Optimal Financial Transaction Taxes*

Eduardo Dávila†

September 2017

Abstract

This paper characterizes the optimal transaction tax in an equilibrium model of competitive financial markets. As long as investors hold heterogeneous beliefs that are orthogonal to their fundamental trading motives and the planner calculates welfare using any single belief, a strictly positive tax is optimal, regardless of the magnitude of fundamental trading. The optimal tax, which depends on investors' beliefs and asset-demand sensitivities to tax changes, can be implemented by adjusting its value until total volume equals fundamental volume. Under some conditions, the optimal tax is independent of the belief used by the planner to calculate welfare. A calibration of the model that is consistent with empirically estimated volume sensitivities to tax changes and that features a 20% share of non-fundamental trading is associated with an optimal tax on the order of 17bps.

JEL Classification: H21, D61, G18

Keywords: transaction taxes, optimal taxation, behavioral public economics, Tobin tax, belief disagreement

*First draft: May 2013. I am very grateful to my discussant at the BFI Advances in Price Theory Conference, Eric Budish, as well as Philippe Aghion, Adrien Auclert, Robert Barro, Roland Bénabou, Markus Brunnermeier, John Campbell, Raj Chetty, Peter Diamond, Emmanuel Farhi, Xavier Gabaix, Gita Gopinath, Robin Greenwood, Sam Hanson, Oliver Hart, Nathan Hendren, David Laibson, Sendhil Mullainathan, Adriano Rampini, David Scharfstein, Florian Scheuer, Andrei Shleifer, Alp Simsek, Jeremy Stein, Adi Sunderam, Glen Weyl, Wei Xiong, and seminar participants at many institutions and conferences for helpful comments. Financial support from Rafael del Pino Foundation is gratefully acknowledged.

†New York University and NBER. Email: edavila@nyu.edu

1 Introduction

Whether to tax financial transactions or not remains an important open question for public economics that periodically gains broad relevance after periods of economic turmoil. For instance, the collapse of the Bretton Woods system motivated James Tobin's well-known 1972 speech — published as Tobin (1978) — endorsing a tax on international transactions. The 1987 crash encouraged Stiglitz (1989) and Summers and Summers (1989) to argue for implementing a transaction tax, while the 2008 financial crisis spurred further public debate on the issue, leading to a contested tax proposal by the European Commission. However, with the lack of formal normative studies of this topic, a financial transaction tax may still seem like "the perennial favorite answer in search of a question", per Cochrane (2013).

In this paper, I study the welfare implications of taxing financial transactions in an equilibrium model in which financial markets play two distinct roles. On the one hand, financial markets allow investors to conduct *fundamental trading*. Fundamental trading allows the transfer of risks towards those investors more willing to bear them. It also allows for trading on information, liquidity, or life-cycle considerations, as well as trading for market-making or limited arbitrage purposes. On the other hand, financial markets also allow investors to engage in betting or gambling, which I refer to as *non-fundamental trading*.

I model non-fundamental trading by assuming that investors' trades are partly motivated by differences in beliefs, while the planner calculates welfare using a single belief. The discrepancy between the planner's belief and investors' beliefs implies that corrective policies, which can involve taxes or subsidies depending on the primitives, are generically optimal. Three main results emerge from the optimal taxation exercise.

First, the optimal transaction tax can be expressed as a function of investors' beliefs and asset demand sensitivities to tax changes. Specifically, the optimal tax corresponds to one-half of the difference between a weighted average of buyers' beliefs and a weighted average of sellers' beliefs. Beliefs are the key sufficient statistic for the optimal tax, due to the corrective nature of the policy. In general, corrective policies are set to correct marginal distortions, which in this case arise from investors' differences in beliefs.

Second, a simple condition involving the cross-sectional covariance between investors' beliefs and their condition as net buyers or net sellers in the laissez-faire economy determines the sign of the optimal corrective policy. Importantly, I show that when investors' beliefs are orthogonal to their fundamental motives for trading, this condition holds in equilibrium, implying that the optimal policy is a strictly positive tax. Therefore, as long as the planner is aware of the existence of some belief-driven trades, under the assumption that these trades are orthogonal to other fundamental trading motives, a positive transaction tax is optimal.[1] Intuitively, even though introducing a (small) transaction tax shifts investors' portfolio allocations towards their no-trade positions, the reduction in fundamental trading creates a second-order welfare loss — because those trades are done optimally — while the reduction in non-fundamental trading creates a first-order welfare gain. While the mere existence of some form of non-fundamental trading is sufficient to determine the sign of the optimal policy, the magnitude of the optimal tax depends on the structure of the model. In particular, it depends on the relative importance of non-fundamental trading.

Third, the optimal tax turns out to be independent of the belief used by the planner to calculate welfare under certain conditions. This result relies on two assumptions: traded assets are in fixed supply and the planner does not seek to redistribute resources across investors. Intuitively, because the source of welfare losses in this model comes from a distorted allocation of risk, the dispersion in investors' beliefs — but not the absolute level of beliefs — determines the optimal tax.

Because optimal tax characterizations are inherently local to the optimum, I also study the convexity properties of the planner's problem. I establish that the planner's problem is in general non-convex, which implies that different levels of transaction taxes may yield similar levels of welfare. However, I show that this phenomenon can only arise when the composition of marginal investors varies with the tax level. Importantly, when the distribution of trading motives is symmetric — a plausible benchmark — the planner's problem is well-behaved and has a unique optimum.

To further enhance the practical applicability of the results, I provide an implementation of the optimal policy that uses trading volume as an intermediate target, simply by adjusting the tax level until observed volume equals fundamental volume. This alternative approach shifts the planner's informational requirements from measuring investors' beliefs to finding an appropriate estimate of fundamental volume. This approach relies on a novel decomposition of trading volume into fundamental volume, non-fundamental volume, and the tax-induced volume reduction. When the optimal tax is close to zero — the relevant case in practice — I also derive a simple back-of-the-envelope approximation for the optimal tax that relies exclusively on two objects of the laissez-faire economy: the sensitivity of trading volume to tax changes and the share of non-fundamental trading volume.

Next, to provide explicit comparative static results of the optimal tax with respect to primitives and to illustrate the quantitative implications of the model, I parametrize the distribution

[1] Although orthogonality between trading motives is a plausible sufficient condition for an optimal positive tax, it is by no means necessary.

of fundamental and non-fundamental trading motives. Consistent with the main results, when fundamental and non-fundamental trading motives are jointly normally distributed and uncorrelated, the optimal tax is positive. The optimal positive tax is increasing in the ratio of non-fundamental to fundamental trading for any correlation level between trading motives. Moreover, I show that a mean-preserving spread of the distribution of investors' beliefs is associated with a higher optimal tax. The optimal tax is not very sensitive to moderate changes in the correlation between trading motives.

Although the parametrized model is stylized, it is worthwhile to provide a sense of the magnitudes that it generates for different parameters. A calibration of the model that is consistent with empirically estimated volume sensitivities to tax changes and that features a 20% share of non-fundamental trading is associated with an optimal tax on the order of 17bps (0.173%). Although these are plausible values, which could be refined by future measurement work, I conduct a sensitivity analysis and provide a menu of optimal taxes for different shares of non-fundamental trading. For instance, when the share of non-fundamental trading volume is 50% or 10%, the model predicts optimal taxes of 57bps (0.57%) or 8bps (0.08%) respectively.[2]

Finally, I establish the robustness of the results. I first characterize the optimal tax for more general specifications of beliefs and utility. The optimal tax formula of the baseline model remains valid as a first-order approximation to the optimal tax in the general case, validating the analysis in the rest of the paper up to a first-order for any specification of beliefs and preferences. I briefly describe in the paper how the results extend to environments with pre-existing trading costs, imperfect tax enforcement, multiple traded assets, production, and dynamics. I study these and other extensions in detail in the Online Appendix.

Related Literature This paper contributes to the growing literature on behavioral welfare economics, recently synthesized and expanded in Mullainathan, Schwartzstein and Congdon (2012). This paper is related to Gruber and Koszegi (2001) and O'Donoghue and Rabin (2006), who characterize optimal corrective taxation when agents misoptimize because of self-control or limited foresight. Within this literature, the work by Sandroni and Squintani (2007) and Spinnewijn (2015), who characterize optimal corrective policies when agents have distorted beliefs, is closely related. While those papers respectively study optimal policies in insurance markets and frictional labor markets, this paper derives new insights in the context of financial market trading. The recent work by Farhi and Gabaix (2015) systematically studies optimal taxation

[2]Given existing estimates of trading volume elasticities to tax changes in stock markets, the simplest rule-of-thumb implies an optimal tax of the same order of magnitude of the share of non-fundamental trading, when expressed in basis points. That is, a 10%, 20%, or 40% share of non-fundamental volume is associated (approximately) with an optimal tax of 10bps, 20bps, or 40bps.

with behavioral agents, while Campbell (2016) advocates for incorporating behavioral insights into optimal policy prescriptions.[3]

This paper belongs to the literature that follows Tobin's proposal of introducing transaction taxes to improve the societal performance of financial markets. Although Tobin's speech largely focused on foreign exchange markets, it has become customary to refer to any tax on financial transactions as a "Tobin tax". Stiglitz (1989) and Summers and Summers (1989) verbally advocate for a financial transaction tax, with Ross (1989) taking the opposite view. Roll (1989) and Schwert and Seguin (1993) contrast costs and benefits of such proposal. Umlauf (1993), Campbell and Froot (1994), several chapters in ul Haq, Kaul and Grunberg (1996), and Jones and Seguin (1997) are representative samples of empirical work in the area. See McCulloch and Pacillo (2011) and Burman et al. (2016) for recent surveys. See also the recent theoretical work by Dang and Morath (2015).

The theory in this paper differs substantially from that in Tobin (1978). Tobin postulates that prices are excessively volatile and that a transaction tax is a good instrument to reduce price volatility. This paper shows instead that transaction taxes are a robust instrument to reduce trading volume, but that their effect on asset prices is a priori indeterminate. The normative results in this paper rely on the fact that a reduction of trading volume improves the allocation of risk in the economy from the planner's perspective. An alternative leading argument defending the desirability of a transaction tax argues that it can prevent investors from learning information which will eventually be publicly revealed. See Stiglitz (1989) for an elaboration of this "foreknowledge" argument and Budish, Cramton and Shim (2015) for a recent assessment in the context of high-frequency trading.[4]

This paper is most directly related to the growing literature that evaluates welfare under belief disagreements in financial markets. Weyl (2007) is the first to study the efficiency of arbitrage in an economy in which some investors have mistaken beliefs. Brunnermeier, Simsek and Xiong (2014) propose a criterion to evaluate welfare in models with belief heterogeneity: they assess efficiency by evaluating welfare under a convex combination of the beliefs of the investors in the economy. Gilboa, Samuelson and Schmeidler (2014) and Gayer et al. (2014) present refined Pareto criteria that identify negative-sum betting situations. No-Betting Pareto requires that there exists a single belief that, if shared, implies that all agents are better off by trading. Unanimity Pareto requires that every agent perceives, using his own belief, that all agents are better off by trading. These papers seek to identify outcomes related to zero-sum

[3]The recent work by Gerritsen (2016), Lockwood (2016), Lockwood and Taubinsky (2017) and Moser and Olea de Souza e Silva (2017), focused on optimal income and capital taxation, also features behavioral agents.

[4]There are other theories in which financial market intervention is optimal — see, among many others, Scheuer (2013) or Dávila and Korinek (2017). However, these arguments do not necessarily rely on transaction taxes.

speculation, but do not discuss policy measures to limit trading, which is the raison d'être of this paper. In the same spirit, Posner and Weyl (2013) advocate for financial regulation grounded on price-theoretic analysis, which is exactly my goal with this paper.[5] Blume et al. (2013) propose a different criterion in which a planner evaluates welfare under the worst case scenario among all possible belief assignments. They quantitatively analyze several restrictions on trading but do not characterize optimal policies. Heyerdahl-Larsen and Walden (2014) propose an alternative criterion in which the planner does not have to take a stand on which belief to use, within a reasonable set, to assess efficiency. I relate my results to these criteria when appropriate.

Many papers explore the positive implications of speculative trading due to belief disagreements, following Harrison and Kreps (1978). Scheinkman and Xiong (2003) analyze the positive implications of a transaction tax in a model with belief disagreement, but they do not draw normative conclusions. Panageas (2005) and Simsek (2013) study implications for production and risk sharing of speculative trading motives. Xiong (2012) surveys this line of work.[6] In the sense that some trades are not driven by fundamental considerations, this paper also relates to the literature on noise trading that follows Grossman and Stiglitz (1980). However, the standard noise trading formulation makes it hard to understand how noise traders react to taxes and how to evaluate their welfare. By using heterogeneous beliefs to model non-fundamental trading, this paper sidesteps these concerns. Given the additive nature of corrective taxes (see e.g. Sandmo (1975); Kopczuk (2003)), there is little loss of generality by not directly incorporating dispersed information to the model. In recent work, Dávila and Parlatore (2016) systematically show that transaction costs/taxes do not affect information aggregation under appropriate conditions, although they discourage endogenous information acquisition.[7]

Finally, the literature on transaction costs is formally related to this paper, since a transaction tax is similar to a transaction cost from a positive point of view. This literature studies the positive effects of transaction costs on portfolio choices and equilibrium variables like prices and volume. I refer the reader to Vayanos and Wang (2012) for a recent comprehensive survey. While those papers focus on the positive implications of exogenously given transaction costs/taxes, this paper

[5] A growing literature exploits market design tools to study normative issues in market microstructure. See, in particular, Budish, Cramton and Shim (2015), who briefly discuss the possibility of taxing financial transactions, as well as Baldauf and Mollner (2014, 2015).

[6] This paper is also directly related to the vast literature on behavioral finance, which includes Black (1986), De Long et al. (1990), Barberis, Shleifer and Vishny (1998), and Hong and Stein (1999), among many others. This body of work is recently surveyed by Barberis and Thaler (2003) and Hong and Stein (2007).

[7] Therefore, without a specific prior on whether there is too much or too little information acquisition in the laissez-faire economy, all the qualitative insights of the paper about the sign of the optimal tax would go through unchanged. In recent work, Vives (2017) studies an environment in which a positive tax is welfare improving by correcting information acquisition.

studies the welfare effects of a transaction tax and its optimal determination. I explicitly relate the positive results of the paper to this work in the text when appropriate.

Outline Section 2 introduces the model and Section 3 studies its positive predictions. Section 4 conducts the normative analysis, delivering the main results. Under specific parametric assumptions on trading motives, Section 5 provides explicit comparative statics for the optimal tax and illustrates the quantitative predictions of the model, while Section 6 studies the robustness of the results. Section 7 concludes. Proofs and derivations are in the Appendix.

2 Model

In the absence of transaction taxes, the environment of this paper resembles Lintner (1969), who relaxes the CAPM by allowing for heterogeneous beliefs among investors.

Investors There are two dates $t = 1, 2$ and there is a unit measure of investors. Investors are indexed by i and distributed according to a continuous probability distribution F such that $\int dF(i) = 1$.

Investors choose their portfolio optimally at date 1 and consume at date 2. They maximize expected utility with preferences that feature constant absolute risk aversion. Therefore, each investor maximizes

$$\mathbb{E}_i \left[U_i \left(W_{2i}\right)\right] \quad \text{with} \quad U_i \left(W_{2i}\right) = -e^{-A_i W_{2i}}, \tag{1}$$

where (1) already imposes that investors consume all terminal wealth, that is $C_{2i} = W_{2i}$. The parameter $A_i > 0$, which represents the coefficient of absolute risk aversion $A_i \equiv -\frac{U_i''}{U_i'}$, can vary in the distribution of investors. The expectation is indexed by i because investors hold heterogeneous beliefs.

Market structure and beliefs There is a riskless asset in elastic supply that offers a gross interest rate normalized to 1. There is a single risky asset in exogenously fixed supply $Q \geq 0$. The price of the risky asset at date 1 is denoted by P_1 and is quoted in terms of an underlying good (dollar), which acts as numeraire. The initial holdings of the risky asset at date 1, given by X_{0i}, are arbitrary across the distribution of investors. As a whole, investors must hold the total supply Q, therefore $\int X_{0i} dF(i) = Q$. Investors face no constraints when choosing portfolios: they can borrow and short sell freely.

The risky asset yields a dividend D at date 2, which is normally distributed with some mean and variance $\mathbb{V}ar[D]$. An investor i believes that D is normally distributed with mean $\mathbb{E}_i[D]$ and variance $\mathbb{V}ar[D]$, that is,

$$D \sim_i N\left(\mathbb{E}_i[D], \mathbb{V}ar[D]\right).$$

Investors do not learn from each other, or from the price, and agree to disagree in the Aumann (1976) sense.[8] For now, the pattern of belief disagreement, which is a primitive of the model, can vary in the distribution of investors.[9] Nothing prevents investors from having correct beliefs; those investors can represent market makers or (limited) arbitrageurs.

Hedging needs Every investor has a stochastic endowment at date 2, denoted by E_{2i}, which is normally distributed and potentially correlated with D. This endowment captures the fundamental risks associated with the normal economic activity of the investor. A given investor's exposure to those risks is captured by the covariance $\mathbb{C}ov\left[E_{2i}, D\right]$, which is known to all investors. The magnitudes of the hedging needs are arbitrary across the distribution of investors. Without loss of generality, I assume that $\mathbb{E}\left[E_{2i}\right] - \frac{A_i}{2}\mathbb{V}ar\left[E_{2i}\right] = 0$ and normalize the initial endowment E_{1i} to zero for all investors.

Trading motives Summing up, there are four reasons to trade in this model:

(i) Different hedging needs: captured by $\mathbb{C}ov\left[E_{2i}, D\right]$

(ii) Different risk aversion: captured by A_i

(iii) Different initial asset holdings: captured by X_{0i}

(iv) Different beliefs: captured by $\mathbb{E}_i\left[D\right]$

The first three correspond to fundamental reasons for trading: sharing risks among investors, transferring risks to those more willing to bear them, or trading for life cycle or liquidity needs. Trading on different beliefs is the single source of non-fundamental trading in the model. For positive purposes, all four reasons are equally valid: the assumed welfare criterion makes the last reason non-fundamental.[10] I assume throughout that all four cross-sectional distributions have bounded moments and that the cross-sectional dispersion of risk aversion coefficients is small.

At times, to sharpen several results, I impose Assumption [S]. I explicitly state when Assumption [S] is used.

[8] Two arguments can be used to justify the assumption of investors who disagree about the mean — not other moments — of the distribution of payoffs. First, it is commonly argued that second moments are easier to learn. In particular, with Brownian uncertainty, second moments can be learned instantly. Second, as shown in Section 6, up to a first-order approximation, only the mean of the distribution of payoffs appears explicitly in the approximated optimal tax formula.

[9] A common prior model in which investors receive a purely uninformative signal (noise), but pay attention to it, maps one-to-one to the environment in this paper. Alternatively, investors could neglect the informational content of prices, as in the cursed equilibrium model of Eyster and Rabin (2005).

[10] Having multiple sources of fundamental trading, while not necessary, is important to show that all sources enter symmetrically in optimal tax formulas.

Assumption. [S] **(Symmetry)** *The cross-sectional distribution of beliefs, hedging needs, and initial asset holdings is symmetric. Investors have identical preferences, so* $A_i = A$.

Assumption [S] simplifies the solution of the model and allows for sharper characterizations. It does not restrict the levels of fundamental trading, non-fundamental trading, or the cross-sectional correlation between fundamental and non-fundamental trading motives.

Policy instrument: a linear financial transaction tax

This paper follows the Ramsey approach of solving for an optimal policy under a restricted set of instruments. The single policy instrument available to the planner is an anonymous linear financial transaction tax τ paid per dollar traded in the risky asset. A change in asset holdings of the risky asset $|X_{1i} - X_{0i}|$ at a price P_1 faces a total tax in terms of the numeraire, due at the time the transaction occurs, for both buyers and sellers, of

$$\tau \left|P_1\right| \left|\Delta X_{1i}\right|, \tag{2}$$

where $|\Delta X_{1i}| \equiv |X_{1i} - X_{0i}|$. Total tax revenue generated by the transaction is thus $2\tau \left|P_1\right| \left|\Delta X_{1i}\right|$. I restrict τ to be in a closed interval $[\underline{\tau}, \overline{\tau}]$ such that $-1 \leq \underline{\tau}$ and $\overline{\tau} \leq 1$. In general, the use of the absolute value for the price in (2) is necessary because asset prices can be negative in this model, as they are in many markets for derivative contracts. I soon will restrict the analysis to situations with strictly positive prices.

Analytically, nothing prevents the tax from being negative (a subsidy). However, an anonymous linear trading subsidy can never be implemented: investors would continuously exchange assets, making infinite profits. Therefore, any reference to trading subsidies in this paper implicitly assumes that the planner can rule out these "wash trades".

Linearity, anonymity, and enforcement I restrict the analysis to linear taxes with the intention of being realistic. The conventional justification for the use of linear (as opposed to non-linear) taxes in this environment is that linear taxes are the most robust to sophisticated trading schemes. For example, a lump-sum tax per trade creates incentives to submit a single large order. Alternatively, quadratic taxes create incentives to split orders into infinitesimal pieces. These concerns, which are shared with other non-linear tax schemes, are particularly relevant for financial transaction taxes, given the high degree of sophistication of many players in financial markets and the negligible costs of splitting orders given modern information technology.

I assume that transaction taxes must apply across-the-board to all market participants and cannot be conditioned on individual characteristics, which implies that the planner's problem is a second-best problem. A planner with the ability to distinguish good trades from bad trades could achieve the first-best by taxing harmful trades on an individual basis: this is an implausible assumption.

Furthermore, I assume that investors cannot avoid paying transaction taxes, either by trading secretly or by moving to a different exchange. This behavior is optimal when the penalties associated with evasion are sufficiently large, provided the taxable event is appropriately defined. I discuss the implications of imperfect tax enforcement for the optimal tax policy in the Online Appendix.

Revenue rebate and redistribution Lastly, since this paper focuses on the corrective (Pigovian) effects of transaction taxes and not on the ability of this tax to raise fiscal revenue, I assume that tax proceeds are rebated lump-sum to investors.[11] Under CARA utility, the rebate that each investor receives is irrelevant to determine trading behavior, although variations in the individual level of the transfers impact wealth and marginal utility. For clarity, I assume that every (group of) investor(s) i receives a rebate T_{1i} equal to his (their) own tax liability, that is $T_{1i} = \tau |P_1| |\Delta X_{1i}|$. Investors do not internalize the rebate since they are assumed to be small. It is important that tax revenue is rebated and not wasted. This paper does not address how to spend tax revenues.

To separate efficiency considerations from distributional considerations, one could assume that the planner has access to lump-sum transfers to redistribute wealth across investors ex-ante. This is a standard assumption in models of corrective taxation with concave utility, and corresponds to a Kaldor-Hicks interpretation — see Weyl (2016) for a description of the approach and references. Instead, until I revisit this issue in Section 5, I assume that the planner maximizes the sum of investors' certainty equivalents, which does not require the use of ex-ante transfers. Both approaches yield identical results.

Investors' budget constraints Consumption/wealth of a given investor i at $t = 2$ is composed of the stochastic endowment E_{2i}, the stochastic payoff of the risky asset $X_{1i}D$ and the return on the investment in the riskless asset. This includes the proceeds from the net purchase/sale of the risky asset $(X_{0i} - X_{1i}) P_1$, the total tax liability $-\tau |P_1| |\Delta X_{1i}|$, and the lump-sum transfer T_{1i}. It can be expressed as

$$W_{2i} = E_{2i} + X_{1i}D + (X_{0i}P_1 - X_{1i}P_1 - \tau |P_1| |\Delta X_{1i}| + T_{1i}) . \tag{3}$$

[11]Broadly defined, there are two types of taxes: those levied with the aim of raising revenue and those levied with the aim of correcting distortions. This paper exclusively studies corrective taxation. Sandmo (1975) shows that corrective taxes and optimal revenue raising taxes are additive; see also Kopczuk (2003). This paper does not consider the additional benefits of corrective taxes generated by "double-dividend" arguments. Those arguments, surveyed by Goulder (1995) in the context of environmental taxation, apply directly to transaction taxes.

For the rest of the paper, I assume that the fundamentals of the economy are such that the price of the risky asset is strictly positive, that is $P_1 > 0$. The Online Appendix provides a sufficient condition. Hence, I use P_1 instead of $|P_1|$. This assumption simplifies the number of cases to consider and is without loss of generality.

Definition. (Equilibrium) A competitive equilibrium with taxes is defined as a portfolio allocation X_{1i} for every investor, a price P_1, and a set of lump-sum transfers T_{1i} such that: a) investors maximize expected utility in X_{1i}, subject to their budget constraint (3), b) the price P_1 is such that the market for the risky asset clears, that is, $\int \Delta X_{1i} dF(i) = 0$, and c) tax revenues are rebated lump-sum to investors.

3 Equilibrium

I first solve for investors' portfolio demands. Subsequently, I characterize the equilibrium price and allocations.

Investors' problem In this model, every investor effectively solves the following mean-variance problem to determine his risky asset demand

$$\max_{X_{1i}} \left[\mathbb{E}_i[D] - A_i \mathbb{C}ov[E_{2i}, D] - P_1\right] X_{1i} - \tau P_1 |\Delta X_{1i}| - \frac{A_i}{2} \mathbb{V}ar[D] X_{1i}^2. \tag{4}$$

As formally shown in the Appendix, the problem solved by investors is well-behaved. Given a price P_1, investor i's optimal net asset demand $\Delta X_{1i}(P_1) = X_{1i}(P_1) - X_{0i}$ is given by

$$\Delta X_{1i}(P_1) = \begin{cases} \Delta X_{1i}^{+}(P_1) = \frac{\mathbb{E}_i[D] - A_i \mathbb{C}ov[E_{2i}, D] - P_1(1+\tau)}{A_i \mathbb{V}ar[D]} - X_{0i}, & \text{if } \Delta X_{1i}^{+}(P_1) > 0 \quad \text{Buying} \\ 0, & \text{if } \Delta X_{1i}^{+}(P_1) \leq 0,\ \Delta X_{1i}^{-}(P_1) \geq 0 \quad \text{No Trade} \\ \Delta X_{1i}^{-}(P_1) = \frac{\mathbb{E}_i[D] - A_i \mathbb{C}ov[E_{2i}, D] - P_1(1-\tau)}{A_i \mathbb{V}ar[D]} - X_{0i}, & \text{if } \Delta X_{1i}^{-}(P_1) < 0 \quad \text{Selling.} \end{cases} \tag{5}$$

Figure 1 illustrates the optimal portfolio demand $X_{1i}(P_1)$ for an investor i as a function of the asset price P_1.[12] The presence of linear transaction taxes modifies the optimal portfolio allocation along two dimensions. First, a transaction tax is reflected as a higher price $P_1(1+\tau)$ paid by buyers and a lower price $P_1(1-\tau)$ received by sellers. Hence, for a given price P_1, a higher tax reduces the net demand of both buyers and sellers at the intensive margin.

Second, a linear tax implies that some investors decide not to trade altogether, creating an inaction region. If the initial holdings of the risky asset X_{0i} are not too far from the optimal allocation without taxes $\frac{\mathbb{E}_i[D] - A_i \mathbb{C}ov[E_{2i}, D] - P_1}{A_i \mathbb{V}ar[D]}$, an investor decides not to trade. Only when $\tau = 0$

[12] I use both in the exposition net and gross asset demands $\Delta X_{1i}(P_1)$ and $X_{1i}(P_1)$. They are related through the expression $\Delta X_{1i}(P_1) = X_{1i}(P_1) - X_{0i}$.

the no-trade region ceases to exist. The envelope theorem, which plays an important role when deriving the optimal tax results, is also key to generating the inaction region, as originally shown in Constantinides (1986). Intuitively, an investor with initial asset holdings close to his optimum experiences a second-order gain from a marginal trade but suffers a first-order loss when a linear tax is present, making no-trade optimal.

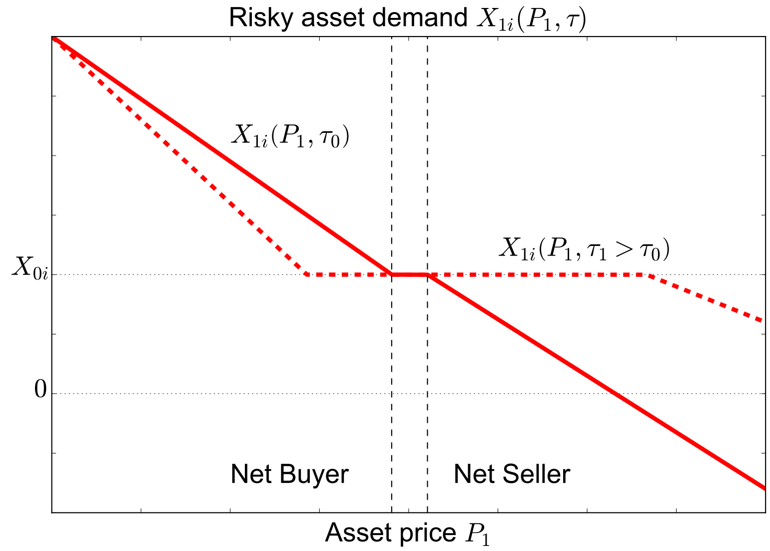

Figure 1: Risky asset demand

Equilibrium characterization Given the optimal portfolio allocation derived in (5) and the market clearing condition $\int \Delta X_{1i}(P_1)\, dF(i) = 0$, the equilibrium price of the risky asset satisfies the following implicit equation for P_1

$$P_1 = \frac{\int_{i\in\mathcal{T}(P_1)} \left(\frac{\mathbb{E}_i[D]}{\mathcal{A}_i} - A\left(\mathbb{C}ov\left[E_{2i}, D\right] + \mathbb{V}ar\left[D\right] X_{0i}\right)\right) dF(i)}{1+\tau\left(\int_{i\in\mathcal{B}(P_1)} \frac{1}{\mathcal{A}_i} dF(i) - \int_{i\in\mathcal{S}(P_1)} \frac{1}{\mathcal{A}_i} dF(i)\right)}, \tag{6}$$

where $A \equiv \left(\int_{i\in\mathcal{T}(P_1)} \frac{1}{A_i} dF(i)\right)^{-1}$ is the harmonic mean of risk aversion coefficients for active investors and $\mathcal{A}_i \equiv \frac{A_i}{A}$ is the quotient between the risk aversion coefficient of investor i and the harmonic mean.[13] The notation $i \in \mathcal{T}(P_1)$ indicates that the domain of integration is the set of investors who actively trade in equilibrium at a price P_1. Analogously, the notation $\mathcal{B}(P_1)$ and $\mathcal{S}(P_1)$ respectively denotes the set of buyers and sellers at a given price P_1. Equation (5) determines the identity of the investors in each of the sets. Because the sets $\mathcal{T}(P_1)$, $\mathcal{B}(P_1)$, and $\mathcal{S}(P_1)$, as well as A and $\mathcal{A}_i$, depend on the equilibrium price, Equation (6) provides an implicit characterization of P_1. Intuitively, only marginal investors directly determine the equilibrium price. As shown in Lemma 1 below, Equation (6) has a unique solution for P_1 whenever there is trade in equilibrium.

[13] It should be clear from their definitions that $\mathcal{A}_i$ and A are also functions of P_1.

The numerator of the equilibrium price has two components. The first term is a weighted average of the expected payoff of the risky asset. The second term is a risk premium, determined by the product of price and quantity of risk. The price of risk is given by the harmonic mean of risk aversion coefficients A. The quantity of risk consists of two terms. The first one is the sum of covariances of the risky asset with the endowments $\int_{i\in\mathcal{T}(P_1)} \mathbb{C}ov\left[E_{2i}, D\right] dF\left(i\right)$. The second one is the product of the variance of the risky asset $\mathbb{V}ar\left[D\right]$ with the number of shares initially held by investors $\int_{i\in\mathcal{T}(P_1)} X_{0i} dF\left(i\right)$.

Lemma 1 synthesizes the main positive results of the model. Lemma 1 shows that the model is well-behaved and that a transaction tax is a robust instrument to reduce trading volume. More broadly, it suggests that theories in which transaction taxes are desirable must rely on a mechanism by which reducing trading volume is welfare improving.[14]

Lemma 1. (Competitive equilibrium with taxes)

a) (Existence/Uniqueness) An equilibrium always exists for a given τ. The equilibrium is (essentially) unique.

b) (Volume response) Trading volume is decreasing in τ.

c) (Price response) The asset price P_1 increases (decreases) with τ if

$$\int_{i\in\mathcal{B}(P_1)} \frac{1}{\mathcal{A}_i} dF\left(i\right) \leq (\geq) \int_{i\in\mathcal{S}(P_1)} \frac{1}{\mathcal{A}_i} dF\left(i\right). \tag{7}$$

Under Assumption [S], the asset price P_1 is invariant to the level of the transaction tax.

Lemma 1 shows that an equilibrium always exists and that it is essentially unique. Intuitively, existence is guaranteed because asset demands are everywhere downward sloping. Trading volume is uniquely pinned down in any equilibrium. The equilibrium price P_1 is also uniquely pinned down in any equilibrium with positive trading volume. Every no-trade equilibrium is associated with a range of prices consistent with such equilibrium. Because of this single dimension of indeterminacy, I say that the equilibrium is essentially unique.

Trading volume always goes down when transaction taxes increase. Even though a change in the transaction tax can change the asset price and induce some sellers to sell more (and some buyers to buy more), this effect is never strong enough to overcome the direct substitution effect induced by the tax.

[14] Existing empirical evidence is consistent with the prediction that trading volume decreases after an increase in transaction taxes/costs, although tax evasion may be at times a confounding factor. The empirical evidence regarding its effect on prices is mixed. Some studies find an increase in price volatility, but others find no significant change or even a reduction. Asset prices usually fall at impact following a tax increase, but seem to recover over time. See the review articles by Campbell and Froot (1994), Habermeier and Kirilenko (2003), McCulloch and Pacillo (2011), Burman et al. (2016), and the recent work on the European Transaction Tax by Colliard and Hoffmann (2013) and Coelho (2014).

The condition that determines the sign of $\frac{dP_1}{d\tau}$ in Equation (7) corresponds to the difference between buyers' and sellers' price elasticities. When this term is positive, increasing τ reduces the buying pressure by more than the selling pressure, reducing the equilibrium price, and vice versa. When the difference between buyers' and sellers' elasticities is zero, the equilibrium price is independent of the tax. In particular, for the symmetric benchmark in which Assumption [S] holds, buyers' and sellers' price elasticities are everywhere identical, implying that the equilibrium price is invariant to the tax level.

4 Normative analysis

After solving for the equilibrium allocations and the equilibrium price for a given tax, I characterize the welfare maximizing value of τ.

4.1 Welfare criterion

In order to aggregate individual preferences, I assume that the planner focuses on maximizing the sum of investors' certainty equivalents. This approach is standard in normative problems. However, to conduct any normative analysis in this paper, one must also take a stand on how to evaluate social welfare when investors hold heterogeneous beliefs, which is a controversial issue.[15]

I assume that the planner calculates indirect utilities using a single probability distribution about payoffs. This distribution will be necessarily different than the one held by most investors. Initially, I solve the case where the planner maximizes welfare using an arbitrary distribution. Subsequently, I point out the conditions under which the optimal policy does not depend on the distribution used by the planner. In those cases, only the consistency requirement that there exists a single distribution of payoffs is relevant.

This approach is paternalistic because the planner does not respect subjective beliefs. Interestingly, when the belief chosen by the planner does not affect the optimal policy, criticisms of paternalistic policies on the grounds that the planner must be better informed than the individuals do not apply.

Belief disagreement among investors can be interpreted as a device to model departures from full rationality in information processing. Instead of modeling a specific rationality failure, this

[15] In addition to the work discussed in the literature review, see Kreps (2012), Cochrane (2014), and Duffie (2014) for some reflections on this topic. Duffie (2014), in particular, challenges policy treatments of speculative trading motivated by differences in beliefs. He raises philosophical/axiomatic challenges and a practical challenge. This paper directly addresses the practical challenge, which questions the ability of enforcement agencies to set policies when some trades are belief-motivated while other trades arise from welfare enhancing activities.

paper takes the distribution of beliefs as a primitive. Consistent with that interpretation, two arguments justify the welfare criterion adopted in this paper. The first relies on the idea that rational investors cannot agree to disagree when their posteriors are common knowledge — see the discussion in Morris (1995). How can the planner respect investors' beliefs when they are inconsistent with one another? If we assume that there is a single correct belief but different investors hold different beliefs, all of them (but one) must be wrong. Alternatively, a veil of ignorance interpretation is also consistent with this welfare criterion. If investors acknowledge that they may wrongly hold different beliefs, they would happily implement, at an ex-ante stage, a tax policy that curtails trading.

Moreover, perhaps the strongest defense of this welfare criterion comes from considering altruism. In this economy, given his own belief, leaving aside price changes, each investor perceives that a transaction tax reduces his individual welfare. However, if investors are at all altruistic towards others, they will agree on implementing a positive tax. An altruistic investor perceives that a small tax creates a first-order gain for all other investors in the economy, at the cost of a second-order private loss. This approach is consistent with the political philosophy tradition of deliberative democracy, in which individuals think and decide together about what serves the common interest, provided this common interest does not harm much any given individual. The approach followed by the planner in this paper fits legal traditions that consider speculation as fraudulent, because each individual perceives a gain at the expense of others, as well as religious precepts questioning gambling.

Remark. General approach to normative problems with heterogeneous beliefs. The normative approach used in this paper can be applied to other environments. Every normative problem in which a planner does not respect investors' beliefs can be approached in two stages. First, one can characterize the solution to the planner's problem for a given planner's belief. This exercise identifies wedges in investors' decisions, as well as the optimal policy. Although overruling investors' beliefs creates a mechanical rationale for intervention, the form of the intervention and the welfare losses induced by belief distortions are not obvious and must be studied on a case-by-case basis. This paper focuses on this first stage. Next, if the optimal policy turns out to be independent of the belief used by the planner — as in Proposition 1 below— no further analysis is needed. If not, a second stage involves choosing the belief used by the planner. In those cases, the welfare criteria proposed in Brunnermeier, Simsek and Xiong (2014) or Gilboa, Samuelson and Schmeidler (2014), among others, can be used. The two approaches are complementary.

4.2 Optimal transaction tax

After introducing the welfare criterion used by the planner, I characterize the properties of the optimal tax policy. The main inputs to the planner's objective function are investors' certainty equivalents from the planner's perspective. These are denoted by $\hat{V}_i(\tau)$ and correspond to

$$\hat{V}_i(\tau) \equiv \left(\mathbb{E}[D] - A_i \mathbb{C}ov\left[E_{2i}, D\right] - P_1(\tau)\right) X_{1i}(\tau) + P_1(\tau) X_{0i} - \frac{A_i}{2} \mathbb{V}ar[D] \left(X_{1i}(\tau)\right)^2,$$

where $X_{1i}(\tau)$ and $P_1(\tau)$ represent equilibrium outcomes that are in general functions of τ. Note that the expectation used to calculate the individual certainty equivalents does not have an individual subscript i, because it is taken using the planner's belief, which correspond to $\mathbb{E}[D]$.

Social welfare, denoted by $V(\tau)$, corresponds to the sum of investors' certainty equivalents and is given by

$$V(\tau) = \int \hat{V}_i(\tau)\, dF(i),$$

The optimal tax is given by $\tau^* = \arg\max_\tau V(\tau)$. Proposition 1 introduces the main results of the paper.

Proposition 1. (Optimal financial transaction tax)

a) [Optimal tax formula] The optimal financial transaction tax τ^* satisfies

$$\tau^* = \frac{\Omega_{\mathcal{B}(\tau^*)} - \Omega_{\mathcal{S}(\tau^*)}}{2}, \tag{8}$$

where $\Omega_{\mathcal{B}(\tau)}$ is a weighted average of buyers' expected returns, given by

$$\Omega_{\mathcal{B}(\tau^*)} \equiv \int_{i \in \mathcal{B}(\tau^*)} \omega_i^{\mathcal{B}}(\tau^*) \frac{\mathbb{E}_i[D]}{P_1(\tau^*)} dF(i), \quad \textit{with} \quad \omega_i^{\mathcal{B}}(\tau^*) \equiv \frac{\frac{dX_{1i}}{d\tau}(\tau^*)}{\int_{i \in \mathcal{B}(\tau^*)} \frac{dX_{1i}}{d\tau}(\tau^*)\, dF(i)}, \tag{9}$$

and $\Omega_{\mathcal{S}(\tau)}$ is a weighted average of sellers' expected returns, analogously defined.

b) [Sign of the optimal tax] In general, a positive tax is optimal when optimistic investors are net buyers and pessimistic investors are net sellers in the laissez-faire economy. Formally,

$$\textit{if } \left.\frac{dV}{d\tau}\right|_{\tau=0} = \mathbb{C}ov_F\left(\mathbb{E}_i[D], -\left.\frac{dX_{1i}}{d\tau}\right|_{\tau=0}\right) > 0, \textit{ then } \tau^* > 0. \tag{10}$$

As long as some investors have heterogeneous beliefs and fundamental and non-fundamental trading motives are orthogonally distributed across the population of investors, this condition is endogenously satisfied, implying that the optimal corrective policy is a strictly positive tax.

c) [Irrelevance of planner's belief] The optimal financial transaction tax does not depend on the distribution of beliefs used by the planner to calculate welfare.

Optimal tax formula Proposition 1a) shows that the expression for the optimal tax formula can be written exclusively as a function of investors' beliefs and demand sensitivities. Because the equilibrium price, asset demand sensitivities, and the identity of the active traders are endogenous to the level of the tax, Equation (8) only provides an implicit representation for τ^*. This is a standard feature of optimal taxation exercises. Below, I provide conditions under which the optimal tax formula has a unique solution.

The corrective (Pigovian) nature of the tax explains why investors' beliefs and demand sensitivities are the relevant sufficient statistics to set the optimal tax. Pigovian logic suggests that corrective taxes must be set to target marginal distortions, which in this particular case arise from investors' beliefs. Ideally, the planner would like to target each individual belief distortion with an investor-specific tax.[16] However, because the planner employs a second-best policy instrument — a single linear tax — demand sensitivities $\frac{dX_{1i}}{d\tau}$ determine the weights given to individual beliefs in the optimal tax formula. The planner gives more weight to the distortions of the most tax-sensitive investors.[17] Note that the weights assigned to buyers $\omega_i^{\mathcal{B}}$ and sellers $\omega_i^{\mathcal{S}}$ add up to one and that investors who do not trade do not affect the optimal tax at the margin.

When Assumption [S] holds, there exists a unique optimal tax that satisfies the simpler condition

$$\tau^* = \frac{\mathbb{E}_{\mathcal{B}(\tau^*)}\left[\frac{\mathbb{E}_i[D]}{P_1}\right] - \mathbb{E}_{\mathcal{S}(\tau^*)}\left[\frac{\mathbb{E}_i[D]}{P_1}\right]}{2}, \tag{11}$$

where $\mathbb{E}_{\mathcal{B}(\tau^*)}[\cdot]$ and $\mathbb{E}_{\mathcal{S}(\tau^*)}[\cdot]$ respectively denote cross-sectional expectations for the set of active buyers and sellers at the optimal τ^*. In this case, demand sensitivities drop out of the optimal tax formula, providing a tractable benchmark in which the optimal tax is exclusively a function of the average belief of buyers and sellers.

It is clear that if all investors agree about the expected payoff the risky asset, so that $\mathbb{E}_i[D]$ is constant, the optimal tax is $\tau^* = 0$. Equations (8) and (11) suggest that an increase in the dispersion of beliefs across investors, by widening the gap between buyers' and sellers' expected returns, calls for a higher optimal transaction tax. In Section 5, I explicitly link the value of the optimal tax to primitives of the distribution of fundamental and non-fundamental motives for trading.

Convexity Although Equation (8) must be satisfied at the level of τ that maximizes the planner's objective function, the planner's problem may have multiple local optima. I formally

[16] See the Online Appendix for a characterization of the first-best policy with unrestricted instruments, which calls for investor-specific corrective policies.

[17] The presence of demand sensitivities in optimal corrective tax formulas goes back to Diamond (1973), who analyzes corrective taxation with restricted instruments in a model of consumption externalities. See also Rothschild and Scheuer (2016) for a recent application of similar principles.

show that the planner's objective function is concave when investors only adjust their trading behavior on the intensive margin. I also show that non-concavities on the planner's objective function can only arise when the composition of investors who actively trade varies with the tax level. Under symmetry, changes in the composition of marginal investors cancel out, guaranteeing that the planner's objective function is concave. I summarize these results in the following Lemma.

Lemma 2. (Convexity of planning problem)

a) The planner's objective function may be non-concave only if the composition of marginal investors varies with the tax level.

b) Assumption [S] is a sufficient condition for the planner's problem to be well-behaved. In that case, Equation (8) *characterizes the unique optimal transaction tax.*

Although in my numerical simulations I focus on the symmetric case, which is a reasonable benchmark, the fact that the planner's problem is non-convex could have economic significance. A planner should always consider whether a marginal tax increase improves social welfare given the set of marginal investors. Suppose a small positive tax is optimal, due to some non-fundamental trading in the economy, but that increasing the optimal tax from low to intermediate levels reduces social welfare by primarily restricting fundamental trading. However, if for even higher tax levels fundamental investors drop out of the market and all marginal trades are driven by heterogeneous beliefs, it may be optimal to increase the tax level further. In such cases, both low and high taxes may be associated with comparable welfare levels.

Finally, note that quadratic taxes, often used as a tractable approximation to linear taxes, do not generate extensive margin adjustments, since it is generically optimal for all investors to trade. A model with quadratic taxes would fail to capture the possibility of non-concavities in the planner's objective function.

Sign of the optimal tax Proposition 1b) shows that the optimal policy corresponds to a strictly positive tax when, in the laissez-faire economy, optimistic investors (those with a high $\mathbb{E}_i[D]$) are on average net buyers (for which $-\frac{dX_{1i}}{d\tau}\big|_{\tau=0} > 0$) of the risky asset, while pessimistic investors are on average net sellers. When Assumption [S] holds, Equation (10) simplifies to the more intuitive condition for a positive tax:

$$\text{if} \quad \mathbb{E}_{\mathcal{B}(\tau=0)}\left[\frac{\mathbb{E}_i[D]}{P_1}\right] > \mathbb{E}_{\mathcal{S}(\tau=0)}\left[\frac{\mathbb{E}_i[D]}{P_1}\right], \text{ then } \tau^* > 0, \tag{12}$$

which highlights that identifying the difference in beliefs between buyers and sellers in the zero tax economy is sufficient to establish the sign of the tax.

If all trading is driven by disagreement, Equation (10) trivially holds — optimists buy and pessimists sell. However, because investors also trade due to fundamental reasons, it is possible for an optimistic investor to be a net seller in equilibrium and vice versa.

Proposition 1b) not only establishes the necessary condition for the optimal tax to be positive, but it also provides a natural sufficient condition for (12) to be satisfied, justifying the language of this paper based on the positive tax case. As long as some investors hold heterogeneous beliefs, and if the distribution of beliefs across investors is independent of the distribution of fundamental trading motives (risk aversion, hedging needs, and initial positions), a strictly optimal tax is positive.[18] Intuitively, in expectation, an optimistic (pessimistic) investor is more likely to be a buyer (seller) in equilibrium. Hence, unless the pattern of fundamental trading specifically counteracts this force, we expect the covariance in (12) to be negative. Orthogonality between fundamental and non-fundamental trading motives is a sufficient condition for an optimal positive tax, but it is not necessary.

This result puts fundamental and non-fundamental trading on different grounds when setting the optimal tax. The presence of non-fundamental trades orthogonal to fundamental trades implies that it is optimal to have a positive tax, regardless of the relative importance of both types of motives. That is, a positive tax is optimal even when most trading is fundamental. Intuitively, a small transaction tax equally reduces fundamental and non-fundamental trades. However, the reduction in trading generates a first-order gain for optimistic buyers and pessimistic sellers, while the same reduction in trading only generates a second-order loss to fundamental investors.

Under which conditions could a trading subsidy be optimal? If many optimists happen to be sellers of the risky asset in the laissez-faire equilibrium, instead of buyers, the optimal policy may be a subsidy. A documented example of this trading pattern involves workers who are overoptimistic about their own company's performance and who fail to sufficiently hedge their labor income risk. They are natural sellers of the risky asset, as hedgers, but they sell too little of it. In that case, a transaction tax, by pushing them towards no-trade, has a negative first-order welfare effect on them. When Equation (10) holds, this phenomenon is not too prevalent among investors.

Irrelevance of planner's belief Proposition 1c) establishes that the optimal tax is independent of the belief used by the planner to calculate welfare. This is a surprising and appealing result because, even though the planner does not respect investors' beliefs when assessing welfare, the planner does not have to impose a particular belief, but only a consistency condition on beliefs. This result also avoids the use of criteria that use a convex combination of beliefs, like Brunnermeier, Simsek and Xiong (2014), a single belief that if shared, makes every agent better off, as in Gilboa, Samuelson and Schmeidler (2014), or worst case scenarios, like

[18] Alternatively, one could argue on empirical grounds that Equation (10) holds. The evidence accumulated in the behavioral finance literature, recently surveyed by Barberis and Thaler (2003) and Hong and Stein (2007), suggests that investors' beliefs drive a non-negligible fraction of purchases/sales.

Blume et al. (2013). Any belief used by the planner is associated with the same optimal policy.

Two features of the economic environment are essential to this result. First, the risky asset is in fixed supply, which implies that if one investor holds more shares of the risky asset, some other investor must be holding less. Formally, $\int \frac{dX_{1i}}{d\tau} dF(i) = 0$. In that case, only relative asset holdings matter for welfare. Intuitively, the key economic outcome in this model corresponds to the allocation of risk among investors, which is determined by the dispersion on beliefs, but not by the average belief.

Second, the planner does not use the transaction tax with the purpose of redistributing resources across investors. Intuitively, the linearity of investors' certainty equivalents on beliefs combined with the fact that the planner equally weights welfare gains/losses across investors in certainty equivalent terms guarantee that the optimal tax does not depend on the difference between investors and the planner's belief, but only on the belief dispersion among investors. See Section 5 for how introducing distributional concerns affects this result.

4.3 An alternative implementation using trading volume

As described above, the distribution of beliefs is the key sufficient statistic that determines the optimal tax. Given that directly recovering investors' beliefs is challenging, I now propose an alternative approach that implements the optimal policy using trading volume as an intermediate target.[19] Under this alternative approach, the planner must adjust the tax rate until total trading volume equals fundamental volume.

Trading volume, measured in dollars and expressed as a function of the tax level, formally corresponds to

$$\mathcal{V}(\tau) = P_1(\tau) \int_{i \in \mathcal{B}(\tau)} \Delta X_{1i}(\tau) \, dF(i), \tag{13}$$

where only the net trades of buyers are considered, to avoid double counting. Proposition 2 provides a decomposition of trading volume into different components and describes a new implementation of the optimal policy that compares total trading volume with fundamental volume.

Proposition 2. (Trading volume implementation)

a) [Trading volume decomposition] Trading volume, as defined in (13), can be decomposed as follows

$$\underbrace{\mathcal{V}(\tau)}_{\text{Total volume}} = \underbrace{\Theta_F(\tau)}_{\text{Fundamental volume}} + \underbrace{\Theta_{NF}(\tau)}_{\text{Non-fundamental volume}} - \underbrace{\Theta_\tau(\tau)}_{\text{Tax-induced volume reduction}},$$

[19] I use the intermediate target nomenclature by analogy to the literature on optimal monetary policy. In this model, equilibrium trading volume becomes an intermediate target to implement optimal portfolio allocations.

where $\Theta_F(\tau)$, $\Theta_{NF}(\tau)$, and $\Theta_\tau(\tau)$ are defined in the Appendix for the general case. Under Assumption [S], they correspond to

$$\Theta_F(\tau) = \frac{1}{2}\left|\frac{dX_{1i}}{d\tau}\right| A\left(\int_{i\in\mathcal{S}} \left(\mathbb{C}ov\left[E_{2i}, D\right] - \mathbb{V}ar\left[D\right] X_{0i}\right) dF(i) - \int_{i\in\mathcal{B}} \left(\mathbb{C}ov\left[E_{2i}, D\right] - \mathbb{V}ar\left[D\right] X_{0i}\right) dF(i)\right)$$

$$\Theta_{NF}(\tau) = \frac{1}{2}\left|\frac{dX_{1i}}{d\tau}\right| \left(\int_{i\in\mathcal{B}} \mathbb{E}_i\left[D\right] dF(i) - \int_{i\in\mathcal{S}} \mathbb{E}_i\left[D\right] dF(i)\right)$$

$$\Theta_\tau(\tau) = \tau P_1 \left|\frac{dX_{1i}}{d\tau}\right| \int_{i\in\mathcal{B}} dF(i).$$

b) [Optimal policy implementation] The planner can implement the optimal corrective policy by adjusting the tax rate until trading volume equals fundamental volume. Formally,

$$\tau^* \textit{ is optimal} \iff \mathcal{V}(\tau^*) = \Theta_F(\tau^*).$$

c) [Approximation for small taxes under symmetry] Under Assumption [S], when the optimal tax takes values close to zero, it can be approximated using two variables from the laissez-faire economy: the semi-elasticity of trading volume to tax changes and the share of non-fundamental volume. Formally τ^* must satisfy

$$\underbrace{\left|\frac{d\log\mathcal{V}}{d\tau}\right|_{\tau=0}}_{\textit{Volume semi-elasticity}} \tau^* \approx \underbrace{\frac{\Theta_{NF}(0)}{\Theta_F(0)+\Theta_{NF}(0)}}_{\textit{Non-fundamental volume share}} \tag{14}$$

Proposition 2a) provides a novel decomposition of trading volume into three components. The first component of trading volume is a function of investors' initial asset holdings, risk aversion, and hedging needs. I refer to this component as fundamental volume. The second component of trading volume is a function of investors' beliefs. I refer to this component as non-fundamental volume. The third component of trading volume is a function of the tax level. I refer to this component as the tax-induced volume reduction. Note that when $\tau = 0$, this last component is zero, and all volume can be attributed to fundamental and non-fundamental components. The ability to decompose trading volume allows us to develop alternative implementations.

Proposition 2b) shows that, if the planner can credibly predict the amount of fundamental trading volume, it can adjust the optimal tax until observed volume is commensurate with the appropriate amount of fundamental trading. This new approach is appealing because it shifts the informational requirements for the planner from recovering investors' beliefs to construct a model that predicts the appropriate amount of fundamental volume. Alternatively, one can also reinterpret the optimal policy as setting a tax rate such that the tax-induced volume reduction equals non-fundamental volume, that is, setting τ^* so that $\Theta_\tau(\tau^*) = \Theta_{NF}(\tau^*)$.

This alternative implementation, which is not a direct consequence of classic Pigovian logic, relies on the ability to relate total trading volume to belief differences (the marginal distortion) and to the impact of the tax on trading (the effect of the policy instrument). Importantly, in the same way that poor measures of investors' beliefs are associated with an optimal tax that is too high or too low, overestimating (underestimating) the amount of non-fundamental trading is associated with an optimal tax that is too high (low) relative to the correct one.

Finally, Proposition 2c) provides a new alternative implementation that exploits the definition of trading volume. The upshot of this new approximation is that it provides a simple "back-of-the-envelope" solution to the model based exclusively on information from the laissez-faire economy: the semi-elasticity of trading volume to a tax change when $\tau \approx 0$ and the share of non-fundamental trading volume (or fundamental, given that they must add up to one) without intervention. Intuitively, it states that the reduction in trading volume caused by a tax change of size τ must correspond to the share of non-fundamental trading. For instance, an economy in which a 100bps tax increase reduces trading volume by 40%, and whose share of non-fundamental volume is 20% will feature an (approximate) optimal tax of 0.5%. The next section further illustrates the quantitative implications of these results.

5 Gaussian trading motives and quantitative implications

The results in Propositions 1 and 2 apply to any distribution of trading motives. In this section, I explicitly parametrize the cross-sectional distribution of beliefs and hedging needs. This allows us to better understand how changes in the composition of trading motives affect the optimal tax. I assume that investors' beliefs and hedging needs are jointly normally distributed, as formally described in Assumption [G].[20]

Assumption. [G] **(Gaussian trading motives)**

a) Investors' beliefs and hedging needs are jointly distributed across the population of investors according to

$$\mathbb{E}_i [D] \sim \mu_d + \varepsilon_{di}$$
$$A\mathbb{C}ov [E_{2i}, D] \sim \mu_h + \varepsilon_{hi},$$

where $\mu_d \geq 0$ *and* $\mu_h = 0$*. The random variables* ε_{hi} *and* ε_{di} *are jointly normally distributed as*

[20] Assuming that beliefs and hedging needs are normally distributed simplifies computations and allows us to find additional analytical results in Proposition 3. It is a common choice in related environments, like Athanasoulis and Shiller (2001) and Simsek (2013).

follows, where $\rho \in [-1, 1]$ and $\sigma_d^2, \sigma_h^2 \geq 0$.[21]

$$\begin{pmatrix} \varepsilon_{di} \\ \varepsilon_{hi} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_d^2 & \rho\sigma_d\sigma_h \\ \rho\sigma_d\sigma_h & \sigma_h^2 \end{pmatrix} \right) \tag{15}$$

b) Investors have identical preferences $A_i = A$ and hold identical initial asset positions $X_{0i} = X_0$.

By varying the variances of beliefs σ_d^2 and hedging needs σ_h^2 we can parametrize the relative importance of fundamental versus non-fundamental trading in the cross-section of investors. I refer to $\frac{\sigma_d}{\sigma_h}$ as the ratio of non-fundamental to fundamental trading. At times, it is more convenient to focus on the share of non-fundamental trading in total volume, defined by $\chi \equiv \frac{\sigma_d^2}{\sigma_d^2 + \sigma_h^2} = \frac{1}{1 + \left(\frac{\sigma_d}{\sigma_h}\right)^{-2}}$. When $\chi = 0$, investors have identical beliefs and all trade is fundamental. When $\chi = 1$, all trade is driven by investors' beliefs. The parameter ρ determines the correlation between both motives to trade across the population. A positive (negative) value of ρ implies that optimistic investors are more likely to be sellers (buyers) for fundamental reasons.

Making investors' preferences identical and assuming that they have identical asset holdings of the risky asset eliminates other reasons for trading. These assumptions can be relaxed without impact on the insights. Because the normal distribution is symmetric, Assumption [G] implies that Assumption [S] is satisfied, which guarantees that the planner's problem has a unique optimum.

Theoretical results First, I characterize several theoretical results that arise from restricting the distribution of trading motives. Subsequently, I study the quantitative predictions of the model.

Under Assumption [G], the equilibrium price in this economy is constant for any value of τ and can be expressed as a function of primitives. It corresponds to

$$P_1 = \mu_d - A\mathbb{V}ar[D]\,Q. \tag{16}$$

As described in the Online Appendix, there exist explicit expressions for individual equilibrium allocations, trading volume, and the fraction of buyers, sellers, and inactive investors. There also exist explicit expressions for fundamental and non-fundamental trading volume, as well as for the tax-induced volume reduction. The optimal tax satisfies a non-linear equation involving the inverse-Mills ratio of the normal distribution. The following results emerge.

Proposition 3. (Optimal tax and comparative statics with Gaussian trading motives)

a) Under Assumption [G], as long as some investors have heterogeneous beliefs ($\sigma_d > 0$) and investors' beliefs and hedging needs are not positively correlated ($\rho \leq 0$), it is optimal to set a strictly positive tax.

[21] Note that $\mathbb{V}ar[D]$ refers to the variance of the asset payoff while σ_d^2 corresponds to the cross-sectional dispersion of beliefs about expected payoffs.

b) When positive, the optimal tax is increasing in the ratio of non-fundamental trading to fundamental trading $\frac{\sigma_d}{\sigma_h}$ *for any correlation level* ρ*. Consequently, a mean-preserving spread in investors' beliefs is associated with a higher optimal tax.*

c) When $\frac{\sigma_d}{\sigma_h} < 1$*, the optimal policy corresponds to a subsidy (*$\tau^* < 0$*) when* $\rho > \frac{\sigma_d}{\sigma_h}$*. When* $\frac{\sigma_d}{\sigma_h} > 1$*, the optimal tax is maximal (*$\tau^* = \infty$*) when* $\rho \geq \frac{\sigma_h}{\sigma_d}$*.*

The result for the case in which $\rho = 0$ is a special case of the general result in Proposition 1, which guarantees the optimality of a positive tax when fundamental and non-fundamental motives for trade are orthogonal to each other and there exists some non-fundamental trading. With Gaussian trading motives, assuming that fundamental and non-fundamental trading motives are negatively correlated further increases the rationale for taxation, since it implies that optimistic (pessimistic) investors are also more likely to be buyers (sellers) for fundamental reasons. Figure 2, described below, clearly illustrates that in many instances in which $\rho > 0$, the optimal tax can still be positive and finite.

The optimal tax increases with the share of non-fundamental trading in the relevant region in which the tax is positive. Consequently, a mean-preserving spread of investors' beliefs is associated with a higher optimal tax. Intuitively, an increase in belief dispersion makes optimistic (pessimistic) investors more likely to be buyers (sellers), increasing non-fundamental trading volume and the motive to tax by the planner.

Proposition 3c) concludes that subsidies or arbitrarily large taxes are optimal when ρ takes large positive values. When $\sigma_d < \sigma_h$, and ρ is close 1, the planner knows that fundamental reasons drive the direction of trades, at the same time that he is aware that sellers are most likely to have optimistic beliefs and vice versa. In that case, it is clear that a trading subsidy will be welfare improving by inducing sellers to sell more and buyers to buy more. On the contrary, when $\sigma_d > \sigma_h$, the planner is aware that beliefs are the main driver of trading and that fundamental reasons reinforce the direction of investors' fundamental trades. In that case, it may be optimal to fully shut-down trade (or reduce it as much as possible).[22]

Parametrization While the model is stylized, it is worthwhile to provide a sense of the magnitudes that it generates for different parameters. The two key inputs that determine the sign and magnitude of the optimal tax are $\frac{\sigma_d}{\sigma_h}$ and ρ. Given the difficulty of finding direct empirical evidence regarding both parameters, I explore the sensitivity of the model predictions for different combinations of $\frac{\sigma_d}{\sigma_h}$ and ρ.

[22]For technical reasons, I've assumed throughout that there is a maximum feasible bound $\overline{\tau}$ for the tax. One should interpret the region in which $\tau^* = \infty$ as corresponding with this bound. Similarly, as described in Section 2, one should associate an optimal policy involving a subsidy to no intervention.

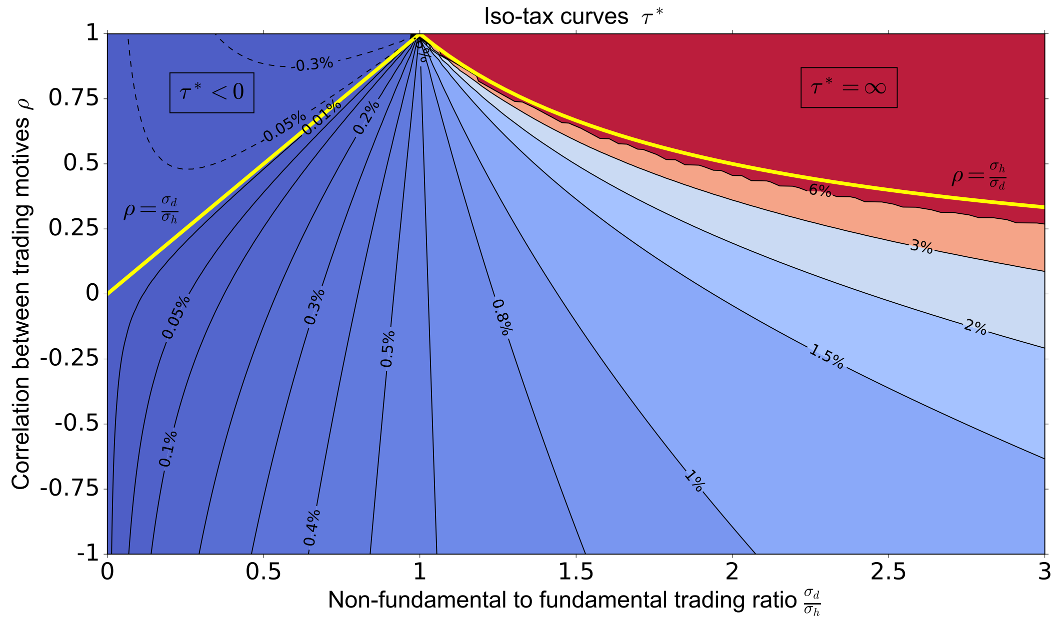

Figure 2: Iso-tax curves for combinations of $\frac{\sigma_d}{\sigma_h}$ and ρ.

When needed, I adopt $\rho = 0$ and $\frac{\sigma_d}{\sigma_h} = 0.5$ as reference values. Without direct evidence, assuming that the different types of trading motives are uncorrelated is a plausible justification for $\rho = 0$. The choice of $\frac{\sigma_d}{\sigma_h} = 0.5$, which implies that 20% of laissez-faire trade is non-fundamental, is debatable. Some may argue that up to 90% of trading volume is non-fundamental, while others strongly defend a value that is close to zero — see the discussion in Hong and Stein (2007). For instance, using a structural approach, Koijen and Yogo (2015) are only able to explain 40% of asset holdings, leaving 60% of investors' portfolio holdings unexplained, which sets an upper bound for $\frac{\sigma_d}{\sigma_h}$. My reference choice is thus a conservative one — implying that a non-negligible fraction of trades is non-fundamental, while erring on the side of attributing most trades to fundamental reasons.

For given values of $\frac{\sigma_d}{\sigma_h}$ and ρ, I formally show in Lemma 3 in the Online Appendix that the magnitude of the optimal tax is fully determined by two high-level scale invariant variables: i) *turnover* in the laissez-faire economy and ii) the *risk premium*. Calibrating the model through scale-invariant variables sidesteps common concerns associated with the lack of scale-invariance of CARA calibrations (see, e.g., Campbell (2017)) and allows us to conjecture that the quantitative insights should remain valid more generally.

First, regarding laissez-faire turnover, I assume that quarterly turnover corresponds to 25% of total float. This value is consistent with the long-run historical average of turnover among NYSE stocks, as reported by Hong and Stein (2007) — see also the NYSE Factbook. Next, I calibrate the quarterly risk premium to a standard value of $6\%/4$. As explained in detail in the Online Appendix, I adopt a quarterly calibration to be able to generate plausible values for the

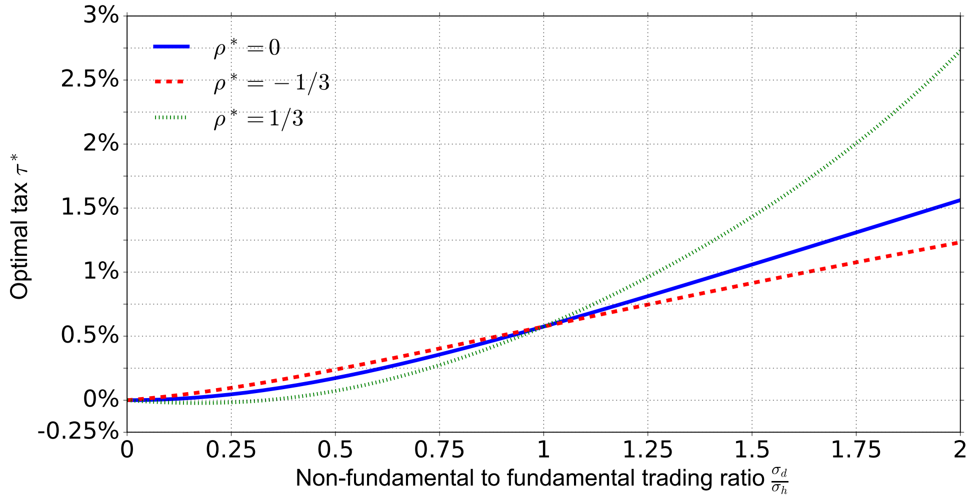

Figure 3: Optimal tax

semi-elasticity of trading volume to tax changes.[23]

Quantitative results Under the parametrization I just described, I rely on Figures 2 to 4 to illustrate the theoretical results as well as to explore the magnitudes implied by the model.

Figure 2 illustrates the quantitative results by showing a contour plot of iso-tax curves for different combinations ρ and $\frac{\sigma_d}{\sigma_h}$. The two solid lines delimiting the upper left and right corners differentiate the special regions in which the optimal policy takes the form of a subsidy or in which it is optimal to set the highest positive tax. The region between both lines delimits the area of interest in which the optimal tax is positive but finite. The steepness of iso-tax curves measures the sensitivity of the optimal tax to changes in the correlation level, illustrated in greater detail in Figure 3.

Figure 3 illustrates how the optimal tax changes as a function of the ratio of non-fundamental to fundamental trading for different correlations. For low values of $\frac{\sigma_d}{\sigma_h}$, the optimal tax increases slowly with the level of non-fundamental trading. For values of $\frac{\sigma_d}{\sigma_h}$ less than unity — when non-fundamental volume is less than 50% — the value of the optimal tax is not particularly sensitive to the value of the correlation between trading motives. Higher values of $\frac{\sigma_d}{\sigma_h}$ imply higher optimal tax rates, as well as amplified effects due to changes in the value of ρ. More specifically, when

[23] Colliard and Hoffmann (2013) find estimates for the volume semi-elasticity to tax changes using the recent French experience that correspond to $\left|\frac{d\log\mathcal{V}}{d\tau}\right| = 100$: they find that a 20bps tax increase (0.02%) reduces trading volume — persistently, not at impact — by 20%. Coelho (2014) finds comparable estimates. I reference average values for the semi-elasticity, although these papers find a range of semi-elasticities for different investors and market structures. Using data from the Swedish experience in the 80's, Umlauf (1993) finds that a 1% tax increase is associated with a decline in turnover of more than 60%, which corresponds to a semi-elasticity $\left|\frac{d\log\mathcal{V}}{d\tau}\right| = 60$. The benchmark parametrization, which implies a semi-elasticity of $\left|\frac{d\log\mathcal{V}}{d\tau}\right|_{\tau=0} = 133$, yields a conservative estimate of the optimal tax, by slightly over-estimating the behavioral market response to tax changes.

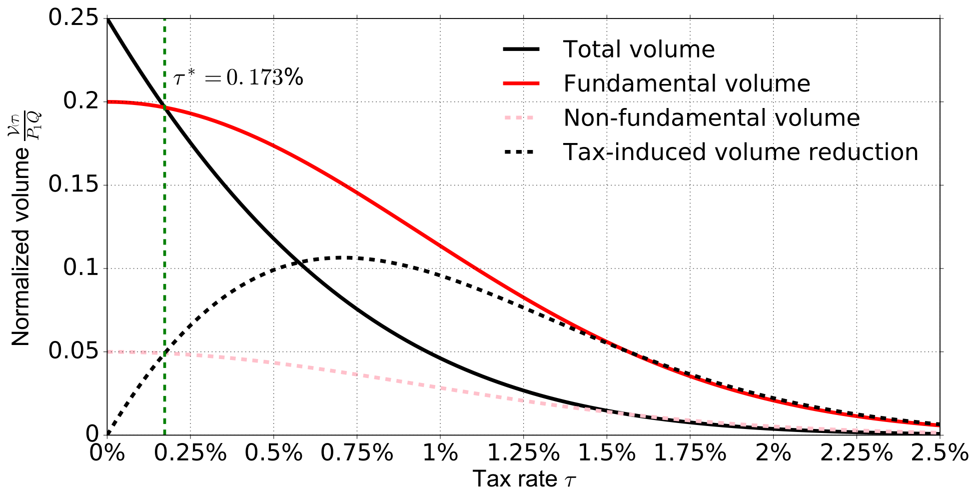

Figure 4: Trading volume implementation when $\frac{\sigma_d}{\sigma_h} = 0.5$ and $\rho = 0$.

$\frac{\sigma_d}{\sigma_h} = 0.5$, implying a share of non-fundamental trading of 20%, the optimal tax corresponds to $\tau^* = 0.173\%$. When the share of non-fundamental trading corresponds to 50% ($\frac{\sigma_d}{\sigma_h} = 1$) or 10% ($\frac{\sigma_d}{\sigma_h} = 0.33$), the optimal tax respectively increases to 0.57% or decreases to a value of 0.08%.[24]

While Figures 2 and 3 provide quantitative results and illustrate the results of Propositions 1 and 3, Figure 4 graphically illustrates how to make use of the results of Proposition 2. It shows the differential behavior of the three components of trading volume to changes in the tax rate. The reduction in fundamental and non-fundamental volume, which is monotonic, is driven by extensive margin changes in the composition of active investors. Meanwhile, the tax-induced component of volume grows rapidly at first before it starts decreasing monotonically, due to the overall trading reduction on the extensive and intensive margins. Total and fundamental volume intersect at the optimal tax level of $\tau^* = 0.173\%$. As expected, non-fundamental volume and the tax-induced volume component intersect as well at the same tax level.

We can also verify the validity of Proposition 2c in this particular calibration. The benchmark parametrization implies a volume semi-elasticity of $\left|\frac{d \log \mathcal{V}}{d\tau}\right|_{\tau=0} = 133$, which, combined with a ratio of fundamental to non-fundamental trading corresponding to 20% of total laissez-faire volume, yields an (approximate) optimal tax rate of $\tau^* \approx \frac{0.2}{133} = 0.15\%$, close to the exact value found. Assuming a volume semi-elasticity of 100, the simplest rule-of-thumb derived from this paper associates the percentage of non-fundamental trades to the optimal tax, when expressed in basis points. That is, a 20%, 40%, or 60% share of non-fundamental volume is approximately associated with an optimal tax of 20, 40, or 60bps.

Two final remarks are worth emphasizing. First, allowing investors to trade dynamically

[24]Interestingly, the optimal tax is convex in the level of $\frac{\sigma_d}{\sigma_h}$, which suggests that uncertainty about the level of non-fundamental trading calls for higher optimal taxes.

at different frequencies or introducing technological trading costs will require recalibrating the model. However, as long as the new calibration is consistent with the observed semi-elasticity of output to tax changes, one would expect to find comparable values for the optimal tax for a given share of non-fundamental volume. Consequently, finding improved measures of volume elasticities to tax changes as well as estimates of non-fundamental volume shares are necessary measurement efforts to refine optimal tax prescriptions. Second, the quantitative results in this Section assume that the volume reduction associated with a tax increase represents a behavioral response and not tax avoidance. This is consistent with most of the derivations in the paper and is a reasonable assumption for small taxes. However, if tax enforcement is imperfect, one must interpret the optimal taxes reported in this section as upper bounds, and refine the analysis along the lines of Section C.2 in the Online Appendix.

Distributional concerns Finally, I briefly explain how distributional concerns on the part of the planner affect the determination of the optimal tax. The Online Appendix contains a more detailed discussion of distributional issues.[25]

So far, in order to facilitate the aggregation of preferences, the planner has maximized either a sum of certainty equivalents or a sum of indirect utilities jointly with ex-ante transfers. Now, I assume that the planner maximizes the sum of investors' indirect utilities without ex-ante transfers, which introduces an endogenous desire for redistribution towards poorer investors. For clarity, I continue to assume that investors receive a rebate equal to their tax liabilities, which is small when the total tax liability is small.

In the scenario in which buyers and sellers are equally well-off on average ex-post,[26] the optimal transaction tax satisfies

$$\tau^* = \frac{\mathbb{E}_{\mathcal{B}(\tau^*)}\left[\frac{\mathbb{E}_i[D]}{P_1}\right] - \mathbb{E}_{\mathcal{S}(\tau^*)}\left[\frac{\mathbb{E}_i[D]}{P_1}\right]}{2} + \frac{\mathbb{C}ov_{\mathcal{B}(\tau^*)}\left[\frac{h_i(\tau^*)}{\mathbb{E}_{\mathcal{B}}[h_i(\tau^*)]}, \frac{\mathbb{E}_i[D]}{P_1}\right] - \mathbb{C}ov_{\mathcal{S}(\tau^*)}\left[\frac{h_i(\tau^*)}{\mathbb{E}_{\mathcal{S}(\tau^*)}[h_i(\tau^*)]}, \frac{\mathbb{E}_i[D]}{P_1}\right]}{2}, \quad (17)$$

where $h_i(\tau^*) \equiv \mathbb{E}\left[U'\left(W_{2i}(\tau^*)\right)\right]$ measures investors' average marginal utility from the planner's standpoint. The first term in Equation (17) is identical to one derived without distributional concerns. The second term, which is driven by the endogenous correlation between expected marginal utilities h_i and investors' beliefs $\frac{\mathbb{E}_i[D]}{P_1}$, is expected to be positive.

We can draw some intuition from the generalized optimal tax formula. Intuitively, investors with extreme beliefs tend to have lower average wealth and expected utility from the planner's viewpoint, inducing a positive correlation between beliefs and marginal social weights for buyers

[25] See Lockwood and Taubinsky (2017) for a study of the redistributive impact of corrective sin taxes in the context of general commodity taxation.

[26] Even under Assumption [G], which guarantees that the model is symmetric from a positive standpoint, sellers are on average worse off, since they face a higher exposure to outside risks. These effects are small in the baseline calibration and zero when the risky asset is in zero net supply.

and a negative correlation for sellers. This correlation implies a positive second term in Equation (17), so a planner with distributional concerns implements a higher optimal tax. However, in the baseline calibration for a planner with a belief equal to the average belief, these distributional concerns are quantitatively small, and the difference between optimal taxes with and without distributional concerns is minimal.

In the Online Appendix, I derive a more general version of Equation (17), and provide a second stage analysis for different planner's beliefs, which now do matter to determine the optimal tax. The quantitative results suggest that the optimal tax prescription from the benchmark model remains accurate if distributional concerns are present when the planner's belief is close to the average belief of investors. I also illustrate how investors with extreme beliefs experience the largest welfare gains from the policy and discuss the implementation of ex-ante transfers if desired.

6 Robustness of the results

Before concluding, I show that the optimal tax from the CARA-Normal setup remains valid as a first-order approximation to the optimal tax under more general assumptions on preferences and beliefs. This result shows that the analysis of the paper is of first-order importance more generally. Based on results described in the Online Appendix, I also discuss several extensions to the baseline model.

6.1 General utility and arbitrary beliefs

This section extends the main results to an environment in which investors have general utility specifications and disagree about probability distributions in an arbitrary way. Investors' beliefs are now modeled as a change of measure with respect to the planner's probability measure, which (jointly) determines the realization of all random variables — asset payoffs and endowments — in the model. The beliefs of investor i about date 2 uncertainty are determined by a Radon-Nikodym derivative Z_i, which is absolutely continuous with respect to the planner's probability measure. This random variable Z_i captures any discrepancy between probability assessments made by the planner and those made by investors.

Investors thus maximize

$$\max_{X_{1i}} \mathbb{E}_i \left[U_i \left(W_{2i} \right) \right],$$

where $U_i(\cdot)$ satisfies standard regularity conditions, subject to a wealth accumulation constraint

$$W_{2i} = E_{2i} + X_{1i} D + \left(X_{0i} P_1 - X_{1i} P_1 - \tau P_1 \left| \Delta X_{1i} \right| + T_{1i} \right),$$

which normalizes investors' initial endowment of consumption good to zero. I continue to assume that investors receive a rebate equal to their tax liabilities — if the total tax liability is small, these income effects are negligible.

In this model, when investors trade, their optimal portfolio decision satisfies a modified Euler equation

$$\mathbb{E}_i \left[U_i' \left(W_{2i}\right) \left(D - P_1 \left(1 + \tau \operatorname{sgn}\left(\Delta X_{1i}\right)\right)\right)\right] = 0. \tag{18}$$

Again, some investors may decide not to trade at all when their optimal asset holdings are close to their initial asset position.

As in the baseline model, the planner uses his own belief to calculate welfare. In this case, he maximizes the sum of investors' indirect utilities. Proposition 4 characterizes the exact optimal tax in the general case, as well as its approximation when risks are small. Importantly, the approximated optimal tax characterization in this general model turns out to be equal to the exact optimal tax characterization in the baseline model, lending support to the results derived in previous sections of this paper.

Proposition 4. (General utility and arbitrary beliefs)

a) The optimal financial transaction tax τ^* satisfies

$$\tau^* = \frac{\int \mathbb{E}\left[U_i'\left(W_{2i}\right)\left(Z_i - 1\right)\left(\frac{D}{P_1} - 1\right)\right] \frac{dX_{1i}}{d\tau} dF\left(i\right) + \frac{d \log P_1}{d\tau} \int \mathbb{E}\left[U_i'\left(W_{2i}\right)\right] \Delta X_{1i} dF\left(i\right)}{\int \mathbb{E}_i\left[U_i'\left(W_{2i}\right)\right] \operatorname{sgn}\left(\Delta X_{1i}\right) \frac{dX_{1i}}{d\tau} dF\left(i\right)}, \tag{19}$$

where $W_{2i}\left(\tau^*\right)$, $P_1\left(\tau^*\right)$, $\frac{dX_{1i}}{d\tau}\left(\tau^*\right)$, $\frac{dP_1}{d\tau}\left(\tau^*\right)$, and $\Delta X_{1i}\left(\tau^*\right)$ are implicit functions of τ^*.

b) When risks are small, implying that marginal utilities are approximately constant, the optimal financial transaction tax τ^* (approximately) satisfies

$$\tau^* \approx \frac{\Omega_{\mathcal{B}(\tau^*)} - \Omega_{\mathcal{S}(\tau^*)}}{2},$$

where $\Omega_{\mathcal{B}(\tau^*)}$ and $\Omega_{\mathcal{S}(\tau^*)}$ are described in Equations (8) and (9). This expression is identical to the one in Proposition 1.

The optimal tax in the general case satisfies a highly implicit condition. Investors' probability assessments, Z_i, asset demand sensitivities $\frac{dX_{1i}}{d\tau}$, and investors' marginal valuations, which depend on investors' wealth and marginal utility through $U_i'\left(W_{2i}\right)$, are the key determinants of the optimal tax. Since the denominator of Equation (19) turns out to be negative under mild regularity conditions, the two terms in the numerator of Equation (19) pin down the sign of τ^*. The first term corresponds to the difference in beliefs between the planner and the investors. Intuitively, if Z_i is high when $\frac{D}{P_1}$ is high — which denotes optimism — for a buyer, for which we expect a negative $\frac{dX_{1i}}{d\tau}$, we expect a positive tax. This effect is amplified by the marginal valuation in a given state $U_i'(W_{2i})$. The second term corresponds to the distributive pecuniary effects of the policy, which are zero-sum in dollars and cancel out under complete markets.[27]

Proposition 4b) provides clearer insights of the determinants of Equation (19), showing that the optimal tax satisfies, up to a first-order approximation, the same condition as in the CARA-Normal case. Hence, we can interpret the results of the paper as valid more generally up to a first-order approximation. This result is related to the classic Arrow-Pratt approximation — see Arrow (1971) and Pratt (1964) — which shows that the solution to the CARA-Normal portfolio problem is an approximation to any portfolio problem for small gambles, but it is not identical, since Proposition 4 directly approximates the optimal tax formula, while the standard approximation is done over investor's optimality conditions. Interestingly, only the first moment of the distribution of beliefs appears explicitly in the approximated optimal tax formula: this result motivates the sustained assumption in the paper restricting belief differences to the first-moment of the distribution of payoffs.

6.2 Extensions

In the Online Appendix, I study several extensions. Analogous expressions for the optimal tax remain valid in the more general environments.

Within the static model, I first show that the optimal tax formula from Proposition 1 remains valid when there are pre-existing trading costs, as long as these are compensation for the use of economic resources, not economic rents. Perhaps counter-intuitively, when pre-existing trading costs reduce the share of fundamental trading, the optimal transaction tax can be increasing in the level of trading costs and vice versa. Second, I show that the sign of the optimal tax is independent of whether tax enforcement is perfect or imperfect. However, I show that the magnitude of the optimal tax is decreasing in the investor's ability to avoid paying taxes. Third, in an environment with multiple risky assets, the optimal tax becomes a weighted average of the optimal tax for each asset, with higher weights given to those assets whose volume is more sensitive to tax changes. This result follows from the second-best Pigovian nature of the policy. Fourth, I show that investor-specific taxes are needed to implement the first-best outcome. Finally, I provide a formula for the upper bound of welfare losses induced by a marginal tax change when all trades are deemed fundamental.

In a q-theory production economy, I show that a transaction tax generates additional first-order gains/losses as long as the planner's belief differs from the average belief of investors. In

[27] See Dávila and Korinek (2017) for a study of these effects in a broad class of incomplete market models. An earlier version of this draft included a more detailed analysis of how transaction taxes affect welfare in the presence of distribute pecuniary externalities.

addition to the allocation of risk among investors, the level of aggregate risk and investment in the economy now also affects welfare. If a marginal tax increase reduces (increases) investment at the margin when investors are too optimistic (pessimistic) relative to the planner, a positive tax is welfare improving, and vice versa. In principle, the optimal tax formula in a production economy depends on the belief used by the planner. However, if the planner uses the average belief of investors in the economy to calculate welfare, there is no additional rationale for taxation due to production. Access to an additional policy instrument that targets aggregate investment would be optimal in this environment, allowing the planner to set the optimal transaction tax as in the baseline model, for any planner's belief.

Finally, I study how allowing for dynamic trading affects the magnitude of the optimal tax. A transaction tax is more effective with forward-looking investors who buy and sell at high frequencies, since the anticipation of future taxes reduces the incentives to trade. Buy-and-hold investors are barely sensitive to a transaction tax so, if they predominate, a larger optimal tax is needed, holding disagreement constant. This result is consistent with Tobin's insight that high-frequency trading is more affected by a transaction tax, although this paper shifts the emphasis towards identifying first the source of the trading distortion and then adjusting the magnitude of the optimal corrective tax depending on investors' asset demand sensitivities.

7 Conclusion

This paper studies the welfare implications of taxing financial transactions in an equilibrium model in which financial market trading is driven by both fundamental and non-fundamental motives. While a transaction tax is a blunt instrument that distorts both fundamental and non-fundamental trading, the welfare implications of reducing each kind of trading are different. As long as a fraction of investors hold heterogeneous beliefs that are orthogonal to their fundamental motives to trade, a utilitarian planner who calculates social welfare using a single belief will find a strictly positive tax optimal. Interestingly, the optimal tax may be independent of the belief used by the planner to calculate welfare.

The optimal transaction tax can be expressed as a function of investors' beliefs and asset demand sensitivities. Alternatively, the planner can determine the optimal tax level by directly equating the level of fundamental volume to total trading volume. When quantifying the model's implications, finding measures of the sensitivity of trading volume to tax levels, as well as finding estimates of the fraction of non-fundamental trading in the laissez-faire economy are the key measurement efforts necessary to determine the magnitude of the optimal tax.

Although the Online Appendix includes multiple extensions, there are additional extensions that could refine optimal tax prescriptions that are worth exploring. Understanding the normative implications of taxing financial transactions in models with endogenous learning dynamics or rich wealth dynamics, when markets are decentralized, or when some investors have market power are fruitful avenues for further research.

Appendix

Section 3: Proofs and derivations

Properties of investors' problem Given a price P_1 and a tax τ, investors solve $\max_{X_{1i}} J(X_{1i})$, where $J(X_{1i})$ denotes the objective function of investors, introduced in Equation (4) in the text.[28] The first and second order conditions in the regions in which the problem is differentiable respectively are

$$\begin{aligned} J'(X_{1i}) &= [\mathbb{E}_i[D] - A_i \mathbb{C}ov[E_{2i}, D] - P_1] - \tau |P_1| \operatorname{sgn}(\Delta X_{1i}) - A_i \mathbb{V}ar[D] X_{1i} = 0 \\ J''(X_{1i}) &= -A_i \mathbb{V}ar[D] < 0. \end{aligned}$$

Note that $\lim_{X_{1i} \to X_{0i}^-} J'(X_{1i}) > \lim_{X_{1i} \to X_{0i}^+} J'(X_{1i})$, which implies that the transaction tax generates a concave kink at $X_{1i} = X_{0i}$. The existence of a concave kink combined with the fact that $J''(\cdot) < 0$ jointly imply that the solution to the investors' problem is unique, and that it can be reached either at an interior optimum or at the kink. Equation (5) provides a full characterization of the solution. When taxes are positive, for a given price P_1, an individual investor i decides not to trade when

$$\left| \frac{\mathbb{E}_i[D] - A_i \mathbb{C}ov[E_{2i}, D] - A_i \mathbb{V}ar[D] X_{0i}}{P_1} - 1 \right| \leq \tau.$$

Lemma 1. (Competitive equilibrium with taxes)

a) [Existence/Uniqueness] For given primitives and a tax level τ, let us define an excess demand function $Z(P_1) \equiv \int_{i \in \mathcal{T}(P_1)} \Delta X_{1i}(P_1) dF(i)$, where net demands $\Delta X_{1i}(P_1)$ are determined by Equation (5) and $\mathcal{T}(P_1)$ denotes the set of traders with non-zero net trading demands for a given price P_1. A price P_1' is part of an equilibrium if $Z(P_1') = 0$, which guarantees that market clearing is satisfied. The continuity of $Z(P_1)$ follows trivially. It is equally straightforward to show that $\lim_{P_1 \to \infty} Z(P_1) = -\infty$ and $\lim_{P_1 \to -\infty} Z(P_1) = \infty$. These three properties are sufficient to establish that an equilibrium always exist, applying the Intermediate Value Theorem.

To establish uniqueness, we must study the properties of $Z'(P_1)$, given by

$$Z'(P_1) = \int_{i \in \mathcal{T}(P_1)} \frac{\partial X_{1i}(P_1)}{\partial P_1} dF(i) = -\int_{i \in \mathcal{T}(P_1)} \frac{1 + \operatorname{sgn}(\Delta X_{1i}) \tau}{A_i \mathbb{V}ar[D]} \leq 0,$$

[28] In particular, $J(X_{1i}) = [\mathbb{E}_i[D] - A_i \mathbb{C}ov[E_{2i}, D] - P_1] X_{1i} - \tau |P_1| |\Delta X_{1i}| - \frac{A_i}{2} \mathbb{V}ar[D] X_{1i}^2$. When needed for any formal statement, I allow the price P_1 to be negative. A sufficient (but not necessary) condition for P_1 to be strictly positive is that the expected dividend of every investor is large enough when compared to his risk bearing capacity, that is: $\mathbb{E}_i[D] > A_i (\mathbb{C}ov[E_{2i}, D] + \mathbb{V}ar[D] Q)$, $\forall i$. Note also that one must assume that $\mathbb{V}ar[E_{2i}]$ is sufficiently large to guarantee that the variance-covariance matrix of the joint distribution of E_{2i} and D is positive semi-definite.

where the first equality follows from Leibniz's rule. Because the distribution of investors is continuous, $Z(P_1)$ is differentiable.[29] Note that $Z'(P_1)$ is strictly negative when the region $\mathcal{T}(P_1)$ is non-empty. This is sufficient to conclude that if there exists a price P_1' that satisfies $Z(P_1') = 0$ and that implies that the set of investors who actively trade has positive measure, the equilibrium must be unique, because $Z'(P_1') < 0$ at that point and $Z'(P_1) \leq 0$ everywhere else. However, a price P_1' that satisfies $Z(P_1') = 0$ but that implies that the set of investors who actively trade has zero measure can also exist. In that case, there will generically be a range of prices that are consistent with no-trade.

Therefore, trading volume is always pinned down, although there is an indeterminacy in the set of possible asset prices when there is no trade in equilibrium. In that sense, the equilibrium is essentially unique.

b) [Volume response] The change in trading volume is given by $\frac{d\mathcal{V}}{d\tau} = \int_{i\in\mathcal{B}(P_1)} \frac{dX_{1i}}{d\tau} dF(i)$. It follows that $\frac{dX_{1i}}{d\tau} = \frac{\partial X_{1i}}{\partial\tau} + \frac{\partial X_{1i}}{\partial P_1}\frac{dP_1}{d\tau}$ can be expressed as

$$\frac{dX_{1i}}{d\tau} = \frac{\partial X_{1i}}{\partial\tau} \underbrace{\left[1 - (\operatorname{sgn}(\Delta X_{1i}) + \tau) \frac{\int_{i\in\mathcal{T}(P_1)} \frac{\operatorname{sgn}(\Delta X_{1i})}{A_i} dF(i)}{\int_{i\in\mathcal{T}(P_1)} \frac{1+\operatorname{sgn}(\Delta X_{1i})\tau}{A_i} dF(i)}\right]}_{\equiv\varepsilon_i}, \tag{20}$$

where $\frac{\partial X_{1i}}{\partial\tau} = \frac{-P_1 \operatorname{sgn}(\Delta X_{1i})}{A_i \mathbb{V}ar[D]}$, $\frac{\partial X_{1i}}{\partial P} = \frac{-(1+\operatorname{sgn}(\Delta X_{1i})\tau)}{A_i \mathbb{V}ar[D]}$, and it is straightforward to show that $\varepsilon_i > 0$ for both buyers and sellers. Equation (20) implies that $\frac{dX_{1i}}{d\tau} < 0$ for buyers, while $\frac{dX_{1i}}{d\tau} > 0$ for sellers, implying that trading volume decreases with τ.

c) [Price response] The price P_1 is continuous and differentiable in τ when the distribution of investors is continuous. Using again Leibniz's rule, the derivative $\frac{dP_1}{d\tau}$ can be expressed as

$$\frac{dP_1}{d\tau} = \frac{\int_{i\in\mathcal{T}(P_1)} \frac{\partial X_{1i}}{\partial\tau} dF(i)}{-\int_{i\in\mathcal{T}(P_1)} \frac{\partial X_{1i}}{\partial P} dF(i)} = \frac{-\left(\int_{i\in\mathcal{B}(P_1)} \frac{P_1}{A_i \mathbb{V}ar[D]} dF(i) - \int_{i\in\mathcal{S}(P_1)} \frac{P_1}{A_i \mathbb{V}ar[D]} dF(i)\right)}{\int_{i\in\mathcal{T}(P_1)} \frac{1+\operatorname{sgn}(\Delta X_{1i})\tau}{A_i \mathbb{V}ar[D]} dF(i)} \tag{21}$$

It follows that $\frac{dP_1}{d\tau} < 0$ if $\int_{i\in\mathcal{B}(P_1)} \frac{1}{A_i} dF(i) > \int_{i\in\mathcal{S}(P_1)} \frac{1}{A_i} dF(i)$ and vice versa. Under Assumption [S], which implies that $\frac{1}{A_i}$ is constant and the share of buyers equals the share of sellers, the numerator of Equation (21) is zero, implying that $\frac{dP_1}{d\tau} = 0$.

[29] When the distribution of investors F is continuous, $P_1(\tau)$, $X_{1i}(\tau)$, and $V(\tau)$ are continuously differentiable. All the economic insights from this paper remain valid when the distribution of investors can have mass points. Assuming a continuous probability distribution simplifies all formal characterizations by preserving differentiability.

Section 4: Proofs and derivations

Proposition 1. (Optimal financial transaction tax)

a) [Optimal tax formula] The derivative of the planner's objective function is given by $\frac{dV}{d\tau} = \int \frac{d\hat{V}_i}{d\tau} dF(i)$, where $\frac{d\hat{V}_i}{d\tau}$ corresponds to

$$\frac{d\hat{V}_i}{d\tau} = \left[\mathbb{E}[D] - \mathbb{E}_i[D] + \operatorname{sgn}(\Delta X_{1i}) P_1 \tau\right] \frac{dX_{1i}}{d\tau} - \Delta X_{1i} \frac{dP_1}{d\tau}.$$

This derivation uses the envelope theorem at the extensive margin between trading and no trading. Note that $\frac{d\hat{V}_i}{d\tau} = 0$ for investors who do not trade at the margin, because $\frac{dX_{1i}}{d\tau} = 0$ and $\Delta X_{1i} = 0$. We can further express the change in social welfare as

$$\frac{dV}{d\tau} = \int \left[-\mathbb{E}_i[D] + \operatorname{sgn}(\Delta X_{1i}) P_1 \tau\right] \frac{dX_{1i}}{d\tau} dF(i), \tag{22}$$

where Equation (22) follows from market clearing, which implies $\int \Delta X_{1i} dF(i) = 0$ and $\int \frac{dX_{1i}}{d\tau} dF(i) = 0$.[30]

We can express τ^* as

$$\begin{aligned}
\tau^* &= \frac{\int \frac{\mathbb{E}_i[D]}{P_1} \frac{dX_{1i}}{d\tau} dF(i)}{\int \operatorname{sgn}(\Delta X_{1i}) \frac{dX_{1i}}{d\tau} dF(i)} = \frac{1}{2} \frac{\int \frac{\mathbb{E}_i[D]}{P_1} \frac{dX_{1i}}{d\tau} dF(i)}{\int_{i \in \mathcal{B}(\tau)} \frac{dX_{1i}}{d\tau} dF(i)} \\
&= \frac{1}{2} \left[\overbrace{\int_{i \in \mathcal{B}(\tau)} \frac{\mathbb{E}_i[D]}{P_1} \underbrace{\frac{\frac{dX_{1i}}{d\tau}}{\int_{i \in \mathcal{B}(\tau)} \frac{dX_{1i}}{d\tau} dF(i)}}_{\omega_i^{\mathcal{B}}} dF(i)}^{\Omega_{\mathcal{B}}} - \overbrace{\int_{i \in \mathcal{S}(\tau)} \frac{\mathbb{E}_i[D]}{P_1} \underbrace{\frac{\frac{dX_{1i}}{d\tau}}{\int_{i \in \mathcal{S}} \frac{dX_{1i}}{d\tau} dF(i)}}_{\omega_i^{\mathcal{S}}} dF(i)}^{\Omega_{\mathcal{S}}} \right].
\end{aligned}$$

This derivation exploits the fact that $\int \frac{\mathbb{E}_i[D]}{P_1} \frac{dX_{1i}}{d\tau} dF(i) = \int_{i \in \mathcal{B}(\tau)} \frac{\mathbb{E}_i[D]}{P_1} \frac{dX_{1i}}{d\tau} dF(i) + \int_{i \in \mathcal{S}(\tau)} \frac{\mathbb{E}_i[D]}{P_1} \frac{dX_{1i}}{d\tau} dF(i)$, as well as the fact that $\int_{i \in \mathcal{B}(\tau)} \frac{dX_{1i}}{d\tau} dF(i) = -\int_{i \in \mathcal{S}(\tau)} \frac{dX_{1i}}{d\tau} dF(i)$.

b) [Sign of the optimal tax] Given the properties of the planner's problem, established below, it is sufficient to show that $\left.\frac{dV}{d\tau}\right|_{\tau=0} > 0$ to guarantee that the optimal policy is a positive tax. We can express $\left.\frac{dV}{d\tau}\right|_{\tau=0}$ as follows

$$\begin{aligned}
\left.\frac{dV}{d\tau}\right|_{\tau=0} &= -\int \mathbb{E}_i[D] \left.\frac{dX_{1i}}{d\tau}\right|_{\tau=0} dF(i) = -\mathbb{C}ov_F\left(\mathbb{E}_i[D], \left.\frac{dX_{1i}}{d\tau}\right|_{\tau=0}\right) \\
&= \frac{P_1}{\mathbb{V}ar[D]} \left[\mathbb{C}ov_F\left(\mathbb{E}_i[D], \frac{\mathbb{I}\left[\Delta X_{1i}|_{\tau=0} > 0\right]}{A_i}\right) \varepsilon_B - \mathbb{C}ov_F\left(\mathbb{E}_i[D], \frac{\mathbb{I}\left[\Delta X_{1i}|_{\tau=0} < 0\right]}{A_i}\right) \varepsilon_S\right].
\end{aligned}$$

[30] Under the assumption that the planner maximizes a weighted sum of indirect utilities with access to ex-ante lump-sum transfers, $V = \int \lambda_i V_i dF(i)$, where investor i indirect utility corresponds to $V_i(\tau) = \mathbb{E}[U_i(W_{2i}(\tau))] = e^{-A_i \hat{V}_i(\tau)}$. Access to lump-sum transfers endogenously guarantees that $\lambda_i \mathbb{E}[U_i'(W_{2i})]$ is constant across investors, making Equation (22) equally applicable.

Hence, $\frac{dV}{d\tau}\big|_{\tau=0}$ is positive if $\mathbb{C}ov_F\left(\mathbb{E}_i[D], \frac{\mathbb{I}\left[\Delta X_{1i}|_{\tau=0}>0\right]}{A_i}\right) > 0$, since that result directly implies that $\mathbb{C}ov_F\left(\mathbb{E}_i[D], \frac{\mathbb{I}\left[\Delta X_{1i}|_{\tau=0}<0\right]}{A_i}\right) < 0$. Under the assumption that all cross-sectional distributions are independent, we can decompose equilibrium net trading volume as

$$\Delta X_{1i} = \underbrace{\frac{\mathbb{E}_i[D] - \mathbb{E}_F[\mathbb{E}_i[D]] + A\mathbb{E}_F[\mathbb{C}ov[E_{2i}, D]] + A\mathbb{V}ar[D]Q}{A_i\mathbb{V}ar[D]}}_{\equiv Z_1} \underbrace{-\frac{\mathbb{C}ov[E_{2i}, D]}{\mathbb{V}ar[D]} - X_{0i}}_{\equiv Z_2}, \tag{23}$$

where Z_1 and Z_2 are defined in Equation (23) and $A \equiv \left(\mathbb{E}_F\left[\frac{1}{A_i}\right]\right)^{-1}$. For a low cross-sectional dispersion of risk tolerances/risk aversion coefficients, that is, when $\mathbb{V}ar\left[\frac{1}{A_i}\right] \approx 0$,[31] the sign of the covariance of interest is identical to sign of the following expression

$$\mathbb{C}ov_F(Z_1, g(Z_1 + Z_2)),$$

where Z_1 and Z_2, given their definition above, are independent random variables, and $g(\cdot)$ is an increasing function. It then follows directly from the FKG inequality (Fortuin, Kasteleyn and Ginibre, 1971) that $\mathbb{C}ov_F(Z_1, g(Z_1 + Z_2))$ is positive, which allows us to conclude that $\frac{dV}{d\tau}\big|_{\tau=0} > 0$ when fundamental and non-fundamental motives to trade are orthogonally distributed across the population.

c) [Irrelevance of the planner's belief] The claim follows directly from Equation (22). The fact the risky asset is in fixed supply, which implies that $\int \frac{dX_{1i}}{d\tau} dF(i) = 0$, combined with the linearity of investors certainty equivalents are necessary for the irrelevance result to hold.

Properties of planner's problem The planner's objective function $V(\tau)$ is continuous and differentiable at all interior points of its domain $[\underline{\tau}, \overline{\tau}]$ when the distribution of investors is also continuous. Hence, the extreme value theorem guarantees that there exists a maximum. The first order condition of the planner's problem is given by Equation (22).[32]

Establishing the uniqueness of the optimum and its properties requires the study of $\frac{d^2V}{d\tau^2}$. I show that the planner's objective function is concave on the intensive margin, although changes in the composition of marginal investors on the extensive margin cause non-concavities. Formally, the second order condition of the planner's problem is given by

$$\begin{aligned}\frac{d^2V}{d\tau^2} &= \int \operatorname{sgn}(\Delta X_{1i}) \frac{d(P_1\tau)}{d\tau}\frac{dX_{1i}}{d\tau} dF(i) + \int [-\mathbb{E}_i[D] + \operatorname{sgn}(\Delta X_{1i}) P_1\tau] \frac{d^2 X_{1i}}{d\tau^2} dF(i) + e.m. \\ &= \frac{d(P_1\tau)}{d\tau}\int \operatorname{sgn}(\Delta X_{1i}) \frac{dX_{1i}}{d\tau} dF(i) + 2\frac{dP_1}{d\tau}\frac{1}{P_1}\int [-\mathbb{E}_i[D] + \operatorname{sgn}(\Delta X_{1i}) P_1\tau] \frac{dX_{1i}}{d\tau} dF(i) + e.m.,\end{aligned} \tag{24}$$

[31] This condition is not required when the risky asset is in zero net supply.

[32] I say that an objective function is concave when its second derivative is negative. I say that a well-behaved problem with a concave objective function optimized over a convex set is convex.

where $e.m.$ denotes terms that involve marginal effects of changes in the composition of marginal investors, which can take any sign and whose value is determined by the underlying cross sectional distribution of investors. Under symmetry, it's trivially the case that $e.m. = 0$.[33] It follows from Equation (24) that, at the optimum,

$$\left.\frac{d^2V}{d\tau^2}\right|_{\tau=\tau^*,\, e.m.=0} = \left.\frac{d(P_1\tau)}{d\tau}\int \operatorname{sgn}(\Delta X_{1i})\frac{dX_{1i}}{d\tau}dF(i)\right|_{\tau=\tau^*} \leq 0,$$

because $\frac{d(P_1\tau)}{d\tau} = P_1\left(1 - \frac{\tau\int\frac{\operatorname{sgn}(\Delta X_{1i})}{A_i}dF(i)}{\int\frac{1+\operatorname{sgn}(\Delta X_{1i})\tau}{A_i}dF(i)}\right) > 0$. Because $\frac{dV}{d\tau}$ is differentiable, this result implies that, when there are are no extensive margin effects (or when they are small), any interior optimum must be a maximum. If extensive margin effects are large, there could potentially be multiple interior optima — see the working paper version of this paper for an example. Because there are no extensive margin changes when $\tau < 0$, $e.m. = 0$ in that case and the planner's problem is convex in that region, implying that $\left.\frac{dV}{d\tau}\right|_{\tau=0} > 0$ is a sufficient condition for $\tau^* > 0$. The optimum can be interior or reached at the maximum (minimum) feasible tax $\overline{\tau}$ ($\underline{\tau}$).

Proposition 2. (Volume implementation)

a) [Trading volume decomposition] Trading volume (in dollars) is defined by[34]

$$\mathcal{V}(\tau) \equiv P_1\int_{i\in\mathcal{B}(\tau)}\Delta X_{1i}dF(i) = \frac{1}{2}\left(\int_{i\in\mathcal{B}}P_1\Delta X_{1i}dF(i) - \int_{i\in\mathcal{S}}P_1\Delta X_{1i}dF(i)\right).$$

We can express the individual net trade (in dollars) as

$$P_1\Delta X_{1i} = \frac{P_1}{A_i\mathbb{V}ar[D]}\left(\mathbb{E}_i[D] - A_i\mathbb{C}ov[E_{2i},D] - P_1(1+\operatorname{sgn}(\Delta X_{1i})\tau) - A_i\mathbb{V}ar[D]X_{0i}\right),$$

which allows us to write trading volume as

$$\begin{aligned}
\mathcal{V}(\tau) &= -\frac{1}{2}\left[\int_{i\in\mathcal{T}}\left(\frac{\partial X_{1i}}{\partial\tau}\left(\mathbb{E}_i[D] - A_i\mathbb{C}ov[E_{2i},D] - P_1(1+\operatorname{sgn}(\Delta X_{1i})\tau) - A_i\mathbb{V}ar[D]X_{0i}\right)\right)dF(i)\right] \\
&= -\frac{1}{2}\left[\begin{array}{c}\int_{i\in\mathcal{T}}\left(\frac{dX_{1i}}{d\tau}\left(\mathbb{E}_i[D] - A_i\mathbb{C}ov[E_{2i},D] - P_1\operatorname{sgn}(\Delta X_{1i})\tau - A_i\mathbb{V}ar[D]X_{0i}\right)\right)dF(i) \\ +\frac{dP_1}{d\tau}\int_{i\in\mathcal{T}}\left(-\frac{\partial X_{1i}}{\partial P_1}\right)A_i\mathbb{V}ar[D]\Delta X_{1i}dF(i)\end{array}\right] \\
&= -\frac{1}{2}\left[\int_{i\in\mathcal{T}}\left(\frac{dX_{1i}}{d\tau}\left(\mathbb{E}_i[D] - A_i\mathbb{C}ov[E_{2i},D] - P_1\operatorname{sgn}(\Delta X_{1i})\tau - A_i\mathbb{V}ar[D]X_{0i}\right)\right)dF(i)\right] - \frac{dP_1}{d\tau}\frac{1}{P_1}\tau\mathcal{V}(\tau),
\end{aligned}$$

using the fact that

$$-\int_{i\in\mathcal{T}}\frac{\partial X_{1i}}{\partial P_1}A_i\mathbb{V}ar[D]\Delta X_{1i}dF(i) = \int_{i\in\mathcal{T}}(1+\operatorname{sgn}(\Delta X_{1i})\tau)\Delta X_{1i}dF(i) = \frac{2\tau}{P_1}\mathcal{V}(\tau).$$

[33] The following results are useful: $\frac{d^2P_1}{d\tau^2} = 2\left(\frac{dP_1}{d\tau}\right)^2\frac{1}{P_1} + e.m.$ and $\frac{d^2X_{1i}}{d\tau^2} = 2\frac{dX_{1i}}{d\tau}\frac{dP_1}{d\tau}\frac{1}{P_1} + e.m.$ Also $\frac{d\left(\frac{\partial X_{1i}}{\partial\tau}\right)}{d\tau} = \frac{-P_1\operatorname{sgn}(\Delta X_{1i})}{A_i\mathbb{V}ar[D]}\frac{dP_1}{d\tau}\frac{1}{P_1} = \frac{\partial X_{1i}}{\partial\tau}\frac{dP_1}{d\tau}\frac{1}{P_1}$ and $\frac{d\left(\frac{\partial X_{1i}}{\partial P}\right)}{d\tau} = \frac{-\operatorname{sgn}(\Delta X_{1i})}{A_i\mathbb{V}ar[D]} = \frac{\partial X_{1i}}{\partial\tau}\frac{1}{P_1}$.

[34] All limits of integration are a function of τ, as in $\mathcal{B}(\tau)$, $\mathcal{S}(\tau)$, and $\mathcal{T}(\tau)$. The same applies to $P_1(\tau)$. To simplify the notation, I suppress the explicit dependence in most derivations.

Therefore, we define $\kappa\left(P_1,\tau\right)\equiv\frac{1}{1+\frac{dP_1}{d\tau}\frac{\tau}{P_1}}$, and express trading volume as

$$\begin{aligned}\mathcal{V}\left(\tau\right)&=\frac{\kappa\left(P_1,\tau\right)}{2}\int_{i\in\mathcal{T}}\left(\left(-\frac{dX_{1i}}{d\tau}\right)\left(\mathbb{E}_i\left[D\right]-A_i\mathbb{C}ov\left[E_{2i},D\right]-P_1\operatorname{sgn}\left(\Delta X_{1i}\right)\tau-A_i\mathbb{V}ar\left[D\right]X_{0i}\right)\right)dF\left(i\right)\\&=\Theta_F\left(\tau\right)+\Theta_{NF}\left(\tau\right)-\Theta_\tau\left(\tau\right),\end{aligned}$$

where

$$\begin{aligned}\Theta_F\left(\tau\right)&\equiv\frac{\kappa\left(P_1,\tau\right)}{2}\int_{i\in\mathcal{T}}\left(-\frac{dX_{1i}}{d\tau}\right)\left(-A_i\mathbb{C}ov\left[E_{2i},D\right]-A_i\mathbb{V}ar\left[D\right]X_{0i}\right)dF\left(i\right)\\\Theta_{NF}\left(\tau\right)&\equiv\frac{\kappa\left(P_1,\tau\right)}{2}\int_{i\in\mathcal{T}}\left(-\frac{dX_{1i}}{d\tau}\right)\mathbb{E}_i\left[D\right]dF\left(i\right)\\\Theta_\tau\left(\tau\right)&\equiv\frac{\kappa\left(P_1,\tau\right)}{2}\tau P_1\int_{i\in\mathcal{T}}\left(-\frac{dX_{1i}}{d\tau}\right)\operatorname{sgn}\left(\Delta X_{1i}\right)dF\left(i\right).\end{aligned}$$

When Assumption [S] holds, $\frac{dX_{1i}}{d\tau}$ is constant across investors and $\kappa\left(P_1,\tau\right)=1$, justifying the expressions in the text.

b) [Optimal policy implementation] Note that the optimality condition for the planner obtained in Proposition 1 can be expressed as

$$\int_{i\in\mathcal{T}}\frac{dX_{1i}}{d\tau}\mathbb{E}_i\left[D\right]dF\left(i\right)=\tau P_1\int_{i\in\mathcal{T}}\operatorname{sgn}\left(\Delta X_{1i}\right)\frac{dX_{1i}}{d\tau}dF\left(i\right),$$

which is satisfied when $\Theta_{NF}\left(\tau^*\right)=\Theta_\tau\left(\tau^*\right)$ or, alternatively, when $\mathcal{V}\left(\tau^*\right)=\Theta_F\left(\tau^*\right)$.

c) [Approximation for small taxes under symmetry] One can always express $\Theta_\tau\left(\tau\right)$ as $\Theta_\tau\left(\tau\right)=-\tau P_1\frac{d\mathcal{V}}{d\tau}$. We know that, at the optimum $\Theta_\tau\left(\tau^*\right)=\Theta_{NF}\left(\tau^*\right)$, which allows us to write

$$\tau^*=\frac{\frac{\Theta_{NF}\left(\tau^*\right)}{\Theta_F\left(\tau^*\right)+\Theta_{NF}\left(\tau^*\right)-\Theta_\tau\left(\tau^*\right)}}{\left|\frac{d\log\mathcal{V}}{d\tau}\right|_{\tau^*}},$$

which can be approximated when $\tau^*\approx0$ as $\tau^*\approx\frac{\frac{\Theta_{NF}(0)}{\Theta_F(0)+\Theta_{NF}(0)}}{\left|\frac{d\log\mathcal{V}}{d\tau}\right|_{\tau^*=0}}$.

The remaining proofs and derivations are in the Online Appendix.

References

Arrow, Kenneth J. 1971. "The theory of risk aversion." *Essays in the theory of risk-bearing*, 90–120.

Athanasoulis, S., and R. Shiller. 2001. "World Income Components: Measuring and Exploiting Risk-Sharing Opportunities." *The American Economic Review.*

Auerbach, Alan J, and James R Hines Jr. 2002. "Taxation and economic efficiency." *Handbook of public economics*, 3: 1347–1421.

Aumann, Robert J. 1976. "Agreeing to Disagree." *The Annals of Statistics*, 4(6): 1236–1239.

Baldauf, Markus, and Joshua Mollner. 2014. "High-Frequency Trade and Market Performance." *Working Paper.*

Baldauf, Markus, and Joshua Mollner. 2015. "Fast Traders Make a Quick Buck: The Role of Speed in Liquidity Provision." *Working Paper.*

Barberis, Nicholas, Andrei Shleifer, and Robert Vishny. 1998. "A model of investor sentiment." *Journal of financial economics*, 49(3): 307–343.

Barberis, Nicholas, and Richard Thaler. 2003. "A survey of behavioral finance." *Handbook of the Economics of Finance*, 1: 1053–1128.

Black, Fischer. 1986. "Noise." *The Journal of Finance*, 41(3): 529–543.

Blume, Lawrence E, Timothy Cogley, David A Easley, Thomas J Sargent, and Viktor Tsyrennikov. 2013. "Welfare, Paternalism and Market Incompleteness." *Working Paper.*

Brunnermeier, Markus K, Alp Simsek, and Wei Xiong. 2014. "A Welfare Criterion For Models With Distorted Beliefs." *The Quarterly Journal of Economics*, 129(4): 1753–1797.

Budish, Eric, Peter Cramton, and John Shim. 2015. "Editor's Choice The High-Frequency Trading Arms Race: Frequent Batch Auctions as a Market Design Response." *The Quarterly Journal of Economics*, 130(4): 1547–1621.

Burman, Leonard E, William G Gale, Sarah Gault, Bryan Kim, Jim Nunns, and Steve Rosenthal. 2016. "Financial Transaction Taxes in Theory and Practice." *National Tax Journal*, 69(1): 171–216.

Campbell, John Y. 2016. "Richard T. Ely Lecture Restoring Rational Choice: The Challenge of Consumer Financial Regulation." *The American Economic Review*, 106(5): 1–30.

Campbell, John Y. 2017. *Financial Decisions and Markets: a Course in Asset Pricing.* Princeton University Press.

Campbell, J.Y., and K.A. Froot. 1994. "International experiences with securities transaction taxes." In *The internationalization of equity markets.* 277–308. University of Chicago Press.

Cochrane, John. 2013. "Finance: Function Matters, Not Size." *The Journal of Economic Perspectives*, 27(2): 29–50.

Cochrane, John. 2014. "Challenges for Cost-Benefit Analysis of Financial Regulation." *Journal of Legal Studies*, Forthcoming.

Coelho, Maria. 2014. "Dodging Robin Hood: Responses to France and Italy's Financial Transaction Taxes." *UC Berkeley Working Paper.*

Colliard, Jean-Edouard, and Peter Hoffmann. 2013. "Sand in the chips: Evidence on taxing transactions in an electronic market." European Central Bank Working Paper.

Constantinides, G.M. 1986. "Capital market equilibrium with transaction costs." *The Journal of Political Economy*, 842–862.

Dang, Tri Vi, and Florian Morath. 2015. "The Taxation of Bilateral Trade with Endogenous Information." *Working Paper.*

Dávila, Eduardo, and Anton Korinek. 2017. "Pecuniary Externalities in Economies with Financial Frictions." *The Review of Economic Studies*, rdx010.

Dávila, Eduardo, and Cecilia Parlatore. 2016. "Trading Costs and Informational Efficiency." *NYU Stern Working Paper.*

De Long, J Bradford, Andrei Shleifer, Lawrence H Summers, and Robert J Waldmann. 1990. "Noise Trader Risk in Financial Markets." *Journal of Political Economy*, 98(4): 703–738.

Diamond, Peter A. 1973. "Consumption externalities and imperfect corrective pricing." *The Bell Journal of Economics and Management Science*, 526–538.

Duffie, Darrell. 2014. "Challenges to a Policy Treatment of Speculative Trading Motivated by Differences in Beliefs." *The Journal of Legal Studies*, 43(S2): S173–S182.

Eyster, Erik, and Matthew Rabin. 2005. "Cursed equilibrium." *Econometrica*, 73(5): 1623–1672.

Farhi, Emmanuel, and Xavier Gabaix. 2015. "Optimal Taxation with Behavioral Agents." *Working Paper.*

Fortuin, Cees M, Pieter W Kasteleyn, and Jean Ginibre. 1971. "Correlation inequalities on some partially ordered sets." *Communications in Mathematical Physics*, 22(2): 89–103.

Gayer, Gabrielle, Itzhak Gilboa, Larry Samuelson, and David Schmeidler. 2014. "Pareto Efficiency with Different Beliefs." *The Journal of Legal Studies*, 43(S2): S151–S171.

Gerritsen, Aart. 2016. "Optimal taxation when people do not maximize well-being." *Journal of Public Economics*, 144: 122–139.

Gilboa, Itzhak, Larry Samuelson, and David Schmeidler. 2014. "No-Betting-Pareto Dominance." *Econometrica*, 82(4): 1405–1442.

Glosten, L.R., and P.R. Milgrom. 1985. "Bid, ask and transaction prices in a specialist market with heterogeneously informed traders." *Journal of financial economics*, 14(1): 71–100.

Goulder, Lawrence H. 1995. "Environmental taxation and the double dividend: a reader's guide." *International Tax and Public Finance*, 2(2): 157–183.

Greene, William H. 2003. *Econometric Analysis.* Upper Saddle River, NJ.

Grossman, Sanford J., and Joseph E. Stiglitz. 1980. "On the impossibility of informationally efficient markets." *American Economic Review*, 70(3): 393–408.

Gruber, Jonathan, and Botond Koszegi. 2001. "Is addiction "rational"? Theory and evidence." *The Quarterly Journal of Economics*, 116(4): 1261–1303.

Habermeier, K., and A.A. Kirilenko. 2003. "Securities Transaction Taxes and Financial Markets." *IMF Staff Papers*, 165–180.

Harberger, Arnold C. 1964. "The measurement of waste." *The American Economic Review*, 58–76.

Harrison, J.M., and D.M. Kreps. 1978. "Speculative investor behavior in a stock market with heterogeneous expectations." *The Quarterly Journal of Economics*, 92(2): 323–336.

Heyerdahl-Larsen, Christian, and Johan Walden. 2014. "Efficiency and Distortions in a Production Economy with Heterogeneous Beliefs." *Working Paper.*

Hong, Harrison, and Jeremy C Stein. 1999. "A unified theory of underreaction, momentum trading, and overreaction in asset markets." *The Journal of Finance*, 54(6): 2143–2184.

Hong, Harrison, and Jeremy C Stein. 2007. "Disagreement and the stock market." *The Journal of Economic Perspectives*, 21(2): 109–128.

Jones, C.M., and P.J. Seguin. 1997. "Transaction costs and price volatility: evidence from commission deregulation." *The American Economic Review*, 728–737.

Koijen, Ralph SJ, and Motohiro Yogo. 2015. "An equilibrium model of institutional demand and asset prices." NBER Working Paper.

Kopczuk, Wojciech. 2003. "A note on optimal taxation in the presence of externalities." *Economics Letters*, 80(1): 81–86.

Kreps, David M. 2012. *Microeconomic Foundations I: Choice and Competitive Markets.* Princeton University Press.

Lintner, John. 1969. "The aggregation of investors' diverse judgments and preferences in purely competitive security markets." *Journal of Financial and Quantitative Analysis*, 4(4): 347–400.

Lockwood, Benjamin B. 2016. "Optimal Income Taxation with Present Bias." *Working Paper.*

Lockwood, Benjamin B, and Dmitry Taubinsky. 2017. "Regressive sin taxes." *Working Paper.*

McCulloch, Neil, and Grazia Pacillo. 2011. "The Tobin Tax - A Review of the Evidence." *Working Paper, Institute of Development Studies.*

Morris, Stephen. 1995. "The common prior assumption in economic theory." *Economics and philosophy*, 11: 227–227.

Moser, Christian, and Pedro Olea de Souza e Silva. 2017. "Optimal Paternalistic Savings Policies." *Working Paper.*

Mullainathan, Sendhil, Joshua Schwartzstein, and William Congdon. 2012. "A Reduced-Form Approach to Behavioral Public Finance." *Annual Review of Economics*, 4: 511–540.

O'Donoghue, Ted, and Matthew Rabin. 2006. "Optimal sin taxes." *Journal of Public Economics*, 90(10): 1825–1849.

Panageas, Stavros. 2005. "The Neoclassical Theory of Investment in Speculative Markets." *Working paper.*

Posner, Eric, and Glen Weyl. 2013. "Benefit-Cost Analysis for Financial Regulation." *American Economic Review, P&P*, 103(3).

Pratt, John W. 1964. "Risk aversion in the small and in the large." *Econometrica: Journal of the Econometric Society*, 122–136.

Roll, R. 1989. "Price volatility, international market links, and their implications for regulatory policies." *Journal of Financial Services Research*, 3(2): 211–246.

Ross, S.A. 1989. "Commentary: using tax policy to curb speculative short-term trading." *Journal of Financial Services Research*, 3: 117–120.

Rothschild, Casey, and Florian Scheuer. 2016. "Optimal taxation with rent-seeking." *The Review of Economic Studies*, 83(3): 1225–1262.

Sandmo, Agnar. 1975. "Optimal taxation in the presence of externalities." *The Swedish Journal of Economics*, 86–98.

Sandmo, Agnar. 1985. "The effects of taxation on savings and risk taking." *Handbook of Public Economics*, 1: 265–311.

Sandroni, Alvaro, and Francesco Squintani. 2007. "Overconfidence, insurance, and paternalism." *The American Economic Review*, 97(5): 1994–2004.

Santos, Tano, and Jose A Scheinkman. 2001. "Competition among exchanges." *The Quarterly Journal of Economics*, 116(3): 1027–1061.

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

Scheuer, Florian. 2013. "Optimal Asset Taxes in Financial Markets with Aggregate Uncertainty." *Review of Economic Dynamics*, 16(3): 405–420.

Schwert, G. William, and Paul J. Seguin. 1993. "Securities transaction taxes: an overview of costs, benefits and unresolved questions." *Financial Analysts Journal*, 27–35.

Simsek, Alp. 2013. "Speculation and Risk Sharing with New Financial Assets." *The Quarterly Journal of Economics.*

Spinnewijn, Johannes. 2015. "Unemployed but optimistic: Optimal insurance design with biased beliefs." *Journal of the European Economic Association*, 13(1): 130–167.

Stiglitz, J.E. 1989. "Using tax policy to curb speculative short-term trading." *Journal of Financial Services Research*, 3(2): 101–115.

Summers, L.H., and V.P. Summers. 1989. "When financial markets work too well: a cautious case for a securities transactions tax." *Journal of financial services research*, 3(2): 261–286.

Tobin, J. 1978. "A proposal for international monetary reform." *Eastern Economic Journal*, 4(3/4): 153–159.

ul Haq, M., I. Kaul, and I. Grunberg. 1996. *The Tobin tax: coping with financial volatility.* Oxford University Press, USA.

Umlauf, S.R. 1993. "Transaction taxes and the behavior of the Swedish stock market." *Journal of Financial Economics*, 33(2): 227–240.

Vayanos, D., and J. Wang. 2012. "Liquidity and asset returns under asymmetric information and imperfect competition." *Review of financial studies*, 25(5): 1339–1365.

Vives, Xavier. 2017. "Endogenous Public Information and Welfare in Market Games." *The Review of Economic Studies*, 84(2): 935–963.

Weyl, E. Glen. 2007. "Is Arbitrage Socially Beneficial?" *Princeton Working Paper.*

Weyl, E. Glen. 2016. "Price theory." *Journal of Economic Literature, Forthcoming.*

Xiong, W. 2012. "Bubbles, Crises, and Heterogeneous Beliefs." *Handbook for Systemic Risk.*

Online Appendix (not for publication)

A Section 5: Proofs and derivations

The variance-covariance matrix, given by Equation (15) in the text, is positive semi-definite when $\sigma_d^2 \sigma_h^2 > (\sigma_{dh})^2$, where σ_{dh} defines the covariance between both random variables, given by $\sigma_{dh} = \rho \sigma_d \sigma_h$. This restriction is implied by the bounds on the correlation coefficient $\rho \in [-1, 1]$.

Because the cross-sectional distribution of beliefs is symmetric, the equilibrium price is constant for any value of τ and is given by

$$\begin{aligned} P_1 &= \int \left(\mathbb{E}_i [D] - A \left(\mathbb{C}ov [E_{2i}, D] + \mathbb{V}ar [D] X_{0i} \right) \right) dF(i) \\ &= \mu_d - \mu_h - A \mathbb{V}ar [D] Q, \end{aligned}$$

which corresponds to Equation (16) in the text, since $\mu_h = 0$. Note that P_1 is fully characterized as a function of primitives. We can express equilibrium net trades (when non-zero) as

$$\begin{aligned} \Delta X_{1i}^{+,-} (P_1) &= \frac{\mathbb{E}_i [D] - A \mathbb{C}ov [E_{2i}, D] - P_1 (1 + \operatorname{sgn}(\Delta X_{1i}) \tau) - A \mathbb{V}ar [D] X_{0i}}{A \mathbb{V}ar [D]} \\ &= \frac{\left(\mathbb{E}_i [D] - \int \mathbb{E}_i [D] \, dF(i) \right) - \left(A \mathbb{C}ov [E_{2i}, D] - \int A \mathbb{C}ov [E_{2i}, D] \right) - \operatorname{sgn}(\Delta X_{1i}) \tau P_1}{A \mathbb{V}ar [D]} \\ &= \frac{\varepsilon_{di} - \varepsilon_{hi} - \operatorname{sgn}(\Delta X_{1i}) \tau P_1}{A \mathbb{V}ar [D]}. \end{aligned}$$

In particular, following the formulation in Equation (5) in the text, we can write the distribution of equilibrium latent net trades in the population as

$$\begin{aligned} \Delta X_{1i}^{+} (P_1) &= \frac{\varepsilon_{di} - \varepsilon_{hi} - \tau P_1}{A \mathbb{V}ar [D]} \sim N \left(\frac{-\tau P_1}{A \mathbb{V}ar [D]}, \frac{\sigma_d^2 + \sigma_h^2 - 2 \rho \sigma_d \sigma_h}{(A \mathbb{V}ar [D])^2} \right) \\ \Delta X_{1i}^{-} (P_1) &= \frac{\varepsilon_{di} - \varepsilon_{hi} + \tau P_1}{A \mathbb{V}ar [D]} \sim N \left(\frac{\tau P_1}{A \mathbb{V}ar [D]}, \frac{\sigma_d^2 + \sigma_h^2 - 2 \rho \sigma_d \sigma_h}{(A \mathbb{V}ar [D])^2} \right), \end{aligned}$$

where we use the fact that $\mathbb{C}ov \left[\frac{\varepsilon_{di}}{A \mathbb{V}ar[D]}, \frac{-\varepsilon_{hi}}{A \mathbb{V}ar[D]} \right] = \frac{-\sigma_{dh}}{(A \mathbb{V}ar[D])^2}$. I refer to ΔX_{1i}^{+} and ΔX_{1i}^{-} as latent net buying/selling positions. Figures (A.1) and (A.2) below explicitly show a realization of the joint distribution of ε_{di} and ε_{hi}, as well as net trades.

Note that $\mathbb{V}ar [\varepsilon_{di} - \varepsilon_{hi}] = \sigma_d^2 + \sigma_h^2 - 2 \rho \sigma_d \sigma_h$ can take different values depending on the correlation between both motives for trading. Specifically

$$\mathbb{V}ar [\varepsilon_{di} - \varepsilon_{hi}] = \begin{cases} (\sigma_d + \sigma_h)^2, & \text{if } \rho = -1 \\ \sigma_d^2 + \sigma_h^2, & \text{if } \rho = 0 \\ (\sigma_d - \sigma_h)^2, & \text{if } \rho = 1. \end{cases}$$

Note that when $\rho \approx -1$, there is maximum trade. Note also that, when $\rho \approx 1$, if $\sigma_d = \sigma_h$, there is no trade in equilibrium.

Volume/Share of buyers and sellers Because there is a continuum of investors, the equilibrium level of trading volume is deterministic in this model. It corresponds to $\mathcal{V}$ in Equation (13):

$$\begin{aligned}\mathcal{V} &= P_1 \int_{i\in\mathcal{B}} \Delta X_{1i} dF(i) = P_1 \cdot \mathbb{P}\left[\Delta X_{1i}^{+} > 0\right] \cdot \mathbb{E}\left[\Delta X_{1i}^{+} > 0\right] \\ &= \frac{P_1}{A\mathbb{V}ar[D]} (1-\Phi(\alpha)) \left(-\tau P_1 + \sqrt{\sigma_d^2 + \sigma_h^2 - 2\rho\sigma_d\sigma_h}\lambda(\alpha)\right),\end{aligned}$$

where α is defined in Equation (25). The fraction of buyers (and sellers, by symmetry) corresponds to $\mathbb{P}\left[\Delta X_{1i}^{+} > 0\right]$. Following the results in Greene (2003) for truncated normal distributions, stated below for completeness, we can express both elements of $\mathcal{V}$ as follows:

$$\begin{aligned}\mathbb{P}\left[\Delta X_{1i}^{+} > 0\right] &= 1-\Phi(\alpha) \\ \mathbb{E}\left[\Delta X_{1i}^{+} > 0\right] &= \frac{1}{A\mathbb{V}ar[D]} \left(-\tau P_1 + \sqrt{\sigma_d^2 + \sigma_h^2 - 2\rho\sigma_d\sigma_h}\lambda(\alpha)\right),\end{aligned}$$

where $\lambda(\alpha) = \frac{\phi(\alpha)}{1-\Phi(\alpha)}$ and the argument α corresponds to[35]

$$\alpha = \frac{-\mu}{\sigma} = \frac{\tau P_1}{\sqrt{\sigma_d^2 + \sigma_h^2 - 2\rho\sigma_d\sigma_h}}. \tag{25}$$

Hence, we can express trading volume as a function of primitives as follows

$$\begin{aligned}\mathcal{V} &= \frac{P_1}{A\mathbb{V}ar[D]} (1-\Phi(\alpha)) \left(-\tau P_1 + \sqrt{\sigma_d^2 + \sigma_h^2 - 2\rho\sigma_d\sigma_h}\lambda(\alpha)\right) \\ &= \frac{P_1}{A\mathbb{V}ar[D]} \left(-\tau P_1 (1-\Phi(\alpha)) + \sqrt{\sigma_d^2 + \sigma_h^2 - 2\rho\sigma_d\sigma_h}\phi(\alpha)\right).\end{aligned}$$

The response of volume to a tax change corresponds to[36]

$$\begin{aligned}\frac{d\mathcal{V}}{d\tau} &= \frac{P_1}{A\mathbb{V}ar[D]} \left(-P_1(1-\Phi(\alpha)) + \tau P_1 \phi(\alpha)\frac{d\alpha}{d\tau} + \sqrt{\sigma_d^2 + \sigma_h^2 - 2\rho\sigma_d\sigma_h}\phi'(\alpha)\frac{d\alpha}{d\tau}\right) \\ &= \frac{P_1}{A\mathbb{V}ar[D]} \left(-P_1(1-\Phi(\alpha)) + \tau P_1 \phi(\alpha)\frac{d\alpha}{d\tau} - \alpha\sqrt{\sigma_d^2 + \sigma_h^2 - 2\rho\sigma_d\sigma_h}\phi(\alpha)\frac{d\alpha}{d\tau}\right) \\ &= \frac{P_1}{A\mathbb{V}ar[D]} \left(-P_1(1-\Phi(\alpha)) + \left(\tau P_1 - \alpha\sqrt{\sigma_d^2 + \sigma_h^2 - 2\rho\sigma_d\sigma_h}\right)\phi(\alpha)\frac{d\alpha}{d\tau}\right) \\ &= -P_1 \frac{P_1}{A\mathbb{V}ar[D]}(1-\Phi(\alpha)),\end{aligned}$$

which takes strictly negative values. Note that we can write

$$\begin{aligned}\frac{d\mathcal{V}}{d\tau}\frac{1}{\mathcal{V}} &= \frac{-P_1 \frac{P_1}{A\mathbb{V}ar[D]}(1-\Phi(\alpha))}{\frac{P_1}{A\mathbb{V}ar[D]}(1-\Phi(\alpha))\left(-\tau P_1 + \sqrt{\sigma_d^2 + \sigma_h^2 - 2\rho\sigma_d\sigma_h}\lambda(\alpha)\right)} = \\ &= \frac{-P_1}{-\tau P_1 + \sqrt{\sigma_d^2 + \sigma_h^2 - 2\rho\sigma_d\sigma_h}\lambda(\alpha)}.\end{aligned}$$

Therefore

$$\left|\frac{d\log\mathcal{V}}{d\tau}\right|_{\tau=0} = \frac{P_1}{\sqrt{\sigma_d^2 + \sigma_h^2 - 2\rho\sigma_d\sigma_h}\lambda(0)}.$$

[35] The function $\lambda(\cdot)$ is known as the Inverse Mills Ratio. It satisfies the following properties: $\lambda(z) \geq 0$, $\lambda'(z) > 0$, $\lim_{z\to-\infty}\lambda'(x) = 0$, and $\lim_{x\to\infty}\lambda'(x) = 1$. Also $\lambda(0) = \sqrt{\frac{2}{\pi}}$.

[36] Where we use the fact that $\phi'(\alpha) = -\alpha\phi(\alpha)$.

Optimal tax Because Assumption [S] is satisfied, the optimal tax formula corresponds to

$$\tau^* = \frac{\mathbb{E}_{\mathcal{B}(\tau^*)}\left[\frac{\mathbb{E}_i[D]}{P_1}\right] - \mathbb{E}_{\mathcal{S}(\tau^*)}\left[\frac{\mathbb{E}_i[D]}{P_1}\right]}{2}.$$

We can calculate the average belief of the buyers as

$$\begin{aligned}\mathbb{E}_{\mathcal{B}}\left[\mathbb{E}_i\left[D\right]\right] &= \mu_d + \mathbb{E}_{\mathcal{B}}\left[\varepsilon_{di}\right] = \mu_d + \mathbb{E}\left[\varepsilon_{di}\middle|\,\Delta X_{1i}^{+} > 0\right] = \mu_d + \mathbb{E}\left[\varepsilon_{di}\middle|\,\varepsilon_{di} - \varepsilon_{hi} - \tau P_1 > 0\right] \\ &= \mu_d + \mathbb{E}\left[Y\middle|\,Z^{+} > 0\right].\end{aligned}$$

Similarly for sellers

$$\begin{aligned}\mathbb{E}_{\mathcal{S}}\left[\mathbb{E}_i\left[D\right]\right] &= \mu_d + \mathbb{E}_{\mathcal{S}}\left[\varepsilon_{di}\right] = \mu_d + \mathbb{E}\left[\varepsilon_{di}\middle|\,\Delta X_{1i}^{-} < 0\right] = \mu_d + \mathbb{E}\left[\varepsilon_{di}\middle|\,\varepsilon_{di} - \varepsilon_{hi} + \tau P_1 < 0\right] \\ &= \mu_d + \mathbb{E}\left[Y\middle|\,Z^{-} < 0\right],\end{aligned}$$

where we can define variables Y, Z^+, and Z^- as follows

$$\begin{aligned}Y &= \varepsilon_{di} \\ Z^{+} &= \varepsilon_{di} - \varepsilon_{hi} - \tau P_1 \\ Z^{-} &= \varepsilon_{di} - \varepsilon_{hi} + \tau P_1.\end{aligned}$$

These are (jointly) distributed as follows[37]

$$\begin{pmatrix} Y \\ Z^{+} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ -\tau P_1 \end{pmatrix}, \begin{pmatrix} \sigma_d^2 & \sigma_d^2 - \rho\sigma_d\sigma_h \\ \vdots & \sigma_d^2 + \sigma_h^2 - 2\rho\sigma_d\sigma_h \end{pmatrix}\right)$$
$$\begin{pmatrix} Y \\ Z^{-} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ \tau P_1 \end{pmatrix}, \begin{pmatrix} \sigma_d^2 & \sigma_d^2 - \rho\sigma_d\sigma_h \\ \vdots & \sigma_d^2 + \sigma_h^2 - 2\rho\sigma_d\sigma_h \end{pmatrix}\right).$$

The correlation coefficient between Y and Z^+ (or Z^-) is

$$\rho_{YZ^{-}} = \rho_{YZ^{+}} = \frac{\sigma_d^2 - \rho\sigma_d\sigma_h}{\sigma_d\sqrt{\sigma_d^2 + \sigma_h^2 - 2\rho\sigma_d\sigma_h}} = \frac{\sigma_d - \rho\sigma_h}{\sqrt{\sigma_d^2 + \sigma_h^2 - 2\rho\sigma_d\sigma_h}}.$$

Using again the results from Greene (2003) for truncated normal distributions,

$$\mathbb{E}\left[Y\middle|\,Z^{+} > 0\right] = \mu_y + \rho\sigma_y\lambda\left(\alpha^{+}\right) = \frac{\sigma_d\left(\sigma_d - \rho\sigma_h\right)}{\sqrt{\sigma_d^2 + \sigma_h^2 - 2\rho\sigma_d\sigma_h}}\frac{\phi\left(\alpha^{+}\right)}{1 - \Phi\left(\alpha^{+}\right)} > 0$$

$$\mathbb{E}\left[Y\middle|\,Z^{-} < 0\right] = \mu_y + \rho\sigma_y\lambda_{-}\left(\alpha^{+}\right) = \frac{\sigma_d\left(\sigma_d - \rho\sigma_h\right)}{\sqrt{\sigma_d^2 + \sigma_h^2 - 2\rho\sigma_d\sigma_h}}\frac{\phi\left(\alpha^{-}\right)}{-\Phi\left(\alpha^{-}\right)} < 0,$$

where

$$\begin{aligned}\alpha^{+} &= \frac{-\mu}{\sigma} = \frac{\tau P_1}{\sqrt{\sigma_d^2 + \sigma_h^2 - 2\rho\sigma_d\sigma_h}} \\ \alpha^{-} &= \frac{-\mu}{\sigma} = \frac{-\tau P_1}{\sqrt{\sigma_d^2 + \sigma_h^2 - 2\rho\sigma_d\sigma_h}} = -\alpha^{+}.\end{aligned}$$

[37]Note that $\mathbb{C}ov\left[\varepsilon_{di}, \varepsilon_{di} - \varepsilon_{hi}\right] = \sigma_d^2 - \rho\sigma_d\sigma_h = \sigma_d\left(\sigma_d - \rho\sigma_h\right)$.

Note that naturally $\mathbb{E}_{\mathcal{B}}[\mathbb{E}_i[D]]$ is increasing in τ while $\mathbb{E}_{\mathcal{S}}[\mathbb{E}_i[D]]$ is decreasing — the average belief of marginal buyers increases with the tax level while the average belief of sellers decreases. Combining all results, we can express the numerator of the τ^* formula as

$$\begin{aligned}
\mathbb{E}_{\mathcal{B}(\tau^*)}[\mathbb{E}_i[D]] - \mathbb{E}_{\mathcal{S}(\tau^*)}[\mathbb{E}_i[D]] &= \frac{\sigma_d(\sigma_d - \rho\sigma_h)}{\sqrt{\sigma_d^2 + \sigma_h^2 - 2\rho\sigma_d\sigma_h}}\left(\frac{\phi(\alpha^+)}{1-\Phi(\alpha^+)} + \frac{\phi(\alpha^-)}{\Phi(\alpha^-)}\right) \\
&= 2\frac{\sigma_d(\sigma_d - \rho\sigma_h)}{\sqrt{\sigma_d^2 + \sigma_h^2 - 2\rho\sigma_d\sigma_h}}\frac{\phi(\alpha^+)}{1-\Phi(\alpha^+)} \\
&= 2\frac{\sigma_d(\sigma_d - \rho\sigma_h)}{\sqrt{\sigma_d^2 + \sigma_h^2 - 2\rho\sigma_d\sigma_h}}\lambda(\alpha^+),
\end{aligned}$$

where we use the fact that

$$\frac{\phi(\alpha^+)}{1-\Phi(\alpha^+)} + \frac{\phi(\alpha^-)}{\Phi(\alpha^-)} = \frac{\phi(\alpha^+)}{1-\Phi(\alpha^+)} + \frac{\phi(-\alpha^+)}{\Phi(-\alpha^+)} = \frac{\phi(\alpha^+)}{1-\Phi(\alpha^+)} + \frac{\phi(\alpha^+)}{1-\Phi(\alpha^+)} = 2\frac{\phi(\alpha^+)}{1-\Phi(\alpha^+)}.$$

We can therefore write τ^* as

$$\begin{aligned}
\tau^* &= \frac{\mathbb{E}_{\mathcal{B}(\tau^*)}\left[\frac{\mathbb{E}_i[D]}{P_1}\right] - \mathbb{E}_{\mathcal{S}(\tau^*)}\left[\frac{\mathbb{E}_i[D]}{P_1}\right]}{2} \\
&= \frac{\sigma_d(\sigma_d - \rho\sigma_h)}{\sqrt{\sigma_d^2 + \sigma_h^2 - 2\rho\sigma_d\sigma_h}}\lambda\left(\frac{\tau^* P_1}{\sqrt{\sigma_d^2 + \sigma_h^2 - 2\rho\sigma_d\sigma_h}}\right)\frac{1}{P_1}.
\end{aligned}$$

We can rearrange this expression to find that

$$\begin{aligned}
\tau^* P_1\sqrt{\sigma_d^2 + \sigma_h^2 - 2\rho\sigma_d\sigma_h} &= \sigma_d(\sigma_d - \rho\sigma_h)\lambda\left(\frac{\tau^* P_1}{\sqrt{\sigma_d^2 + \sigma_h^2 - 2\rho\sigma_d\sigma_h}}\right) \\
\frac{\tau^* P_1}{\sqrt{\sigma_d^2 + \sigma_h^2 - 2\rho\sigma_d\sigma_h}} &= \frac{\sigma_d(\sigma_d - \rho\sigma_h)}{\sigma_d^2 + \sigma_h^2 - 2\rho\sigma_d\sigma_h}\lambda\left(\frac{\tau^* P_1}{\sqrt{\sigma_d^2 + \sigma_h^2 - 2\rho\sigma_d\sigma_h}}\right),
\end{aligned}$$

which allows us to define $\tilde{\tau}^* \equiv \frac{\tau^* P_1}{\sqrt{\sigma_d^2+\sigma_h^2-2\rho\sigma_d\sigma_h}}$, implying that one must solve the fixed point

$$\tilde{\tau}^* = \underbrace{\frac{\sigma_d(\sigma_d - \rho\sigma_h)}{\sigma_d^2 + \sigma_h^2 - 2\rho\sigma_d\sigma_h}}_{\equiv\delta}\lambda(\tilde{\tau}^*), \tag{26}$$

which can be equivalently written as a function of the ratio $\frac{\sigma_h}{\sigma_d}$ and ρ as follows[38]

$$\tilde{\tau}^* = \frac{1 - \rho\frac{\sigma_h}{\sigma_d}}{1 + \left(\frac{\sigma_h}{\sigma_d}\right)^2 - 2\rho\frac{\sigma_h}{\sigma_d}}\lambda(\tilde{\tau}^*).$$

Note that $\tilde{\tau}^*$ is exclusively a function of $\frac{\sigma_d}{\sigma_h}$ and ρ. We can recover τ^* as

$$\tau^* = \tilde{\tau}^*\frac{\sqrt{\sigma_d^2 + \sigma_h^2 - 2\rho\sigma_d\sigma_h}}{P_1}.$$

[38]For numerical purposes, it is often more stable to solve $\tilde{\tau}^*(1-\Phi(\tilde{\tau}^*)) = \frac{1-\rho\frac{\sigma_h}{\sigma_d}}{1+\left(\frac{\sigma_h}{\sigma_d}\right)^2-2\rho\frac{\sigma_h}{\sigma_d}}\phi(\tilde{\tau}^*)$.

The value of δ in Equation (26) determines whether τ^* is found at an interior optimum. For extreme values of ρ, δ takes the following values

$$\delta = \begin{cases} \frac{\sigma_d^2+\sigma_d\sigma_h}{\sigma_d^2+\sigma_h^2+2\sigma_d\sigma_h}, & \text{if } \rho = -1 \\ \frac{\sigma_d^2}{\sigma_d^2+\sigma_h^2}, & \text{if } \rho = 0 \\ \frac{\sigma_d^2-\sigma_d\sigma_h}{\sigma_d^2+\sigma_h^2-2\sigma_d\sigma_h}, & \text{if } \rho = 1. \end{cases}$$

Note that $\delta \geq 1$ if $\rho \geq \frac{\sigma_h}{\sigma_d}$, implying that $\tau^* = \infty$ (the actual tax would be set at the assumed upper bound $\overline{\tau}$). Note also that $\delta < 0$ if $\rho > \frac{\sigma_d}{\sigma_h}$, which implies that $\tau^* < 0$, formally

$$\text{if} \quad \begin{cases} \rho \leq \frac{\sigma_d}{\sigma_h} & \Rightarrow \tau^* \geq 0 \\ \rho > \frac{\sigma_d}{\sigma_h} & \Rightarrow \tau^* < 0. \end{cases}$$

Alternatively, we can see that the sign of the tax is determined by

$$\left.\frac{dV}{d\tau}\right|_{\tau=0} = \frac{\sigma_d\left(\sigma_d - \rho\sigma_h\right)}{\sqrt{\sigma_d^2 + \sigma_h^2 - 2\rho\sigma_d\sigma_h}}\lambda\left(0\right)\frac{1}{P_1},$$

whose sign is a function of $\sigma_d - \rho\sigma_h$.

Trading volume implementation

Under the new parametric assumption, it is possible to provide explicit expressions for the trading volume decomposition. First, we can express total trading volume in dollars as

$$\begin{aligned} \mathcal{V}\left(\tau\right) &= \frac{1}{2}\int_{i\in\mathcal{T}}\left(\left(-\frac{\partial X_{1i}}{\partial\tau}\right)\left(\mathbb{E}_i\left[D\right] - A_i\mathbb{C}ov\left[E_{2i}, D\right] - P_1\,\text{sgn}\left(\Delta X_{1i}\right)\tau - A_i\mathbb{V}ar\left[D\right]X_{0i}\right)\right)dF\left(i\right) \\ &= \Theta_F\left(\tau\right) + \Theta_{NF}\left(\tau\right) - \Theta_\tau\left(\tau\right), \\ &= \frac{P_1}{A\mathbb{V}ar\left[D\right]}\left(1-\Phi\left(\alpha\right)\right)\left[\frac{\sigma_d\left(\sigma_d-\rho\sigma_h\right)}{\sqrt{\sigma_d^2+\sigma_h^2-2\rho\sigma_d\sigma_h}}\lambda\left(\alpha\right) + \frac{-\sigma_h\left(\rho\sigma_d-\sigma_h\right)}{\sqrt{\sigma_d^2+\sigma_h^2-2\rho\sigma_d\sigma_h}}\lambda\left(\alpha\right) - \tau P_1\right] \\ &= \frac{P_1}{A\mathbb{V}ar\left[D\right]}\left(1-\Phi\left(\alpha\right)\right)\left(-\tau P_1 + \sqrt{\sigma_d^2+\sigma_h^2-2\rho\sigma_d\sigma_h}\lambda\left(\alpha\right)\right). \end{aligned}$$

The fundamental component of trading volume can be expressed as

$$\begin{aligned} \Theta_F\left(\tau\right) &= \frac{1}{2}\int_{i\in\mathcal{T}}\left(\frac{\text{sgn}\left(\Delta X_{1i}\right)P_1}{A\mathbb{V}ar\left[D\right]}\right)\left(-A\mathbb{C}ov\left[E_{2i}, D\right]\right)dF\left(i\right) \\ &= -\frac{1}{2}\frac{P_1}{A\mathbb{V}ar\left[D\right]}\left(\int_{i\in\mathcal{B}}A\mathbb{C}ov\left[E_{2i}, D\right]dF\left(i\right) - \int_{i\in\mathcal{S}}A\mathbb{C}ov\left[E_{2i}, D\right]dF\left(i\right)\right) \\ &= -\frac{1}{2}\frac{P_1}{A\mathbb{V}ar\left[D\right]}\int_{i\in\mathcal{B}}dF\left(i\right)\left(\mathbb{E}_\mathcal{B}\left[A\mathbb{C}ov\left[E_{2i}, D\right]\right] - \mathbb{E}_\mathcal{S}\left[A\mathbb{C}ov\left[E_{2i}, D\right]\right]\right) \\ &= -\frac{1}{2}\frac{P_1}{A\mathbb{V}ar\left[D\right]}\left(1-\Phi\left(\alpha\right)\right)\left(\mathbb{E}_\mathcal{B}\left[A\mathbb{C}ov\left[E_{2i}, D\right]\right] - \mathbb{E}_\mathcal{S}\left[A\mathbb{C}ov\left[E_{2i}, D\right]\right]\right) \\ &= -\frac{P_1}{A\mathbb{V}ar\left[D\right]}\left(1-\Phi\left(\alpha\right)\right)\frac{\sigma_h\left(\rho\sigma_d-\sigma_h\right)}{\sqrt{\sigma_d^2+\sigma_h^2-2\rho\sigma_d\sigma_h}}\frac{\phi\left(\alpha^+\right)}{1-\Phi\left(\alpha^+\right)} \\ &= \frac{P_1}{A\mathbb{V}ar\left[D\right]}\frac{\sigma_h\left(\sigma_h-\rho\sigma_d\right)}{\sqrt{\sigma_d^2+\sigma_h^2-2\rho\sigma_d\sigma_h}}\phi\left(\alpha\right). \end{aligned}$$

The non-fundamental component of trading volume can be expressed as

$$
\begin{aligned}
\Theta_{NF}(\tau) &= \frac{1}{2}\int_{i\in\mathcal{T}}\left(\frac{\text{sgn}(\Delta X_{1i})P_1}{A\mathbb{V}ar[D]}\right)\mathbb{E}_i[D]\,dF(i) \\
&= \frac{1}{2}\frac{P_1}{A\mathbb{V}ar[D]}\int_{i\in\mathcal{B}} dF(i)\left(\mathbb{E}_{\mathcal{B}}[\mathbb{E}_i[D]]-\mathbb{E}_{\mathcal{S}}[\mathbb{E}_i[D]]\right) \\
&= \frac{1}{2}\frac{P_1}{A\mathbb{V}ar[D]}(1-\Phi(\alpha))\left(\mathbb{E}_{\mathcal{B}}[\mathbb{E}_i[D]]-\mathbb{E}_{\mathcal{S}}[\mathbb{E}_i[D]]\right) \\
&= \frac{P_1}{A\mathbb{V}ar[D]}\frac{\sigma_d(\sigma_d-\rho\sigma_h)}{\sqrt{\sigma_d^2+\sigma_h^2-2\rho\sigma_d\sigma_h}}\phi(\alpha).
\end{aligned}
$$

The tax component of trading volume can be expressed as

$$
\begin{aligned}
\Theta_{\tau}(\tau) &= \frac{1}{2}\tau P_1\int_{i\in\mathcal{T}}\left(\frac{\text{sgn}(\Delta X_{1i})P_1}{A\mathbb{V}ar[D]}\right)\text{sgn}(\Delta X_{1i})\,dF(i) \\
&= \frac{1}{2}\tau P_1\frac{P_1}{A\mathbb{V}ar[D]}\int_{i\in\mathcal{T}} dF(i) \\
&= \tau P_1\frac{P_1}{A\mathbb{V}ar[D]}\int_{i\in\mathcal{B}} dF(i) \\
&= \tau P_1\frac{P_1}{A\mathbb{V}ar[D]}(1-\Phi(\alpha)).
\end{aligned}
$$

Note that the results may become less intuitive when $\rho > \frac{\sigma_d}{\sigma_h}$ or $\rho < \frac{\sigma_d}{\sigma_h}$. To derive the expression for $\Theta_{NF}(\tau)$, we use the fact that we can define variables Y, Z^+, and Z^- as follows

$$
\begin{aligned}
Y &= \varepsilon_{hi} \\
Z^+ &= \varepsilon_{di}-\varepsilon_{hi}-\tau P_1 \\
Z^- &= \varepsilon_{di}-\varepsilon_{hi}+\tau P_1.
\end{aligned}
$$

Their joint distribution is[39]

$$
\begin{aligned}
\begin{pmatrix} Y \\ Z^+ \end{pmatrix} &\sim N\left(\begin{pmatrix} 0 \\ -\tau P_1 \end{pmatrix},\begin{pmatrix} \sigma_h^2 & \sigma_h(\rho\sigma_d-\sigma_h) \\ \vdots & \sigma_d^2+\sigma_h^2-2\rho\sigma_d\sigma_h \end{pmatrix}\right) \\
\begin{pmatrix} Y \\ Z^- \end{pmatrix} &\sim N\left(\begin{pmatrix} 0 \\ \tau P_1 \end{pmatrix},\begin{pmatrix} \sigma_h^2 & \sigma_h(\rho\sigma_d-\sigma_h) \\ \vdots & \sigma_d^2+\sigma_h^2-2\rho\sigma_d\sigma_h \end{pmatrix}\right),
\end{aligned}
$$

which allows us to calculate

$$
\begin{aligned}
\mathbb{E}_{\mathcal{B}}[A\mathbb{C}ov[E_{2i},D]] = \mathbb{E}_{\mathcal{B}}[\varepsilon_{di}] &= \mathbb{E}\left[\varepsilon_{di}\middle|\,\Delta X_{1i}^+>0\right] = \mathbb{E}[\varepsilon_{di}|\,\varepsilon_{di}-\varepsilon_{hi}-\tau P_1>0] \\
&= \rho\sigma_h\lambda\left(\alpha^+\right) = \frac{\sigma_h(\rho\sigma_d-\sigma_h)}{\sqrt{\sigma_d^2+\sigma_h^2-2\rho\sigma_d\sigma_h}}\frac{\phi(\alpha^+)}{1-\Phi(\alpha^+)}.
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}_{\mathcal{S}}[A\mathbb{C}ov[E_{2i},D]] = \mathbb{E}_{\mathcal{S}}[\varepsilon_{di}] &= \mathbb{E}\left[\varepsilon_{di}\middle|\,\Delta X_{1i}^-<0\right] = \mathbb{E}[\varepsilon_{di}|\,\varepsilon_{di}-\varepsilon_{hi}+\tau P_1<0] \\
&= \rho\sigma_h\lambda_-\left(\alpha^+\right) = \frac{\sigma_h(\rho\sigma_d-\sigma_h)}{\sqrt{\sigma_d^2+\sigma_h^2-2\rho\sigma_d\sigma_h}}\frac{\phi(\alpha^-)}{-\Phi(\alpha^-)}.
\end{aligned}
$$

[39]Note that $\mathbb{C}ov[\varepsilon_{hi},\varepsilon_{di}-\varepsilon_{hi}] = -\sigma_h^2+\rho\sigma_d\sigma_h = \sigma_h(\rho\sigma_d-\sigma_h)$.

with

$$\alpha^{+} = \frac{-\mu}{\sigma} = \frac{\tau P_1}{\sqrt{\sigma_d^2 + \sigma_h^2 - 2\rho\sigma_d\sigma_h}}$$

$$\alpha^{-} = \frac{-\mu}{\sigma} = \frac{-\tau P_1}{\sqrt{\sigma_d^2 + \sigma_h^2 - 2\rho\sigma_d\sigma_h}} = -\alpha^{+}.$$

We have to calculate

$$\begin{aligned}\mathbb{E}_{\mathcal{B}}\left[A\mathbb{C}ov\left[E_{2i}, D\right]\right] - \mathbb{E}_{\mathcal{S}}\left[A\mathbb{C}ov\left[E_{2i}, D\right]\right] &= \frac{\sigma_h\left(\rho\sigma_d - \sigma_h\right)}{\sqrt{\sigma_d^2 + \sigma_h^2 - 2\rho\sigma_d\sigma_h}}\frac{\phi\left(\alpha^{+}\right)}{1 - \Phi\left(\alpha^{+}\right)} + \frac{\sigma_h\left(\rho\sigma_d - \sigma_h\right)}{\sqrt{\sigma_d^2 + \sigma_h^2 - 2\rho\sigma_d\sigma_h}}\frac{\phi\left(\alpha^{-}\right)}{\Phi\left(\alpha^{-}\right)} \\ &= 2\frac{\sigma_h\left(\rho\sigma_d - \sigma_h\right)}{\sqrt{\sigma_d^2 + \sigma_h^2 - 2\rho\sigma_d\sigma_h}}\frac{\phi\left(\alpha\right)}{1 - \Phi\left(\alpha\right)},\end{aligned}$$

where $\alpha = \frac{\tau P_1}{\sqrt{\sigma_d^2 + \sigma_h^2 - 2\rho\sigma_d\sigma_h}}$.

Finally, note that one can move from the ratio of fundamental to non-fundamental trading to a share $\chi \equiv \frac{\sigma_d^2}{\sigma_d^2 + \sigma_h^2}$ by using the following two expressions

$$\chi \equiv \frac{\sigma_d^2}{\sigma_d^2 + \sigma_h^2} = \frac{1}{1 + \frac{1}{\left(\frac{\sigma_d}{\sigma_h}\right)^2}} \quad \text{and} \quad \frac{\sigma_d}{\sigma_h} = \frac{1}{\sqrt{\frac{1}{\chi} - 1}}.$$

Results from Greene (2003)

The following results from Greene (2003) are useful. Let's respectively denote by $\phi(\cdot)$ and $\Phi(\cdot)$ the pdf and cdf of the standard normal distribution.

Fact 1. *If $X \sim N\left(\mu, \sigma^2\right)$, then*

$$\mathbb{E}\left[X \left| X > a\right.\right] = \mu + \sigma\lambda\left(\alpha\right), \quad \textit{where} \quad \lambda\left(\alpha\right) = \frac{\phi\left(\alpha\right)}{1 - \Phi\left(\alpha\right)} \textit{ and } \alpha = \frac{a - \mu}{\sigma}.$$

Fact 2. *If Y and Z have a bivariate normal distribution with correlation ρ, then*

$$\mathbb{E}\left[Y \left| Z > a\right.\right] = \mu_y + \rho_{xz}\sigma_y\lambda\left(\alpha_z\right), \quad \textit{where} \quad \lambda\left(\alpha_z\right) = \frac{\phi\left(\alpha_z\right)}{1 - \Phi\left(\alpha_z\right)} \quad \textit{and} \quad \alpha_z = \frac{a - \mu_z}{\sigma_z}$$

$$\mathbb{E}\left[Y \left| Z < a\right.\right] = \mu_y + \rho_{xz}\sigma_y\lambda_{-}\left(\alpha_z\right), \quad \textit{where} \quad \lambda_{-}\left(\alpha_z\right) = -\frac{\phi\left(\alpha_z\right)}{\Phi\left(\alpha_z\right)} \quad \textit{and} \quad \alpha_z = \frac{a - \mu_z}{\sigma_z}.$$

Proposition 3. (Optimal tax and comparative statics with Gaussian trading motives)

a) It easily follows that, when $\rho \leq 0$ and $\sigma_d > 0$, Equation (26) has a non-negative solution, since $\delta > 0$.

b) From Equation (26), for any $\tau^* \geq 0$, the sign of $\frac{d\tilde{\tau}^*}{d\left(\frac{\sigma_d}{\sigma_h}\right)}$ is determined by the sign of $-\frac{d\delta}{d\left(\frac{\sigma_h}{\sigma_d}\right)}$.We can express $\frac{d\delta}{d\left(\frac{\sigma_h}{\sigma_d}\right)}$ as follows

$$\frac{d\delta}{d\left(\frac{\sigma_h}{\sigma_d}\right)} = \frac{-\rho\left(1 + \left(\frac{\sigma_h}{\sigma_d}\right)^2 - 2\rho\frac{\sigma_h}{\sigma_d}\right) - \left(1 - \rho\frac{\sigma_h}{\sigma_d}\right)2\left(\frac{\sigma_h}{\sigma_d} - \rho\right)}{\left(1 + \left(\frac{\sigma_h}{\sigma_d}\right)^2 - 2\rho\frac{\sigma_h}{\sigma_d}\right)^2} = \frac{\rho\left(\frac{\sigma_h}{\sigma_d}\right)^2 - 2\frac{\sigma_h}{\sigma_d} + \rho}{\left(1 + \left(\frac{\sigma_h}{\sigma_d}\right)^2 - 2\rho\frac{\sigma_h}{\sigma_d}\right)^2}.$$

When $\rho \leq 0$ this expression is everywhere negative implying that $\frac{d\tilde{\tau}^*}{d\left(\frac{\sigma_d}{\sigma_h}\right)}$ is positive. Note that we can find the following cases, for $\rho \geq 0$:

$$\text{if} \begin{cases} \frac{\sigma_d}{\sigma_h} \leq \rho, & \tilde{\tau}^* < 0 \\ \rho < \frac{\sigma_d}{\sigma_h} < \frac{1}{\rho}, & \tilde{\tau}^* \geq 0 \\ \frac{\sigma_d}{\sigma_h} \geq \frac{1}{\rho}, & \tilde{\tau}^* = \infty. \end{cases}$$

It can be established that $\rho\left(\frac{\sigma_h}{\sigma_d}\right)^2 - 2\frac{\sigma_h}{\sigma_d} + \rho$ is negative in between its two roots when $\rho \in (0,1]$. Its roots are given by $\frac{1\pm\sqrt{1-\rho^2}}{\rho}$. It is sufficient to show that $\frac{1-\sqrt{1-\rho^2}}{\rho} < \rho$ and that $\frac{1}{\rho} < \frac{1+\sqrt{1-\rho^2}}{\rho}$, which are trivially satisfied for any $\rho \in (0,1]$. This fact is sufficient to show the desired comparative static for the $\rho > 0$ case.

c) It follows from Equation (26) that $\tau^* < 0$ if $1 - \rho\frac{\sigma_h}{\sigma_d} < 0$. A necessary and sufficient condition for $\tau^* = \infty$ is that $\delta \geq 1$, which is satisfied when $\rho \geq \frac{\sigma_h}{\sigma_d}$.

Parametrization

The next Lemma introduces the key result that supports the calibration of the optimal tax.

Lemma 3. (Required inputs for calibration) *Knowledge of four variables is sufficient to calibrate the optimal tax. These are*

i) the turnover of the risky asset in the laissez-faire economy, expressed as a fraction of the total number of shares, defined by $\Xi \equiv \frac{\mathcal{V}(0)}{Q}\frac{1}{P_1}$*,*

ii) the risk premium, defined by $\Pi \equiv \frac{\mu_d}{P_1} - 1$*,*

iii) the share of non-fundamental trading, through $\frac{\sigma_d}{\sigma_h}$*,*

iv) the correlation between trading motives, ρ*.*

Proof. The first target corresponds to the turnover of the risky asset. Using the fact that $\phi(0) = \frac{1}{\sqrt{2\pi}}$, we can express turnover in the laissez-faire economy as

$$\Xi = \frac{1}{A\mathbb{V}ar[D]Q}\sqrt{\sigma_d^2 + \sigma_h^2 - 2\rho\sigma_d\sigma_h}\frac{1}{\sqrt{2\pi}}$$

which implies that

$$\sqrt{\sigma_d^2 + \sigma_h^2 - 2\rho\sigma_d\sigma_h} = \Xi A\mathbb{V}ar[D]Q\sqrt{2\pi}.$$

The second target is the risky asset risk premium $\Pi \equiv \frac{\mu_d}{P_1} - 1$. Note that $P_1 = \mu_d - A\mathbb{V}ar[D]Q$, which allows us to write

$$\Pi = \frac{A\mathbb{V}ar[D]Q}{P_1},$$

which implies that knowledge of the risk premium Π is sufficient to pin down $\frac{A\mathbb{V}ar[D]Q}{P_1}$.

Combining both relations, the relation between the optimal tax τ^* and the solution to Equation (26) can be expressed as

$$\tau^* = \tilde{\tau}^*\frac{\sqrt{\sigma_d^2 + \sigma_h^2 - 2\rho\sigma_d\sigma_h}}{P_1} \Rightarrow \tau^* = \tilde{\tau}^*\frac{\Xi A\mathbb{V}ar[D]Q\sqrt{2\pi}}{P_1} \Rightarrow \tau^* = \tilde{\tau}^*\Xi\Pi\sqrt{2\pi}.$$

Given that, as shown above, $\tilde{\tau}^*$ is exclusively a function of $\frac{\sigma_d}{\sigma_h}$ and ρ, for given values of $\frac{\sigma_d}{\sigma_h}$ and ρ, the optimal τ^* is fully determined by information on the laissez-faire economy turnover Ξ as well as the risk premium Π. $\square$

When interested in using the approximation derived in Proposition 2, an important object of interest is $\left.\frac{d\log\mathcal{V}}{d\tau}\right|_{\tau=0}$, which can be expressed as

$$\left.\frac{d\log\mathcal{V}}{d\tau}\right|_{\tau=0} = \frac{-P_1}{\sqrt{\sigma_d^2 + \sigma_h^2 - 2\rho\sigma_d\sigma_h}\lambda(0)} = -\frac{1}{2}\frac{1}{\Xi\Pi}. \quad (27)$$

Equation (27) shows that once one adopts targets for turnover Ξ and the risk premium Π, the model cannot freely determine the elasticity of trading volume to tax changes. Adopting a quarterly calibration allows us to match at the same time the values of turnover and the risk premium, as wells as the volume semi-elasticity, as described in the text.

Note that we can also express (relative) trading volume using only information on $\frac{\sigma_d}{\sigma_h}$, ρ, Ξ and Π. Formally, relative total volume corresponds to

$$\begin{aligned}\frac{\mathcal{V}(\tau)}{P_1 Q} &= (1-\Phi(\alpha))\left(-\tau\frac{P_1}{A\mathbb{V}ar[D]Q}+\frac{\sqrt{\sigma_d^2+\sigma_h^2-2\rho\sigma_d\sigma_h}}{A\mathbb{V}ar[D]Q}\lambda(\alpha)\right)\\ &= (1-\Phi(\alpha))\left(-\frac{\tau}{\Pi}+\Xi\sqrt{2\pi}\lambda(\alpha)\right).\end{aligned}$$

The fundamental, non-fundamental, and tax component of trading volume can be expressed as

$$\begin{aligned}\frac{\Theta_F(\tau)}{P_1 Q} &= \Xi\sqrt{2\pi}\frac{1-\rho\frac{\sigma_d}{\sigma_h}}{1+\frac{\sigma_d}{\sigma_h}-2\rho\frac{\sigma_d}{\sigma_h}}\phi(\alpha)\\ \frac{\Theta_{NF}(\tau)}{P_1 Q} &= \Xi\sqrt{2\pi}\frac{1-\rho\frac{\sigma_h}{\sigma_d}}{1+\left(\frac{\sigma_h}{\sigma_d}\right)^2-2\rho\frac{\sigma_h}{\sigma_d}}\phi(\alpha)\\ \frac{\Theta_\tau(\tau)}{P_1 Q} &= \frac{\tau}{\Pi}(1-\Phi(\alpha)),\end{aligned}$$

where the argument α is given by

$$\alpha=\frac{\tau P_1}{\sqrt{\sigma_d^2+\sigma_h^2-2\rho\sigma_d\sigma_h}}=\tau\frac{P_1}{A\mathbb{V}ar[D]Q\Xi\sqrt{2\pi}}=\frac{\tau}{\Pi\Xi\sqrt{2\pi}}.$$

Distributional concerns: optimal tax formula

In this section, I assume that the planner maximizes the sum of investors' indirect utilities, without ex-ante transfers. Because investors have concave utility, this new approach endogenously generates a desire for redistribution. In the rest of the paper, I have assumed that the planner maximizes the sum of investors' certainty equivalents or, equivalently, that it maximizes the sum of investors' direct utilities combined with ex-ante transfers.

Formally, investor i's indirect utility from the planner's perspective now corresponds to V_i, where

$$V_i \equiv \mathbb{E}[U_i(W_{2i})] = -e^{-A_i\hat{V}_i},$$

and $\hat{V}_i$ is given by

$$\begin{aligned}\hat{V}_i &= \mathbb{E}[W_{2i}(X_{1i},P_1)]-\frac{A_i}{2}\mathbb{V}ar[W_{2i}(X_{1i},P_1)]. \qquad (28)\\ &= (\mathbb{E}[D]-A_i\mathbb{C}ov[E_{2i},D]-P_1)X_{1i}+P_1X_{0i}-\frac{A_i}{2}\mathbb{V}ar[D](X_{1i})^2+\hat{T}_i,\end{aligned}$$

where $\hat{T}_i$ corresponds to a potential transfer set by the planner above and beyond the tax rebate, which is already accounted for. Exploiting the fact that $\mathbb{E}[U_i'(W_{2i})]=-A_i\mathbb{E}[U_i(W_{2i})]$, we can express $\frac{dV_i}{d\tau}$ as follows

$$\frac{dV_i}{d\tau}=A_ie^{-A_i\hat{V}_i}\frac{d\hat{V}_i}{d\tau}=\mathbb{E}\left[U_i'(W_{2i})\right]\frac{d\hat{V}_i}{d\tau}=-A_i\mathbb{E}[U_i(W_{2i})]\frac{d\hat{V}_i}{d\tau},$$

where, as above, $\frac{d\hat{V}_i}{d\tau}$ corresponds to

$$\frac{d\hat{V}_i}{d\tau}=[(\mathbb{E}[D]-\mathbb{E}_i[D])+\operatorname{sgn}(\Delta X_{1i})P_1\tau]\frac{dX_{1i}}{d\tau}-\Delta X_{1i}\frac{dP_1}{d\tau}.$$

Under Assumption [G], $\frac{dP_1}{d\tau} = 0$, we can express $\frac{dV}{d\tau} = \int \frac{dV_i}{d\tau} dF(i)$ as follows:

$$\frac{dV}{d\tau} = \int \underbrace{\mathbb{E}\left[U'\left(W_{2i}\right)\right]}_{\equiv h_i} \left[\left(\mathbb{E}[D] - \mathbb{E}_i[D]\right) + \operatorname{sgn}\left(\Delta X_{1i}\right) P_1 \tau\right] \frac{dX_{1i}}{d\tau} dF(i).$$

More generally, Assumption [G] guarantees that the equilibrium price is independent of the tax level and potential transfers. It also implies that equilibrium portfolio allocations, and consequently the identities of buyers/sellers are exclusively a function of τ and primitives, but not of transfers $\hat{T}_i$. At the optimum, τ^* satisfies

$$\tau^* = \frac{\int h_i \left(\mathbb{E}[D] - \mathbb{E}_i[D]\right) \frac{dX_{1i}}{d\tau} dF(i)}{-P_1 \int h_i \operatorname{sgn}\left(\Delta X_{1i}\right) \frac{dX_{1i}}{d\tau} dF(i)},$$

which can be expressed, using the fact that $\frac{dX_{1i}}{d\tau} = \frac{-\operatorname{sgn}(\Delta X_{1i}) P_1}{A \mathbb{V}ar[D]}$, as

$$\begin{aligned}
\tau^* &= \frac{\int h_i \left(\mathbb{E}[D] - \mathbb{E}_i[D]\right) \operatorname{sgn}\left(\Delta X_{1i}\right) dF(i)}{-P_1 \int h_i dF(i)} \\
&= \frac{\int_{\mathcal{B}} h_i \left(\frac{\mathbb{E}_i[D]}{P_1} - \frac{\mathbb{E}[D]}{P_1}\right) dF(i) - \int_{\mathcal{S}} h_i \left(\frac{\mathbb{E}_i[D]}{P_1} - \frac{\mathbb{E}[D]}{P_1}\right) dF(i)}{\int_{\mathcal{B}} h_i dF(i) + \int_{\mathcal{S}} h_i dF(i)},
\end{aligned}$$

where all right-hand side variables are functions of τ. After being rearranged, the optimal tax satisfies the following equation

$$\tau^* = \nu \mathbb{E}_{\mathcal{B}}^{h} \left[\frac{\mathbb{E}_i[D]}{P_1} - \frac{\mathbb{E}[D]}{P_1}\right] - (1-\nu) \mathbb{E}_{\mathcal{S}}^{h} \left[\frac{\mathbb{E}_i[D]}{P_1} - \frac{\mathbb{E}[D]}{P_1}\right], \tag{29}$$

where we define a variable $\nu \in [0,1]$,

$$\nu = \frac{\int_{\mathcal{B}} h_i dF(i)}{\int_{\mathcal{B}} h_i dF(i) + \int_{\mathcal{S}} h_i dF(i)} \quad \text{and} \quad 1 - \nu = \frac{\int_{\mathcal{S}} h_i dF(i)}{\int_{\mathcal{B}} h_i dF(i) + \int_{\mathcal{S}} h_i dF(i)},$$

and cross-sectional expectations under a modified measure indexed by h, given by

$$\begin{aligned}
\mathbb{E}_{\mathcal{B}}^{h} \left[\frac{\mathbb{E}_i[D]}{P_1} - \frac{\mathbb{E}[D]}{P_1}\right] &= \int_{\mathcal{B}} \frac{h_i}{\int_{\mathcal{B}} h_i dF(i)} \left(\frac{\mathbb{E}_i[D]}{P_1} - \frac{\mathbb{E}[D]}{P_1}\right) dF(i) \\
\mathbb{E}_{\mathcal{S}}^{h} \left[\frac{\mathbb{E}_i[D]}{P_1} - \frac{\mathbb{E}[D]}{P_1}\right] &= \int_{\mathcal{S}} \frac{h_i}{\int_{\mathcal{S}} h_i dF(i)} \left(\frac{\mathbb{E}_i[D]}{P_1} - \frac{\mathbb{E}[D]}{P_1}\right) dF(i),
\end{aligned}$$

where the variables $\frac{h_i}{\int_{\mathcal{B}} h_i dF(i)}$ and $\frac{h_i}{\int_{\mathcal{S}} h_i dF(i)}$ induce a change of measure in the cross-sectional distribution of investors, giving more weight to those investors with higher (expected) marginal utility. Note that Equation (29) is equivalent to Equation (11) in the limit when $h_i \to h$, with h being any constant. Formally,

$$\lim_{h_i \to h} \nu = \lim_{h_i \to h} 1 - \nu = \frac{1}{2},$$

$$\lim_{h_i \to h} \mathbb{E}_{\mathcal{B}}^{h} \left[\frac{\mathbb{E}_i[D]}{P_1} - \frac{\mathbb{E}[D]}{P_1}\right] - \mathbb{E}_{\mathcal{S}}^{h} \left[\frac{\mathbb{E}_i[D]}{P_1} - \frac{\mathbb{E}[D]}{P_1}\right] = \mathbb{E}_{\mathcal{B}} \left[\frac{\mathbb{E}_i[D]}{P_1}\right] - \mathbb{E}_{\mathcal{S}} \left[\frac{\mathbb{E}_i[D]}{P_1}\right],$$

so we recover an optimal tax formula that corresponds to Equation (11) in the text, under Assumption [G], that is

$$\tau^* = \frac{\mathbb{E}_{\mathcal{B}} \left[\frac{\mathbb{E}_i[D]}{P_1}\right] - \mathbb{E}_{\mathcal{S}} \left[\frac{\mathbb{E}_i[D]}{P_1}\right]}{2}.$$

When $\nu \approx \frac{1}{2}$, which occurs when the average welfare of buyers and sellers is approximately equal$\int_{\mathcal{B}} h_i dF(i) \approx \int_{S} h_i dF(i)$, the optimal tax can be expressed as

$$\tau^* \approx \frac{\mathbb{E}_{\mathcal{B}}\left[\frac{\mathbb{E}_i[D]}{P_1}\right] - \mathbb{E}_{\mathcal{S}}\left[\frac{\mathbb{E}_i[D]}{P_1}\right]}{2} + \frac{\mathbb{C}ov_{\mathcal{B}}\left[\frac{h_i}{\mathbb{E}_{\mathcal{B}}[h_i]}, \frac{\mathbb{E}_i[D]}{P_1}\right] - \mathbb{C}ov_{\mathcal{S}}\left[\frac{h_i}{\mathbb{E}_{\mathcal{S}}[h_i]}, \frac{\mathbb{E}_i[D]}{P_1}\right]}{2}.$$

This approximation is accurate for the baseline calibration when the planner's belief corresponds to the average belief — the difference between $\int_{\mathcal{B}} h_i dF(i)$ and $\int_{S} h_i dF(i)$ is of approximately 5%. This approximation is exact when the planner's belief corresponds to the average belief, $\mathbb{E}[D] = \mu_d$, and the risky asset is in zero net supply.

Distributional concerns: quantification

Given that the weights h_i assigned to each investor are complex endogenous objects, it is worth providing a quantitative analysis of the results in this case. As shown above, we can express equilibrium prices and relative (dividing by investors' initial holdings) net trading positions as[40]

$$P_1 = \mu_d - \mu_h - A\mathbb{V}ar[D]\,Q, \quad \text{and}$$

$$\frac{\Delta X_{1i}^{+}(P_1)}{Q} = \frac{\varepsilon_{di} - \varepsilon_{hi} - \tau P_1}{A\mathbb{V}ar[D]\,Q} \sim N\left(\frac{-\tau P_1}{A\mathbb{V}ar[D]\,Q}, \frac{\sigma_d^2 + \sigma_h^2 - 2\rho\sigma_d\sigma_h}{(A\mathbb{V}ar[D]\,Q)^2}\right)$$
$$\frac{\Delta X_{1i}^{-}(P_1)}{Q} = \frac{\varepsilon_{di} - \varepsilon_{hi} + \tau P_1}{A\mathbb{V}ar[D]\,Q} \sim N\left(\frac{\tau P_1}{A\mathbb{V}ar[D]\,Q}, \frac{\sigma_d^2 + \sigma_h^2 - 2\rho\sigma_d\sigma_h}{(A\mathbb{V}ar[D]\,Q)^2}\right).$$

Given the baseline parametrization of the model featuring $\frac{\sigma_d}{\sigma_h} = 0.5$ and $\rho = 0$, Figures A.1 to A.5 illustrate the behavior of the model and of the modified optimal tax. Unless explicitly stated, all calculations assume that the planner's belief corresponds to the average belief, so $\mathbb{E}[D] = \mu_d$.

Figure A.1 shows a random draw from the joint normal distribution of beliefs and hedging needs.[41] The different colors indicate which investors become buyers, sellers, or remain inactive. Investors with high (low) belief realizations ε_{di} and low (high) hedging realizations ε_{hi} become natural buyers (sellers) of the risky asset. Holding constant investors' beliefs ε_{di}, the realization of ε_{hi} generates residual variation in net trade positions.

Figure A.2 illustrates that the value of the difference $\varepsilon_{di} - \varepsilon_{hi}$ is the key determinant of trading positions, supporting the analytical results derived above. Investors with high (low) realizations of $\varepsilon_{di} - \varepsilon_{hi}$ become buyers (sellers) in equilibrium. Investors with $\varepsilon_{di} - \varepsilon_{hi}$ close to zero, optimally decide not to trade. As exploited throughout the paper, the net trading positions of investors are fully symmetric.

[40] Note that only information on Π and Ξ is again necessary to pin down relative net trading positions. Formally

$$\frac{\Delta X_{1i}^{-}(P_1)}{Q} \sim N\left(-\frac{\tau}{\Pi}, 2\pi\Xi^2\right) \quad \text{and} \quad \frac{\Delta X_{1i}^{+}(P_1)}{Q} \sim N\left(\frac{\tau}{\Pi}, 2\pi\Xi^2\right).$$

[41] Figures A.1 through A.3 and A.5 employ 5,000 investors and use $\tau = 0.173\%$. Figure A.4 employs 5,000 investors and reports the average of optimal taxes for 200 realizations. When needed, I normalize $P_1 = Q = 1$, and assume that $A = 10^{-6}$ and $\mathbb{V}ar[D] = (6\%)^2$.

Figure A.1: Realization of beliefs ε_{di} and hedging needs ε_{hi}

Figure A.3 illustrates how investor's welfare in certainty equivalents varies with $\varepsilon_{di} - \varepsilon_{hi}$, where this figure assumes that $\mathbb{E}[D] = \mu_d$.[42] The fact that the middle plot has an inverted U-shape implies that investors with extreme beliefs tend to be worse than investors with average beliefs. It also shows that natural sellers due to hedging motives are on average worse than buyers (right plot). However, this effect reverses for extreme sellers, implying that investors who sell much of the risky asset are better off than moderate sellers. Once both effects are combined, the leftmost figure shows that buyers are on average better off than sellers.

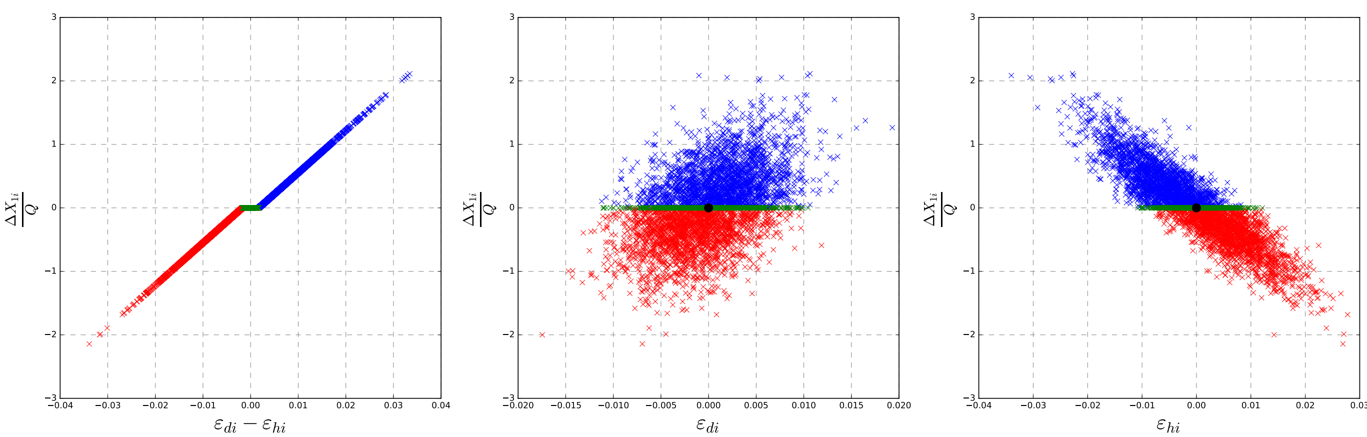

Figure A.2: Net trading positions

It is worth providing some analytical insight behind these welfare patterns. We can express $\hat{V}_i$, as defined in Equation (28), as

$$\hat{V}_i = \left(\mathbb{E}[D] - A\mathbb{C}ov[E_{2i}, D] - P_1\right) X_{1i} - \frac{A}{2}\mathbb{V}ar[D]\left(X_{1i}\right)^2 + P_1 X_{0i}.$$

[42] I multiply by A only for scaling purposes. This does not affect any relative welfare comparisons. Note that h_i is related to $A_i\hat{V}_i$ through $h_i = A_i e^{-A_i\hat{V}_i}$. For this particular parametrization, the realizations of $\hat{V}_i$ are such that h_i and $\hat{V}_i$ satisfy an almost linear relation.

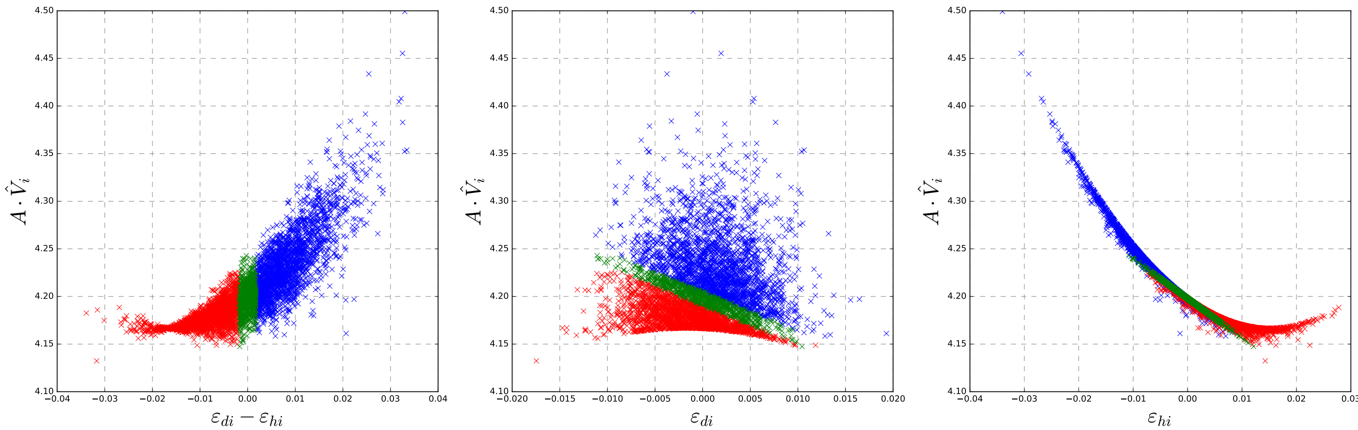

Figure A.3: Investors' certainty equivalents

In particular, we can express the certainty equivalent of investors who decide not to trade as

$$\hat{V}_i\left(\Delta X_{1i}=0\right)=\left(\mathbb{E}\left[D\right]-P_1-\underbrace{A\mathbb{C}ov\left[E_{2i},D\right]}_{\varepsilon_{hi}}-\frac{A}{2}\mathbb{V}ar\left[D\right]X_{0i}\right)X_{0i}+P_1X_{0i}.$$

In autarky, investors with a high ε_{hi} (natural sellers of the risky asset) are in general worse off. They have inherited positive asset holdings of the risky asset, and increasing $\mathbb{C}ov\left[E_{2i},D\right]$ increases the total risky associated with that position. Interestingly, conditional on being inactive, investors' beliefs are irrelevant for the planner's welfare assessment. Inactive investors with a high ε_{di} are worse off only because they must also have a high ε_{hi}, but not independently.

Similarly, we can express the certainty equivalent of investors who decide to trade as follows:

$$\begin{aligned}\hat{V}_i&=\left(\mathbb{E}\left[D\right]-\mathbb{E}_i\left[D\right]\right)X_{1i}+\frac{1}{2}\left(\mathbb{E}_i\left[D\right]-A\mathbb{C}ov\left[E_{2i},D\right]-P_1\left(1+\tau\,\mathrm{sgn}\left(\Delta X_{1i}\right)\right)\right)X_{1i}+P_1X_{0i}\\&=\frac{1}{2}\left[\underbrace{\left(\mathbb{E}\left[D\right]-\mathbb{E}_i\left[D\right]\right)X_{1i}}_{\text{Belief Difference}}+\left(\underbrace{\overbrace{\mathbb{E}\left[D\right]-P_1\left(1+\tau\,\mathrm{sgn}\left(\Delta X_{1i}\right)\right)}^{\text{Risk Premium}}-\overbrace{A\mathbb{C}ov\left[E_{2i},D\right]}^{\text{Covariance Risk}}}_{\text{Risk Compensation (net of of taxes)}}\right)X_{1i}\right]+P_1X_{0i},\end{aligned}\qquad(30)$$

where X_{1i} satisfies

$$X_{1i}=\frac{\mathbb{E}_i\left[D\right]-A\mathbb{C}ov\left[E_{2i},D\right]-P_1\left(1+\tau\,\mathrm{sgn}\left(\Delta X_{1i}\right)\right)}{A\mathbb{V}ar\left[D\right]}.$$

Intuitively, when $\mathbb{E}\left[D\right]=\mathbb{E}_i\left[D\right]$, and $\tau=0$, $\hat{V}_i$ corresponds to

$$\hat{V}_i=\frac{1}{2}\frac{1}{A\mathbb{V}ar\left[D\right]}\left(A\mathbb{V}ar\left[D\right]Q-\underbrace{A\mathbb{C}ov\left[E_{2i},D\right]}_{\varepsilon_{hi}}\right)^2+P_1X_{0i},$$

whose minimum is found when $\varepsilon_{hi}=A\mathbb{V}ar\left[D\right]Q$, which for the current parametrization corresponds to 0.015, explaining where the lower values of $\hat{V}_i$ in the right plot of Figure are found around the value $\varepsilon_{hi}=0.015$. Note that when $Q=X_{0i}=0$ and $\mathbb{E}\left[D\right]=\mu_d$, Equation (30) is symmetric in investors' beliefs.

If the planner had access to ex-ante transfers, Figure A.3 allows to easily recover the pattern of required ex-ante transfers. Investors with $\hat{V}_i$ higher than a threshold will have negative ex-ante transfers

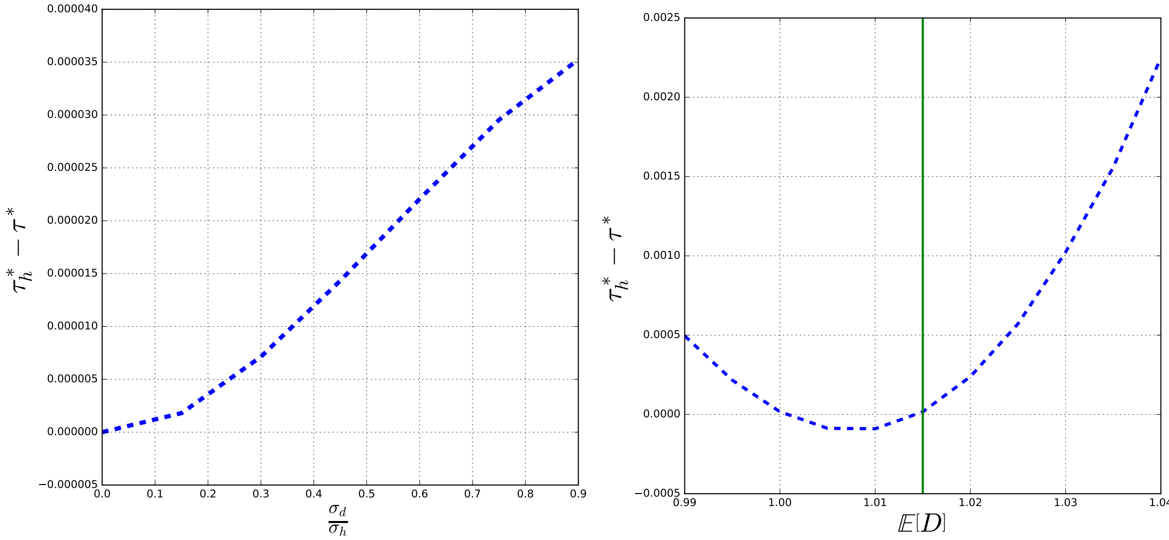

Figure A.4: Difference between optimal taxes with and without distributional motives $\tau_h^* - \tau^*$

until their welfare equals the threshold, while those below such threshold will receive transfers. The threshold is chosen such that total transfers add up to zero, that is, $\sum_i \hat{T}_i = 0$. Given Figure A.3, one expects buyers to face negative ex-ante transfer (payments), while sellers receive positive transfers. If the risky asset is in zero-net supply, recovering symmetry, one expects investors who trade more to receive positive transfers, while those who trade less to face negative transfers.

Figure A.4 shows the difference between the optimal tax with distributional concerns, which is denoted by τ_h^* and that corresponds to Equation (29), and the optimal tax without distributional concerns, which is denoted by τ^* and that corresponds to Equation (8) in the text.

The left plot in Figure A.4 illustrates how the difference between optimal taxes varies with the level of non-fundamental volume, parametrized by $\frac{\sigma_d}{\sigma_h}$. It shows that the desire to increase the optimal tax increases due to distributional concerns beyond the Pigovian rationale are increasing in $\frac{\sigma_d}{\sigma_h}$. Intuitively, as per the middle plot in Figure A.3, a relatively higher $\frac{\sigma_d}{\sigma_h}$ exacerbates the belief dispersion, causing investors with very extreme beliefs to have low indirect utilities. As illustrated in Equation (17) in the text, this generates a new rationale to set a higher tax.

The right plot in Figure A.4 shows the difference between the optimal tax with distributional concerns and without them, respectively denoted by $\tau_h^* - \tau^*$, for different values of $\mathbb{E}[D]$. When $\mathbb{E}[D] = \mu_d$ — highlighted by the solid green vertical line — the tax differential $\tau_h^* - \tau^*$ is positive by a small amount, consistently with the left plot. A slightly pessimistic planner on average relative to investors puts more weight on the welfare losses of buyers, who are on average better off, making optimal to reduce the tax. When the planner is very optimistic (pessimistic) relative to investors, the planner becomes more and more concerned about the low expected marginal utility/wealth of sellers (buyers) relative to their marginal portfolio holding distortions. Therefore, purely to avoid having a group of investors with very low marginal utility given his belief, the planner decides to set a higher optimal tax when his belief is too different from the average belief of investors.

Figure A.4 implies that distributional concerns don't seem to be of great quantitative importance for the preferred calibration in the paper. When $\mathbb{E}[D] = \mu_d$, as shown in the left plot, the change in $\tau_h^* - \tau^*$ for different values of $\frac{\sigma_h}{\sigma_d}$ is several orders of magnitude too small when compared to τ^*. Although

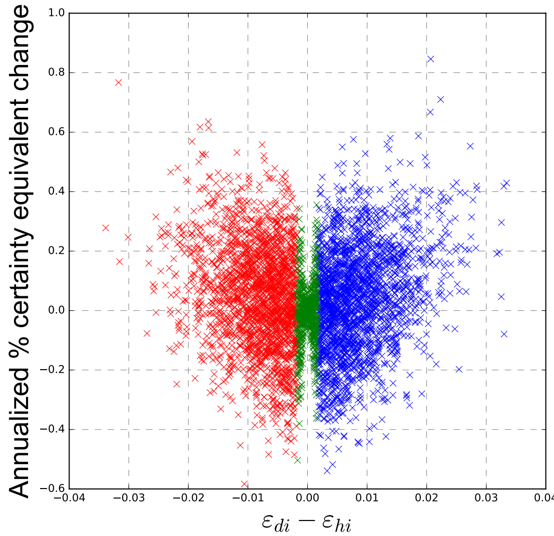

Figure A.5: Individual welfare changes

assuming extreme differences between $\mathbb{E}[D]$ and μ_d may change the optimal tax prescription, as long as the planner's belief does not differ much from the average belief, the optimal tax without redistributional concerns remains approximately optimal.

Finally, Figure A.5 illustrates the pattern of winners and losers associated with a transaction tax policy (this figure assumes that $\mathbb{E}[D] = \mu_d$). It shows the annualized change in investors' certainty equivalents from the planner's perspective in scenarios with and without a transaction tax ($\tau = 0$). I assume that τ^* corresponds to the optimal tax without distributional concerns, but the insights are equivalent when using τ_h^*. The "heart shape" of Figure A.5 clearly illustrates i) that most investors benefit from the optimal tax policy, and ii) that investors with extreme beliefs ε_{di} benefit the most on average from the policy. Intuitively, investors who trade the least are the least affected by the policy.

B Section 6: Proofs and derivations

Proposition 4. (Optimal tax approximation with general utility and arbitrary beliefs)[43]

a) Social welfare is given by $V(\tau) = \int V_i dF(i)$, where V_i denotes indirect utility from the planner's perspective, that is:

$$V_i \equiv \mathbb{E}[U_i(W_{2i})],$$

[43] For notational simplicity, I removed the explicit dependence of all equilibrium variables on τ.

where $W_{2i} = E_{2i} + X_{1i}D + (X_{0i}P_1 - X_{1i}P_1 - \tau |P_1| |\Delta X_{1i}| + T_{1i})$, $T_{1i} = \tau |P_1| |\Delta X_{1i}|$, and X_{1i} and P_1 are equilibrium determined. We can express $\frac{dV_i}{d\tau}$ as follows:

$$\begin{aligned}\frac{dV_i}{d\tau} &= \mathbb{E}\left[U_i'(W_{2i})\frac{dW_{2i}}{d\tau}\right] = \mathbb{E}\left[U_i'(W_{2i})\left[(D-P_1)\frac{dX_{1i}}{d\tau} - \Delta X_{1i}\frac{dP_1}{d\tau}\right]\right] \\ &= \mathbb{E}\left[U_i'(W_{2i})(D-P_1)\right]\frac{dX_{1i}}{d\tau} - \mathbb{E}\left[U_i'(W_{2i})\right]\Delta X_{1i}\frac{dP_1}{d\tau} \\ &= \left[-\mathbb{C}ov\left[Z_i, U_i'(W_{2i})(D-P_1)\right] + \mathbb{E}_i\left[U_i'(W_{2i})\right]\tau P_1 \operatorname{sgn}(\Delta X_{1i})\right]\frac{dX_{1i}}{d\tau} - \mathbb{E}\left[U_i'(W_{2i})\right]\Delta X_{1i}\frac{dP_1}{d\tau},\end{aligned}$$

where the last line exploits investors' optimality condition in Equation (18), which implies that

$$\mathbb{E}\left[U_i'(W_{2i})(D-P_1)\right] = -\mathbb{C}ov\left[Z_i, U_i'(W_{2i})(D-P_1)\right] + \mathbb{E}_i\left[U_i'(W_{2i})\right]\tau P_1 \operatorname{sgn}(\Delta X_{1i}).$$

Hence, we can express $\frac{dV}{d\tau} = \int \frac{dV_i}{d\tau} dF(i)$ as follows:

$$\begin{aligned}\frac{dV}{d\tau} = &\int \left[-\mathbb{C}ov\left[Z_i, U_i'(W_{2i})(D-P_1)\right] + \mathbb{E}_i\left[U_i'(W_{2i})\right]\tau P_1 \operatorname{sgn}(\Delta X_{1i})\right]\frac{dX_{1i}}{d\tau} dF(i) \\ &- \frac{dP_1}{d\tau}\int \mathbb{E}\left[U_i'(W_{2i})\right]\Delta X_{1i} dF(i).\end{aligned}$$

Hence, the optimal tax must satisfy

$$\tau^* = \frac{\int \mathbb{C}ov\left[Z_i, U_i'(W_{2i})(D-P_1)\right]\frac{dX_{1i}}{d\tau}dF(i) + \frac{dP_1}{d\tau}\int \mathbb{E}\left[U_i'(W_{2i})\right]\Delta X_{1i} dF(i)}{P_1 \int \mathbb{E}_i\left[U_i'(W_{2i})\right]\operatorname{sgn}(\Delta X_{1i})\frac{dX_{1i}}{d\tau}dF(i)},$$

or equivalently

$$\tau^* = \frac{\int \mathbb{E}\left[(Z_i - 1)U_i'(W_{2i})(D-P_1)\right]\frac{dX_{1i}}{d\tau}dF(i) + \frac{dP_1}{d\tau}\int \mathbb{E}\left[U_i'(W_{2i})\right]\Delta X_{1i} dF(i)}{P_1 \int \mathbb{E}_i\left[U_i'(W_{2i})\right]\operatorname{sgn}(\Delta X_{1i})\frac{dX_{1i}}{d\tau}dF(i)}.$$

b) When risks are small, and marginal utilities are approximately constant, implying that $U_i'(W_{2i}) \approx 1$, the optimal tax in Equation (8) (approximately) corresponds to

$$\tau^* \approx \frac{\int \mathbb{C}ov\left[Z_i, \frac{D}{P_1}\right]\frac{dX_{1i}}{d\tau}dF(i)}{\int \operatorname{sgn}(\Delta X_{1i})\frac{dX_{1i}}{d\tau}dF(i)} = \frac{1}{2}\frac{\int \frac{\mathbb{E}_i[D]}{P_1}\frac{dX_{1i}}{d\tau}dF(i)}{\int_{i\in\mathcal{B}(\tau)}\frac{dX_{1i}}{d\tau}dF(i)} = \frac{\Omega_{\mathcal{B}(\tau^*)} - \Omega_{\mathcal{S}(\tau^*)}}{2},$$

where $\Omega_{\mathcal{B}(\tau^*)}$ and $\Omega_{\mathcal{S}(\tau^*)}$ are described in Equations (8) and (9) in Proposition 1. This expression follows by taking the limit in Equation (8) when $U_i'(W_{2i}) \to 1$ (or any other constant), and uses the fact that $\mathbb{C}ov(Z_i, D) = \mathbb{E}_i[D] - \mathbb{E}[D]$.

C Extensions[44]

I now study multiple extensions of the benchmark model. Earlier versions of this paper contained additional and more detailed extensions.

[44]Once again, to ease notation, I avoid explicit dependence of endogenous variables on P_1 and τ.

C.1 Pre-existing trading costs

Because actual investors face trading costs even when there are no taxes, one could wonder about the validity of the results derived around the point $\tau = 0$. The optimal tax formula is still valid as long as transaction costs are a mere compensation for the use of economic resources.[45]

Assumptions Investors now face transaction costs, regardless of the value of τ. These represent costs associated with trading, like brokerage commissions, exchange fees or bookkeeping costs. Investors must pay a quadratic cost, parametrized by α, a linear cost η on the number of shares traded, and a linear cost ψ on the dollar volume of the transaction. These trading costs are paid to a new group of agents (intermediaries) which facilitate the process of trading. Crucially, I assume that intermediaries make zero profits in equilibrium. Hence, wealth at date 2 for an investor i is now given by:

$$W_{2i} = E_{2i} + X_{1i}D + \left(X_{0i}P_1 - X_{1i}P_1 - |\Delta X_{1i}| \, |P_1| \, (\tau + \psi) - \eta \, |\Delta X_{1i}| - \frac{\alpha}{2} \left(\Delta X_{1i}\right)^2 + T_{1i} \right). \quad (31)$$

The transfer rebates tax revenues, but not trading costs, to investors.

Results The demand for the risky asset takes a similar form as in the baseline model, featuring also an inaction region, now determined jointly by the trading costs and the transaction tax. The optimal portfolio given prices can be compactly written in the trade region as:

$$X_{1i} = \frac{\mathbb{E}_i [D] - A_i \mathbb{C}ov [E_{2i}, D] - P_1 (1 + \text{sgn} (\Delta X_{1i}) (\tau + \psi)) - \text{sgn} (\Delta X_{1i}) \, \eta + \alpha X_{0i}}{A_i \mathbb{V}ar [D] + \alpha}.$$

All three types of trading costs — quadratic, linear in shares and linear in dollar value — shift investors' portfolios towards their initial positions. The equilibrium price is a slightly modified version of (6).

When calculating welfare, the planner takes into account that investors must incur in these costs when trading — this is the natural constrained efficient benchmark. The optimal tax formula remains unchanged when investors face transaction costs, as long as these trading costs represent exclusively a compensation for the use of economic resources.

Proposition 5. (Pre-existing trading costs) *When investors face trading costs as specified in (31), the expression for τ^* is identical to the one in Proposition 1.*

The intuition behind Proposition 5 is similar to the baseline case. An envelope condition eliminates any term regarding transaction costs from $\frac{dV}{d\tau}$, because the planner must also face such costs, so the optimal tax looks identical to the one in the baseline model. This relies on the assumption that the economic profit made by the intermediaries who receive the transaction

[45] There is scope to study in more detail the interaction of trading costs in models that deliver endogenous bid-ask spreads in models with differentially informed traders, like that of Glosten and Milgrom (1985).

costs is zero — there cannot be economic rents.[46]

This result has further implications. First, although the optimal tax formula does not vary, an economy with transaction costs has less trade in equilibrium than one without transaction costs. Depending on whether this reduction in trading is of the fundamental type or not, the optimal tax may be larger or smaller. Transaction costs affect the optimal tax through changes in the identity of the marginal investors. Second, the mere existence of transaction costs does not provide a new rationale for further discouraging non-fundamental trading. Welfare losses must be traced back to wedges derived from portfolio distortions. Third, if transactions costs, that is, ψ, η and α, were endogenously functions of τ, as in richer models of the market microstructure, the planner would have to take into account those effects when solving for optimal taxes.[47] For instance, if a transaction tax endogenously increases trading costs, the optimal tax may be very small. However, if endogenously determined transaction costs are efficiently determined, the envelope theorem would still apply, leaving Proposition 5 unchanged.

C.2 Imperfect tax enforcement

All the results in the paper have been derived under the assumption of perfect tax enforcement. I now show how introducing imperfect tax enforcement does not change the main qualitative predictions of the paper. I also show that imperfect enforcement is associated with lower optimal taxes.[48]

Assumptions Investors can now trade in two different markets, A and B. Market A captures existing venues for trading, and all trades in that market face a transaction tax τ. Market B seeks to represent trading venues that cannot be monitored by authorities. In market B, investors face instead a quadratic cost of trading, parametrized by α. When $\alpha \to 0$, avoiding the tax is costless, and all trades move to market B for any values of τ, so for regularity purposes, α must be sufficiently large, which is consistent with the empirical evidence discussed in the text.[49] Varying α modulates the costs of evasion. To simplify notation, at times I define $\hat{\alpha} = \frac{\alpha}{A_i \mathbb{V}ar[D]}$. Investors' initial endowments are X_{0i}^A market A shares and $X_{0i}^B = 0$ market B shares (without loss of generality).

[46] Does Proposition 5 imply that if there were two authorities with taxation power, they would both impose the same τ^* twice? That is not case. Assume for simplicity that they set taxes sequentially. The first authority would set the optimal tax according to Proposition 1. The second authority, internalizing that the pre-existing tax is a mere transfer and does not correspond to a compensation for costs of trading, would set a zero tax. Alternatively, τ^* would characterize the sum of both taxes.

[47] The Walrasian approach of this paper does not capture market microstructure effects. There is scope for understanding how transaction taxes affect market making and liquidity provision in greater detail, introducing, for instance, imperfectly competitive investors, search or network frictions. The results of this paper would still be present regardless of the specific trading microstructure.

[48] There is scope for further research understanding the specificities of tax avoidance/evasion when there is competition among exchanges, as in Santos and Scheinkman (2001).

[49] Assuming a different form of costs yields similar insights.

Hence, wealth at date 2 for an investor i is now given by:

$$W_{2i} = E_{2i} + X_{1i}D + \left(-\Delta X_{1i}^{A} P_1^{A} - \tau \left|P_1^{A}\right| \left|\Delta X_{1i}^{A}\right| - \Delta X_{1i}^{B} P_1^{B} - \frac{\alpha}{2} \left(\Delta X_{1i}^{B}\right)^2 + T_{1i}\right),$$

where I define $X_{1i} = X_{1i}^A + X_{1i}^B$, same for $t = 0$. The transfer rebates tax revenues, but not the costs of trading in the B market, to investors. Note that the linear tax only affects trading in market A, while the quadratic cost only affects trading in market B.

Results Now investors must formulate demands for both markets. Investors' optimality conditions correspond to

$$X_{1i} = X_{1i}^A + X_{1i}^B = \frac{\left[\mathbb{E}_i\left[D\right] - A_i \mathbb{C}ov\left[E_{2i}, D\right] - P_1^A\right] - \tau \left|P_1^A\right| \operatorname{sgn}\left(\Delta X_{1i}^A\right)}{A_i \mathbb{V}ar\left[D\right]} \tag{32}$$

$$X_{1i} = X_{1i}^A + X_{1i}^B = \frac{\left[\mathbb{E}_i\left[D\right] - A_i \mathbb{C}ov\left[E_{2i}, D\right] - P_1^B\right]}{A_i \mathbb{V}ar\left[D\right]} - \hat{\alpha} X_{1i}^B. \tag{33}$$

Note that we generically expect $\Delta X_{1i}^B \neq 0$. Whenever $\Delta X_1^A \neq 0$, by combining Equations (32) and (33), X_1^B must satisfy

$$X_1^B = \frac{\tau}{\hat{\alpha}} \frac{\left|P_1\right| \operatorname{sgn}\left(\Delta X_{1i}^A\right)}{A_i \mathbb{V}ar\left[D\right]}.$$

Whenever investors are inactive in market A, so $\Delta X_1^A = 0$, X_1^B is given by

$$X_{1i}^B = \frac{1}{1+\hat{\alpha}} \left(\frac{\left[\mathbb{E}_i\left[D\right] - A_i \mathbb{C}ov\left[E_{2i}, D\right] - P_1^B\right]}{A_i \mathbb{V}ar\left[D\right]} - X_{0i}^A \right).$$

Note that, regardless of the scenario, $\frac{dX_{1i}}{d\tau}$ is weakly lower in absolute value. To simplify the exposition, I impose assumption S from now on, which guarantees that $P_1^A = P_1^B = P_1$, although similar results can be found for the general case.

Proposition 6. (Imperfect tax enforcement) *When investors can trade in an alternative market without facing the tax, the sign of the optimal is given by the sign of*

$$\left.\frac{dV}{d\tau}\right|_{\tau=0} = -\int \mathbb{E}_i\left[D\right] \left.\frac{dX_{1i}}{d\tau}\right|_{\tau=0} dF\left(i\right), \tag{34}$$

where $\frac{dX_{1i}}{d\tau} = \frac{dX_{1i}^A}{d\tau} + \frac{dX_{1i}^B}{d\tau}$. *The expression for* τ^* *corresponds to*

$$\tau^* = \frac{\int \mathbb{E}_i\left[D\right] \frac{dX_{1i}}{d\tau} dF\left(i\right)}{\int \operatorname{sgn}\left(\Delta X_{1i}\right) P_1 \frac{dX_{1i}^A}{d\tau} dF\left(i\right)}. \tag{35}$$

Now the numerator of the optimal tax formula accounts for the change in volume in both markets, while the denominator only accounts for changes in the market in which the optimal tax is paid. Intuitively, now the same tax change is less effective in reducing non-fundamental trades, so the optimal tax chosen by a planner is lower than before. Note that the condition that determines the sign of the optimal tax is identical with and without perfect enforcement after accounting for the total volume reduction in both markets. Since it can be shown that $\frac{dX_{1i}}{d\tau}$ has the same sign as in the model with perfect enforcement, the qualitative insights for the sign of the optimal tax go through unchanged.

C.3 Multiple risky assets

Assumptions The results of the baseline model extend naturally to an environment with multiple assets. Now there are J risky assets in fixed supply, in addition to the riskless asset. The $J \times 1$ vectors of total shares, equilibrium prices and dividend payments are respectively denoted by $\mathbf{q}$, $\mathbf{p}$ and $\mathbf{d}$.[50] Every purchase or sale of a risky asset faces an identical linear transaction tax τ. This is a further restriction on the planner's problem, since belief disagreements can vary across different assets, but the tax must be constant. Allowing for different taxes for different (groups of) assets is conceptually straightforward, following the logic of Section C.4.

The distribution of dividends $\mathbf{d}$ paid by the risky assets is a multivariate normal with a given mean and variance-covariance matrix $\mathbb{V}ar\left[\mathbf{d}\right]$. All investors agree about the variance, but an investor i believes that the mean of $\mathbf{d}$ is $\mathbb{E}_i\left[\mathbf{d}\right]$. We can thus write:

$$\mathbf{d} \sim_i N\left(\mathbb{E}_i\left[\mathbf{d}\right], \mathbb{V}ar\left[\mathbf{d}\right]\right),$$

where risk aversion A_i, and the vectors of initial asset holdings $\mathbf{x}_{0i}$, hedging needs $\mathbb{C}ov\left[E_{2i}, \mathbf{d}\right]$ and beliefs $\mathbb{E}_i\left[\mathbf{d}\right]$ are arbitrary across the distribution of investors. The wealth at $t = 2$ of an investor i is thus given by:

$$W_{2i} = E_{2i} + \mathbf{x}_{1i}'\mathbf{d} + \left(\mathbf{x}_{0i}'\mathbf{p} - \mathbf{x}_{1i}'\mathbf{p} - \left|\mathbf{x}_{1i}' - \mathbf{x}_{0i}'\right|\mathbf{p}\tau + T_{1i}\right).$$

Results The first order condition (36) characterizes the solution of this problem for the set of assets traded:

$$\mathbf{x}_{1i} = \left(A_i \mathbb{V}ar\left[\mathbf{d}\right]\right)^{-1}\left(\mathbb{E}_i\left[\mathbf{d}\right] - A_i \mathbb{C}ov\left[E_{2i}, \mathbf{d}\right] - \mathbf{p} - \hat{\mathbf{p}}_i\tau\right), \tag{36}$$

where $\hat{\mathbf{p}}_i$ is a $J \times 1$ vector where row j is given by $\text{sgn}\left(\Delta X_{1ij}\right) p_j$ and p_j denotes the price of asset j. If an asset j is not traded by an investors i, then $X_{1ij} = X_{0ij}$. If asset returns are independent, the portfolio allocation to every asset can be determined in isolation. Equilibrium prices are the natural generalization of the baseline model.

Proposition 7. (Multiple risky assets) *The optimal tax when investors can trade J risky assets is given by:*

$$\tau^* = \sum_{j=1}^{J} \omega_j \tau_j^*, \tag{37}$$

with weights ω_j and individual-asset taxes τ_j^* given by $\omega_j \equiv \frac{p_j \int \text{sgn}(\Delta X_{1ij}) \frac{dX_{1ij}}{d\tau} dF(i)}{\sum_{j=1}^{J} p_j \int \text{sgn}(\Delta X_{1ij}) \frac{dX_{1ij}}{d\tau} dF(i)}$ *and* $\tau_j^* \equiv \frac{\int \frac{\mathbb{E}_i\left[D_j\right]}{p_j} \frac{dX_{1ij}}{d\tau} dF(i)}{\int \text{sgn}(\Delta X_{1ij}) \frac{dX_{1ij}}{d\tau} dF(i)}$.

The formula for τ_j^* is identical to the one in an economy with a single risky asset. The optimal tax in a model with J risky assets is simply a weighted average of all τ_j^*. The weights

[50] I use bold lower-case letters to denote vectors but, for consistency, I keep the upper-case notation for holdings of a single asset.

are determined by the relative marginal changes in (dollar) volume. Those assets whose volume responds more aggressively to tax changes carry higher weights when determining the optimal tax and vice versa.[51]

C.4 Asymmetric taxes/Multiple tax instruments

In the baseline model, the only instrument available to the planner is a single linear financial transaction tax which applies symmetrically to all investors. However, the planner could set different (linear) taxes for buyers and sellers. Or, at least theoretically, even investor-specific taxes. In general, more sophisticated policy instruments bring the outcome of the planner's problem closer to the first-best, at the cost of increasing informational requirements.

Asymmetric taxes on buyers versus sellers

Assume now that buyers pay a linear tax τ_B in the dollar volume of the transaction while sellers pay τ_S. Hence, total tax revenue is given by $(\tau_B + \tau_S)\, P_1\, |\Delta X_{1i}|$. Outside of the inaction region, the optimal portfolio demand is given by:

$$X_{1i} = \frac{\mathbb{E}_i\,[D] - A_i \mathbb{C}ov\,[E_{2i}, D] - P_1\,(1 + \mathbb{I}\,[\Delta X_{1i} > 0]\,\tau_B + \mathbb{I}\,[\Delta X_{1i} < 0]\,\tau_S)}{A_i \mathbb{V}ar\,[D]},$$

where $\mathbb{I}\,[\cdot]$ denotes the indicator function. This expression differs from (5) in that buyers now face a different tax than sellers. The equilibrium price is a natural extension of the one in the baseline model.

Proposition 8. (Asymmetric taxes on buyers versus sellers) *The pair of optimal taxes $\{\tau_B^*, \tau_S^*\}$ is characterized by the solution of the following system of non-linear equations:*

$$\tau_B^* + \tau_S^* = \frac{\int \frac{\mathbb{E}_i[D]}{P_1} \frac{dX_{1i}}{d\tau_B} dF\,(i)}{\int_{\mathcal{B}} \frac{dX_{1i}}{d\tau_B} dF\,(i)}, \qquad \tau_B^* + \tau_S^* = \frac{\int \frac{\mathbb{E}_i[D]}{P_1} \frac{dX_{1i}}{d\tau_S} dF\,(i)}{\int_{\mathcal{B}} \frac{dX_{1i}}{d\tau_S} dF\,(i)}. \tag{38}$$

The economic forces that shape the optimal values for τ_B^* and τ_S^* are the same as in the baseline model. Once again, the planner's belief is irrelevant for the optimal policy, which shows that that results is not sensitive to the use of more sophisticated policy instruments. Intuitively, the change in portfolio allocations induced by a marginal change in any instrument must cancel out in the aggregate. Equation (38) provides intuition for why all taxes in the baseline model are divided by 2; in that case, there exists a single optimality condition and $2\tau^* = \tau_B^* + \tau_S^*$.

As long as there are more than two investors, this system has at least a solution. When there are two investors, the system is indeterminate and only the sum $\tau_B^* + \tau_S^*$ is pinned down. In that case, $\tau_B^* + \tau_S^* = \frac{\mathbb{E}_{\mathcal{B}}[D] - \mathbb{E}_{\mathcal{S}}[D]}{P_1}$.

[51] The model with many assets contains interesting predictions for how the set of assets traded in equilibrium endogenously adjust to tax changes. There is scope for further analysis on that question.

Individual taxes/First-best

Assume now that the planner can set investor specific taxes. This is an interesting theoretical benchmark, despite being unrealistic. For simplicity, I now assume that there is a finite number N of (types of) investors in the economy.

Proposition 9. (Individual taxes/First-best)

a) The first-best can be implemented with a set of investor specific taxes given by:

$$\tau_i^* = \operatorname{sgn}(\Delta X_{1i}) \frac{\mathbb{E}_i[D] - \Upsilon}{P_1}, \forall i = 1, \ldots, N, \tag{39}$$

where Υ is any real number; a natural choice for Υ is $\mathbb{E}[D]$.

b) The planner only needs $N-1$ taxes to implement the first-best in an economy with N investors.

Proposition 9a follows standard Pigovian logic. The planner sets optimal individual taxes so that investors portfolio choices replicate those of a economy with homogeneous beliefs. Note that the planner can use any belief Υ to implement the first-best allocation, as long as it is the same for all investors. In a production economy, the natural choice would be $\Upsilon = \mathbb{E}[D]$. Finally, because P_1 is a function of all taxes, Equation (39) also defines a system of non-linear equations.

Proposition 9b shows that the first-best could be implemented with $N-1$ taxes. This occurs because the risky asset is in fixed supply. The logic behind this result is similar to Walras' law. For instance, when $N = 2$, a single tax which modifies directly the allocation of one of the investors necessarily changes the allocation of the other one through market clearing.

C.5 Production

The results derived so far rely on the assumption that assets are in fixed supply. I now study how optimal policies vary when financial markets determine production by influencing the intertemporal investment decision in a standard price-taking environment — this is the role explored in classic q-theory models.

Assumptions There is a new group of agents in the economy who were not present in the baseline model: identical competitive producers in unit measure. Producers are indexed by k and maximize well-behaved time separable expected utility, with flow utility given by $U_k(\cdot)$. They have exclusive access to a technology $\Phi(S_{1k})$, which allows them to issue or dispose of S_{1k} shares of the risky asset at date 1.[52] I refer to S_{1k}, which can be negative, as investment. The function $\Phi(\cdot)$ is increasing and strictly convex; that is, $\Phi'(\cdot) > 0$, $\Phi''(\cdot) > 0$. To ease the exposition, I

[52] A "tree" analogy can be helpful here. Assume that a share of the risky asset (i.e., a tree) entitles the owner to a dividend payment D (fruit). Producers can plant new trees or chop them at a cost $\Phi(S_{1k})$, which they sell or buy at a price P_1. Producers would be willing to create trees until the marginal cost of producing a new tree/chopping and old tree $\Phi'(S_{1k})$ equals the marginal benefit of selling/buying P_1. For consistency, any normalization concerning Q must also normalize $\Phi(\cdot)$.

assume throughout that $\Phi(S_{1k}) = \gamma_1 |S_{1k}| + \frac{\gamma_2}{2} |S_{1k}|^2$, with $\gamma_1, \gamma_2 > 0$. Producers are initially endowed with E_{1k} units of consumption good (dollars) and can only borrow or save in the riskless asset at a (gross) rate $R = 1$. Their endowment E_{2k} at date 2 is stochastic and follows an arbitrary distribution.

To avoid distortions in primary markets, the planner does not tax the issuance of new shares. Importantly, market clearing is now given by $\int X_{1i} dF(i) = Q + S_{1k}$. Total output at date 2 in this economy is endogenous and given by $D(Q + S_{1k})$.

Positive results Producers thus maximize:

$$\max_{C_{1k}, C_{2k}, S_{1k}, Y_k} U_k(C_{1k}) + \mathbb{E}[U_k(C_{2k})].$$

With budget constraints $Y_k + C_{1k} = E_{1k} + P_1^s S_{1k} - \Phi(S_{1k})$ and $C_{2k} = E_{2k} + Y_k$, where Y_k denotes the amount saved in the riskless asset and P_1^s denotes the price faced by producers — the superscript s stands for supply. The optimality conditions for producers are given by:

$$U_k'(C_{1k}) = \mathbb{E}[U_k'(C_{2k})] \quad \text{and} \quad P_1^s = \Phi'(S_{1k}).$$

The first condition is a standard Euler condition for the riskless asset. The second condition provides a supply curve for the number of shares. Combining this supply curve with the portfolio choices of investors, generates the following equilibrium price:

$$P_1 = (1 - \alpha)\gamma_1 + \alpha P_1^e,$$

where the weight $\alpha \in [0, 1]$ — defined in the Appendix — is higher when the adjustment cost is very concave (γ_2 is large) and P_1^e is essentially the same expression for the price that would prevail in an exchange economy, which is given in Equation (6). Intuitively, the equilibrium price is a weighted average of the exchange economy price and γ_1, which is the replacement cost of the risky asset with linear adjustments costs.

Allowing for production does not affect those positive properties of the model that matter for the determination of the optimal tax. An increase in the transaction tax can increase, reduce or keep equilibrium prices (and investment) constant, but all buyers buy less and all sellers sell less.

Normative results Accounting for producers, social welfare is now defined, using indirect utilities, as:

$$V(\tau) = \int \lambda_i V_i dF(i) + \lambda_k V_k,$$

where V_k and λ_k respectively denote the indirect utility and the welfare weight of producers. The change in producers' welfare induced by a marginal change in the tax is given by:

$$\frac{dV_k}{d\tau} = U_k'(C_{1k}) \left[\frac{dP_1}{d\tau} S_{1k} + [P_1 - \Phi'(S_{1k})] \frac{dS_{1k}}{d\tau} - \frac{dY_k}{d\tau} \right] + \mathbb{E}[U_k'(C_{2k})] \frac{dY_k}{d\tau} = \mathbb{E}[U_k'(C_{2k})] \frac{dP_1}{d\tau} S_{1k},$$

where the second line follows by substituting producers' optimality conditions. Intuitively, because producers do not pay taxes and invest optimally given prices, a marginal tax change only modifies their welfare through the distributive price effects on the shares they issue/repurchase. When P_1 is high, producers enjoy a better deal selling shares than when P_1 is low. The envelope theorem eliminates from $\frac{dV_k}{d\tau}$ the direct effects caused by changes in producers portfolio or investment choices.

Proposition 10 assumes that the planner accounts for producers' certainty equivalents, and characterizes the optimal tax.

Proposition 10. (Optimal tax in production economies) *The optimal tax in a production economy is given by*:

$$\tau^* = \frac{\int \left(\frac{\mathbb{E}[D]-\mathbb{E}_i[D]}{P_1}\right) \frac{dX_{1i}}{d\tau} dF(i)}{-\int \operatorname{sgn}(\Delta X_{1i}) \frac{dX_{1i}}{d\tau} dF(i)} = (1-\omega)\,\tau^*_{exchange} + \omega \tau^*_{production}, \tag{40}$$

where $\tau^*_{exchange} = \frac{\zeta(\tau)\mathbb{C}ov_{F,\mathcal{T}}\left[\frac{\mathbb{E}_i[D]}{P_1}, \frac{dX_{1i}}{d\tau}\right]}{2\int_{i\in\mathcal{B}} \frac{dX_{1i}}{d\tau} dF(i)}$, $\tau^*_{production} = \frac{\mathbb{E}[D]-\mathbb{E}_{F,\mathcal{T}}[\mathbb{E}_i[D]]}{P_1}$ *and* $\omega < 1$ *is given in the Appendix* (*ω is small in magnitude when* $\frac{dS_{1k}}{d\tau} \approx 0$ *and close to unity when* $\left|\frac{dS_{1k}}{d\tau}\right|$ *is large*). $\mathbb{E}_{F,\mathcal{T}}[\mathbb{E}_i[D]]$ *denotes the average belief in the population of active investors,* $\mathbb{C}ov_{F,\mathcal{T}}\left[\mathbb{E}_i[D], \frac{dX_{1i}}{d\tau}\right]$ *denotes a cross-sectional covariance among active investors and* $\zeta(\tau) \equiv \int_{i\in\mathcal{T}} dF(i)$ *is the fraction of active investors.*

The optimal tax can be expressed as a linear combination between the optimal tax in a (fictitious) exchange economy and the optimal tax in a (fictitious) production economy with a single investor with belief $\mathbb{E}_{F,\mathcal{T}}[\mathbb{E}_i[D]]$. The sensitivity of investment with respect to a tax change determines the relative importance of each term.

Market clearing now implies that $\int \frac{dX_{1i}}{d\tau} dF(i) = \frac{dS_{1k}}{d\tau}$, which can take any positive or negative value. Hence, in production economies, the belief used by the planner to calculate welfare matters in general for the optimal policy. However, if the planner uses investors' average belief to calculate welfare, the belief used by the planner drops out of the optimal tax expression. Because of its importance, I state this result as a corollary of Proposition 10.

Corollary. (τ^* may depend on the planner's belief) *The optimal financial transaction tax in a production economy depends on the distribution of payoffs assumed by the planner. However, if the planner uses the average belief across investors, that is,* $\mathbb{E}[D] = \mathbb{E}_{F,\mathcal{T}}[\mathbb{E}_i[D]]$ *at the optimum, the optimal tax is identical to the one in the exchange economy and independent of the belief used by the planner.*[53]

The numerator in Equation (40), which evaluated at $\tau = 0$ determines the sign of the optimal

[53] The average belief may change if there are changes in the composition of marginal investors. For the irrelevance result to hold without further qualifications, the average belief for marginal investors must be invariant to the level of τ.

tax can be decomposed in two terms:

$$\int \left(\mathbb{E}\left[D\right]-\mathbb{E}_i\left[D\right]\right)\frac{dX_{1i}}{d\tau}dF\left(i\right)=-\zeta\left(\tau\right)\underbrace{\mathbb{C}ov_{F,\mathcal{T}}\left[\mathbb{E}_i\left[D\right],\frac{dX_{1i}}{d\tau}\right]}_{\text{Belief dispersion}}+\underbrace{\left(\mathbb{E}\left[D\right]-\mathbb{E}_{F,\mathcal{T}}\left[\mathbb{E}_i\left[D\right]\right]\right)\frac{dS_{1k}}{d\tau}}_{\substack{\text{Aggregate belief difference}\times\\ \text{Investment response}}},\tag{41}$$

Because the second term in (41) is in general non-zero when $\tau = 0$, we can say that belief distortions in production economies have an additional first-order effect on welfare. Again asset prices do not appear in optimal tax formulas, despite playing a role in determining allocations. All welfare losses must be traced back to distortions in "quantities", either in portfolio allocations, captured by $\frac{dX_{1i}}{d\tau}$, or in production decisions, captured by $\frac{dS_{1k}}{d\tau}$.

Intuitively, the optimal tax corrects two wedges created by heterogeneous beliefs. First, given an amount of aggregate risk, the optimal tax seeks to reduce the asset holding dispersion induced by disagreement — some investors are holding too much risk and some others too little risk. This is the same mechanism present in exchange economies. Second, as long as the average belief differs from the one used by the planner, the level of production in the economy is too high (low) when investors are on average too optimistic (pessimistic). This provides a second rationale for taxation. Intuitively, the investors in the economy hold too much aggregate risk when they are on average optimistic or too little when they are pessimistic.[54]

Sign of $\omega\tau^*_{\textbf{production}}$ Belief dispersion is not sufficient anymore to pin down the sign of the optimal tax, which now also depends on whether $\mathbb{E}\left[D\right]-\mathbb{E}_F\left[\mathbb{E}_i\left[D\right]\right]$ and $\frac{dS_{1k}}{d\tau}$ have the same or opposite signs. Intuitively, if a marginal tax increase reduces (increases) investment at the margin when investors are too optimistic (pessimistic), a positive tax is welfare improving, and vice versa. Table 5 summarizes the conditions that determine the sign of the term associated to production.

	Aggregate optimism	Aggregate pessimism
	$\mathbb{E}_F\left[\mathbb{E}_i\left[D\right]\right]>\mathbb{E}\left[D\right]$	$\mathbb{E}_F\left[\mathbb{E}_i\left[D\right]\right]<\mathbb{E}\left[D\right]$
$\int_{\mathcal{B}}\frac{1}{A_i}dF\left(i\right)>\int_{\mathcal{S}}\frac{1}{A_i}dF\left(i\right)$	$\omega\tau^*_{\text{production}}>0$	$\omega\tau^*_{\text{production}}<0$
$\int_{\mathcal{B}}\frac{1}{A_i}dF\left(i\right)<\int_{\mathcal{S}}\frac{1}{A_i}dF\left(i\right)$	$\omega\tau^*_{\text{production}}<0$	$\omega\tau^*_{\text{production}}>0$

Table 1: Sign of $\omega\tau^*_{\text{production}}$

Unlike in the exchange economy, in which the orthogonality of beliefs justifies that the belief dispersion term is negative, it is not obvious whether we should expect $\omega\tau^*_{\text{production}}$ to be positive or negative. For investment to be reduced (increased) at the margin by a tax increase, it has to be the case that the (risk aversion adjusted) mass of buyers is larger (smaller) than the mass of

[54]If there were many produced risky assets, the welfare losses would capture the idea that belief distortions misallocate real investment across sectors in the economy. These results are available under request.

sellers. In principle, the relation between the average belief distortion and the relative mass of buyers/sellers need not be linked, so the sign of $\omega\tau^*_{\text{production}}$ is theoretically ambiguous.[55]

Many informal discussions regarding the convenience of a transaction tax, following Tobin (1978) revolve around the notion that it would help reduce price volatility . Implicit in those discussions is the notion that high volatility is bad. The results in this section show that it is not price volatility, a variance, but whether investment (through prices) is lower when investors are optimistic and vice versa, a covariance, what captures the welfare consequences of a transaction tax in a production context.

C.6 Tax on the number of shares

I assume in the baseline model that the tax is levied on the dollar value of a trade rather than on the number of shares traded to prevent investors from circumvent it by varying the effective number of shares traded — through a reverse split. All results apply to taxes that depend on the number of shares with minor modifications.

When P_1 is exactly zero, a tax based on the dollar volume of the transaction is ineffective. However, a tax based on the number of shares traded $|\Delta X_{1i}|$ can be introduced to effectively tax the notional value of the contract. I extend here Proposition 1 to the case of taxes levied on the number of shares traded. In this case, the distinction between buyers and sellers is somewhat arbitrary, giving support to the idea that both sides of the market should face the same tax.

In the trade region, the optimal portfolio choice of an investor can be expressed as: $X_{1i} = \frac{\mathbb{E}_i[D]-A_i\mathbb{C}ov[E_{2i},D]-P_1-\text{sgn}(\Delta X_{1i})\tau}{A_i\mathbb{V}ar[D]}$. The equilibrium price becomes:

$$P_1 = \int_{i\in\mathcal{T}} \left(\frac{\mathbb{E}_i[D]}{\mathcal{A}_i} - A\left(\mathbb{C}ov\left[E_{2i}, D\right] - \mathbb{V}ar\left[D\right] X_{0i}\right) - \frac{\text{sgn}\left(\Delta X_{1i}\right)}{\mathcal{A}_i}\tau \right) dF\left(i\right).$$

The price correction is now additive rather than multiplicative. The value of $\frac{dV}{d\tau}$ corresponds to $\frac{dV}{d\tau} = \int \left[-\mathbb{E}_i\left[D\right] + \text{sgn}\left(\Delta X_{1i}\right)\tau\right] \frac{dX_{1i}}{d\tau} dF\left(i\right)$. The optimal tax now satisfies:

$$\tau^* = \frac{\int \mathbb{E}_i\left[D\right] \frac{dX_{1i}}{d\tau} dF\left(i\right)}{\int \text{sgn}\left(\Delta X_{1i}\right) \frac{dX_{1i}}{d\tau} dF\left(i\right)}.$$

This shows that optimal taxes in the paper are written in terms of returns because they are levied on the dollar value of the transaction. When they are levied on the number of shares, the dispersion in expected payoffs rather than the dispersion in expected returns becomes the welfare relevant variable.

C.7 Harberger calculation

The results derived so far rely on the assumption that the planner maximizes welfare using a single belief. However, it is straightforward to quantify the welfare loss induced by a tax increase

[55] Additional policy instruments, like short-sale of borrowing constraints, investment taxes, or active monetary policy can be used to target the production distortion induced by beliefs, allowing the transaction tax to be exclusively focused again on belief dispersion.

assuming that all investors hold correct beliefs or that the planner assess social welfare respecting individual beliefs. Under either of these assumptions, all trades are regarded as fundamental, so any tax induces a welfare loss. I derive a result analogous to Harberger (1964), whose triangle analysis can be traced back to Dupuit (1844).

Proposition 11. (Harberger (1964) revisited)

a) When investors hold identical beliefs or the planner respects individual beliefs when calculating social welfare, the marginal welfare loss generated by increasing the transaction tax at a level $\tilde{\tau}$, expressed as a money-metric (in dollars) at $t = 1$, is given by:

$$\int \left.\frac{d\hat{V}_i}{d\tau}\right|_{\tau=\tilde{\tau}} dF(i) = 2\tilde{\tau} P_1 \int_{i\in\mathcal{B}} \left.\frac{dX_{1i}}{d\tau}\right|_{\tau=\tilde{\tau}} dF(i) \leq 0, \tag{42}$$

where $i \in \mathcal{B}$ denotes that the integration is made only over the set of buyers and $\hat{V}_i$ denotes investors' certainty equivalents.

b) The marginal welfare loss of a small tax change around $\tau = 0$ can be approximated, using a second order Taylor expansion, by:

$$dV = \int d\hat{V}_i\Big|_{\tau=0} dF(i) \approx \tau^2 P_1 \int_{i\in\mathcal{B}} \left.\frac{dX_{1i}}{d\tau}\right|_{\tau=0} dF(i)$$

This result provides a measure of welfare losses as a function of observables for any tax intervention. Given the money-metric correction, investors in this economy are willing to pay $\mathcal{L}(\tau)$ dollars to prevent a change in the tax rate. Note that this happens to correspond to the marginal change in revenue raised. Equation (42) derives an upper bound for the size of the welfare losses induced by taxation in the case in which all trades are deemed to be fundamental.

Equation (42) resembles the classic Harberger (1964) result about welfare losses in the context of commodity taxation.[56] However, the welfare loss in this case is given by twice the size of the tax, because the portfolio holdings of both buyers and sellers are distorted. Taxing a commodity distorts the amount consumed of a given good, reducing welfare. Taxing financial transactions distorts portfolio allocations, inducing investors to hold more or less risk than they should, also reducing welfare. The distortion created by a tax (approximately) grows with the square in this context of the model studied in this paper.

An older version of the paper analyzed the effects of transaction taxes in a general environment with incomplete markets but without heterogeneous beliefs. In that environment, when markets are complete, a transaction tax is always welfare reducing. However, when markets are incomplete, the economy is constrained inefficient and it may be the case that a transaction tax or subsidy improve welfare. Social welfare gains or losses are driven by pecuniary effects (the induced effects by price responses to a tax change $\frac{dP}{d\tau}$). My results show that it is not consumption volatility,

[56] Although this result is intuitive, to my knowledge, it had not been derived before in the context of a portfolio choice problem. See Auerbach and Hines Jr (2002) for a comprehensive analysis of tax efficiency results and Sandmo (1985) for a survey of results on how taxation affects portfolio allocations.

but the appropriate covariance of price changes with investors' stochastic discount factors what determines the exact effect of transaction taxes on welfare. These results can be interpreted as an extension of the Harberger results to the incomplete market case.

C.8 Linear combination between planner's belief and investors' beliefs

Throughout the paper, I assume that the planner maximizes welfare using a single distribution of payoffs for all investors. It is straightforward to generalize the results to a planner that puts weight α on his own belief and weight $1-\alpha$ on the belief of each investor. In that case, the new optimal tax τ_α^* looks turns out to be a linear combination of both taxes. In the baseline model, because the optimal tax for a planner that respects investors' beliefs is $\tau^*=0$, the optimal tax becomes:

$$\tau_\alpha^* = \alpha\tau^*,$$

where τ^* is given by Equation (8). The same logic applies to other extensions of the baseline model. The case with $\alpha=1$ is the leading case analyzed in the paper. This approach should be appealing to readers who prefer to have a partially paternalistic planner.

C.9 Disagreement about other moments

Assumptions Motivated by Proposition 4, in the baseline model, investors only disagree about the expected value of the payoff of the risky asset. I now assume that investors also hold distorted beliefs about their hedging needs $\mathbb{C}ov_i\left[E_{2i}, D\right]$ and about the variance of the payoff of the risky asset $\mathbb{V}ar_i\left[D\right]$.

Results The optimality condition presented in (5) applies directly, after using the individual beliefs of each investor. Hedging needs enter additively, but perceived individual variances modify the sensitivity of portfolio demands with respect to the baseline case.

Market clearing determines the equilibrium price, given now by:

$$P_1 = \frac{\int_{i\in\mathcal{T}}\left(\frac{\mathbb{E}_i[D]}{\mathcal{AV}_i} - AV\left(\beta_{ii} + X_{0i}\right)\right)dF\left(i\right)}{1+\tau\int_{i\in\mathcal{T}}\frac{\operatorname{sgn}(\Delta X_{1i})}{\mathcal{AV}_i}dF\left(i\right)},$$

where $AV \equiv \left(\int_{i\in\mathcal{T}}\frac{1}{A_i\mathbb{V}ar_i[D]}dF\left(i\right)\right)^{-1}$ is the harmonic mean of risk aversion coefficients and perceived variances for active investors and $\mathcal{AV}_i \equiv \frac{A_i\mathbb{V}ar_i[D]}{AV}$ is the quotient between investor i risk aversion times perceived variance and the harmonic mean. I define the regression coefficient (beta) of individual endowments E_{2i} on payoffs D perceived by investors by $\beta_{ii} = \frac{\mathbb{C}ov_i[E_{2i},D]}{\mathbb{V}ar_i[D]}$. Again, $\mathcal{T}$ denotes the set of active investors.

Proposition 12. (Disagreement about second moments)

a) The marginal change in social welfare from varying the financial transaction tax when investors disagree about second moments is given by:

$$\frac{dV}{d\tau}=\int\left[\left(-r_i\mathbb{E}_i\left[D\right]-A_i\mathbb{C}ov\left[E_{2i},D\right]\left(1-\frac{\beta_{ii}}{\beta_i}\right)+P_1r_i\left(1+\operatorname{sgn}\left(\Delta X_{1i}\right)\tau\right)\right)\frac{dX_{1i}}{d\tau}\right]dF\left(i\right), \tag{43}$$

where $r_i\equiv\frac{\mathbb{V}ar[D]}{\mathbb{V}ar_i[D]}$, $\beta_{ii}\equiv\frac{\mathbb{C}ov_i[E_{2i},D]}{\mathbb{V}ar_i[D]}$ *and* $\beta_i\equiv\frac{\mathbb{C}ov[E_{2i},D]}{\mathbb{V}ar[D]}$*. Note that* $r_i\in(0,\infty)$ *and* $\beta_i,\beta_{ii}\in(-\infty,\infty)$*.*

b) The optimal tax when investors disagree about second moments is given by:

$$\tau^*=\frac{\int\left(r_i\mathbb{E}_i\left[D\right]+A_i\mathbb{C}ov\left[E_{2i},D\right]\left(1-\frac{\beta_{ii}}{\beta_i}\right)\right)\frac{dX_{1i}}{d\tau}dF\left(i\right)}{P_1\int\left(r_i\left(1+\operatorname{sgn}\left(\Delta X_{1i}\right)\right)\right)\frac{dX_{1i}}{d\tau}dF\left(i\right)}.$$

The formula for the optimal tax now incorporates hedging needs and modifies the weights given to investors' beliefs. An investor with correct beliefs about second moments has $r_i=1$ and $\beta_{ii}=\beta_i$; in that case, we recover (8). When investors perceive a high variance, that is, r_i is close to 0, they receive less weight in the optimal tax formula. The opposite occurs when they perceive a low variance. Intuitively, lower perceived variances amplify distortions in expected payoffs, and vice versa.

As in the baseline model, the planner does not need to know the value of $\mathbb{E}\left[D\right]$ to implement the optimal tax. However, if investors hold distorted beliefs about their hedging needs, the planner needs to know explicitly the magnitude of the mistake. Intuitively, there is no mechanism in the model which cancels out the mistakes in hedging made by investors. The sign of the optimal tax depends directly on the errors made by investors when hedging.

There are two interesting parameters restrictions. First, when investors with correct expected payoffs and hedging betas, that is $\frac{\beta_{ii}}{\beta_i}=1$, disagree about variances, the optimal tax τ^* turns out to be:

$$\tau^*=\frac{\mathbb{E}\left[D\right]\int r_i\frac{dX_{1i}}{d\tau}dF\left(i\right)}{P_1\int r_i\left(1+\operatorname{sgn}\left(\Delta X_{1i}\right)\right)\frac{dX_{1i}}{d\tau}dF\left(i\right)}.$$

The dispersion of variances, given by $\mathbb{C}ov_{F,\mathcal{T}}\left[r_i,\frac{dX_{1i}}{d\tau}\right]$, determines now the sign of the optimal tax. When r_i is constant (although not necessarily equal to one), the optimal tax becomes zero. This reinforces the intuition that belief dispersion is what matters for optimal taxes in an exchange economy. Intuitively, when buyers, with $\frac{dX_{1i}}{d\tau}<0$, are relatively aggressive, that is, r_i is large, they are buying too much of the risky asset, so $\mathbb{C}ov_{F,\mathcal{T}}\left[r_i,\frac{dX_{1i}}{d\tau}\right]$ is negative and the optimal tax is positive, and vice versa.

Second, when investors have correct beliefs about the mean and the variance of expected returns, but hedge incorrectly, the optimal tax becomes:

$$\tau^*=\frac{\mathbb{V}ar\left[D\right]\int A_i\left(\beta_i-\beta_{ii}\right)\frac{dX_{1i}}{d\tau}dF\left(i\right)}{P_1\int\operatorname{sgn}\left(\Delta X_{1i}\right)\frac{dX_{1i}}{d\tau}dF\left(i\right)}.$$

The optimal tax now has the opposite sign of $\mathbb{C}ov_{F,\mathcal{T}}\left[A_i\left(\beta_i-\beta_{ii}\right),\frac{dX_{1i}}{d\tau}\right]$. Intuitively, when buyers, with $\frac{dX_{1i}}{d\tau}<0$, overestimate their need for hedging and end up buying too much of the risky asset — this occurs when $\beta_i-\beta_{ii}<0$ — the optimal tax is positive, and vice versa.

C.10 Portfolio constraints: short-sale and borrowing constraints

Assumptions Although participants in financial markets face short-sale and borrowing constraints, investors in the baseline model face no restrictions when choosing portfolios. I now introduce trading constraints into the model as a pair of functions $\overline{g}_i(\cdot)$ and $\underline{g}_i(\cdot)$ for every investor i, which can depend on equilibrium prices,[57] such that:

$$\underline{g}_i(P_1) \leq X_{1i} \leq \overline{g}_i(P_1) \tag{44}$$

Both short-sale constraints and borrowing constraints are special cases of (44). Short-sale constraints are in general price independent and can be expressed as $X_{1i} \geq 0$. Borrowing constraints can be modeled by choosing $\overline{g}_i(P_1)$ appropriately, such that $X_{1i} \leq \overline{g}_i(P_1)$. Intuitively, an investor who wants to sufficiently increase his holdings of the risky asset must rely on borrowing. Hence, a borrowing limit is equivalent to an upper bound constraining the amount held of the risky asset.

Results The optimal portfolio is identical to the one in the baseline model, unless a constraint binds. In that case, X_{1i} equals the trading limit. The equilibrium price is a slightly modified version of (6).

Proposition 13. (Trading constraints) *The optimal tax when investors face trading constraints is given by:*

$$\tau^* = \frac{\int \mathbb{E}_i[D] \frac{dX_{1i}}{d\tau} dF(i) - \int_{i=\mathcal{C}} A_i \mathbb{V}ar[D] \left(\hat{X}_{1i} - g_i(P_1)\right) g_i'(P_1) \frac{dP_1}{d\tau} dF(i)}{P_1 \int \operatorname{sgn}(\Delta X_{1i}) \frac{dX_{1i}}{d\tau} dF(i)},$$

where $i = \mathcal{C}$ denotes the set of investors with binding trading constraints and $\tilde{X}_{1i}$ denotes the optimal unconstrained portfolio holding for a constrained investor, given in the Appendix. Note that I have used the fact that $\frac{dX_{1i}}{d\tau} = g_i'(P_1) \frac{dP_1}{d\tau}$ for constrained investors.

When trading constraints do not depend on prices that is, $g_i'(P_1) = 0$, the optimal tax formula is identical to the one of the baseline model. In those cases, changes in taxes do not modify the portfolio allocation of constrained investors, leaving their welfare unchanged, i.e., for those investors $\frac{dX_{1i}}{d\tau} = 0$. Intuitively, investors with price independent trading constraints are inframarginal for price determination.

When trading constraints depend on prices, the optimal policy takes these effects into account. A marginal tax change modifies asset prices and consequently portfolio allocations for constrained investors; this portfolio change has a first-order effect on welfare. The size of the correction has three determinants. First, it depends on how far the actual portfolio allocation is from the unconstrained portfolio allocation, given by how much the constrained allocation $g_i(P_1)$ differs from the optimal unconstrained allocation $\hat{X}_{1i}$. Second, it depends on how sensitive the equilibrium restriction is with respect to asset prices — this is captured by $g_i'(P_1)$. Third, it

[57] I assume that investors' choice sets remain convex after imposing (44). By $g_i(\cdot)$, I denote either $\overline{g}_i(\cdot)$ or $\underline{g}_i(\cdot)$.

depends on how equilibrium prices react to tax changes $\frac{dP_1}{d\tau}$. If prices remain constant after varying τ, that is $\frac{dP_1}{d\tau}=0$, the optimal tax formula does not change.

D Extensions: Proofs and derivations

Proposition 5. (Pre-existing trading costs)

Given investors' optimal portfolios, stated in the main text, it is straightforward to derive the equilibrium price, which is given by:

$$P_1 = \frac{\int_{i\in\mathcal{T}} \frac{\mathbb{E}_i[D]-A_i\mathbb{C}ov[E_{2i},D]-\operatorname{sgn}(\Delta X_{1i})\eta+\alpha X_{0i}}{A_i\mathbb{V}ar[D]+\alpha} - \int_{i\in\mathcal{T}} X_{0i}}{\int_{i\in\mathcal{T}} \frac{(1+\operatorname{sgn}(\Delta X_{1i})(\tau+\psi))}{A_i\mathbb{V}ar[D]+\alpha}}.$$

Indirect utility for an investor i from the planner's perspective is given by:

$$V_i = -e^{-A_i\left((\mathbb{E}[D]-A_i\mathbb{C}ov[E_{2i},D]-P_1)X_{1i}+P_1X_{0i}-|\Delta X_{1i}|P_1\psi-\frac{\alpha}{2}(\Delta X_{1i})^2-\frac{A_i}{2}\mathbb{V}ar[D](X_{1i})^2\right)}.$$

Note that only the resources corresponding to the transaction tax are rebated back to investors. All resources devoted to transaction costs are a compensation for the use of resources, so the planner does not have to account for them explicitly, since they form part of a zero profit condition.

Hence, the marginal change in welfare for an investor i is given by:

$$\frac{dV_i}{d\tau} = \mathbb{E}\left[U_i'(W_{2i})\right]\left[\begin{array}{c}(\mathbb{E}[D]-A_i\mathbb{C}ov[E_{2i},D]-P_1-\operatorname{sgn}(\Delta X_{1i})P_1\psi-\eta\operatorname{sgn}(\Delta X_{1i}))\frac{dX_{1i}}{d\tau} \\ -(\alpha\Delta X_{1i}+A\mathbb{V}ar[D]X_{1i})\frac{dX_{1i}}{d\tau}-\Delta X_{1i}\frac{dP_1}{d\tau}\end{array}\right].$$

By substituting investors' first order conditions, we find:

$$\frac{dV_i}{d\tau} = \mathbb{E}\left[U_i'(W_{2i})\right]\left[\left[\mathbb{E}[D]-\mathbb{E}_i[D]+\operatorname{sgn}(\Delta X_{1i})P_1\tau\right]\frac{dX_{1i}}{d\tau}-\Delta X_{1i}\frac{dP_1}{d\tau}\right].$$

It follows that the optimal tax has the same expression as in Proposition 1.

Proposition 6. (Imperfect tax enforcement)

After eliminating terms that do not affect the maximization problem, investors solve:

$$\max_{X_{1i}^A,X_{1i}^B}\left[\mathbb{E}_i[D]-A_i\mathbb{C}ov[E_{2i},D]-P_1\right]\left(X_{1i}^A+X_{1i}^B\right)-\tau|P_1|\left|\Delta X_{1i}^A\right|-\frac{A_i}{2}\mathbb{V}ar[D]\left(X_{1i}^A+X_{1i}^B\right)^2-\frac{\alpha}{2}\left(\Delta X_{1i}^B\right)^2,$$

with interior optimality conditions shown in the text.

The change in investors' certainty equivalents is given by

$$\frac{d\hat{V}_i}{d\tau} = \left[\mathbb{E}[D]-\mathbb{E}_i[D]\right]\frac{dX_{1i}}{d\tau}+\operatorname{sgn}(\Delta X_{1i})P_1\tau\frac{dX_{1i}^A}{d\tau}-\Delta X_{1i}^A\frac{dP_1^A}{d\tau}-\Delta X_{1i}^B\frac{dP_1^B}{d\tau},$$

where the last two terms are zero under symmetry. Equations (34) and (35) follow immediately.

Proposition 7. (Multiple risky assets)

After eliminating terms that do not affect the maximization problem, investors solve:

$$\max_{\mathbf{x}_{1i}} \mathbf{x}_{1i}' \left(\mathbb{E}_i\left[\mathbf{d}\right] - A_i \mathbb{C}ov\left[E_{2i}, \mathbf{d}\right] - \mathbf{p}\right) - \left|\mathbf{x}_{1i}' - \mathbf{x}_{0i}'\right| \mathbf{p}\tau - \frac{A_i}{2} \mathbf{x}_{1i}' \mathbb{V}ar\left[\mathbf{d}\right] \mathbf{x}_{1i}.$$

Where I use $\left|\mathbf{x}_{1i}' - \mathbf{x}_{0i}'\right|$ to denote the vector of absolute values of the difference between both vectors. This problem is convex, so the first order condition fully characterizes investors' optimal portfolios as long as they trade a given asset j:

$$\mathbf{x}_{1i} = \left(A_i \mathbb{V}ar\left[\mathbf{d}\right]\right)^{-1} \left(\mathbb{E}_i\left[\mathbf{d}\right] - A_i \mathbb{C}ov\left[E_{2i}, \mathbf{d}\right] - \mathbf{p} - \hat{\mathbf{p}}_i \tau\right),$$

where $\hat{\mathbf{p}}_i$ is a $J \times 1$ vector where a given row j is given by $\operatorname{sgn}\left(\Delta X_{1ij}\right) p_j$. If an asset j is not traded by an investor i, then $X_{1ij} = X_{0ij}$. The inaction regions are defined analogously to the one asset case. Note that there exists a way to write optimal portfolio choices only with matrix operations; however, the notation turns out to be more cumbersome. The equilibrium price vector is given by:

$$\mathbf{p} \hat{\times} \int \left(1 + \frac{\mathbf{s}_i}{A_i}\tau\right) dF\left(i\right) = \int \frac{\mathbb{E}_i\left[\mathbf{d}\right]}{A_i} dF\left(i\right) - \int \left(\mathbb{C}ov\left[E_{2i}, \mathbf{d}\right] + \mathbb{V}ar\left[\mathbf{d}\right] \mathbf{x}_{0i}\right) dF\left(i\right).$$

Where I denote element-by-element multiplication as $y \hat{\times} z$ and use $\mathbf{s}_i$ to denote a $J \times 1$ vector given by $\operatorname{sgn}\left(\Delta X_{1ij}\right)$.

$$\frac{dV}{d\tau} = \int \left(\mathbb{E}\left[\mathbf{d}\right] - \mathbb{E}_i\left[\mathbf{d}\right] + \hat{\mathbf{p}}_i \tau\right)' \frac{d\mathbf{x}_{1i}}{d\tau} dF\left(i\right).$$

The marginal effect of varying taxes in social welfare is given by:

$$\frac{dV}{d\tau} = \int \lambda_i \mathbb{E}\left[U_i'\left(W_{2i}\right)\right] \left[\left(\mathbb{E}\left[\mathbf{d}\right] - \mathbb{E}_i\left[\mathbf{d}\right] + \hat{\mathbf{p}}_i \tau\right)' \frac{d\mathbf{x}_{1i}}{d\tau} - \left(\mathbf{x}_{1i} - \mathbf{x}_{0i}\right)' \frac{d\mathbf{p}}{d\tau}\right] dF\left(i\right).$$

This is a generalization of the one asset case. We can write in product notation:

$$\int \sum_{j=1}^{J} \left(-\mathbb{E}_i\left[D_j\right] + \operatorname{sgn}\left(\Delta X_{1ij}\right) p_j \tau\right) \frac{dX_{1ij}}{d\tau} dF\left(i\right) = 0.$$

So the optimal tax becomes:

$$\tau^* = \frac{\sum_{j=1}^{J} \int \mathbb{E}_i\left[D_j\right] \frac{dX_{1ij}}{d\tau} dF\left(i\right)}{\sum_{j=1}^{J} \int \operatorname{sgn}\left(\Delta X_{1ij}\right) p_j \frac{dX_{1ij}}{d\tau} dF\left(i\right)}.$$

Which can be rewritten as:

$$\tau^* = \frac{\sum_{j=1}^{J} \int \operatorname{sgn}\left(\Delta X_{1ij}\right) p_j \frac{dX_{1ij}}{d\tau} dF\left(i\right) \tau_j^*}{\sum_{j=1}^{J} \int \operatorname{sgn}\left(\Delta X_{1ij}\right) p_j \frac{dX_{1ij}}{d\tau} dF\left(i\right)}.$$

Where $\tau_j^* = \frac{\int \frac{\mathbb{E}_i\left[D_j\right]}{p_j} \frac{dX_{1ij}}{d\tau} dF(i)}{\int \operatorname{sgn}(\Delta X_{1ij}) \frac{dX_{1ij}}{d\tau} dF(i)}$. And by defining weights $\omega_j = \frac{\int \operatorname{sgn}(\Delta X_{1ij}) p_j \frac{dX_{1ij}}{d\tau} dF(i)}{\sum_{j=1}^{J} \int \operatorname{sgn}(\Delta X_{1ij}) p_j \frac{dX_{1ij}}{d\tau} dF(i)}$, we recover Equation (37).

Proposition 8. (Asymmetric taxes on buyers versus sellers)

The budget constraint for an investor in this case can be expressed as:

$$W_{2i} = E_{2i} + X_{1i}D + \left(X_{0i}P_1 - X_{1i}P_1 - \tau_B \left|P_1\right| \left|\Delta X_{1i}\right|_+ - \tau_S \left|P_1\right| \left|\Delta X_{1i}\right|_- + T_{1i}\right).$$

The first order condition becomes:

$$X_{1i} = \frac{\mathbb{E}_i \left[D\right] - A_i \mathbb{C}ov \left[E_{2i}, D\right] - P_1 \left(1 + \mathbb{I}\left[\Delta X_{1i} > 0\right]\tau_B + \mathbb{I}\left[\Delta X_{1i} < 0\right]\tau_S\right)}{A_i \mathbb{V}ar \left[D\right]}.$$

With an equilibrium price:

$$P_1 = \frac{\int_{i\in\mathcal{T}} \left(\frac{\mathbb{E}_i[D] - A_i \mathbb{C}ov[E_{2i}, D]}{A_i} - \mathbb{V}ar \left[D\right] X_{0i}\right) dF\left(i\right)}{\int_{i\in\mathcal{T}} \frac{1}{A_i} + \tau_B \int_{i\in\mathcal{B}} \frac{1}{A_i} - \tau_S \int_{i\in\mathcal{S}} \frac{1}{A_i} dF\left(i\right)}.$$

In this case we can write: $X_{1i}\left(\tau_i, P_1\left(\{\tau_j\}\right)\right)$, where $\{\tau_j\}$ denotes a vector of taxes. This implies that $\frac{dX_{1i}}{d\tau_j} = \frac{\partial X_{1i}}{\partial \tau_j} + \frac{\partial X_{1i}}{\partial P_1}\frac{dP_1}{d\tau_j}$. The change in social welfare for an investor i when varying a tax τ_j, from a planner's perspective, is given by:

$$\begin{aligned}\frac{dV_i}{d\tau_j} &= \mathbb{E}\left[U_i'\left(W_{2i}\right)\right]\left[\left(\mathbb{E}\left[D\right] - A_i\mathbb{C}ov\left[E_{2i}, D\right] - P_1 - A_i X_{1i}\mathbb{V}ar\left[D\right]\right)\frac{dX_{1i}}{d\tau_j} - \Delta X_{1i}\frac{dP_1}{d\tau_j}\right]\\ &= \mathbb{E}\left[U_i'\left(W_{2i}\right)\right]\left(\left(\mathbb{E}\left[D\right] - \mathbb{E}_i\left[D\right] + P_1\left(\mathbb{I}\left[\Delta X_{1i} > 0\right]\tau_B - \mathbb{I}\left[\Delta X_{1i} < 0\right]\tau_S\right)\right)\frac{dX_{1i}}{d\tau_j} - \Delta X_{1i}\frac{dP_1}{d\tau_j}\right).\end{aligned}$$

Social welfare is then:

$$\frac{dV}{d\tau_j} = \int \lambda_i \mathbb{E}\left[U_i'\left(W_{2i}\right)\right]\left(\left(\mathbb{E}\left[D\right] - \mathbb{E}_i\left[D\right] + P_1\begin{pmatrix}\mathbb{I}\left[\Delta X_{1i} > 0\right]\tau_B\\ -\mathbb{I}\left[\Delta X_{1i} < 0\right]\tau_S\end{pmatrix}\right)\frac{dX_{1i}}{d\tau_j} - \Delta X_{1i}\frac{dP_1}{d\tau_j}\right)dF\left(i\right).$$

Any tax change has two direct effects. First, it marginally affects those investors who pay that tax at the margin. Second, it moves prices. This price change creates two effects. There is a first effect working through terms-of-trade. A second effect works through demand changes. Under the usual differentiability and convexity assumptions, the optimal tax is characterized by $\frac{dV}{d\tau_j} = 0$, $\forall j$. This yields a system of equation in the vector of taxes:

$$0 = \int \lambda_i \mathbb{E}\left[U_i'\left(W_{2i}\right)\right]\left(\left(\mathbb{E}\left[D\right] - \mathbb{E}_i\left[D\right] + P_1 \operatorname{sgn}\left(\Delta X_{1i}\right)\tau_i\right)\frac{dX_{1i}}{d\tau_j} - \Delta X_{1i}\frac{dP_1}{d\tau_j}\right)dF\left(i\right),\ \forall j$$

This equation characterizes a system of equations in τ_B and τ_S.

We can write:

$$\frac{dV}{d\tau_j} = \int\left(\mathbb{E}\left[D\right] - \mathbb{E}_i\left[D\right]\right)\frac{dX_{1i}}{d\tau_j}dF\left(i\right) + P_1\left(\tau_B\int_{i\in\mathcal{B}}\frac{dX_{1i}}{d\tau_j}dF\left(i\right) - \tau_S\int_{i\in\mathcal{S}}\frac{dX_{1i}}{d\tau_j}dF\left(i\right)\right).$$

Using market clearing, we can find:

$$\frac{dV}{d\tau_j} = -\int \mathbb{E}_i\left[D\right]\frac{dX_{1i}}{d\tau_j}dF\left(i\right) + P_1\left(\tau_B + \tau_S\right)\int_{i\in\mathcal{B}}\frac{dX_{1i}}{d\tau_j}dF\left(i\right).$$

Solving for $\tau_B + \tau_S$:

$$\tau_B + \tau_S = \frac{\int \mathbb{E}_i [D] \frac{dX_{1i}}{d\tau_j} dF(i)}{P_1 \int_{i \in \mathcal{B}} \frac{dX_{1i}}{d\tau_j} dF(i)}, \forall_j.$$

In general this gives a system of non-linear equations in $\tau_B + \tau_S$. When there are two investors, the two equations become collinear, because of market clearing:

$$\int \mathbb{E}_i [D] \frac{dX_{1i}}{d\tau_j} dF(i) = (\mathbb{E}_{\mathcal{B}} [D] - \mathbb{E}_S [D]) \int_{i \in \mathcal{B}} \frac{dX_{1i}}{d\tau_j} dF(i).$$

In that case only the sum of taxes is pinned down:

$$\tau_B + \tau_S = \frac{\mathbb{E}_{\mathcal{B}} [D] - \mathbb{E}_{\mathcal{S}} [D]}{P_1}.$$

Proposition 9. (Individual taxes/First-best)

a) In the case with I taxes and I investors, the first order conditions for the planner become:

$$\frac{dV}{d\tau_j} = \sum_i (-\mathbb{E}_i [D] + P_1 \operatorname{sgn}(\Delta X_{1i}) \tau_i) \frac{dX_{1i}}{d\tau_j} F(i) = 0, \ \forall j.$$

This system of equations characterizes the set of optimal taxes. Note that one solution to this system is given by:

$$-\mathbb{E}_i [D] + P_1 \operatorname{sgn}(\Delta X_{1i}) \tau_i = -F.$$

Where F is an arbitrary real number. Rearranging this expression we can find Equation (39).

b) Starting from the system of equations which characterizes the optimal set of taxes, we can write, using market clearing $F(j) \frac{dX_{1i}}{d\tau_j} + \sum_{i \neq j} \frac{dX_{1i}}{d\tau_j} F(i) = 0$, the following set of equations:

$$\sum_{i \neq j} (\mathbb{E}_j [D] - \mathbb{E}_i [D]) \frac{dX_{1i}}{d\tau_j} F(i) + P_1 \begin{pmatrix} -\operatorname{sgn}(\Delta X_{1j}) \tau_j \sum_{i \neq j} \frac{dX_{1i}}{d\tau_j} F(i) \\ + \sum_i \left(\operatorname{sgn}(\Delta X_{1i}) \tau_i \frac{dX_{1i}}{d\tau_j} \right) F(i) \end{pmatrix} = 0.$$

For all equations but for the one with respect to tax j. To show that this system only depends on $N-1$ taxes, we simply need to show that all $\frac{dX_{1i}}{d\tau_j}$ do not depend on the tax τ_j. Note that $\frac{dX_{1i}}{d\tau_j} = \frac{\partial X_{1i}}{\partial \tau_j} + \frac{\partial X_{1i}}{\partial P_1} \frac{dP_1}{d\tau_j}$. But when $i \neq j$ then $\frac{dX_{1i}}{d\tau_j}$ only depends on $\frac{dP_1}{d\tau_j}$ because $\frac{\partial X_{1i}}{\partial \tau_j}$ equals zero and $\frac{\partial X_{1i}}{\partial P_1}$ does not depend on τ_j. We just need to show that $\frac{dP_1}{d\tau_j}$ can be expressed as a function of all other taxes but τ_j. This can be easily shown combining the expressions used to show Lemma 1 with market clearing conditions.

Proposition 10. (Optimal tax with production)

The expression for the asset price in (45) now yields a demand curve for shares.

$$P_1^e = \frac{\int_{i \in \mathcal{T}} \left(\frac{\mathbb{E}_i[D]}{\mathcal{A}_i} - A \left(\mathbb{C}ov [E_{2i}, D] + \mathbb{V}ar [D] X_{0i} \right) \right) dF(i) - A \mathbb{V}ar [D] S_{1k}}{1 + \tau \int_{i \in \mathcal{T}} \frac{\operatorname{sgn}(\Delta X_{1i})}{\mathcal{A}_i} dF(i)} \tag{45}$$

The demand by investors for the risky asset is identical to the baseline model. The equilibrium price is now determined by the intersection of Equation (45) and the supply curve, given by $P_1^s = \gamma_1 + \gamma_2 S_{1k}$.

After writing the market clearing condition as $\int (X_{1i} - X_{0i})\, dF(i) = F(P_1)$, where $F(\cdot) = \Phi'^{-1}(\cdot)$ is an upward sloping function, we can derive $\frac{dP_1}{d\tau} = \frac{\int \frac{\partial X_{1i}}{\partial \tau} dF(i)}{F'(P_1) - \int \frac{\partial X_{1i}}{\partial P_1} dF(i)} = \frac{-P_1 \int \frac{\text{sgn}(\Delta X_{1i})}{A_i \mathbb{V}ar[D]} dF(i)}{F'(P_1) + \int \frac{(1+\text{sgn}(\Delta X_{1i})\tau)}{A_i \mathbb{V}ar[D]} dF(i)}$. $\frac{dP_1}{d\tau}$ can have any sign, depending on its numerator. We can write $\frac{dX_{1i}}{d\tau} = \frac{\partial X_{1i}}{\partial \tau} \varepsilon_i$, where ε_i, which is constant within buyers/sellers, can be expressed as $\varepsilon_i = 1 - (\text{sgn}(\Delta X_{1i}) + \tau) \frac{1-H}{\frac{\mathbb{V}ar[D]F'(P_1)}{\int_{\mathcal{B}} \frac{1}{A_i} dF(i)} + 1 + \tau + H(1-\tau)}$ and $H \equiv \frac{\int_{i\in\mathcal{S}} \frac{1}{A_i} dF(i)}{\int_{i\in\mathcal{B}} \frac{1}{A_i} dF(i)} \in (0,\infty)$. It is easy to show that $\varepsilon_i > 0$, which proves the result.

The marginal change in social welfare is given by:

$$\frac{dV}{d\tau} = \int \lambda_i \mathbb{E}\left[U_i'(W_{2i})\right] \left[(\mathbb{E}[D] - \mathbb{E}_i[D] + \text{sgn}(\Delta X_{1i}) P_1 \tau) \frac{dX_{1i}}{d\tau} - \Delta X_{1i} \frac{dP_1}{d\tau}\right] dF(i) + \lambda_k \mathbb{E}\left[U_k'(C_{2k})\right] \frac{dP_1}{d\tau} S_{1k}$$

Using market clearing, the marginal change in social welfare as:

$$\frac{dV}{d\tau} = \int (\mathbb{E}[D] - \mathbb{E}_i[D] + \text{sgn}(\Delta X_{1i}) P_1 \tau) \frac{dX_{1i}}{d\tau} dF(i)$$

Solving for τ^* in the previous expression yields Equation (40). We can re-write the numerator of the optimal tax as:

$$\begin{aligned}
\int (\mathbb{E}[D] - \mathbb{E}_i[D]) \frac{dX_{1i}}{d\tau} dF(i) &= \zeta(\tau) \mathbb{E}_{F,\mathcal{T}}\left[(\mathbb{E}[D] - \mathbb{E}_i[D]) \frac{dX_{1i}}{d\tau}\right] \\
&= \zeta(\tau) \left(\begin{array}{c} \mathbb{C}ov_{F,\mathcal{T}}\left[\mathbb{E}[D] - \mathbb{E}_i[D], \frac{dX_{1i}}{d\tau}\right] \\ + \mathbb{E}_{F,\mathcal{T}}\left[\mathbb{E}[D] - \mathbb{E}_i[D]\right] \mathbb{E}_{F,\mathcal{T}}\left[\frac{dX_{1i}}{d\tau}\right] \end{array} \right) \\
&= -\zeta(\tau) \mathbb{C}ov_{F,\mathcal{T}}\left[\mathbb{E}_i[D], \frac{dX_{1i}}{d\tau}\right] + (\mathbb{E}[D] - \mathbb{E}_{F,\mathcal{T}}[\mathbb{E}_i[D]]) \frac{dS_{1k}}{d\tau}
\end{aligned}$$

Where we define $\zeta(\tau) \equiv \int_{i\in\mathcal{T}} dF(i)$; this normalization by the number of active investors is necessary to use expectation and covariance operators. Using the fact that $\int_{i\in\mathcal{S}} \frac{dX_{1i}}{d\tau} dF(i) = \frac{dS_{1k}}{d\tau} - \int_{i\in\mathcal{B}} \frac{dX_{1i}}{d\tau} dF(i)$, the denominator in (40) can be expressed as: $\int \text{sgn}(\Delta X_{1i}) \frac{dX_{1i}}{d\tau} dF(i) = \int_{i\in\mathcal{B}} \frac{dX_{1i}}{d\tau} dF(i) - \int_{i\in\mathcal{S}} \frac{dX_{1i}}{d\tau} dF(i) = 2\int_{i\in\mathcal{B}} \frac{dX_{1i}}{d\tau} dF(i) - \frac{dS_{1k}}{d\tau}$. By substituting and rearranging the previous two expressions in the optimal tax formula, we can write τ^*as:

$$\tau^* = \underbrace{\frac{-2\int_{i\in\mathcal{B}} \frac{dX_{1i}}{d\tau} dF(i)}{-2\int_{i\in\mathcal{B}} \frac{dX_{1i}}{d\tau} dF(i) + \frac{dS_{1k}}{d\tau}}}_{\equiv 1-\omega} \underbrace{\frac{-\zeta(\tau) \mathbb{C}ov_{F,\mathcal{T}}\left[\frac{\mathbb{E}_i[D]}{P_1}, \frac{dX_{1i}}{d\tau}\right]}{-2\int_{i\in\mathcal{B}} \frac{dX_{1i}}{d\tau} dF(i)}}_{\equiv \tau^*_{\text{exchange}}} + \underbrace{\frac{\frac{dS_{1k}}{d\tau}}{-2\int_{i\in\mathcal{B}} \frac{dX_{1i}}{d\tau} dF(i) + \frac{dS_{1k}}{d\tau}}}_{\equiv \omega} \underbrace{\frac{\mathbb{E}[D] - \mathbb{E}_{F,\mathcal{T}}[\mathbb{E}_i[D]]}{P_1}}_{\equiv \tau^*_{\text{production}}}$$

Proposition 11. (Harberger (1964) revisited)

a) When there are no belief differences between investors and the planner or the planner assesses social welfare respecting individual beliefs, we can write the marginal change in welfare as a money-metric (divided by investors' marginal utility) as:

$$\left.\frac{d\hat{V}_i}{d\tau}\right|_{\tau=\tilde{\tau}} \equiv \frac{\frac{dV_i}{d\tau}}{\mathbb{E}\left[U_i'\left(W_{2i}\right)\right]} = \operatorname{sgn}\left(\Delta X_{1i}\right)\tilde{\tau}P_1\frac{dX_{1i}}{d\tau} - \Delta X_{1i}\frac{dP_1}{d\tau}$$

Adding up across all investors, and using the fact that $\int \operatorname{sgn}\left(\Delta X_{1i}\right)\frac{dX_{1i}}{d\tau}dF\left(i\right) = 2\int_{i\in\mathcal{B}}\frac{dX_{1i}}{d\tau}dF\left(i\right)$, we then recover Equation (42).

b) The result in a) is an exact expression. However, we can write a second order approximation around $\tau = 0$ of the marginal change in social welfare. Note all terms corresponding to terms-of-trade cancel out after imposing market clearing, so I do not consider them. The first term of the Taylor expansion is given above. The derivative of the second term of the Taylor expansion is given by: $\operatorname{sgn}\left(\Delta X_{1i}\right)P_1\frac{dX_{1i}}{d\tau}+\operatorname{sgn}\left(\Delta X_{1i}\right)\tau P_1\frac{d^2X_{1i}}{d\tau^2}$. Around $\tau = 0$, this becomes $\operatorname{sgn}\left(\Delta X_{1i}\right)P_1\frac{dX_{1i}}{d\tau}$. Hence, when $\tau = 0$ we can write:

$$\begin{aligned}\int \left. d\hat{V}_i\right|_{\tau=0} dF\left(i\right) \approx & \int \operatorname{sgn}\left(\Delta X_{1i}\right)\tau P_1\left.\frac{dX_{1i}}{d\tau}\right|_{\tau=0} dF\left(i\right)\left(d\tau\right) \\ & + \frac{1}{2}\int\left.\left(\operatorname{sgn}\left(\Delta X_{1i}\right)P_1\frac{dX_{1i}}{d\tau} + \operatorname{sgn}\left(\Delta X_{1i}\right)\tau P_1\frac{d^2X_{1i}}{d\tau^2}\right)\right|_{\tau=0} dF\left(i\right)\left(d\tau\right)^2 \\ = & P_1\tau^2\int_{i\in\mathcal{B}}\left.\frac{dX_{1i}}{d\tau}\right|_{\tau=0} dF\left(i\right).\end{aligned}$$

Proposition 12. (Disagreement about second moments)

The optimal portfolio allocation for an investor i in his trade region is given by:

$$X_{1i} = \frac{\mathbb{E}_i\left[D\right] - A_i\mathbb{C}ov_i\left[E_{2i},D\right] - P_1\left(1+\operatorname{sgn}\left(\Delta X_{1i}\right)\tau\right)}{A_i\mathbb{V}ar_i\left[D\right]}.$$

The marginal change in welfare for an investor i is given by:

$$\frac{\frac{dV_i}{d\tau}}{\mathbb{E}\left[U_i'\left(W_{2i}\right)\right]} = \left[\begin{array}{c}\left(\mathbb{E}\left[D\right] - A_i\mathbb{C}ov\left[E_{2i},D\right] - P_1\right)\frac{dX_{1i}}{d\tau} \\ -r_i\left(\mathbb{E}_i\left[D\right] - A_i\mathbb{C}ov_i\left[E_{2i},D\right] - P_1\left(1+\operatorname{sgn}\left(\Delta X_{1i}\right)\tau\right)\right)\frac{dX_{1i}}{d\tau} - \Delta X_{1i}\frac{dP_1}{d\tau}\end{array}\right].$$

Where $r_i \equiv \frac{\mathbb{V}ar[D]}{\mathbb{V}ar_i[D]}$. The change in social welfare can then be written as:

$$\frac{dV}{d\tau} = \int\left(-r_i\mathbb{E}_i\left[D\right] - A_i\mathbb{C}ov\left[E_{2i},D\right]\left(1-\frac{\beta_{ii}}{\beta_i}\right) + P_1 r_i\left(1+\operatorname{sgn}\left(\Delta X_{1i}\right)\tau\right)\right)\frac{dX_{1i}}{d\tau}dF\left(i\right).$$

Solving for τ in this equation, which corresponds to Equation (43) in the paper, delivers the expression for the optimal tax in Proposition 12b).

Proposition 13. (Trading constraints)

The equilibrium price is given by:

$$P_1 = \frac{\int_{i\in\mathcal{T},U}\frac{\mathbb{E}_i[D]-A_i\mathbb{C}ov[E_{2i},D]}{A_i\mathbb{V}ar[D]} - \int_{i\in\mathcal{T}}X_{0i} - \int_{i\in\mathcal{C}}g_i\left(P_1\right)}{\int_{i\in\mathcal{T},U}\frac{1+\operatorname{sgn}(\Delta X_{1i})\tau}{A_i\mathbb{V}ar[D]}},$$

where $i \in \mathcal{T}, U$ denotes the set of active unconstrained investors.

The change in social welfare for an investor i, from a planner's perspective, is given by:

$$\frac{dV_i}{d\tau} = \mathbb{E}\left[U_i'\left(W_{2i}\right)\right]\left[\left[\mathbb{E}\left[D\right] - A_i\mathbb{C}ov\left[E_{2i}, D\right] - P_1 - A_i X_{1i}\mathbb{V}ar\left[D\right]\right]\frac{dX_{1i}}{d\tau} - \Delta X_{1i}\frac{dP_1}{d\tau}\right].$$

I use $i = U$ to denote unconstrained investors and $i = C$ for constrained investors. Substituting the optimality condition:

$$\left.\frac{dV_i}{d\tau}\right|_{i=U} = \mathbb{E}\left[U_i'\left(W_{2i}\right)\right]\left[\left[\mathbb{E}\left[D\right] - \mathbb{E}_i\left[D\right] + \operatorname{sgn}\left(\Delta X_{1i}\right)P_1\tau\right]\frac{dX_{1i}}{d\tau} - \Delta X_{1i}\frac{dP_1}{d\tau}\right].$$

$$\left.\frac{dV_i}{d\tau}\right|_{i=C} = \mathbb{E}\left[U_i'\left(W_{2i}\right)\right]\left[\left[\mathbb{E}\left[D\right] - A_i\mathbb{C}ov\left[E_{2i}, D\right] - P_1 - A_i g\left(P_1\right)\mathbb{V}ar\left[D\right]\right]\frac{dX_{1i}}{d\tau} - \Delta X_{1i}\frac{dP_1}{d\tau}\right].$$

We can show that the term multiplying $\frac{dX_{1i}}{d\tau}$ for constrained investors is positive when

$$g\left(P_1\right) < \frac{\mathbb{E}\left[D\right] - A_i\mathbb{C}ov\left[E_{2i}, D\right] - P_1}{A_i\mathbb{V}ar\left[D\right]},$$

and negative otherwise. The welfare change for constrained investors can be rewritten, by substituting the (shadow) first order condition as:

$$\left.\frac{dV_i}{d\tau}\right|_{i=C} = \mathbb{E}\left[U_i'\left(W_{2i}\right)\right]\left[\left[\mathbb{E}\left[D\right] - \mathbb{E}_i\left[D\right] + A_i\mathbb{V}ar\left[D\right]\left(\hat{X}_{1i} - g\left(P_1\right)\right) + P_1\operatorname{sgn}\left(\Delta X_{1i}\right)\tau\right]\frac{dX_{1i}}{d\tau} - \Delta X_{1i}\frac{dP_1}{d\tau}\right].$$

We can write social welfare as:

$$\frac{dV}{d\tau} = \int\left[-\mathbb{E}_i\left[D\right] + \operatorname{sgn}\left(\Delta X_{1i}\right)P_1\tau\right]\frac{dX_{1i}}{d\tau}dF\left(i\right) + \int_{i\in\mathcal{C}} A_i\mathbb{V}ar\left[D\right]\left(\hat{X}_{1i} - g\left(P_1\right)\right)\frac{dX_{1i}}{d\tau}dF\left(i\right).$$

Where I use $\mathcal{C}$ to denote the set of constrained investors and $\hat{X}_{1i}$ is given by the individual first order condition in Equation (5). After substituting for constrained investors $\frac{dX_{1i}}{d\tau} = g'\left(P_1\right)\frac{dP_1}{d\tau}$, we can write the optimal tax as:

$$\int\left[-\mathbb{E}_i\left[D\right] + \operatorname{sgn}\left(\Delta X_{1i}\right)P_1\tau\right]\frac{dX_{1i}}{d\tau}dF\left(i\right) + \int_{i\in\mathcal{C}} A_i\mathbb{V}ar\left[D\right]\left(\hat{X}_{1i} - g\left(P_1\right)\right)g'\left(P_1\right)\frac{dP_1}{d\tau}dF\left(i\right) = 0.$$

$$\tau^* = \frac{\int\mathbb{E}_i\left[D\right]\frac{dX_{1i}}{d\tau}dF\left(i\right) - \int_{i\in\mathcal{C}} A_i\mathbb{V}ar\left[D\right]\left(\hat{X}_{1i} - g\left(P_1\right)\right)g'\left(P_1\right)\frac{dP_1}{d\tau}dF\left(i\right)}{P_1\int\operatorname{sgn}\left(\Delta X_{1i}\right)\frac{dX_{1i}}{d\tau}dF\left(i\right)}.$$