulation. We show the CMRS of regulated agents are equal to the allocative value of a market to their regulator, plus taxes of all other regulators. This results in a similar targeting rule and partial regulation targeting rule as in the baseline model.

**Related literature.** Our theoretical results build on two literatures, namely those on (i) misallocation in disaggregated economies, and (ii) corrective instruments and regulation. We apply our theory to policy questions on unregulated finance, in particular in the context of shadow banks and capital flows.

*Misallocation.* The study of misallocation has a long history in economics, tracing its origin from Dupuit (1844) and Jenkin (1872) to Marshall (1890) and later Harberger (1964). A recent wave of papers has renewed interest in misallocation and potential gains from allocative efficiency with contributions, among many others, from Restuccia and Rogerson (2008), Hsieh and Klenow (2009), Jones (2013), Midrigan and Xu (2014), Bigio and La'O (2020), Liu (2019), Baqaee and Farhi (2020), and Dávila and Schaab (2022). Relative to these papers, we focus on the interplay between misallocation and optimal regulation in economies with externalities. The paper closest to ours is Liu (2019) who studies the distortionary effects of market imperfections in an input-output production network. Liu (2019)'s result that distortions in sectoral size are a sufficient statistic for the social value and welfare relevance of that sector are similar in spirit to our results characterizing the allocative value of markets. Liu (2019) is part of a broader literature concerned with aggregation in disaggregated economies.[4] While much of this literature is concerned with positive aggregation in disaggregated production networks with input-output links, our focus is on welfare aggregation in exchange (market) economies. We deliberately abstract from bilateral customer-supplier relationships, recognizing that important parts of the modern economic system are governed by market-based instead of relationship-based interactions. In particular, our focus on market networks seems particularly appropriate in the context of studying financial markets and their regulation.

Our characterization of deviations from allocative efficiency due to pecuniary externalities and misallocation is similar to Baqaee and Farhi (2020), who show that Hulten's theorem breaks down in inefficient production economies, with the impact of technology shocks now comprising both a direct effect and an indirect effect through changes in allocative efficiency. Unlike Baqaee

[4] Hulten (1978) shows that, in efficient economies, the effect of sectoral technology shocks on aggregate activity is proportional to that sector's sales share. Subsequent work has emphasized that aggregation in richer environments, for example in inefficient economies, requires additional information about the microeconomic details of the economic environment (Baqaee and Farhi, 2019, 2020; Bigio and La'O, 2020; Liu, 2019).

and Farhi (2020), however, our environment allows for both arbitrary exogenous distortions and pecuniary externalities that take the form of endogenous wedges. In particular, we show that in the presence of market failures the network structure and interconnectedness of markets becomes a key determinant of allocative value and potential gains from changes in allocative efficiency.

*Corrective instruments.* A growing literature studies incomplete corrective regulation, particularly in the context of unregulated finance.[5] Closest to us in this literature is the independent and contemporaneous work of Dávila and Walther (2022), who study second-best corrective policy with imperfect instruments. 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 characterizes externalities and misallocation that arise in general equilibrium with complete and incomplete regulation.[6] We use our characterizations to develop a classification scheme for new entrants and new regulations based on simple statistics arising from our model, and relate these classifications concretely to misallocation.

The leading applications of our framework address policy questions on unregulated finance, and so relate in particular to the growing normative literature studying the consequences of unregulated institutions/capital flows for different forms of regulation. These contributions include studying impacts on capital requirements/debt taxes (Plantin, 2014; Huang, 2018; Martinez-Miera and Repullo, 2019; Bengui and Bianchi, 2019; Begenau and Landvoigt, 2020), liquidity regulation (Grochulski and Zhang, 2019), and reserve requirements (He et al., 2018). Ordoñez (2018) and Farhi and Tirole (2021) study incentives for a bank to choose to become regulated, with the former proposing explicit subsidies and the latter proposing pairing regulation with fiscal backstops. Our classification scheme informs the social value of new entrants and new regulation, and in conjunction with a large positive literature on determinants of shadow banking and capital flows can provide guidance to regulators on how to think about welfare in a complex financial system.[7]

## 2 Model

Our economy is populated by a unit continuum of agents indexed by a finite set of types $i \in \mathcal{I}$, each of measure $\mu_i$. Agents trade with each other in markets for $M$ goods. We denote by $I_{im}$ agent

[5] In addition to the incomplete regulation literature, a large outstanding literature studies regulation with pecuniary externalities, such as Caballero and Krishnamurthy 2001, Bianchi 2011, Bianchi et al. 2018, Dávila and Korinek 2018, Farhi et al. 2009, and Lorenzoni 2008.

[6] Our focus on the interplay between regulation, externalities, and misallocation also has similarities to Hebert (2022), who show how arbitrage opportunities arise because regulation of intermediaries to mitigate externalities distorts risk sharing between intermediaries and households.

[7] The literature on positive determinants of shadow banking include (Acharya et al., 2013; Claessens et al., 2012; Gorton et al., 2010; Buchak et al., 2018; Moreira and Savov, 2017; Chretien and Lyonnet, 2020; Coppola, 2021), while the positive literature on determinants of capital flows include (Avdjiev et al. 2018, Shen 2019, Milesi-Ferretti and Tille 2011, Forbes and Warnock 2012, Broner et al. 2013, Caballero and Simsek 2019,Maggiori et al. 2019, De Marco et al. 2019, Coppola et al. 2019).

$i$'s trade in market $m$, and refer to $I_{im} > 0$ (or $I_{im} < 0$) as a purchase (or sale) of good $m$. The vector $I_i = \{I_{i1}, \ldots, I_{iM}\}$ summarizes $i$'s trades in all $M$ markets. Goods in market $m$ have a *market price* $q_m$, and we denote $q = \{q_1, \ldots, q_M\}$ to be the vector of market prices. Taking good 1 as our numeraire, we adopt the normalization $q_1 = 1$. Each agent is endowed with tradeable wealth $w_i$.

**Agent $i$'s decision problem.** Agent $i$'s preferences over goods trades are ordered according to a utility function $U_i(I_i)$. Agents face distortions (wedges) $\tau_i = \{\tau_{i1}, \ldots, \tau_{iM}\}$ when trading with each other, so that $i$'s effective *transaction price* in market $m$ is $p_{im} = q_m + \tau_{im}$, with $p_i$ denoting the vector of transaction prices across markets. We take $\tau_i$ as exogenous for now and will later treat it as a regulatory instrument of a planner. Without loss of generality, we assume the market for the numeraire good 1 to be undistorted, with $\tau_{i1} = 0$ for all $i$.[8]

Denoting by $w_i$ the tradable wealth of agent $i$, her budget constraint takes the standard form

$$p_i I_i \leq w_i, \tag{1}$$

where $p_i I_i = \sum_m p_{im} I_{im}$ denotes total expenditures. Agent $i$ additionally faces a set of restrictions on her trades that may depend on market prices $q$, taking the form

$$\Gamma_i(I_i, q) \leq 0. \tag{2}$$

The constraint set embeds constraints such as technological constraints (production technologies) and collateral constraints.

The decision problem of agent $i$ is therefore to maximize her utility, subject to constraints (1) and (2). We denote by $I_i(p_i, q, w_i)$ agent $i$'s Marshallian demand over goods trades in $M$ markets when she faces transaction prices $p_i$, market prices $q$, and has tradable wealth $w_i$. Her Marshallian demand is given by

$$I_i(p_i, q, w_i) = \arg\max_{I_i} \; U_i(I_i) \quad \text{s.t.} \quad p_i I_i \leq w_i, \quad \Gamma_i(I_i, q) \leq 0. \tag{3}$$

Finally, we denote $i$'s indirect utility function by

$$V_i(p_i, q, w_i) = U_i(I_i(p_i, q, w_i)). \tag{4}$$

**Market clearing and competitive equilibrium.** Traded goods are in zero net supply in the economy. Market clearing therefore requires that $\sum_i \mu_i I_{im} = 0$ for all $m$, or in vector form

$$\sum_i \mu_i I_i = 0. \tag{5}$$

[8] We can re-represent distortions in the numeraire as a different vector of (relative) price distortions and wealth levels. Formally for any $p_{i1} \neq 1$, define $\hat{p}_i = \frac{p_i}{p_{i1}}$ and $\hat{w}_i = \frac{w_i}{p_{i1}}$. The problem over $(\hat{p}_i, \hat{w}_i)$ then features an undistorted numeraire.

Note that market clearing sums to zero because we have defined $I$ as trades and directly encoded any endowments of goods into the utility or constraint set of agents.

We can now define competitive equilibrium in our economy.

**Definition 1.** Given $\{\tau_i, w_i\}_{i\in\mathcal{I}}$, a competitive equilibrium is market prices $q$ and allocations $\{I_i\}_{i\in\mathcal{I}}$ such that: (i) agents optimize, i.e., equation (3) holds; and (ii) markets clear, i.e., equation (5) holds.

### 2.1 Constrained Marginal Rates of Substitution and Misallocation

The marginal rate of substitution (MRS) of agent $i$ between two goods, $m$ and $n$, is defined as $MRS_i^{mn} = \frac{\partial U_i/\partial I_{im}}{\partial U_i/\partial I_{in}}$, which in economies without binding collateral constraints is equal to the relative transaction price of the two goods, $p_{im}/p_{in}$. Given that our economy features constraints at the agent level, the MRS is no longer generally equal to the relative purchase price of goods. This reflects that an agent with a high MRS may not be able to capitalize on that MRS due to binding constraints.

We now define a notion of marginal rate of substitution that accounts for binding constraints. Denoting agent $i$'s Lagrangian as $\mathcal{L}_i = U_i(I_i) + \lambda_i[w_i - p_i I_i] - \Lambda_i \Gamma_i(I_i, q)$, we can define agent $i$'s *constrained utility* as $\mathcal{U}_i(I_i, q) = U_i(I_i) - \Lambda_i \Gamma_i(I_i, q)$, which encodes the constraint set as a utility penalty. We define the *constrained marginal rate of substitution (CMRS)* as the marginal rate of substitution of constrained utility,

$$CMRS_i^{mn} = \frac{\partial \mathcal{U}_i/\partial I_{im}}{\partial \mathcal{U}_i/\partial I_{in}} \tag{6}$$

Intuitively, CMRS is a marginal rate of substitution that accounts for the implied penalty from binding constraints: even if an agent has a high MRS, a constraint may prevent that agent from capitalizing on the MRS. CMRS is equal to MRS for agents that do not face binding constraints (i.e., $\mathcal{U}_i = U_i$). From the Lagrangian we obtain

$$CMRS_i^{mn} = \frac{p_{im}}{p_{in}},$$

that is an optimizing agent has a CMRS equal to the relative transaction price between the two goods.

A particular CMRS of interest in our framework is the CMRS between good $m$ and the numeraire, which we define as

$$\chi_i = CMRS_i^{m1}. \tag{7}$$

Since the numeraire is undistorted, we have $\chi_i = p_i$. We henceforth will use CMRS to refer to $\chi_i$.

**Misallocation.** Our results to come will give a central role to misallocation, that is, when one agent has a high value for a good relative to another agent. Given binding constraints, misallocation in our model is misallocation that can be corrected without violating a binding constraint. This notion of misallocation in our setting is reflected in different CMRS across agents. Intuitively if $\chi_{im} > \chi_{jm}$ for two agents, then agent $i$ places a higher value on good $m$ (relative to the numeraire) in constrained utility. Thus there is a welfare-improving transaction between $i$ and $j$, formalized in the following Lemma.

**Lemma 2** (Misallocation and CMRS)**.** *If $\chi_{im} > \chi_{jm}$, then a marginal transaction between i and j whereby m sells $\hat{p}_m \in (\chi_{jm}, \chi_{im})$ units of the numeraire to j in exchange for a unit of good m improves the welfare of both i and j.*

Lemma 2 formalizes a simple notion of misallocation in our setting, that is the existence of a transaction that increases constrained utility. Misallocation arises when two different agents have different CMRS for the same good. In this case, the agent with lower CMRS should sell the good to the agent with higher CMRS, at a transaction price intermediate to the two.

## 2.2 Welfare Impact of Price Changes

We define aggregate welfare in terms of the social welfare function

$$W(q, \tau, w) = \sum_i \mu_i \omega_i V_i(q + \tau_i, q, w_i), \tag{8}$$

where $\omega_i$ is an individual-specific Pareto weight and $w = \{w_1, \ldots, w_I\}$ is the vector of agents' tradable wealth.

A key focus of this paper will be the welfare impact of changes in a market price $q_m$. The following (reduced-form) result characterizes this social welfare impact.

**Lemma 3.** *The social welfare impact of a price change $dq_m$ in market m is given by*

$$\frac{\partial W}{\partial q_m} = \sum_i \mu_i \omega_i \left[ -\lambda_i I_{im} + \frac{\partial \mathcal{U}_i}{\partial q_m} \right], \tag{9}$$

*where recall that $\lambda_i > 0$ is the Lagrange multiplier on the budget constraint and $\mathcal{U}_i$ is constrained utility.*

Lemma 3 characterizes the spillovers that arise from a price change $dq(m)$ in market $m$. The first term captures a motivation for redistribution: a price increase redistributes wealth from buyers to sellers. It will be expositionally convenient to rule out pure redistribution by selecting relative weights $\omega_i \lambda_i = \omega_j \lambda_j$ for all $i, j$, so that $\sum_i \mu_i \omega_i \lambda_i I_{im} = 0 = \sum_i \mu_i I_{im} = 0$, although we can obtain generalized versions of our results that account for redistribution absent this. The second term captures (non-redistributive) pecuniary externalities that arise because prices appear in constraints, and hence in constrained utility. These externalities arise because a price increase alters the feasible set of allocations when prices appear in constraints. Note that $\frac{\partial W}{\partial q_n}$ is expressed in utils, but we will re-express it in wealth-equivalent in Section 3.

### 2.3 Exchange Economies as Networks

We now characterize key statistics on demand and price responses that will play important roles in the results we present below.

We define the matrix $\boldsymbol{D}_i = \frac{dI_i}{dq}$ to be the *demand response* of agent $i$ to a market price change $dq$. The $jk$th element of the demand response matrix is given by

$$\boldsymbol{D}_i(j,k) = \frac{\partial I_{ij}}{\partial p_{ik}} + \frac{\partial I_{ij}}{\partial q_k},$$

where the total derivative encompasses the transaction and market price partial derivatives. For any subset $\mathcal{J} \subset \mathcal{I}$, we define the *aggregate demand response* of $\mathcal{J}$ by $\boldsymbol{D}_{\mathcal{J}} = \sum_{i \in \mathcal{J}} \mu_i \boldsymbol{D}_i$. Formally, $\boldsymbol{D}_{\mathcal{J}}(j,k)$ captures the total change in demand for good $j$ to a change in price of good $k$ by all agents $i \in \mathcal{J}$.

Next, we define the matrix $\boldsymbol{P} = \boldsymbol{D}_{\mathcal{I}}^{-1}$ to be the *price response matrix*. Formally, element $(j,k)$ of $\boldsymbol{P}$ measures how a change in total aggregate demand in market $k$ induces a change in the equilibrium price of market $j$ in order for markets to clear. $\boldsymbol{P}$ therefore provides a mapping from changes in demand, $dI$, to changes in prices, $dp$, given by $dp = \boldsymbol{P}' dI$.

## 3 Markets, Prices, and Allocative Efficiency

In this section, we develop a characterization of allocative inefficiency with general distortions, pecuniary externalities, and market network structures. Our results here take as given exogenously specified wedges $\tau_i$ in transaction prices that distort allocations but do not destroy wealth (Baqaee and Farhi, 2020; Liu, 2019).

### 3.1 The Allocative Value of Markets

Markets are efficient when the private competitive market equilibrium allocates resources to their socially most valuable uses. Our characterization of allocative efficiency below takes the perspective of a potential new producer entering the economy. This hypothetical producer evaluates her production decision based on the transaction prices she observes across markets, which determine the private marginal benefit she would obtain from market entry. Formally, suppose that a new producer entered market $m$ with $e_m$ units of good $m$. Her entry has two immediate consequences.

First, the condition for market clearing in market $m$ becomes $\sum_i \mu_i I_{im} = e_m$, accounting for the new resources brought to market. Second, total wealth in the economy upon distribution of the producer's surplus becomes

$$\sum_i \mu_i w_i \le q_m e_m + \sum_i \mu_i \sum_m \tau_{im} I_{im}(p_i, q, w_i). \tag{10}$$

According to equation (10), total tradable wealth of agents in the economy must equal the new endowment brought into the market—which has market value $q_m e_m$—plus revenues remitted from the distortionary wedges $\tau_i$ in transaction prices.[9]

It will be helpful to define the social welfare impact of price changes in wealth equivalent (rather than in utils). Denote $\mathcal{E}_m = \frac{1}{\lambda}\frac{\partial W}{\partial q_m}$, where $\lambda > 0$ is the social value of wealth, i.e., the Lagrange multiplier on (10). Further denote $\mathcal{E} = \{\mathcal{E}_1, \ldots, \mathcal{E}_M\}$.

We define the *allocative value* of market $m$ as the social marginal value (in wealth equivalent) of our hypothetical producer bringing a marginal unit of good $m$ to market. That is, the allocative value of market $m$ is

$$\vartheta_m = \frac{1}{\lambda}\frac{\partial W}{\partial e_m}\bigg|_{e_m=0}.$$

We refer to the vector $\vartheta = \{\vartheta_1, \ldots, \vartheta_M\}$ as the allocative value of markets. If all agents in the economy (as well as hypothetical new producers) faced transaction prices equal to $\vartheta$, then private and social marginal values would align and markets would efficiently allocate resources. We now state our first main result, which characterizes the allocative value of markets in inefficient economies.

**Proposition 4** (Allocative Value of Markets)**.** *The allocative value of markets in economies with general distortions and pecuniary externalities is*

$$\vartheta = \mathbb{E}_i(p_i) + \boldsymbol{P}\,\mathbb{C}ov_i(\boldsymbol{D}_i, \chi_i) + \boldsymbol{P}\,\mathcal{E}. \tag{11}$$

Proposition 4 characterizes the allocative value of a market and develops a three-way decomposition into a direct effect and two indirect effects. The direct value of a marginal increase in $e_m$ is the average price at which it is transacted, $\mathbb{E}_i(p_{im})$. The first of the two indirect terms identifies cross-sectional misallocation across market participants, which is proportional to $\mathbb{C}ov_i(\boldsymbol{D}_i, \chi_i)$. Finally, the second indirect effect summarizes the effect on pecuniary externalities. Both indirect effects together comprise a measure of the change in allocative efficiency.

[9] A useful illustrative case is symmetric distortions, $\tau_i = \tau$, as an illustrative special case of our framework. The resource constraint in this case becomes $\sum_i \mu_i w_i \le q_m e_m + \tau \sum_i \mu_i I_i = (q_m + \tau_m) e_m = p_m e_m$, using market clearing. Intuitively, this constraint tells us that the total (wealth) value of endowment $e_m$ is the transaction price $p_m$ at which agents privately value the endowment. This reflects the usual Walrasian logic, except that here the transaction price comprises the market price, $q_m$, plus the distortion, $\tau_m$, that is remitted lump-sum.

Cross-sectional misallocation arises in our model when different agents face different constrained marginal rates of substitution for goods, arising from different transaction prices. It is characterized by the covariance operator

$$\mathbb{C}ov_i(\boldsymbol{D}_i, \chi_i) \equiv \sum_i \mu_i \boldsymbol{D}_i \chi_i - \sum_i \mu_i \boldsymbol{D}_i \sum_i \mu_i \chi_i = \sum_i \mu_i \boldsymbol{D}_i (\chi_i - \mathbb{E}_i \chi_i).$$

Intuitively, $\chi_i - \mathbb{E}_i \chi_i$ represents the gap between $i$'s CMRS and the average CMRS in the market. When agent $i$ has a high (low) CMRS in a good relative to the market, there is social value to reallocating consumption in that good toward (away from) $i$. Reallocation is captured in the demand response $\boldsymbol{D}_i$, that is how much agent $i$ changes demand in response to a price change. From the perspective of cross-sectional misallocation, a price change is valuable when it increases (decreases) demand by agents with a high (low) CMRS relative to the market. Hence, a large covariance results in a greater impact on allocative value because it reflects greater ability for changes in market prices to elicit demand responses in agents with larger deviations in CMRS from the average. Finally, the pre-multipliation by $\boldsymbol{P}$ reflects the conversion from a new good in a market to a change in prices, giving the full chain of how allocative value of a market is impacted by cross-sectional heterogeneity in CMRS.

The second indirect effect comes from endogenous wedges that result from pecuniary externalities. First, a change in a price in market $m$ induces pecuniary externalities $\mathcal{E}_m$, as characterized in Lemma 3. Premultiplication by $\boldsymbol{P}$ characterizes the change in market prices that arise because the new producer changes total supply in market $m$. Intuitively, equilibrium prices must change in order for agents in the economy to accommodate that change in supply. These changes in market prices in turn generate pecuniary externalities. Pecuniary externalities thus give rise to an endogenous effect similar to misallocation from exogenous wedges. Exogenous wedges lead different agents to have different CMRS, and a price increase in valuable when it reallocates demand from agents with low CMRS to agents with high CMRS. In a similar vein, endogenous wedges lead some agents to have particularly binding constraints. A price increase is most valuable when it relaxes constraints of particularly constrained agents, that is it generates large pecuniary externalities.

To further unpack the economic forces that determine the allocative value of markets, we next discuss two illustrative special cases of Proposition 4.

**Corollary 5** (Efficient Markets)**.** *Absent distortions and pecuniary externalities, $\tau_i = \mathcal{E} = 0$, the allocative value of market $m$ is its transaction price, $\vartheta_m = p_m$.*

Corollary 5 characterizes the benchmark case of efficient markets and represents a form of the first welfare theorem in our environment. It establishes clearly the classical result that transaction prices determined in a competitive market equilibrium are efficient in allocating resources to those uses with highest social marginal value. In other words, the transaction price that our hypothetical producer faces when deciding whether to enter a market is reflective of the social marginal value. There is no market failure, and the price system allocates resources efficiently. As we discuss below, it is noteworthy that the allocative value of market $m$ in this case depends neither on the network structure and interconnectedness of markets nor on the identities of market participants, except through the transaction price. In other words, the transaction price alone becomes a sufficient statistic for the allocative value of a market.

It is worth mentioning that even without distortions or externalities, CMRS and MRS do not generally coincide. The reason is that absence of pecuniary externalities eliminates price-based constraints, but does not necessarily eliminate all constraints (such as technology constraints). If some agents are constrained, price mechanisms are constrained efficient in that CMRS is equalized, even though MRS is not equalized.

We next turn to a second illustrative special case of Proposition 4.

**Corollary 6** (Symmetric Distortions)**.** *Suppose that there are no pecuniary externalities, $\mathcal{E} = 0$, and distortions are symmetric across agents, $\tau_i = \tau$. Then the allocative value of market m is again its transaction price price, $\vartheta_m = p_m$.*

More generally, with symmetric distortions, $\tau_i = \tau$ for all $i$, we immediately see from Proposition 4 that cross-sectional misallocation vanishes and $\vartheta = p + \boldsymbol{P}\mathcal{E}$. When $\mathcal{E} = 0$ and there are no externalities, then we have $\vartheta_m = p_m$. This can be viewed as a classical case of a differentiation between before- and after-tax prices: given before-tax prices $q$ and taxes $\tau$, the prices all agents face are simply the after-tax prices $\tau$. In this environment where taxes are applied symmetrically to all agents and given taxes are rebated lump sum, the environment of Corollary 6 achieves the same efficient allocation as the efficient markets benchmark of Corollary 5 with an irrelevant tax-and-rebate scheme that leaves the final purchase prices of agents (equivalently, sale prices of producers) undistorted. This highlights the crucial property that absent pecuniary externalities, exogenous market distortions in this setting affect the allocative value of markets to the extent they generate *cross-sectional* misallocation. This is exactly the covariance term in Proposition 4.

**Welfare aggregation.** Proposition 4 can be interpreted as a *welfare aggregation result*. It characterizes how perturbations in a single market $m$ transmit to other markets and affect aggregate welfare both directly and indirectly.

Our result is in the same spirit as the positive aggregation results that have been the subject of the positive literature on misallocation and production networks. It is well understood, for example, that in efficient economies and to first order the impact of a sectoral technology shock on aggregate economic activity is proportional to the Domar weight of that sector (Hulten, 1978; Baqaee and Farhi, 2019). So does Corollary 5 establish in our setting that—to first order around the efficient allocation—the aggregate welfare effect of an endowment shock in market $m$ is proportional to that market's transaction price. Corollary 5 can therefore be interpreted as the analog of Hulten's theorem for welfare aggregation in our setting. And transaction prices serve the same role for welfare aggregation as Domar weights do for positive aggregation of production functions: the transaction price of market $m$ becomes a sufficient statistic for its allocative value and welfare relevance.

In their seminal contribution, Baqaee and Farhi (2020) show that Hulten's theorem breaks down in inefficient production economies with general exogenous wedges: the impact of sectoral technology shocks on aggregate economic activity now comprises both a direct effect and an indirect effect through changes in allocative efficiency. Proposition 3 represents the counterpart of their result for welfare aggregation in exchange economies: the welfare consequence of a shock to market $m$ comprises both a direct effect—proportional to the transaction price in market $m$—and indirect effects that represent gains and losses through allocative efficiency. Unlike in their paper, we allow for both arbitrary exogenous distortions and pecuniary externalities that represent endogenous wedges. In much of Sections 4 and 5 below, we focus on characterizing the implications of these two determinants of allocative efficiency.

**Importance of network structure.** In the presence of market failures, the network structure of markets becomes a key determinant of allocative value and potential gains from changes in allocative efficiency. This observation again echoes similar insights from the literature on positive production aggregation. In efficient economies and to a first order, Domar weights are sufficient statistics for aggregation. In other words, the aggregate consequences of sectoral shocks can be summarized using information about only that sector. So here can the aggregate welfare impact of a shock to market $m$ be summarized using only that market's transaction price according to Corollary 5. Information about neither the network structure and interconnectedness of markets nor the identities of market participants is required to determine and aggregate the welfare effects of shocks to market $m$. In inefficient exchange economies, however, welfare aggregation requires information about the network structure and interconnectedness of markets according to Proposition 3. In particular, the allocative value of markets depends on both $\boldsymbol{P}$ and the distribution of $\boldsymbol{D}_i$.

### 3.2 Welfare Implications of Market Transactions

Proposition 4 characterizes the allocative value of a market in terms of a potential new producer. We now characterize the allocative value of a *potential new trader*. A potential new trader shows up to the market and sells $e_m$ units of good $m$ in exchange for $q_m e_m$ units of the numeraire, where $q_m$ is the market price. Note that this trade creates no new wealth, but alters market clearing. We denote $\theta_m = \frac{\partial W}{\partial t_m}$ to be the *allocative value* of a trade $t_m$ of selling in market $m$ to purchase the numeraire. Note that these trades form a basis for all other possible trades, a point we revisit after characterizing $\theta$. We obtain the following result.

**Proposition 7** (Allocative Value of Trades)**.** *The allocative value of trades is given by*

$$\theta = \underbrace{\mathbb{E}_i(p_i) - q}_{\text{Average Distortion}} + \boldsymbol{P}\,\mathbb{C}ov_i(\boldsymbol{D}_i, \chi_i) + \boldsymbol{P}\,\mathcal{E}. \tag{12}$$

Proposition 7 characterizes the allocative value of a zero-wealth-position trade of selling in market $m$ to buy the numeraire. Intuitively, a zero-wealth-position trade generates the same indirect effects as a new seller in market $m$, as it increases supply in market $m$ which forces a change in equilibrium prices.

Unlike a new good, the market value of a trade generates a direct value equal to the average distortion of transaction prices from the market price, $\mathbb{E}_i p_i - q$. This means that the direct value of a trade can be positive if there are positive average distortions, $\mathbb{E}_i p_i - q > 0$, but negative if there are negative average distortions, $\mathbb{E}_i p_i - q < 0$. Intuitively if $\mathbb{E}_i p_{im} > q_m$, then the average positive distortion means agents on average value good $m$ more than its market price. Therefore, there is positive value to a new trader selling that good to agents at the market price $q$ in exchange for the numeraire. Conversely if $\mathbb{E}_i p_{im} < q_m$, then agents value the good less than the market price, and hence there is a loss from a new seller showing up and forcing agents to purchase more of it.

Whether or not there are distortions or externalities, the trade produces responses in equilibrium prices. If there are no distortions or externalities, these price changes have no net welfare consequence. However when there are distortions or externalities, the trade has indirect welfare effects through the equilibrium price changes. These indirect welfare consequences are precisely the indirect terms of the allocative value of a potential producer from Proposition 4.

Finally, it is helpful to show how the allocative value of trades form a basis across markets. First, we know that the allocative value of a trade is simply the allocative value of a market plus the market price, that is $\vartheta = q + \theta$. Second, we know that we can construct a zero-wealth trade by selling $e_m$ units of good $m$, buying $q_m e_m$ units of the numeraire, and then selling $q_m e_m$ units of the numeraire to buy $\frac{q_m}{q_n} e_m$ units of good $n$. Thus, the allocative value of a trade selling in market $m$ and buying in market $n$, which we denote $\theta_{m,n}$, is given by $\theta_{m,n} = \vartheta_m - \frac{q_m}{q_n}\vartheta_n$. Intuitively, trading between two markets gives a direct effect $q_m\left(\mathbb{E}_i \frac{p_{im}}{q_m} - \mathbb{E}_i \frac{p_{in}}{q_n}\right)$, which simply measures the difference in how much the purchase price deviates from the market price between the two markets.

## 4 Complete Regulation

We now turn to the question of optimal regulation. A planner who maximizes social welfare chooses the wedges $\tau$, which we treated as exogenously given in Section 3. We begin with the benchmark of complete regulation: the planner can freely set an agent-specific wedge $\tau_{im}$ on every agent $i$ for every market $m$. Our main result in this section is that optimal regulation with complete instruments sets wedges equal to the allocative value of trades in each market. Optimal policy therefore features equal treatment, i.e., $\tau_{im} = \tau_m$ and all agents face the same taxes. No tradeoff emerges between misallocation and pecuniary externalities when the planner has complete instruments.

Formally, our planner chooses complete regulation $\tau$ in order to maximize social welfare, $W$. The planner takes as constraints the resource constraint (10), the Marshallian demand functions of agents $I_i(p_i, q, w_i)$, and that market prices $q$ must be consistent with market clearing. We obtain the following characterization of optimal regulation in this environment.

**Proposition 8** (Optimal Regulation with Complete Instruments)**.** *With complete regulation:*

1. *The CMRS $\chi_i = p_i$ of all agents is equal to the allocative value of a market,*

$$\chi_i = \vartheta.$$

2. *Optimal regulation sets uniform taxes across agents that are equal to the allocative value of a trade, that is,*

$$\tau_i = \theta.$$

3. *Both cross-sectional misallocation and average pecuniary externalities are zero, that is*

$$\mathbb{C}ov_i\left(\boldsymbol{D}_i, \chi_i\right) = \mathcal{E} = 0.$$

Proposition 8 characterizes how complete regulation implements an allocation with no cross-sectional misallocation and no average pecuniary externalities. As a result, transaction prices are equal to the allocative value of markets, $\vartheta = p^*$, in accordance with the symmetric distortions environment of Corollary 6. The efficient transaction price therefore reflects the correct notion of allocative efficiency. Correspondingly, the CMRS of every agent is set equal to this price vector, $\chi_i = p^*$.

To achieve both no misallocation and no average externalities, the planner needs to separate the transaction price from the market price. Intuitively, the transaction price $p$ reflects allocative efficiency, whereas the market price $q$ slackens collateral constraints. Separating these two objects involves setting a tax equal to the difference, that is $\tau = p - q$. Moreover because there is no misallocation and no externalities in this setting, the allocative value of a trade is also solely the direct effect, $\theta = p - q$. Therefore, the optimal tax $\tau$ is also equal to the allocative value of a trade.

Proposition 4 and Corollary 5 make clear that differences in CMRS across agents lead to misallocation. From Lemma 2, we know that misallocation leads to existence of privately efficient transactions between agents for some transaction price. As a result, under complete regulation the planner seeks to eliminate cross-sectional misallocation. The planner does so by imposing *equal* *treatment*: the efficient tax $\tau_i$ is constant across all $i$.[10] Equal treatment in this environment prevents misallocation and contributes to efficiency.

In sum, Proposition 8 reflects the key insight that optimal regulation with complete instruments induces no trade-off between misallocation and pecuniary externalities. This is because with complete regulation, the planner is able to equalize CMRS across agents (equal treatment), eliminating misallocation. The planner can then use the average level of distortions to separately control transaction and market prices, eliminating pecuniary externalities while maintaining allocative efficiency.

## 5 Incomplete Regulation

Section 4 studies the benchmark of complete regulation and shows that (i) optimal taxes are equal to the allocative value of a trade, and (ii) no trade-off emerges between correcting externalities and cross-sectional misallocation. In this section, we study incomplete regulation. We start with a case where the planner has control over wedges for only a subset of agents, taking as given wedges on all remaining agents, and then allow for more general partial regulation over these latter agents.[11]

We develop our main result on incomplete regulation in Section 5.1: Optimal taxes are still equal to the allocative value of trades. The insight that allocative value is a sufficient statistic for optimal regulation is therefore robust to the (in)completeness of policy instruments. With incomplete instruments, however, a tradeoff now emerges between correcting externalities and cross-sectional misallocation: regulation results in different transaction prices between regulated and unregulated agents, leading to misallocation between the regulated and unregulated. We show that this gives rise to a targeting rule for optimal regulation, which reflects the planner's balancing of misallocation against externalities.

We then use our results to unpack regulatory arbitrage by the unregulated, and study how it impacts the direction of misallocation, i.e., whether regulated or unregulated agents have higher CMRS.

Finally, we study partial regulation of agents and markets – for example, introducing a uniform tax on a market or regulating a subsidiary. We show that optimal regulation is determined by a two step process: first, the planner uses partial instruuments to reduce misallocation between the regulated and a partially regulated agent; then, the planner uses the targeting rule of equation (13) to trade off misallocation against pecuniary externalities.[12]

[10] Equal regulatory treatment has been emphasized in policy debates on both bank regulation and capital flow management. 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).

[11] Although our paper focuses on pecuniary externalities, we provide a brief characterization of optimal incomplete regulation with non-pecuniary externalities in Appendix B.

[12] Note that the results on partial regulation apply even if no agent is subject to complete regulation. In this case, we can interpret the two step process as corresponding to correcting misallocation relative to a *hypothetical* fully regulated

### 5.1 Optimal Regulation

Suppose there is now a subset $S \subset \mathcal{I}$ of agents for whom the planner takes $\tau_i$ as exogenously specified, while maintaining the ability to freely choose $\tau_i$ for $i \notin S$. We think of agents $i \in S$ as, for example, shadow banking institutions or unregulated capital flows, and develop applications to both these settings in Section 7. The social planning problem is otherwise identical to Section 4, up to the reduction in available instruments. The following result characterizes the properties of the social optimum in this setting.

**Proposition 9** (Optimal Incomplete Regulation)**.** *With incomplete regulation:*

1. *The CMRS $\chi_i = p_i$ of regulated agents is equal to the allocative value of a market,*

$$\chi_i = \vartheta \quad \forall i \notin \mathcal{S}.$$

2. *Regulated agents face uniform taxes equal to the allocative value of a trade, that is*

$$\tau_i = \theta \quad \forall i \notin \mathcal{S},$$

3. *The social optimum encodes a targeting rule:*

$$0 = \underbrace{\overbrace{\boldsymbol{D}_S}^{\textit{Amount of Misallocation}} \times \overbrace{\mathbb{E}_{i \in S}(\chi_i - \vartheta)}^{\textit{Cost of Misallocation}}}_{\text{Misallocation Between Regulated/Unregulated}} + \underbrace{\mu_S \mathbb{C}ov_{i \in S}\left(\boldsymbol{D}_i, \chi_i - \vartheta\right)}_{\text{Misallocation Within Unregulated}} + \mathcal{E}. \tag{13}$$

Proposition 9 shows that optimal regulation equalizes the constrained marginal rate of substitution of regulated agents with the allocative value of a market, that is $\chi_i = \vartheta$. It achieves this by setting the taxes $\tau$ equal to the allocative value of a trade, $\theta$. This parallels the characterization of complete regulation in Proposition 8, that is the same fundamental objects determine optimal regulation. Equation (13) provides a "targeting rule" for optimal policy that balances the indirect effects of a change in market prices $q$. This targeting rule trades off the costs of misallocation against the costs of pecuniary externalities.

The first two elements of the targeting rule correspond to misallocation that arises due to incomplete regulation. We showed in Section 4 that under complete regulation the constrained marginal rates of substitution are equalized across all agents. This equal treatment property of complete regulation ensured there was no misallocation. Equation (13) thus recovers Proposition 8 under complete regulation: if all agents are regulation, then CMRS is equalized across agents, there is no misallocation, and hence the targeting rule specifies $\mathcal{E} = 0$.

agent, and then choosing complete regulation of that hypothetical agent to trade off misallocation against externalities.

Absent complete regulation, two terms arise due to misallocation. First, there is *between* (average) misallocation between regulated and unregulated agents. Between misallocation arises because the CMRS of regulated agents is $\chi_i = \vartheta$, but the CMRS of unregualted agents is not. For example if unregulated agents are fully undistorted, $\tau_i = 0$ for $i \in S$, then their CMRS is $\chi_i = q$ for $i \in S$. Between misallocation is reflected by $\mathbb{E}_{i \in S}(\chi_i - \vartheta)$. Intuitively if $\mathbb{E}_{i \in S}(\chi_i - \vartheta) < 0$, then regulated agents have a higher CMRS for good $m$ than regulated agents, and misallocation costs arise if a unit of demand for good $m$ is shifted from regulated agents to unregulated agents. The cost of this shift in demand is precisely the gap in CMRS, that is $\mathbb{E}_{i \in S}(\chi_i - \vartheta)$. Conversely is $\mathbb{E}_{i \in S}(\chi_i - \vartheta) > 0$, then on average regulated agents have a lower CMRS than regulated agents, and misallocation dictates there is value to shifting a unit of demand from unregulated agents to regulated agents.

While $\mathbb{E}_{i \in S}(\chi_i - \vartheta)$ encodes the cost of misallocation, $\boldsymbol{D}_S$ encodes the *amount* of misallocation that arises from a change in market prices $q$. Intuitively, it captures the total change in demand of unregulated agents $i \in S$ in response to price changes. Thus, the average cost of misallocation from a price change is simply the amount of misallocation times the cost of misallocation.

The second term that arises is cross-sectional misallocation within unregulated agents that arises when unregulated agents have different CMRS. In many environments, this term is in fact equal to zero. For example if agents $i \in S$ are fully unregulated ($\tau_i = 0$) or are subject to the same suboptimal regulation ($\tau_i = \underline{\tau} \neq \theta$), then this second effect is zero. To the extent it is nonzero, its intuition reflects that of the first term: misallocation effects are particularly potent when misallocation is concentrated in unregulated agents with particularly large deviations of CMRS from $\vartheta$.

In sum, the targeting rule (13) reflects the trade-off the planner faces between misallocation and pecuniary externalities. We now use this targeting rule to answer the important question: how does regulatory arbitrage affect the nature of misallocation in the optimum.

### 5.2 Regulatory Arbitrage and Direction of Misallocation

We no look to shed light on the important question of the *direction* of misallocation. In particular, we seek to understand whether regulated or unregulated agents have higher constrained marginal rates of substitution under optimal policy. This question is nontrivial because the targeting rule for misallocation depends not only on the sign of pecuniary externalities, but also the demand response of unregulated agents to changes in prices. This demand response of unregulated agents to prices has a natural interpretation as *regulatory arbitrage* in our setting: when a planner implements changes in market prices through a change in regulation, $\boldsymbol{D}_i$ for $i \in S$ captures how unregulated agents change demand in response to those market prices. Thus, the question of direction of misallocation is analogously a question of the form regulatory arbitrage takes.

Formally, we can tackle this question by studying the targeting rule (13). Let us suppose for simplicitly that there is no within (unregulated) misallocation, so that the targeting rule specifies

$$\boldsymbol{D}_S(\chi_S - \vartheta) + \mathcal{E} = 0,$$

where we have denoted $\chi_S = \mathbb{E}_{i\in S}\chi_i$ for notational compactness. To build intuition, consider the own-price component term of the targeting rule, $D_{S,mm}(\chi_{Sm} - \vartheta_m) + \mathcal{E}_m$. For expositional purposes, we will focus discussion on the case of a negative externality $\mathcal{E}_m < 0$, with all discussion holding with sign reversed in the case of a positive externality.

Suppose first that the own price demand response is negative, $D_{s,mm} < 0$. This means that on average, unregulated agents respond to an increase in price $q_m$ by reducing their demand for good $m$. If the direction of misallocation were $\chi_{im} - \vartheta_m > 0$, that is unregulated agents have higher CMRS, then $\boldsymbol{D}_{S,mm}(\chi_{Sm} - \vartheta_m) < 0$, and both direct components of the targeting rule are negative. This pushes for a direction of misallocation $\chi_{Sm} - \vartheta_m < 0$, that is regulated agents have high CMRS relative to unregulated agents. Economically, negative pecuniary externalities give value to a tax for that good, which depresses the market price by discouraging regulated agents from purcashing it. This induces regulatory arbitrage in the sense that unregulated agents with $D_{S,mm} < 0$ respond by increasing demand in response to the positive price. Because unregulated agents have a low CMRS relative to regulated agents, this regulatory arbitrage exacerbates misallocation. Note that if the externality were instead positive, $\mathcal{E}_m > 0$, then the direction of misallocation goes in the opposite direction: regulated agents are subsidized to inflate the market price, and unregulated agents reduce demand and counteract the price increase.

Suppose, alternatively, that the own price demand response is positive, $D_{S,mm} > 0$, and so an increase in market price $q_m$ on average leads unregulated agents to *increase* demand. Following the same logic, this is a force for the direction of misallocation to be $\chi_{Sm} - \vartheta_m > 0$, which in the targeting rule counteracts the negative externality. In this case, the regulatory arbitrage of *increasing* demand in response to marker price increases leads regulated agents to have *lower* CMRS than unregulated agents, despite the negative externality.

The above logic reflects that the sign of the regulatory arbitrage response is crucial for determining the direction of misallocation. It is intuitive to conjecture that an unregulated agent $i$ has a negative own-price demand response, $\boldsymbol{D}_{i,mm} < 0$: an increase in the price $q_m$ reduces demand for good $m$ owing to a classical substitution effect. This would imply that a positive pecuniary externality leads to misallocation costs because regulated agents have a low CMRS relative to unregulated agents. We now do a careful decomposition of the demand response of an unregulated agent, and show that models of pecuniary externalities naturally give rise to a counterveiling force that motivates a *postive* demand response $\boldsymbol{D}_{i,mm} > 0$ to price increase, rather than a negative response. This changes the direction of misallocation and implies that regulated agents have a *low* CMRS relative to unregulated agents. The intuition is that positive pecuniary externalities slacken the constraint sets of agents, implying their choice set expands and letting them increase demand (despite the higher transaction cost). We provide a specific example of this form after presenting the general result.

We focus on unpacking the own-price response, $\boldsymbol{D}_{i,mm}$, to determine its sign. To do so, we employ tools from price theory to give a decomposition into income and substitution effects. Our decomposition differs from the canonical one because prices appear not only in the budget constraint (transaction price), but also in the constraint set (market price). We define Hicksian demand $h_i(p_i, q, \overline{U}_i)$ and the expenditure function $e_i(p_i, q, \overline{U}_i)$ from the expenditure minimization problem.[13] From here, we obtain the formal decomposition of $\boldsymbol{D}_{i,mm}$.

**Proposition 10.** *A price-theory decomposition of agent i's demand in market m yields*

$$\boldsymbol{D}_{i,mm} = \underbrace{\frac{\partial h_{im}}{\partial p_{im}}}_{\substack{\text{Standard Substitution Effect} \\ <0}} + \underbrace{\frac{\partial h_{im}}{\partial q_m}}_{\substack{\text{Constraint Set Substitution Effect} \\ \lesseqgtr 0}} + \underbrace{\frac{\partial I_{im}}{\partial w_i}\mathcal{E}_i}_{\substack{\text{Total Income Effect} \\ \lesseqgtr 0}} \tag{14}$$

*where* $\mathcal{E}_i = -h_i + \frac{1}{\lambda_i}\nabla_q \mathcal{U}_i$ *is the vector of pecuniary externalities on agent i.*

Proposition 10 tells us that the sign of $\boldsymbol{D}_{i,mm}$ generally depends on a combination of income and substitution effects. The right-hand side of equation (14) consists of three terms. The first term, $\frac{\partial h_{im}}{\partial p_{im}}$, is the standard substitution effect in the transaction price. This term is negative, reflecting the usual effect that a compensated increase in a good's own price shifts demand away from that good. This term leads regulatory arbitrage to cause positive (negative) pecuniary externalities to be associated with regulated agents having lower (higher) CRMS than unregulated agents.

However, there are two additional forces. The first is the substitution effect in the collateral price. Unlike the transaction price substitution effect, the collateral price substitution effect cannot be signed in general. However, it can be positive in natural models of pecuniary externalities. For example, an increase in the collateral price may encourage unregulated agents to sell less of an asset if debt-back rollover is tied to the market value of collateral, resulting in $\frac{\partial h_{im}}{\partial q_m} > 0$. In this case, the collateral price substitution effect pushes misallocation in the *opposite* direction: positive (negative) pecuniary externalities push for regulated agents to have a *higher* (*lower*) CMRS than unregulated agents.

Finally, there is also an income effect. Absent prices in constraints, the traditional income effect is based on demand $I_{im}$: an increase in the transaction price reduces the effective wealth of agent $i$ because it costs more to purchase the same amount $I_{im}$. With prices in constraints, the income effect

[13] The expenditure minimization problem is $\min p_i h_i$ subject to $U_i(h_i) \geq \overline{U}_i$ and $\Gamma(h_i, q) \leq 0$, Hicksian demand is the solution $h_i(p_i, q, \overline{U}_i)$ to this expenditure minimization problem, and the expenditure function is $e_i(p_i, q, \overline{U}_i) = p_i h_i(p_i, q, \overline{U}_i)$. 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.

is instead given by $\mathcal{E}_{im}$, which is the total wealth-equivalent value to agent $i$ of the price change. $\mathcal{E}_{im}$ incorporates not only the classical income effect, but also an additional income effect that arises from changing tightness in the constraint set (pecuniary externality). This changing constraint set tightness is equivalent to an increase in wealth, and hence incorporated into the wealth equivalent measure $\mathcal{E}_{im}$. If the total pecuniary externality is *positive*, in line with above discussion, if if $m$ is a normal good, that is $\frac{\partial I_{im}}{\partial w_i} > 0$, then the total income effect *increases* demand, even when the classical income effect reduces it. This again leads to positive (negative) pecuniary externalities being associated with regulated agents having *higher* (*lower*) CMRS than unregulated agents.

Taken together, Proposition 10 tells us that although the classical substitution effect from the transaction price can lead to misallocation whereby positive (negative) pecuniary externalities lead regulated agents have a low (high) CMRS relative to unregulated agents, both the collateral price substitution effect and the (pecuniary externality) income effect can push misallocation in the opposite direction.

**Example: the liquidity model.** We provide a simple example to highlight how pecuniary externalities can lead to positive responses $\boldsymbol{D}_{i,mm}$ that shift the direction of misallocation. Suppose there is a three-date economy, $t = 1, 2, 3$, with three markets ($M = 3$): the numeraire, investment at date 1 (capital good trade), and investment at date 2 (capital good trade), with $I$ ordered accordingly. A specific agent $i$ has an endowment $A_i$ of the date 1 investment good, a budget constraint $p_i I_i = 0$, and the following two constraints. Agent $i$ has utility over consumption of the numeraire date 2. Her final consumption is her final production plus purchases of the numeraire, $U_i(I_i) = I_{i1} + R(A_i + I_{i2} + I_{i3})$, where $R > 1$ is the final return on investment. She faces a simple constraint: she has a transitory liquidity surplus/shortfall $\rho_i$ at date 2, which requires her to set $q_3 I_{i3} = \rho_i$. We can interpret $\rho_i < 0$ as forced sellers (Holmström and Tirole, 1998) and $\rho_i > 0$ as arbitrageurs with limited wealth (Allen and Gale, 1994). We colloquially refer to this model as the "liquidity model."

In this simple example, Proposition 10 yields that $\boldsymbol{D}_{i,33} = -\frac{1}{q_3^2}\rho_i$ is negative if $\rho_i > 0$ and positive if $\rho_i < 0$.[14] Thus, the direction of misallocation induced by agent $i$ depends on whether agent $i$ has a liquidity surplus or shortage. If agent $i$ has a liquidity *shortage*, then she is a forced seller at date 2. In this event, a higher price $q_3$ of the date 2 good *increases* her demand for that good, that is it reduces how much she has to sell. Intuitively, this is a pecuniary externality operating through her constraints: as the price rises, she has to sell less in order to meet her same liquidity shortage. If the economy-wide pecuniary externality is positive (by relaxing forced seller constraints), this pushes the direction of misallocation to have a higher CMRS for regulated agents than for unregulated agents. By contrast if the economy-wide pecuniary externality is negative, the direction of misallocation goes in the opposite direction.

In contrast to forced sellers, a constrained buyer has fixed liquidity, meaning the amount she

[14] From the constraint, agent $i$ has Marshallian demand $I_{i3}(p_{i3}, q_3, w_i) = \frac{1}{q_3}\rho_i$ from which the derivative follows.

can purchase *declines* in the price. Intuitively, her ability to purchase assets at date 1 declines for her given wealth as its price rises. Conversely to the forced seller, the constrained buyer therefore has a *negative* demand response to the price increase, owing to a negative pecuniary externality. The constrained buyer therefore pushes the direction of misallocation such that positive (negative) pecuniary externalities lead to a lower (higher) CMRS for regulated agents than for unregulated agents.

### 5.3 Measuring Misallocation in the Data

An important question is how to quantify regulatory arbitrage and misallocation and take it to the data. We show that regulatory arbitrage in our setting is quantifiable in terms of sufficient statistics that are in principle estimable in the data.

Our results imply that misallocation comprises two terms: the amount and cost. Start first with the cost of misallocation. In our model, the cost of misallocation for agent $i$ is simply $\chi_i - \vartheta$, which is the differences in CMRS between unregulated agent $i$ and regulated agents. As discussed in Section 2, this is equivalent to the difference in the transaction prices of unregulated and regulated agents. Thus, we can measure the *cost* of misallocation as the difference in transaction price of agent $i$ relative to the transaction price of a regulated agent.

Second, consider the amount of misallocation, that is regulatory arbitrage. Our model implies that the amount of misallocation is simply the demand response of $i$ to price changes, $\boldsymbol{D}_i$, that is regulatory arbitrage by $i$. We now show that we can decompose regulatory arbitrage into a combination of aggregate financial flows and micro estimable price elasticities. In particular, element $\boldsymbol{D}_{i,mn}$ can be represented as

$$\boldsymbol{D}_{i,mn} = \underbrace{\frac{1}{q_m}}_{\substack{\text{Market}\\\text{Price}}} \times \underbrace{\xi_{i,mn}}_{\substack{\text{Price}\\\text{Elasticities}}} \times \underbrace{I_{in}}_{\substack{\text{Aggregate}\\\text{Flows}}} \tag{15}$$

This decomposition depends on three objects that are in principle empirically observable. The first is the price $q_m$ in the market. The second are the micro price elasticities $\xi_{i,mn}$ of flows by an agent of type $i$. The third are the aggregate flow positions $I_{in}$ of a type $i$ agent, which could be drawn for example from the Flow of Funds.

### 5.4 Partial Regulation of Agents and Activities

Our analysis has so far focused on complete regulation of all or a subset of agents, with exogenous (possibly zero) wedges on other agents. We classify such regulation as *identity-based* regulation. Regulation can also target specific markets or subsets of markets (i.e., activities). We classify such regulation as *activity-based* regulation. For example, the planner might directly regulate the market for a specific form of debt, such as repurchase agreements (for example, a uniform tax in a market across all unregulated agents). Alternatively, a planner might target only the domestic operations of a foreign company (such as bank subsidiary regulation), which involves taxes of only a subset of an agent's activities. We might call such regulation *partial identity-based regulation*. The distinction between these types of regulations is not only of interest from a theoretical point of view but also highly relevant in practice.[15]

In this subsection, we characterize optimal partial regulation. Formally, suppose that the planner can choose a subset of instruments $\{\tau_{im}\}$ for agents $i \in S$.[16] The results to come will give an important role to demand responses of agents to transaction price changes $p_i$. We will denote $\boldsymbol{D}_i^p = \nabla_{p_i} I_i$ to be the matrix of demand responses of agent $i$ to changes in her transaction price vector $p_i$, so that for example $\boldsymbol{D}_{i,m} = \nabla_{p_{im}} I_i$ is the set of demand responses across all goods to a change in transaction price of good $m$.

We begin by characterizing optimal regulation for any instrument the planner has.

**Proposition 11.** *Suppose the planner has complete instruments for $i \notin S$ and incomplete instruments for $i \in S$. Then:*

1. *Optimal regulation of $i \notin S$ follows as in Proposition 9, and the targeting rule of equation (13) applies.*
2. *Optimal regulation of any $i \in S$ for available instrument $\tau_{im}$ satisfies*

$$0 = \boldsymbol{D}_{im}^p \Big( \chi_i - \vartheta \Big), \tag{16}$$

Proposition 11 provides optimal regulation with any instrument the planner possesses. Intuitively, Proposition 11 divides the problem into two interrelated components.

First, for a given CMRS of regulated agents, $\vartheta$, equation (16) describes a targeting rule for misallocation for partial instruments on agents $i \in S$. This targeting rule trades off misallocation between agent $i \in S$ and regulated agents. Note that if the planner had complete instruments over $i$, equation (16) collapses to $\chi_i = \vartheta$, recovering that a regulated agent has no misallocation relative to other regulated agents.

Without complete regulation over $i \in S$, equation (13) provides a targeting rule based on weighted average of misallocation across different markets for agent $i$. The targeting rule sums

[15] 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 (market-based regulation).

[16] Results of this section are easily generalized to the case where for unregulated agents, the planner chooses a subset $\tau^1$ of wedges optimally and where the remaining wedges $\tau^2(\tau^1)$ are functions of $\tau^1$. In this case, we can simply recharacterize Proposition 11 accounting for the additional misallcoation effects $\sum_{j \in S} \mu_j \nabla_{\tau^1_{im}} \tau^2_j \boldsymbol{D}^p(\chi_j - \vartheta_j)$ across all unregulated agents. For expositional simplicity we do not focus on such characterizations except in specific examples, such as uniform activity regulation.

together the product of the demand response of $i$ in market $m$ times the misallocation of $i$ in market $m$. In the limiting case where cross-price elasticities are zero, the targeting rule collapses to $\chi_{im} = \vartheta_m$ for any instrument the planner possesses, and so corrects misallocation in those markets. Outside of the limiting case, Proposition 11 illustrates that there is no longer necessarily equal treatment: partially regulated agents may be subject to different regulation for the same activity.

Is is important to recognize that the weights $\boldsymbol{D}^p_{im}$ are the demand responses of agent $i$ to *transaction* prices, holding fixed market prices. Intuitively, this reflects the ability of regulation to directly affect the transaction price alone.

It is interesting to note that the targeting rule of equation (16) only reflects misallocation, and does not directly account for pecuniary externalities. Intuitively, this reflects the two-step nature of this problem. Once optimal misallocation for agent $i \in S$ relative to regulated agents has been determined for given $\vartheta$, we can use the targeting rule of equation (13) to determine the socially optimal value of $\vartheta$. This determination then trades off pecuniary externalities against total misallocation in the normal fashion.

In sum, Proposition 11 provides a two-step process for thinking about partial regulation. First, the planner uses the targeting rule of equation (16) to determine the optimal level of misallocation between agents subject to complete regulation and an agent subject to partial regulation. Then, the planner uses the targeting rule of equation (13) to determine the social trade-off between misallocation and externalities.

**Uniform activity regulation.** A particularly important application is uniform regulation of a specific activity (market), that is $\tau_{im} = \tau_m$ is constant for all $i \in S$. Analogously to equation (16), we obtain

$$0 = \boldsymbol{D}^p_{Sm}\mathbb{E}_{i\in S}\left[\chi_i - \vartheta\right] + \mu_S \mathbb{C}ov_{i\in S}\left(\boldsymbol{D}^p_{im}, \chi_i - \vartheta\right).$$

If there is no within-unregulated misallocation, then this collapses to a simple average rule $0 = \boldsymbol{D}^p_{Sm}\left[\chi_S - \vartheta\right] = 0$. Intuitively, the planner uses the market-based instrument to correct a weighted average misallocation across agents and markets. If all cross-price elasticities are zero, then the optimal rule simple states $p_{im} = \vartheta_m$, and hence activity regulation achieves no misallocation for the market it targets.

**Subsidiary regulation.** A second case of interest is when the planner can regulate a "subsidiary" of agent $i$, that is a subset $\hat{M} \subset M$ of agent $i$'s activities. For example, this could be a foreign-country subsidiary of a domestic bank. In this case, the targeting rule in matrix form is

$$0 = \boldsymbol{D}^p_{i\hat{M}}\left(\chi_i - \vartheta\right).$$

Intuitively, this gives the planner more degrees of freedom with which to reduced $i$'s misallocation relative to a regulated agent. If the subsidiary fully independent, that is $\boldsymbol{D}_i^p = \begin{pmatrix} \boldsymbol{D}_{i\hat{M}}^p & 0 \\ 0 & \boldsymbol{D}_{i,M\setminus\hat{M}}^p \end{pmatrix}$ is block diagonal, then we have $\chi_{i\hat{M}} = \vartheta_{\hat{M}}$. In this case, it is as-if the independent subsidiary is subject to complete regulation whereas the rest of the entity is unregulated. As a result, the subsidiary is regulated in the same manner as regulated agents. This insight provides a possible efficiency based rationale for ring fencing foreign bank subsidiaries and subjecting them to the same regulation as domestic banks in the case that the foreign bank subsidiaries' activities are sufficiently independent from those of the banking group.

### 5.5 Impact of newly unregulated agents.

Two closely related regulatory concerns in practice are that an agent in a regulated industry might reclassify to an unregulated industry to escape regulation, or that a new entrant might establish in an unregulated industry to escape regulation. We now shed light on the important question of how reclassification (or entry) of agents to unregulated status affects optimal regulation.

In particular, suppose we start from a case where a subset $S$ of agents are unregulated. Now expand the unregulated set to $S_1$, and denote $dS = S_1 \setminus S$ the newly unregulated agents. It is easy to see we can write the new targeting rule as

$$0 = (1-\alpha_S) \underbrace{\boldsymbol{D}_{dS} \times \mathbb{E}_i(\chi_i - \vartheta \mid i \in dS)}_{\text{Average Misallocation in dS}} + \alpha_S \overbrace{\underbrace{\boldsymbol{D}_S \mathbb{E}(\chi_i - \vartheta \mid i \in S)}_{\text{Average Misallocation in S}} + \mathcal{E}}^{\text{Old Targeting Rule}},$$

where $\alpha_S = \frac{\mu_S}{\mu_{S_1}}$ is the relative share of unregulated agents in $S$. Thus, an expansion of the unregulated set provides a clean and separable addition to the targeting rule. The targeting rule weights misallocation between the old and new sets based on the relative size of the two sets.

## 6 A Classification Scheme for Unregulated Finance

After the 2008 financial crisis, a new regulatory regime for financial stability has emerged. While conventional bank holding companies face tighter regulatory requirements, many other financial institutions that conduct similar activities remain unregulated. This has raised questions about the efficiency and efficacy of current regulatory policies. An active debate has ensued about whether and how to start regulating the unregulated financial sector. 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.[17] At the same time, there is little consensus on what constitutes a "shadow bank" and which parts of

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

the unregulated sector, if any, we should subject to new regulation.

In this section, we leverage the theoretical framework of Sections 2 through 5 to develop a regulatory classification scheme for unregulated finance that can be used to evaluate targets for new regulation. Our classification scheme directly identifies agents and activities whose regulation would attain the largest welfare gains. Importantly, this assessment is based on estimable sufficient statistics, allowing us to determine the attributes of agents and activities that are the most valuable targets for new 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 "shadow banks" from the perspective of a regulator designing new regulation.

Our theory directly speaks to important policy questions. In Section 6.1, we address how a policymaker should evaluate the welfare benefits of new unregulated entrants in a market. For example, an increasingly prominent policy question is whether and how to regulate new FinTech companies that enter a market either because of a technological advantage relative to existing players or because regulation has divorced the transaction price of regulated agents from the market price. We show that the welfare benefits of an entrant is the sum of its profits plus the social value of its activities, evaluated according to the allocative value of a trade.

We then in Sections 6.2 and 6.3 use our classification scheme to characterize the tradeoffs between using identity-based (Section 6.2) and activity-based (Section 6.3) regulation to extend the current regulatory framework to unregulated finance or "shadow banks".

### 6.1 Welfare Impact of a New Unregulated Player

What are the welfare consequences of a new (unregulated) market entrant such as a FinTech company? Suppose there is a new agent $I+1$ that enters the economy producing new goods. To simplify the exposition, we assume that this agent enjoys no direct utility from consumption and instead remits its profits to existing agents. In other words, we think of agent $I+1$ as a new producer that is directly owned by existing agents.

The new player (agent $I+1$) produces a vector of endowments $e^u$ by engaging in a set of zero-wealth trades $t^u$ (recall that trades are defined to be zero-wealth).[18] Observe that: (i) if $e^u_m + t^u_m > 0$, then the new player is a net producer of good $m$; (ii) if $e^u_m + t^u_m = 0$, then the new player has no net position in $m$; (ii) if $e^u_m + t^u_m < 0$, the new player is a net user of good $m$. Define $n^u = e^u + t^u$ to be the player's *net production*. The following result generically characterizes the social welfare impact of this new player with either complete or incomplete regulation.

[18] Note that this formulation in principle allows for the new player to be resource-destroying: it can have (some) negative endowments and potentially have negative profits. Although our results accommodate such cases, we focus exposition on cases of nonnegative endowments and profits.

**Proposition 12.** *The first order social welfare impact of a new player is*

$$\Delta_{I+1} = \underbrace{qe^u}_{\text{Profit}} \quad + \quad \underbrace{\tau n^u}_{\text{Social Value of Net Production}}$$

According to Proposition 12, the welfare impact of a new entrant comprises two terms. First, $qe^u$ is the profit of the agent. Since trades sum to zero, new profits are simply the market value of the new endowments, i.e., the direct value added to the economy relative to existing resources. As long as the new agent's profit is positive, the direct social welfare impact of the new player is always positive. An immediate corollary is that it is always *privately* efficient for this new player to enter the market when its profit is positive, since it provides direct positive value to its owners.

The second term reflects the indirect social value of the new player's *net production* $n^u = e^u + t^u$ across markets: new production in market $m$ net of purchases in market $m$. For example if the player does produce good $m$ ($e^u_m = 0$) but uses it in production ($t^u_m < 0$), then net production in good $m$ is negative ($n^u_m < 0$).

The welfare consequence of net production in a market is ambiguous even when the new player makes positive profits. Net production in market *m increases* social welfare if $\tau_m > 0$, that is regulated agents are *taxed* for purchasing goods in that market. In contrast, net production in market *m reduces* welfare if regulated agents are subsidized for purchasing goods in that market. The intuition is that if regulated agents are taxed in a market $m$, then the allocative value of a trade is positive in that market: the planner is encouraging these agents to sell in the market, rather than buy in it. Thus, new entrants that *sell* in markets with positive regulation are particularly welfare enhancing. New entrants that *buy* in these markets are welfare reducing. The opposite results hold for subsidized purchases in a market.

Proposition 12 provides a simple way for evaluating the welfare consequences (to first order) of a new player in a market. The new player's profit is sufficient to determine the *private* value of the new player. The (indirect) social value of the new player is evaluated by adding together its net production in each market weighted by the optimal tax on regulated agents for that market. This gives regulators a simple tool for evaluating the social desirability of a new player, such as a new FinTech company, by adding together its profits and the social value of its net production.

### 6.2 Regulatory Classification of Agents

Section 6.1 classified the welfare benefits of a new entrant. This section develops a regulatory classification of unregulated agents: that is, which agents are the most valuable new targets for regulation. Crucially, we show that the agents that are key contributors to regulatory arbitrage are *not necessarily* the most valuable targets for new regulation because the relevant demand responses differ between the two exercises.

Formally, we introduce the following exercise: suppose that the planner proposes a change in regulation, $d\hat{\tau}$, for agent $i$. The following result characterizes the first-order welfare gains from such new regulation.

**Proposition 13.** *To first order, the welfare gains from new regulation $\hat{\tau}$ of agent i are*

$$\Delta_i = d\hat{\tau} \boldsymbol{D}_i^p \Big( \chi_i - \vartheta \Big).$$

Proposition 13 provides a simple regulatory classification of unregulated agents. It tells us that the welfare benefits of the new regultaion are simply the demand responses generated by that regulation for agent $i$, $\boldsymbol{D}_i^p$, times the cost of misallocation for that agent, $\chi_i - \vartheta$. This means that new regulation is particularly value when it introduces a demand response in an activity associated with large existing misallocation.

It is important to note that the demand response that is relevant for the regulatory classification is only the demand response to the transaction price. This happens because introduction of new regulation only affects agent $i$'s transaction price, and not her market price. This leads to an crucial observation: $\boldsymbol{D}_i^p$ constitutes the relevant notion of regulatory arbitrage by an agent from the perspective of the regulatory classification, whereas $\boldsymbol{D}_i$ constitutes the relevant notion for the targeting rule. This means that the regulatory classification embeds a notion of regulatory arbitrage that only considers the classical income and substitution effects, and not income and substitution effects from changes in collateral prices. In the context of Proposition 10, this means that agents that matter for regulatory arbitrage in the targeting rules are not necessarily the most valuable targets for new regulation (or may matter in different ways). We make this point cleanly in an example by revisiting the example of the liquidity model.

**Example: Liquidity model revisited.** The contrast between the two notions of regulatory arbitrage manifests starkly in the liquidity shock model of Section 5.2. While an extreme case, this simple model helps further clarify the difference between the two policy questions addressed in Sections 5 and 6. In the liquidity model from Section 5.2, we emphasized that regulatory arbitrage through collateral price effects in market 3 was a driver of misallocation. And yet in this same example, we have $\boldsymbol{D}_{i3}^p = 0$ and hence $\Delta_i = 0$ for any regulation of agent $i$ that targets market 3. The intuition is that the binding constraint of the liquidity model, $q_3 I_{i3} = \rho_i$, implies that the allocation in market 3 is *entirely* determined by the collateral price, and not at all by the transaction price. Therefore, from the perspective of regulatory classifcatino, there is no point to extending regulation along this dimension to this agent.

This example provides a simple illustration of how agents that matter for regulatory arbitrage may not necessarily be desirable targets for regulation. It amounts to the principle that regulatory arbitrage in the targeting rule accounts for the set of collateral price effects on demand, whereas regulatory arbitrage from the perspective of regulatory classifciation only accounts for the classical income and substitution effects operating through transaction prices. It highlights the two different objects for a regulator to measure in the two different exercises.

### 6.3 Regulatory Classification of Markets

We now present a regulatory classification scheme for markets, which identifies which markets are the most valuable targets for new activity-based regulation. This exercise parallels our exercise for the reglatory classifciation of agents. Formally, we now instead suppose that the planner imposes a uniform change in tax $d\hat{\tau}_m$ on market $m$ across all agents (it will not matter for welfare to first order if regulated agents are exempt since for them $\chi_i = \vartheta$). The following result characterizes the welfare gains from such new regulation.

**Proposition 14.** *To first order, the welfare gains from new regulation $d\hat{\tau}_m$ of market $m$ across unregulated agents is*

$$\Delta_m = d\hat{\tau}_m \mathbb{E}_i \left[ \boldsymbol{D}^p_{im} \left( \chi_i - \vartheta \right) \right].$$

The intuition of Proposition 14 is close to that of Proposition 13. The welfare gain from a new uniform tax in market $m$ is simply the sum of agent demand responses weighted by agent misallocation. In the special case where the CMRS is equal across unregulated agents $i \in S$, we can rewrite the welfare gains from new regulation equivalently as

$$\Delta_m = d\hat{\tau}_m \boldsymbol{D}^p_{Sm} \mathbb{E}_{i \in S} \left[ \chi_i - \vartheta \right].$$

Thus, the classification of a market adds up the demand response of unregulated agents to the new regulation, and multiplies that regulatory arbitrage by the average misallocation of unregulated agents relative to regulated agents. Activity-based 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.

**Activity regulation as implicit discrimination.** An interesting observation is that activity-based regulation act as a discriminatory tax against certain agents and business models, even though it in principle applies equally to all agents. The intuition is seen most starkly when a subset $\hat{S}$ of unregulated agents do not participate in market $m$. In this case, their demand responses to market $m$ are zero, and we have

$$\Delta_m = d\hat{\tau}_m \boldsymbol{D}_{S \setminus \hat{S}} \mathbb{E}_{i \in S \setminus \hat{S}} \left[ \chi_i - \vartheta \right].$$

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 (in principle) uniform tax nevertheless implicitly discriminates against participants in market $m$, 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 7.2.

**Identity-based regulation versus activity-based regulation.** The regulatory classifications of Propositions 13 and 14 identify the potential welfare gains from new identity- and activity-based regulations. 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 $d\hat{\tau}$ to agent $i$ against activity-based regulation $d\hat{\tau}_m$ in market $m$ to all agents. From above, and for simplicity focusing on the case of no underlying cross-sectional misallocation, we have

$$\Delta_i - \Delta_m = \left[ d\hat{\tau} \boldsymbol{D}_i^p - d\hat{\tau}_m \boldsymbol{D}_{Sm}^p \right] \mathbb{E}_{i \in S} \left[ \chi_i - \vartheta \right]$$

The trade-off revolves around the ability to target many activities within a single agent versus the ability to target a single activity across agents. The first term reflects the difference in arbitrage responses in these two cases. In the first case, arbitrage responses are generated for all prices in agent $i$. In the latter case, arbitrage responses are generated for all agents with respect to price $m$. This difference in regulatory arbitrage between the two cases is then multiplied by the same cost of misallocation. Intuitively, this tells us that market regulation is valuable when it can correct regulatory arbitrage in markets with large misallocation. By contrast, agent regulation is valuable when it can correct regulatory arbitrage in an agent with large misallocation.

## 6.4 Non-Regulatory Interventions

In practice, proposals for interventions in unregulated finance include both regulatory interventions and fiscal "support programs," such as access to the lender of last resort (LOLR). For example, support programs may look to bolster rollover by allowing financial institutions 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 market price of agent $i$, $q_i$, in her constraints while holding fixed the transaction price.

We can analogously characterize the welfare gains from a support program. Formally, this exercise supposes that the planner is able to boost the market price $q_m$ (for collateral) for agent $i$ in market $m$ to $q_{im} = q_m + dq_{im}$. It should be recognized that our analysis takes into account the benefits of such an intervention, but abstracts away from costs. It should also be noted that our analysis will implicity allow for regulatory arbitrage in response to that market price increase.

It will be helpful to define $\boldsymbol{D}_i^q = \boldsymbol{D}_i - \boldsymbol{D}_i^p$, which by construction is the derivative of demand in the market price while holding fixed the transaction price. We obtain the following result.

**Proposition 15.** *To first order, the welfare gains from a support program $dq_{im}$ for any agent $i$ is*

$$\Delta^q_{im} = \underbrace{\frac{1}{\lambda}\omega_i \frac{\partial \mathcal{U}_i}{\partial q_m}}_{\text{Direct Effect}} + \boldsymbol{D}^q_{im}[\chi_i - \vartheta].$$

Intuitively, a support program for agent $i$ has two effects. The first effect is a direct effect: the increase in the market price $i$ faces for collateral alters her set of feasible allocations, represented through its impact on constrained utility. This effect is precisely the pecuniary externality underlying Lemma 3. Thus if a support program were extended to all agents, the sum of direct effects would add to $\mathcal{E}_m$ (assuming no redistributive motive). The direct effect is strong when $i$ is particularly subject to the pecuniary externality in that market.

Second, there is an indirect effect: the support program generates a set of demand responses $\boldsymbol{D}^q_{im}$ in agent $i$ to the market price change. These demand response changes are beneficial to the extent they increase demand in markets where $i$ has a high CMRS relative to regulated agents. Intuitively, if agent $i$ has a high CMRS because she faces binding constraints, then a support program that relaxes those constraints and allows her to purchase more has additional value in correcting misallocation.

It is interesting, however, to connect back to Proposition 10, which suggested that a positive pecuniary externality that generated a positive $\boldsymbol{D}_{i,mm}$ meant that the direction of misallocation was $\chi_{im} - \vartheta < 0$. Intuitively, under these conditions the indirect effect is actually *negative*, and partially offsets the benefit of the support program. The intuition is that in the presence of positive externalities and large market price effects on demand, the planner has optimally implemented regulation that taxes regulated agents and so raises their CMRS relative to the unregulated. Thus the indirect effect of the support program actually *exacerbates* misallocation, even while it relaxes constraints.

More broadly, Proposition 15 suggests that if an agent has a large demand respoinse, $\boldsymbol{D}$, it is strong potential candidate for either new regulation (large transaction price response $\boldsymbol{D}^p_i$) or for a support program (large market price response $\boldsymbol{D}^q_i$), or both. This tells regulators that agents associated with large amounts of regulatory arbitrage are strong candidates for *some* form of intervention. However, such agents are not unambiguously good targets for intervention, as the indirect effects of support programs can counteract the direct effects through misallocation.

# 7 Applications

In this section, we apply our theory to two primary applications. In our first application in Section 7.1, 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 7.2, 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.

### 7.1 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.[19] 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 $\pi_H$. 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 $q_0$, and we denote as $q_1$ the price vector at date 1 in the low state.

At date 0, shadow banks (unregulated agents) can finance a project by purchasing the capital good, $I_{i0}$, at price $q_0$, where $I_{i0} > 0$ denotes a purchase of the capital good. At date 0, shadow bank $i$ can use the capital good to create $R(s)\Phi_i(I_{i0})$ 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 $\Phi_i$. Shadow banks can sell the capital good at date 1, denoted by $I_{i1}$, where $I_{i1} < 0$ denotes selling the capital good. The resale price in the low state is $q_1 \leq 1$, while the resale price in the high state $s_H$ is constant at 1.

Shadow banks can also issue debt, $D_{i0}$ and $D_{i1}$, and consume $C_{it}$. Shadow banks can trade the consumption good at date 0, $c_{i0}$, 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

$$D_{i0} = c_{i0} + p_{i0} I_{i0} - w_i,$$

where $w_i$ is the tradeable wealth level and $p_{i0}$ is the transaction price. 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 $D_{i1} = D_{i0}$ and $C_{i2}(s_H) = R_H \Phi_i(I_{i1}) - D_{i0}$ 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 $D_{i1} \leq 0$. As a result, in the low state debt repayment must be done using asset liquidations,

[19] Our simple model is in the spirit of standard macro-finance models such as Kiyotaki and Moore (1997) and Lorenzoni (2008).

$q_1 I_{i1} = -D_{i0}$. Hence, consumption in the low state is $C_{i2}(s_L) = R_L \Phi_i(I_{i0}) + I_{i1}$, since the entire debt level is repaid at date 1 through asset liquidations. Substituting $q_1 I_{i1} = -D_{i0}$ in to preferences and the budget constraint, we obtain that the bank's object is to maximize constrained utility,

$$\mathcal{U}_i = c_{i0} + \Phi_i(I_{i0}) + \Big[\pi_H q_1 + \pi_L\Big] I_{i1},$$

subject to the budget constraint

$$c_{i0} + p_{i0} I_{i0} + p_{i1} I_{i1} = w_i$$

and the non-negativity constraint $c_0^i \geq 0$.[20] In the general notation, $M = 3$, with $c_{i0}$ being the numeraire and $I_{i0}$, $I_{i1}$ being the two other traded goods. 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 $c_0^i = 0$, since if $c_{i0} > 0$. We will assume this is the case throughout the remainder of this section.

From here, we obtain the following result.

**Proposition 16.** *The regulatory classification of shadow bank i is proportional to* $-\xi_i I_{i0}$, *where* $\xi_i = \frac{p_{i0}}{I_{i0}} \frac{\partial I_{i0}}{\partial p_{i0}} = \frac{\Phi_i'(I_{i0})}{\Phi_i''(I_{i0}) I_{i0}}$ *is the elasticity of shadow bank investment at date* 0 *to the date* 0 *price.*

Proposition 16 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 $I_{i0}$. 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 $\Phi_i(I_{i0}) = A_i(I_{i0})^{\alpha_i}$, 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, $I_{i0}$, and (ii) a large illiquid investment factor share, $\alpha_i$. These are in principal estimable sufficient statistics, with for example illiquid investment directly observable from balance sheet data.

[20] Note that the appearance of $p_{i1}$ is the budget constraint and $q_1$ in (constrained) utility implies that the shadow bank $i$ has to cover debt at the market price, that is extra funds raised from the transaction price do not relax the constraint.

The results of this section provide a simple way to think about classifying shadow banking institutions as valuable targets for regulation. We have shown here in a simple environment that the institutions that are the best targets for new regulation are those with high illiquid investment levels and factor shares. This provides guidance to regulators on what to look for in the unregulated financial sector when evaluating new targets for regulation.

### 7.2 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 (domestic) 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. A key takeaway from our framework is that the relative value of inflow and outflow regulation can be summarized by differences in misallocation at both dates, and differences in investor propensity for flight and retrenchment. We also highlight how an outflow tax can be valuable as an implicitly discriminatory tax against undesirable flighty investors.

There are three periods, $t = 0, 1, 2$. The SOE faces aggregate uncertainty that is realized at date 1, with $s \in \{s_H, s_L\}$ ("high state" $H$ and "crisis" $L$). The probability of the high state, $s_H$, is $\pi_H$. The economy has $N$ domestic capital goods which can be purchased and sold at both date 0 and date 1. We denote market prices in period 0 as $q_{0n}$. At date 1, the market price of capital good $n$ is denoted $q_{1n}$ if the economy is in the crisis state, $s = s_L$. Market prices are constant and normalized to 1 in the high state, $s = s_H$. We denote $q_0$ and $q_1$ to be the vectors of date 0 and date 1 prices.[21]

At date 0, international investor $i$ (i.e., unregulated agent $i$) can purchase a vector $I_{i0}$ of domestic capital goods, with $I_{i0n}$ denoting purchases of good $n$. If the high state is realized at date 1, international investor $i$ earns a high payoff $F_{iH}(I_{i0})$ from her investment in units of the consumption good, at which point her project ends. If instead the crisis state is realized at date 1, the project yields nothing at date 1 and a lower final value $F_{iL}(I_{i0}, I_{i1})$ 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.[22] $I_{i1}$ is the endogenous vector of date 1 flows into or out of domestic capital goods during the crisis state, with $I_{i1n}$ denoting flows in good $n$.

International investors are deep-pocketed at date 0, which makes it convenient to interpret wealth at date 0 as wealth alloated to purchases in the SOE (coming out of $i$'s international resoruces).

[21] Note that in the general notation, the index $m$ corresponds to pairs $(t, n)$. Thus we can index $(0, 1)$ by $m = 2$, $(0, n)$ by $m = n + 1$, $(1, 1)$ by $m = n + 2$, and so on. It is expositionally clearer in this example to maintain pair dependence $(t, n)$ as opposed to the index $m$.

[22] 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 by the real exchange rate $\epsilon_L$ in the low state.

The utility of international investors from their investments in the SOE can then be written as

$$U_i = \pi_H F_{iH}(I_{i0}) + (1-\pi_H)\Big(\lambda_{i1} c_{i1} + F_{iL}(I_{i0}, I_{i1})\Big).$$

We denote by $\lambda_{i1}$ the marginal value of repatriated wealth at date 1 in the crisis state. 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

$$c_{i1} + \hat{p}_{i0} I_{i0} + p_{i1} I_{i1} \leq \hat{w}_i$$

where $\hat{p}_{i0} = \frac{p_{i0}}{1-\pi_H}$ denotes probability-normalized prices. 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 $F_{iH}(I_{i0}) = \sum_n F_{iHn}(I_{i0n})$ and $F_{iL}(I_{i0}, I_{i1}) = \sum_n F_{iLn}(I_{i0n}, I_{i1n})$. In this case, cross-price elasticities between different goods $n$ are zero. However, 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

$$\omega_{in} \equiv -\frac{\partial I_{i1n}/\partial p_{i0n}}{\partial I_{i0n}/\partial p_{i0n}}. \tag{17}$$

Intuitively, $\omega in$ measures the fraction of a new inflow $dI_{i0n}$ at date 1 that ends up withdrawing from the SOE as an outflow at date 1. It is natural for $\omega_{in} \geq 0$ when an increase in inflows is associated with an increase in outflows. Note that the negative sign on the right hand side of equation (17) appears because an increase in inflows in our model is a more positive value of $I_{i0n}$, whereas an increase in outflows is a more negative value of $I_{i1n}$.

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

$$\zeta_{in} \equiv -\frac{\partial I_{i0n}/\partial p_{i1n}}{\partial I_{i1n}/\partial p_{i1n}}. \tag{18}$$

Intuitively, $\zeta_{in}$ measures how much an increase $dI_{i1n}$ in outflows at date 1 results in an increase in inflows $dI_{i0n}$ at date 0. When $\zeta_{in}$ 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.[23]

Investor flight $\omega_{in}$ and investor retrenchment $\zeta_{in}$ 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

[23] These notions of flight and retrenchment are generalizations of similar concepts from Caballero and Simsek (2019).

between maintaining the project and fleeing in the low state might have a large tendency for flight, $\omega_{in}$. By contrast, that same investor would likely have a low tendency for retrenchment, $\zeta_{in}$. 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 $\frac{\partial I_{i1n}}{\partial p_{i1n}}$ 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 $\frac{\partial I_{i0n}}{\partial p_{i1n}}$ is low. Put together, this means that retrenchment $\zeta_{in}$ is low.

**Unregulated regulatory arbitrage.** Given that we have written the model without cross-price elasticities across goods, the demand response matrix $\boldsymbol{D}_i$ of international investor $i$ can be written as the block diagonal matrix of demand responses $\boldsymbol{D}_{in}$ in good $n$. That is, we have

$$\boldsymbol{D}_{in} = \begin{pmatrix} \frac{\partial I_{i0n}}{\partial p_{i0n}} & -\omega_{in}\frac{\partial I_{i0n}}{\partial p_{i0n}} \\ -\zeta_{in}\frac{\partial I_{i1n}}{\partial p_{i1n}} & \frac{\partial I_{i1n}}{\partial p_{i1n}} \end{pmatrix}.$$

This represents off-diagonal elements of $\boldsymbol{D}_{in}$ 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 of all international investors given by

$$\boldsymbol{D}_{Sn} = \begin{pmatrix} \frac{\partial I_{S0n}}{\partial p_{0n}} & -\omega_{Sn}\frac{\partial I_{S0n}}{\partial p_{0n}} \\ -\zeta_{Sn}\frac{\partial I_{S1n}}{\partial p_{1n}} & \frac{\partial I_{S1n}}{\partial p_{1n}} \end{pmatrix}$$

where $\omega_{Sn} = \sum_i \beta_{i0n}\omega_{in}$ is the average flight across international investors and $\zeta_{Sn} = \sum_i \beta_{i1n}\zeta_{in}$ the average retrenchment. The weight $\beta_{i0n} = \frac{\mu_i \xi_{i0n}\alpha_{i0n}}{\sum_j \mu_j \xi_{j0n}\alpha_{j0n}}$ reflects the relative inflow elasticity of $i$ in good $n$ weighted by $i$'s share of inflows, and similarly where $\beta_{i1n} = \frac{\mu_i \xi_{i1n}\alpha_{i1n}}{\sum_j \mu_j \xi_{j1n}\alpha_{j1n}}$ is the analogous weighting measure for outflows. Therefore, aggregate flight $\omega_{Sn}$ is high when flighty investors with high $\omega_{in}$ also have high inflow elasticities and high market shares of inflows. Similarly, aggregate retrenchment $\omega_{Sn}$ is high when retrenching investors with high $\zeta_{in}$ have high outflow elasticities and high market shares of outflows.

**Example: safe and flighty investors.** Suppose that there are only two types of investors: fully safe, $i = s$, and fully flighty, $i = f$. Fully safe investors inelastically set $I_{s1} = 0$, while fully flighty investors inelastically set $I_{f1} = -I_{f0}$. Their measures are $\mu_s$ and $\mu_f$, respectively. In this case, we have $\omega_s = \zeta_s = 0$ and $\omega_f = \zeta_f = 1$. Moreover, we have $\frac{\partial I_{s1}}{\partial p_{s1}} = 0$ and $\frac{\partial I_{s1}}{\partial p_{s1}} = -\frac{\partial I_{s0}}{\partial p_{s1}} = \frac{\partial I_{s0}}{\partial \hat{p}_{s0}}$.[24] Therefore, we have $\omega_{Sn} = \zeta_{Sn} = \frac{\mu_f \xi_{f0n}\alpha_{f0n}}{\mu_f \xi_{f0n}\alpha_{f0n} + \mu_s \xi_{s0n}\alpha_{s0n}}$, 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 $\xi_{s0} = \xi_{f0} = \xi_{S0}$, then we have $\omega_{Sn} = \zeta_{Sn} = \frac{\mu_f \alpha_{f0n}}{\mu_f \alpha_{f0n} + \mu_s \alpha_{s0n}}$ is the share of flows of flighty

[24] This means that we have $\boldsymbol{D}_{sn} = \begin{pmatrix} \frac{\partial I_{i0n}}{\partial p_{i0n}} & 0 \\ 0 & 0 \end{pmatrix}$ and $\boldsymbol{D}_{fn} = \begin{pmatrix} \frac{\partial I_{i0n}}{\partial p_{i0n}} & -\frac{\partial I_{i0n}}{\partial p_{i0n}} \\ \frac{\partial I_{i0n}}{\partial p_{i0n}} & \frac{\partial I_{i0n}}{\partial p_{i0n}} \end{pmatrix}$.

investors relative to safe investors. Finally, observe that we have $\frac{\partial I_{S0n}}{\partial p_{0n}} = \mu_s \frac{\partial I_{0sn}}{\partial p_{i0n}} + \mu_f \frac{\partial I_{f0n}}{\partial p_{f0n}}$ but $\frac{\partial I_{S1n}}{\partial p_{1n}} = \mu_f \frac{\partial I_{f0n}}{\partial p_{f0n}}$, reflecting that only flighty investors engage in outflows.

### 7.2.1 Optimal Regulation with Unregulated Capital Flows

From here, we can characterize the impact of unregulated capital flows on optimal financial regulation. Assume for now that all capital flows are unregulated, so there is no cross-sectional misallocation. Because $\boldsymbol{D}_S$ is block diagonal, the targeting rule of equation (13) is also block diagonal, meaning we can characterize the targeting rule for each $n$, $\boldsymbol{D}_{Sn}\left[\chi_{Sn} - \vartheta_n\right] = \mathcal{E}_n$. We thus obtain the following result characterizing optimal domestic regulation.

**Proposition 17.** *Optimal regulation with unregulated capital flows yields a targeting rule for misallocation*

$$\chi_{S0n} - \vartheta_{0n} = \overbrace{\frac{1}{1-\omega_{Sn}\zeta_{Sn}}}^{\text{Amplification}} \left( \frac{1}{\frac{\partial I_{S0n}}{\partial p_{0n}}} \mathcal{E}_{0n} + \overbrace{\omega_{Sn} \frac{1}{\frac{\partial I_{S1n}}{\partial p_{1n}}} \mathcal{E}_{1n}}^{\text{Flight}} \right)$$

$$\chi_{S1n} - \vartheta_{1n} = \underbrace{\frac{1}{1-\omega_{Sn}\zeta_{Sn}}}_{\text{Amplification}} \left( \underbrace{\frac{1}{\frac{\partial I_{S0n}}{\partial p_{0n}}} \zeta_{Sn} \mathcal{E}_{0n}}_{\text{Retrenchment}} + \frac{1}{\frac{\partial I_{S1n}}{\partial p_{1n}}} \mathcal{E}_{1n} \right)$$

Although Proposition 17 appears complicated, it is in fact intuitive. We begin by discussing optimal misallocation at date 0. Taxing regulated agents at date 0 reduces demand and increases the price $q_{0n}$. This price leads to regulatory arbitrage in the form of inflows, which is captured by $\frac{1}{\frac{\partial I_{S0n}}{\partial p_{0n}}}$. When this response is large and capital flows respond aggressively to price changes, changes in demand of regulated agents only translates to a small change in equilibrium prices. Thus higher degrees of regulatory arbitrage dampen the efficacy of regulation, as the increase in capital inflows counteracts the decrease in demand from regulated agents, leading the planner to prefer smaller magnitudes of misallocation.

When $\omega_{Sn} = \zeta_{Sn} = 0$, that is there is no flight or retrenchment, this direct effect is the only additional component of the targeting rule. However, when there is flight and retrenchment, $\omega_{Sn}\zeta_{Sn} > 0$, this is not the entire effect. First, there is a counterveiling *amplification* effect, measured by $\frac{1}{1-\omega_{Sn}\zeta_{Sn}} > 1$, 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, which in turn fuels a decrease in inflows due to retrenchment. This partially offsets regulatory arbitrage at date 0. Thus surprisingly, flight and retrenchment indirectly counteract the dampening effect of regulatory arbitrage on optimal tax rates.

The effect of inflow regulatory arbitrage on outflows is captured by the "Flight" term in the targeting rule for date 0. Intuitively, an increase in the date 0 tax increases outflows via flight, captured by $\omega_{Sn}$. Flight is costly to the extent it generates large pecuniary externalities, either through outflow arbitrage or externalities. Thus interestingly, the cost of flight is low (for given externalities) when outflows are highly elastic to price changes: highly elastic outflows mean the price impact is small.

If both date 0 and date 1 externalities have the same sign, then flight amplifies the targeting rule and promotes accepting greater misallocation. Intuitively, this is because the date 0 tax has an amplified benefit of indirectly targeting outflows. Conversely if the externalities have opposite signs, the two effects partially offset one another. Intuitively if date 0 investment is taxed but date 1 investment subsidized (e.g., to prevent fire sales), then the two forces offset one another and misallocation is smaller in magnitude than the direct effect. Intuitively, this is because a higher tax on regulated agents promotes regulatory arbitrage by flighty international investors, which then results in flight and destructive outflows. Therefore, a date 0 tax becomes less desirable.

Conversely, there is an equivalent "Retrenchment" term in the date 1 targeting rule. A higher date 1 subsidy for regulated agents at date 1 that promotes a higher date 1 price, leads to retrenchment by international investors and higher inflows ex ante. This retrenchment is measured by $\zeta_{Sn}$, and produces through regulatory arbitrage a change in price at date 0. Interestingly again, the retrenchment effect is muted when date 0 capital flow elasticities are muted. As with Flight, Retrenchment amplifies misallocation when externalities have the same sign, and dampens misallocation when externalities have opposite signs.

#### 7.2.2 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 or outflows. This parallels the classification exercise of markets of Section 6, in particular that of Proposition 14. 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.

Following Section 6, we define $\Delta_{0n}$ as the first-order money-metric welfare gain to the SOE from imposing a uniform 1% ad valorem inflow tax at date 0 on good $n$ across international investors, and similar $\Delta_{1n}$ the gains from an outflow tax (inflow subsidy) at date 1. We assume capital flows were previously unregulated. The money-metric measure is the date 0 change in wealth that would yield the same welfare as imposing the tax. Note that $\Delta_{0n}$ is the gain from an inflow tax and $\Delta_{1n}$ from an outflow tax. We characterize these welfare gains in the following proposition.

**Proposition 18.** *The money-metric first-order welfare gains for capital controls are:*

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

$$\Delta_{0n} = \frac{\partial I_{S0n}}{\partial p_{0n}} \Big( \chi_{S0n} - \vartheta_{0n} - \omega_{Sn}(\chi_{S1n} - \vartheta_{1n}) \Big). \tag{19}$$

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

$$\Delta_{1n} = \frac{\partial I_{S1n}}{\partial p_{1n}} \Big( \zeta_{Sn}(\chi_{S0n} - \vartheta_{0n}) - (\chi_{S1n} - \vartheta_{1n}) \Big) \tag{20}$$

Proposition 18 shows that the value of applying an inflow control depends on two terms. First, there is the direct effect: an inflow tax discourages inflows as long as $\frac{\partial I_{S0n}}{\partial p_{0n}} < 0$. Thus, the capital control tax has greater efficacy when inflow elasticities or inflow volumes are high. This total flow change is multiplied by misallocation at both dates. First, there is misallocation at date zero, $\chi_{S0n} - \vartheta_{0n}$. The tax generates welfare gains if $\chi_{S0n} - \vartheta_{0n} < 0$, that is there is misallocation in that the CMRS of regulated domestic agents is high relative to international investors. In addition to the direct effect, there is an indirect effect through international flightiness. Discouraging the inflow discourages outflows in proportion to Flight, $\omega_{Sn}$, which is in turn valuable when the CMRS of international investors is high relative to domestic agents. Thus even if there is no misallocation at date 0, an inflow control can be valuable to the extent it discourages flighty outflows. This motivation rises in the average flightiness of investors, $\omega_{Sn}$, for example when flighty investors constitute a higher share of the market.

Proposition 18 shows that the value of outflow controls also depends on a direct and indirect effect. The direct effect is analogous to that of inflows, but for otuflows. The indirect effect reflects that an outflow tax affects misallocation at date 0 through investor retrenchment. Intuitively if $\zeta_{Sn} > 0$, then taxing outflows discourages inflows because investors value the ability to repatraite capital, for example due to low domestic productivity or high marginal value of wealth.

**Inflow versus outflow taxes.** Proposition 18 highlights an important difference between the efficacies of inflow and outflow regulation. Let us assume that $\chi_{S0n} - \vartheta_{0n} < 0$ and $\chi_{S1n} - \vartheta_{1n} > 0$, consistent with taxes on inflows and outflows (rather than subsidies) being optimal, and noting that all results that follow apply if we swap sign and apply a subsidy rather than a tax. Suppose the two taxes are equally effective in discouraging quantities, that is $\frac{\partial I_{S0n}}{\partial p_{0n}} = \frac{\partial I_{S1n}}{\partial p_{1n}} < 0$. In this case, Proposition 18 tells us that an inflow tax is more valuable than an outflow tax if

$$\left| \frac{\chi_{S1n} - \vartheta_{1n}}{\chi_{S0n} - \vartheta_{0n}} \right| < \frac{1 - \zeta_{Sn}}{1 - \omega_{Sn}}. \tag{21}$$

Equation (21 tells us that the relative value of the two taxes is determined by a comparison of misallocation against the two dates against the indirect targeting of the instrument. Intuitively, an inflow control is more likely to be valuable than an outflow control if misallocation is larger at date 0 than at date 1. Notably, greater misallocation at date 0 is neither necessary nor sufficient for an inflow tax to dominate an outflow tax, as the relative desirability also depends on investor flight and retrenchment. If there is no flight or retrenchment, $\omega_{Sn} = \zeta_{Sn} = 0$, then the extent of misallocation summarizes the benefits of the taxes.

Equation (21 shows that inflow taxes are valuable when investor flight is high relative to investor retrenchment, $\omega_{Sn} > \zeta_{Sn}$, potentially even if date 1 misallocation is larger. The intuition can be seen from the limiting case of $\omega_{Sn} = 0$: if investors are perfectly flighty, then an inflow tax also functions as an outflow tax because a unit of inflows translates one-for-one to a unit of outflows.

Unsurprisingly, misallocation at date 0, $\chi_{S0n} - \vartheta_{0n} < 0$, pushes for inflow taxes to be desirable, while misallocation at date 1, $\chi_{S1n} - \vartheta_{S1n} > 0$, pushes for outflow taxes to be desirable. What is notable about this relationship is that the relative desirability of the two depends on Flight and Retrenchment.

If there is no flight or retrenchment, $\omega_{Sn} = \zeta_{Sn} = 0$, then the relative benefit of each tax is simply the degree of misallocation in that market. Conversely at the other extreme, $\omega_{Sn} = \zeta_{Sn} = 1$, *there is no difference between the two taxes*. Intuitively this arises because with complete flight, a marignal until of inflows leads one-to-one to a marignal unit of outflows. An inflow tax thus dominates an outflow tax. By contrast, an outflow tax is particularly valuable if retrenchment is strong relative to flight.

Although Proposition 18 provides analysis of capital control benefits based on *average* flight and retrenchment across investors, it also in the background provides an interesting perspective on differential targeting of inflow and outflow taxes among heterogeneous (e.g., safe and flighty) investors. Inflow taxes target all investors at date 0 but have a strong effect on the outflows of flighty investors at date 1 relative to safe investors. In contrast, outflow taxes target all investors at date 1 but have a strong effect on inflows of retrenching investors at date 0. If an SOE knows the composition of investors but cannot easily identify what specific investors are of what type, choice between the two instruments implicitly describes a choice of what type(s) of investors to discriminate against. For example, an outflow tax provides an incentive compatible method of differentially screening out risky investors in favor of safe investors.[25]

A clean example of this idea comes from revisiting the example of safe and flighty investors from above. It is easy to see here that outflow regulation only affects flighty investors, since they are the sole drivers of date 1 outflows. 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. It suggests a potential advantage of outflow regulation as a method of screening out flighty investors in favor

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

of safe investors, in environments where inflows are socially benefical but outflows are socially destructive.

# 8 Incomplete Regulation with Multiple Regulators and Common Agency

A common scenario for incomplete regulation arises when there are multiple regulators and common agency. For example, in international settings countries typically have regulatory jurisdiction over domestic agents, as well as partial regulatory jurisdiction over foreign agents for their domestic activities. In this section, we extend our setting to study the problem of multiple regulators with common agency.[26]

## 8.1 Model with Multiple Regulators

We model $K > 1$ independently operating regulators. Regulator $k \in K$ possesses complete regulatory jurisdiction over agents $i \in \mathcal{I}_k \subset \mathcal{I}$. We denote $S_k = \mathcal{I} \backslash \mathcal{I}_k$ the unregulated set of $k$ and $S \equiv \cap S_k$ the set of agents that are unregulated by any regulator. The sets $\mathcal{I}_k$ need not be disjoint. Regulator $k$ also has a (possibly empty) partial set of instruments she can choose optimally over her unregulated set $S_k$. We denote $p_i = q + \sum_k \tau_{k,i}$, where $\tau_{k,i}$ is taxes applied by regulator $k$ to agent $i$. We denote $\tau_{-k,i} = \sum_{\ell \neq k} \tau_{\ell,i}$ the vector of taxes applied by all regulators apart from $k$ to agent $i$.

We denote $\mathcal{V}_k$ to be the set of "valued" agents that appear in regulator $k$'s objective function, that is

$$W_k = \sum_{i \in \mathcal{V}_k} \mu_i \omega_{k,i} V_i.$$

We assume that the set of valued agents is a superset of regulated agents, $\mathcal{I}_k \subset \mathcal{V}_k$.[27] We assume finally that $\cup \mathcal{V}_k = \mathcal{I}$, that is all agents are valued by some regulator, consistent with ruling out collective redistribution motives. We rule out redistributive motives through appropriate welfare weights within the set of agents regulator $k$ values.

The model of an individual regulator proceeds as in the baseline model. We assume all taxes are revenue-generating for individual regulators to streamline analysis. Assuming revenue neutrality for the unregulated set of non-valued agents would simply introduce extra motivations for excess regulation of unvalued agents (Clayton and Schaab 2022).

## 8.2 Allocative Values

We begin by characterizing the allocative value of a market from the perspective of regulator $k$. Formally, this is the same new producer exercise as in Proposition 4, where the new producer is

[26] In doing so, we build upon work by Korinek (2017) and Clayton and Schaab (2022). The former allows for limited instruments but not common agency, whereas the latter allows for a specific form of common agency.

[27] For example, this assumption is common in settings where multiple countries have benevolent governments that both value and fully regulate their domestic agents (e.g., Korinek (2017), Clayton and Schaab (2022)).

owned by the set of agents $\mathcal{V}_k$ that regulator $k$ values. We obtain the following result.

**Proposition 19.** *The allocative value of a market to regulator k is*

$$\vartheta_k = \mathbb{E}_i\big(p_i - \tau_{-k,i}\big) + \boldsymbol{P}\,\mathbb{C}ov_i\big(\boldsymbol{D}_i\,, \chi_i - \tau_{-k,i}\big) + \boldsymbol{P}\,\mathcal{E}_k$$

Proposition 19 generalizes the definition of allocative value of a market to the environment with multiple regulators. There are two essential distinctions from the single regulator case.

The first difference is that when calculating the allocative value, regulator $k$ uses the transaction price/CMRS *net* of other regulator's taxes. Intuitively, regulator $k$ treats the taxes of other regulators as if they were higher marginal costs that push down the effective CMRS to an agent. Thus, the regulator $k$ has a lower allocative value on average when other regulators are applying positive taxes.

The second key difference is that regulator $k$ only directly internalizes externalities $\mathcal{E}_k$ that apply to domestic agents. Note that the planner does in fact partially (indirectly) internalize spillovers to agents $i \notin \mathcal{V}_k$ that arises through their demand functions $\boldsymbol{D}_i$, which accounts for market price responses.

Given the allocative value of a market, the allocative value of a trade is defined analogously: $\theta_k = \vartheta_k - q$.

### 8.3 Optimal Regulation

The problem of regulator $k$ is now necessarily a problem of incomplete regulation, as in Section 5. The problem proceeds similarly to Section 5, except that there are now three types of agents regulator $k$ could apply regulation to. First, regulator $k$ can apply complete regulation to the subset $\mathcal{I}_k$ of regulated, valued agents. Second, regulator $k$ can apply partial regulation to the subset $\mathcal{V}_k \cap S_k$ of unregulated, valued agents. Third, regulator $k$ can apply partial regulation to the subset $S_k \backslash \mathcal{V}_k$ of unregulated, unvalued agents. The following result characterizes the regulatory choices of regulator $k$ for each type of agent.

**Proposition 20.** *With multiple regulators and incomplete regulation:*

1. *The CMRS $\chi_i = p_i$ of regulated, valued agents is equal to the allocative value of a market for their regulator plus taxes of other regulators,*

$$\chi_i = \vartheta_k + \tau_{-k,i}, \qquad \forall i \in \mathcal{I}_k.$$

2. *Regulator k's optimal policy encodes a targeting rule:*

$$0 = \boldsymbol{D}_{S_k}\mathbb{E}_{i\in S_k}[\chi_i - (\vartheta_k + \tau_{-k,i})] + \mu_S\mathbb{C}ov_{i\in S_k}\big(\boldsymbol{D}_i\,, \chi_i - (\vartheta_k + \tau_{-k,i})\big) + \mathcal{E}_k.$$

3. *For any instrument $\tau_{k,im}$ on a valued unregulated agent,*

$$0 = \boldsymbol{D}^p_{im}[p_i - (\vartheta_k + \tau_{-k,i})]$$

4. *For any instrument $\tau_{k,im}$ on an unvalued unregulated agent,*

$$0 = I_{im} + \boldsymbol{D}^p_{im}[p_i - (\vartheta_k + \tau_{-k,i})].$$

Proposition 20 extends the analysis of our baseline model to environments with multiple regulators. It highlights both simiilarities and important differences relative to the baseline model.

As in our baseline model, regulator $k$ sets the CMRS of a regulated valued agent equal to that agent's transaction price. In this environment, that is equal to the allocative value of a market to regulator $k$, plus taxes applied by all other regulators. Unlike in the baseline model, the CMRS of regulated agents is not necessarily equated. In particular if two regulated agents feature different total regulation $\tau_{-k,i}$ from other regulators than $k$, then the two agents have different CMRS in equilibrium. The intuition is that from Proposition 19: regulator $k$ perceives $\tau_{-k,i}$ as additional costs. Regulator $k$ thus starts from her baseline allocative value of a market, $\vartheta_k$, and then adds in the taxes applied by other regulators to find the desired CMRS of a regulated agent.

Although regulator $k$ does not necessarily equalize the CMRS across regulated agents, nevertheless her targeting rule takes an analogous form to the targeting rule of equation (13): it trades off misallocation among unregulated agents against externalities. Misallocation, however, takes a subtly different form from the baseline model, $\chi_i - (\vartheta_k + \tau_{-k,i})$. Intuitively, this formulation of misallocation states that misallocation of agent $i$ is relative to a *hypothetical* regulated agent who faced the same total taxes $\tau_{-k,i}$ from other regulators. Thus the targeting rule captures the same intuition as in the baseline model, up to the adjustment that each agent's misallocation is evaluated relative to the distortions introduced by other regulators.

Interestingly, because the social optimum sets the CMRS of agent $i$ equal to the allocative value of the market plus that of other regulators, this unwinds the effect from the allocative value of the market that the planner discounted taxes applied by other regulators. This means that as before, regulated agents drop out of the targeting rule. In the language of the previous paragraph, every regulated agent is also its own hypothetical regulated agent, meaning its misallocation is zero.

As in Proposition 11, the optimal rule of regulator $k$ is divided into two components. The above describes how the optimal allocative value $\vartheta_k$ is determined from the targeting rule. The second component of the decision problem involves using partial regulations of unregulated agents to minimize misallocation for partially regulated agents.

The third part of Proposition 20 parallels the targeting rule of Proposition 11: the targeting rule for optimal partial regulation of an unregulated valued agent is a targeting rule for her misallocation relative to a hypothetical regulated agent with her same external taxes from other regulators. The intutiion follows analogously to Proposition 11 and to the logic above.

The final part of Proposition 20 is the targeting rule for partial regulation of an unregulated unvalued agent. This final component is nearly identical to the targeting rule for an unregulated valued agent, except it adds in the direct effect of the tax revenue $I_{im}$ collected. Intuitively, because agent $i$ is not valued, regulator $k$ internalizes the revenue benefit but does not net out the revenue cost to the agent. This leads to a distortion in the targeting rule relative to that for unregulated valued agents, that increases with the size of the position that agent takes in market $m$.

# 9 Conclusion

We study regulation with externalities and regulation in a general equilibrium economy. Regulators with complete instruments face no trade-off between managing misallocation and managing externalities. With incomplete instruments, we characterize the trade-off regulators face a trade-off between managing misallocation and managing externalities. Interestingly, regulation takes a two-part approach: partial regulation reduces misallocation between unregulated and fully regulated agents, while complete regulation of fully regulated agents trades off misallocation against externalities.

We leverage our framework to address important policy questions. Foremost, we provide a classification scheme that allows regulators to evaluate the welfare consequences of new unregulated entrants or of extending new regulation to existing players or activities. We show that each of these classifications is characterized by simple and estimable statistics of agent's activities and arbitrage responses. 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, Viral V, Philipp Schnabl, and Gustavo Suarez**. 2013. Securitization without risk transfer. *Journal of Financial Economics*, 107(3):515–536.

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

**Avdjiev, Stefan, Bryan Hardy, Sebnem Kalemli-Ozcan, and Luis Servén**. 2018. Gross Capital Flows by Banks, Corporates and Sovereigns.

**Baqaee, David Rezza and Emmanuel Farhi**. 2019. The macroeconomic impact of microeconomic shocks: Beyond Hulten's theorem. *Econometrica*, 87(4):1155–1203.

—. 2020. Productivity and misallocation in general equilibrium. *The Quarterly Journal of Economics*, 135(1):105–163.

**Begenau, Juliane and Tim Landvoigt**. 2020. Financial Regulation in a Quantitative Model of the Modern Banking System.

**Bengui, Julien and Javier Bianchi**. 2019. Macroprudential Policy with Leakages.

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

**Bianchi, Javier, Enrique G Mendoza, Samuel G Hanson, Andrei Shleifer, Jeremy C Stein, and Robert W Vishny**. 2018. Optimal Time-Consistent Macroprudential Policy. *Journal of Financial Economics*, 126(2):588–634.

**Bigio, Saki and Jennifer La'O**. 2020. Distortions in production networks. *The Quarterly Journal of Economics*, 135(4):2187–2253.

**Broner, Fernando, Tatiana Didier, Aitor Erce, and Sergio L Schmukler**. 2013. Gross capital flows: Dynamics and crises. *Journal of Monetary Economics*, 60(1):113–133.

**Buchak, Greg, Gregor Matvos, Tomasz Piskorski, and Amit Seru**. 2018. Fintech, regulatory arbitrage, and the rise of shadow banks. *Journal of Financial Economics*, 130(3):453–483.

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

**Caballero, Ricardo J and Alp Simsek**. 2019. A Model of Fickle Capital Flows and Retrenchment. *Journal of Political Economy. Forthcoming*.

**Chretien, Edouard and Victor Lyonnet**. 2020. Are Traditional and Shadow Banks Symbiotic?

**Claessens, Stijin, Zoltan Pozsar, Lev Ratnovski, and Manmohan Singh**. 2012. Shadow Banking: Economics and Policy. *IMF Staff Discussion Note*.

**Clayton, Christopher and Andreas Schaab**. 2022. Multinational Banks and Financial Stability. *Quarterly Journal of Economics*, 137(3):1681–1736.

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

**Coppola, Antonio, Matteo Maggiori, Brent Neiman, and Jesse Schreger**. 2019. Redrawing the Map of Global Capital Flows : The Role of Cross-Border Financing and Tax Havens.

**Davila, Eduardo**. 2021. Optimal Financial Transaction Taxes.

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

**Dávila, Eduardo and Andreas Schaab**. 2022. Welfare Accounting.

**Dávila, Eduardo and Ansgar Walther**. 2022. Corrective Regulation with Imperfect Instruments. Technical report, National Bureau of Economic Research.

**De Marco, Filippo, Marco Macchiavelli, and Rosen Valchev**. 2019. International Diversification and Information Heterogeneity.

**Dupuit, Jules**. 1844. On the measurement of the utility of public works. *International Economic Papers*, 2(1952):83–110.

**ECB**. 2018. SSM Supervisory Manual.

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

**Farhi, Emmanuel, Mikhail Golosov, and Aleh Tsyvinski**. 2009. A theory of liquidity and regulation of financial intermediation. *Review of Economic Studies*, 76(3):973–992.

**Farhi, Emmanuel and Jean Tirole**. 2021. Shadow Banking and the Four Pillars of Traditional Financial Intermediation. *Review of Economic Studies*, 88(6):2622–2653.

**Feldman, Ron J and Ken 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.

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

**Gorton, Gary, Andrew Metrick, Andrei Shleifer, and Daniel K Tarullo**. 2010. Regulating the shadow banking system. *Brookings Papers on Economic Activity*, (2):261–312.

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

**Harberger, Arnold C**. 1964. The measurement of waste. *The American Economic Review*, 54(3):58–76.

**He, Ping, Zehao Liu, and Chengbo Xie**. 2018. In the Shadow of Shadow Banking.

**Hebert, Benjamin M.** 2022. Externalities as Arbitrage.

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

**Hsieh, Chang-Tai and Peter J Klenow**. 2009. Misallocation and manufacturing TFP in China and India. *The Quarterly journal of economics*, 124(4):1403–1448.

**Huang, Ji**. 2018. Banking and shadow banking. *Journal of Economic Theory*, 178:124–152.

**Hulten, Charles R**. 1978. Growth accounting with intermediate inputs. *The Review of Economic Studies*, 45(3):511–518.

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

**Jenkin, Fleeming H.C.** 1872. No Title. *Proceedings of the Royal Society of Edinburgh*, pp. 618–31.

**Jones, Charles I**. 2013. Misallocation, input-output economics, and economic growth. In *Advances in economics and econometrics: Tenth world congress*, volume 2, pp. 419–458. Cambridge University Press Cambridge, UK.

**Kiyotaki, Nobuhiro and John Moore**. 1997. Credit Cycles. *Journal of Political Economy*, 105(2):211–248.

**Korinek, Anton**. 2017. Currency Wars or Efficient Spillovers? A General Theory of International Policy Cooperation.

**Liu, Ernest**. 2019. Industrial policies in production networks. *The Quarterly Journal of Economics*, 134(4):1883–1948.

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

**Maggiori, Matteo, Brent Neiman, and Jesse Schreger**. 2019. International Currencies and Capital Allocation. *Journal of Political Economy. Forthcoming*.

**Marshall, Alfred**. 1890. Principles of Economics, 8th edn (1920). *London, Mcmillan*.

**Martinez-Miera, David and Rafael Repullo**. 2019. Markets, Banks, and Shadow Banks.

**Midrigan, Virgiliu and Daniel Yi Xu**. 2014. Finance and misallocation: Evidence from plant-level data. *American economic review*, 104(2):422–458.

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

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

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

**Plantin, Guillaume**. 2014. Shadow Banking and Bank Capital Regulation.

**Restuccia, Diego and Richard Rogerson**. 2008. Policy distortions and aggregate productivity with heterogeneous establishments. *Review of Economic dynamics*, 11(4):707–720.

**Shen, Leslie Sheng**. 2019. Global vs. Local Banking: Firm Financing in a Globalized Financial System.

# A Proofs

## A.1 Proof of Lemma 2

Consider such a transaction, denoted $t_m$. The welfare benefit to agent $i$ is

$$\frac{\partial \mathcal{L}_i}{\partial t_m} = \frac{\partial \mathcal{U}_i}{\partial I_{im}} - \hat{p}_m \frac{\partial \mathcal{U}_i}{\partial I_{i1}} = \chi_i - \hat{p}_m > 0$$

for any $\hat{p}_m \in (\chi_{jm}, \chi_{im})$. Similarly, the welfare benefit to agent $j$ is

$$\frac{\partial \mathcal{L}_j}{\partial t_m} = \hat{p}_m \frac{\partial \mathcal{U}_j}{\partial I_{i1}} - \frac{\partial \mathcal{U}_i}{\partial I_{im}} = \hat{p}_m - \chi_{jm} > 0.$$

Thus, both parties benefit from the transaction.

## A.2 Proof of Lemma 3

By Envelope Theorem, we can write

$$\frac{\partial V_i}{\partial q_m} = -\lambda_i I_{im} + \frac{\partial \mathcal{U}_i}{\partial q_m}.$$

The result then comes from summing across agents. Note that we have $\frac{\partial \mathcal{U}_i}{\partial q_m} = \frac{\partial \Lambda_i \Gamma_i}{\partial q_m}$.

## A.3 Proof of Proposition 4

The social planner of this simple problem picks the vector of prices $q$ and wealth distribution $w_i$ to maximize welfare, taking as given the exogenous wedges $\tau_i$, the wealth constraint, and market clearing. Formally, this Lagrangian is

$$\mathcal{L}(q, \tau, w, e) = W(q, \tau, w, e) + \lambda \left[ qe + \sum_i \mu_i \tau_i I_i - \sum_i w_i \right] + \sum_m \nu_m \left[ e_m - \sum_i \mu_i I_{im} \right],$$

where $I_i = I_i(q + \tau_i, q, w_i)$. Therefore, we have

$$\vartheta_m = \frac{\partial \mathcal{L}}{\partial e_m} = \lambda q_m + \nu_m$$

Next we characterize the Lagrange multiplier $\nu_m$. Taking the vector of derivatives in $q$ of the Lagrangian and evaluating around $e = 0$,

$$0 = \nabla_q W + \sum_i \mu_i \nabla_q I_i \left[ \lambda \tau_i - \nu \right]$$

Using definitions $\boldsymbol{D}_i = \nabla_q I_i$, $\mathcal{E} = \frac{1}{\lambda}\nabla_q W$ and $\boldsymbol{P} = \boldsymbol{D}_{\mathcal{I}}^{-1}$, then we have

$$\nu = \lambda \boldsymbol{P}\Big(\mathcal{E} + \sum_i \mu_i \boldsymbol{D}_i \tau_i\Big)$$

Thus substituting in and dividing through by $\lambda$ obtains

$$\vartheta = q + \boldsymbol{P}\sum_i \mu_i \boldsymbol{D}_i \tau_i + \boldsymbol{P}\mathcal{E}.$$

Thus rewriting the second term decomposed into an expectation and covariance gives the result.

### A.4 Proof of Corollary 5

The result is an immediate corollary of Proposition 4, since $\tau_i = 0$ implies $\chi_i = p_i = q$ is constant across agents and hence the covariance is zero, while $\mathcal{E} = 0$ implies the last term is zero.

### A.5 Proof of Corollary 6

The result is an immediate corollary of Proposition 4, since $\tau_i = \tau$ implies $\chi_i = p_i = q + \tau$ is constant across agents and hence the covariance is zero, while $\mathcal{E} = 0$ implies the last term is zero.

### A.6 Proof of Proposition 7

The proof follows from Proposition 4. A trade $t_m$ is equivalent to a new producer in $m$ combined with a new negative producer $-q$ in the numeraire. Thus we have $\theta = \vartheta - q\vartheta_1 = \vartheta - q$, giving the result (since the numeraire is not distorted, $\vartheta_1 = 1$).

### A.7 Proof of Proposition 8

The social planner's Lagrangian is analogous to Proposition 4, but with $e = 0$ and with flexible choice over $\tau_i$ for all $i$. The social first-order condition for $\tau_i$ is, by Envelope Theorem,

$$0 = -\mu_i \omega_i \lambda_i I_{im} + \lambda \mu_i I_{im} + \lambda \mu_i \nabla_{p_i} I_i \tau_i - \mu_i \nabla_{p_i} I_i \nu.$$

Thus using weights $\omega_i \lambda_i = \lambda$ across agents, we have $\tau_i = \frac{1}{\lambda}\nu$. Next, we can apply an analogous derivative in $q$ to obtain, as in Proposition 4,

$$0 = \mathcal{E} + \sum_i \boldsymbol{D}_i [\lambda \tau_i - \nu].$$

Thus having set $\tau_i = \frac{1}{\lambda}\nu$ for all $i$, we obtain $\mathcal{E} = 0$. Finally, let us return to the definition of allocative value of a market from Proposition 4,

$$\vartheta = \mathbb{E}_i\big(p_i\big) + \boldsymbol{P}\,\mathbb{C}ov_i\big(\boldsymbol{D}_i\,,\chi_i\big) + \boldsymbol{P}\,\mathcal{E}.$$

Since $\chi_i = p_i = q + \tau_i = q + \frac{1}{\lambda}\nu$ is constant across $i$, then the covariance is zero. Since $\mathcal{E} = 0$, the last term is zero. Finally since the transaction price is constant across agents, the expected transaction price is equal to the transaction price for any individual agent, and hence we have $p_i = \vartheta$ for all $i$.

Lastly, the fact that the tax is equal to the allocative value of a trade follows frm Proposition 7. Since misallocation and externalities are zero, we have $\theta = p_i - q_i = \tau_i$ for all $i$, completing the proof.

## A.8 Proof of Proposition 9

The social planner's Lagrangian is analogous to Proposition 8, but with flexible choice over $\tau_i$ only for $i \notin S$. As in the proof of Proposition 8, we have $\tau_i = \frac{1}{\lambda}\nu$. As in the proof of Proposition 4, we have $\nu = \lambda \boldsymbol{P}\Big(\mathcal{E} + \sum_i \mu_i \boldsymbol{D}_i \tau_i\Big)$. Therefore, addint $q$ to both sides and substituting in, we have

$$p_i = \boldsymbol{P}\sum_j \mu_i \boldsymbol{D}_j p_j + \boldsymbol{P}\mathcal{E}, \qquad i \notin S.$$

Finally, note that from the proof of Proposition 4, the LHS is the allocative value of a market, and therefore $p_i = \vartheta$. Since $\chi_i = p_i$, then $\chi_i = \vartheta$ for $i \notin S$. The allocative value of a trade follows as before. Finally, we can form the targeting rule from the definition of allocative value,

$$\vartheta = \boldsymbol{P}\sum_i \mu_i \boldsymbol{D}_i \chi_i + \boldsymbol{P}\mathcal{E}.$$

Substituting in $\chi_i = \vartheta$ for $i \notin S$ and subtracting, we obtain

$$0 = \boldsymbol{P}\sum_i \mu_i \boldsymbol{D}_i[\chi_i - \vartheta] + \boldsymbol{P}\mathcal{E}.$$

Since $\boldsymbol{P}$ drops out (full rank among non-numeraire goods), and using the decomposition

$$\sum_i \mu_i \boldsymbol{D}_i[\chi_i - \vartheta] = \mu_S \sum_i \frac{\mu_i}{\mu_S}\boldsymbol{D}_i\Big[\chi_i - \vartheta - \mathbb{E}_{i\in S}\Big[\chi_i - \vartheta\Big] + \mathbb{E}_{i\in S}\Big[\chi_i - \vartheta\Big]\Big] = \boldsymbol{D}_S \mathbb{E}_{i\in S}[\chi_i - \vartheta] + \mu_S \mathbb{C}ov_i\big(\boldsymbol{D}_i\,,\chi_i\big),$$

which gives the result.

### A.9 Proof of Proposition 10

Define the expenditure function $e_i(p_i, q, \overline{U}_i)$ and Hicksian demand $h_i(p_i, q, \overline{U}_i)$ as the solutions to the expenditure minimization problem. As usual, Marshallian and Hicksian demand are related by

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

From here, we use $p_i = q + \tau_i$ and totally differentiate the above equation for good $m$ in $q_m$ to obtain

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

$$\frac{\partial I_{im}}{\partial p_{im}} + \frac{\partial I_{im}}{\partial q_m} + \frac{\partial I_{im}}{\partial w_i}\left(\frac{\partial e_i}{\partial p_{im}} + \frac{\partial e_i}{\partial q_m}\right) = \frac{\partial h_{im}}{\partial p_{im}} + \frac{\partial h_{im}}{\partial q_m}.$$

Recall that $\frac{\partial I_{im}}{\partial p_{im}} + \frac{\partial I_{im}}{\partial q_m} = \boldsymbol{D}_{i,mm}$ by definition. It remains only to characterize the derivatives of the expenditure function. The Lagrangian of the expenditure minimization problem is

$$\mathcal{L}^E = p_i h_i - \frac{1}{\lambda_i}(U_i(h_i) - \overline{U}) - \frac{1}{\lambda_i}\Lambda_i \Gamma_i(h_i, q),$$

which can be rewritten as

$$\mathcal{L}^E = p_i h_i - \frac{1}{\lambda_i}(\mathcal{U}_i(h_i, q) - \overline{U}).$$

Note that this is the dual problem of utility maximization, and hence constrained utility here is analogous to constrained utility when we use $\overline{U}_i = V_i(p_i, q, w_i)$. Thus, the Envelope Theorem tells us that the total derivative of the expenditure function in $q_m$ is

$$\frac{de_i}{dq_m} = h_{im} - \frac{1}{\lambda_i}\frac{\partial \mathcal{U}_i}{\partial q_m},$$

so that substituting back in obtains the result.

### A.10 Proof of Proposition 11

The social planner's Lagrangian is analogous to Proposition 9, but with additional choice of a set of instruments for $i \in S$. The exact same steps as the proof of Proposition 9 shows that for $i \notin S$, we have $p_i = \vartheta$. Therefore for any instrument $\tau_{im}$ for $i \in S$, we obtain by Envelope Theorem the analogous first order condition

$$0 = -\mu_i \omega_i \lambda_i I_{im} + \lambda \mu_i I_{im} + \lambda \mu_i \nabla_{p_{im}} I_i \tau_i - \mu_i \nabla_{p_{im}} I_i \nu$$

and once again using $\omega_i \lambda_i = \lambda$, defining $\boldsymbol{D}^p_{im} = \nabla_{p_{im}} I_i$, and using $\theta = \frac{1}{\lambda}\nu$, we obtain

$$0 = \boldsymbol{D}^p_{im}[\tau_i - \theta] = \boldsymbol{D}^p_{im}[\chi_i - \vartheta],$$

which gives the result.

### A.11 Proof of Proposition 12

From Propositions 4 and 7, we have that the welfare gains from the new player are $\vartheta e^u + \theta t^u$. From Proposition 9, we have $\vartheta = q + \tau$ and $\theta = \tau$. Therefore, we have welfare gains $(q+\tau)e^u + \tau t^u = qe^u + \tau n^u$.

### A.12 Proof of Proposition 13

Take the planner's Lagrangian. The analogous arguments from the proof of Proposition 11 tell us that, by Envelope Theorem,

$$\Delta_i = \frac{1}{\lambda}\frac{\partial \mathcal{L}}{\partial d\hat{\tau}_i} = -\frac{1}{\lambda}\omega_i\lambda_i I_i + \frac{1}{\lambda}\lambda I_i + \frac{1}{\lambda}\lambda d\hat{\tau}\boldsymbol{D}_i^p\tau_i - \frac{1}{\lambda}d\hat{\tau}\boldsymbol{D}_i^p\nu = d\hat{\tau}\boldsymbol{D}_i^p[\tau_i - \theta] = d\hat{\tau}\boldsymbol{D}_i^p[\chi_i - \vartheta],$$

which completes the proof.

### A.13 Proof of Proposition 14

Take the planner's Lagrangian. The analogous arguments from the proof of Proposition 11 and Proposition 14 tell us that, by Envelope Theorem,

$$\Delta_m = \frac{1}{\lambda}\frac{\partial \mathcal{L}}{\partial d\hat{\tau}_m} = \sum_i \left[-\frac{1}{\lambda}\omega_i\lambda_i I_{im} + \frac{1}{\lambda}\lambda I_{im} + \frac{1}{\lambda}\lambda d\hat{\tau}\boldsymbol{D}_{im}^p\tau_i - \frac{1}{\lambda}d\hat{\tau}\boldsymbol{D}_{im}^p\nu\right] = d\hat{\tau}_m\sum_i \boldsymbol{D}_{im}^p[\chi_i - \vartheta],$$

which completes the proof.

### A.14 Proof of Proposition 15

Suppose an intervention targets the collateral price (holding fixed $p_i$). By Envelope Theorem,

$$\Delta_{im}^q = \frac{1}{\lambda}\frac{\partial \mathcal{L}}{\partial q_m} = \frac{1}{\lambda}\omega_i\frac{\partial \mathcal{U}_i}{\partial q_m} + \frac{1}{\lambda}\lambda\boldsymbol{D}_{im}^q\tau_i - \frac{1}{\lambda}\boldsymbol{D}_{im}^p\nu = \frac{1}{\lambda}\omega_i\frac{\partial \mathcal{U}_i}{\partial q_m} + \boldsymbol{D}_{im}^q[\chi_i - \vartheta],$$

which gives the result.

### A.15 Proof of Proposition 19

We have the Lagrangian

$$\mathcal{L} = \sum_{i\in\mathcal{V}_k}\mu_i\omega_i V_i + \lambda_k\left[q_m e_m + \sum_{i\in\mathcal{I}}\mu_i\tau_{k,i}I_i - \sum_{i\in\mathcal{V}_k}\mu_i w_i\right] + \sum_m \nu_{k,m}\left[e_m - \sum_{i\in\mathcal{I}}\mu_i I\right].$$

Note that the regulator $k$ only values agents $i \in \mathcal{V}_k$, but collects taxes from all agents and internalizes market clearing from all agents. Thus, we have

$$\vartheta_{k,m} = q_m + \frac{1}{\lambda_k}\nu_{k,m}.$$

Next, taking the derivative in $q_m$, we obtain

$$0 = \nabla_q W_k + \sum_{i\in\mathcal{I}} \mu_i \boldsymbol{D}_i[\lambda_k \tau_{k,i} - \nu_k].$$

Therefore defining $\mathcal{E}_k = \frac{1}{\lambda_k}\nabla_q W_k$, we obtain

$$\vartheta_k = \boldsymbol{P}\sum_{i\in\mathcal{I}} \mu_i \boldsymbol{D}_i[q + \tau_{k,i}] + \boldsymbol{P}\mathcal{E}_k.$$

Finally, we substitute in $p_i = q + \sum_k \tau_{k,i}$ and apply the usual covariance decomposition to obtain

$$\vartheta_k = \mathbb{E}_i\big(p_i - \tau_{-k,i}\big) + \boldsymbol{P}\,\mathbb{C}ov_i\big(\boldsymbol{D}_i\,, \chi_i - \tau_{-k,i}\big) + \boldsymbol{P}\,\mathcal{E}_k,$$

where $\tau_{-k,i} = \sum_{\ell\neq k} \tau_{\ell,i}$.

### A.16 Proof of Proposition 20

Given regulator $k$'s Lagrangian,

$$\mathcal{L} = \sum_{i\in\mathcal{V}_k} \mu_i\omega_i V_i + \lambda_k\left[\sum_{i\in\mathcal{I}} \mu_i\tau_{k,i}I_i - \sum_{i\in\mathcal{V}_k}\mu_i w_i\right] + \sum_m \nu_{k,m}\left[-\sum_{i\in\mathcal{I}}\mu_i I\right],$$

then for a regulated agent $i \in \mathcal{I}_k$, we have by Envelope Theorem

$$0 = -\mu_i\omega_i\lambda_i I_i + \lambda_k\left[\mu_i I_i + \mu_i \boldsymbol{D}_i^p \tau_{k,i}\right] - \mu_i \boldsymbol{D}_i^p \nu$$

and therefore we have with $\omega_i\lambda_i = \lambda_k$, $\tau_{k,i} = \frac{1}{\lambda_k}\nu_k$. Following the proof of Proposition 19, we therefore have $\vartheta_{k,m} = q_m + \tau_{k,m}$. Thus, adding in taxes of other regulators, we have

$$p_i = \vartheta_{k,m} + \tau_{-k,i}$$

We can now use this in the targeting rule. Rearranging the allocative value of a market,

$$0 = \boldsymbol{P}\sum_{i\in\mathcal{I}} \mu_i \boldsymbol{D}_i[q + \tau_{k,i} - \vartheta_k] + \boldsymbol{P}\mathcal{E}_k.$$

Now, we know that we have for $i \in \mathcal{I}_k$

$$q + \tau_{k,i} - \vartheta_k = q + \tau_{k,i} - p_i + \tau_{-k,i} = p_i - p_i = 0.$$

Thus, the targeting rule is given by

$$0 = \boldsymbol{P} \sum_{i \in S_k} \mu_i \boldsymbol{D}_i [q + \tau_{k,i} - \vartheta_k] + \boldsymbol{P}\mathcal{E}_k.$$

Now, adding and subtracting $\tau_{-k,i}$ inside the sum, we have

$$0 = \boldsymbol{P} \sum_{i \in S_k} \mu_i \boldsymbol{D}_i [p_i - \vartheta_k - \tau_{-k,i}] + \boldsymbol{P}\mathcal{E}_k.$$

Now taking a valued unregulated agent, we have for $\tau_{k,im}$ by the usual steps

$$0 = \boldsymbol{D}^p_{im} [\lambda_k \tau_{k,i} - \nu_k].$$

And so substituting in $\vartheta_k = q + \frac{1}{\lambda}\nu_k$ and adding and subtracting $\tau_{-k,i}$,

$$0 = \boldsymbol{D}^p_{im} [p_i - (\vartheta_k + \tau_{-k,i})].$$

Finally for an unvalued unregulated agent, we have for $\tau_{k,im}$

$$0 = \lambda_k \boldsymbol{I}_{im} + \boldsymbol{D}^p_{im} [\lambda_k \tau_{k,i} - \nu_k]$$

which reflects that the usual cost no longer appears. Through usual substitution, we thus have

$$0 = I_{im} + \boldsymbol{D}^p_{im} [p_i - (\vartheta_k + \tau_{-k,i})].$$

# B 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 agents, and moreover considers pecuniary externalities. We now allow for non-pecuniary externalities.

Incorporating non-pecuniary externalities requires only a slight modification to the framework above. Agents maximize $U_i(I_i, P)$ subject to the constraint set $\Gamma_i(I_i, P, w_i - \tau_i I_i) \geq 0$, 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. We denote $\mathcal{U}_i = U_i + \Lambda_i \Gamma_i$ to be constrained utility.

The equilibrium aggregates are defined by an equilibrium relationship

$$\Phi(I, P) = 0 \tag{22}$$

where $I = \{I_j\}_j$ is the activities of all agents in the economy. This notation captures the model with pecuniary externalities when equation (22) is a set of market clearing conditions and $P$ is market prices.

**Aggregate Response Matrices.** In the baseline model, we defined $\boldsymbol{D}_i$ as price response matrices. We now need to define an analogous notion, which is $\boldsymbol{D}_i^P = \nabla_P I_i$, that is derivatives in equilibrium aggregate. Similarly, we can define an unregulated demand impact matrix

$$\Theta_S = \sum_{i \in S} \boldsymbol{D}_i^P \nabla_{I_i} \Phi.$$

These are the impacts of changes in $P$ on the constraint set through the activities of unregulated agents. In the baseline model, the impact $\nabla_{I_i}\Phi$ is simply the measure $\mu_i$, and $\Theta_S$ collapses to $\boldsymbol{D}_S$.

**Example 21** (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 agents to absorb residual demand in a market, even though those agents may not be on their first order conditions.[28] 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 agents are forced by their constraint sets to absorb.

## B.1 Optimal Incomplete Regulation

We can now characterize optimal incomplete regulation

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

**Proposition 22.** *In this environment, optimal regulation of a regulated agent is*

$$\tau_i = -\frac{1}{\lambda_i}\nabla_{I_i}\Phi\left(\nabla_P\Phi + \Theta_S\right)^{-1}\mathcal{E}$$

*where* $\mathcal{E} = \nabla_P W$.

Optimal regulation accounts for externalities resulting from changes in aggregates $P$. A change in activities of $i$ leads to an effect $\nabla_{I_i}\Phi$ 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 unregulated agents, given by $\Theta_S$. When aggregates appear in the constraint set only through unregulated demand, as with prices and market clearing, we have $\nabla_P\Phi = 0$ and are left only with the indirect effect, as in baseline model. The total change in both current and future prices multiplies the vector $\mathcal{E}$ of externalities arising from changes in aggregates.

## B.2 Proof of Proposition 22

The Lagrangian of the social planner is given by

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

We obtain by the usual steps

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

We have

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

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

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

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

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