HIGHER SCHOOL OF ECONOMICS
NATIONAL RESEARCH UNIVERSITY

Sylvain Carre, Pierre Collin-Dufresne,
Franck Gabriel

INSIDER TRADING WITH PENALTIES

Working Paper WP9/2021/02
Series WP9
Research of economics and finance

Moscow
2021

УДК 330.4
ББК 65.01
С21

Editor of the Series WP9
"Research of economics and finance"
Maxim Nikitin

Carre, S., Collin-Dufresne, P., Gabriel, F.
C21 Insider Trading with Penalties [Electronic resource] : Working paper WP9/2021/02 / S. Carre, P. Collin-Dufresne, F. Gabriel ; National Research University Higher School of Economics. – Electronic text data (1,2 Mb). – Moscow : Higher School of Economics Publ. House, 2021. – (Series WP9 "Research of economics and finance"). – 72 p.

We establish existence and uniqueness of equilibrium in a generalized one-period Kyle (1985) model where the insider's trades can be subject to a cost that is non-decreasing in the trade size – a "penalty". The result is obtained for uniform noise and holds for general penalty functions. Uniqueness is among all non-decreasing strategies. We construct the equilibrium price and optimal insider trading policy explicitly and find that, except for quadratic penalties, both are non-linear functions of, respectively, the trading volume and the liquidation value.

Our tractable framework can be used in a variety of contexts. As an application, we characterize the set of optimal penalties a regulator would choose to maximize price informativeness for a given level of expected uninformed traders' losses. Efficient penalties eliminate small rather than large trades. We extend the analysis to the case where implementable regulations are constrained by a budget constraint. We evidence the robustness of our main findings to our distributional assumption.

УДК 330.4
ББК 65.01

Keywords: Kyle model; non-linear equilibria; existence and uniqueness; market microstructure; insider trading; market regulation; efficient penalties

JEL classification: C72; G14

Sylvain Carre, ICEF, National Research University Higher School of Economics (HSE University), Russian Federation; Office S708, 11 Pokrovsky Boulevard, 109028 Moscow, Russian Federation; E-mail: scarre@hse.ru

Pierre Collin-Dufresne, Swiss Finance Institute and Ecole Polytechnique Federale de Lausanne; E-mail: pierre.collin-dufresne@epfl.ch

*Franck Gabriel**, Department of Mathematics, Ecole Polytechnique Federale de Lausanne; E-mail: franck.gabriel@epfl.ch

* Supported by the ERC CG CRITICAL, held by Pr. Hairer.

1 Introduction

Kyle's (1985) model of insider trading, in which a risk-neutral informed agent submits market orders along side uninformed 'noise traders' to risk-neutral market makers, has been a central work-horse of microstructure. The model assumes that trading by the insider is frictionless, so that information asymmetry is the only source of illiquidity. The equilibrium characterized by Kyle is linear, in that the equilibrium price is linear in the total order flow and the optimal trading strategy of the insider is proportional to the liquidation value known only to him.

In this paper we extend Kyle's (1985) seminal one-period model to the case where the insider's trading may be subject to an additional trading cost, which is increasing in the size of her trades and which we call 'trading penalty'. Such a penalty may reflect non-informational components of trading costs, for example related to market maker's inventory concerns or market power. It may also reflect the costs faced by legal insiders of a firm who are subject to costly disclosure requirements, or the expected cost of prosecution faced for trading on illegal insider information.[1]

We establish the existence and uniqueness of the equilibrium and characterize analytically the equilibrium price and optimal strategy of the insider for an arbitrary (non-decreasing) penalty function under the assumption that noise-trading and fundamental liquidation value have a uniform distribution.

[1] It is noteworthy that Kyle's 'insider trading' model does not distinguish between legal or illegal insider trading. In practice, most countries with active financial markets make a clear distinction between the two and, indeed, severely punish investors who trade in secondary markets based on 'material non-public' insider information - see e.g. `https://www.investor.gov/introduction-investing/investing-basics/glossary/insider-trading`. While data based on SEC prosecutions is often used in the literature to investigate predictions of usual asymmetric information models (e.g. Kacperczyk and Pagnotta (2019), Ahern (2020), Cornell and Sirri (1992)) it seems natural to expect that the nature of the information, legal versus illegal, should affect the optimal trading strategy of the informed agent since her payoffs are different due to the potential costs associated with their prosecution in case of detection. Similarly, legal insiders of the firm are subject to specific trading and disclosure rules that make it more costly for them to trade and thus would likely affect their optimal trading strategy relative to an unconstrained investor (see, e.g., Shkilko (2019)).

Doing so is challenging because a priori it implies solving a fixed point problem on a functional space where linear guesses do not turn out to be valid candidates. Moreover, we are able to go one step further by endogenizing the cost function itself — a maximization over the function space of penalties.

Kyle (1985) assumes a Gaussian distribution and Bagnoli, Viswanathan, and Holden (2001) show that linear equilibria in the Kyle-type model can obtain for a number of distributions including the uniform case. Uniqueness of the linear equilibrium has been established for the Gaussian case (Boulatov, Kyle, and Livdan (2013)).[2] However, all these papers focus on the case where there is no penalty cost.

The equilibrium we characterize is, except for quadratic penalty functions, non-linear in that the insider's trading rule is a non-linear function of the liquidation value and the equilibrium price is a non-linear function of total trading volume. Further, we establish uniqueness among all non-decreasing strategies. Instrumental to our proof is the surprising finding that under the uniform distribution assumption, even though the equilibrium price is a non-linear function of the total trading volume for most penalty functions, the *insider's expectation of the price* is always a linear function of her trades. This result is specific to the uniform distribution. However, we solve the Gaussian case numerically and compare the solution for many different penalty functions to the uniform case. We find similar non-linear shapes of the optimal policy and price functions in the usual Gaussian case.

Because, as emphasized above, trading costs take various forms and are prevalent in financial markets, and because our model solution is tractable, our main theorem has a large array of applications, including insider trading regulation but also models for a variety of transaction fees. So far, there was no tool in the literature to handle general cost functions in asymmetric information frameworks à la Kyle.

[2] See also McLennan, Monteiro, and Tourky (2017). Rochet and Vila (1994) also prove existence and uniqueness for any distribution in a modified one-period Kyle model where the insider observes noise trading. However, their approach does not extend straightforwardly to the case with penalties.

As an application of our main theorem, we characterize which penalty functions provide the highest price informativeness for a given level of market liquidity. We interpret those as regulatory fines on insider trading and then find out how the previously found efficient penalties modify with the introduction of different budget constraints.

There is a long debate in the literature, going back at least to Manne (1966), about the pros and cons of insider-trading regulations.[3] Most of this literature compares economies with insider trading to economies without in terms of welfare and stresses the trade-off between more informative prices and secondary market liquidity. Allowing insider trading typically leads to more informationally efficient prices. At the same time, the presence of insiders increases adverse selection risk for uninformed traders and thus harms market liquidity.[4]

We note that regulators typically motivate insider trading rules in terms of 'establishing a level-playing field' for investors or in terms of market 'fairness'.[5] In the context of our model, fairness and market liquidity are substitutes as higher market illiquidity translates directly into higher trading losses for uninformed investors.

Without taking a stand on which objective (price informativeness or market liquidity) is more important, we identify the set of 'efficient penalties' which maximize price efficiency for a given level of market liquidity. Specifically, we measure price efficiency by the expected post-trade standard deviation of the asset fundamental value in the market maker filtration and we measure the cost of illiquidity by the ex-ante expected losses of uninformed

[3] A non exhaustive list includes Manne (1966); Carlton and Fischel (1983); Easterbrook (1985); Glosten (1989); Bajeux and Rochet (1989); Manove (1989); Leland (1992); Shin (1996); DeMarzo, Fishman, and Hagerty (1998).

[4] Consequences for the efficiency of production and investment decisions can be ambiguous as more price efficiency typically leads to better production/investment decisions (Leland (1992)), whereas more adverse selection risk can decrease prices and lead to less efficient investment decisions Manove (1989)).

[5] See e.g. `https://www.sec.gov/spotlight/insidertrading/cases.shtml` and the reference in footnote 1.

traders.

We characterize explicitly the set of efficient penalties and show analytically that 'efficient penalties' eliminate small rather than large trades. In cases where the market's expectation of the fundamental value is sufficiently far away from the true value, efficient penalties are such that an insider will still find it optimal to trade. Indeed, in these cases she trades as much as she would in the absence of regulation, despite the fact that expected penalties (to be paid if detected) are large for large trading volumes. Although such trades—if they occur—are costly for uninformed traders, the regulator finds it optimal to not rule out all trading when market prices are severely biased, since these trades bring a lot of information to the market and help correct severe price inefficiencies.

Interestingly, we prove that quadratic penalties, which are the only penalties who lead to a linear Kyle-like equilibrium, are the most inefficient type of penalty functions a regulator could choose. Yet, the existing literature on illegal insider trading has often considered (exogenous) quadratic penalities presumably because of the tractability it affords due to the implied linear equilibrium (Shin (1996), Kacperczyk and Pagnotta (2019)).

The relative performances of the different cost functions are also shown to be robust to the distributional assumption, as very similar results are obtained numerically in the Gaussian case.

DeMarzo, Fishman, and Hagerty (1998) also study the design of optimal insider trading regulation. They consider the problem of a regulator who must choose the optimal investigation policy and pecuniary fines levied on the insider, in order to minimize losses incurred by uninformed traders. The regulator has finite resources which constrains the number of (costly) investigations she can perform. Their paper predicts that the regulator should investigate only following large trading volume, consistent with empirical evidence.[6] However, a surprising prediction of their model is that the regulator will initiate costly investigations even though she never collects any penalty

[6] See e.g. Augustin and Subrahmanyam (2019).

in equilibrium. Indeed, their optimal penalties involve zero fines on small trades, so that insiders make small enough trades to avoid all sanctions, even if investigated.

In contrast, the optimal penalties in our model are such that an insider will typically trade when her signal is very informative, i.e., when market expectations are far away from fundamentals. Contrary to DeMarzo, Fishman, and Hagerty (1998), our model generates insider demand schedules which are monotonic in the ex-ante mispricing (in line with the empirical evidence in Frino, Satchell, Wong, and Zheng (2013)) and predicts the existence of large insider trades as well as a positive rate of conviction in equilibrium. Intuitively, the difference between the two models' predictions arise because in our setting we allow the regulator to also put some weight in her objective function on price informativeness.

Of course, investigations in insider trading cases are typically very costly. For instance, Augustin and Subrahmanyam (2019) report in their comprehensive survey on insider trading that from 2011 to 2019, the SEC has spent $ 300 million in whistleblower rewards as part of its efforts to "pursue high profile insider traders."[7]

Adding a budget constraint for the regulator may significantly distort the set of implementable regulations. Furthermore, whether prosecution of insider trading is carried out at the civil or penal level has different implications. Indeed, the ability to levy pecuniary fines can help relax the regulator's budget constraint. To incorporate these considerations, we extend our setting to the case where regulators are subject to a budget constraint; and consider both the cases of pecuniary and non-pecuniary penalties. We prove that in the non-pecuniary case, the set of efficient penalty functions is simply a truncation of the set in the absence of any budgetary consideration. Intuitively, either a level of expected noise traders losses is too low to be implemented (as it would require a cost of investigation exceeding the

[7]Further evidence of the constraints faced by US regulators is provided in Newkirk and Robertson (1998).

budget constraint), or it can be implemented in which case the regulator uses the same penalty function as in the unconstrained case. Instead, in the pecuniary case, the regulator may select different penalties because they help relax the budget constraint. New patterns emerge in the demand schedules of the insider trader and the associated price functions.

Finally, we compare both prosecution modes by considering a regulator facing the following policy alternative in order to balance her budget: either be less active in investigating and use non-pecuniary fines; or investigate more often and use monetary penalties. We find that no policy is uniformly superior to the other, as the regulator's optimal choice depends on her preferences. Using pecuniary penalties is useful when the regulator puts more emphasis on market liquidity/fairness than on price efficiency but is limited in her ability to implement strong restrictions on insider trading because of the high cost of investigation. Instead, if the regulator puts more weight on price efficiency and thus does not wish to implement such strong restrictions on insider trading, then non-pecuniary penalties perform better.

2 A One-Period Kyle Model with Penalties

As in the one-period version of Kyle (1985), the model features a risk-neutral insider trader (IT), noise traders (NT) and a competititve risk-neutral market maker (MM). Agents are trading an asset with fundamental value v. The IT perfectly observes v and places an order $X(v)$. NT have a stochastic demand u independent of v. MM observes the total demand $X(v) + u$ and executes orders at a price P such that she breaks even on average.

Our model differs from the original Kyle (1985) model in two respects. First, we consider uniform—instead of Gaussian—noise:

$$u, v \sim U(-1, 1), \qquad u \perp v.$$

Remark 1 *The choice of* $[-1, 1]$ *as the support for* u, v *simplifies the expo-*

sition but is without loss of generality. All our results carry over to the case $u \sim U(-a, a)$ and $v \sim U(b, c)$.

Second we assume that a trade of size x involves a cost $C(x)$. C can be interpreted as a transaction cost or as an expected penalty imposed on illegal insider trades: $C = \alpha\tilde{C}$. In this intepretation, which we investigate in Section 4, α is the probability that the regulator starts and successfully completes an investigation, while $\tilde{C}(x)$ is the cost imposed on the IT conditional on the investigation being successful and the order of the IT being x. More generally, our approach allows the ex-post cost to depend on factors other than x as long as the ex-ante (expected) cost is a function of x only.

2.1 Benchmark Equilibrium without Penalties

In the absence of penalties, the IT solves

$$\max_{x \in I} x\mathbb{E}_u[v - P(x + u)] \tag{1}$$

taking the price function P of the MM as given. The MM breaks even on average:

$$P(d) = \mathbb{E}[v|X(v) + u = d]. \tag{2}$$

An equilibrium is a pair (X, P) that satisfies (1) and (2). For example, (X, P) defined by

$$X(v) = v \tag{3}$$

$$P(x + u) = \frac{x + u}{2}, \tag{4}$$

is an equilibrium of the one-period Kyle model without penalty. We call it the *(linear) mimicking equilibrium* since $X(v)$ and u are equal in distribution. We will prove later that this equilibrium is unique among all equilibria featuring a non-decreasing demand whose image lies in $[-1, 1]$. (With penalties, the optimal demand is no longer mimicking the random demand u.)

2.2 Equilibrium with Penalties

The IT solves

$$\max_{x \in I} x\mathbb{E}_u[v - P(x+u)] - C(x), \tag{5}$$

taking the price function P of the MM as given. The MM breaks even on average:

$$P(d) = \mathbb{E}[v|X(v) + u = d]. \tag{6}$$

This game involving the IT and the MM is denoted $\mathcal{K}(C)$. An equilibrium of $\mathcal{K}(C)$ is a pair (X, P) such that X solves (5) and P satisfies (6).
The interval $I \subset \mathbb{R}$ in the maximization program (5) is the set of admissible insider's demands. We will use the following assumption:

Assumption 1 $I = [-1, 1]$.

The bounds of I are natural since they are obtained in the linear mimicking equilibrium without penalty function. A demand function X whose image is not contained in $[-1, 1]$ would imply that for some values of the fundamental v, the magnitude of the IT order is *higher* when there is a penalty, compared to the linear equilibrium without penalty.[8]

To conclude this section, we state two remarks and introduce some notation.

(i) The data of a strategy X implies a pricing function P via Equation (6). We denote the pricing function associated with a demand schedule X by P_X.

(ii) In the IT's maximization program (5), the pricing function P only appears via the *expected price function*, denoted $\hat{P}$ and defined by

$$\hat{P}(x) = \mathbb{E}_u[P(x+u)]. \tag{7}$$

$\hat{P}$ represents the price that the IT will face on average if she places an order

[8] An interesting question is whether there exists an equilibrium for which Assumption 1 does not hold.

x. The program (5) can be rewritten in terms of the expected price function only:

$$\max_{x \in I} x(v - \hat{P}(x)) - C(x). \tag{8}$$

2.3 Out-of-Equilibrium Pricing

The noise u we consider has bounded support. Moreover, from Assumption 1, the equilibrium demand functions X we consider satisfy $|X| \leq 1$. This means that the aggregate order, $d = X(v) + u$ belongs to a bounded set D. The conditional expectation in (6) is not defined for values of $d \notin D$, meaning that we must make an assumption on the out-of-equilibrium pricing of the MM:

Assumption 2 *For any equilibrium (X, P) of $\mathcal{K}(C)$ we consider, with X non-decreasing and $X([-1, 1]) \subset [-1, 1]$, we always impose the following out-of-equilibrium pricing (letting $x_M = X(1)$ and $x_m = X(-1)$):*

$$\begin{aligned} P(d) &= 1 \quad \textit{for} \quad d > 1 + x_M, \\ P(d) &= -1 \quad \textit{for} \quad d < -1 + x_m. \end{aligned}$$

This assumption states that when the MM observes an aggregate order larger than the maximal possible equilibrium order, he prices the asset as if it had realized at its maximal value, $v = 1$. Similarly, when the aggregate order is smaller than the minimal possible equilibrium order, the MM prices as if $v = -1$. When verifying that (X, P) is an equilibrium, one must not only check that $X(v)$ maximizes the IT's program (5) among all x in the candidate support $[x_m, x_M]$, but also among values of x in $I \setminus [x_m; x_M]$. For these values of x, the aggregate order $d = x + u$ realizes in the out-of-equilibrium region with positive probability, in which case Assumption 2 defines the price $P(d)$.

Finally, notice that Assumption 2 fully characterizes out-of-equilibrium pricing: indeed, any $d \in [-1+x_m, 1+x_M]$ belongs to the support of $u+X(v)$, since u is $U(-1, 1)$ and $-1 \leq x_m \leq x_M \leq 1$.

2.4 Indistinguishable Equilibria

In the presence of penalties, we should expect the existence of realizations of v such that the IT is indifferent between two strategies: placing a small order and undergoing a small cost, or placing a larger order associated with a larger cost. However, as long as the set of v such that the maximization program of the IT (5) admits several solutions has measure zero, these indifference points will almost surely not be reached. The equilibrium will therefore be independent of the choice of the maximizer $X(v)$, in the sense that any *ex post* model observable, such as the IT demand $X(v)$ or the observed price $P(d)$, is almost surely the same, and any *ex ante* model quantity, such as the IT expected profit, is the same. In that case, we wish to consider that any choice of maximizer X induces the same equilibrium. We formalize this by introducing an equivalence relation between equilibria that we call indistinguishability.

Definition 1 *Let* (X, P) *and* (X', P') *be two equilibria of* $\mathcal{K}(C)$*. We say that* (X, P) *and* (X', P') *are* indistinguishable *if* X *and* X' *agree outside of a countable set.*

From now on, we identify an equilibrium of $\mathcal{K}(C)$ to its equivalence class. Definition 1 is useful because we will see that maximizers of (5) agree outside of a countable set, so the equilibria they induce belong to the same equivalence class.

2.5 Admissible penalty functions

We require that our cost functions solely depends in a non-decreasing manner on the magnitude of the order of the insider trader.

Definition 2 $C : [-1; 1] \to \mathbb{R}_+$ *is a penalty function if it is symmetric and non-decreasing, left-continuous over* $[0; 1]$ *and satisfies* $C(0) = 0$*. The set of penalty functions is denoted* $\mathcal{C}$*.*

The left-continuity assumption makes sure that the supremum of the possible IT profits is attainable.

3 Existence and uniqueness of equilibrium for $\mathcal{K}(C)$

In this section, we set out to prove our main theorem:

Theorem 2 *For any $C \in \mathcal{C}$, the Kyle game $\mathcal{K}(C)$ with penalty function C admits a unique equilibrium (X, P) such that X is non-decreasing. In general, X and P are non-linear functions.*

3.1 Analysis of the expected price function

3.1.1 The expected price function is linear regardless of the IT demand

Lemma 1 contains the key observation at the root of our analysis.

Lemma 1 *Let $X : [-1, 1] \to [-1, 1]$ be a non-decreasing function, $x_M = X(1)$ and $x_m = X(-1)$. The expected price function $\hat{P}$ associated with X is linear on $[x_m, x_M]$:*

$$\hat{P}(x) = \frac{x}{2}.$$

Lemma 1 states the surprising finding that even though the equilibrium price is a non-linear function of the total trading volume for most penalty functions, the *insider's expectation of the price* is always a linear function of her trades. Specifically, in equilibrium the expected price function is $\hat{P}(x) = x/2$ for $x_m \leq x \leq x_M$; neither the form of C nor guesses about X or P are needed to derive this result. In fact, using Assumption 2, we will prove that the equality $\hat{P}(x) = x/2$ holds for any admissible demand x.

Hence, the equilibrium demand of the IT, $X(v)$, must be a maximizer of

$$\psi_C(.,v) : x \mapsto x\left(v - \frac{x}{2}\right) - C(x). \tag{9}$$

If X is such a maximizer, we claim that (X, P_X) is the unique equilibrium of $\mathcal{K}(C)$ such that X is non-decreasing. To reach this conclusion, several issues remain to be addressed. First, we need to show that $\hat{P}(x) = x/2$ for any x as claimed above. Second, to make sure that Lemma 1 applies, we must check that any (selection of) maximizer is non-decreasing. Third, in order to obtain uniqueness, we need to show that $\psi_C(.,v)$ admits a unique maximizer except for a countable number of values of v.

We now provide the proof of this lemma. Section 3.1.2 clarifies the main intuitions.

Proof of Lemma 1. We use the notation $p(.)$ for a density and $p(.|.)$ for a conditional density. Write

$$\begin{aligned} p(v|d) &\propto p(d|v)p(v) \\ &\propto \mathbb{I}_{X(v)\in[d-1;d+1]}\mathbb{I}_{v\in[-1;1]}. \end{aligned}$$

That is, for $-1 + x_m \leq d \leq 1 + x_M$, $v|d$ is uniform over

$$\begin{aligned} \{v \in [-1;1] | X(v) \in [d-1;d+1]\} &= \{v \in [-1;1] | X(v) \in [d-1;d+1] \cap [x_m;x_M]\} \\ &= [(X_\ell^{-1}((d-1)\vee x_m); X_r^{-1}((d+1)\wedge x_M))] \end{aligned} \tag{10}$$

where $X_\ell^{-1}(x) = \inf\{v | X(v) \geq x\}$ and $X_r^{-1}(x) = \sup\{v | X(v) \leq x\}$. X_ℓ^{-1} and X_r^{-1} only disagree when there is v such that $X(v) = x$ and X is locally constant at v, i.e. they agree outside of a countable set. Then, letting $P = P_X$,

$$P(d) = \frac{1}{2}\left(X_\ell^{-1}((d-1)\vee x_m) + X_r^{-1}((d+1)\wedge x_M)\right).$$

Now since

$$\hat{P}(x) = \frac{1}{2}\int_{x-1}^{x+1} P(z)\,\mathrm{d}z,$$

it is enough to show that $P(x+1) - P(x-1) = 1$ a.e.. Using the expression of P

found above, we obtain that for $x_m \leq x \leq x_M$:

$$\begin{aligned} 2(P(x+1) - P(x-1)) &= X_\ell^{-1}(x \vee x_m) + X_r^{-1}((x+2) \wedge x_M) \\ &- X_\ell^{-1}((x-2) \vee x_m) - X_r^{-1}(x \wedge x_M) \\ &= X_r^{-1}(x_M) - X_\ell^{-1}(x_m) \\ &= 2 \end{aligned}$$

a.e.. This is because $X_\ell^{-1} = X_r^{-1}$ a.e., $X_r^{-1}(x_M) = 1$, and $X_\ell^{-1}(x_m) = -1$. ■

Having identified $\hat{P}$, we know that the insider trader's problem is to maximize $\psi_C(., v)$ as defined in (9). Because we will use this function throughout the paper, we repeat its definition here:

Definition 3 *The insider's profit function (under the expected price function $\hat{P}(x) = x/2$) for a demand x when the fundamental value is v is*

$$\psi_C(x, v) = x\left(v - \frac{x}{2}\right) - C(x). \tag{11}$$

3.1.2 Intuition

In order to isolate the intuition behind Lemma 1, let us consider the case where X is continuous and strictly increasing.

Assume that the market maker observes an aggregate order $d > 0$. Since the demand of the noise traders u takes values in $[-1, 1]$, the possible demands of the IT $X(v)$ consistent with the observation of d are exactly the admissible demands such that $d - 1 \leq X(v) \leq d + 1$. Because admissible demands satisfy $X(v) \leq 1$ and $d + 1 > 1$, the information obtained by the market maker when he observes d is that $X(v) \geq d - 1$. Thus, he knows that $v \geq X^{-1}(d-1)$. Intuitively, the fact that the aggregate order is positive rules out extreme negative values of v and the MM deduces a lower bound on v, $X^{-1}(d-1)$.

Moreover, due to the uniform noise assumption, all values of v above this lower bound are equally likely. Therefore, the price $P(d)$ is given by the midpoint of the interval $\left[X^{-1}(d-1), 1\right]$.

In a similar manner, when $d < 0$, the price $P(d)$ is given by the midpoint of the interval $\left[-1, X^{-1}(d+1)\right]$.

Now, assume that the IT wants to place an order x. The IT is only concerned by the expected price impact, $\hat{P}(x)$, which is a uniform average of the $P(d)$ over $d \in [x-1, x+1]$, the set of possible aggregate demands given an IT demand x. If, instead, the IT decides to place an order $x + \Delta x$, the set of possible aggregate demands d is $d \in [x-1+\Delta x, x+1+\Delta x]$: see Figure 1.

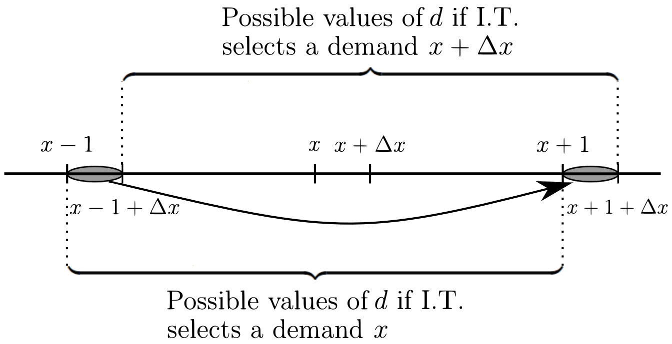


Figure 1: Marginal expected price impact of an increase in x

Thus, the only contribution to the marginal increase in expected price $\hat{P}(x+\Delta x) - \hat{P}(x)$ is due to the fact that the weight that was attributed to the interval $[x-1, x-1+\Delta x]$ is now attributed to the interval $[x+1, x+1+\Delta x]$. Crucially, this weight is the same due to the uniform noise assumption. Considering a vanishing Δx, one concludes that the marginal impact of increasing demand on expected price is proportional to $P(x+1) - P(x-1)$.

We have seen above that $P(x+1)$ is the midpoint of $\left[X^{-1}((x+1)-1), 1\right] = \left[X^{-1}(x), 1\right]$, and that $P(x-1)$ is the midpoint of $\left[-1, X^{-1}((x-1)+1)\right] =$

$\left[-1, X^{-1}(x)\right]$. Therefore, the marginal impact on the expected price is proportional to the distance between these two midpoints:

$$\frac{d}{dx}\hat{P}(x) \propto P(x+1) - P(x-1) = \frac{1 + X^{-1}(x)}{2} - \frac{X^{-1}(x) - 1}{2} = 1. \quad (12)$$

Figure 2 provides an illustration of this result. From (12), we see that the expected price function is linear.

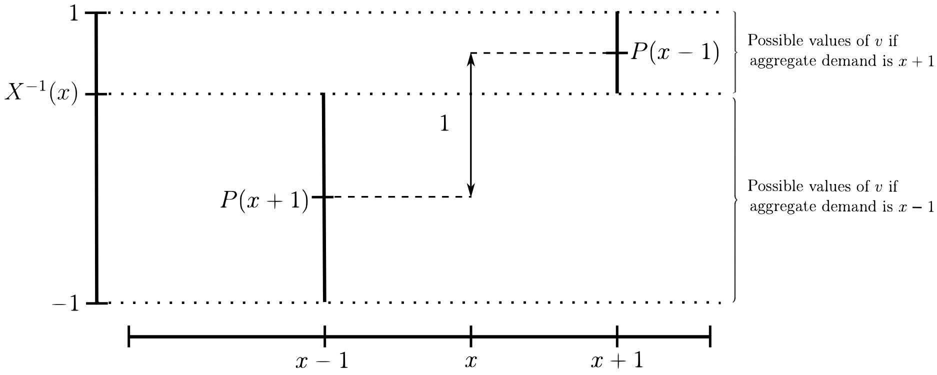


Figure 2: The marginal expected price impact is constant

3.2 Candidate optimal demands are unique up to changes on a countable set

In this section, we make sure that the maximizer X of ψ_C defined in (11) unambiguously defines a strategy, in the sense that for (almost all) v, $X(v)$ is unique.

Recall that if V and I are two intervals of $\mathbb{R}$, a correspondence $\mathcal{X} : V \to \mathcal{P}(I) \setminus \emptyset$ is said to be *non-decreasing* if for any $v_1 < v_2$ in V, $\sup \mathcal{X}(v_1) \leq \inf \mathcal{X}(v_2)$. (If $\mathcal{X}$ is a one-to-one mapping, we recover the usual notion of a non-decreasing function.)

For a given penalty $C \in \mathcal{C}$, let $\mathcal{X}_C$ be the correspondence mapping $v \in [-1;1]$ to the set of maximizers of the insider's profit when she observes the realization v:

$$\mathcal{X}_C(v) = \arg\max_x \psi_C(x, v).$$

We can now show that the maximizer of the IT's expected profit is unique except for a countable number of values of v:

Lemma 2 *For any $v \in [-1, 1]$, $\mathcal{X}_C(v) \neq \emptyset$, and $\mathcal{X}_C$ is a non-decreasing correspondence. This implies that there exists a non-decreasing function X_C such that for all $v \in [-1, 1]$ except on a countable set, $\mathcal{X}_C(v) = \{X_C(v)\}$.*

Proof. See Appendix A.1. ■

As we identify equilibria in a same equivalence class, as introduced in Definition 1, we do not need to specify which particular X_C we consider: we can unambiguously talk about *the* maximizer of the expected profit. We are now ready to prove our main result.

3.3 Existence and uniqueness of the equilibrium of $\mathcal{K}(C)$

We recast the statement of Theorem 2 using the expression of $\hat{P}$ obtained in Lemma 1.

Let $C \in \mathcal{C}$ and $X_C(v)$ be the maximizer of $x \mapsto x\left(v - \frac{x}{2}\right) - C(x)$. Then (X_C, P_{X_C}) is an equilibrium of $\mathcal{K}(C)$. This is the unique equilibrium among the pairs (X, P) such that $X : [-1, 1] \to [-1, 1]$ is non-decreasing.

Proof of Theorem 2. From Lemma 1, $\hat{P}(x) = \frac{x}{2}$ for $x_m \leq x \leq x_M$. Since $X_C(v)$ is a maximizer of $x\left(v - \frac{x}{2}\right) - C(x)$, $x = X_C(v)$ is an optimal response to the expected price function $\hat{P}$ among all $x \in [x_m, x_M]$. To confirm that (X_C, P_{X_C}) is an equilibrium, we need to check what happens if the IT makes a choice outside of the candidate support $[x_m, x_M]$, knowing that the out-of-equilibrium pricing is defined by Assumption 2. Consider for instance the case $x \in (x_M, 1]$, as the case $x \in [-1, x_m]$ works identically. Then

$$\begin{aligned}
\hat{P}(x) &= \frac{1}{2}\int_{x-1}^{x+1} P(z)\,\mathrm{d}z \\
&= \frac{1}{2}(x - x_M) + \frac{1}{2}\int_{x_M-1}^{x_M+1} P(z)\,\mathrm{d}z - \frac{1}{2}\int_{x_M-1}^{x-1} P(z)\,\mathrm{d}z \\
&= \frac{1}{2}(x - x_M) + \hat{P}(x_M) - \frac{1}{2}\int_{x_M-1}^{x-1} P(z)\,\mathrm{d}z \\
&= \frac{1}{2}(x - x_M) + \frac{x_M}{2} - \frac{1}{2}\int_{x_M-1}^{x-1} P(z)\,\mathrm{d}z \\
&= \frac{x}{2}. \qquad (13)
\end{aligned}$$

This is because when $z \in [x_M - 1, x - 1]$, $z - 1 < x - 2 \leq -1 \leq x_m$ and $z + 1 \geq x_M$ so from (10), $v|z$ is uniform over $[-1, 1]$ and $P(z) = 0$.

As $X_C(v)$ maximizes $x \mapsto x\left(v - \frac{x}{2}\right) - C(x)$, and $\hat{P}(x) = \frac{x}{2}$ for $x \in [-1, x_m) \cup (x_M, 1]$, $X_C(v)$ maximizes $x \mapsto x(v - \hat{P}(x)) - C(x)$ over $[-1, 1]$: (X_C, P_{X_C}) is an equilibrium.

We now prove uniqueness. Let $X' : [-1, 1] \to [x'_m, x'_M]$ be a non-decreasing strategy of the IT. By Lemma 1, the expected price $\hat{P}'$ associated with X' is $\frac{x}{2}$ for $x \in [x'_m, x'_M]$. But the computation of $\hat{P}'$ outside of $[x'_m, x'_M]$ is the same as the computation of $\hat{P}$ in (13). Hence, for all $x \in [-1, 1], \hat{P}'(x) = \frac{x}{2}$. So, if $(X', P_{X'})$ is an equilibrium of $\mathcal{K}(C)$ such that X' is non-decreasing, X_C and X' maximize the same objective ψ_C over $[-1; 1]$. By Lemma 2 the maximizers agree outside of a countable set, hence so do X_C and X'. In turn, we have $P_{X'} = P_{X_C}$. Therefore (X_C, P_{X_C}) and $(X', P_{X'})$ are the same equilibrium, which establishes uniqueness.

■

Since $\psi_C(x, v) = \psi_C(-x, -v)$, the theorem implies that the equilibrium demand function of the IT must be an odd function. In particular we know that the minimal demand x_m equals $-x_M$.

3.4 Examples of equilibria

Our proof of Theorem 2 shows how to explicitly construct the equilibrium IT demand function and equilibrium price for a specific penalty function. To see how the equilibrium is affected by the presence of a penalty, in this section we consider three examples of penalty functions: quadratic, linear, and constant over large trades. These type of cost functions have been widely studied in partial equilibrium models of optimal portfolio choice in the presence of transaction costs.[9] This literature typically assumes exogenous price dynamics and solves for the optimal demand of an investor facing specific transaction costs. In each example we also derive the optimal demand of the investor, but in addition deliver the price function consistent with the equilibrium demand. Further, unlike in the partial equilibrium models, in our model the IT faces, in addition to the exogenous trading penalty, also an endogenous adverse-selection trading cost component, which is driven by the market makers optimal price-setting mechanism in response to the observed order-flow.

Consistent with intuition, we find that penalties typically reduce the demand of the IT. In general, the presence of penalty generates non-linear IT demand schedules as well as non-linear price functions, which can be very flat in some regions and increase sharply in others. In particular, the price impact of a marginal uninformed trade $\frac{d}{du}P(X(v)+u)$ strongly depends on both the realizations of u and v. By constrast, in the mimicking equilibrium of the model without penalties, this price impact is constant, regardless of the distributional assumptions on the noise.

[9] For example, Garleanu and Pedersen (2013), Collin-Dufresne, Daniel, and Saglam (2020) use quadratic t-costs, Constantinides (1986) and Davis and Norman (1990) use linear transaction costs, Lynch and Tan (2010) and Liu (2004) consider fixed transaction costs.

3.4.1 Quadratic penalty

In this particular instance, X remains linear after the introduction of the penalty (but not P).[10] Imposing quadratic costs is akin to increasing the perceived expected price impact. Since this cost is in x^2 while the gross gains of trading are in x, the IT always trades as soon as $v \neq 0$, and the magnitude of the trade increases with the absolute value of v.

[10] In the special case of a quadratic penalty, a tractable Kyle-type linear-equilibrium also obtains when asset value and noise trading have a Gaussian distribution (see Shin (1996)).

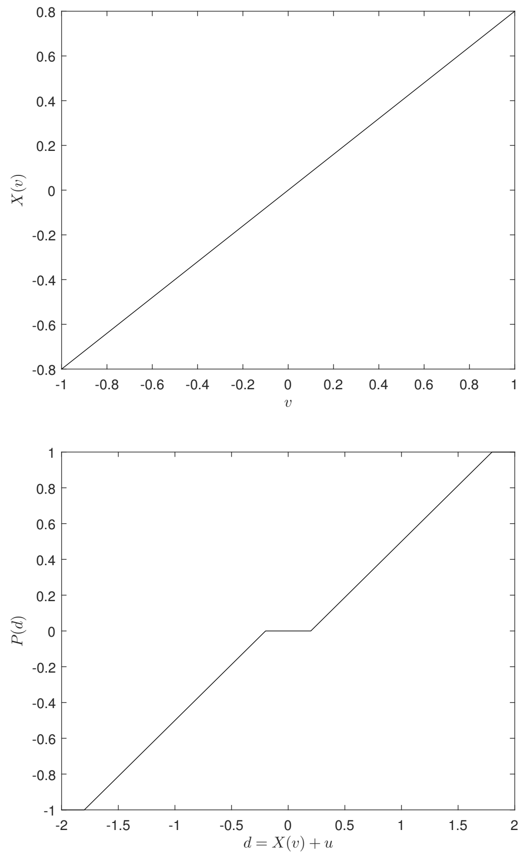


Figure 3: Insider's demand and pricing under quadratic penalty

$C(x) = \alpha x^2$, $\alpha = 0.125$. Left panel: IT demand X. Right panel: price function P.

Due to the presence of the penalty, the insider trades less than in the linear mimicking equilibrium, so that $X(1) = x_M < 1$ $(= 0.8$ in this example).

When $|d| \leq 1 - x_M (= 0.2)$, any demand of the IT is compatible with the observed aggregate order, so all values v remain equally likely, as explained in Section 3.1.2. No information is incorporated and the price remains at the initial expected value of the asset: 0. When $d > 1 - x_M$, one knows that v has not realized at a very low value. This provides a lower bound on v and the price becomes positive. As d increases, so do the lower bound and the price, until $d = 1 + x_M (= 1.8)$. In that case, one knows for sure that the IT has placed an order x_M, which means that $v = 1$, and P reaches 1. The situation is symmetrical for values of d below $x_M - 1(= -0.2)$.

3.4.2 Linear penalty

When the penalty is linear, $C(x) = \alpha|x|$, for positive values of v, the maximization program of the IT can be rewritten as

$$\max_{x\in[0,1]} x\left((v-\alpha) - \frac{x}{2}\right).$$

If $v \geq \alpha$, one sees that a linear cost has the same effect as reducing the value of the fundamental v by an amount α, and having no cost. Therefore, the strategy of the IT for values $v \in [\alpha, 1]$ is a translation of the linear mimicking strategy over $v \in [0, 1-\alpha]$. Similarly, the strategy of the IT for values $v \in [-1, -\alpha]$ is a translation of the linear mimicking strategy over $v \in [\alpha - 1, 0]$. This creates the two increasing linear segments in the left panel of Figure 4. In the flat middle section, v is not sufficient to cover the penalty: the IT does not trade.

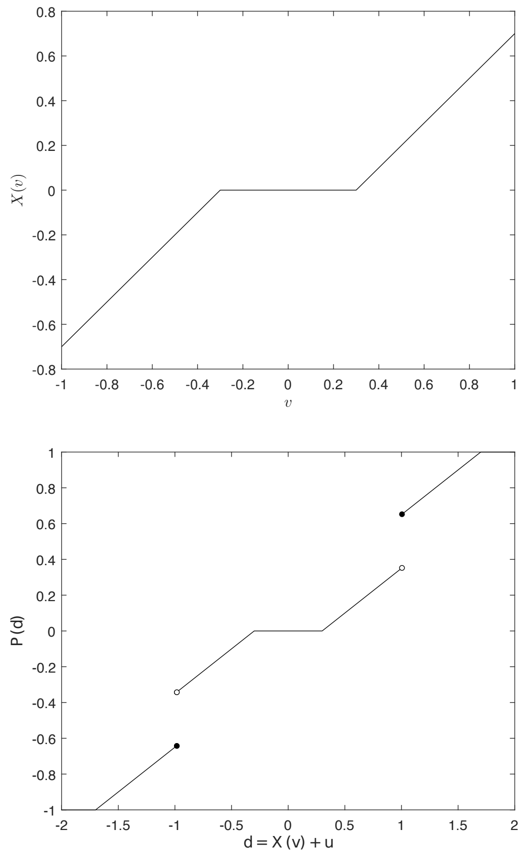


Figure 4: Insider's demand and pricing under linear penalty

$C(x) = \alpha|x|$, $\alpha = 0.3$. Left panel: IT demand X. Right panel: price function P.

The price function depicted in the right panel of Figure 4 exhibits a flat section in the center surrounded by increasing linear segments. The intuition is similar to the quadratic penalty case: when the magnitude of d is small

($|d|\leq \alpha(=0.3)$), all values of v remain equally possible and no information is incorporated. As d grows, a lower bound on v can be deduced and the price increases. The key difference with the quadratic penalty case is that the price function jumps at $d = \pm 1$. Indeed, when $d > 1$, the market maker knows for sure that the insider has placed a positive order. But the IT only does so when $v > \alpha$. By contrast, if $d = 1^-$, $X(v) = 0$ remains possible, so we can only deduce that $v > -\alpha(= -0.3)$.

3.4.3 Constant cost on trades of magnitude larger than x_0

Absent penalties, the IT picks $X(v) = v$. Hence, if she is sanctioned only for trades of magnitude larger than x_0, she will not change her demand as long a $|v|\leq x_0$: this corresponds to the increasing linear section in the middle of Figure 5. For intermediate values of v, the IT prefers to block her demand at the value x_0 (or $-x_0$) in order to avoid the penalty: this corresponds to the flat sections in Figure 5. When v becomes large enough ($|v|> \sqrt{2K}(\approx 0.63)$), the penalty is recouped in expectation by using the strategy that prevails in the absence of costs: it appears as a sunk cost and the IT selects again the demand $X(v) = v$. This corresponds to the increasing linear sections at the left and right of Figure 5.

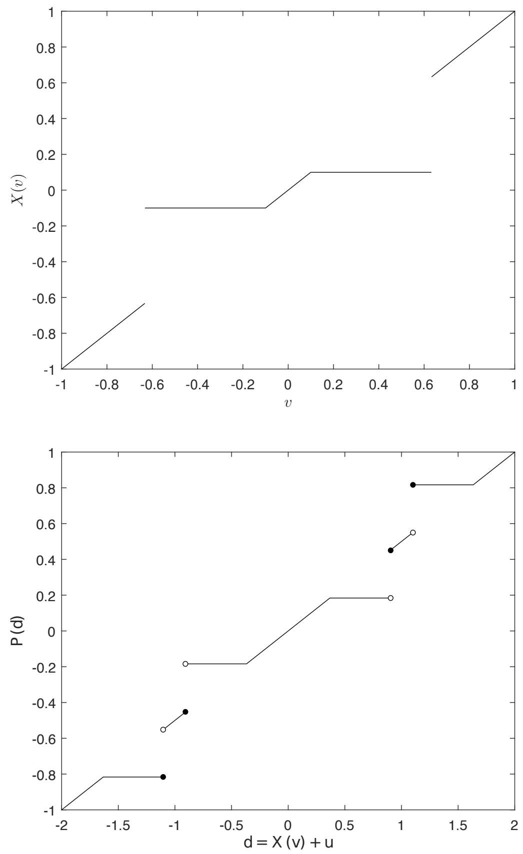


Figure 5: Insider's demand and pricing under constant penalty on large trades

$$C(x) = K\mathbb{I}_{|x|>x_0},\ K = 0.2,\ x_0 = 0.1.$$

Left panel: IT demand X. Right panel: price function P.

The price function jumps at $d = \pm(1-x_0)(= \pm 0.9)$ and $d = \pm(1+x_0)(= \pm 1.1)$. The intuition is as in the linear penalty case. When d exceeds $1-x_0$, the MM knows that the demand of the IT was larger than $-x_0$ which rules out all values of v at the left of $-\sqrt{2K}$, the left jump of X. Similarly, when d exceeds $1 + x_0$, the MM knows that the demand of the IT was larger than x_0, which rules out all values of v at the left of $\sqrt{2K}$, the right jump of X.[11]

Remark 3 *The counterparts of these examples for the model with Gaussian noise are reported in Section 5.1.1. Most of the properties detailed above carry through.*

4 An Application: Efficient Penalties for Insider trading regulation

Having shown that equilibrium outcomes depend heavily on the functional form of the penalty function, in this section we ask what type of penalties are optimal from the point of view of a regulator who trades-off price efficiency and market liquidity. A long literature, going back at least to Manne (1966) and already mentioned above, has argued about the pros and cons of insider-trading regulations. Most of this literature compares economies with insider trading to economies without in terms of welfare and stresses the trade-off between more informative prices and secondary market liquidity. Allowing insider trading typically leads to more informationally efficient prices. At the same time, the presence of insiders increases adverse selection risk for uninformed traders and thus harms market liquidity.

Without taking a stand on which objective (price informativeness or market liquidity) is more important, we identify the set of 'efficient penalties' which maximize price efficiency for a given level of market liquidity. Specifically, we measure price efficiency by the expected post-trade standard devi-

[11] The left and right-continuities of P are determined by the preceding reasoning and represented graphically by the filled and empty dots. By contrast, at a jump point v, X can indifferently be taken to be X_{v-} or X_{v+} since it generates maximizers X in the same equivalence class. Hence left- and right- continuities do not need be specified on the graph of X.

ation of the asset fundamental value in the market maker filtration and we measure the cost of illiquidity by the ex-ante expected losses of uninformed traders.

For this section it is useful to interpret the function C as a product $C = \alpha\tilde{C}$, where α is the probability that an investigation starts and succeeds, and $\tilde{C}$ is the penalty imposed upon the insider trader conditional on investigation success. For now we treat α as a constant; Section 5.2 discusses the case where α depends on the total order flow, capturing the idea that investigation is more likely following the observation of large trading volumes.

We first consider the unconstrained regulator's optimal tradeoff between the two objectives and then add a resource constraint for the regulator to see how limited investigative resources might affect the optimal penalties.

4.1 The unconstrained regulator's problem

The regulator seeks an 'efficient penalty' which trades-off:

(i) the post-trade standard deviation of the fundamental, $\sigma(v|d)$,

(ii) the profits of the uninformed traders:

$$g(u, v) = u(v - P(X(v) + u)). \tag{14}$$

Quantity (i) captures the residual uncertainty about v. If it is small then prices are informative. Quantity (ii) captures liquidity in financial markets. In a liquid market, agents who have to trade for non-fundamental reasons do not experience high losses. This corresponds to a situation where g is not too negative. The trade-off arises because improving (i) typically causes (ii) to worsen.

Let

$$S = \mathbb{E}[\sigma(v|d)] \tag{15}$$

be the expectation of the post-trade standard deviation of v and

$$G = \mathbb{E}[g(u, v)] \tag{16}$$

denote the expected profit of the NT.

The objective of the regulator can now be stated as the characterization of the *efficient frontier*, with the following definition:

Definition 4 *(i) A point (G, S) is implementable if it is the outcome of an equilibrium of $\mathcal{K}(C)$ for some admissible penalty C.*

(ii) An implementable point (G, S) is dominated by (G', S') if (G', S') is implementable and $G' \geq G$, $S' \leq S$ with at least one strict inequality.

(iii) The set of implementable non-dominated points is called the **efficient frontier***.*

In Section 4.3.2, we will need the following refinement of (ii):

(ii') An implementable point (G, S) belonging to some subset of the plane H is dominated in H by (G', S') if (G', S') is implementable, $G' \geq G$, $S' \leq S$ with at least one strict inequality and $(G', S') \in H$.

Points outside the efficient frontier are irrelevant from the regulator's perspective, as it can improve upon one of its objectives without harming the other. By contrast, any point belonging to the efficient frontier could be picked by a regulator for a suitable weighting of the objectives.[12] Our goal is to characterize the efficient frontier and the penalties that implement it. We do so in three different settings: without a budget constraint (Section

[12] Our approach has the advantage that we take no stance on the specific preferences of the regulator.

4.2), under a budget constraint with non-pecuniary (Section 4.3.1) and pecuniary (Section 4.3.2) penalties. First, we introduce some useful notation.

Let

$$\pi^N(v) := \psi_C(X(v), v) \tag{17}$$

be the expected net profit of the insider trader in state v,

$$\Pi^N := \mathbb{E}_v[\pi^N(v)] \tag{18}$$

be the overall expected net profit (after fine, if any), and

$$F := \mathbb{E}[C(X(v))] \tag{19}$$

be the expected penalty levied on the insider.

Observe that we can write

$$|G| = \int_0^1 X(v)\left(v - \frac{X(v)}{2}\right) \mathrm{d}v = \underbrace{\int_0^1 \frac{v^2}{2}\,\mathrm{d}v}_{1/6} - \frac{1}{2}\int_0^1 (v - X(v))^2 \,\mathrm{d}v. \tag{20}$$

This way of seeing the expected losses of the uninformed traders as (an affine transformation of) the L^2 distance between X and the identity will be useful in Section 4.3.1.

4.2 Efficient frontier without a budget constraint

Theorem 4 *When the regulator does not face a budget constraint, the equation of the efficient frontier is*

$$S = \frac{1}{\sqrt{3}}(1 + 2G), \quad -\frac{1}{6} \leq G \leq 0.$$

The set of penalties that implements the efficient frontier is exactly the class $\mathcal{O}$ defined as

$$\mathcal{O} = \left\{ C \in \mathcal{C},\ \exists K \in [0, 1/2], \quad \begin{array}{ll} C(x) \geq x\left(\sqrt{2K} - \frac{x}{2}\right) & \text{for } 0 \leq x \leq \sqrt{2K}, \\ C(x) = K & \text{for } \sqrt{2K} < x \leq 1 \end{array} \right\} \tag{12}$$

When $C \in \mathcal{O}$, the demand of the insider writes

$$X_K(v) = \begin{cases} 0 & |v| \leq \sqrt{2K}, \\ v & |v| > \sqrt{2K}, \end{cases}$$

for the $K \in [0, 1/2]$ associated with C.

Figure 6 gives a graphical representation of functions in $\mathcal{O}$.

If two penalties in $\mathcal{O}$ are associated with the same K, they implement the same demand schedule X_K. Moreover, it is easy to see that any point on the efficient frontier is implemented by X_K for exactly one value of K.[13] Therefore, K parametrizes the efficient frontier. Points associated with a small (respectively large) K are selected by a regulator who puts more weight on price efficiency (respectively on market liquidity or'fairness' for uninformed traders).

Any regulator who puts some positive weight on both objectives will seek to reduce insider trading somewhat, but not entirely. As we shall prove later, the optimal solution is to allow some large trades for large realizations of $|v|$, because they incorporate a lot of information; more precisely, the

[13] The penalty $C(x) = K\mathbb{I}_{x \neq 0}$ belongs to $\mathcal{O}$ and implements X_K. A direct calculation shows that the P& L G of the uninformed traders under the demand X_K is $-\frac{1}{6}\left(1 - (2K)^{3/2}\right)$. Hence, the value of K that implements the point (G, S) of the efficient frontier is the solution to $G = -\frac{1}{6}\left(1 - (2K)^{3/2}\right)$.

regulator wants to implement $X(v) = v$ for large values of $|v|$. The cutoff point $\sqrt{2K}$ in the schedule X_K then appears as the solution to the equation $\frac{v^2}{2} = K$. (Recall that $\frac{v^2}{2}$ is the profit of the IT when there is no penalty). This characterizes the magnitude of v above which the penalty appears as a sunk cost to the insider, who then effectively optimizes as if there was no penalty and selects the mimicking demand $X(v) = v$.

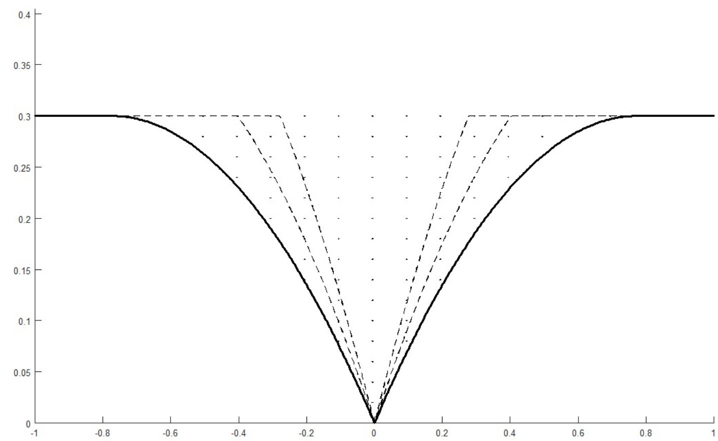


Figure 6: Some penalty functions in $\mathcal{O}$.

In Figure 6, the thick line represents the lower bound in the definition of $\mathcal{O}$ when $K = 0.3$ (then, $\sqrt{2K} \approx 0.77$) : any penalty in $\mathcal{O}$ must be above this line. Given that a penalty is symmetrical and non-decreasing over $[0, 1]$, the graph of a function in $\mathcal{O}$ must be included in the dotted area. The two dashed lines represent two such functions.

4.2.1 Preliminary results on the regulator's objective

Lemma 3 *In equilibrium, the net profits satisfy*

$$\pi^N(v) = \int_0^v X(s)\,\mathrm{d}s, \tag{22}$$

$$\Pi^N = \int_0^1 (1-v)X(v)\,\mathrm{d}v. \tag{23}$$

Proof. Consider the parametrized objective function

$$\psi_C : [0,1] \times [0,1] \to \mathbb{R}$$

defined in (11). Notice that (i) $\psi_C(x,.)$ is linear in v and therefore absolutely continuous, (ii) $|\partial_v \psi_C(x,v)| = |x| \leq 1$. (i) and (ii) guarantee that the assumptions of Theorem 2 in Milgrom and Segal (2002) are satisfied. In the present case, this theorem tells us that we can write:

$$\begin{aligned} \pi^N(v) &= \pi^N(0) + \int_0^v \partial_2 \psi_C(X(s), s)\,\mathrm{d}s \\ &= \int_0^v X(s)\,\mathrm{d}s, \end{aligned}$$

since the insider does not make any profit when the fundamental v is 0. Finally,

$$\begin{aligned} \Pi^N = \frac{1}{2}\int_{-1}^1 \pi^N(v)\,\mathrm{d}v &= \frac{1}{2}\int_{-1}^1 \int_0^v X(y)\,\mathrm{d}y\,\mathrm{d}v \\ &= \int_0^1 \int_0^v X(y)\,\mathrm{d}y\,\mathrm{d}v \\ &= \int_0^1 (1-v)X(v)\,\mathrm{d}v. \end{aligned}$$

■

Lemma 4 expresses the expected post-trade standard deviation as a function of the demand profile X.

Lemma 4 *The expected post-trade standard deviation satisfies*

$$S = \frac{1}{\sqrt{3}}\left(1 - \int_0^1 vX(v)\,\mathrm{d}v\right). \tag{24}$$

Proof. By the proof of Lemma 1, $v|d$ is uniform over

$$I_X(d) \equiv [(X_\ell^{-1}\left((d-1)\vee(-x_M)\right); X_r^{-1}\left((d+1)\wedge x_M\right))].$$

Since the standard deviation of a uniform variable over $[a;b]$ equals $\frac{1}{2\sqrt{3}}(b-a)$, Lemma 4 is an immediate consequence of the following result: *if X is an odd non-decreasing function from $[-1;1]$ to $[-x_M;x_M]$, then the expected length of the interval $I_X(X(v)+u)$ equals* $2\left(1-\int_0^1 vX(v)\,\mathrm{d}v\right)$, which we must now prove.

For $v \in [-1;1]$, define

$$\begin{aligned} Y_v &= X_r^{-1}\left((X(v)+u+1)\wedge x_M\right) \\ Z_v &= X_l^{-1}\left((X(v)+u-1)\vee(-x_M)\right). \end{aligned}$$

What we need to prove is that $\mathbb{E}_{v,u}[Y_v - Z_v] = 2\left(1-\int_0^1 vX(v)\,\mathrm{d}v\right)$. By symmetry, $\mathbb{E}_{v,u}[Z_v] = -\mathbb{E}_{v,u}[Y_v]$, thus, it remains to prove that:

$$\mathbb{E}_{v,u}[Y_v] = 1 - \int_0^1 vX(v)\,\mathrm{d}v.$$

Let us consider v fixed. The random variable Y_v takes values in $[-1,1]$: using Fubini theorem,

$$\mathbb{E}[Y_v] = \mathbb{E}\left[\int_{-1}^1 \mathbb{I}_{-1\le y\le Y_v}\,\mathrm{d}y\right] - 1 = \int_{-1}^1 \mathbb{P}\left(y \le Y_v\right)\,\mathrm{d}y - 1.$$

By definition of X_r^{-1}, if $X(y) \le (X(v)+u+1)\wedge x_M$ then $y \le Y_v$. Besides, if $y < Y_v$, then using the fact that X is non decreasing, $X(y) \le (X(v)+u+1)\wedge x_M$. Thus:

$$\{y \le Y_v\} \setminus \{X(y) \le (X(v)+u+1)\wedge x_M\} \subset \{y = Y_v\}.$$

Let us remark that $Y_v = y$ can hold for two different values of u if and only if X

is discontinuous at y or $y = 1$. In particular,

$$\{y \neq 1 | \mathbb{P}(y = Y_v) > 0\} \subset \{y | X(y^-) \neq X(y^+)\}.$$

It follows from this discussion that :

$$\left| \mathbb{E}[Y_v] - \int_{-1}^{1} \mathbb{P}(X(y) \leq X(v) + u + 1) \, \mathrm{d}y + 1 \right| \leq \int_{-1}^{1} \mathbb{P}(Y_v = y) \, \mathrm{d}y \leq \mu \left(\{y | X(y^-) \neq X(y^+)\} \right),$$

where μ is the Lebesgue measure on $[-1, 1]$. Since X is non-decreasing, it has a countable number of discontinuity points. In particular $\mu \left(\{y | X(y^-) \neq X(y^+)\} \right) = 0$ and:

$$\mathbb{E}[Y_v] = \int_{-1}^{1} \mathbb{P}(X(y) \leq X(v) + u + 1) \, \mathrm{d}y - 1.$$

Now,

$$\begin{aligned} \mathbb{P}(X(y) \leq X(v) + u + 1) &= \mathbb{P}(u \geq X(y) - X(v) - 1) \\ &= 1 + \left(\frac{1}{2}(X(v) - X(y)) \wedge 0 \right). \end{aligned}$$

Going back to the expression of $\mathbb{E}[Y_v]$, we obtain

$$\mathbb{E}[Y_v] = 1 - \frac{1}{2} \int_{v}^{1} X(y) \, \mathrm{d}y + \frac{1}{2}(1 - v)X(v).$$

Integrating over v:

$$\begin{aligned} \mathbb{E}_{v,u}[Y_v] &= 1 - \frac{1}{4} \int_{-1}^{1} \int_{v}^{1} X(y) \, \mathrm{d}y \, \mathrm{d}v + \frac{1}{4} \int_{-1}^{1} (1 - v)X(v) \, \mathrm{d}v \\ &= 1 - \frac{1}{4} \int_{-1}^{1} (v + 1)X(v) \, \mathrm{d}v + \frac{1}{4} \int_{-1}^{1} (1 - v)X(v) \, \mathrm{d}v \\ &= 1 - \frac{1}{2} \int_{-1}^{1} vX(v) \, \mathrm{d}v \\ &= 1 - \int_{0}^{1} vX(v) \, \mathrm{d}v, \end{aligned}$$

where in line 3, we used the fact that X is odd. This concludes the proof. ■

One consequence of Lemma 4 is that large orders associated with large values of the fundamental are the ones that contribute the most to incorporating information into prices. Indeed, the values of v such that the product $vX(v)$ is large have the strongest negative impact on S, as can be seen from (24).

4.2.2 Shape of the efficient frontier and efficient demand functions

In this section, we give the shape of the efficient frontier and explain what demand schedules are compatible with it. We call these schedules *efficient demand functions.*

Lemma 5 *Let C be a penalty function in $\mathcal{C}$. In the equilibrium of $\mathcal{K}(C)$,*

$$S \geq \frac{1}{\sqrt{3}}(1+2G)$$

with equality if and only if there is $v^* \in [0,1]$ such that $X(v) = 0$ for $|v| < v^*$ and $X(v) = v$ for $|v| > v^*$.

Proof. Due to Lemma 4, what we need to show is that

$$-\int_0^1 vX(v)\,\mathrm{d}v \geq -2\int_0^1 X(v)\left(v - \frac{X(v)}{2}\right)\mathrm{d}v.$$

This is equivalent to

$$\int_0^1 vX(v)dv \geq \int_0^1 X(v)^2\,\mathrm{d}v,$$

or

$$\int_0^1 X(v)(v - X(v))\,\mathrm{d}v \geq 0 \tag{25}$$

which holds because $0 \leq X(v) \leq v$ for $v \in [0;1]$.

For the equality to hold, it is necessary and sufficient to have $X(v) = 0$ or $X(v) = v$ almost everywhere. Since X is non-decreasing, it is equivalent to $X(v) = 0$ for $|v| < v^*$ and $X(v) = v$ for $|v| > v^*$, where $v^* = \sup\{v,\, X(v) = 0\}$. ■

Equation (25) is particularly convenient because it immediately indicates what type of demand function is needed to implement the efficient frontier. Of course, X is an endogenous outcome: what remains to be seen is what regulations implement the efficient demand functions.

4.2.3 Implementation of the efficient demand functions

Lemma 6 *The efficient demand functions derived in Lemma 5 are implemented exactly by the penalties $C \in \mathcal{O}$.*

Proof. See Appendix A.3. ■

By construction, penalties in $\mathcal{O}$ increase quickly as $|v|$ departs from 0 and are flat for large values of $|v|$ (see Figure 6). Intuitively, this is what is required to implement the efficient demand functions. Indeed, for small values of $|v|$, increasing demand has a large marginal impact on the expected penalty, so that the IT prefers to refrain from trading. For $|v|$ large, however, since the penalty schedule become flat for large demand, a large order allows to cover the expected fine, which appears as a sunk cost. The IT then optimizes as in the linear mimicking equilibrium and demands $X(v) = v$ for sufficiently large v.

The proof of Theorem 4 is complete: Lemma 5 characterizes the efficient frontier and due to Lemma 6, achieving the efficient frontier can only be done by selecting a cost $C \in \mathcal{O}$, characterized by a $K \in [0, 1/2]$.

4.2.4 Illustrations and discussion

Varying K between 0 and $1/2$ allows to cover the entire efficient frontier. As K increases, the losses $(-G)$ of the uninformed traders decrease from $\int_0^1 \frac{v^2}{2}\,\mathrm{d}v = \frac{1}{6} \approx 0.167$ to 0, while the expected post-trade standard deviation increases from $\frac{1}{\sqrt{3}}(1 - 2/6) = \frac{2}{3\sqrt{3}} \approx 0.385$ to $\frac{1}{\sqrt{3}} \approx 0.577$.

Each point of Figure 7 corresponds to a penalty function C; it represents the outcomes $(S, -G)$ in the unique equilibrium of $\mathcal{K}(C)$. For a fixed y-coordinate (a fixed S) the preferred option of the regulator is to select a point with the smallest x-coordinate (that minimizes $-G$). Consistent with Theorem 4, penalties in $\mathcal{O}$ achieve the efficient frontier, which is linear.

Outcomes $(S, -G)$ corresponding to quadratic and linear penalties ($C(x) = \alpha x^2$, $C(x) = \alpha|x|$ for varying $\alpha \geq 0$) are also reported in Figure 7. As one can see, they perform significantly worse than penalties $C \in \mathcal{O}$. This is also the case of penalties with no cost on small trades and big costs on large trades, $C(x) = K^H \mathbb{I}_{|x|>x_0}$. Here K^H is a constant large enough so that the insider never chooses to trade more than x_0. The fact that these particular penalty functions perform poorly compared to penalties in $\mathcal{O}$ is consistent with the intuition given below Lemma 4. Indeed, they imply that $X(v) = v$ for $|v|$ small and $|X|$ stops growing for $|v|$ large (the opposite of the demand functions implied by $C \in \mathcal{O}$), so that the reduction of the expected standard deviation, measured by the term $\int_0^1 vX(v)\,\mathrm{d}v$, is low.

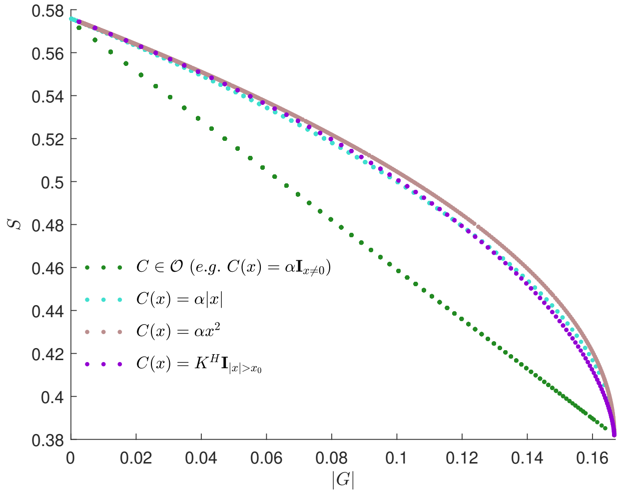


Figure 7: Locus of $(S, -G)$ for some penalty functions.
Each dot corresponds to a different value of α, for the constant, linear and quadratic costs; or to a different value of x_0, for the fourth functional form considered. (Recall that in this case, K^H is fixed to a large value such that insiders decide never to trade more than x_0. Of course when $x_0 \to 0$, trading disappears, and the points obtained converge the top-left point, with no losses for the noise traders and no incorporation of information.)

Figure 7 shows that quadratic costs are the most inefficient among the considered costs. In fact, they have the worst performance among *all penalty functions*:

Proposition 1 *Quadratic penalties implement the upper frontier of the locus of outcomes (S, G) generated by all penalty functions in $\mathcal{C}$, i.e. they induce the highest possible expected post-trade standard deviation for a given P&L of the uninformed traders.*

Proof. See Appendix A.2. ■

This result is interesting given that quadratic penalties are precisely those that have been used in extant models of insider trading (Shin (1996), Kacperczyk and Pagnotta (2019)). Granted, these models use a Gaussian setting where the quadratic penalty assumption leads to a Kyle-type linear equilibrium. However, as indicated by Figure 15 in the Section 5.1.2, our finding that quadratic penalties are bad at solving the regulator's trade-off extends to the Gaussian case.

Remark 5 *Figure 7 is replicated in Section 5.1.2 in the case of normal instead of uniform noise. The result is strikingly similar.*

4.3 The constrained regulator's efficient frontier

So far, our analysis has not considered limited resources regulators have to run investigations. And yet, insider trading investigations are typically very costly, as they require significant time, financial and human resources. For instance, Augustin and Subrahmanyam (2019) report that from 2011 to 2019, the SEC has spent \$ 300 million in whistleblower rewards as part of its efforts to "pursue high profile insider traders." How do the regulator's efficient frontier change in the presence of a binding budget constraint?

In Section 4.3.1, we consider the case of non-pecuniary penalties: the regulator cannot use fines to soften the budget constraint. In our model, this caps the maximal expected penalty that can be imposed on insider trades. Hence, in Section 4.3.1, the regulator faces the same problem as before—maximizing the efficiency of prices for a given level of uninformed traders' losses—yet its set of admissible strategies is narrower.

In Section 4.3.2, we study pecuniary fines. We shall assume that the regulator's initial budget is insufficient to cover its investigation expenditures: it needs to levy fines to break even. Hence, in Section 4.3.2, the regulator faces a richer problem than before—it now needs to consider a third criterion—yet, this time, there is no restriction *a priori* on its set of admissible strategies.

Section 4.3.3 compares the efficiency properties of pecuniary and non-pecuniary fines. We ask whether a constrained regulator aiming at balancing her budget should decide to investigate less and use non-pecuniary fines or investigate more and break even by collecting monetary penalties.

4.3.1 Non-pecuniary penalties

Let $B > 0$ be the budget allotted to the regulator and $\kappa > 0$ denote the investigation cost. Here, we work under the assumption that the investigation probability α is constant. The regulator operates under the constraint

$$\alpha\kappa \leq B. \tag{26}$$

Recall that the insider trader optimizes under an expected penalty schedule $C = \alpha\tilde{C}$, where $\tilde{C}$ is the actual sanction conditional on investigation success.

First, consider the case where there is no constraint on $\tilde{C}$. Then, the regulator can trivially get around its budget constraint (26) by reducing α and increasing $\tilde{C}$. In that case, we are back to the unconstrained problem solved in Section 4.2.

We now consider the more interesting case where there is an upper bound $\tilde{C}^M$ on $\tilde{C}$.

The existence of $\tilde{C}^M$ can be justified on several grounds: (i) stronger sanctions involve a larger burden of proof to be enforced. Newkirk and Robertson (1998) report that in the Netherlands, a law was passed at the end of the XXth century, presented as the "toughest in the world" against insider trading. The result was that in the following years, virtually no conviction was possible—not because there were no cases of insider trading suspicions, but because evidence was almost never strong enough "to satisfy the heavy burden of proof that must be met to support a criminal conviction". Hence, $\tilde{C}^M$ can be seen as a threshold penalty above which the sanction is effectively not enforceable. (ii) Alternatively, $\tilde{C}^M$ can be seen as the maximal disutility that a non-pecuniary sanction (such as lifetime imprisonment) can impose upon a human being.

From (26), the insider trader faces an expected penalty

$$C = \alpha\tilde{C} \leq K := \frac{B}{\kappa}\tilde{C}^M. \tag{27}$$

The constraint (27) means that we have restricted the regulator's set of admissible penalties:[14]

Definition 5 *In the non-pecuniary case, the set of admissible penalties with a budget constraint is*

$$\mathcal{C}_K = \{C \in \mathcal{C},\, C(1) \leq K\}.$$

Note that $C(1) \leq K$ is equivalent to (27) because any penalty in $\mathcal{C}$ is symmetrical and non-decreasing over $[0, 1]$. Moreover, the budget constraint is an actual constraint for $K \in [0, 1/2)$; for $K \geq 1/2$, $\mathcal{C}_K = \mathcal{C}$.

What happens when one restricts the set of admissible penalties? First, some previously efficient points may no longer be feasible. Second, some points that were not previously efficient may no longer be dominated by any point still implementable under the budget constraint. We recast Definition 4 in this new setting:

Definition 6 *In the non-pecuniary case, the efficient frontier under a budget constraint is the set of points* (G, S) *implementable by a penalty in* $\mathcal{C}_K$ *that are not dominated by any point implementable by a penalty in* $\mathcal{C}_K$*.*

Denote

$$\mathcal{O}_K = \mathcal{O} \cap \mathcal{C}_K.$$

$\mathcal{O}_K$ is the set of efficient penalties derived in Section 4.2.3, that are still feasible under the budget constraint. These penalties are still efficient under the budget constraint. Moreover, by direct computation, we obtain that as

[14] Of course, we could obtain the same constraint by ignoring investigation costs, setting $\alpha = 1$ and assuming that the cap $\tilde{C}^M$ on $\tilde{C} = C$ is below $1/2$. The idea here is that if investigation was systematic, the bound $\tilde{C}^M$ would likely be non-binding. It only becomes binding because investigation is costly, which reduces the expected penalty that the IT faces. The extent to which it binds depends on the budget-relevant parameters B and κ: see (27).

C varies in $\mathcal{O}_K$, $|G|$ describes the interval $\left[|G|_{\min}(K), \frac{1}{6}\right]$, where

$$|G|_{\min}(K) := \frac{1}{6}\left(1 - (2K)^{3/2}\right). \tag{28}$$

The truncation of the previously efficient frontier at the right (in the $(|G|, S)$ plane) of $|G|_{\min}(K)$ is part of the efficient frontier under the budget constraint. The key question is to know what happens at the left of $|G|_{\min}(K)$. Theorem 7 shows that no penalty in $\mathcal{C}_K$ can implement $|G|< |G|_{\min}(K)$. This immediately implies the characterization of the constrained efficient frontier:

Theorem 6 *The efficient frontier under the constraint $C \leq K$ is the truncation $|G| \geq |G|_{min}(K)$ of the efficient frontier of Theorem 4 and is implemented exactly by penalties in $\mathcal{O}_K$.*

As explained above, Theorem 6 is a consequence of:

Theorem 7 *Let $K \leq 1/2$. Under the constraint $C \leq K$, the expected losses of the uninformed traders are at least*

$$|G| \geq |G|_{min}(K).$$

Proof. See Appendix A.4, where we provide the proof as well as several intuitions and graphical interpretations. ■

One consequence of Theorem 6 is that it is not possible to infer from a regulator's choice of penalty whether it is constrained or not. In the non-pecuniary case, a regulator subject to a binding budget constraint effectively behaves like an unconstrained regulator that would assign less weight to curtailing the losses of the uninformed traders. In the next Section, we study the case of pecuniary penalties and show that, by contrast to the previous result, the introduction of the constraint creates new efficient points.

4.3.2 Pecuniary penalties

We now consider pecuniary penalties, collected by the regulator. We maintain the assumption of a constant α. We suppose that the regulator must have a balanced budget in expectation. The budget constraint (26) transforms into

$$\alpha\kappa \le B + \underbrace{\mathbb{E}[C(X(v))]}_{F}. \tag{29}$$

If $B \ge \alpha\kappa$, we are back to the case studied in Section 4.2. We now consider the case $B < \alpha\kappa$ and aim at characterizing the new efficient frontier under the constraint (29). This frontier will obtain by projection once we determine the *efficient surface*:

Definition 7 *The efficient surface Σ is the locus of points (G, S, F) generated by any $C \in \mathcal{C}$ such that no $C' \in \mathcal{C}$ can weakly (i) increase G, (ii) decrease S, (iii) increase F with at least one among (i), (ii) or (iii) being strict.*

Define the set of indices $J := \{(x, y),\, 0 \le y/(1+y) \le x \le y \le 1\}$.

Theorem 8 *The efficient surface Σ in the space (G, S, F) is constructed exactly by the demand schedules $(X_{v_1,v_2})_{(v_1,v_2)\in J}$, that are implemented by the penalties $(C_{v_1,v_2})_{(v_1,v_2)\in J}$:*

$$X_{v_1,v_2}(v) = \begin{cases} 0 & v \in [0, v_1] \\ \frac{v_2}{v_2 - v_1}(v - v_1) & v \in (v_1, v_2] \\ v & v \in (v_2, 1] \\ -X_{v_1,v_2}(-v) & v < 0. \end{cases} \quad ; \quad C_{v_1,v_2}(x) = \begin{cases} v_1|x| - \frac{v_1}{2v_2}x^2 & |x| \le v_2 \\ \frac{v_1 v_2}{2} & |x| > v_2. \end{cases}$$

Proof. See Appendix A.5. ■

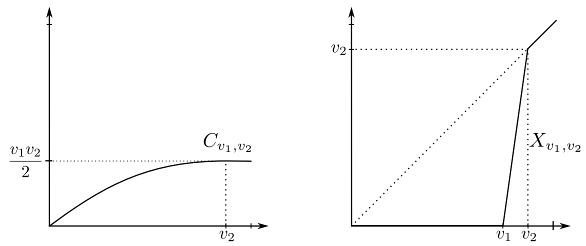


Figure 8: Efficient demand schedule and penalty function under a budget constraint with pecuniary fines.

There is a key difference between proving Theorem 7 and Theorem 8. Here, the proof can be reduced to a routine pointwise minimization, because the pointwise minimizer of the weighted regulatory objective (see the proof in Appendix A.5) turns out to be an implementable demand schedule. Reducing the problem to a pointwise minimization exercise was not possible to prove Theorem 7.

We can now turn to the characterization of the efficient (G, S) frontiers for various regulator's budgets. The budget constraint can be written as $F = \mathbb{E}[C(X(v))] \geq F_{\min} := \alpha\kappa - B$. Hence, we can give the following definition.

Definition 8 *The F_{min}-efficient frontier is the set of non-dominated points in*

$$\mathcal{F}(F_{min}) := \{(G(X), S(X)),\ X \text{ implemented by some } C \in \mathcal{C} \text{ with } \mathbb{E}[C(X(v))] \geq F_{min}\}\,.$$

We can now obtain the $F_{\min}$-efficient frontiers from the efficient surface Σ by projection. Denote $\pi_{GS} : (G, S, F) \mapsto (G, S)$ the projection on the (G, S)-plane: the $F_{\min}$-efficient frontier is the set of points of $\pi_{GS}(\Sigma \cap \{F \geq F_{\min}\})$ that are not dominated in $\pi_{GS}(\Sigma \cap \{F \geq F_{\min}\})$.

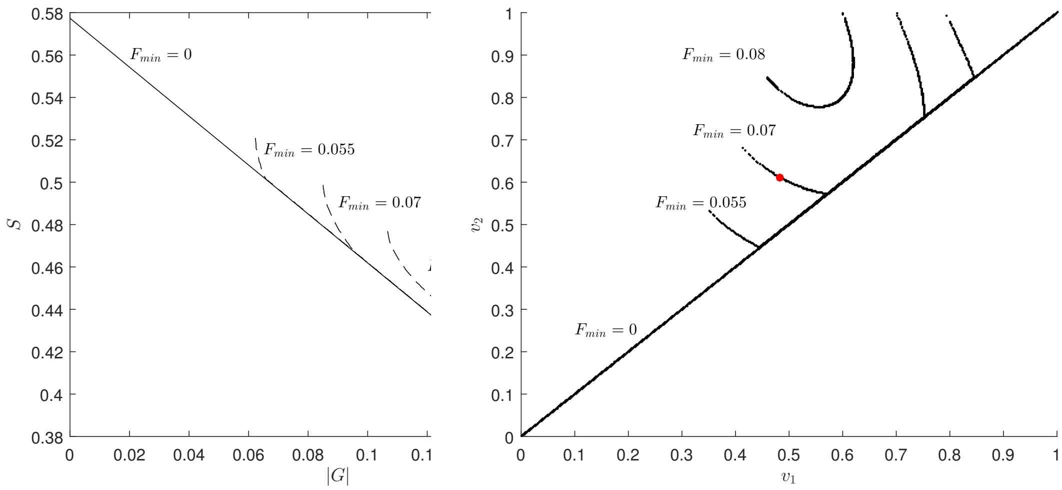
Of course, the penalties in $\mathcal{O}$ that implement $F \geq F_{\min}$ are still part of the $F_{\min}$-efficient frontier, but new constrained efficient points emerge (dotted arcs in Figure 9a), which are associated with penalty functions and demand schedules that were not previously optimal. The $F_{\min}$-efficient frontier does not even intersect the unconstrained frontier for $F_{\min}$ very large.[15]

A lax regulation of insider trading obviously does not allow to levy a lot of fines. But a very strict regulation does not either, because it deters insiders from trading and reduces the rate of conviction. Hence, the regulator must pick some "intermediate" level of regulation when it operates under a tight budget constraint. This is illustrated by the dotted frontiers in Figure 9a, which correspond to "intermediate" levels of $|G|$ and S. One takeaway is that if regulators need fines to balance their budget then completely ruling out insider trading may not be a credible goal.

Points of the $F_{\min}$-efficient frontier correspond to points in Σ, which means that they are associated with demand schedules of the form X_{v_1,v_2} defined in Theorem 8. To understand how the budget constraint $F \geq F_{\min}$ modifies the nature of the optimal strategies, Figure 9b plots the (v_1, v_2) used on the $F_{\min}$-efficient frontier for various values of $F_{\min}$. For example, the red dot, $(v_1, v_2) \approx (0.48, 0.61)$ represents $X_{0.48,0.61}$ and indicates that this demand schedule implements one point of the efficient frontier when the budget constraint of the regulator is such that $F_{\min} = 0.07$.

When $F_{\min} = 0$, we obtain the line $v_2 = v_1$, in which case X_{v_1,v_2} is implemented by a penalty $C \in \mathcal{O}$, consistent with Section 4.2.3. We observe that as $F_{\min}$ increases, one needs to widen the gap $v_2 - v_1$. The intuition is that the linear section over $[v_1, v_2]$ of the demand schedule X_{v_1,v_2} best resolves the trade off between large fines and large trade volumes of the insider trader and allows to collect a relatively high amount of fines in expectation. As an example, recall that the demand schedule that implies the highest expected fine (1/12) had $v_2 - v_1 = \frac{1}{2}$.

[15] Indeed, it is easy to show that the maximal expected fine under a penalty in $\mathcal{O}$ is 2/27, so if $F_{\min} > \frac{2}{27}$, no penalty in $\mathcal{O}$ allows to balance the regulator's budget. Using penalties in $\mathcal{C}$, the regulator can levy up to 1/12 in expectation, e.g. using $C_{\frac{1}{2},1}$. If $F_{\min} > \frac{1}{12}$, the regulator cannot balance its budget.



(a) Efficient frontiers

(b) Indices (v_1, v_2)

Figure 9: Efficient $(|G|, S)$ frontiers and indices (v_1, v_2) of the efficient demand functions X_{v_1,v_2}, for various constraints $F \geq F_{\mathbf{min}}$

Price functions are also modified by the introduction of a budget constraint with pecuniary fines.

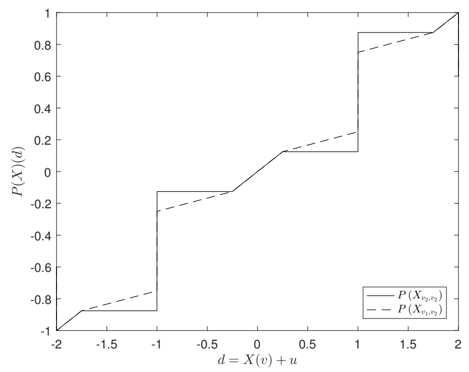


Figure 10: New price functions.
$v_1 = 0.5$ and $v_2 = 0.75$.

Figure 10 compares the price functions implied by an efficient demand schedule absent a budget constraint, X_{v_2,v_2} and a constrained efficient demand schedule X_{v_1,v_2}. Contrary to $P(X_{v_2,v_2})$, $P(X_{v_1,v_2})$ has no flat sections and is everywhere increasing. In particular, in the unconstrained case, the random price is partly discrete: with positive probability, it will be equal to one of the ordinates of the flat sections of $P_{X_{v_2,v_2}}$. Conversely, in the case of a strong budget constraint, the random price has a continuous density.

4.3.3 Comparison between pecuniary or non-pecuniary penalties

Consider a regulator who is constrained in the sense that her budget B does not allow to always investigate, i.e. to set $\alpha = 1$. This corresponds to the case where the investigation cost κ exceeds B. The regulator can contemplate the two following strategies:

- investigate less often and impose non-pecuniary fines;
- break even by levying pecuniary fines.

These policies involve a different distortion of the efficient frontier.

In the first case, the regulator accepts to only investigate with probability $\alpha = \frac{B}{\kappa}$. As proven in Section 4.3.1, the regulator will not be able to implement points that are the most favourable to liquidity traders ($|G|$ small), because she cannot investigate enough; however, if the regulator does not wish to implement such a point, imposing non-pecuniary penalties is optimal, since we know that the non-pecuniary-constrained efficient frontier is a truncation of the unconstrained one.

In the second case, the regulator needs to levy $F_{\min} = \kappa - B$ in fines in order to be able to always investigate. Pecuniary penalties may help implement points that are not achievable under non-pecuniary penalties.

Figure 11 evidences by a a numerical example where $\kappa > B$ that none of the two strategies uniformly dominates the other. Which policy to adopt when facing a budget constraint depends on the preferences of the regulator.

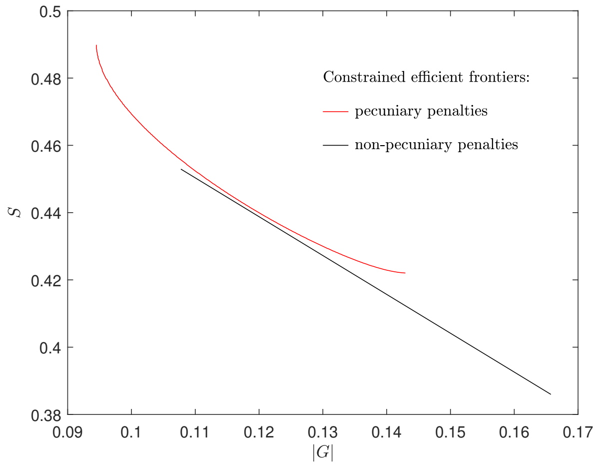


Figure 11: Available efficient frontiers for a constrained regulator.
$\kappa = 0.15$, $B/\kappa = 50\%$, $\tilde{C}^M = 1/2$.

A regulator facing a strong budget constraint but nevertheless desiring to severely restrict insider trading would opt for pecuniary penalties. Otherwise, there is no need to investigate all the time and therefore it is better for the regulator to simply be less active rather than subjecting herself to an additional constraint—which prevents her from resolving the trade-off between S and $|G|$ optimally.

5 Robustness and Discussion

In this section we investigate the robustness of our findings with respect to the distributional assumption. Since the literature typically assumes a Gaussian distribution rather than uniform we investigate numerically the properties of the Gaussan equilibrium and compare it to the uniform case, characterized analytically in the previous sections.

Then we discuss possible extensions of the penalty models considered in the main text.

5.1 Equilibrium with Penalties under Gaussian noise

In this section, we show that the main features of the equilibrium insider's demand X and of the pricing function P are preserved if one assumes Gaussian instead of uniform noise. The same holds true regarding the efficiency properties of standard cost functions. Hence our initial distributional assumption, while technically convenient, does not drive our main results.

We do not prove formally the existence of an equilibrium in the case of Gaussian noise; instead we run a fixed-point algorithm—detailed in appendix B—on Equations (5) and (6).

5.1.1 Properties of the equilibrium (X, P)

Figures 12, 13 and 14 display the equilibrium (X, P) for the model with Gaussian noise ($u, v \sim N(0, 1)$). For comparison purposes, we consider the same penalties C as in Section 3.4 namely: quadratic, linear and constant on large trades.[16]

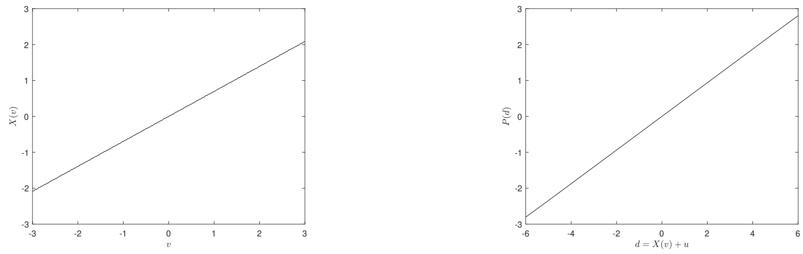


Figure 12: IT demand and pricing under quadratic penalty, Gaussian case
$C(x) = \alpha x^2$, $\alpha = 0.25$. Left panel: IT demand X. Right panel: price function P.

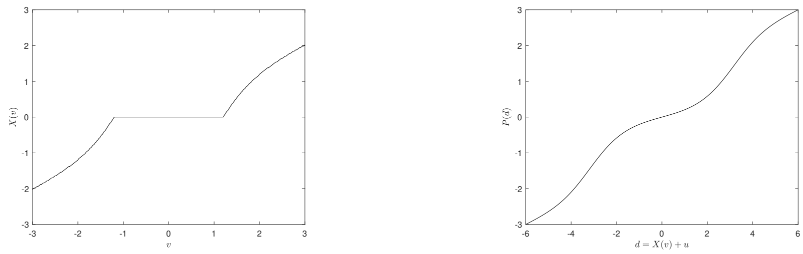


Figure 13: IT demand and pricing under linear penalty, Gaussian case
$C(x) = \alpha|x|$, $\alpha = 1.2$. Left panel: IT demand X. Right panel: price function P.

The Gaussian equilibrium bears many similarities with the the uniform equilibrium. First, under quadratic costs, X is also linear. Second, jumps in X, being driven by the economic forces described in Section 3.4, are not linked to a particular distributional choice and still exist in the Gaussian

[16] Of course, in the quadratic case, there is no need to resort to a numerical solution.

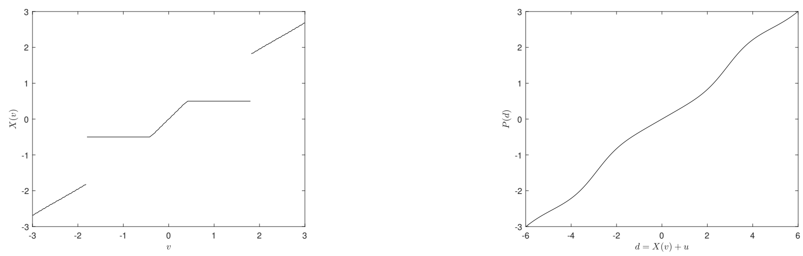


Figure 14: IT demand and pricing under constant penalty, Gaussian case

$C(x) = K\mathbb{I}_{|x|>x_0}$, $K = 1$, $x_0 = 0.5$.
Left panel: IT demand X. Right panel: price function P.

case. Third, P is in general non-linear. However, because of its infinite support, the Gaussian distribution case leads to a continuous price function (by contrast the uniform density is discontinuous at ± 1). The jumps which appeared in the uniform case are replaced by strong non-linearities in the price function.

5.1.2 Efficiency properties of different penalty costs under Gaussian noise

We construct numerically the equivalent of Figure 7 (which was obtained analytically) when noise is normal instead of uniform: $u, v \sim N(0; 1)$. We obtain Figure 15. The constant costs upon nonzero trades $C(x) = K\mathbb{I}_{x\neq 0}$ are doing best among the penalty functions considered. This is consistent with the results in the uniform noise case. Other penalties are suboptimal, as before, and the locus of points $(S, -G)$ they generate is very similar in shape. Interestingly, the quadratic penalty costs, which as discussed before have often been used in previous literature, perform the worst of all the considered cost functions. This confirms that the result on the suboptimality of quadratic penalties proved in Proposition 1 appears not to be driven by the uniform distribution assumption.

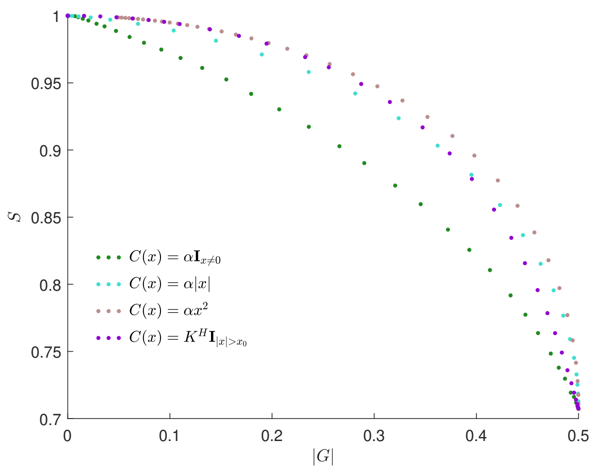


Figure 15: Locus of $(S, -G)$ for different penalty functions - Gaussian noise.

Each dot corresponds to a different value of α, for the constant, linear and quadratic costs; or to a different value of x_0, for the fourth functional form considered. (Recall that in this case, K^H is fixed to a large value such that insiders decide never to trade more than x_0. Of course when $x_0 \to 0$, trading disappears, and the points obtained converge the top-left point, with no losses for the noise traders and no incorporation of information.)

5.2 Extensions of the penalty model

State-dependent investigation probabilities. In Section 4, we viewed the penalty function C as the product of a constant investigation probability, α, and a penalty conditional on conviction, $\tilde{C}$. However, both our equilibrium characterization and the solution of the unconstrained regulatory problem carry through when α is a function of the total order flow $x + u$:

Remark 9 *We can consider ex-post costs of the form $C(x, u) = \alpha(x + u)\tilde{C}(x)$ with $\alpha(\cdot), \tilde{C}(\cdot)$ symmetric and non-decreasing over their positive domain. This is useful if one thinks the probability of investigation is increasing in total volume (as argued in De Marzo et al. (1998)).*

The reason—detailed in Appendix A.6—is that the ex-ante expected penalty, $C(x) \equiv \mathbb{E}_u[C(x,u)]$ satisfies Definition 2. Hence, one can consider such state-contingent investigation probabilities and retain our results on the Kyle equilibrium under penalties. Moreover, the unconstrained regulatory problem is unchanged, as the set of ex-ante penalty costs, $\mathcal{C}$, in unchanged.

Hence, our optimal resolution of the regulator's trade-off does not hinge on assuming constant investigation probabilities. In our framework with a risk-neutral informed agent, policies of selecting an informed order-dependent penalty schedule or a total order flow-dependent investigation probability are substitutes.

Generalization of the class of penalty functions. It is also possible to consider penalties under which—if convicted—the insider must not only pay a fine but also forfeit part or all of his trading profit. Indeed, if α denotes the probability of conviction, the insider gets the usual Kyle profit $\pi = x(v-p)$ with probability $1-\alpha$, and a fraction $\theta\pi$ of this, minus a penalty $C(x)$, with probability α. Hence, the IT maximizes $(1-\alpha+\alpha\theta)\pi - \alpha C(x)$, or equivalently, $\pi - \frac{\alpha}{1-\alpha+\alpha\theta}C(x)$:[17] we are back to the case where the penalty cost only depends on x. However, for some realizations of u and v, the IT makes losses. In order for the previous reasoning to hold, we need to assume that in such states, there is a transfer *from* the regulator to the insider.

A more realistic specification would be to impose that part or all of the profit is forfeited only conditional on this profit being positive. In that case, and more broadly when the penalty is a general function of x, but also u and v, the methods developed in this article do not apply, at least in their current form.

[17] Shin (1996) uses a similar argument in his model of insider trading under quadratic costs.

6 Conclusion

We have proven existence and uniqueness of the equilibrium in a Kyle-type model of insider trading under the assumption of uniformly—instead of normally—distributed noise. Importantly our result holds when insider trading may be subject to an additional trading penalty that may be an arbitrary non-decreasing function of total order flow and/or the insider's trade. As an application, we characterize the set of efficient insider trading penalties for a regulator who seeks to maximize market liquidity (or minimize uninformed trader losses) for a given level of price (informational) efficiency. We show that efficient penalties penalize small rather than large trades. In equilibrium, it is optimal for a regulator who puts some weight on price efficiency to set penalties such that insiders retain an incentive to trade when ex-ante market prices are far from fundamentals. We also consider how efficient penalties are affected if the regulator has limited investigative resources and consider both cases of pecuniary and non-pecuniary penalties. Non-pecuniary penalties simply truncate the unconstrained efficient frontier. Instead, pecuniary penalties may lead to a different resolution of the trade-off between liquidity and price efficiency as they help relax the budget constraint of the regulator. Especially when regulators put a lot of weight on minimizing uninformed trader losses can pecuniary penalties be 'superior'.

Our main results appear robust to our distributional assumption as many of the properties of the uniform equilibrium carry over to the Gaussian distribution case, which we investigate numerically.

References

Ahern, K. (2020): "Do Proxies for Informed Trading Measure Informed Trading? Evidence from Illegal Insider Trades," *The Review of Asset Pricing Studies*. [3]

Augustin, P., and M. Subrahmanyam (2019): "Informed Options Trading Before Corporate Events," *Annual Review of Financial Economics*. [6, 7, 40]

Bagnoli, M., S. Viswanathan, and C. Holden (2001): "On the Existence of Linear Equilibria in Models of Market Marking," *Mathematical Finance*, 11(1), 1–31. [4]

Bajeux, I., and J.-C. Rochet (1989): "Operations d'inities: Une analyse de surplus," *Finance*, 10, 7–19. [5]

Boulatov, A., A. Kyle, and D. Livdan (2013): "Uniqueness of Equilibrium in the single period Kyle '85 model," *Working paper*. [4]

Carlton, D., and D. Fischel (1983): "The Regulation of Insider Trading," *Stanford Law Review*, 35, 857–895. [5]

Collin-Dufresne, P., K. Daniel, and M. Saglam (2020): "Liquidity regimes and optimal dynamic asset allocation," *Journal of Financial Economics*, 136(2), 379–406. [20]

Constantinides, G. M. (1986): "Capital Market Equilibrium with Transaction Costs," *Journal of Political Economy*, 94, 842–862. [20]

Cornell, B., and E. Sirri (1992): "The Reaction of Investors and Stock Prices to Insider Trading," *Journal of Finance*, 47(3). [3]

Davis, M., and A. Norman (1990): "Portfolio Selection with Transaction Costs," *Math. Operations Res.*, 15, 676–713. [20]

DeMarzo, P., M. Fishman, and K. Hagerty (1998): "The Optimal Enforcement of Insider Trading Regulations.," *Journal of Political Economy*, 106(3), 602–632. [5, 6, 7]

Easterbrook, F. H. (1985): "Insider Trading as an Agency Problem," *in Principals and Agents: The Structure of Business, John W. Pratt and Richard Zeckhauser, eds., Harvard Business School Press*. [5]

Frino, A., S. Satchell, B. Wong, and H. Zheng (2013): "How much does an Illegal Insider Trade?," *International Review of Finance*, 13(2), 241–263. [7]

Garleanu, N., and L. Pedersen (2013): "Dynamic Trading with Predictable Returns and Transaction Costs," *Journal of Finance*, 68(6). [20]

Glosten, L. R. (1989): "Insider Trading, Liquidity, and the Role of the Monopolist Specialist," *The Journal of Business*, 62(2), 211–235. [5]

Kacperczyk, M., and E. Pagnotta (2019): "Becker Meets Kyle: Inside Insider Trading," *Working paper*. [3, 6, 39]

Kyle, A. (1985): "Continuous Auctions and Insider Trading," *Econometrica*, 53(6), 1315–1335. [8]

Leland, H. (1992): "Insider Trading: Should It Be Prohibited?," *Journal of Political Economy*, 100(4), 859–887. [5]

Liu, H. (2004): "Optimal Consumption and Investment with Transaction Costs and Multiple Risky Assets," *Journal of Finance*, 59(1), 289–338. [20]

Lynch, A., and S. Tan (2010): "Multiple Risky Assets, Transaction Costs, and Return Predictability: Allocation Rules and Implications for U.S. Investors," *Journal of Financial and Quantitative Analysis*, 45(4), 1015–1053. [20]

Manne, H. (1966): "Insider Trading and the Stock Market.," *The Free Press, New York*. [5, 27]

Manove, M. (1989): "The Harm from Insider Trading and Informed Speculation," *Quarterly Journal of Economics*, 104(4), 823–845. [5]

McLennan, A., P. Monteiro, and R. Tourky (2017): "On uniqueness of equilibrium in the Kyle model," *Math. Finan. Econ.*, 11, 161–172. [4]

MILGROM, P., AND I. SEGAL (2002): "Envelope Theorems for Arbitrary Choice Sets," *Econometrica*, 70(2), 583–601. [33, 63]

NEWKIRK, T., AND M. ROBERTSON (1998): "Insider Trading: A US Perspective," *Speech by SEC Staff*. [7, 41]

ROCHET, J.-C., AND J.-L. VILA (1994): "Insider Trading without Normality," *Review of Economic Studies*, 61, 131–152. [4]

SHIN, J. (1996): "The Optimal Regulation of Insider Trading," *Journal of Financial Intermediation*, 5, 49–73. [5, 6, 21, 39, 54]

SHKILKO, A. (2019): "Insider Trading Under the Microscope," *Working paper*. [3]

Online Appendix

A Additional Proofs

A.1 Lemma 2

First, let us show that $\mathcal{X}_C(v)$ is never empty. Let $v \in [-1,1]$, the function $\psi_C(.,v)$ has a finite upper bound as $C \geq 0$. Let $M = \sup_x \psi_C(x,v) < \infty$ and (x_n) such that $\psi_C(x_n,v) \to M$. There is an extraction of (x_n), still denoted (x_n), such that x_n converges to x and either (i) (x_n) is increasing or (ii) (x_n) is decreasing. By symmetry, we can assume without loss of generality that $x > 0$ or $x = 0$ and the case (ii) holds. Let us first consider case (i). Since C is left-continuous and $x \mapsto x\left(v - \frac{x}{2}\right)$ is continuous, $\psi_C(x_n,v)$ converges to $\psi_C(x,v)$: therefore $\psi_C(x,v) = M$ and $x \in \mathcal{X}_C(v)$. Let us now consider case (ii). Since C is non decreasing, it has a right limit at x denoted by $C(x^+)$ which is greater than $C(x)$. Taking the limit in the definition of $\psi_C(x_n,v)$, the value of $\psi_C(x_n,v)$ converges to $x\left(v - \frac{x}{2}\right) - C(x^+) \leq x\left(v - \frac{x}{2}\right) - C(x)$. Using the fact that $\psi_C(x_n,v)$ converges to M, we conclude that $C(x^+) = C(x)$ and $\psi_C(x,v) = M$.

Now, let us show that $\mathcal{X}_C$ is a non-decreasing correspondence. Let $v_1 < v_2$ in $[-1;1]$ and $x_1^* \in \mathcal{X}_C(v_1)$ and $x_2^* \in \mathcal{X}_C(v_2)$. For any $x \in [-1,1]$:

$$\psi_C(x,v_2) = \psi_C(x,v_1) + (v_2 - v_1)x.$$

Using the fact that $x_1^* \in \mathcal{X}_C(v_1)$ and $v_1 < v_2$, for any $x < x_1^*$,

$$\psi_C(x, v_2) < \psi_C(x_1^*, v_1) + (v_2 - v_1)x_1^* = \psi_C(x_1^*, v_2).$$

By definition, $\psi_C(x_2^*, v_2) \geq \psi_C(x_1^*, v_2)$, thus $x_2^* \geq x_1^*$. Since this inequality holds for any $x_1^* \in \mathcal{X}_C(v_1)$ and $x_2^* \in \mathcal{X}_C(v_2)$, we get that $\sup \mathcal{X}_C(v_1) \leq \inf \mathcal{X}_C(v_2)$: the correspondence $\mathcal{X}_C$ is non-decreasing.

But we also know that if $\mathcal{X} : V \to \mathcal{P}(I) \setminus \emptyset$ is a non-decreasing correspondence, then for all v in V except on a countable set, $\mathcal{X}(v)$ is a singleton. This implies the second part of the Lemma and concludes the proof.

A.2 Proposition 1

Using Lemma 4, we can write

$$\begin{aligned} -G &= \int_0^1 X(v)\left(v - \frac{X(v)}{2}\right) \mathrm{d}v \\ &= 1 - \sqrt{3}S - \frac{1}{2}\int_0^1 X(v)^2 \, \mathrm{d}v. \end{aligned} \tag{30}$$

By Cauchy-Schwarz inequality

$$\left(\int_0^1 vX(v)dv\right)^2 \leq \int_0^1 v^2 dv \int_0^1 X(v)^2 \, \mathrm{d}v \tag{31}$$

$$\begin{aligned} &\leq \frac{1}{3}\int_0^1 X(v)^2 \, \mathrm{d}v \\ -\frac{1}{2}\int_0^1 X(v)^2 \, \mathrm{d}v &\leq -\frac{3}{2}\left(\int_0^1 vX(v)\right)^2 \mathrm{d}v = -\frac{3}{2}(1-\sqrt{3}S)^2. \end{aligned}$$

Plugging this into (30), we obtain

$$G \geq \sqrt{3}S - 1 + \frac{3}{2}(1-\sqrt{3}S)^2. \tag{32}$$

This inequality determines the highest possible S given G. But there is equality in (32) if and only if there is equality in the Cauchy-Schwarz bound (31). This is the case if and only if the two functions in the left-hand side are colinear, i.e. if $X(v)$ is proportional to v: $X(v) = \beta v$. Since $0 \leq X(v) \leq 1$ for $0 \leq v \leq 1$, $\beta \in [0;1]$. We conclude by noting that if $\beta \in [0;1]$ and $\gamma \in [0;\infty]$ is defined by $\gamma = \frac{1}{2\beta} - \frac{1}{2}$, the quadratic penalty $C(x) = \gamma x^2$ implements $X(v) = \beta v$.

A.3 Lemma 6

We first need to introduce some definitions:

Let f be a function defined over $[0,1]$ and $x \in [0,1]$. We define:

$$\begin{aligned}\overline{D}^{-} f(x) &= \limsup_{x' \nearrow x} \frac{f(x') - f(x)}{x' - x}, \\ \underline{D}^{-} f(x) &= \liminf_{x' \nearrow x} \frac{f(x') - f(x)}{x' - x},\end{aligned}$$

One can define similarly $\overline{D}^{+} f(x)$ and $\underline{D}^{+} f(x)$. Let us recall the first order conditions satisfied by a function at a local maximum.

If x^* is a local maximum of f, then:

$$\begin{aligned}&\overline{D}^{+} f(x^*) \le 0, \\ &\underline{D}^{-} f(x^*) \ge 0\end{aligned}$$

We will also use the following real analysis result:

Lemma 7 *Any continuous function f on $]0,1]$ with a null left derivative is contant.*

Let C be a penalty function such that the strategy of the IT satisfies that for any $v \in [0,1]$, $X(v)$ is either 0 or v. Since the strategy of the IT is non-decreasing, there exists v_0 such that $X(v) = 0$ for any $v \in [0, v_0[$ and $X(v) = v$ for any $v \in]v_0, 1]$.

Besides, the penalty function C must be continuous on $]v_0, 1]$. Indeed, if $v' > v \ge v_0$, using the fact that $X(v') = v'$,

$$v\left(v' - \frac{v}{2}\right) - C(v) \le v'\left(v' - \frac{v'}{2}\right) - C(v'),$$

thus, since C is non-decreasing,

$$0 \le C(v') - C(v) \le v'\left(v' - \frac{v'}{2}\right) - v\left(v' - \frac{v}{2}\right).$$

Taking the limit as v' goes to v, we see that C is right continuous at v. Since by hypothesis it is left continuous on $[0,1]$, the penalty function C is continuous on $]v_0, 1]$.

Let us show that C has a null left derivative on $]v_0, 1]$. If $v \in]v_0, 1]$, we know that v is a profit maximizer at v: $v \in \arg\max_x f_v(x)$:. Using the first order condition for the lower left derivative $\underline{D}^{-}$ recalled above, at v, $\underline{D}^{-} f_v(v) \ge 0$. Since $\underline{D}^{-} f_v(v) = -\overline{D}^{-} C(v)$, we obtain $\overline{D}^{-} C(v) \le 0$. Yet, C is increasing, so the lower and upper left derivatives must be positive : $0 \le \underline{D}^{-} C(v) \le \overline{D}^{-} C(v)$. Thus:

$$\underline{D}^{-} C(v) = \overline{D}^{-} C(v) = 0.$$

This means that the cost function C admits a left derivative at any $v \in]v_0, 1]$, and the value of this left derivative is zero.

Thus C is continuous and has a null left derivative on $]v_0, 1]$. Using Lemma 7, we obtain that C is constant on $]v_0, 1]$. Let us denote by K the value of C on this interval.

The IT does not trade for $v \in [0, v_0)$. In that case, since we know that $0 \leq X(v) \leq v$, we must have

$$\forall\, x \in [0, v], \quad x\left(v - \frac{x}{2}\right) \leq C(x).$$

By continuity of the left-hand term and the fact that the right-hand term is non-decreasing, we obtain

$$\forall\, x \in [0, v_0], \quad x\left(v_0 - \frac{x}{2}\right) \leq C(x).$$

There must be equality for $x = v_0$, because otherwise it would not be optimal to select $X(v) = v$ on the right neighborhood of v_0. For the same reason, C can not jump at v_0. This implies that $v_0\left(v_0 - \frac{v_0}{2}\right) = K$, or $v_0 = \sqrt{2K}$ and therefore C must belong to $\mathcal{O}$.

Assume conversely that $C \in \mathcal{O}$. Then for $0 \leq v < v_0$, the insider trader will make negative expected profits if she trades, so that $X(v) = 0$. For $v > v_0$, there are two cases to consider. (i) The IT plays $x \geq v_0$. In that case, K appears as a sunk cost and the best choice is $x = v$, leading to a net profit of $\frac{v^2}{2} - K$. (ii) The IT plays $x \in [0, v)$. The net profit is then

$$\begin{aligned}
x\left(v - \frac{x}{2}\right) - C(x) &= x\left(v_0 - \frac{x}{2}\right) - C(x) + x(v - v_0) \\
&\leq x(v - v_0) \\
&\leq v_0(v - v_0)
\end{aligned}$$

where the second line uses the fact that $C \in \mathcal{O}$. Since

$$\begin{aligned}
\frac{v^2}{2} - K &= \frac{v^2}{2} - \frac{v_0^2}{2} \\
&= \frac{1}{2}(v + v_0)(v - v_0) \\
&> v_0(v - v_0),
\end{aligned}$$

choice (i) is always preferred. Hence, if $C \in \mathcal{O}$, $X(v) = 0$ for $|v| < v_0$ and $X(v) = v$ for $|v| > v_0$, which concludes the proof.

A.4 Theorem 7

First, define for $0 \leq \alpha \leq 1 - \sqrt{2K}$:

$$X_\alpha(v) = \begin{cases} v & 0 \le v \le \alpha \\ \alpha & \alpha < v \le \alpha + \sqrt{2K} \\ v & v > \alpha + \sqrt{2K} \\ -X_\alpha(-v) & v < 0. \end{cases}$$

We will see that the X_α for $0 \le \alpha \le 1 - \sqrt{2K}$ are exactly the demand schedules that implement the lower bound in Theorem 7. (Note that these demand schedules are implemented by the penalties C_α where $C_\alpha(x) = K\mathbb{I}_{|x|>\alpha}$.)

Step 1: transformation of the problem into a constrained problem of L^2 distance maximization.

Recall Equation (20):

$$|G| = \frac{1}{6} - \frac{1}{2}\int_0^1 (v - X(v))^2 \, \mathrm{d}v.$$

This means that obtaining the bound of the theorem is equivalent to showing

$$\max_{C \in \mathcal{C}_K} \int_0^1 (v - X(v))^2 \, \mathrm{d}v = \frac{(2K)^{3/2}}{3}, \tag{33}$$

subject to the constraint that $X(v)$ maximizes the net profit $\psi_C(., v)$.

Let $g(v) = v - X(v)$, so that we are looking for an upper bound of $\int_0^1 g^2$. By Lemma 3 and under the constraint $C \le K$, we obtain:

$$\begin{aligned} \int_0^1 g &= \int_0^1 v \, \mathrm{d}v - \int_0^1 X(v) \, \mathrm{d}v \\ &= \frac{1}{2} - \pi^N(1) \\ &\le K. \end{aligned} \tag{34}$$

This is because, when $v = 1$, the IT can achieve at least a net profit of $\frac{1}{2} - C(1) \ge \frac{1}{2} - K$. Therefore, the maximum in (33) is less or equal to

$$\sup \int_0^1 g^2$$

subject to the constraints (i) $\int_0^1 g \le K$, and (ii) $g(0) = 0 \le g(v)$ and $v \mapsto v - g(v)$ is non-decreasing. (i) comes from (34), and (ii) is an immediate consequence of the properties of an optimal demand schedule X.

Notice how crucial Lemma 3 is, and therefore how effective the result of Milgrom and Segal (2002) is. Once noted that $C(1) \leq K$ implies a lower bound on the net profit at 1, Lemma 3 allows (i) to incorporate the constraint the X is a maximizer in a parsimonious way, (ii) to reduce the two constraints–$C \leq K$ and X must maximize ψ_C–into a single condition, $\int g \leq K$, which is particularly convenient, as it is a L^1 bound in a L^2 maximization problem.

Absent the fact that X must be non-decreasing, which translates into the fact that $v \mapsto v - g(v)$ is non-decreasing, the maximization of $\int g^2$ subject to $\int g = K$ (and $0 \leq g(v) \leq v$) would be standard: to "spread mass as unenvenly as possible", one would pick $g(v) = v\mathbb{I}_{v \geq v^*}$ with $\int_{v^*}^1 v \, \mathrm{d}v = K$. This is not feasible, however, because it violates the monotonicity constraint. The $g_\alpha : v \mapsto v - X_\alpha(v)$ are then natural candidate maximizers, as they are constructed in a similar spirit of variance maximization, but respect the monotonicity constraint.

The g_α all have the same L^2 norm, but are away from zero over different intervals. This hints at the fact that for a general function g, when trying to find a bound on $\int g^2$, we will have no way to know where g must be small or large, and therefore little grip on g. The idea is then to consider the repartition function φ of g, because (i) one can reconstruct the moments of g with those of φ (see Step 3) and (ii) it does not matter *where* g is large, only how often it is large. In fact, all the g_α have the same repartition function, which suggests that this is the correct perspective to adopt.

For any function f and $x \neq y$, let

$$\tau_{x,y} f = \frac{f(y) - f(x)}{y - x}.$$

Since X is non-decreasing, we have

$$\tau_{x,y} g \leq 1 \tag{35}$$

for all $x \neq y$. Now, define

$$\varphi(z) = \mu\left(\{x, \, g(x) \geq z\}\right).$$

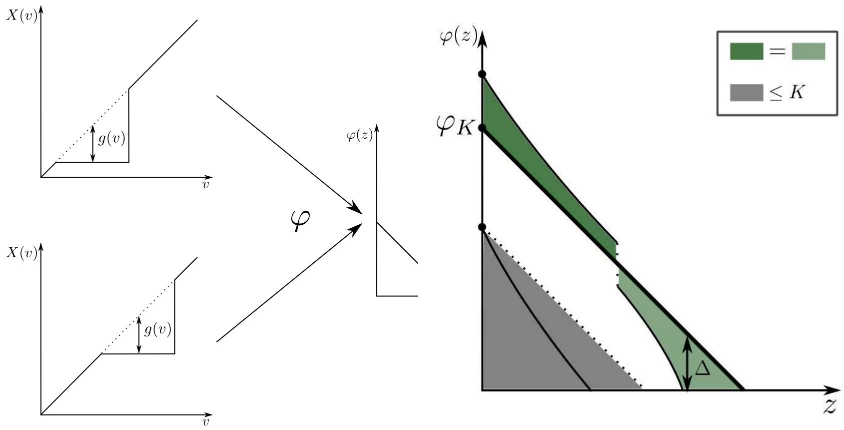
Figure 16a provides a graphical representation.

Step 2: (35) implies

$$\tau_{x,y} \varphi \leq -1 \tag{36}$$

for all $x < y$ such that $\varphi(y) > 0$.

g is subject to a monotonicity constraint (namely $v \mapsto v - g(v)$ must be non-decreasing), which we need to transform into a constraint for φ. Clearly, if g increases at speed 1, φ decreases at speed 1. What we show here is that if g increases at speed less than 1 then



(a) Using the repartition function to transform g

(b) The transform φ of a maximizer g must be φ_K.

Figure 16: Illustrating some steps of the proof

φ decreases at speed larger than 1.

Since $\varphi(y) > 0$, the set $\{u,\, g(u) \geq y\}$ is nonempty, so we can consider

$$u^+ = \inf\{u,\, g(u) \geq y\}.$$

Since $g(0) = 0 \leq x$ we can also define

$$u^- = \sup\{u \leq u^+,\, g(u) \leq x\}.$$

Because of (35), the function g can not jump upwards, hence $g(u^-) = x$ and $g(u^+) = y$. By construction of u^- and u^+, we have:

$$[u^-, u^+) \subset \{u,\, g(u) \in [x, y)\}. \tag{37}$$

Since $\tau_{u^-,u^+} g \leq 1$, we have:

$$u^+ - u^- \geq g(u^+) - g(u^-) = y - x, \tag{38}$$

We can now obtain (36):

$$\begin{aligned}
\tau_{x,y}\varphi &= \frac{\mu\left(\{u,\, g(u)\geq y\}\right)-\mu\left(\{u,\, g(u)\geq x\}\right)}{y-x}\\
&= -\frac{\mu\left(\{u,\, g(u)\in[x,y)\}\right)}{y-x}\\
&\leq -\frac{\mu\left([u^-,u^+)\right)}{y-x}\\
&\leq -1.
\end{aligned}$$

Line 3 uses (37) and Line 4 is a consequence of (38).

Step 3: expression of the moments of g as a function of the moments of φ.

Recall that

$$\begin{aligned}
\int_0^1 g &= \int_0^1 \varphi\\
\int_0^1 g^2 &= 2\int_0^1 y\varphi(y)\,\mathrm{d}y.
\end{aligned} \tag{39}$$

Indeed,

$$\begin{aligned}
\int_0^1 g^2(y)\,\mathrm{d}y &= \int_0^1\int_0^1 \mathbb{I}_{0\leq s\leq g^2(y)}\,\mathrm{d}s\,\mathrm{d}y\\
&= \int_0^1 \mu\left(\left\{u,\, g^2(u)\geq s\right\}\right)\,\mathrm{d}s\\
&= \int_0^1 \mu\left(\left\{u,\, g(u)\geq \sqrt{s}\right\}\right)\,\mathrm{d}s\\
&= 2\int_0^1 y\varphi(y)\,\mathrm{d}y,
\end{aligned}$$

by using the change of variable $y=\sqrt{s}$. The other equality in (39) is proven similarly.

Step 4: translation into a functional maximization problem with respect to the transform φ.

Using the previous discussion,

$$\begin{aligned}
\sup_{C\in\mathcal{C}_K}\int_0^1 (v-X(v))^2\,\mathrm{d}v &\leq 2\sup_{\varphi\in\Phi_K^{\leq}}\int_0^1 y\varphi(y)\,\mathrm{d}y\\
&\leq 2\sup_{\varphi\in\Phi_K}\int_0^1 y\varphi(y)\,\mathrm{d}y
\end{aligned} \tag{40}$$

where $\Phi_K^{\leq}$ is the set of measurable functions $\left\{\varphi : [0,1] \to [0,1],\, \sup_{x,y} \tau_{x,y}\varphi \leq -1,\, \int_0^1 \varphi(y)\,\mathrm{d}y \leq K\right\}$ and $\Phi_K = \{\varphi \in \Phi_K^{\leq}, \int_0^1 \varphi = K\}$. Clearly, in (40) the right-hand-side of Line 1 equals the term in Line 2.

Define $\varphi_K(z) = \max\left\{\sqrt{2K} - z,\, 0\right\}$ for $0 \leq z \leq 1$. Note that $\varphi_K \in \Phi_K$. If $\varphi \in \Phi_K$, $\varphi(0) \geq \varphi_K(0)$. Otherwise, using the fact that $\tau_{0,y}\varphi \leq -1$,

$$\varphi(y) \leq \varphi(0) - y < \varphi_K(0) - y \leq \varphi_K(y).$$

Hence, $\int_0^1 \varphi(y)\,\mathrm{d}y$ would be strictly less than $K = \int_0^1 \phi_K(y)\,\mathrm{d}y$.
Define $\Delta = \varphi - \varphi_K$: we proved that $\Delta(0) > 0$. Besides, by construction, $\int_0^1 \Delta(y)\,\mathrm{d}y = 0$. Define

$$y_0 = \inf\{y,\, \Delta(y) \leq 0\}.$$

Because $\tau_{y_0,y}\varphi \leq -1$, we have $\Delta(y) \leq 0$ for $y > y_0$ and $\Delta(y) \geq 0$ for $y < y_0$. Hence:

$$\begin{aligned}
\int_0^1 y\varphi(y)\,\mathrm{d}y - \int_0^1 y\varphi_K(y)\,\mathrm{d}y &= \int_0^1 y\Delta(y)\,\mathrm{d}y \\
&= \int_0^{y_0} y\Delta(y)\,\mathrm{d}y + \int_{y_0}^1 y\Delta(y)\,\mathrm{d}y \\
&\leq y_0 \int_0^{y_0} \Delta(y)\,\mathrm{d}y + y_0 \int_{y_0}^1 \Delta(y)\,\mathrm{d}y \leq 0.
\end{aligned}$$

Thus, the supremum in (40) is attained only by the function φ_K and equal to

$$\begin{aligned}
2\int_0^1 y\varphi_K(y)\,\mathrm{d}y &= \int_0^{\sqrt{2K}} y(\sqrt{2K} - y)\,\mathrm{d}y \\
&= \frac{(2K)^{3/2}}{3},
\end{aligned}$$

which establishes the bound of the theorem.

Figure 16b provides some intuition: (i) starting from a point $\varphi(0) < \varphi_K(0)$ (lowest thick dot on the y-axis), φ (solid black curved line) remains below the dotted line and its integral is therefore smaller than the area of the grey region, itself below K. (ii) After crossing φ_K, φ must remain below φ_K. Here, the crossing occurs through a downwards jump of φ.
Step 5: The maximum in (33) is attained exclusively by the demand schedules $(X_\alpha)_{\alpha \in [0, 1-\sqrt{2K}]}$ defined in the theorem.

First, it is easy to see that these demand schedules achieve the maximum in (33). It remains to show that they are the only one to do so. Let X be a demand schedule obtained under a penalty $C \in \mathcal{C}$, $C \leq K$. Let us suppose that it achieves the maximum in (33). Consider, as in step 2, the function φ associated with $g(v) = v - X(v)$. The function φ is then a supremum of (40) and by step 3, $\varphi = \varphi_K$. Since

$$\sup_x g(x) \geq \sup\{x,\, \varphi(x) > 0\} = \sup\{x,\, \varphi_K(x) > 0\} = \sqrt{2K},$$

the supremum of $g(v)$ is at least $\sqrt{2K}$. Let us remark that:

$$\sup_v g(v) = \sup_v \sup_{s\in[0,v]} g(s).$$

Since $\tau_{.,.}g \leq 1$, the function $\overline{g}(v) = \sup_{s\in[0,v]} g(s)$ is continuous: the supremum of $\overline{g}(v)$ and thus of $g(v)$ is attained at a point v_0. Since $\tau_{.,.}g \leq 1$, $v_0 \geq \sqrt{2K}$ and for $v \in [v_0 - \sqrt{2K}, v_0]$, $g(v) \geq v - v_0 + \sqrt{2K}$. Since $g \geq 0$, we obtain

$$\begin{aligned}
\int_0^1 g &\geq \int_{v_0-\sqrt{2K}}^{v_0} g \\
&\geq \int_{v_0-\sqrt{2K}}^{v_0} (v - v_0 + \sqrt{2K})\,\mathrm{d}v \\
&\geq K
\end{aligned}$$

with equality if and only if $g = 0$ outside $[v_0 - \sqrt{2K}, v_0]$ and $g(v) = v - v_0 + \sqrt{2K}$ over $[v_0 - \sqrt{2K}, v_0]$. But there must be equality because $g \in \Phi_K$. Hence g has the above form, and the demand function X, given by $X(v) = v - g(v)$, is equal to X_α as stated in the theorem, with $\alpha = v_0 - \sqrt{2K}$.

A.5 Theorem 8

As a consequence of Lemma 3, in equilibrium the expected fine satisfies

$$\mathbb{E}[C(X(v))] = \int_0^1 X(v)\left(v - \frac{X(v)}{2}\right)\mathrm{d}v - \int_0^1 (1-v)X(v)\,\mathrm{d}v,$$

and we are working under a constraint $\mathbb{E}[C(X(v))] \geq K_1$.

By Lemma 4, an upper bound constraint on the expected post-trade standard deviation translates into a constraint

$$\int_0^1 vX(v)dv \geq K_2.$$

This leads us to consider the following minimization problem:

$$\min_X \quad \int_0^1 X(v)\left(v - \frac{X(v)}{2}\right) \mathrm{d}v + \gamma\left(K_1 - \int_0^1 X(v)\left(v - \frac{X(v)}{2}\right) \mathrm{d}v + \int_0^1 (1-v)X(v)\,\mathrm{d}v\right)$$
$$+ \quad \eta\left(K_2 - \int_0^1 vX(v)\,\mathrm{d}v\right),$$

for some weights $\gamma, \eta \geq 0$. Gathering terms, we obtain that this program is equivalent to

$$\min_X \int_0^1 X(v)\left(\gamma + (1-2\gamma-\eta)v + \frac{\gamma-1}{2}X(v)\right) \mathrm{d}v \tag{41}$$

For $0 \leq v \leq 1$, define

$$\begin{aligned} P_v : [0,v] \quad &\to \mathbb{R} \\ x \quad &\mapsto x\left(\gamma + (1-2\gamma-\eta)v + \frac{\gamma-1}{2}x\right) \end{aligned}$$

Case 1: $\gamma > 1$. P_v is the restriction to $[0,v]$ of a second-order polynomial with positive leading coefficient. Therefore it reaches its minimum at either 0, v, or when the first order condition is satisfied, say at $x_0(v)$, and $x_0(v)$ achieves the minimum as soon as $0 \leq x_0(v) \leq v$. Given that

$$x_0(v) = \frac{(2\gamma+\eta-1)v - \gamma}{\gamma-1},$$

algebra shows that

$$\arg\max\, P_v = \begin{cases} 0 & v \leq \frac{\gamma}{2\gamma+\eta-1} \\ x_0(v) & \frac{\gamma}{2\gamma+\eta-1} \leq v \leq \frac{\gamma}{\gamma+\eta} \\ v & v > \frac{\gamma}{\gamma+\eta}. \end{cases}$$

Let $v_1 = \frac{\gamma}{2\gamma+\eta-1}$ and $v_2 = \frac{\gamma}{\gamma+\eta}$. We have obtained that with the function X_{v_1,v_2} given in the theorem, the equality

$$\arg\max P_v = X_{v_1,v_2}(v)$$

holds. Direct calculations show that X_{v_1,v_2} is implemented by C_{v_1,v_2}. This means that we have found an implementable demand schedule that maximizes the integral in (41) pointwise, which implies that X_{v_1,v_2} is a minimizer of the program (41), and it is the only one because the pointwise minimization of the integral in (41) has a unique solution.

Case 2: $\gamma \leq 1$. P_v is now either linear or with a negative leading coefficient, meaning that its minimum is attained either at 0 or v. Algebra shows that $\arg\max P_v = v$ (for $0 \leq v \leq 1$) if and only if

$$\gamma + 2\eta \geq 1 \tag{42}$$

and

$$v \geq v^* := \frac{\gamma}{\eta + \frac{3\gamma}{2} - \frac{1}{2}},$$

where, by condition (42), $v^* \in [0,1]$. With $v_1 = v_2 = v^*$ we conclude as before that X_{v_1,v_2} is the unique minimizer of (41). Finally, if (42) is not satisfied, the minimizer of (41) is identically zero, which corresponds to $X_{1,1}$ defined in the theorem.

Finally, it is easy to see that the (v_1, v_2) constructed above describe the set J as $\gamma, \eta \geq 0$ vary, and J is the family of indices specified in the theorem. So any index in J corresponds to an efficient demand function. This shows that $(X_{v_1,v_2})_{(v_1,v_2)\in J}$ is the family of efficient demand functions.

A.6 Remark 9

Let $\alpha : [-2,2] \to [0,1]$—with $\alpha(0) = 0$, $\alpha(x) = \alpha(-x)$ for any x, α non-decreasing over $[0,2]$—be a probability of investigation that depends on the total volume. $\tilde{C}$ denotes the penalty conditional on successful investigation. With $\tilde{\alpha}(x) \equiv \mathbb{E}_u[\alpha(x+u)]$, the insider faces ex-ante an expected penalty $C(x) = \tilde{\alpha}(x)\tilde{C}(x)$. We need to check that if $\tilde{C}$ is an admissible penalty (an element of $\mathcal{C}$ introduced in Definition 2) , then so is C. The regularity condition, the facts that $C(0) = 0$ and that C is even are clear. Now for $1 \geq x \geq y \geq 0$,

$$\begin{aligned}
\tilde{\alpha}(x) - \tilde{\alpha}(y) &= \int_{-1}^{1} \alpha(x+u)du - \int_{-1}^{1} \alpha(y+u)du \\
&= \int_{y+1}^{x+1} (\alpha(u) - \alpha(u-2))du \\
&\geq 0.
\end{aligned}$$

The last line is because for any $u \in [y+1, x+1]$, $|u| \geq |u-2|$ and so $\alpha(u) \geq \alpha(u-2)$. This proves the monotonicity condition.

B Computation of the equilibrium under Gaussian noise

We consider a normalized Gaussian noise and create grids for u and v as fine discretizations of $[-5,5]$, so that the probability that u or v realizes outside of this range is neglectible.

The fixed-point algorithm to solve (5) and (6) with Gaussian distributions is the following:

1. Start from an initial guess $X(v) = v$.

2. For a given X, the price function satisfies the theoretical relationship

$$P(d) = \frac{\int_{\mathbb{R}} v\phi(v)\phi(d - X(v))dv}{\int_{\mathbb{R}} \phi(v)\phi(d - X(v))dv},$$

where ϕ is the standard Gaussian p.d.f.. Both integrals in this ratio can be computed numerically using our discretizations of v and X for any d.

3. Similarly, compute the conditional mean square error as

$$Q(d) = \frac{\int_{\mathbb{R}} (v - P(d))^2\phi(v)\phi(d - X(v))dv}{\int_{\mathbb{R}} \phi(v)\phi(d - X(v))dv}.$$

4. Given P, compute $\tilde{P}(x) = \mathbb{E}_u[P(x + u)]$ (a straightforward numerical integration).

5. Find the maximizer of $x \mapsto x(v - \tilde{P}(x)) - C(x)$.

6. Doing so defines a new X; iterate the procedure from point 2 until convergence.

Convergence is strikingly quick: in 10 steps it is always virtually perfect. This gives us the numerical equilibrium (X, P) and Q.

The expected post-trade standard deviation S is obtained by averaging under the Gaussian noise the quantity $Q^{1/2}(X(v) + u)$. The losses of the uninformed traders $-G$ are obtained by averaging under the Gaussian noise the quantity $-u(v - P(X(v) + u))$.

Карр, С.Ж.П, Коллин-Дюфресн, П., Габриэль, Ф.
Инсайдерская торговля со штрафами [Электронный ресурс] : препринт WP9/2021/02 / С.Ж.П. Карр, П. Коллин-Дюфресн, Ф. Габриэль ; Нац. исслед. ун-т «Высшая школа экономики». – Электрон. текст. дан. (1,2 Мб). – М. : Изд. дом Высшей школы экономики, 2021. – (Серия WP9 «Исследования по экономике и финансам»). – 72 с. (На англ. яз.)

Мы устанавливаем существование и уникальность равновесия в обобщенной однопериодной модели Кайла (1985), в которой инсайдерские сделки могут облагаться издержками, не уменьшающимися в размере сделки, - «штрафом». Результат получен для равномерного шума и справедлив для общих штрафных функций. Уникальность среди всех неубывающих стратегий. Мы строим в явном виде равновесную цену и оптимальную политику инсайдерской торговли и обнаруживаем, что, за исключением квадратичных штрафов, оба являются нелинейными функциями соответственно от объема торгов и ликвидационной стоимости.

Наша гибкая структура может использоваться в различных контекстах. В качестве приложения мы характеризуем набор оптимальных штрафов, которые регулирующий орган выберет, чтобы максимизировать информативность цены для заданного уровня ожидаемых потерь неинформированных трейдеров. Эффективные штрафы устраняют мелкие, а не крупные сделки. Мы расширяем анализ на случай, когда реализуемые нормативы ограничены бюджетным ограничением. Мы подтверждаем устойчивость наших основных выводов к нашему предположению о распределении.

Карр Сильвен Жан Паскаль, МИЭФ, Национальный исследовательский университет «Высшая школа экономики»; Российская Федерация, 109028, Москва, Покровский бульвар, д. 11, каб. S708; Email: scarre@hse.ru

Коллин-Дюфресн Пьер, Швейцарский финансовый институт и Федеральная политехническая школа Лозанны; E-mail: pierre.collin-dufresne@epfl.ch

Габриэль Франк, департамент математики, Федеральная политехническая школа Лозанны; E-mail: franck.gabriel@epfl.ch

Препринты Национального исследовательского университета «Высшая школа экономики» размещаются по адресу: http://www.hse.ru/org/hse/wp

Препринт WP9/2021/02
Серия WP9
Исследования по экономике и финансам

Карр Сильвен Жан Паскаль, Коллин-Дюфресн Пьер,
Габриэль Франк

Инсайдерская торговля со штрафами

(*на английском языке*)

Изд. № 2336