

Learning in Crowded Markets *

Péter Kondor[†]

Adam Zawadowski[‡]

London School of Economics

Central European University

March 12, 2016

Abstract

We develop a model of capital reallocation to analyze whether the presence of more arbitrageurs improves capital allocation and welfare. While trades can become crowded due to imperfect information and externalities, arbitrageurs can devote resources to flexibly learn about the number of earlier entrants. Above a threshold, increasing the number of arbitrageurs does not affect capital allocation: whether there is eventually too little or too much capital allocated to the trade is solely determined by the parameters of the market. The flexibility in the learning technology is key to this insight. However, the presence of more arbitrageurs decreases welfare, as they use more aggregate resources to learn about each others' position. When both sophisticated and unsophisticated arbitrageurs are present, increasing the share of sophisticated arbitrageurs might be welfare reducing.

*For comments and suggestions, we are grateful to Patrick Bolton, Thomas Chemmanur, Johannes Hörner, Victoria Vanasco and seminar participants at Boston University, Central European University, ESSET 2014 (Gerzensee), Brigham Young University, GLMM 2015 (Boston), Paul Woolley Conference 2015 (LSE). We thank Pellumb Reshidi for his excellent research assistance. The support of the European Research Council Starting Grant #336585 is acknowledged.

[†]e-mail: p.kondor@lse.ac.uk, <http://personal.lse.ac.uk/kondor>

[‡]e-mail: zawa@ceu.edu, <http://www.people.bu.edu/zawa>

1 Introduction

Global markets are increasingly dominated by hedge funds, a group often thought of as sophisticated arbitrageurs. A fundamental question in finance is whether an increasing presence of such traders makes markets more efficient and increases welfare. As [Stein \(2009\)](#) points out, a major challenge for these funds is that many of the trading strategies they follow are “unanchored”, i.e., a given manager cannot easily tell from prices how many others have already taken the same position. A prime example for such investment opportunities are emerging markets or new technologies. Thus there is a growing concern among analysts and practitioners that these trading strategies can become “crowded” as managers inflict negative externalities on each other.¹ On the other hand, managers devote large amounts of resources to analyzing trading opportunities in order to avoid “crowded” markets with low payoffs. In this paper, we focus on how the opportunity to learn changes the welfare effects of the increasing presence of arbitrageurs.

We analyze this question through a capital reallocation problem with learning and potential externalities.² Our main observation is that more arbitrageurs having the opportunity to enter a trade neither improves the efficiency of capital allocation (market efficiency) nor does it aggravate crowding. In fact, whether there is eventually too little or too much capital allocated to the trade is determined solely by the technology in that sector, the cost of learning, the depth of the market, and the severity of the potential shocks, but not the mass of sophisticated arbitrageurs present. Being able to learn in a flexible way is key to this insight. On the other hand, the presence of more arbitrageurs does decrease welfare, even though aggregate entry remains unchanged. The insight is that as more arbitrageurs are present, to avoid entering in overcrowded market, they invest increasing amounts of resources in learning about each others’ position without any added social value. We also show that increasing

¹The August 2007 crash of quant funds is commonly attributed to over-crowded strategies, see [Khandani and Lo \(2011\)](#). In an early paper, [MacKenzie \(2003\)](#) also attributes the 1998 LTCM episode to crowded trades. Furthermore, [Sun et al. \(2012\)](#) show that trades followed by many hedge fund managers do not perform as well as ones that are not imitated that much.

²The idea of modelling arbitrage as cross-market capital reallocation of capital goes back to at least [Gromb and Vayanos \(2002\)](#). For a survey on financially constrained arbitrage, see [Gromb and Vayanos \(2010\)](#).

the share of more sophisticated arbitrageurs, while keeping the number of potential entrants fixed, might also lower welfare since less sophisticated arbitrageurs leave the market and more sophisticated arbitrageurs engage in a vicious race of costly learning.

In our model, arbitrageurs' main problem is to decide whether to invest in a new sector with scarce capital and thus high marginal product of capital. The random order in which each arbitrageur learns about the existence of this trading opportunity is its unknown type. Because of decreasing returns to scale, trades can become crowded and arbitrageurs who invest earlier (lower types) benefit the most. The characteristics of the new sector and the potential shocks determine the relative payoff to early and late entrants (low and high types, respectively). If the marginal product of capital is strongly decreasing in the new sector, early entrants gain more relative to late entrants. Also, a market with more arbitrageurs provides an easier exit opportunities for those hit by an idiosyncratic liquidity shock. On the other hand, in a market with more arbitrageurs, an aggregate liquidity shock creates a more severe price crash making exit more costly for all. Nevertheless, the average type would always prefer a smaller number of entrants: this is the crowding externality.

We allow arbitrageurs to learn about their type, which in equilibrium is equivalent to learning about how many arbitrageurs have already entered the trade. Following the rational inattention approach of [Sims \(1998\)](#), [Woodford \(2008\)](#) and [Yang \(2015a\)](#), we allow arbitrageurs to acquire optimally chosen flexible signals about their type, subject to a constant marginal cost of reducing information entropy. The arbitrageur's problem thus reduces to choosing a function which maps its ex ante unknown type into a probability of entry. For example, entering with a constant probability regardless of its type is free of learning cost as this function has the maximum entropy. However, entering only when very few entered before is costly, because this requires a large reduction in the entropy. We think of a high marginal cost μ trade, such as investing in an emerging market, as one for which prices give little information on how many arbitrageurs are already following that strategy. Apart from the parsimony of jointly capturing choice and learning, the main advantage of this learning specification is that it allows for full flexibility in learning and also has an axiomatic foundation based on information theory.

Our main focus is to analyze how the allocation of capital and welfare changes as we increase the total mass of arbitrageurs who eventually learn about the investment opportunity. It is not surprising that when the total mass of arbitrageurs is small, all arbitrageurs enter, therefore they do not engage in costly learning. At that stage, more arbitrageurs lead to more entrants, more efficient allocation of capital and higher welfare. However, when the total mass of arbitrageurs is sufficiently large, arbitrageurs learn and differentiate their entry probability by their type. There are two main effects which both get stronger as the number of arbitrageurs increases. First, the relative benefit of entering early (low type) compared to late (high type) is increasing because there are more arbitrageurs who may potentially enter before any given type. Thus arbitrageurs choose entry strategies that are more contingent on their types. We refer to this as the “rat race effect”. Second, the negative effect from crowding for the average entrant is increasing. This pushes average entry of a given arbitrageur down. We refer to this as the “crowding effect”. While the rat race has little effect on average entry, the crowding effect decreases average entry. Due to optimal signal choice, as we increase the mass of arbitrageurs, the average entry decreases proportionally leading to constant aggregate entry as a function of the mass of arbitrageurs. That is, a larger mass of sophisticated arbitrageurs does not improve the efficiency of capital allocation.

As increasing the mass of arbitrageurs does not affect aggregate entry, whether there is over- or under-entry compared to the planners’ solution is independent of the mass of arbitrageurs. Instead, there is permanent over- or under-entry, depending on the parameters. Less idiosyncratic and more aggregate shocks, financial markets with less depth, less strongly decreasing returns to scale technology and higher cost of learning all make over-entry more likely and/or more severe.

Importantly, as the number of potential entrants increase, welfare decreases. This is because the rat race effect increases the private incentives to learn, but does not change the social value of learning. In fact, the rat race effect captures the zero-sum part of the trading process: early entrants gain at the expense of late entrants. In contrast, the planner is interested only in aggregate entry (as opposed to the order of entry), because this determines the efficiency of capital allocation. Therefore, even

though in the presence of more arbitrageurs, aggregate entry does not increase, they do invest more in learning whether to enter. This leads to a wasteful use of resources and decreasing welfare.

Allowing arbitrageurs to flexibly choose their information (and thus their entry) is crucial in generating the result of constant entry as the mass of arbitrageurs increases. With flexible learning the arbitrageurs can optimally devise their strategy to exactly enter at a constant aggregate rate. We demonstrate this in an extension in which arbitrageurs can only buy Gaussian signals about their type subject to the same entropy cost as before. Aggregate entry is not constant in this case, we show that it converges to the equilibrium with no information acquisition. Thus with fixed information structure the entry of arbitrageurs becomes too constrained and they end up not purchasing information. Since arbitrageurs in the real world are quite inventive in gathering and processing information, we believe the results for flexible information structure are important in understanding market behavior.

We also extend our model to the case in which there is heterogeneity across arbitrageurs: some are more sophisticated, others are less sophisticated. Keeping the mass of all arbitrageurs fixed but increasing the share of more sophisticated arbitrageurs might also decrease welfare. While having some sophisticated arbitrageurs increases welfare since it raises the average sophistication of arbitrageurs and this can alleviate over-entry due to the crowding externality. As more sophisticated arbitrageurs abound, less sophisticated arbitrageurs afraid of being ripped off exit the market, leaving more sophisticated arbitrageurs in a vicious race of spending increasing resources on learning.

We show that our general model allows for a wide range of other applications differing in the source of the externalities and their welfare implications. For example, firms in industries with knowledge spillovers benefit from others following their location choices. Academics benefit from entering into fields early which later become very popular and thus increasing the number of their citations. Welfare consequences of crowding can be very different across these applications. The main reason is that in most other applications than our baseline of capital reallocation, there is a group of arbitrageurs whose welfare increases when others enter too much. For example, the over-entry of firms in an industry with knowledge spillover might decrease consumer prices in that industry below marginal cost. This

harms firms but helps consumers. In our analysis, this effect makes markets with a higher learning costs more attractive from a welfare point of view.

Our main contribution is to embed learning in a model of capital allocation. Our paper is connected to various branches of literature. First, there is a recent and growing literature on slow moving capital, see [Duffie \(2010\)](#), [Duffie and Strulovici \(2012\)](#), and [Greenwood et al. \(2015\)](#). They also assume that capital does not immediately move to markets where it is scarce but focus on the asset pricing implications. We focus on the endogenous choice of the amount of capital transferred and show that even though increasing the amount of arbitrageurs might make capital transfer faster, markets do not converge to efficiency and welfare deteriorates.

Second, there is a literature analyzing entry/exit in the presence of externalities from other investors. [Stein \(2009\)](#) introduces a simple model of crowded markets but leaves the effect of learning in such models for future research. [Abreu and Brunnermeier \(2003\)](#) and [Moinas and Pouget \(2011\)](#) show that the inability to learn about one's relative position versus that of other investors' is a key ingredient in sustaining excessive investment in bubbles. This highlights our contribution in adding learning to a model of crowded markets with potential over-entry.

Third, a growing literature analyzes the consequences of limited information processing capacity based on the rational inattention approach pioneered by [Sims \(1998\)](#) and [Sims \(2003\)](#). [Maćkowiak and Wiederholt \(2009\)](#), [Hellwig and Veldkamp \(2009\)](#) and [Kacperczyk et al. \(2015\)](#) study the allocation of limited attention across signals but restrict the signals to be Gaussian. Fully flexible information acquisition in rational inattention models is employed by [Matějka and McKay \(2015\)](#), [Woodford \(2008\)](#), [Yang \(2015a\)](#) and [Yang \(2015b\)](#). In a coordination game, [Yang \(2015a\)](#) points out that constraining the signal structure instead of choosing it in a fully flexible way is not without loss of generality. While methodologically related to ours, none of the above papers directly analyze capital allocation.

Fourth, there are numerous papers showing excessive investment in learning or effort. In models of high frequency trading, [Budish et al. \(2015\)](#) and [Biais et al. \(2015\)](#) show that there is excessive

investment in speed if trading is continuous in time. Our framework is conceptually different: investors cannot change their type (speed) they can only learn about it. Also, our insights work on longer time horizons, e.g. in the setting of investing in emerging markets. [Hirshleifer \(1971\)](#) shows that in endowment economies, learning about information that will eventually be revealed is privately profitable but socially wasteful. This is the case e.g. in [Grossman and Stiglitz \(1980\)](#), though no formal welfare analysis is possible because of the presence of non-optimizing noise traders. [Glode et al. \(2012\)](#) also focus on inefficient over-learning and how it may lead to market breakdowns by exacerbating asymmetric information. More generally, socially inefficient effort choice has also been emphasized in very different settings: [Tullock \(1967\)](#) and [Krueger \(1974\)](#) show it in a rent seeking setting, while [Loury \(1979\)](#) shows that excessive resources are devoted to research in a patent race. Our focus is different compared to the above papers: we are interested in how learning affects capital allocation.

The rest of the paper is structured as follows. In [Section 2](#) we present our reduced form model and also give a structural microfoundation. In [Section 3](#) we first solve the problem under no information and full information. Then we present the general results for the entry and learning decision and analyze how this relates to aggregate entry and welfare. In [Section 4](#) we analyze an extension of the model with two types of arbitrageurs. In [Section 5](#) we check how important flexible learning is in generating our results by comparing it to a model with Gaussian signals. In [Section 6](#) we present further microfoundations of the reduced form model and discuss how our main insights are affected by these different settings.

2 Learning and investing in crowded markets

In this part we describe our set up. We first present the reduced form setup we use in the paper and then describe a micro-foundation.

2.1 Payoffs

The heart of our model is an entry game with a continuum (mass M) arbitrageurs, each with a type $\theta \in [0, 1]$. Each arbitrageur can decide to take an action: whether to enter in a market or not. θ is interpreted as the time when arbitrageur θ can make this decision. The utility gain (or loss, if negative) from entry is given by

$$\Delta u(\theta) = 1 - \beta \cdot b(\theta) + \alpha \cdot a(\theta) \tag{1}$$

where α and β are constants, $a(\theta)$ is the mass of entrants whose type is higher than θ (the arbitrageurs who enter *after* arbitrageur θ), while $b(\theta)$ is the mass of entrants with a type lower than θ (arbitrageurs who enter *before* arbitrageur θ). We show in the microfoundation that the following two assumptions are natural. First, $\beta + \alpha > 0$, such that entering earlier is better than later: we call this property *rat-race*. Second, $\beta - \alpha > 0$, such that the average entrant imposes a negative externality on others: we call this property *crowding*. The two assumptions together imply that $\beta > 0$ while α could be positive or negative. As we specify below, players do not know their type, but can gather information about it through a costly learning process.

While throughout the paper we work with the reduced form payoff (1), to clarify the economic interpretation of the parameters α and β it is useful to develop a fully specified economic model microfounding the reduced form (1). In the next part we present such a model in the context of capital arbitrage: this is our leading microfoundation. There are many other potential microfoundations of this reduced-form model, such as academic publication tournament, production with externalities and behavioral utility functions that reward early adoption of a trend; several of these are described in Section 6.

2.2 Microfoundation: Capital arbitrage

2.2.1 Baseline: No shocks

There are two islands A and B indexed by $i \in \{A, B\}$. There are two types of agents: a worker on each island and a continuum of arbitrageurs with mass M uniformly distributed over types denoted by $\theta \in [0, 1]$. There are three types of goods: capital, a specialized consumption good produced by capital, and a numeraire good. The numeraire good is used as a method of exchange and all arbitrageurs are endowed with sufficient numeraire goods to make transactions possible. Time is continuous and denoted by $t \in [0, 1]$. At $t = 0$ capital is inefficiently distributed, the worker on island A is endowed with $k_{A,0}$ capital, while the worker on island B has none ($k_{B,0} = 0$). We think of island B as an emerging idea/industry/country representing a profitable investment opportunity found sequentially by arbitrageurs.

At time t , arbitrageur of type $\theta = t$ has the opportunity to buy one unit of capital on island A , transport it to island B and sell it there. For simplicity, we assume workers do not have any bargaining power and buy and sell capital at a price equal to its marginal product at the given capital level at the time.³ Production only happens at $t = 1$: on island i , k_i capital produces c_i consumption good according to the production function

$$c_i = \gamma \cdot k_i - \delta_i \cdot \frac{k_i^2}{2} \quad (2)$$

thus on island i the marginal product of capital, given the capital level at time t , becomes:

$$\frac{dc_i}{dk_{i,t}} = \gamma - \delta_i \cdot k_{i,t}. \quad (3)$$

In our baseline case, on both islands $\delta_i = \delta > 0$. All consumption happens at $t = 1$ after production has been completed. Arbitrageurs only consume the n_j numeraire goods with utility $U_{arbitrageur} = n_j$.

³This is a shortcut to capture that workers know even less than arbitrageurs and thus cannot capture any of the surplus generated by trade.

Worker i consumes n_i numeraire goods it possesses and c_i consumption goods it produces with utility $U_i = n_i + c_i$.

Denote by $b(t)$ the mass of arbitrageurs who chose to enter (i.e. engage in capital transport) before time t and $a(t)$ the mass of arbitrageurs who enter after time t . Thus $k_{A,t} = k_{A,0} - b(t)$ and $k_{B,t} = b(t)$: there is more capital on island B if already $b(t)$ arbitrageurs have decided to transport capital there from island A . Then, the revenue of an arbitrageur that chooses to transport capital at time t is

$$\underbrace{\left[\gamma - \delta \cdot b(t) \right]}_{\text{sell price}} - \underbrace{\left[\gamma - \delta \cdot [k_{A,0} - b(t)] \right]}_{\text{buy price}} = \delta \cdot k_{A,0} - 2\delta \cdot b(t)$$

Choosing $k_{A,0} = \frac{1}{\delta}$, the expected payoff of arbitrageur θ from transporting capital (given that arbitrageur θ can enter at time t) simplifies to (1) with:

$$\alpha = 0, \quad \beta = 2\delta. \tag{4}$$

Resulting in strictly positive crowding and rat-race parameters of:

$$\beta - \alpha = \alpha + \beta = 2\delta > 0 \tag{5}$$

We refer to this baseline case as the no-shock case. Note that both our rat-race and our crowding effect is driven solely by δ which measures the extent of decreasing return to scale. That is, entering early is beneficial because the investment opportunity is more profitable when not many arbitrageurs have entered yet. This shows that in financial market it is natural to assume that crowding comes hand-in-hand with a rat race even in the absence of any externalities.

2.2.2 Introducing shocks

To explore the effect of various externalities in our problem, we also consider the possibility that the capital reallocation is subject to shocks. In particular, suppose that the transport is successful only

with probability $1 - \nu$, with probability $\nu \geq 0$, the arbitrageur is hit by an idiosyncratic (“liquidity”) shock and has to go back to island A and sell the capital there at $t = 1$. Furthermore, with ex ante probability $\eta \geq 0$, even if the capital transfer is successful, upon arriving on island B , there is an aggregate shock (“crisis”) and all arbitrageurs have to sell their capital in a fire sale. While on island A , we still always have $\delta_i = \delta$, on island B , $\delta_i = \delta$ only if there is no crisis, but $\delta_i = \delta_c + \delta$ if there is a crisis (aggregate shock) where $\delta_c > 0$.

Note that unless there is an idiosyncratic shock, arbitrageurs buy capital at its marginal return on island A and sell capital at its marginal return on island B . Whenever this activity is profitable, it also decreases the difference between the marginal return on capital across the two islands, that is, it increases market efficiency. Idiosyncratic shock complicates this picture only to the extent that it introduces some redistribution among arbitrageurs; an element which washes out by aggregation. Therefore, we can interpret the aggregate revenue of arbitrageurs as a measure of market efficiency.

We interpret our deep parameters as follows. δ captures the extent of decreasing return to scale in each market. Conceptually, this is a technological parameter of the sectors or firms which are subject to the capital reallocation. δ_c characterizes the depth of the financial market where claims on these firms and sectors are traded. When δ_c is large, a sudden selling pressure of the participating arbitrageurs drives the price down significantly. In contrast, η and ν characterizes the arbitrageurs as opposed to the markets. A large η is interpreted as a large probability for a common liquidity shock for all arbitrageurs, for example, because of large common exposure to risk factors outside of the model. A large ν is a large probability of an idiosyncratic liquidity shock, for example, because arbitrageurs are professional investors with a volatile investor base.

Now we recalculate α and β with shocks. Note that with probability ν the arbitrageur is reverted to island A and sells at the average marginal product of capital: $\nu \cdot (a(t) + b(t))$ capital is sold by arbitrageurs at $t = 1$ in random order.⁴ In case of a crisis, $(1 - \nu) \cdot (a(t) + b(t))$ capital is sold on island

⁴On average half of the entrants sell before the arbitrageur in a fire sale. Thus the expected price is higher than if it was sold at a market clearing price in e.g. an auction. This assumption simplifies the analysis by not allowing workers to capture any of the surplus.

B in a fire sale, in random order. Overall, the revenue of an arbitrageur that chooses to transport capital at time t is:

$$\begin{aligned}
& (1 - \eta) \cdot (1 - \nu) \cdot \underbrace{\left[\gamma - \delta \cdot (1 - \nu) \cdot b(t) \right]}_{\text{sell price (no shock)}} + \eta \cdot (1 - \nu) \cdot \underbrace{\left[\gamma - (\delta_c + \delta) \cdot (1 - \nu) \cdot \frac{a(t) + b(t)}{2} \right]}_{\text{sell price (crisis)}} + \\
& \nu \cdot \underbrace{\left[\gamma - \delta \cdot \left(k_{A,0} - [a(t) + b(t)] + \nu \cdot \frac{a(t) + b(t)}{2} \right) \right]}_{\text{sell price (idiosyncratic shock)}} - \underbrace{\left[\gamma - \delta \cdot [k_{A,0} - b(t)] \right]}_{\text{buy price}} \quad (6)
\end{aligned}$$

Choosing $k_{A,0} = \frac{1}{\delta \cdot (1 - \nu)}$, the expected payoff of arbitrageur θ from transporting capital (given that arbitrageur θ can enter at time t) simplifies to (1) if

$$\alpha = \left(\frac{1}{2} + (1 - \nu)^2 \left(\frac{1 - \eta}{2} - 1 \right) \right) \cdot \delta - \frac{1}{2} (1 - \nu)^2 \cdot \eta \cdot \delta_c \quad (7)$$

$$\beta = \left(\frac{1}{2} + (1 - \nu)^2 \left(\frac{1 - \eta}{2} + 1 \right) \right) \cdot \delta + \frac{1}{2} (1 - \nu)^2 \cdot \eta \cdot \delta_c \quad (8)$$

Resulting in positive crowding and rat-race parameters of:

$$\beta - \alpha = (1 - \nu)^2 \cdot (\eta \cdot \delta_c + 2\delta) > 0 \quad (9)$$

$$\alpha + \beta = (1 + (1 - \nu)^2 (1 - \eta)) \cdot \delta > 0 \quad (10)$$

for all parameter values.

To interpret α and β , it is useful to first consider the case without idiosyncratic shock ($\nu = 0$). In this case, $\alpha = -\frac{1}{2}\eta(\delta + \delta_c)$ and $\beta = 2\delta + \frac{1}{2}\eta(\delta_c - \delta)$, implying $\beta - \alpha = 2\delta + \eta \cdot \delta_c$ and $\alpha + \beta = (2 - \eta) \cdot \delta$. Note that the crowding parameter, $\beta - \alpha$, is increasing in the probability of the aggregate liquidity shock, η , and the illiquidity of the market, δ_c . That is, entrants impose a negative externality on each other, because it is more costly to exit when more arbitrageurs want to exit at the same time. Finally, note that α is negative without idiosyncratic shock, because of the same logic: late entrants harm

early entrants because they aggravate crowding. Without idiosyncratic shock, the effect of more late entrants in a liquidity crisis is the same as the effect of more entrants, late or early.

While the introduction of idiosyncratic shock affects all our reduced form parameters, its main qualitative effect is that it changes the sign of α . Indeed, α is monotonically increasing in ν , reaching $\frac{1}{2} \cdot \delta > 0$ when $\nu = 1$. The intuition is that for large ν early entrants benefit from late entrants since if they have to liquidate their position, they can do so at a higher price. This means that α is likely to be positive in markets where entrants need enough subsequent liquidity to exit at a reasonable price.

Comparing the cases with and without shocks, it is clear that in both cases (1) (with positive crowding and rat-race parameters) is a suitable reduced form representation of payoffs. The main qualitative difference is that without shocks $\alpha = 0$, while with shocks α can be both negative and positive. We show in Section 3.3 that in the case with shocks, changing the parameters affecting α and β help understand how pay-off externalities affect the learning and entry decisions of our arbitrageurs.

2.3 Learning cost based on entropy

Before entry, arbitrageurs can engage in costly learning about their type. We specify the learning cost to be proportional to the reduction in entropy.⁵ Sims (1998) argues that the advantage of such a specification is that it both allows for flexible information acquisition and can be derived based on information theory. Note that the payoff (1) for a given θ in our model is linear in entry. Woodford (2008) derives the optimal signal structure and entry decision rule for such problems which we restate in the lemma below.

Lemma 1. *The optimal signal structure is binary: given their type θ , arbitrageurs choose to receive signal $s = 1$ with probability $m(\theta)$ and $s = 0$ with probability $1 - m(\theta)$. The optimal entry decision conditional on the signal is: enter if $s = 1$, stay out if $s = 0$.*

⁵The entropy of a discrete variable is defined as $\sum_x P(x) \log \frac{1}{P(x)}$, where the random variable takes on the value x with probability $P(x)$, see MacKay (2003).

Thus, similar to [Yang \(2015a\)](#), $m(\theta)$ is the only choice variable (learning and entry strategy combined) which in turn is the conditional probability of entry. The intuition for the binary signal structure is that the only reason arbitrageurs want to learn about θ is to be able to make a binary decision of whether or not to enter. Given the linearity of the problem, the “cheapest” signal to implement the optimal entry strategy is also binary, it simply tells the arbitrageur whether or not to enter.

We now write the cost of learning, defined by the reduction in entropy, in case of a binary information structure. Denote the marginal cost of learning as μ and define the amount of learning L using the informational entropy reduction due to the signal,

$$L(m) = \underbrace{\left(p \log \left[\frac{1}{p} \right] + (1-p) \log \left[\frac{1}{1-p} \right] \right)}_{\text{entropy of no info signal } p} - \int_0^1 \underbrace{\left(m(\theta) \log \left[\frac{1}{m(\theta)} \right] + (1-m(\theta)) \log \left[\frac{1}{1-m(\theta)} \right] \right)}_{\text{entropy of signal } m(\theta)} d\theta, \quad (11)$$

where p denotes the unconditional probability of entry and is defined by:

$$p = \int_0^1 m(\tilde{\theta}) d\tilde{\theta}. \quad (12)$$

The expression for learning [\(11\)](#) can be understood in the following way. Entropy is always highest for the signal that reveals the least amount of information, in this case this is the uninformative signal which prompts the arbitrageur to enter with probability p unconditional on its type θ . Thus learning depends on how much lower the entropy is for the signal $m(\theta)$ than for the flat (uninformative) signal p that gives the same amount of average entry. Intuitively, the steeper $m(\theta)$ becomes in θ (keeping average entry p constant), the more the arbitrageur is differentiating its entry decision according to its type and the higher the entropy reduction, thus the higher the learning cost. Note that the minimum of L is zero when $m(\theta) = p$ and L is bounded from above but might generate infinite marginal cost of learning.

If all arbitrageurs choose the same amount of learning and entry (i.e. the same $m(\theta)$), the mass of lower types entering (“before” arbitrageur θ) becomes:

$$b(\theta) = M \cdot \int_0^\theta m(\tilde{\theta})d\tilde{\theta} \quad (13)$$

mass of higher types entering (“after” arbitrageur θ):

$$a(\theta) = M \cdot \int_\theta^1 m(\tilde{\theta})d\tilde{\theta} \quad (14)$$

thus $M \cdot p = b(\theta) + a(\theta)$ is the aggregate entry of arbitrageurs.

We assume that arbitrageurs have to decide about the amount of information acquisition ex ante without any knowledge about the action of others. We interpret this as the cost of building an information gathering and evaluation “machine” which includes the costs of gathering the right data, building the right contacts, research group, devising the best institutional practices, etc. An alternative would be to think of capacity as limited and μ being the Lagrange multiplier of the capacity constraint. We choose to use a fixed μ instead of a fixed capacity because we think in this context learning capacity can be expanded freely at a fixed marginal cost: e.g. it is always possible to hire new staff or allocate more attention to this specific trade at the expense of other trades.

We interpret μ as a characteristic of the trading strategy. With low μ it is easy for the arbitrageur to determine how many arbitrageurs have entered before, e.g. because the price’s relation to the fundamentals reveals this. Examples for a trade like this would be that of twin stocks or on-the-run-off-the-run bonds: it is clear from the price difference whether an arbitrageur is early (large price gap) or late (small price gap). Another example is merger arbitrage, where the price offered by the bidder is known. On the other hand, with high μ , it is very hard for the arbitrageur to determine whether to enter, e.g. because there is no clear price signal whether the trade is still profitable. In the language of [Stein \(2009\)](#), high μ represents unanchored strategies. Examples for such trades include: emerging markets, carry trade, January effect. While the assumption that learning from prices is also a part of

a more general learning process makes the problem very tractable, it is not without loss of generality. It might be that certain θ 's are easier to learn from prices than others. We leave this theoretical question for future research.

2.4 Definition of allocative efficiency and welfare

The expected revenue of an arbitrageur (which it then consumes in the numeraire good), before taking into account the cost of information acquisition, is:

$$R \equiv \int_0^1 m(\theta) \cdot \Delta u(\theta) \cdot d\theta \quad (15)$$

Recall from section 2.2 that in our leading application, arbitrageurs gain in the aggregate if and only if they reduce the difference in marginal returns of capital across locations, i.e. equate the price of capital in the two regions. Therefore, in this economy, aggregate revenue $M \cdot R$ can be also interpreted as a measure of allocative efficiency.

The total expected payoff (value) per unit of arbitrageur is the revenue from entering, net of the ex ante cost of learning:

$$V \equiv R - \mu \cdot L \quad (16)$$

which is what arbitrageurs maximize. In equilibrium all surplus ends up with the arbitrageur: workers capture none, since they are myopic and sell/buy capital at a price equal to its marginal product. Thus the overall welfare in the whole economy can be computed as $W \equiv M \cdot V$.

3 Model Solution

In this section we present our main results. First, we establish benchmark results for markets with full information when learning is free, and those no information at all when learning is prohibitively expensive. Second, we derive the optimal strategies of arbitrageurs for general levels of learning cost.

Third, we analyze how aggregate entry and welfare changes as the mass of arbitrageurs increases. All proofs are relegated to Appendix A.

3.1 Full information and no information

To better understand the optimal strategies, we first analyze the extreme cases of $\mu = 0$ (full information) and $\mu = \infty$ (no information).

Lemma 2. *For full information ($\mu = 0$), both the competitive and social planner's entry functions $m(\theta)$ are step functions, resulting in the first p arbitrageurs entering. The competitive and social planner's optimum are, respectively:*

$$p|_{\mu=0} = \min\left(\frac{1}{M \cdot \beta}, 1\right) \quad (17)$$

$$p_s|_{\mu=0} = \min\left(\frac{1}{M \cdot (\beta - \alpha)}, 1\right) \quad (18)$$

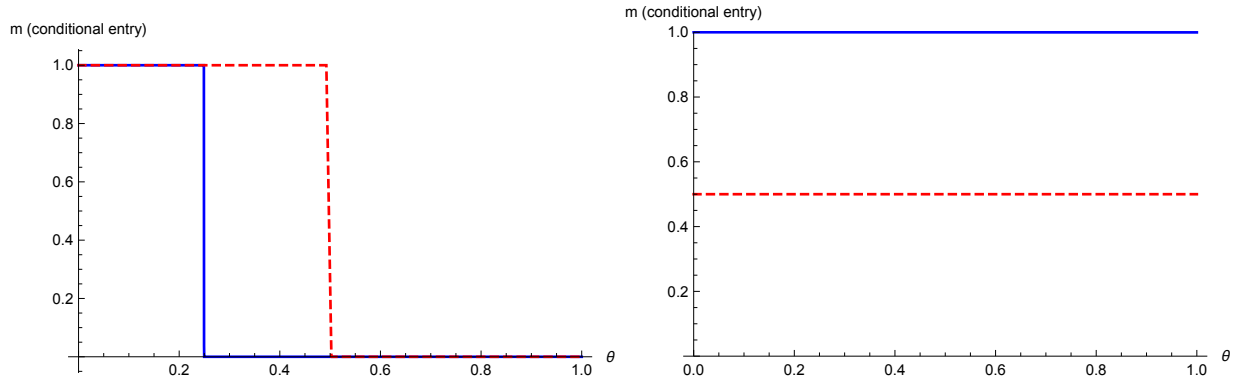
For no information ($\mu \rightarrow \infty$), both the competitive and social planner's entry functions $m(\theta)$ are flat. All arbitrageurs enter with the same unconditional probability:

$$p|_{\mu \rightarrow \infty} = \min\left(\frac{2}{M \cdot (\beta - \alpha)}, 1\right) \quad (19)$$

$$p_s|_{\mu \rightarrow \infty} = \min\left(\frac{1}{M \cdot (\beta - \alpha)}, 1\right). \quad (20)$$

Under full information, whether there is under- or over-entry (compared to that under the social planner) depends on the sign of α , i.e. the type of liquidity shock in the structural model. We analyze three cases: First, when idiosyncratic liquidity shocks are important, there is competitive under-entry if $\alpha > 0$, since arbitrageurs with higher θ do not take into account the positive effect of their entry that accrues to entrants with lower θ . When the order of entry is known, arbitrageurs do not internalize that if they enter, market liquidity increases for everyone, that is, a market with more arbitrageurs

Figure 1: **Entry under full and no information**



The first panel illustrates the entry decision under full information ($\mu = 0$), while the second the entry decision under no information ($\mu = \infty$). The competitive choice is the solid line, the social planner's solution the dashed line. Parameters: $\beta = 4$, $\alpha = 2$, $M = 1$.

provides easier exit opportunities for those hit by an idiosyncratic liquidity shock. Second, when aggregate liquidity shocks are more important, $\alpha < 0$ and there is over-entry. This is due to the fact that arbitrageurs do not take into account that as they enter, in case of an aggregate liquidity shock, they all want to sell reducing the price for everyone. This is a fire-sale externality. Finally, when there is neither idiosyncratic, nor aggregate shocks, $\alpha = 0$. Hence, the competitive solution under perfect information is the same as the social planner's solution. The last entrant takes into account the effect of all others but does not impose any externality on anyone else (since $\alpha = 0$), since there are no more entrants afterwards (thus β does not matter).

Under no information, there is competitive over-entry under any structural parameter values: arbitrageurs enter twice as often than they should. The intuition is analogous to the “tragedy of commons”. While each arbitrageur internalizes that if others enter more often, that reduces its own revenue, it does not internalize that when she enters that reduces the benefit of entry for everyone else. Note that as $\beta - \alpha >$ for any of our structural parameters, there is over-entry even if all type of liquidity shocks have zero probability. In this particular case, the “externality” comes from the decreasing returns to scale technology and the lack of a well observable price.

Lemma 2 makes it clear that simply increasing the mass of arbitrageurs in a market does not mean that entry converges to the socially optimal level. Using the above analysis, one can draw implications

about specific markets. Trades where it is easy to learn (low μ) with high α do not have enough arbitrageurs entering. One example for such a market is twin stocks: the price difference between the two stocks reveals whether it is still profitable to enter but early entrants need later entrants to be able to exit the trade with a profit. Thus the model can give a potential explanation of why there is insufficient entry into trades like twin stocks and why mispricing persists. On the other hand, it also shows why trades where it is hard to learn (high μ), such as carry trade or momentum, might see too much entry: in the extreme case of $\mu \rightarrow \infty$ even driving the revenue of arbitrageurs to zero, like in a tragedy of commons game.

We now turn to welfare as a function of the mass of arbitrageurs M . In both extreme cases it is the same as the aggregate revenue of arbitrageurs since arbitrageurs do not spend anything on learning: either there is no learning (under $\mu \rightarrow \infty$) or learning is free ($\mu = 0$). Thus in these extreme cases, welfare equals the measure of the efficiency of capital allocation.

Lemma 3. *With full information ($\mu = 0$):*

$$W = \begin{cases} M - \frac{M^2(\beta-\alpha)}{2} & \text{if } M \leq \frac{1}{\beta} \\ \frac{(\alpha+\beta)}{2\beta^2} & \text{if } M > \frac{1}{\beta}, \end{cases}$$

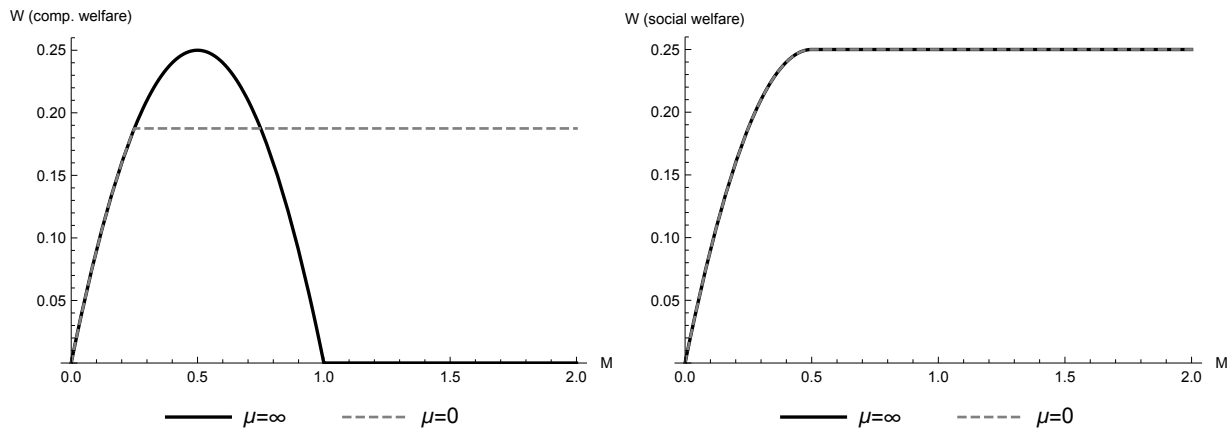
without no information ($\mu \rightarrow \infty$):

$$W = \begin{cases} M - \frac{M^2(\beta-\alpha)}{2} & \text{if } M \leq \frac{2}{\beta-\alpha} \\ 0 & \text{if } M > \frac{2}{\beta-\alpha}. \end{cases}$$

Under the planner, both with and without information:

$$W_s = \begin{cases} M - \frac{M^2(\beta-\alpha)}{2} & \text{if } M \leq \frac{1}{\beta-\alpha} \\ \frac{1}{2(\beta-\alpha)} & \text{if } M > \frac{1}{\beta-\alpha}. \end{cases}$$

Figure 2: **Competitive and social welfare with extreme μ**



Welfare for competitive (left) and social (right) entry as a function of the mass M of arbitrageurs. The no information ($\mu = \infty$) case is represented by the solid line, the full information ($\mu = 0$) case by the dashed line. Parameters: $\beta = 4$, $\alpha = 2$.

Thus increasing the amount of arbitrageurs M in a market has a different effect on welfare depending on the cost of learning μ , see Figure 2. When learning is free, welfare monotonically increases up to the point $M = \frac{1}{\beta}$ and is flat afterwards. Qualitatively, it is a similar picture to the planner's outcome, except that, for $\alpha > 0$ ($\alpha < 0$), we have permanent under-entry (over-entry) leading to reduced welfare compared to the social level. When learning is prohibitively expensive, welfare follows the same increasing path as with full information but continues to grow to a higher level $M = \frac{1}{\beta - \alpha}$ if $\alpha > 0$. The reason is that without information, if $\alpha > 0$, late arbitrageurs cannot figure out they are late and do not stay out of the market. As the mass of arbitrageurs increases further, all revenues are gradually competed away and $W = 0$ for any $M > \frac{2}{\beta - \alpha}$. In this case the model reverts to a simple tragedy of commons. It is immediately clear from the Figure 2 that full information is not always preferred to no information for a given level of the mass of arbitrageurs M , we further analyze this question in a more general setup in Section 6.4.

3.2 Optimal strategies

The private problem of any arbitrageur is to choose its conditional entry $m(\theta)$ to maximize its value function V , which can be written as the following:

$$\max_{m(\theta)} \int_0^1 (m(\theta) \cdot \Delta u(\theta) - \mu \cdot L(m)) d\theta. \quad (21)$$

We contrast the private solution with that of a social planner who can choose the amount of learning and entry for all arbitrageurs. This gives us a benchmark against which we can evaluate learning and entry decisions in the competitive equilibrium. The main difference between the competitive solution and the social planner's one is that the social planner takes into account the externalities that arbitrageurs exert on each other: it takes into account that Δu depends not only on θ but on the choice function of all other arbitrageurs m . The social planner chooses the symmetric function $m_s(\theta)$ to maximize

$$\max_{m_s(\theta)} \int_0^1 (m_s(\theta) \cdot \Delta u(\theta, m_s) - \mu \cdot L(m_s)) d\theta \quad (22)$$

We derive the first order condition (FOC) of these problems using the variation method, i.e. we look for the function $m(\theta)$ such that if we take a very small variation around the function, the value function of the arbitrageurs does not change.

Lemma 4. *Denote the strategy function of all other players as $\tilde{m}(\theta)$. The first-order condition of the private problem is:*

$$M \cdot \alpha \cdot \int_0^1 \tilde{m}(\tilde{\theta}) d\tilde{\theta} - M \cdot \beta \cdot \int_0^\theta \tilde{m}(\tilde{\theta}) d\tilde{\theta} + 1 = \mu \cdot \left[\log \left(\frac{m(\theta)}{1 - m(\theta)} \right) - \log \left(\frac{p}{1 - p} \right) \right]. \quad (23)$$

The first-order condition of the social problem (assuming the same entry function m_s for all arbitrageurs) is:

$$M \cdot (\alpha - \beta) \cdot p_s + 1 = \mu \cdot \left[\log \left(\frac{m_s(\theta)}{1 - m_s(\theta)} \right) - \log \left(\frac{p}{1 - p} \right) \right] \quad (24)$$

Using the FOC one can derive an ordinary differential equation for $m(\theta)$ where the original FOC at $\theta = 0$ (an integral equation) is the boundary condition. The solution of this ordinary differential equation can be expressed up to a constant (boundary value) $m(0)$.

Proposition 1. *If $M \leq \bar{M}$, $m(\theta) = 1$. If $M > \bar{M}$:*

$$m(\theta) = \frac{1}{1 + W_0 \left(e^{M \cdot \frac{\alpha + \beta}{\mu} \cdot \theta + \frac{1 - m(0)}{m(0)} + \log \left(\frac{1 - m(0)}{m(0)} \right)} \right)}, \quad (25)$$

where W_0 denotes the upper branch of the Lambert function⁶ and $m(0)$ is pinned down by the boundary condition ((23) evaluated at $\theta = 0$):

$$M \cdot \alpha \cdot p + 1 = \mu \cdot \left[\log \left(\frac{m(0)}{1 - m(0)} \right) - \log \left(\frac{p}{1 - p} \right) \right]. \quad (26)$$

\bar{M} is pinned down by the following implicit equation:

$$\frac{\bar{M} \cdot (\alpha + \beta)}{\mu} = e^{-\frac{1 - \beta \cdot \bar{M}}{\mu}} - e^{-\frac{1 + \alpha \cdot \bar{M}}{\mu}}. \quad (27)$$

Proposition 2. *If $M \leq \bar{M}_s = \frac{1}{\beta - \alpha}$ then $m_s(\theta) = 1$. If $M > \bar{M}_s$ then:*

$$m_s(\theta) = \frac{1}{M \cdot (\beta - \alpha)}. \quad (28)$$

Note that arbitrageurs want to differentiate between states, but the planner does not. The planner chooses a flat entry function. The reason for this over-learning is the rat race ($\alpha + \beta > 0$) that accompanies the crowding: every arbitrageur wants to know whether he is ahead of the other arbitrageurs even if this is wasteful from the social planner's point of view.

To understand the effect of M on incentives, see Figure 3 which shows the competitive and social planner's optimal entry function for different levels of M . For small $M = 0.5$, arbitrageurs enter for

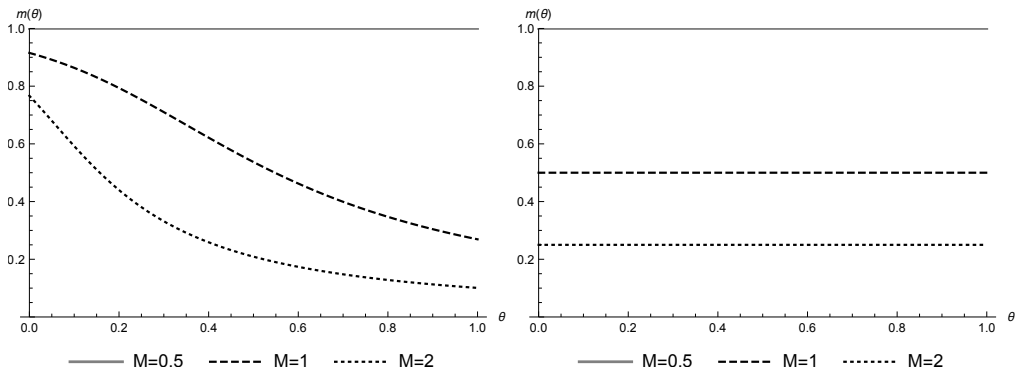
⁶The definition of the upper branch of Lambert function is $z = W_0(z) \cdot e^{W_0(z)}$ if $z > 0$.

sure and there is no need for learning since revenues in the market are high. To gain intuition on how $m(\theta)$ changes as M increases from 1 to 2, consider the effect of larger M on the benefit of entry $\Delta u(\theta)$ for a given arbitrageur, keeping the others' strategy constant. First, note that we can measure the relative incentive for entering earlier by differentiating $\Delta u(\theta)$ in θ , giving $M \cdot (\alpha + \beta) \cdot \tilde{m}(\theta)$. Therefore, keeping other arbitrageur's strategy fixed, the incentive to learn more and follow a more differentiated strategy is increasing in M . Loosely speaking, this results in a steeper $m(\theta)$. Given that this effect is scaled by the rat race parameter, $(\alpha + \beta)$, we refer to this as the rat race effect. Second, note that the benefit of entry for the average arbitrageur is $\Delta u\left(\frac{1}{2}\right) = M \cdot \alpha \cdot \int_{\frac{1}{2}}^1 \tilde{m}(\tilde{\theta}) d\tilde{\theta} - M \cdot \beta \cdot \int_0^{\frac{1}{2}} \tilde{m}(\tilde{\theta}) d\tilde{\theta}$, hence, keeping others' strategy constant

$$\frac{\partial \Delta u\left(\frac{1}{2}\right)}{\partial M} = \alpha \cdot \int_{\frac{1}{2}}^1 \tilde{m}(\tilde{\theta}) d\tilde{\theta} - \beta \cdot \int_0^{\frac{1}{2}} \tilde{m}(\tilde{\theta}) d\tilde{\theta} < (\alpha - \beta) \frac{p}{2} < 0$$

where the first inequality comes from the fact that $\tilde{m}(\theta)$ is decreasing in equilibrium. This suggests that for the average arbitrageur, increasing M is decreasing the incentive to enter. Given that this effect is scaled by the crowding parameter $(\alpha - \beta)$, we refer to this as the crowding effect. While in equilibrium the strategy of other arbitrageurs, $\tilde{m}(\tilde{\theta})$, also changes, implying further adjustments, as Figure 3 demonstrate, the total effect is still driven by this intuition.

Figure 3: **Competitive and social entry functions for different M**



Entry functions for the competitive entry (m , left panel) and the social planners entry function (m_s , right panel) for different levels of M . Parameters: $\beta = 4$, $\alpha = 2$, $\mu = 1$.

3.3 Allocative efficiency: over- and under-entry

Since the allocative efficiency of capital in the markets depends on the overall entry of all arbitrageurs, in this subsection we analyze how aggregate entry $M \cdot p$ changes as the mass of arbitrageurs M grows.

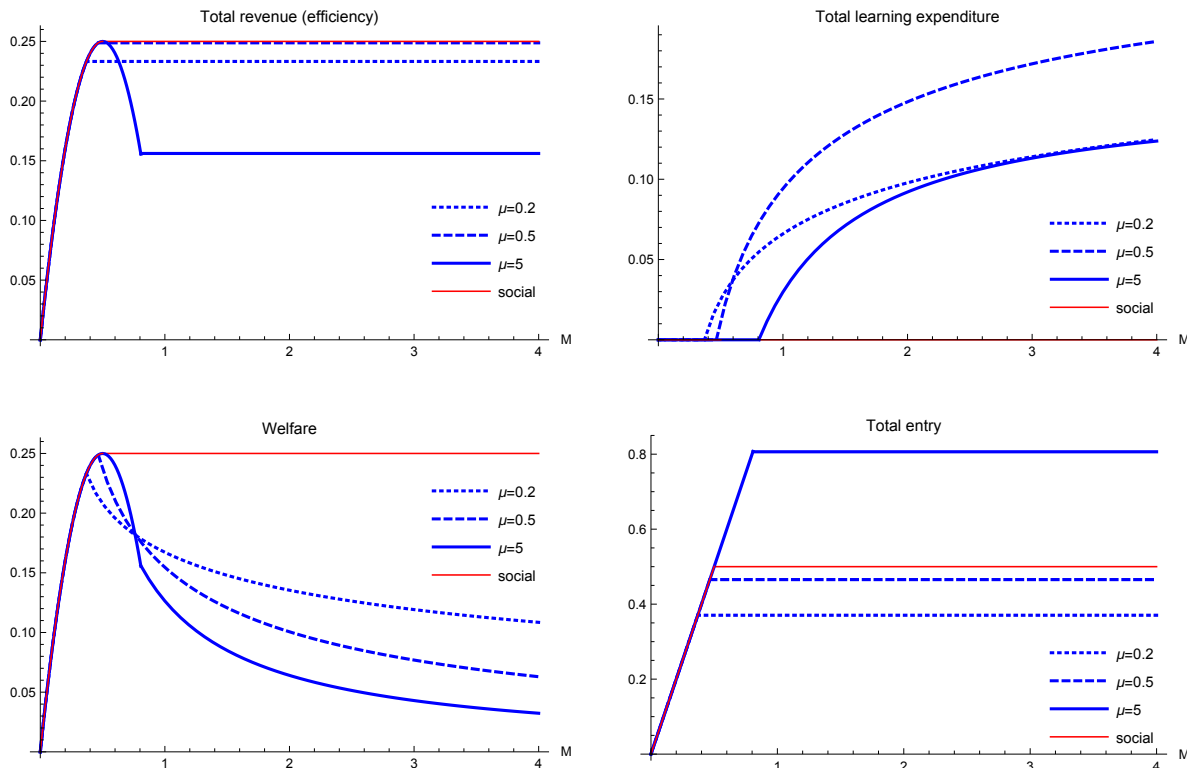
Proposition 3. *If $M \leq \bar{M}$, the competitive aggregate entry is $M \cdot p = M$. If $M > \bar{M}$, aggregate entry is constant in M at $M \cdot p = \bar{M}$. The aggregate entry in the social planner's solution is $M \cdot p_s = \min\left(M, \frac{1}{\beta - \alpha}\right)$.*

Just as in the planner's solution, arbitrageurs in the competitive equilibrium also enter with probability one for small M and aggregate entry $M \cdot p$ is constant when M is large. However, that constant level, \bar{M} , is different from the social optimum \bar{M}_s . That is, whenever $M > \bar{M}$, increasing the number of arbitrageurs neither improves the efficiency of capital allocation, nor does it lead to additional crowding. Figure 4 illustrates this part by showing the amount of total entry $M \cdot p$ as a function of the mass of arbitrageurs M .

For the intuition, recall that changing M changes the optimal strategy $m(\theta)$ for every arbitrageur through the rat race effect and the crowding effect. As the rat race effect primarily affects the slope of the entry function, as opposed to its level, it has little influence on $M \cdot p$. In contrast, due to the crowding effect the average entry, p , decreases. In equilibrium, the decrease in p is exactly proportional to the increase in M , keeping $M \cdot p$ constant.

Allowing arbitrageurs to flexibly choose their information structure is crucial in generating the result of constant entry as the mass of arbitrageurs increases. With flexible learning the arbitrageurs can optimally devise their information to exactly counter the increase in the mass of arbitrageurs and thus enter at a constant aggregate rate. When learning is constrained, this is not necessarily the case: we demonstrate this in Section 5 in which arbitrageurs can only buy Gaussian signals about their type subject to the same entropy cost as before. While the exact proportionality is also due to the properties of entropy, the choice of entropy as measuring the amount of learning has a sound axiomatic foundation.

Figure 4: Varying the number of arbitrageurs M



Outcomes as a function of the mass M of arbitrageurs allowed to invest. The thin solid line is the social optimum. Parameters: $\beta = 4$, $\alpha = 2$. μ takes three different values: $\mu = 0.2$ (dotted line), $\mu = 0.5$ (dashed line), $\mu = 5$ (solid line). Parameters: $\beta = 4$, $\alpha = 2$.

Proposition 3 shows that whether there is under- or over-entry for $M > \bar{M}$ is independent of the mass of arbitrageurs considering to enter. Instead, as we state in Proposition 4, it is determined by all the other characteristics of the capital reallocation problem. More frequent aggregate liquidity shocks (larger η) and less market depth (higher δ_c) make markets more crowded since they increase fire sales externalities. More costly information leads to more crowding, because the game is closer to a tragedy of commons problem as explained in Section 3.1. A slower decrease in marginal product of capital (higher δ) in the technology also makes the market more crowded in the presence of aggregate

shocks. On the other hand, more frequent idiosyncratic liquidity shocks (larger ν) makes the market less crowded since it leads to under-entry due to late entrants not internalizing the positive effect they have on earlier entrants.

Proposition 4. *If there are is a sufficient mass $M > \max[\bar{M}, \bar{M}_s]$ of arbitrageurs, the relative amount of competitive aggregate entry to social aggregate entry $\frac{\bar{M}}{\bar{M}_s}$ is*

1. *increasing in μ , the marginal cost of information*
2. *decreasing in δ , the rate of decreasing returns to scale of the technology, if $\eta > 0$*
3. *increasing in δ_c , i.e. decreasing in market depth in crisis, if $\eta > 0$*
4. *increasing in η , the probability of aggregate shocks*
5. *decreasing in ν , the probability of idiosyncratic shocks*

Note that the cost of learning μ does not only influence the amount of entry but also the welfare for a given mass M of arbitrageurs. Figure 4 gives us some insights: First, For high levels of M , easier learning (lower μ) means higher welfare W . This holds irrespective of the fact that a very low μ might lead to less entry than the social optimum. The reason is that with many arbitrageurs, they all have to spend an increasing fraction of their revenues on learning in order to stabilize entry and this is more costly if the marginal cost of learning μ is high. This also highlights that the possibility to learn is beneficial from a welfare point of view, especially if the cost is not that high. Second, for lower mass M of arbitrageurs, welfare might be higher when learning is more expensive. The intuition here is that for $\alpha > 0$, higher learning cost μ deters learning and thus helps avert under-entry.

3.4 Decoupling of welfare and allocative efficiency

Now we turn to the question of welfare as more and more sophisticated arbitrageurs enter. We show that the presence of some arbitrageurs ($M < \bar{M}$) unambiguously increases welfare in the competitive equilibrium. Note that in the small M case, welfare does not depend on μ since no resources are spent

on learning. The total mass of sophisticated investors is small in this range, hence they do not try to beat each other by learning about their relative type. Instead, all decide to enter without putting resources in learning. As their mass is marginally increasing, in terms of our microfoundation, they are able to allocate more capital to the new market, which increases the efficiency of capital allocation. Thus allocative efficiency and welfare go hand-in-hand. The above insight that larger M means (at least weakly) higher welfare and a more efficient capital allocation remains true in case of the social planner's optimum since no learning is chosen in that case.

In the competitive equilibrium, raising M above \bar{M} leads to decoupling of welfare and allocative (or market) efficiency: while allocative efficiency stays constant (even though at a suboptimal level), welfare decreases, see Figure 4. The reason is that as the amount of arbitrageurs in the market grows, these start worrying about crowding and thus their relative type θ , inducing them to learn about it. A rat race ensues with increasing amounts invested in learning costs and reduced welfare. Thus an increasing mass of sophisticated arbitrageurs leads to a drop in welfare not because of crowding (the total amount of arbitrageurs entering is constant) but because of increased spending on learning.

Proposition 5. *If $M > \bar{M}$, the efficiency of capital allocation (aggregate revenue of arbitrageurs) stays constant as we increase M . However, welfare becomes decoupled from allocative efficiency, welfare converges to zero from above as $M \rightarrow \infty$:*

$$W(\bar{M}) > W(M \rightarrow \infty) = 0 \tag{29}$$

In general, one can write:

$$W(M) = M \cdot \int_0^1 \log \left(\frac{1-p}{1-m(\theta)} \right) d\theta \tag{30}$$

The welfare in the social planner's optimum for $M > \bar{M}_s$ is constant:

$$W(M) = \frac{1}{2 \cdot (\beta - \alpha)}. \tag{31}$$

Note that learning is useful in limiting crowding, albeit at a cost. One can see this by comparing the above positive welfare to zero welfare for $M > \frac{2}{\beta-\alpha}$ in the case where learning is prohibitively expensive as $\mu \rightarrow \infty$. In the context of our structural model, the above result also implies that increasing the mass of arbitrageurs e.g. from $M < \bar{M}$ to $M > \bar{M}$ might make markets more efficient from an allocative point of view and decrease welfare at the same time. This is due to the fact that increased allocative efficiency is achieved at the cost of spending on learning.

If there are other ways to limit entry without learning, that might be welfare improving. In fact if we limit the mass of arbitrageurs indiscriminately before they learn their type one can improve welfare. The intuition is that limiting entry decreases the effective M and thus limits the incentive to learn, see the discussion following Proposition 2. Note that in a more general model where workers (or consumers) benefit from entry of arbitrageurs, the welfare analysis changes, see Section 6.4.

4 Heterogenous arbitrageurs

In this section we consider an extension with heterogenous arbitrageurs to analyze how changing the composition of arbitrageurs, instead of the total mass, influences allocative efficiency and welfare. This is interesting, as the level of sophistication among traders is very heterogenous: e.g. pension funds might be less sophisticated than hedge funds. We consider two groups: $\omega \cdot M$ mass of arbitrageurs is sophisticated and faces a lower learning cost of μ_L , while $(1 - \omega) \cdot M$ mass of arbitrageurs is unsophisticated and faces a higher learning cost of $\mu_H > \mu_L$. Both groups of arbitrageurs have types θ that are evenly distributed over $[0, 1]$. We consider the symmetric equilibrium in which sophisticated arbitrageurs choose the same entry strategy of $m_L(\theta)$, while unsophisticated arbitrageurs choose the same $m_H(\theta)$. To simplify the problem, we assume that the unsophisticated cannot learn at all, i.e. $\mu_H \rightarrow \infty$, resulting in a constant m_H in θ . Otherwise the solution would be a set of two joint differential equations which cannot be easily solved.⁷

⁷See the proof of Proposition 6 for the full problem.

Proposition 6. *If $\mu_H \rightarrow \infty$, the optimal interior solution for $m_L(\theta)$ and m_H is given by the following system of equations: The optimal strategy m_L of the sophisticated is given by*

$$\begin{aligned} \omega \cdot \log\left(\frac{(1-\omega) \cdot m_H + \omega \cdot m_L(\theta)}{m_L(\theta)}\right) + (1-\omega) \cdot m_H \cdot \log\left(\frac{1-m_L(\theta)}{m_L(\theta)}\right) - \frac{M \cdot (\alpha + \beta) \cdot (1-\omega) \cdot m_H \cdot ((1-\omega) \cdot m_H + \omega)}{\mu_L} \cdot \theta = \\ \omega \cdot \log\left(\frac{(1-\omega) \cdot m_H + \omega \cdot m_L(0)}{m_L(0)}\right) + (1-\omega) \cdot m_H \cdot \log\left(\frac{1-m_L(0)}{m_L(0)}\right) \end{aligned} \quad (32)$$

and m_H is given by the indifference condition of the unsophisticated

$$1 - M \cdot (\beta - \alpha) \cdot (1 - \omega) \cdot m_H + M \cdot \omega \cdot \int_0^1 \left[\alpha \cdot \int_{\theta}^1 m_L(\tilde{\theta}) d\tilde{\theta} - \beta \cdot \int_0^{\theta} m_L(\tilde{\theta}) d\tilde{\theta} \right] d\theta = 0 \quad (33)$$

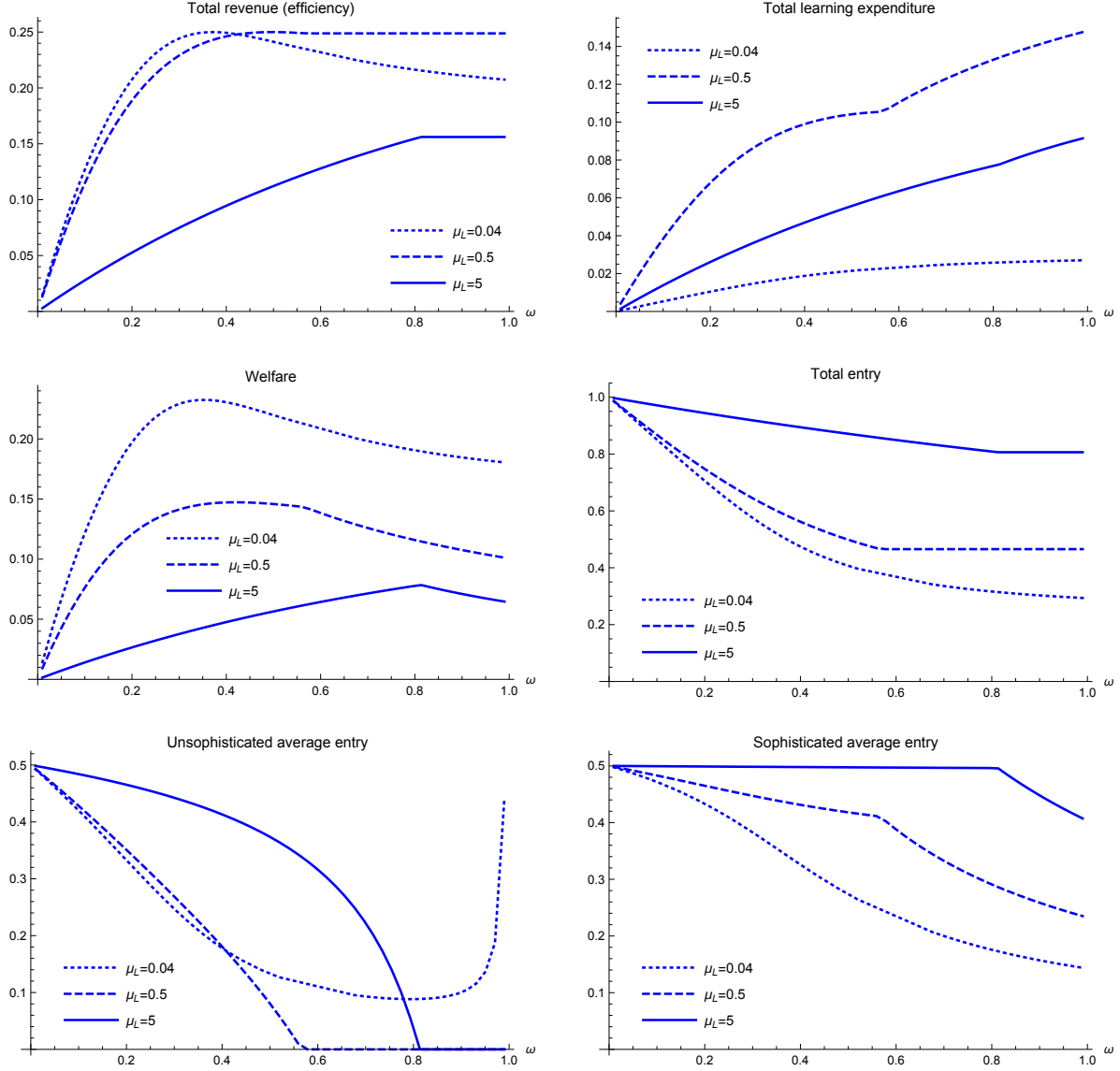
where $m_L(0)$ is pinned down by the boundary condition (the FOC of the sophisticated)

$$M \cdot \alpha \cdot [\omega \cdot p_L + (1 - \omega) \cdot m_H] + 1 = \mu_L \cdot \left[\log\left(\frac{m_L(0)}{1 - m_L(0)}\right) - \log\left(\frac{p_L}{1 - p_L}\right) \right] \quad (34)$$

and $p_L = \int_0^1 m_L(\tilde{\theta}) d\tilde{\theta}$ is the average entry of the sophisticated arbitrageur.

We solve the above set of equations numerically since it is analytically intractable. In Figure 5 we vary the portion ω of sophisticated arbitrageurs who can learn with cost μ_L . Thus the overall mass of arbitrageurs M is kept constant but a growing fraction of arbitrageurs is sophisticated. At $\omega = 0$ only unsophisticated are present and thus they enter until revenue is zero (given that M is large enough), yielding zero welfare. As ω initially increases, welfare increases since the average arbitrageur is more sophisticated. This is very similar to the result in the case of homogenous arbitrageurs that welfare increases as the average sophistication of arbitrageurs increases (i.e. as μ decreases). There are two effects leading to decreasing welfare as ω increases further. First, if the sophisticated are very sophisticated (low μ_L) then having lots of sophisticated leads to under-entry for $\alpha > 0$, thus decreasing welfare. This is like the case of homogenous agents where lowering μ is welfare reducing at low levels of μ since it aggravates under-entry. Second, and more interestingly, welfare can be decreasing in the share of sophisticated agents ω even for high μ_L in the absence of under-entry. The reason is that as

Figure 5: Changing the composition of heterogenous arbitrageurs



Here we change the portion ω of sophisticated arbitrageurs who can learn with low cost μ_L , while $1 - \omega$ cannot learn. The mass of arbitrageurs M is kept constant. Parameters: $\beta = 4$, $\alpha = 2$, $M = 2$. μ takes three different values: $\mu_L = 0.04$ (dotted line), $\mu_L = 0.5$ (dashed line), $\mu_L = 5$ (solid line). In all cases the social planner would allow each agent to enter with probability $\frac{1}{M \cdot (\beta - \alpha)} = \frac{1}{4}$, yielding total entry of $\frac{1}{\beta - \alpha} = \frac{1}{2}$.

ω increases, the unsophisticated are less likely to enter (m_H decreases) and above a threshold they are completely driven out of the market. Once the unsophisticated are not present, welfare is decreasing in ω . The intuition is similar to the baseline result in case of homogenous entrepreneurs that increasing the number of arbitrageurs M , welfare eventually decreases as arbitrageurs spend their revenue on learning.

Figure 5 also highlights the intricate interplay between the entry and learning strategies of the sophisticated and the unsophisticated. First, the remaining unsophisticated are less likely to enter as the fraction ω of sophisticated increases because there is more and more aggregate entry at low θ , cream-skimming the market and leaving less revenues for unsophisticated arbitrageurs who enter indiscriminately. Second, unsophisticated arbitrageurs are completely driven out of the market for high ω when sophisticated arbitrageurs are also not perfectly sophisticated (if $\mu_L > 0$), thus they are competitors of the unsophisticated, cannibalizing their revenues and eventually driving them out. The intuition is similar to that in high frequency trading where some agents may stay out of the market because they are afraid of very fast agents front-running them.

In fact, if the sophisticated arbitrageurs are sophisticated enough (μ_L close to zero), unsophisticated arbitrageurs will never be completely driven out of the market, see Figure 5. The reason is that perfectly sophisticated arbitrageurs follow cutoff strategies with the last entrant at the cutoff getting zero payoff and being indifferent (see $\mu_L = 0$ of the baseline model). If only sophisticated arbitrageurs are present in the market, then an unsophisticated arbitrageur with a uniform prior about its θ knows it can get positive payoff if its θ is smaller than the cutoff of the perfectly sophisticated arbitrageurs and gets zero payoff (equal to that of the last perfectly sophisticated to enter) with θ higher than the cutoff since there are no other entrants with higher θ in equilibrium. Figure 5 shows that in this case with ω close to one, all unsophisticated enter.

The above analysis also highlights how a not very well informed (unsophisticated) arbitrageur should behave if it learns about an arbitrage opportunity. It should enter with relatively high probability if it thinks arbitrageurs in the market are predominantly sophisticated but only if it believes

that the sophisticated arbitrageurs are very sophisticated. On the other hand, it should not enter at all, if it thinks the other sophisticated arbitrageurs are not extremely sophisticated. It may also choose to enter if it thinks that arbitrageurs are predominantly unsophisticated.

5 Gaussian learning

In this section, we investigate how our assumptions in the baseline analysis influence our main results. In particular, we contrast our framework with fully flexible learning with a, perhaps more standard, Gaussian formalization (see e.g. [Hellwig and Veldkamp \(2009\)](#)): suppose that each investor observes a Gaussian signal about its type θ of a chosen precision and enter if and only if this signal is larger than a chosen threshold. We show that as long as we specify the cost of learning analogously to our baseline model, this alternative structure amounts to a restriction on the functional form of $m(\theta)$. We refer to this as the Gaussian problem and show how this restriction affects the results. In the following, $\Phi(\cdot; \sigma)$ and $\phi(\cdot; \sigma)$ denote, respectively, the cdf and the pdf of a normally distributed variable with zero mean and σ standard deviation. $\Phi^{-1}(\cdot; \sigma)$ denotes the inverse of $\Phi(\cdot; \sigma)$.

First, we introduce the transformed type variable $\eta_i = \Phi^{-1}(\theta_i; \sigma_\eta)$. Clearly, as θ_i is uniform on the unit interval, $\eta_i \sim N(0, \sigma_\eta^2)$. Arbitrageur i with type η_i can, at a cost $C(\sigma_{\varepsilon_i})$, choose the standard deviation σ_{ε_i} of a signal $s_i = \eta_i + \varepsilon_i$ about its type where $\varepsilon_i \sim N(0, \sigma_{\varepsilon_i}^2)$ and ε_i is independent of η_i .⁸ After having received the signal s_i , arbitrageur i decides whether to enter.

As for the cost of learning function $C(\sigma_{\varepsilon_i})$, we use the reduction in entropy in order to focus on the effect of flexibility in learning (and thus entry). We assume the cost is identical to our baseline specification (11) with the only exception that the entry function $m(\theta)$ resulting from the entry decision based on the received signal might be restricted. Intuitively, we specify learning as a reduction

⁸This is equivalent to drawing the type and signal from a bivariate normal: $\begin{pmatrix} \eta_i \\ s_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\eta^2 & \sigma_\eta^2 \\ \sigma_\eta^2 & \sqrt{\sigma_\eta^2 + \sigma_{\varepsilon_i}^2} \end{pmatrix}\right)$.

in the entropy of agents' decision, which is not the same as the reduction in the entropy of their knowledge after observing the signal.⁹

Proposition 7. *In a symmetric equilibrium with Gaussian learning, the optimal strategy of the agent can be fully described by a choice of the signal noise σ_ε and the entry cutoff \bar{s} . The agent enters if and only if it receives a signal $s_i < \bar{s}$. The entry function has the form of*

$$m_G(\theta) = \Phi(\bar{s} - \Phi^{-1}(\theta; \sigma_\eta); \sigma_\varepsilon), \quad (35)$$

and the probability of unconditional entry is:

$$p = \Phi\left(\bar{s}; \sqrt{\sigma_\eta^2 + \sigma_\varepsilon^2}\right). \quad (36)$$

The equilibrium σ_ε and \bar{s} are pinned down by their respective first-order conditions (see Appendix).

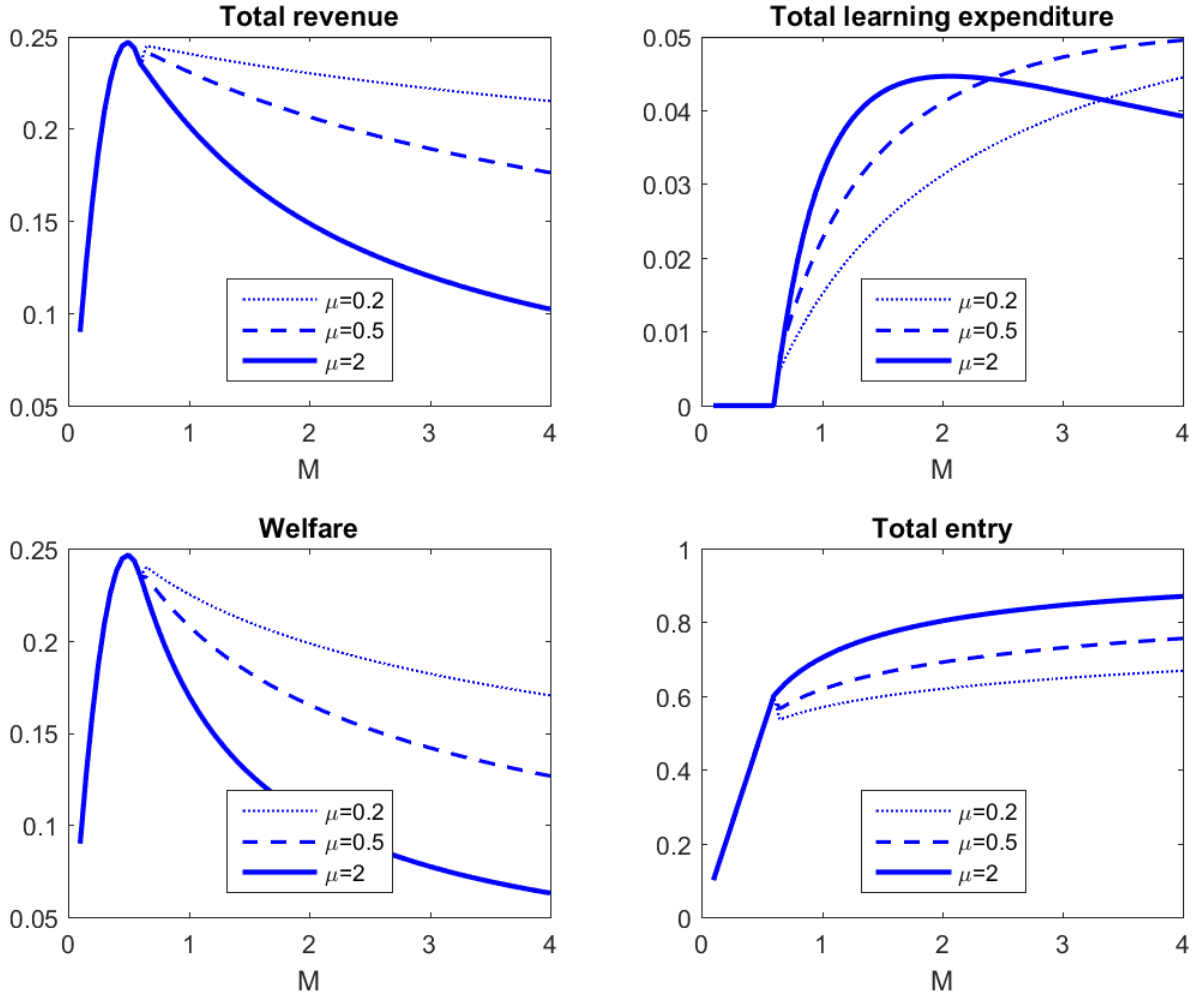
Note that $m_G(\cdot)$ can be fully described using two parameters: the standard deviation of signal noise σ_ε and the entry cutoff \bar{s} . $m_G(\cdot)$ is thus constrained compared to the completely unrestricted choice of $m(\cdot)$ in the baseline model (which yields a Lambert function). The entry function is constrained due to the fact that the signal structure is constrained and that the agent has to decide on entering based on this restricted signal. Thus constraining learning automatically means constraining the entry strategies.

Because a closed form solution is not attainable, we perform a numerical analysis. Figure 6 illustrates the outcome of the Gaussian problem. The outcomes are directly comparable to Figure 4.

Given our analogous specification of the cost function, all differences in the two sets of figures must come from the restriction on the form of the entry function (35) implied by the Gaussian specification.

⁹Another route would be to specify the cost of learning $C(\sigma_{\varepsilon_i})$ as the reduction in entropy in knowledge after the observation of the signal s_i , which, by the property of the normal distribution would give $C(\sigma_{\varepsilon_i}) = \frac{1}{2} \cdot \log\left(1 + \frac{\sigma_{\varepsilon_i}}{\sigma_\eta}\right)$. We do not follow this route in the main text, because that would introduce an additional departure from our baseline model. This specification would automatically imply too much learning as investors are using a continuous signal for a bivariate decision. However, we also worked out this case. It turns out that the results are very close both qualitatively and quantitatively to the ones we report in this subsection.

Figure 6: Varying the number of arbitrageurs M in the Gaussian specification

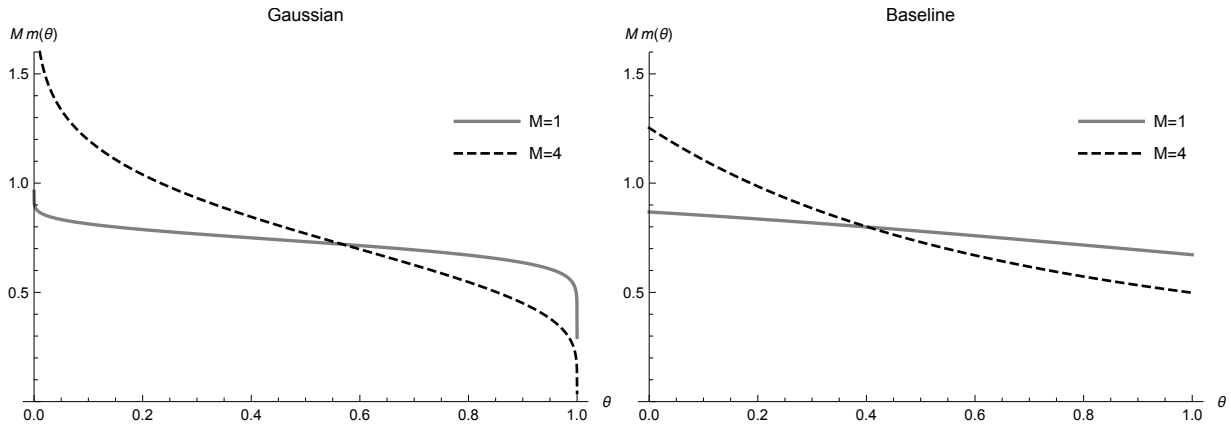


[FIGURE TO BE UPDATED, IT IS ONLY QUALITATIVELY CORRECT] Outcomes as a function of the mass M of arbitrageurs allowed to invest in the Gaussian specification. Parameters: $\beta = 4$, $\alpha = 2$. μ takes three different values: $\mu = 0.2$ (dotted line), $\mu = 0.5$ (dashed line), $\mu = 5$ (solid line). Parameters: $\beta = 4$, $\alpha = 2$.

First of all, observe that total entry is monotonically increasing as an increasing mass of arbitrageurs have the possibility to enter to the new market. As a result, total revenue, and consequently, the efficiency of capital allocation converges to zero. That is, a large mass of arbitrageurs always leads to crowded markets regardless of the parameter values in the Gaussian specification. In fact, as zero total revenue is the theoretical lower bound, the efficiency of capital allocation reaches its theoretical minimum as the mass of arbitrageurs with the possibility to enter grows without bound. Even though, in the social planner's solution, arbitrageurs in expectation are strictly better off due to the emergence

of the new market, in the decentralized solution, so many arbitrageurs decide to enter that the loss of the late ones is just about compensated by the gains of the early ones. This is in sharp contrast with the fully flexible learning case where, depending on the strength of external effects driven by α, β and μ , total entry might be too low or too large even when the mass of arbitrageurs present is very large. Thus while welfare goes to zero both in the baseline setup and the Gaussian setup, the reason is very different: in the baseline setup it is due to over-learning, in the Gaussian setup it is due to over-entry.

Figure 7: **Scaled entry $M \cdot m(\theta)$ under the Gaussian and the baseline specification**



Entry functions scaled by the mass of arbitrageurs allowed to invest $M \cdot m(\theta)$ in the Gaussian specification (left) and in the baseline case (right). Parameters: $\beta = 4$, $\alpha = 2$, $\mu = 4$. M takes two different values: $M = 1$ (solid line), $M = 4$ (dashed line).

To see the intuition why flexibility in entry (and thus learning) is critical, Figure 7 compares the entry functions at different M levels, multiplied by M under the Gaussian specification and under our baseline specifications. We graph $M \cdot m(\theta)$, so the areas under these functions are the total mass of entry in each of these cases. Both panels illustrate that the rat-race and the crowding effects are present. As M increases, regardless of the specification, agents differentiate their entry more w.r.t. θ by learning more, and they also enter less (larger M does not imply a proportional rise in $M \cdot m(\theta)$). In the baseline case, as we proved in Proposition 3, the two effects exactly cancel out in terms of their effect on total entry, i.e. the area below the $M \cdot m(\theta)$ curves do not change. However, in the Gaussian case, total entry clearly increases. This is because the restricted shape of the entry function, m_G , does not allow the agents to respond to both incentives optimally at the same time. So they neither

steepen their function nor reduce it sufficiently. This results in the over entry and the reduction in the efficiency of capital allocation.

To sum up, our exercise in this subsection emphasizes the importance of the flexible specification for entry and learning. While a large part of the literature uses Gaussian specifications for analytical convenience, we show, that, at least in our context, this simplification would fundamentally change our main results. While in our baseline model, when a larger mass of arbitrageurs is present, it neither increases nor decreases the level of entry and the efficiency of capital allocation, in the Gaussian specification it always leads to maximal crowding (i.e. the level of crowding under no information).

6 Further microfoundations

6.1 Consumer/producer surplus

Beyond arbitrageurs one might want to allow for other agents in the economy who benefit from arbitrage. Here we show a few examples where passive arbitrageurs (consumers or producers) benefit from entry and thus the welfare consequences of over-entry change. First we present a model of competition for scarce resources and then a model of publication tournaments. In Section 6.4 we then present the welfare implication of the presence of these arbitrageurs.

Assume there are some passive agents with a payoff (value) of

$$V^P \equiv \alpha_w \cdot \int_0^1 m(\theta) \cdot a(\theta) \cdot d\theta = \alpha_w \cdot \frac{(M \cdot p)^2}{2} \quad (37)$$

with some parameter $\alpha_w \geq 0$. Thus the overall welfare in the whole economy can be computed as:

$$W \equiv M \cdot V + V^P \quad (38)$$

One can interpret all our results up to now as a special case setting $\alpha_w = 0$. Thus the relevant payoff of entry from the social planner's perspective becomes:

$$\Delta u = (\alpha + \alpha_w) \cdot a(\theta) - \beta \cdot b(\theta) + 1 \quad (39)$$

Note that the arbitrageurs' competitive payoff function is unchanged, i.e. it does not include α_w .

6.2 Product market with scarce resources

Consider a single good traded in a competitive market produced by heterogenous producers that compete for scarce resources. Producing the good is also subject to local spill-overs. In this example the producers are the active agents making the entry decision while the consumers are passive.

The goods are purchased by a representative consumer with quadratic utility:

$$U(q) = q - \frac{\alpha_w}{2} q^2 \quad (40)$$

such that $MU = p$ yields linear demand for the good.

$$\text{price}(q) = 1 - \alpha_w \cdot q \quad (41)$$

A unit mass of firms indexed by θ decide to move to a specific area (e.g. the new Silicon Valley) where other firms might also move. Producers are heterogenous θ and lower θ producers have lower costs both because they have better technology and can also secure a better (heterogenous) input. The cost of building the plant is subject to weakly increasing marginal cost (and price of the building):

$$\text{cost}(\theta) = (\alpha + \beta)b(\theta) - (\alpha + \alpha_w)(a(\theta) + b(\theta)) \quad (42)$$

The $(\alpha + \beta)$ term multiplying $b(\theta)$ of the cost function comes from the price of the production input (e.g. land) which increases as more (and worse) type producers enter. Better types can choose a better quality input before the others at a fixed price (given by the value of external use). The $\alpha + \alpha_w$ term multiplying $(a(\theta) + b(\theta))$ captures network externalities that some resources might become cheaper if many producers use it because of economies of scale.

The payoff of firm θ conditional on moving is:

$$\Delta u = \text{price}(q) - \text{cost}(\theta) = 1 - \alpha_w \cdot q - (\alpha + \beta)b(\theta) + (\alpha + \alpha_w)(a(\theta) + b(\theta)) \quad (43)$$

Where in equilibrium the amount of goods produced depends on how many firm decide to produce: $q = a(\theta) + b(\theta)$ yielding the payoff function of Equation 1 in the reduced form.

The payoff of the passive agent in this example is the consumer surplus which can be computed as:

$$\text{CS} = \int_0^q (1 - \alpha_w \cdot \hat{q} - \text{price}(q)) d\hat{q} = \int_0^q \alpha_w \cdot (q - \hat{q}) d\hat{q} = \alpha_w \cdot \frac{q^2}{2} \quad (44)$$

A different way to compute the consumer surplus is to add a term of $\alpha_w \cdot a(\theta)$ (alternatively $\alpha_w \cdot b(\theta)$) to Δu

$$\text{CS} = \int_0^1 m(\theta) \cdot [\alpha_w \cdot a(\theta)] d\theta = -\frac{\alpha_w}{2} a(\theta)^2 \Big|_0^1 = \frac{\alpha_w}{2} a(1)^2 = \alpha_w \cdot \frac{q^2}{2} \quad (45)$$

This means that the objective function for the social planner maximizing welfare is the same as (1) substituting α with $\alpha + \alpha_w$ yielding (39).

6.3 Academic publications

Consider an academic tournament: e.g. the strategic choice of field of an aspiring academic. The academic wants to choose a topic that has not yet been done and that will have many followers who cite him. Both reading through and understanding the previous literature is time consuming and also trying to figure out whether others will find the same topic interesting and cite you. The

academic's payoff is the probability of publishing and being cited. θ can be interpreted as time. If lots of researchers finish their paper before (or other old but similar papers are discovered), the lower your chance of publication: this is captured by β . If lots of researchers write a paper on the same topic afterwards that increases citation and the academic's chance of publication, captured by α . Thus the payoff to the academic (producer of knowledge) is:

$$\Delta u = 1 - \beta \cdot b(\theta) + \alpha \cdot a(\theta) \quad (46)$$

where we assumed that the fixed payoff already incorporates the fixed cost of producing knowledge. This yields Equation 1 in the reduced form model. Adding a consumer of knowledge with quadratic utility:

$$U(q) = q - \frac{\alpha_w}{2} q^2 \quad (47)$$

yields a similar additional term in welfare as in Section 6.2.

6.4 Welfare and opaqueness

Thus crowding is not necessarily welfare decreasing if there is a consumer or producer in the economy who benefits from higher consumption or production. While there is over-entry from the arbitrageurs' point of view, from a welfare point of view this might be beneficial. Since $\alpha + \beta > 0$, opaqueness is never preferred if there is no passive arbitrageur ($\alpha_w = 0$).

Proposition 8. *Welfare is higher in case of no information than in case of full information if*

$$\alpha_w > \frac{(\beta - \alpha)^2(\alpha + \beta)}{\alpha(2\beta - \alpha)} \quad (48)$$

If the denominator is negative, it is true for all $\alpha_w \geq 0$.

7 Conclusions

We develop a model of capital reallocation to analyze whether the presence of more arbitrageurs improve capital allocation and welfare. Trades can become crowded due to imperfect information and externalities but arbitrageurs can devote resources to learn about the number of earlier entrants. In general, more arbitrageurs having the choice to enter into a trade neither improves the efficiency of capital allocation nor does it aggravate crowding. In fact, whether there is eventually too little or too much capital allocated to the new sector is determined solely by the technology in that sector, the cost of learning, the depth of the market, and the severity of the potential shocks, but not the mass of arbitrageurs present. However, the presence of more arbitrageurs decreases welfare, as they use more aggregate resources learning about each others' position.

References

- Abreu, Dilip and Markus K. Brunnermeier**, “Bubbles and Crashes,” *Econometrica*, 2003, *71* (1), 173–204.
- Biais, Bruno, Thierry Foucault, and Sophie Moinas**, “Equilibrium fast trading,” *Journal of Financial Economics*, 2015, *116* (2), 292 – 313.
- Budish, Eric, Peter Cramton, and John Shim**, “The high-frequency trading arms race: Frequent batch auctions as a market design response,” *Quarterly Journal of Economics*, 2015, *130* (4), 1547–1621.
- Duffie, Darrell**, “Asset Price Dynamics with Slow-Moving Capital,” *Journal of Finance*, 2010, *65*, 1238– 1268.
- and **Bruno Strulovici**, “Capital Mobility and Asset Pricing,” *Econometrica*, 2012, *80*, 2469–2509.
- Glode, Vincent, Richard C. Green, and Richard Lowery**, “Financial Expertise as an Arms Race,” *Journal of Finance*, 2012, *67* (5), 1723–1759.
- Greenwood, Robin, Samuel G. Hanson, and Gordon Y. Liao**, “Asset Price Dynamics in Partially Segmented Markets,” 2015. Harvard University Working Paper.
- Gromb, Denis and Dimitri Vayanos**, “Equilibrium and Welfare in Markets with Financially Constrained Arbitrageurs,” *Journal of Financial Economics*, 2002, *66* (2-3), 361–407.
- and — , “Limits of Arbitrage: The State of the Theory,” *Annual Review of Financial Economics*, 2010, *2*, 251–275.
- Grossman, Sanford J. and Joseph E. Stiglitz**, “On the impossibility of informationally efficient markets,” *American Economic Review*, 1980, *70* (3), 393–408.

- Hellwig, Christian and Laura Veldkamp**, “Knowing what others know: Coordination motives in information acquisition,” *Review of Economic Studies*, 2009, *76*, 223–251.
- Hirshleifer, Jack**, “The Private and Social Value of Information and the Reward to Inventive Activity,” *American Economic Review*, 1971, *61* (4), 561–574.
- Kacperczyk, Marcin, Stijn Van Nieuwerburgh, and Laura Veldkamp**, “A Rational Theory of Mutual Funds’ Attention Allocation,” *Econometrica*, *forthcoming*, 2015.
- Khandani, Amir E and Andrew W Lo**, “What happened to the quants in August 2007? Evidence from factors and transactions data,” *Journal of Financial Markets*, 2011, *14* (1), 1–46.
- Krueger, Anne O.**, “The Political Economy of the Rent-Seeking Society,” *American Economic Review*, 1974, *64* (3), pp. 291–303.
- Loury, Glenn C.**, “Market Structure and Innovation,” *Quarterly Journal of Economics*, 1979, *93* (3), 395–41.
- MacKay, David J.C.**, *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press, 2003.
- MacKenzie, Donald**, “Long-Term Capital Management and the sociology of arbitrage,” *Economy and Society*, 2003, *32* (3), 349–380.
- Maćkowiak, Bartosz and Mirko Wiederholt**, “Optimal Sticky Prices under Rational Inattention,” *American Economic Review*, 2009, *99* (3), 769–803.
- Matějka, Filip and Alisdair McKay**, “Foundation for the Multinomial Logit Model,” *American Economic Review*, 2015, *105* (1), 272–98.
- Moinas, Sophie and Sebastien Pouget**, “The Bubble Game: An Experimental Study of Speculation,” 2011. University of Toulouse working paper.

- Sims, Christopher A.**, “Stickiness,” *Carnegie-Rochester Conference Series on Public Policy*, December 1998, *49* (1), 317–356.
- , “Implications of Rational Inattention,” *Journal of Monetary Economics*, 2003, *50* (3), 665–90.
- Stein, Jeremy C.**, “Presidential Address: Sophisticated Investors and Market Efficiency,” *Journal of Finance*, 08 2009, *64* (4), 1517–1548.
- Sun, Zheng, Ashley Wang, and Lu Zheng**, “The Road Less Traveled: Strategy Distinctiveness and Hedge Fund Performance,” *Review of Financial Studies*, 2012, *25* (1), 96–143.
- Tullock, Gordon**, “The welfare costs of tariffs, monopolies, and theft,” *Economic Inquiry*, 1967, *5* (3), 224–232.
- Woodford, Michael**, “Inattention as a Source of Randomized Discrete Adjustment,” 2008. New York University working paper.
- Yang, Ming**, “Coordination with Rational Inattention,” *Journal of Economic Theory*, *forthcoming*, 2015.
- , “Optimality of Debt under Flexible Information Acquisition,” 2015. Duke University working paper.

A Proofs

Proof of Lemma 1

Proof. The proof follows that of Lemma 1 in Woodford (2008). □

Proof of Lemma 2

Proof. Under complete information, in the competitive optimum the last one to enter $\bar{\theta}$ is indifferent between entering and not:

$$-M \cdot \beta \cdot \bar{\theta} + 1 = 0 \tag{49}$$

yielding Eq. 17. For the social planner's optimum one has to find $\bar{\theta}$ that everyone with $\theta < \bar{\theta}$ enters, the others stay out, maximizing:

$$\int_0^{\bar{\theta}} ((M \cdot \alpha) \cdot (\bar{\theta} - \theta) - M \cdot \beta \cdot \theta + 1) d\theta = \frac{M \cdot \alpha - M \cdot \beta}{2} \cdot \bar{\theta}^2 + 1 \cdot \bar{\theta} \tag{50}$$

yielding the interior optimum in Eq. 18 if $M \cdot (\beta - \alpha) > 1$. If on the other hand, $M \cdot (\beta - \alpha) < 1$, everyone enters: $m(\theta) = 1$ is optimal.

Under no information, in the competitive equilibrium every arbitrageur enters with probability p and they are all indifferent given they do not know their θ and use a uniform prior. Expected payoff to entering:

$$\int_0^1 (M \cdot \alpha \cdot (1 - \theta) \cdot p - M \cdot \beta \cdot \theta \cdot p + 1) d\theta = 0 \tag{51}$$

yielding the unconditional entry probability in Eq. 19. If M is low and the implied entry is > 1 , then the revenue is not driven to zero and everyone enters for sure implying $p = 1$. In the social planner's optimum every arbitrageur enters with probability p and they maximize social planner's welfare

$$\int_0^1 p \cdot (M \cdot \alpha \cdot (1 - \theta) \cdot p - M \cdot \beta \cdot \theta \cdot p + 1) d\theta \tag{52}$$

taking derivative w.r.t. p and setting to zero, this implies the entry probability in Eq. 20. As before, if the implied entry probability is > 1 , then everyone enters for sure $m(\theta) = 1$ implying $p_s = 1$. Note that there are infinite other solutions since the social planner does not care about who exactly enters. □

Proof of Lemma 3

Proof. Note that in both benchmark cases $\mu \cdot L = 0$ thus $W = M \cdot R$. The values for R follow from straightforward algebra using the entry functions from Lemma 2. \square

Proof of Lemma 4

Proof. For the private FOC we use a perturbation method similar to the proof in Yang (2015a). In the first order perturbation we set $m(\theta) + \nu \cdot \epsilon(\theta)$ as $m(\theta)$, while we keep the entry decision of the others \tilde{m} fixed:

$$\int_0^1 ((m(\theta) + \nu \cdot \epsilon(\theta)) \cdot \Delta u(\tilde{m}, \theta) - \mu \cdot L(m(\theta) + \nu \cdot \epsilon(\theta))) d\theta. \quad (53)$$

We then take derivative wrt ν and then set $\nu = 0$ yielding the FOC:

$$\int_0^1 \epsilon(\theta) \cdot \left(\Delta u(\tilde{m}, \theta) - \mu \cdot \left[\log \left(\frac{m(\theta)}{1 - m(\theta)} \right) - \log \left(\frac{\int_0^1 m(\tilde{\theta}) d\tilde{\theta}}{1 - \int_0^1 m(\tilde{\theta}) d\tilde{\theta}} \right) \right] \right) d\theta = 0. \quad (54)$$

Since the original equation is an optimum, the above equality has to hold for any $\epsilon(\theta)$: thus the part multiplying $\epsilon(\theta)$ has to be zero for all θ . Setting $\tilde{m} = m$ we arrive at the symmetric solution we get (23).

For the social FOC we also use a perturbation method similar to the proof in Yang (2015a). In the first order perturbation we set $m_s(\theta) + \nu \cdot \epsilon(\theta)$ as $m_s(\theta)$, take derivative w.r.t. ν and then set $\nu = 0$ in order to arrive at the following equation that has to hold for any function $\epsilon(\theta)$:

$$\int_0^1 \epsilon(\theta) \cdot \left(M \cdot \alpha \cdot \int_{\theta}^1 m_s(\tilde{\theta}) d\tilde{\theta} - M \cdot \beta \cdot \int_0^{\theta} m_s(\tilde{\theta}) d\tilde{\theta} - \mu \cdot \left[\log \left(\frac{m_s(\theta)}{1 - m_s(\theta)} \right) - \log \left(\frac{\int_0^1 m_s(\tilde{\theta}) d\tilde{\theta}}{1 - \int_0^1 m_s(\tilde{\theta}) d\tilde{\theta}} \right) \right] \right) d\theta + \quad (55)$$

$$+ \int_0^1 m_s(\theta) \cdot \left(M \cdot \alpha \cdot \int_{\theta}^1 \epsilon(\tilde{\theta}) d\tilde{\theta} - M \cdot \beta \cdot \int_0^{\theta} \epsilon(\tilde{\theta}) d\tilde{\theta} \right) d\theta = 0 \quad (56)$$

We choose $\epsilon(\theta) = \delta_{\hat{\theta}}(\theta)$ where $\delta_{\hat{\theta}}$ is the Dirac-Delta function. Thus $\int_{\theta}^1 \epsilon(\tilde{\theta}) d\tilde{\theta} = \mathbf{1}_{\theta < \hat{\theta}}$ where $\mathbf{1}$ is the heaviside function. Substituting $\hat{\theta} = \theta$, the equation becomes:

$$M \cdot \alpha \cdot \int_{\theta}^1 m_s(\tilde{\theta}) d\tilde{\theta} - M \cdot \beta \cdot \int_0^{\theta} m_s(\tilde{\theta}) d\tilde{\theta} + 1 - \mu \cdot \left[\log \left(\frac{m_s(\theta)}{1 - m_s(\theta)} \right) - \log \left(\frac{\int_0^1 m_s(\tilde{\theta}) d\tilde{\theta}}{1 - \int_0^1 m_s(\tilde{\theta}) d\tilde{\theta}} \right) \right] + \quad (57)$$

$$+ M \cdot \alpha \cdot \int_0^{\theta} m_s(\tilde{\theta}) d\tilde{\theta} - M \cdot \beta \cdot \int_{\theta}^1 m_s(\tilde{\theta}) d\tilde{\theta} = 0 \quad (58)$$

which simplifies to (24). □

Proof of Proposition 1

Proof. Differentiating the FOC (23) we arrive at the following differential equation:

$$(M \cdot \alpha + M \cdot \beta) \cdot \tilde{m}(\theta) = -\frac{\mu \cdot m'(\theta)}{m(\theta) \cdot (1 - m(\theta))}. \quad (59)$$

thus the competitive equilibrium strategy $m(\theta)$ in the symmetric equilibrium ($m = \tilde{m}$) has to solve the above differential equation with the original FOC (e.g. evaluated at $\theta = 0$) as a boundary condition which is (26). If there is an interior solution (s.t. $m(\theta) \neq 1$), it can be written in the form

$$\frac{\frac{1}{m(\theta)} + \log\left(\frac{1-m(\theta)}{m(\theta)}\right)}{M(\alpha + \beta)} = C + \frac{\theta}{\mu} \quad (60)$$

for an appropriate constant C . Setting $\theta = 0$ above and subtracting from the above we can eliminate C and thus arrive at (61).

$$\frac{1}{m(\theta)} + \log\left(\frac{1-m(\theta)}{m(\theta)}\right) - \frac{M(\alpha + \beta)}{\mu} \cdot \theta = \frac{1}{m(0)} + \log\left(\frac{1-m(0)}{m(0)}\right), \quad (61)$$

Taking logs and using the definition of the Lambert function (upper branch if $z > 0$) yields (25).

To calculate the level of \bar{M} we use the observation (independently proven in Proposition 3) that $M \cdot p$ is constant, including in the limit as $M \rightarrow \infty$. At \bar{M} still all arbitrageurs enter with probability 1, thus $p = 1$ and \bar{M} can be expressed as:

$$\bar{M} = \lim_{\mu \rightarrow \infty} (M \cdot p) \quad (62)$$

Thus we focus on expressing $M \cdot p$ in the limit for large M . As a first step note that as $M \rightarrow \infty$, given that $M \cdot p$ is constant, $m(\theta) \rightarrow 0$ for every θ . Thus the implicit equation (61) for $m(\theta)$ can be approximated by

$$\frac{1}{m} - M(\alpha + \beta) \left(C + \frac{\theta}{\mu} \right) = 0 \quad (63)$$

since for $m \approx 0$: $\frac{1}{m} \gg \log\left(\frac{1}{m}\right)$. A closed form solution can be obtained in this limit case:

$$m(\theta) = \frac{\mu}{M(\alpha + \beta)(C\mu + \theta)} \quad (64)$$

for a specific C . By the definition of the average entry p this implies

$$M \cdot p = M \cdot \int_0^1 m(\theta) d\theta = \frac{\mu}{\alpha + \beta} \cdot \log\left(\frac{1}{C\mu} + 1\right). \quad (65)$$

Substituting this into the boundary condition (26) yields:

$$\alpha \frac{\mu}{\alpha + \beta} \log\left(\frac{1}{C\mu} + 1\right) + 1 = \mu \left[\log\left(\frac{1}{CM(\alpha + \beta) - 1}\right) - \log\left(\frac{M(\alpha + \beta) - \mu \log\left(\frac{1}{C\mu} + 1\right)}{\mu \log\left(\frac{1}{C\mu} + 1\right)}\right) \right] \quad (66)$$

Since $M \cdot p$ is a constant for any $M > \bar{M}$, C also has to converge to a finite constant as $M \rightarrow \infty$. Using this insight, one can take the limit of the above equation as $M \rightarrow \infty$:

$$\mu(-\alpha - \beta) \log\left(\frac{1}{C}\right) + (\alpha + \beta) \left(\mu \log\left(\mu \log\left(\frac{1}{C\mu} + 1\right)\right) + 1 \right) + \alpha \mu \log\left(\frac{1}{C\mu} + 1\right) = 0 \quad (67)$$

Using the relation between C and $M \cdot p$ in (65), one can eliminate C :

$$(\alpha + \beta) \left(\mu \log(M \cdot p \cdot (\alpha + \beta)) - \mu \log\left(\mu \left(e^{\frac{M \cdot p \cdot (\alpha + \beta)}{\mu}} - 1\right)\right) + \alpha M \cdot p + 1 \right) = 0 \quad (68)$$

using (62) and rearranging yields equation (27) in the proposition. \square

Proof of Proposition 2

Proof. The derivative of FOC (24) w.r.t. θ delivers the differential equation

$$0 = -\frac{\mu \cdot m'_s(\theta)}{m_s(\theta) \cdot (1 - m_s(\theta))} \quad (69)$$

subject to the boundary condition (setting $\theta = 0$ in (24))

$$M \cdot (\alpha - \beta) \cdot p_s + 1 = \mu \cdot \left[\log\left(\frac{m_s(0)}{1 - m_s(0)}\right) - \log\left(\frac{p_s}{1 - p_s}\right) \right]. \quad (70)$$

This trivially yields

$$m_s(\theta) = C \tag{71}$$

for some constant C , implying $p_s = C$. The boundary condition (70) simplifies to

$$M \cdot (\alpha - \beta)p_s + 1 = 0 \tag{72}$$

implying (28). If the implied entry probability is > 1 , then we have the corner solution that all enter with $m(\theta) = 1$. □

Proof of Proposition 3

Proof. To show that $M \cdot p$ is constant in M once the solution m is interior, first write the system of 3 equations determining p . First, the difference of FOC (23) at $\theta = 0$ and $\theta = 1$.

$$p = \frac{\mu \left(\log \left(\frac{m(0)}{1-m(0)} \right) - \log \left(\frac{m(1)}{1-m(1)} \right) \right)}{M(\alpha + \beta)} \tag{73}$$

Second, the boundary condition (23) at $\theta = 0$

$$\alpha Mp + 1 = \mu \left(\log \left(\frac{m(0)}{1-m(0)} \right) - \log \left(\frac{p}{1-p} \right) \right). \tag{74}$$

Third, the implicit equation for $m(\theta)$ evaluated at $\theta = 1$.

$$\log \left(\frac{m(0)}{1-m(0)} \right) - \log \left(\frac{m(1)}{1-m(1)} \right) = \frac{M(\alpha + \beta)}{\mu} + \frac{1}{m(0)} - \frac{1}{m(1)} \tag{75}$$

Substituting

$$x_0 = \log \left(\frac{m(0)}{1-m(0)} \right) \tag{76}$$

and

$$x_1 = \log \left(\frac{m(1)}{1-m(1)} \right) \tag{77}$$

the system of three equations can be written as:

$$p = \frac{\mu(x_0 - x_1)}{M(\alpha + \beta)} \quad (78)$$

$$\alpha Mp + 1 = \mu \left(x_0 - \log \left(\frac{p}{1-p} \right) \right) \quad (79)$$

$$x_0 - x_1 = \frac{M(\alpha + \beta)}{\mu} + e^{-x_0} - e^{-x_1} \quad (80)$$

Substituting out p from (78), (79), (80) we arrive at a system of two equations:

$$F = \mu \left(x_0 - \log \left(\frac{\mu(x_0 - x_1)}{M(\alpha + \beta) + \mu(x_1 - x_0)} \right) \right) - \left(\frac{\alpha\mu(x_0 - x_1)}{\alpha + \beta} + 1 \right) = 0 \quad (81)$$

$$G = \frac{M(\alpha + \beta)}{\mu} - (x_0 - x_1) + e^{-x_0} - e^{-x_1} = 0 \quad (82)$$

To prove $M \cdot p$ is constant, it is sufficient to prove $\frac{\partial(M \cdot p)}{\partial M} = 0$ which from (78) is equivalent to

$$\frac{\partial x_0}{\partial M} = \frac{\partial x_1}{\partial M} \quad (83)$$

We apply Cramer's rule both for x_0 and x_1 to the system of equations (81) and (82):

$$\frac{\partial x_0}{\partial M} = \frac{\begin{vmatrix} \frac{\partial F}{\partial x_0} & -\frac{\partial F}{\partial M} \\ \frac{\partial G}{\partial x_0} & -\frac{\partial G}{\partial M} \end{vmatrix}}{\begin{vmatrix} \frac{\partial F}{\partial x_0} & \frac{\partial F}{\partial x_1} \\ \frac{\partial G}{\partial x_0} & \frac{\partial G}{\partial x_1} \end{vmatrix}} \quad (84)$$

$$\frac{\partial x_1}{\partial M} = \frac{\begin{vmatrix} -\frac{\partial F}{\partial M} & \frac{\partial F}{\partial x_1} \\ -\frac{\partial G}{\partial M} & \frac{\partial G}{\partial x_1} \end{vmatrix}}{\begin{vmatrix} \frac{\partial F}{\partial x_0} & \frac{\partial F}{\partial x_1} \\ \frac{\partial G}{\partial x_0} & \frac{\partial G}{\partial x_1} \end{vmatrix}} \quad (85)$$

We check numerically that the denominator (which is the same for both derivatives) is non-zero thus the two equations are indeed independent. It is thus sufficient to show that the numerators of the Cramer rule for

the two derivatives are equal, yielding the sufficient condition

$$\frac{(\alpha + \beta)e^{-x_0 - x_1} (e^{x_0 + x_1} (M(\alpha + \beta) + \mu(x_1 - x_0)) - \mu e^{x_0} + \mu e^{x_1})}{M(\alpha + \beta) + \mu(x_1 - x_0)} = 0. \quad (86)$$

It follows from (80) that the denominator is non-zero if $x_0 \neq x_1$. Thus it is sufficient to prove that

$$\frac{M(\alpha + \beta)}{\mu} + (x_1 - x_0) + \frac{1}{e^{x_0}} - \frac{1}{e^{x_1}} = 0, \quad (87)$$

which is exactly the function $G = 0$ defined in (82). Thus the identity holds and we have proved that $M \cdot p$ is constant in M for interior solutions. □

Proof of Proposition 4

Proof. \bar{M} is defined by the implicit equation 27: $F = \frac{\bar{M} \cdot (\alpha + \beta)}{\mu} - e^{-\frac{1 - \beta \cdot \bar{M}}{\mu}} + e^{-\frac{1 + \alpha \cdot \bar{M}}{\mu}} = 0$. Using the result from Eq. 28 that $\bar{M}_s = \frac{1}{\beta - \alpha}$ and the implicit function theorem, the derivative of interest $\frac{\partial \bar{M}}{\partial \cdot}$ for any parameter “.” becomes:

$$\frac{\partial \bar{M}}{\partial \cdot} = \frac{\partial \bar{M}}{\partial \cdot} \cdot (\beta - \alpha) + \frac{\partial(\beta - \alpha)}{\partial \cdot} \cdot \bar{M} = -\frac{\frac{\partial F}{\partial \cdot}}{\frac{\partial F}{\partial \bar{M}}} \cdot (\beta - \alpha) + \frac{\partial(\beta - \alpha)}{\partial \cdot} \cdot \bar{M} \quad (88)$$

Denote

$$A = e^{\frac{\bar{M}(\alpha + \beta)}{\mu}} - 2e^{\frac{\alpha \bar{M} + 1}{\mu}} + 1 \quad (89)$$

and

$$B = \beta \left(e^{\frac{\alpha \bar{M} + 1}{\mu}} - e^{\frac{\bar{M}(\alpha + \beta)}{\mu}} \right) + \alpha \left(e^{\frac{\alpha \bar{M} + 1}{\mu}} - 1 \right). \quad (90)$$

Basic algebra yields:

$$\frac{\partial \bar{M}}{\partial \beta} = -\frac{A}{B} \cdot \alpha \cdot \bar{M} < 0 \quad (91)$$

$$\frac{\partial \bar{M}}{\partial \alpha} = \frac{A}{B} \cdot \beta \cdot \bar{M} > 0 \quad (92)$$

$$\frac{\partial \bar{M}}{\partial \mu} = \frac{B \cdot \bar{M} + e^{\frac{\bar{M}(\alpha + \beta)}{\mu}} - 1}{B} \cdot \frac{\beta - \alpha}{\mu} > 0 \quad (93)$$

We show below that $A > 0$, $B < 0$ and $\frac{B \cdot \bar{M} + e^{\frac{\bar{M}(\alpha + \beta)}{\mu}} - 1}{B > 0} > 0$. The last expression proves the first part of the Proposition. For the other parts, we use the total derivative to get the effect of the parameters of our full

model:

$$\frac{\partial \frac{\bar{M}}{M_s}}{\partial \delta} = \frac{\partial \frac{\bar{M}}{M_s}}{\partial \beta} \cdot \frac{\partial \beta}{\partial \delta} + \frac{\partial \frac{\bar{M}}{M_s}}{\partial \alpha} \cdot \frac{\partial \alpha}{\partial \delta} = - \left(-\frac{A}{B} \cdot \bar{M} \right) \cdot \frac{1}{2} \delta_c \cdot \eta \cdot (1 - \nu)^2 \cdot ((1 - \eta)(1 - \nu)^2 + 1) \leq 0 \quad (94)$$

$$\frac{\partial \frac{\bar{M}}{M_s}}{\partial \delta_c} = \frac{\partial \frac{\bar{M}}{M_s}}{\partial \beta} \cdot \frac{\partial \beta}{\partial \delta_c} + \frac{\partial \frac{\bar{M}}{M_s}}{\partial \alpha} \cdot \frac{\partial \alpha}{\partial \delta_c} = \left(-\frac{A}{B} \cdot \bar{M} \right) \cdot \frac{1}{2} \delta \cdot \eta \cdot (1 - \nu)^2 \cdot ((1 - \eta)(1 - \nu)^2 + 1) \geq 0 \quad (95)$$

with equality if and only if $\eta = 0$. Furthermore,

$$\frac{\partial \frac{\bar{M}}{M_s}}{\partial \eta} = \frac{\partial \frac{\bar{M}}{M_s}}{\partial \beta} \cdot \frac{\partial \beta}{\partial \eta} + \frac{\partial \frac{\bar{M}}{M_s}}{\partial \alpha} \cdot \frac{\partial \alpha}{\partial \eta} = \left(-\frac{A}{B} \cdot \bar{M} \right) \cdot \frac{1}{2} \delta \cdot (1 - \nu)^2 \cdot ((1 - \nu)^2(2\delta + \delta_c) + \delta_c) > 0 \quad (96)$$

$$\frac{\partial \frac{\bar{M}}{M_s}}{\partial \nu} = \frac{\partial \frac{\bar{M}}{M_s}}{\partial \beta} \cdot \frac{\partial \beta}{\partial \nu} + \frac{\partial \frac{\bar{M}}{M_s}}{\partial \alpha} \cdot \frac{\partial \alpha}{\partial \nu} = - \left(-\frac{A}{B} \cdot \bar{M} \right) \cdot \delta \cdot (1 - \nu) \cdot (2\delta + \delta_c \eta) < 0 \quad (97)$$

where we used $\delta > 0$, $A > 0$, $B < 0$ and $\bar{M} > 0$.

We confirm numerically that $A > 0$, $B < 0$, and $B \cdot \bar{M} + e^{\frac{M(\alpha+\beta)}{\mu}} - 1 < 0$ for all admissible values of α, β, μ .

□

Proof of Proposition 5

Proof. By Proposition 1 when $M < \bar{M}$ all arbitrageurs enter with probability 1. Hence, all equilibrium objects are the same for the planner and in the decentralized solution. In particular, average entry of an arbitrageur is $p = 1$ thus expected aggregate entry is M . Total revenue and welfare are

$$M \cdot R = W = M \cdot R_s = W_s = M \cdot \int_0^1 (M \cdot \alpha \cdot (1 - \theta) - M \cdot \beta \cdot \theta + 1) d\theta = M - \frac{M^2(\beta - \alpha)}{2}. \quad (98)$$

To arrive at the formula for $W(M)$ one can rearrange the aggregate learning from (11) to get:

$$M \cdot L = M \int_0^1 m(\theta) \cdot \mu \cdot \left(\log \left(\frac{m(\theta)}{1 - m(\theta)} \right) - \log \left(\frac{p}{1 - p} \right) \right) \cdot d\theta - M \int_0^1 \mu \log \left(\frac{1 - p}{1 - m(\theta)} \right) \cdot d\theta \quad (99)$$

where the interior part of the first integral multiplying $m(\theta)$ can be replaced using the FOC (23) to yield:

$$M \cdot L = M \int_0^1 m(\theta) \cdot \left[M \cdot \alpha \cdot \int_{\theta}^1 \tilde{m}(\tilde{\theta}) d\tilde{\theta} - M \cdot \beta \cdot \int_0^{\theta} \tilde{m}(\tilde{\theta}) d\tilde{\theta} + 1 \right] d\theta - M \int_0^1 \mu \log \left(\frac{1 - p}{1 - m(\theta)} \right) \cdot d\theta \quad (100)$$

thus the first integral is exactly the definition of aggregate revenue. Since $M \cdot p$ is constant if $M \geq \bar{M}$ (Proposition 3), so is aggregate revenue $M \cdot R$. Rearranging yields expression (30) for W in the proposition.

We now show that welfare converges to zero for large M . For large M , $m \approx 0$ and $p \approx 0$ thus in the $M \rightarrow \infty$ limit (30) converges to zero. This convergence happens from above, since the payoff per arbitrageur $\frac{W}{M}$ cannot be negative, otherwise arbitrageurs would choose not to enter.

In the social planner's interior optimum every arbitrageur enters with probability $p = \frac{1}{M \cdot (\beta - \alpha)}$ and thus welfare becomes

$$W_s = M \cdot \int_0^1 p \cdot (M \cdot \alpha \cdot (1 - \theta) \cdot p - M \cdot \beta \cdot \theta \cdot p + 1) d\theta = \frac{1}{2 \cdot (\beta - \alpha)}. \quad (101)$$

□

Proof of Proposition 6

Proof. We first set up the problem for general μ_d before setting the special case of $\mu_d \rightarrow \infty$. In equilibrium, the mass of lower types entering ("before" arbitrageur θ) becomes:

$$b(\theta) = M \cdot \int_0^\theta \omega \cdot m_c(\tilde{\theta}) + (1 - \omega) \cdot m_d(\tilde{\theta}) d\tilde{\theta} \quad (102)$$

the mass of higher types entering ("after" arbitrageur θ):

$$a(\theta) = M \cdot \int_\theta^1 \omega \cdot m_c(\tilde{\theta}) + (1 - \omega) \cdot m_d(\tilde{\theta}) d\tilde{\theta}. \quad (103)$$

Thus the problem the arbitrageurs solve is the joint maximization of two equations $i \in [d, c]$:

$$\max_{m_i(\theta)} \int_0^1 (m_i(\theta) \cdot \Delta u(\theta) - \mu_i \cdot L(m_i)) d\theta. \quad (104)$$

The optimal solution is characterized by the differential equation for m_c

$$(\alpha + \beta) (\omega \tilde{m}_c(\theta) + (1 - \omega) \tilde{m}_d(\theta)) = - \frac{\mu_c m'_c(\theta)}{m_c(\theta) (1 - m_c(\theta))} \quad (105)$$

with the boundary condition:

$$M \cdot \alpha \cdot \int_{\theta}^1 (\omega \cdot \tilde{m}_c(\tilde{\theta}) + (1 - \omega) \cdot \tilde{m}_d(\tilde{\theta})) d\tilde{\theta} - M \cdot \beta \cdot \int_0^{\theta} (\omega \cdot \tilde{m}_c(\tilde{\theta}) + (1 - \omega) \cdot \tilde{m}_d(\tilde{\theta})) d\tilde{\theta} + 1 = \mu_c \cdot \left[\log \left(\frac{m_c(\theta)}{1 - m_c(\theta)} \right) - \log \left(\frac{p_c}{1 - p_c} \right) \right]. \quad (106)$$

Where $p_s = \int_0^1 m_c(\tilde{\theta}) d\tilde{\theta}$ is the average entry of the sophisticated. A symmetric set of equations hold for m_d which we omit for brevity. In equilibrium $\tilde{m}_d = m_d$ and $\tilde{m}_c = m_c$. In general such systems of interlinked differential equations for $m_d(\theta)$ and $m_c(\theta)$ cannot be solved, thus we set $\mu_d \rightarrow \infty$. This means that $m_d(\theta) \equiv m_d$ is a constant chosen such that the revenue of the dumb is exactly zero for interior solutions.

$$M \cdot \alpha \cdot \int_{\theta}^1 (\omega \cdot \tilde{m}_c(\tilde{\theta}) + (1 - \omega) \cdot \tilde{m}_d(\tilde{\theta})) d\tilde{\theta} - M \cdot \beta \cdot \int_0^{\theta} (\omega \cdot \tilde{m}_c(\tilde{\theta}) + (1 - \omega) \cdot \tilde{m}_d(\tilde{\theta})) d\tilde{\theta} + 1 = 0 \quad (107)$$

$m_d = 0$ is chosen if the left hand side of the above equation is negative in such an equilibrium and $m_d = 1$ is chosen if it is positive. Equation 106 can be solved in implicit form and yields Equation in the proposition (evaluated at $\theta = 0$ to substitute out the constant). Equation 106 evaluated at $\theta = 0$ yields the equation for $m_c(0)$ given in Equation 34 in the Proposition. Equation 107 can be simplified to yield 33 in the Proposition. \square

Proof of Proposition 7

Proof. First, note that the expected payoff is monotonously decreasing in either η_i or θ_i . Also, if $s_i > s_j$ then $g(\eta_i | s_i; \sigma_\eta, \sigma_{\varepsilon_i})$ first order stochastically dominates $g(\eta_j | s_j; \sigma_\eta, \sigma_{\varepsilon_i})$ where $g(\cdot | \cdot)$ is the distribution of η_i conditional on the signal s_i . Note that a lower signal means a higher expected payoff. Thus the unique optimal strategy of arbitrageur i is to enter based on a cutoff \bar{s}_i , entering whenever $s_i < \bar{s}_i$. Thus, conjecturing that the choice of σ_{ε_i} and \bar{s}_i are symmetric and dropping the subscripts, each agent solves

$$\max_{\sigma_\varepsilon, \bar{s}} \int_{-\infty}^{\infty} \left[1 + \alpha M \int_{\eta}^{\infty} f(s < \tilde{s} | \eta'; \tilde{\sigma}_\varepsilon, \sigma_\eta) \phi(\eta'; \sigma_\eta) d\eta' - \beta M \int_{-\infty}^{\eta} f(s < \tilde{s} | \eta'; \tilde{\sigma}_\varepsilon, \sigma_\eta) \phi(\eta'; \sigma_\eta) d\eta' \right] \cdot f(s < \bar{s} | \eta; \sigma_\varepsilon, \sigma_\eta) \phi(\eta; \sigma_\eta) d\eta - C(\sigma_\varepsilon) \quad (108)$$

where the tilde denotes the choice of others, and $f(s < \bar{s} | \eta; \sigma_\varepsilon, \sigma_\eta)$ is the conditional probability of $s < \bar{s}$ given η with choice σ_ε .

Note that we can rewrite the expected revenue in (108) as follows

$$\begin{aligned}
& \int_{-\infty}^{\infty} \left(1 + \alpha M \int_{\eta}^{\infty} f(s < \tilde{s}|\eta'; \tilde{\sigma}_{\varepsilon}, \sigma_{\eta}) \phi(\eta'; \sigma_{\eta}) d\eta' - \beta M \int_{-\infty}^{\eta} f(s < \tilde{s}|\eta'; \tilde{\sigma}_{\varepsilon}, \sigma_{\eta}) \phi(\eta'; \sigma_{\eta}) d\eta' \right) \cdot \\
& \quad f(s < \bar{s}|\eta; \sigma_{\varepsilon}, \sigma_{\eta}) \phi(\eta; \sigma_{\eta}) d\eta = \\
& = \int_{-\infty}^{\infty} \left[1 + \alpha M \int_{\eta}^{\infty} \Phi(\tilde{s} - \eta'; \tilde{\sigma}_{\varepsilon}) \phi(\eta'; \sigma_{\eta}) d\eta' - \beta M \int_{-\infty}^{\eta} \Phi(\tilde{s} - \eta'; \tilde{\sigma}_{\varepsilon}) \phi(\eta'; \sigma_{\eta}) d\eta' \right] \cdot \\
& \quad \Phi(\bar{s} - \eta; \sigma_{\varepsilon}) \phi(\eta; \sigma_{\eta}) d\eta = \\
& = \int_0^1 \left[1 + \alpha M \underbrace{\int_{\theta}^1 \Phi(\tilde{s} - \Phi^{-1}(\theta'; \sigma_{\eta}); \tilde{\sigma}_{\varepsilon}) d\theta'}_{a(\theta)} - \beta M \underbrace{\int_0^{\theta} \Phi(\tilde{s} - \Phi^{-1}(\theta'; \sigma_{\eta}); \tilde{\sigma}_{\varepsilon}) d\theta'}_{b(\theta)} \right] \cdot \\
& \quad \Phi(\bar{s} - \Phi^{-1}(\theta; \sigma_{\eta}); \sigma_{\varepsilon}) d\theta. \quad (109)
\end{aligned}$$

In the first equation we used that $f(s < \bar{s}|\eta; \sigma_{\varepsilon}, \sigma_{\eta}) = \Phi(\bar{s} - \eta; \sigma_{\varepsilon})$ based on our assumptions, while in the second equation we used the rule of integration by substitution to replace η with θ .

Note that the last equation has the same form as the expected revenue in our baseline model with the restriction that in the Gaussian problem the entry function $m(\theta)$ is restricted to have the form of

$$m_G(\theta) = \Phi(\bar{s} - \Phi^{-1}(\theta; \sigma_{\eta}); \sigma_{\varepsilon}), \quad (110)$$

The probability of unconditional entry is $p = \Phi(\bar{s}; \sqrt{\sigma_{\eta}^2 + \sigma_{\varepsilon}^2})$ because the standard deviation of the signal is $\sqrt{\sigma_{\eta}^2 + \sigma_{\varepsilon}^2}$ due to the independence of η_i and ε_i .

The cost of learning function $C(\sigma_{\varepsilon_i})$, is:

$$\begin{aligned}
& C(\sigma_{\varepsilon_i}) = \\
& \mu \cdot \left[\left(p \log \left[\frac{1}{p} \right] + (1-p) \log \left[\frac{1}{1-p} \right] \right) - \int_0^1 \left(m_G(\theta) \log \left[\frac{1}{m_G(\theta)} \right] + (1-m_G(\theta)) \log \left[\frac{1}{1-m_G(\theta)} \right] \right) d\theta \right] \\
& \quad (111)
\end{aligned}$$

The solution can be obtained by setting marginal cost and benefit of both parameters σ_{ε} and \bar{s} equal while keeping the others' choice constant and then imposing symmetry, such that $\tilde{\sigma}_{\varepsilon} = \sigma_{\varepsilon}$ and $\tilde{s} = \bar{s}$. this implies the

two first order conditions:

$$\begin{aligned} \frac{\partial C}{\partial \bar{s}} = \mu \cdot \left[\log \left(\frac{1-p}{p} \right) \frac{e^{-\frac{\bar{s}^2}{2(\sigma_\eta^2 + \sigma_\varepsilon^2)}}}{\sqrt{2\pi} \sqrt{\sigma_\eta^2 + \sigma_\varepsilon^2}} - \int_0^1 \log \left(\frac{1 - m_G(\theta)}{m_G(\theta)} \right) \frac{e^{-\frac{(\bar{s} - \Phi^{-1}(\theta))^2}{2\sigma_\varepsilon^2}}}{\sqrt{2\pi} \sigma_\varepsilon} d\theta \right] = \\ \int_0^1 (1 + \alpha \cdot M \cdot a(\theta) - \beta \cdot M \cdot b(\theta)) \frac{e^{-\frac{(\bar{s} - \Phi^{-1}(\theta))^2}{2\sigma_\varepsilon^2}}}{\sqrt{2\pi} \sigma_\varepsilon} d\theta = \frac{\partial R}{\partial \bar{s}} \quad (112) \end{aligned}$$

$$\begin{aligned} \frac{\partial C}{\partial \sigma_\varepsilon} = \mu \cdot \left[-\log \left(\frac{1-p}{p} \right) \frac{e^{-\frac{\bar{s}^2}{2(\sigma_\eta^2 + \sigma_\varepsilon^2)}} \cdot \sigma_\varepsilon}{\sqrt{2\pi} (\sigma_\eta^2 + \sigma_\varepsilon^2)^{\frac{3}{2}}} + \int_0^1 \log \left(\frac{1 - m_G(\theta)}{m_G(\theta)} \right) \frac{e^{-\frac{(\bar{s} - \Phi^{-1}(\theta))^2}{2\sigma_\varepsilon^2}}}{\sqrt{2\pi} \sigma_\varepsilon^2} \cdot (\bar{s} - \Phi^{-1}(\theta)) d\theta \right] = \\ - \int_0^1 (1 + \alpha \cdot M \cdot a(\theta) - \beta \cdot M \cdot b(\theta)) \frac{e^{-\frac{(\bar{s} - \Phi^{-1}(\theta))^2}{2\sigma_\varepsilon^2}} \cdot (\bar{s} - \Phi^{-1}(\theta))}{\sqrt{2\pi} \sigma_\varepsilon^2} d\theta = \frac{\partial R}{\partial \sigma_\varepsilon} \quad (113) \end{aligned}$$

That is one has to find σ_ε and \bar{s} that jointly solves:

$$\begin{aligned} \mu \cdot \left[\log \left(\frac{1 - \Phi(\bar{s}; \sqrt{\sigma_\eta^2 + \sigma_\varepsilon^2})}{\Phi(\bar{s}; \sqrt{\sigma_\eta^2 + \sigma_\varepsilon^2})} \right) \frac{e^{-\frac{\bar{s}^2}{2(\sigma_\eta^2 + \sigma_\varepsilon^2)}}}{\sqrt{2\pi} \sqrt{\sigma_\eta^2 + \sigma_\varepsilon^2}} - \int_0^1 \log \left(\frac{1 - \Phi(\bar{s} - \Phi^{-1}(\theta'; \sigma_\eta); \sigma_\varepsilon)}{\Phi(\bar{s} - \Phi^{-1}(\theta'; \sigma_\eta); \sigma_\varepsilon)} \right) \frac{e^{-\frac{(\bar{s} - \Phi^{-1}(\theta'))^2}{2\sigma_\varepsilon^2}}}{\sqrt{2\pi} \sigma_\varepsilon} d\theta \right] = \\ \int_0^1 \left(1 + \alpha \cdot M \cdot \int_\theta^1 \Phi(\bar{s} - \Phi^{-1}(\theta'; \sigma_\eta); \sigma_\varepsilon) d\theta' - \beta \cdot M \cdot \int_0^\theta \Phi(\bar{s} - \Phi^{-1}(\theta'; \sigma_\eta); \sigma_\varepsilon) d\theta' \right) \frac{e^{-\frac{(\bar{s} - \Phi^{-1}(\theta'))^2}{2\sigma_\varepsilon^2}}}{\sqrt{2\pi} \sigma_\varepsilon} d\theta \quad (114) \end{aligned}$$

$$\begin{aligned} \mu \cdot \left[-\log \left(\frac{1 - \Phi(\bar{s}; \sqrt{\sigma_\eta^2 + \sigma_\varepsilon^2})}{\Phi(\bar{s}; \sqrt{\sigma_\eta^2 + \sigma_\varepsilon^2})} \right) \frac{e^{-\frac{\bar{s}^2}{2(\sigma_\eta^2 + \sigma_\varepsilon^2)}} \cdot \sigma_\varepsilon}{\sqrt{2\pi} (\sigma_\eta^2 + \sigma_\varepsilon^2)^{\frac{3}{2}}} + \int_0^1 \log \left(\frac{1 - \Phi(\bar{s} - \Phi^{-1}(\theta'; \sigma_\eta); \sigma_\varepsilon)}{\Phi(\bar{s} - \Phi^{-1}(\theta'; \sigma_\eta); \sigma_\varepsilon)} \right) \frac{e^{-\frac{(\bar{s} - \Phi^{-1}(\theta'))^2}{2\sigma_\varepsilon^2}} \cdot (\bar{s} - \Phi^{-1}(\theta))}{\sqrt{2\pi} \sigma_\varepsilon^2} d\theta \right] = \\ - \int_0^1 \left(1 + \alpha \cdot M \cdot \int_\theta^1 \Phi(\bar{s} - \Phi^{-1}(\theta'; \sigma_\eta); \sigma_\varepsilon) d\theta' - \beta \cdot M \cdot \int_0^\theta \Phi(\bar{s} - \Phi^{-1}(\theta'; \sigma_\eta); \sigma_\varepsilon) d\theta' \right) \frac{e^{-\frac{(\bar{s} - \Phi^{-1}(\theta'))^2}{2\sigma_\varepsilon^2}} \cdot (\bar{s} - \Phi^{-1}(\theta))}{\sqrt{2\pi} \sigma_\varepsilon^2} d\theta \quad (115) \end{aligned}$$

□

Proof of Proposition 8

Proof. Welfare under no information is better than under full information if and only if:

$$\frac{4\alpha_w}{2(\alpha - \beta)^2} > \frac{M^2(\alpha + \beta + \alpha_w)}{2(\beta M)^2} \quad (116)$$

if $2\beta M > M(\beta - \alpha)$, one can rearrange this as a positive lower bound on α_w given in the proposition. If $2\beta M < M(\beta - \alpha)$, then the above is true for any non-negative α_w . \square