J. Econ. Theory 228 (2025) 106061

Contents lists available at ScienceDirect

Journal of Economic Theory

journal homepage: www.elsevier.com/locate/jet

Insider trading with penalties in continuous time

Umut Çetin

Department of Statistics, London School of Economics and Political Science, 10 Houghton st, London, WC2A 2AE, UK

ARTICLE INFO

Keywords:
Private information
Insider trading
Liquidity
Market regulation
Kyle model
Entropic optimal transport

ABSTRACT

This paper addresses the question of how insiders internalize the additional penalties to trade in a continuous time Kyle model. The penalties can be interpreted as non-adverse selection transaction costs or legal penalties due to illegal insider trading. The equilibrium is established for general asset distribution. In equilibrium, the insider does not disseminate her private information fully into the market prices. Moreover, she always trades a constant multiple of the discrepancy between her own valuation and her forecast of market price right before her private information becomes public. In the particular case of normally distributed asset value, the trades are split evenly over time for sufficiently large penalties, with trade size proportional to the return on the private signal. Although the noise traders lose less when penalties increase, the insider's total penalty in equilibrium is non-monotone since the insider trades little when the penalties surpasses the value of the private signal. As a result, a budget-constrained regulator runs an investigation only if the benefits of the investigation are sufficiently high. Moreover, the optimal penalty policy is reduced to choosing from one of two extremal penalty levels that correspond to high and low liquidity regimes. The optimal choice is determined by the amount of noise trading and the relative importance of price informativeness.

1. Introduction

Kyle's (1985) model (together with its more tractable formulation in continuous time by Back (1992)) is the canonical model for analyzing the dissemination of private information into prices in financial markets. In a market consisting of noise traders, competitive market makers, and an informed trader, it predicts that the informed trader gradually reveals all her private information to the market in equilibrium. At the end of the finite trading horizon, the prices become fully informative in the sense that there is no remaining uncertainty about the private valuation of the informed.

The Kyle model does not take into account non-adverse selection trading costs nor does it distinguish between legal informed trading, for example, the investor paying a fee to obtain private information via fundamental research, and illegal insider trading when the information is typically obtained "in breach of fiduciary duty or other relationship of trust and confidence."[1] As there are no penalties for illegal information acquisition in the Kyle model, both types of informed traders behave identically.

Despite its practical relevance and the special importance attached to it by the regulators,[2] the incorporation of legal risk into the Kyle model is relatively less explored in the literature. This paper analyzes the impact of pecuniary penalties on insider trading in a

E-mail address: u.cetin@lse.ac.uk.

[1] See https://www.investor.gov/introduction-investing/investing-basics/glossary/insider-trading for what constitutes illegal insider trading from the point of view of the US Securities and Exchange Commission (SEC).

[2] For example, the SEC considers "the detection and prosecution of insider trading violations as one of its enforcement priorities." See https://www.investor.gov/introduction-investing/investing-basics/glossary/insider-trading.

https://doi.org/10.1016/j.jet.2025.106061
Received 4 June 2024; Received in revised form 12 May 2025; Accepted 26 July 2025

Available online 31 July 2025

continuous time Kyle model. The previous works typically assume a one-period setting (see, e.g., Shin (1996) and Carré et al. (2022)) except for Kacperczyk and Pagnotta (2019, 2023), who study a two-period model.[3] Although some insiders make block trades, in agreement with the nature of these models, they typically split their orders over a relatively long trading period. Indeed, Kacperczyk and Pagnotta (2019) find that "the median trader splits trades over a period equivalent to 71% of the information horizon." In an earlier empirical study, Fishe and Robe (2004) documented that brokers who had advance access to the "Inside Wall Street" column in *Business Week* made abnormal returns by trading rather frequently before their private information became public.[4] Thus, understanding the deterrence power of legal penalties in continuous time is essential for the regulation of illegal insider trading.

In the present paper, the insiders, who know the fundamental value of the asset, trade dynamically in continuous time as in Back (1992) but are also subject to additional transaction costs that are quadratic in the volume traded. The reason behind these extra costs could be trading frictions associated with the execution of a large portfolio. Another interpretation could be the pecuniary penalties that are paid by illegal insiders if they are caught by the authorities as a result of a successful investigation, in addition to losing all their previous gains from trading. The presence of these additional costs, which will be called penalties in the sequel, makes the resolution of equilibrium in closed form less likely. However, Section 3 describes a novel algorithm to establish the existence of equilibrium, which amounts to finding the fixed point of a nonlinear operator T. Section 4 establishes the existence of a fixed point for T under the assumption that the fundamental value has a finite variance.

Although the introduction of penalties adds considerable difficulty to the computation of equilibrium, the analysis in Section 5 reveals that all equilibrium parameters, such as pricing rule, insider's wealth, noise traders' loss, expected penalties, and price efficiency, are given in closed form up to the determination of the fixed point of the operator T.

A key result of the paper is that, at all times, the insiders trade the same constant multiple of the difference between the fundamental value and what they think the market price will be right before their private information becomes public. This constant is inversely proportional to the rate of penalties; that is, the parameter determining the scale of quadratic transaction costs. When transaction costs are interpreted as expected legal penalties, this parameter is proportional to the success probability of investigations and the rate of pecuniary penalties, in line with the empirical evidence in Frino et al. (2013). The insiders' expectation of the terminal market price, however, is time dependent and is updated as liquidity shocks arrive. As a result, the insider's trade is linear in their private information but nonlinear in the total demand for the asset.[5] This is a clear deviation from the equilibrium without penalties, where the insider's strategy is linear in both. (See Back (1992).)

Another remarkable property of the equilibrium is that the insider's strategy depends on the distribution of the fundamental value of the asset. This is a striking difference from the previous extensions[6] of the Kyle model in continuous time with risk-neutral agents, where the insider always employs the same bridge strategy, irrespective of the asset distribution, since the maximum profit is obtained only if the prices converge to the fundamental value, which is achieved when the terminal demand for the asset equals a specific Gaussian random variable known to the insider. However, when there are additional costs to trades, such a strategy becomes prohibitively expensive and the insider's trading intensity depends on the value of the private information. In particular, the insiders do not disseminate their private information fully into the market prices as opposed to the unconstrained insider in the aforementioned extensions. This deviation in the equilibrium strategy is due to the trade-off between the gains by bringing the market price as close as possible to the fundamental value, and the losses from the penalties for doing so as the liquidity shocks move prices away from the fundamentals. That is, if the expected penalties are relatively large, the insider is less willing to make big trades for relatively small asset value.

The second part of Section 5 is devoted to an in-depth analysis of the equilibrium with Gaussian payoffs, which allows for an explicit representation of equilibrium parameters.[7] As expected, the insider's wealth decreases as the intensity of penalties increases. However, this does not follow from the plausible explanation that the expected penalties that are incurred in equilibrium are quite high due to the higher fines imposed by the regulators. Rather, the presence of legal risk forces the insiders to moderate their trades, and trade particularly small quantities in the case of high penalties. Although the insiders continue to buy (sell) when the market valuation is below (above) the fundamentals, the amount that they trade decreases in the presence of higher legal fines. As a result, the expected penalties in equilibrium remain bounded even if the regulatory fines increase without bounds. That the insiders trade less aggressively aligns with the empirical evidence presented in Kacperczyk and Pagnotta (2019).

Kacperczyk and Pagnotta (2019) also document that insiders split their trades evenly over time. This is at odds with the behavior of Kyle's informed trader, who trades especially frequently and with variance increasing to infinity toward the date of the public announcement of the fundamental value. The model considered in the present paper gives a theoretical justification for this empirical observation of Kacperczyk and Pagnotta. Indeed, when faced with high penalties, insiders split their trades into "almost" equal blocks over the trading horizon with the size of each trade determined by the return on their private signals and the amount of noise trading.

Enforcement of insider trading restrictions is costly in practice. Investigation of potential violations typically involves interviews, examination of trading data, and informal inquiries,[8] and it can lead to a court case when the misconduct warrants it. Section 6 studies

[3] See also DeMarzo et al. (1998) for another study of the regulation of insider trading in a one-period model but outside Kyle's framework.

[4] The average number of daily trades increased from 8.3 to 17.1 from Wednesday to Friday for these brokers, who received the advance copy of the magazine early Thursday afternoon, before its public electronic distribution at 7:00 pm.

[5] The only exception to this is when the asset value is normally distributed.

[6] A non-exhaustive list includes Back (1992), Back and Pedersen (1998), Campi et al. (2011), Campi et al. (2013), Çetin (2018), and Back et al. (2020).

[7] Unreported numerical results show that the qualitative properties of the equilibrium are robust with respect to different distributional assumptions on the fundamental value.

[8] See https://www.sec.gov/enforcement/how-investigations-work for the investigation procedure of the SEC.

the optimal regulation of insider trading by considering a regulator with finite resources. The regulator aims to minimize the losses of uninformed traders but is also concerned with price informativeness. While a high legal penalty will deter illegal trading and reduce the losses of uninformed traders, it will also result in prices being less informationally efficient. Section 6 addresses this dilemma of a regulator, who also faces a budget constraint, by assuming a Gaussian fundamental value. Theorem 6.1 suggests a simple penalty policy that achieves the optimal trade-off between information efficiency and losses of the uninformed: Since the expected penalties in equilibrium are bounded and the insider trades little when facing high penalties, it is optimal for a budget-constrained regulator not to run an investigation if its relative cost exceeds the overall benefits that it provides, even when a whistleblower provides a credible evidence. This particularly occurs if the amount of uninformed trading is relatively low. On the other hand, if the regulator finds the investigation worth pursuing, the budget constraint implies that the optimal penalty rate is reduced to choosing one of two extremal values corresponding to a lower liquidity (higher information efficiency) regime and a higher liquidity (low uninformed losses) regime. A regulator who is more concerned with information efficiency chooses a penalty level that leads to lower liquidity. On the other hand, if the regulator cares more about the losses of the uninformed traders, or, if the uninformed trade volume is relatively large, the penalty level associated with the higher liquidity regime is adopted.

This paper contributes to the theoretical literature on insider trading in the continuous time framework of the Kyle model in two different ways: First, it develops a novel solution methodology for the computation of equilibrium. The quadratic (or, more generally, strictly convex) penalties result in the optimal strategy of the insider being given as the optimizer of a Hamilton-Jacobi-Bellman (HJB) equation as a function of the pricing rule, which will be determined in equilibrium. Section 3 exploits this convexification and shows that the value function of the insider is the sum of two components: While the first component is a solution of a backward linear partial differential equation (PDE) reminiscent of the value function of the Kyle model with no penalties, the second component is given by the solution of a quadratic BSDE (backward stochastic differential equation). Computation of equilibrium then becomes a matter of determining the terminal condition of this BSDE, which is achieved by repeated iterations of an explicit operator T until its fixed point is reached. Moreover, the structure of the operator T makes it amenable to numerical analysis of its fixed point for general distributions. This is fundamentally different from the earlier works[9] on the Kyle model in continuous time, where one needs to make an educated guess for the optimal bridge strategy of the insider and hope that it will also satisfy the market makers in equilibrium.

The existence of a fixed point for the operator T is demonstrated using optimal transport theory. The fixed points of T are identified as *the Schrödinger potentials*, which are linked to the solutions of a system of Schrödinger equations that naturally arise in the context of entropic optimal transport. Optimal transport methods have numerous applications in economics. For instance, Galichon (2016) presents the mathematical theory along with various economic applications. More recent works by Daskalakis et al. (2013) and Daskalakis et al. (2017) utilize optimal transport to explore mechanism design. Kolotilin et al. (2025) address a Bayesian persuasion problem through optimal transport theory. Despite its potential, this theory is scarcely applied in the field of market microstructure. Back et al. (2020) are the first to characterize equilibrium using optimal transport in a continuous-time Kyle model with multiple assets and potentially risk-averse market makers. Kramkov and Xu (2022) investigate equilibrium in a generalization of the single-period model by Rochet and Vila (1994). On the other hand, the theory of entropic optimal transport, a relatively new sub-field of optimal transport in mathematics, has yet to find its way to the economics literature, aside from the application of the Sinkhorn algorithm to approximate optimal transport maps.[10] To the best of author's knowledge, this paper presents the first genuine application of entropic optimal transport in economic theory.

The second theoretical contribution is the incorporation of additional transaction costs (penalties) into the Kyle model in continuous time. The closest works[11] to the model studied herein are Barger and Donnelly (2021), Carré et al. (2022), Kacperczyk and Pagnotta (2019), and Kacperczyk and Pagnotta (2023).

Carré et al. (2022) study a one-period Kyle model that assumes that noise traders and the fundamental value of the asset are uniformly distributed on a bounded interval. However, they consider general convex costs depending on the size of trades that contain quadratic penalties. Carré et al. (2022) shows that constant penalties are the most effective policy for the regulation of insider trading. As explained in Remark 2, the argument does not directly translate to a continuous time setting with dynamic trading, where the market makers infer private information from past trades.

Kacperczyk and Pagnotta perform mainly an empirical analysis of insider trading using the US data in Kacperczyk and Pagnotta (2019) and Kacperczyk and Pagnotta (2023). To provide a theoretical basis for their findings, they consider two-period models with legal penalties on insider trading, where the noise trading has a Gaussian distribution. In Kacperczyk and Pagnotta (2019), the penalties are quadratic in trade size, and an investigation occurs after the trading is complete, possibly following a tip from a whistleblower, as in the present paper. The fundamental value of the asset is also assumed to have a Gaussian distribution. On the other hand, the fundamental value in Kacperczyk and Pagnotta (2023) can have only two distinct values, and the regulators start an investigation if the net instantaneous trade ever becomes too big. As a result of a successful investigation, the insider pays as a penalty a fixed multiple of the total gains, if positive. However, such a regulatory rule will not be meaningful in the continuous time model (such as that of Back (1992)), where the instantaneous trades of the insider, as well as those of noise traders, are infinitesimally small. Moreover, feasibility concerns aside, a penalty on realized profits is in general not an efficient policy for a regulator who is concerned with price informativeness as explained in Remark 3. In contrast to these three models, this paper establishes the existence of equilibrium, and gives its characterization, with general payoffs in continuous time in the presence of quadratic penalties.

[9] See, for example, Back (1992), Back et al. (2000), Back and Pedersen (1998), Back and Baruch (2004), Campi et al. (2011), Campi et al. (2013), Çetin and Danilova (2018), Çetin (2018), and Back et al. (2020), among others.

[10] See Back et al. (2020) for an application.

[11] Other related works include Shin (1996), DeMarzo et al. (1998), and Fishman and Hagerty (1992).

Barger and Donnelly (2021) is the only other work that considers additional transactions costs in a continuous-time Kyle model. They similarly consider quadratic penalties but assume Gaussian asset value. In this sense, the present paper extends their framework as Back (1992) does Kyle (1985). Moreover, the methodology introduced herein allows a more detailed interpretation of equilibrium outcomes as discussed in subsequent sections.

The outline of the paper is as follows. Section 2 presents the model assumptions. Section 3 describes the procedure for constructing the equilibrium and makes the connection with a quadratic BSDE. Section 4 studies the fixed point of a nonlinear operator that yields the terminal condition of the BSDE via equilibrium constraints. Whereas Section 5 analyzes the properties of equilibrium, Section 6 studies the optimal policies for a regulator. Finally, Section 7 concludes and presents directions for future research. Technical results and proofs are contained in the Appendix.

2. The setup

As in Back (1992), the trading will take place on $[0,1]$, and the risk-free interest rate is set to 0. Let $(\Omega, \mathcal{G}, (\mathcal{G}_t)_{t\in[0,1]}, \mathbb{Q})$ be a filtered probability space satisfying the usual conditions. The fundamental value of this asset equals V, which will become public knowledge at time-1. Its distribution has a finite variance but is allowed to contain atoms. Consequently, there exists a non-decreasing f such that $V = f(\eta)$, where η is a standard Normal random variable. The distribution of V is denoted by Π. That is, $\mathbb{Q}(V \in dv) = \Pi(dv)$ for v in the support of V, which is denoted by $f(\mathbb{R})$.

Three types of agents trade in the market, and they differ in their information sets and objectives as follows:

- *Noise traders* are non-strategic, and their total demand at time t is given by σB_t for a standard $(\mathcal{G}_t)$–Brownian motion B independent of V, and constant $\sigma > 0$.
- *Market makers* observe only the net demand

$$Y = \theta + \sigma B,$$

where θ is the demand process of the insider.
The market makers compete in a *Bertrand fashion* and clear the market. Similar to Back (1992), the market makers set the price S as a function of the total order process at time t; that is,

$$S_t = H(t, Y_t), \qquad \forall t \in [0,1). \tag{2.1}$$

- *The informed investor* observes the price process S and V. Different from Back (1992), the informed trader is subject to additional quadratic transaction costs, which implies that her trading strategy is absolutely continuous; that is, $d\theta_t = \alpha_t dt$, for some α adapted to her own filtration. The accumulated transaction cost by time t is defined as

$$C_t := \frac{c}{2}\int_0^t \alpha_s^2 ds,$$

for some $c > 0$, which will be called *the rate* of penalties in the sequel. Since the insider is risk-neutral, the objective is to maximize the expected final wealth; that is,

$$\sup_{\theta\in\mathcal{A}} E^{0,v}\left[W_1^\theta\right], \text{ where} \tag{2.2}$$

$$W_1^\theta = (V - S_{1-})\theta_{1-} + \int_0^{1-} \theta_s dS_s - \frac{c}{2}\int_0^1 \alpha_s^2 ds = \int_0^1 (V - S_s)\alpha_s ds - \frac{c}{2}\int_0^1 \alpha_s^2 ds. \tag{2.3}$$

The set $\mathcal{A}$ above consists of the admissible trading strategies for the given pricing rule, which will be defined in Definition 2.2. The operator $E^{0,v}$ computes the expectation with respect to $P^{0,v}$, which is the regular conditional distribution of $(B_s, V; s \le 1)$ given $B_0 = 0$ and $V = v$, which exists due to Theorem 44.3 in Bauer (1996).
The above quadratic cost can appear due to additional frictions in the execution of the order resulting in non-adverse selection trading costs that are quadratic in trade size.
The expected wealth formulation above may also arise when the insider faces the risk of an investigation that may result in pecuniary penalties. Indeed, suppose that an investigation can successfully identify illegal inside trading with probability p, after which the insider not only loses her gains from trade but also pays a legal penalty of $c_0\int_0^1 \alpha_t^2 dt$. The expected profit of the insider[12] under this scenario is

$$E^{0,v}\left[(1-p)\int_0^1 (V - S_s)\alpha_s ds - pk\int_0^1 \alpha_s^2 ds\right] = (1-p)E^{0,v}\left[\int_0^1 (V - S_s)\alpha_s ds - \frac{pc_0}{1-p}\int_0^1 \alpha_s^2 ds\right].$$

[12] It is assumed that the investigation takes place after trades are over as in the case of a whistleblower.

Thus, the coefficient $\frac{c}{2}$ in (2.3) can be associated with $\frac{pc_0}{1-p}$, which becomes large as the probability of a successful investigation grows, even if the regulator charges a relatively small penalty c_0 per squared trade.

Remark 1. The choice of quadratic penalty is for tractability. As the following sections show, quadratic penalties lead to an explicit solution of equilibrium once the terminal condition of a quadratic BSDE is determined. Moreover, this terminal condition is a fixed point of an operator that admits a numerical solution via a straightforward procedure.

If one opts for a general functional form for the additional transaction costs, it is still possible to characterize the equilibrium via a system of forward and backward SDEs. However, this system will in general not admit a closed-form solution and, thus, won't allow, in particular, for a tractable sensitivity analysis of equilibrium on model parameters.

Remark 2. Carré et al. (2022) show in a one-period model that quadratic penalties are the most inefficient type of penalty for regulators whose objective is to minimize the losses of noise traders for a given level of price efficiency. Interestingly, the optimal choice for the regulators turns out to be a constant penalty for any size of (non-zero) trade. However, when the trading occurs in continuous time, a constant penalty policy cannot be supported in an equilibrium where the insider uses pure strategies. Indeed, such a policy implies that the insider makes only finitely many block trades since, otherwise, a potentially infinite loss is expected. For simplicity assume that $V \in \{0,1\}$ so that the high-type insider buys and the low type sells. On the other hand, since the noise trading is continuous, the market makers will know that all block trades are coming from the insider and set the price equal to V immediately when the insider trades.

Another alternative policy that uses constant penalty is to assign a fixed penalty upon the insider being caught. However, this means that the insider's expected wealth is given by

$$(1-p)E^{0,v}\left[\int_0^1 (V-S_s)\alpha_s ds\right] - pc,$$

where c is the fixed penalty. In this case, unless $pc/(1-p)$ is larger than the maximal wealth of the insider in Back (1992), the insider will employ the same bridge strategy of Back (1992), where the insider trading is not penalized. On the other hand, for large enough values of c, the insider will choose not to trade at all, which will result in a completely inefficient market. Thus, enforcement of fixed penalties in the continuous time Kyle model will result in either the insider trading as though no penalties apply, or not trading at all. Having said that, this conclusion mainly rests on the assumption that the probability of the insider being charged is independent from everything else. If this probably depends on the model parameters such as trade volume. The above arguments break down and one needs a further analysis to investigate the deterrence power of such fixed penalties.

Remark 3. In practice, the regulators assign penalty as a fixed multiple of the profit of the insider. This amounts to the following valuation of the insider's final wealth under the assumption that the success probability of the investigations are independent of model parameters:

$$\begin{aligned}
&(1-p)E^{0,v}\left[\int_0^1 (V-S_s)\alpha_s ds\right] - pcE^{0,v}\left[\left(\int_0^1 (V-S_s)\alpha_s ds\right)^+\right] \\
&= (1-p)E^{0,v}\left[-\left(\int_0^1 (V-S_s)\alpha_s ds\right)^- + \frac{1-p-pc}{1-p}\left(\int_0^1 (V-S_s)\alpha_s ds\right)^+\right],
\end{aligned}$$

where $x^+ = \max\{x,0\}$ and $x^- = -\min\{x,0\}$.

Suppose that V is normally distributed to enable a simple comparison to the Kyle model. Then, if $1-p \le pc$, the insider never trades since $\int_0^1 (V-S_s)\alpha_s^* ds > 0$ for α^* being the optimal strategy of the unconstrained insider in Back (1992). Otherwise, the insider will precisely follow the optimal strategy of Back (1992) despite the legal risk.

Carré et al. (2022) measure the efficiency of a particular penalty policy by comparing the loss of noise traders across penalty functions for a given level of price efficiency. Since the prices are completely inefficient when the insider does not trade, a regulator using the above penalty policy must choose the penalty rate c small enough so that $pc < 1-p$. However, this results in the biggest loss to the noise traders since the insider will trade exactly as the unconstrained insider of Back (1992). Thus, one can argue that the current practice of penalizing the insider activity by a pre-determined multiple of their profit is rather inefficient if one is also concerned with price efficiency.

Given the market structure above, one can now define the filtrations of both the market makers and the insider. To do so, first consider $\mathcal{F} := \sigma(B_t, V; t \le 1)$. Due to the measurability of regular conditional distributions, one can define the probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$ by

$$\mathbb{P}(E) = \int_{f(\mathbb{R})} P^{0,v}(E)\Pi(dv), \tag{2.4}$$

for any $E \in \mathcal{F}$.

The market makers' filtration, denoted by $\mathcal{F}^M$, will be the right-continuous augmentation with the $\mathbb{P}$-null sets of the filtration generated by Y. Insider's filtration, $\mathcal{F}^I$, will be similarly constructed as a right continuous filtration generated by S and V. However, its construction requires more technical care and is therefore delegated to the Appendix. Note that $\mathcal{F}^I = \mathcal{F}^{B,V}$, that is, the filtration generated by B and V, if H is strictly increasing in total demand, which will be the case when H is an *admissible pricing rule*, per Definition 2.1.

An equilibrium is a pair consisting of an *admissible* pricing rule and an *admissible* trading strategy such that: *a)* given the pricing rule, the trading strategy is optimal, and *b)* given the trading strategy, the pricing rule is *rational* in the following sense:

$$H(t, Y_t) = S_t = \mathbb{E}[V|\mathcal{F}_t^M], \tag{2.5}$$

where $\mathbb{E}$ corresponds to the expectation operator under $\mathbb{P}$. To formalize this definition of equilibrium, one must introduce the notions of admissible pricing rules and trading strategies.

Definition 2.1. An *admissible pricing rule* is any function H such that $H \in C^{1,2}([0,1) \times \mathbb{R}) \cap C([0,1] \times \mathbb{R})$, and $x \mapsto H(t,x)$ is strictly increasing for every $t \in [0,1]$.

Definition 2.2. An $\mathcal{F}^{B,V}$-adapted θ is said to be an admissible trading strategy for a given admissible pricing rule H, if the following conditions are satisfied:

(1) θ is absolutely continuous and of finite variation; that is,

$$\theta_t = \int_0^t \alpha_s ds \text{ and } \int_0^1 |\alpha_t| dt < \infty, \ P^{0,v}\text{-a.s.},$$

for some adapted α for each $v \in f(\mathbb{R})$.

(2) No doubling strategies are allowed; that is, for all $v \in f(\mathbb{R})$,

$$E^{0,v} \int_0^1 (H^2(s, Y_s) + \alpha_s^2) ds < \infty. \tag{2.6}$$

The set of admissible trading strategies for the given pricing rule H is denoted by $\mathcal{A}(H)$.

The square integrability condition on H is standard in the literature on the Kyle-Back model. The new condition on α is almost redundant since the expected penalties become infinite when it is not satisfied. However, it could be possible for the insider to manipulate the prices so that the gains from trading $\int_0^1 (V - S_t)\alpha_t dt$ is greater than the penalties. The extra condition prevents such doubling strategies.

On the other hand, the total wealth at the end of trading is given by

$$\int_0^1 (V - H(t, Y_t))\alpha_t dt - \frac{c}{2} \int_0^1 \alpha_t^2 dt \le \int_0^1 \frac{(V - H(t, Y_t))^2}{2} dt + \frac{1-c}{2} \int_0^1 \alpha_t^2 dt.$$

Thus, if the penalties are sufficiently high; that is, $c > 1$, the insider will never use strategies that are not square integrable if the prices are square integrable. In such cases, the condition on α is indeed redundant.

The market equilibrium can now be defined in the following

Definition 2.3. A couple (H^*, θ^*) is said to form an equilibrium if H^* is an admissible pricing rule, $\theta^* \in \mathcal{A}(H^*)$, and the following conditions are satisfied:

(1) *Market efficiency condition:* Given θ^*, H^* is a rational pricing rule; that is, it satisfies (2.5).
(2) *Insider optimality condition:* Given H^*, θ^* solves the insider optimization problem for all $v \in f(\mathbb{R})$:

$$E^{0,v}[W_1^{\theta^*}] = \sup_{\theta \in \mathcal{A}(H^*)} E^{0,v}[W_1^{\theta}] < \infty.$$

3. An algorithm to construct the equilibrium

Since the market makers' pricing rule leads to the martingale property of S and one typically expects the insider's strategy to be inconspicuous – that is, Y is a martingale in its own filtration – in view of the past literature, it is not unreasonable to restrict oneself to H that satisfies a heat equation:

$$H_t + \frac{\sigma^2}{2} H_{yy} = 0, \qquad t \in [0,1]. \tag{3.7}$$

Keeping the equation above in mind, a formal HJB equation for the value function of the insider will be obtained in the first part of this section. To do so, first define

$$J(t,y) := \sup_{\alpha \in \mathcal{A}(H)} E^{0,v}\Big[\int_t^1 (v - H(u,Y_u))\alpha_u du - \frac{c}{2}\int_t^1 \alpha_t^2 dt \Big| Y_t = y\Big].$$

Direct calculations lead to

$$J_t + \frac{\sigma^2}{2} J_{yy} + \sup_\alpha \Big\{\alpha(J_y + v - H) - \frac{c\alpha^2}{2}\Big\} = 0.$$

Note that the presence of the penalty factor $c > 0$ yields that the optimal α can be obtained in the feedback form:

$$\alpha^* = \frac{J_y(t,y) + v - H(t,y)}{c}. \tag{3.8}$$

This leads to the following HJB equation:

$$J_t + \frac{\sigma^2}{2} J_{yy} + \frac{(J_y + v - H)^2}{2c} = 0. \tag{3.9}$$

Recall that in the continuous-time formulation of the Kyle model given in Back (1992), $c = 0$; thus, the optimal strategy cannot be obtained in the feedback form from the HJB equation. Instead, one has $J_y = H - v$, which is what one should expect when allowing for $c \to 0$ in (3.8). (See Section 6.2 in Çetin and Danilova (2018).)

Next, suppose there exists a smooth function J^0 such that

$$\begin{aligned} J^0_t + \frac{\sigma^2}{2} J^0_{yy} &= 0, \\ J^0_y &= H - v, \text{ and} \\ J^0(1,y) &= \int_{h^{-1}(v)}^{y} (h(x) - v)dx + \text{ constant}, \end{aligned} \tag{3.10}$$

for some h to be determined and a constant possibly depending on v. Such a function exists and is the value function of the insider in the Kyle model with $c = 0$ when the pricing rule satisfies (3.7). (See Theorem 5.1 in Çetin and Danilova (2021).) In this case, $h = f(\cdot/\sigma)$, and the constant is zero.

Thus, if one defines $u = J - J^0$ and conjectures that $J(1,\cdot) \equiv 0$, one obtains

$$u_t + \frac{1}{2}\sigma^2 u_{yy} + \frac{1}{2c} u_y^2 = 0, \quad u(1,y) = -j^0(y,v) := -J^0(1,y). \tag{3.11}$$

Note that J^0 would be the value function of the insider had there not been any penalties on her trading strategies; that is, $c = 0$. Thus, $-u$ can be viewed as the expected penalties to the insider in equilibrium.

It is now widely understood that (3.11) has the following BSDE formulation:

$$dU_t = \sigma Z_t dB_t - \frac{1}{2c} Z_t^2 dt, \quad U_1 = u(1, \sigma B_1). \tag{3.12}$$

The quadratic BSDE above is known to have an explicit solution, which, in turn, yields

$$u(t,x) = c\sigma^2 \log \rho(t,x,v), \quad \rho(t,x,v) := E^{0,v}\Big[\exp\Big(-\frac{j^0(\sigma B_1, v)}{c\sigma^2}\Big)\Big|\sigma B_t = x\Big]. \tag{3.13}$$

It immediately follows from Theorem 4.3.6 in Karatzas and Shreve (1991) that for each $v \in f(\mathbb{R})$, $\rho(\cdot,\cdot,v) \in C^{1,2}([0,1) \times \mathbb{R}])$.

In view of (3.8) and (3.10), the above implies that the optimal control of the insider at time $t \in [0,1]$ is $\alpha^*(t, Y_t, V)$, where

$$\alpha^*(t,y,v) := \sigma^2 \frac{\rho_y(t,y,v)}{\rho(t,y,v)}. \tag{3.14}$$

Recall that the considerations above assume that the pricing rules satisfies (3.7). Combined with the requirement that prices follow martingales in market makers' filtration in equilibrium, this yields that Y must be a martingale in its own filtration. (See Lemma 2 in Cho (2003).) Observe that if $\kappa(t)\rho(t,Y_t,v)\Pi(dv) = \mathbb{P}(V \in dv | \mathcal{F}^Y_t)$, for some constant $\kappa(t)$ for each t, then Y is a martingale in its own filtration (see, e.g., Corollary 3.1 in Çetin and Danilova (2018)) since

$$\mathbb{E}\Big[\frac{\rho_y(t,Y_t,V)}{\rho(t,Y_t,V)}\Big|\mathcal{F}^Y_t\Big] = \int_{f(\mathbb{R})} \rho_y(t,Y_t,v)\kappa(t)dv = \frac{d}{dy}\int_{f(\mathbb{R})} \rho(t,y,v)\kappa(t)dv\Big|_{y=Y_t} = 0,$$

provided one can interchange the order of integration and differentiation.

To this end, suppose that $\rho^0(t,y,v) := \kappa(t)\rho(t,y,v)$ is the conditional density of V with respect to the measure Π given $\mathcal{F}^Y_t$ and $Y_t = y$. Then, in view of Theorem 3.3 in Çetin and Danilova (2018), one expects $\rho^0(\cdot,\cdot,v)$ to satisfy for each v,

$$\rho^0_t + \frac{1}{2}\sigma^2\rho^0_{yy} = 0. \tag{3.15}$$

On the other hand, ρ satisfies (3.15) by its construction. Therefore, $\frac{\kappa'}{\kappa} \equiv 0$, and, consequently, κ must be constant since it must be strictly positive. This leads to the following:

Proposition 3.1. *Suppose that there exists a continuous function $j^0 : \mathbb{R}\times f(\mathbb{R}) \to \mathbb{R}$ such that $\kappa\rho(0,0,\cdot) \equiv 1$ for some normalizing constant κ, where ρ is defined via (3.13). Assume further that there exists a unique strong solution on $(\Omega, \mathcal{G}, (\mathcal{G}_t)_{t\in[0,1]}, \mathbb{Q})$ to*

$$Y_t = \sigma B_t + \int_0^t \sigma^2 \frac{\rho_y(s,Y_s,V)}{\rho(s,Y_s,V)} ds$$

such that

$$\mathbb{E}^{\mathbb{Q}}\Big[\int_0^t \Big(\frac{\rho_y(s,Y_s,V)}{\rho(s,Y_s,V)}\Big)^2\Big] < \infty, \quad t\in[0,1]. \tag{3.16}$$

Then

$$\kappa\rho(t,Y_t,v)\Pi(dv) = \mathbb{P}(V\in dv|\mathcal{F}^Y_t), \qquad t\in[0,1].$$

Proof. This is an immediate consequence of Theorem 3.3 in Çetin and Danilova (2018). Note that the process Z therein coincides with V at all times under the setting of the present paper. Thus, the martingale problem $\tilde{A}$, that is, the martingale problem for the pair (Y,V), is trivially well-posed. □

The seemingly strong integrability condition (3.16) is, in fact, a condition on the square integrability of Z in the BSDE representation (3.12). A minimal integrability condition, as in the next result, ensures that it holds.

Proposition 3.2. *Consider ρ defined by (3.13) for some continuous j^0. Assume that*

$$\int_{f(\mathbb{R})} E^{0,v}[|j^0(\sigma B_1, v)|]\Pi(dv) < \infty. \tag{3.17}$$

Then the condition (3.16) holds.

Proof. To ease notation, take $\sigma = 1$. As observed above, it suffices to show that

$$\mathbb{E}^{\mathbb{Q}}\Big[\int_0^1 Z_t^2 dt\Big] < \infty,$$

where (U,Z) is as in (3.12). It follows from the Jensen's inequality and (3.13) that

$$U_t \geq M_t(v) := -E^{0,v}\big[j^0(B_1,v)\big|B_t\big],\ P^{0,v}\text{-a.s.}. \tag{3.18}$$

Since $M(v)$ is a martingale on the time interval $[0,1]$, it follows that U is bounded from below by a uniformly integrable process. Thus, by Fatou's lemma $\int_0^\cdot Z_s dB_s$ in (3.12) is a supermartingale. Therefore,

$$\frac{1}{2c}E^{0,v}\Big[\int_0^1 Z_t^2 dt\Big] \leq E^{0,v}[U_0 - U_1] \leq U_0 + E^{0,v}[j^0(B_1,v)] < \infty.$$

This yields the claim by the independence of B and V under $\mathbb{Q}$, and the hypothesis (3.17). □

The considerations above give the following algorithm for constructing an equilibrium:

(1) Find a continuous function $j^0 : \mathbb{R}\times f(\mathbb{R}) \to \mathbb{R}$ such that i) $\kappa\rho(1,y,v)\Pi(dv)$ is a probability measure for each y, where ρ is given by (3.13) and $\kappa > 0$ is a constant, ii) $\kappa\rho(0,0,\cdot) \equiv 1$, and iii) it is differentiable in its first parameter with $j^0_y(y,v) = h(y) - v$. Note that the second condition entails

$$\kappa \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{y^2}{2\sigma^2}\Big) \exp\Big(-\frac{j^0(y,v)}{\sigma^2 c}\Big) dy = 1, \qquad \forall v \in f(\mathbb{R}). \tag{3.19}$$

(2) Set

$$H_t + \frac{1}{2}\sigma^2 H_{yy} = 0, \qquad H(1,y) = h(y).$$

(3) Show that (H, θ^*) with $d\theta^*_t = \frac{\rho_y(t,Y_t,V)}{\rho(t,Y_t,V)}dt$ is equilibrium, provided they are admissible, by using the candidate value function $J = J^0 - u$ that satisfies (3.9).

In view of the rationality of the pricing rule, that is, (2.1), the steps above appear to assume in equilibrium that the function h also satisfies

$$h(y) = \kappa \int_{f(\mathbb{R})} v\rho(1,y,v)\Pi(dv) = \kappa \int_{f(\mathbb{R})} v \exp\Big(-\frac{j^0(y,v)}{c\sigma^2}\Big)\Pi(dv).$$

However, this already follows from the properties of j^0 in Step (1) alone. Indeed, since $\kappa\rho(1,y,\cdot)$ is a proper density for each y,

$$0 = \frac{d}{dy}1 = \frac{d}{dy}\kappa \int_{f(\mathbb{R})} \exp\Big(-\frac{j^0(y,v)}{\sigma^2 c}\Big)\Pi(dv) = \frac{\kappa}{\sigma^2 c}\int_{f(\mathbb{R})} (v - h(y))\exp\Big(-\frac{j^0(y,v)}{\sigma^2 c}\Big)\Pi(dv),$$

provided one can differentiate under the integral sign. Thus,

$$h(y) = h(y)\kappa \int_{f(\mathbb{R})} \exp\Big(-\frac{j^0(y,v)}{c\sigma^2}\Big)\Pi(dv) = \kappa \int_{f(\mathbb{R})} v\exp\Big(-\frac{j^0(y,v)}{c\sigma^2}\Big)\Pi(dv),$$

as required.

4. A fixed-point iteration

The algorithm from the previous section shows that the equilibrium mainly becomes a matter of finding a function j^0 with certain properties. The condition on its derivative implies that

$$j^0(y,v) = \Psi(v) + \phi(y) - yv, \tag{4.20}$$

where Ψ and ϕ are continuous functions on $f(\mathbb{R})$ and $\mathbb{R}$, respectively. This section will describe an operator on the space of continuous functions, whose fixed point will determine the functions Ψ and ϕ that will appear in equilibrium. The following notation will be used to ease the exposition.

Notation. $\hat{c} := c\sigma^2$.

Since $\exp(-\frac{1}{\hat{c}}j^0(y,\cdot))\kappa$ will be a probability density function itself for all y for some constant κ independent of y, one can without loss of generality take $\kappa = 1$ by incorporating it into Ψ or ϕ. This yields the first relationship,

$$\int_{f(\mathbb{R})} \exp\Big(\frac{yv - \Psi(v)}{\hat{c}}\Big)\Pi(dv) = \exp\Big(\frac{\phi(y)}{\hat{c}}\Big), \qquad y \in \mathbb{R}. \tag{4.21}$$

The second relationship follows from the initial condition (3.19),

$$\int_{\mathbb{R}} \exp\Big(\frac{yv - \phi(y)}{\hat{c}}\Big)\frac{1}{\sqrt{2\pi\sigma^2}}\exp\Big(-\frac{y^2}{2\sigma^2}\Big)dy = \exp\Big(\frac{\Psi(v)}{\hat{c}}\Big), \qquad v \in f(\mathbb{R}). \tag{4.22}$$

Thus, one obtains the following two operators on the space of measurable functions:

$$\begin{aligned} T_1(\phi)(v) &:= \int_{\mathbb{R}} \exp\Big(\frac{yv - \phi(y)}{\hat{c}}\Big)\frac{1}{\sqrt{2\pi\sigma^2}}\exp\Big(-\frac{y^2}{2\sigma^2}\Big)dy = \exp\Big(\frac{\Psi(v)}{\hat{c}}\Big), \qquad v \in f(\mathbb{R}); \\ T_2(\Psi)(y) &:= \int_{f(\mathbb{R})} \exp\Big(\frac{yv - \Psi(v)}{\hat{c}}\Big)\Pi(dv) = \exp\Big(\frac{\phi(y)}{\hat{c}}\Big), \qquad y \in \mathbb{R}. \end{aligned} \tag{4.23}$$

Consequently, one can define an operator T on the space of measurable functions by

$$T\phi = \hat{c}\log T_2(\hat{c}\log T_1(\phi)), \tag{4.24}$$

whose domain consists of functions so that $T(\phi)$ is finite for all $y \in \mathbb{R}$. Clearly, finding a fixed point of T will allow to construct functions ϕ and Ψ so that j^0 defined by (4.20) satisfies the conditions of the algorithm for equilibrium.

The system of equations in (4.21) and (4.22) are the so-called *Schrödinger equations* that have been well-studied in the context of Schrödinger bridges and entropic optimal transport.[13] The elements of the solution pair (ϕ, Ψ) are called Schrödinger potentials. Clearly, the existence of a pair (ϕ, Ψ) solving this system is equivalent to the existence of a fixed point for T, which yields an algorithm, known as *the Sinkhorn algorithm* in the optimal transport literature, amenable to numerical studies.

In the present setting, classical results can be used to establish the existence of a *unique* fixed point for T. The following theorem, whose proof is delegated to the appendix, collects some important properties of the solution that will be useful in the sequel. Recall our standing assumption that V has a finite second moment.

Theorem 4.1. *There exists a unique fixed point ϕ of T with $\phi(0) = 0$. Define $\Psi(v) = \hat{c} \log T_1\phi(v)$, for any $v \in I := \{(\ell, r) : \ell = \inf V, r = \sup V\}$. Then, the following hold, where we denote by $p(\sigma, \cdot)$ the density of a centered normal distribution with variance σ^2.*

(1) $\mathbb{E}[|\phi(\sigma B_1)|] + \mathbb{E}[|\Psi(V)|] < \infty$.
(2) *Ψ is bounded from below by a constant and ϕ is bounded from below by an affine function on $\mathbb{R}$.*
(3) *Ψ is infinitely differentiable and strictly convex. In particular,*

$$\Psi'(v) = \frac{\int_{\mathbb{R}} y \exp\left(\frac{yv-\phi(y)}{\hat{c}}\right) p(\sigma, y) dy}{\int_{\mathbb{R}} \exp\left(\frac{yv-\phi(y)}{\hat{c}}\right) p(\sigma, y) dy}, \text{ and} \tag{4.25}$$

$$\Psi''(v) = \frac{1}{\hat{c}} \frac{\int_{\mathbb{R}} y^2 \exp\left(\frac{yv-\phi(y)}{\hat{c}}\right) p(\sigma, y) dy}{\int_{\mathbb{R}} \exp\left(\frac{yv-\phi(y)}{\hat{c}}\right) p(\sigma, y) dy} - \frac{1}{\hat{c}} \left(\frac{\int_{\mathbb{R}} y \exp\left(\frac{yv-\phi(y)}{\hat{c}}\right) p(\sigma, y) dy}{\int_{\mathbb{R}} \exp\left(\frac{yv-\phi(y)}{\hat{c}}\right) p(\sigma, y) dy}\right)^2. \tag{4.26}$$

(4) *Suppose further that V possesses all exponential moments. Then ϕ is infinitely differentiable and strictly convex. In particular,*

$$\phi'(y) = \frac{\int_{f(\mathbb{R})} v \exp\left(\frac{yv-\Psi(v)}{\hat{c}}\right) \Pi(dv)}{\int_{f(\mathbb{R})} \exp\left(\frac{yv-\Psi(v)}{\hat{c}}\right) \Pi(dv)}, \text{ and} \tag{4.27}$$

$$\phi''(y) = \frac{1}{\hat{c}} \frac{\int_{f(\mathbb{R})} v^2 \exp\left(\frac{yv-\Psi(v)}{\hat{c}}\right) \Pi(dv)}{\int_{f(\mathbb{R})} \exp\left(\frac{yv-\Psi(v)}{\hat{c}}\right) \Pi(dv)} - \frac{1}{\hat{c}} \left(\frac{\int_{f(\mathbb{R})} v \exp\left(\frac{yv-\Psi(v)}{\hat{c}}\right) \Pi(dv)}{\int_{f(\mathbb{R})} \exp\left(\frac{yv-\Psi(v)}{\hat{c}}\right) \Pi(dv)}\right)^2. \tag{4.28}$$

The expressions for the derivatives of ϕ and Ψ establish useful identities for some of the key equilibrium parameters. Observe that $\exp\left(\frac{yv-\Psi(v)}{\hat{c}}\right)\Pi(dv)$ is the time-1 unnormalized conditional density of V in equilibrium for the market makers. Thus, expression (4.27) implies that ϕ' will be the pricing rule at time-1, as anticipated by the equilibrium considerations in the previous section. Consequently, ϕ'' will determine *Kyle's lambda*, and, therefore, the market liquidity at time-1.

Interestingly, $\hat{c}\phi''(y)$ coincides with the variance of V after all trading stops at time-1. This will be used as a measurement of price inefficiency in Section 5.1 and will have an important role in the determination of the optimal penalty policy by the regulators, as discussed in Section 6.

Similarly, $\exp\left(\frac{yv-\phi(y)}{\hat{c}}\right)p(\sigma, y)$ is the unnormalized conditional density of the total demand at time-1 for an insider who observes $V = v$. The above formulae show that, conditional on $V = v$, the net demand at time-1 for the insider has mean $\Psi'(v)$ and variance $\hat{c}\Psi''(v)$, which motivates the expectation that as the rate of penalties vanishes, the net demand at time-1 becomes deterministic for the insider. This is consistent with the bridge strategy of the Kyle model without penalties, where the net demand at time-1 equals $\sigma\eta$,[14] which is constant for the insider.

Example 4.1. Consider the case of a Bernoulli V, where $\Pi(\{1\}) = 1 - \Pi(\{0\}) = p$. Existence of a fixed point for T is ensured by Theorem 4.1. As the support of V contains only two points, an easier representation of ϕ can be found as follows.

In view of the normalization that $\phi(0) = 0$, one can conjecture that

$$\exp\left(\frac{\phi}{\hat{c}}\right) = \exp\left(\frac{y}{\hat{c}}\right)ap + 1 - ap, \text{ and } \Psi(1) = -\hat{c}\log a, \Psi(0) = -\hat{c}\log\frac{1-ap}{1-p}.$$

Thus, it remains to pinpoint the value of a, which is identified by the following nonlinear equation:

[13] See the recent lecture notes by M. Nutz: Introduction to Entropic Optimal Transport, available at https://www.math.columbia.edu/~mnutz/docs/EOT_lecture_notes.pdf.
[14] Recall from Section 2 that $V = f(\eta)$ for some standard Normal η and non-decreasing f.

$$\frac{1}{a} = \int_{\mathbb{R}} \frac{e^{\frac{y}{\hat{c}}}}{ap(e^{\frac{y}{\hat{c}}}-1)+1} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{y^2}{2\sigma^2}\Big)dy.$$

However, this is equivalent to

$$1 = \int_{\mathbb{R}} \frac{e^{\frac{y}{\hat{c}}}}{p(e^{\frac{y}{\hat{c}}}-1)+a^{-1}} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{y^2}{2\sigma^2}\Big)dy.$$

Note that $a \in (0, p^{-1})$. Thus, the right-hand side is increasing from 0 to $1/p$ on $(0, p^{-1})$, and there exists a unique solution belonging to $(0, p^{-1})$.

Example 4.2. Suppose that V has a Gaussian distribution; that is, $V \sim N(\mu, \gamma^2)$ for some $\mu \in \mathbb{R}$ and $\gamma \in (0,\infty)$. In the Kyle model with Gaussian fundamental value (see Back (1992)), the equilibrium price is an affine function of the demand Y. Thus, one expects $H^*(t,y) = \lambda y + \mu$, for some $\lambda > 0$. The conjecture that $H^*(t,0) = \mu$ is a simple consequence of the rationality of the pricing rule that requires $\mu = \mathbb{E}[V] = S_0 = H^*(0,0)$. Combined with the normalization that $\phi(0) = 0$, one expects that $\phi(y) = \frac{\lambda y^2}{2} + \mu y$, for some $\lambda \in \mathbb{R}$.

By completing the squares, it is easily seen that

$$T_1\phi(y) = \frac{\Sigma}{\sigma} \exp\Big(\frac{\Sigma^2(\mu - v)^2}{2\hat{c}^2}\Big),$$

where $\frac{1}{\Sigma^2} = \frac{1}{\sigma^2} + \frac{\lambda}{\hat{c}} = \frac{c+\lambda}{\hat{c}}$. That is,

$$T_1\phi(y) = \frac{\Sigma}{\sigma} \exp\Big(\frac{(\mu - v)^2}{2\hat{c}(c+\lambda)}\Big). \tag{4.29}$$

Thus, to have $T\phi = \phi$, one must have

$$\begin{aligned}
\exp\Big(\frac{\lambda y^2 + 2\mu y}{2\hat{c}}\Big) &= \frac{\sigma}{\Sigma\gamma} \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \exp\Big(\frac{yv}{\hat{c}} - \frac{(\mu-v)^2}{2\hat{c}(c+\lambda)} - \frac{(\mu-v)^2}{2\gamma^2}\Big)dv \\
&= \frac{\sigma\Sigma_0}{\Sigma\gamma} \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\Sigma_0^2}} \exp\Big(\frac{yv}{\hat{c}} - \frac{(\mu-v)^2}{2\Sigma_0^2)}\Big)dv \\
&= \frac{\sigma\Sigma_0}{\Sigma\gamma} \exp\Big(\frac{y\mu}{\hat{c}} + \frac{y^2\Sigma_0^2}{2\hat{c}^2}\Big),
\end{aligned}$$

where $\frac{1}{\Sigma_0^2} = \frac{1}{\hat{c}(c+\lambda)} + \frac{1}{\gamma^2} = \frac{\gamma^2 + \hat{c}(c+\lambda)}{\gamma^2\hat{c}(c+\lambda)}$.

Thus, λ must solve

$$\lambda = \frac{\Sigma_0^2}{\hat{c}} = \frac{\gamma^2(c+\lambda)}{\gamma^2 + \hat{c}(c+\lambda)}. \tag{4.30}$$

Note that one must also have

$$1 = \frac{\sigma^2\Sigma_0^2}{\Sigma^2\gamma^2} = \frac{\sigma^2(c+\lambda)}{\gamma^2}\lambda. \tag{4.31}$$

That is, λ must seemingly satisfy two constraints at the same time. However, (4.30) is equivalent to

$$\lambda c\sigma^2(c+\lambda) = c\gamma^2.$$

Dividing both sides by c yields (4.31). Thus, λ is the unique positive solution of

$$\lambda^2\sigma^2 + \lambda c\sigma^2 - \gamma^2 = 0,$$

which is given by $\lambda = \frac{-c+\sqrt{c^2+4\frac{\gamma^2}{\sigma^2}}}{2}$.

Consequently, in view of (4.29), one obtains

$$\Psi(v) = \hat{c}\log T_1\phi(v) = \frac{\hat{c}}{2}\log\frac{\Sigma^2}{\sigma^2} + \frac{(\mu-v)^2}{2(c+\lambda^*)} = \frac{\hat{c}}{2}\log\frac{c}{c+\lambda^*} + \frac{(\mu-v)^2}{2(c+\lambda^*)}.$$

5. Equilibrium

In the first part of this section, existence of equilibrium for the economy described in Section 2 will be established under the assumption that follows. Note that the mere existence of (ϕ^*, Ψ^*) follows from the standing assumption that V has a finite variance.

The extra smoothness require more from V but is achieved as soon as it possesses exponential moments, e.g. bounded or Gaussian V, in view of Theorem 4.1.

Assumption 5.1. There exists a pair of continuous functions (ϕ^*, Ψ^*) such that ϕ^* is a fixed point of the operator T in (4.24) with $\Phi^*(0) = 0$, and $\Psi^* = c\sigma^2 \log T_1\phi^*$, where T_1 is as in (4.23). The function ϕ^* is twice continuously differentiable, satisfying (4.27) and (4.28).

Recall from Theorem 4.1 that ϕ^* is strictly convex, which will be used frequently in proofs. Moreover, (3.17) is satisfied when j^0 is replaced by

$$j^*(y,v) := \Psi^*(v) + \phi^*(y) - yv. \tag{5.32}$$

In view of the procedure described at the end of Section 3, the candidate pricing rule for equilibrium is the following:

$$H^*(t,y) = \int_{f(\mathbb{R})} \rho^*(t,y,z) z \Pi(dz), \tag{5.33}$$

where

$$\rho^*(t,y,v) := E^{0,v}\left[\exp\left(-\frac{j^*(\sigma B_1, v)}{\hat{c}}\right)\Big|\sigma B_t = y\right]. \tag{5.34}$$

The expectation above is non-zero and finite due to Lemma A.1. Thus, expression (4.27) particularly implies that

$$H^*_t + \frac{\sigma^2}{2} H^*_{yy} = 0, \qquad H^*(1,\cdot) = h^* = \frac{d\phi^*}{dy}. \tag{5.35}$$

Additionally, the discussion in Section 3 predicts that the value function of the insider is given by $J = J^0 + u$,

$$J^0_t + \frac{\sigma^2}{2} J^0_{yy} = 0, \quad J^0(1,y) = j^*(y,v), \tag{5.36}$$

and u is given by (3.13) with j^0 replaced by j^*. These considerations yield the following in view of the PDE (3.11) that u satisfies.

Proposition 5.1. *Under Assumption 5.1, there exists a classical solution to*

$$J_t + \frac{\sigma^2}{2} J_{yy} + \frac{(J_y + v - H^*)^2}{2c} = 0, \qquad J(1,\cdot) = 0, \tag{5.37}$$

where H^* is given by (5.33). In particular, $J = J^0 + \hat{c}\log\rho^*$, where J^0 is given by (5.36).

The key difference between (5.37) and the Bellman equation (18) in Back (1992), which describes the insider's profits in the case of no penalties, is the term $\frac{(J_y+v-H^*)^2}{2c}$. When c approaches 0 in the equation above, to keep the insider's gains finite in any equilibrium, one should have $J_y = H^* - v$ in the limit. This is precisely the other requirement (17) that completes the characterization of the Bellman equation in Back (1992). Thus, two equations align when $c = 0$.[15]

As expected from the formal calculations of Section 3, the function J above can be used to show that an equilibrium is given by the pricing rule H^* and the strategy θ^* with

$$d\theta^*_t = \sigma^2 \frac{\rho^*_y(t, Y_t, V)}{\rho^*(t, Y_t, V)} dt, t \in [0,1], \text{ and } \theta^*_0 = 0. \tag{5.38}$$

The admissibility condition that is required from any equilibrium candidate is the square integrability of the strategy as in (2.6). Proposition A.3 in the Appendix show that θ^* is admissible and, for each v, there exists a unique strong solution to

$$Y^*_t = \sigma B_t + \sigma^2 \int_0^t \frac{\rho^*_y(s, Y^*_s, v)}{\rho^*(s, Y_s, v)} ds, \tag{5.39}$$

where ρ^* is given by (5.34). This leads to the following description of the equilibrium.

Theorem 5.1. *Suppose that Assumption 5.1 holds. Then (H^*, θ^*) is an equilibrium where H^* is given by (5.33) and θ^* follows (5.38). The total demand in equilibrium follows (5.39). Additionally, the following properties of equilibrium hold:*

[15] Indeed, when there are no penalties, the decomposition of J coincides with the one given in Theorem 3.2 in Back et al. (2020).

(1) *The conditional law of* V *is given by* $\mathbb{P}(V \in dv|\mathcal{F}_t^{Y^*}) = \rho^*(t, Y_t^*, v)\Pi(dv)$*, where* ρ^* *is as in* (5.34)*. In particular,* $\frac{Y^*}{\sigma}$ *is a standard Brownian motion in its own filtration.*

(2) *The insider's expected profit is given by*

$$E^{0,v}[W_1^{\theta^*}] = J(0,0) = \Psi^*(v) + E^{0,v}[\phi^*(\sigma B_1)]. \tag{5.40}$$

(3) *Suppose further that one can differentiate inside the expectation*[16] *so that*

$$H_y(t, Y_t^*) = \mathbb{E}\Big[\frac{d^2\phi^*}{dy^2}(Y_1^*)\Big|\mathcal{F}_t^{Y^*}\Big].$$

Then the expected loss of noise traders is given by

$$\sigma^2 \mathbb{E}\Big[\frac{d^2\phi^*}{dy^2}(Y_1^*)\Big]. \tag{5.41}$$

(4) *The insider expects to pay the following amount of penalty in equilibrium:*

$$\frac{c}{2}E^{0,v}\Big[\int_0^1 (\alpha_t^*)^2 dt\Big] = v(\Psi^*)'(v) - \Psi^*(v) - E^{0,v}[\phi^*(Y_1^*)]. \tag{5.42}$$

The theorem above shows that the key equilibrium parameters, such as insider's gain, noise traders' loss, *Kyle's lambda*,[17] and expected penalties are all given explicitly in terms of the functions ϕ^*, Ψ^*, and their derivatives. Note that although the fixed point of T, ϕ^*, is in general not known in closed form, it can be obtained numerically via straightforward policy iterations. Then Ψ^* and the required derivatives follow from numerical integration or Monte-Carlo simulations.

In the continuous time Kyle model (and in other extensions with risk-neutral agents), the insider's strategy in equilibrium is independent of the distribution of V and is given by[18]

$$\alpha_t^K := \frac{\sigma\eta - Y_t^*}{1-t}dt, \tag{5.43}$$

where α^K is used to denote the trading rate in the Kyle model without penalties. When the fundamental value is normally distributed, the expression above can be equivalently rewritten as

$$\alpha_t^K = \frac{\sigma}{\gamma}\frac{V - S_t^K}{1-t}dt,$$

where S^K denotes the equilibrium price. That is, the insider buys if the market price is lower than her own valuation and sell otherwise. Note that such an equivalence only holds if V is Gaussian since the optimal strategy is still given by (5.43) for general V. (See Back (1992).)

Since the conditional distribution of η in the present model is no longer Gaussian in general, the insider's strategy typically will be nonlinear in the total demand. However, as the next proposition demonstrates, the insider buys when her expectation of the terminal market price of the asset is lower than V and sells otherwise.

Proposition 5.2. *Consider the equilibrium of Theorem 5.1, and assume that* h^* *is at most of exponential growth.*[19] *Then*

$$\alpha_t^* := \frac{d\theta_t^*}{dt} = \frac{1}{c}(v - E^{0,v}[h^*(Y_1^*)|\mathcal{F}_t^I]), \tag{5.44}$$

with $E^{0,v}[h(Y_1^*)|\mathcal{F}_t^I] = \mathcal{P}(t, Y_t^*; v)$*, where*

$$\mathcal{P}(t, y; v) = \frac{\int_{\mathbb{R}} h^*(x)\exp(\frac{vx - \phi^*(x)}{\hat{c}})p(\sigma\sqrt{1-t}, x-y))dx}{\int_{\mathbb{R}} \exp(\frac{vx-\phi^*(x)}{\hat{c}})p(\sigma\sqrt{1-t}, x-y))dx}.$$

The proposition above shows that the insider's trading rate is inversely proportional to the rate of penalties, and she trades more when she believes that the market price will be further away from the fundamental value right before the public announcement of V. Interestingly, the number of shares that are traded is the same constant multiple, that is, c^{-1}, of the difference between V and

[16] Recall from (5.35) that $H(t, Y_t^*) = \mathbb{E}[\frac{d\phi^*}{dy}(Y_1^*)|\mathcal{F}_t^{Y^*}]$. A simple sufficient condition for differentiability under the expectation in item (3) above is that the second derivative of ϕ^* has at most exponential growth. This will be the case if V has a bounded support. Indeed, by Theorem 4.1, a fixed point of T exists and has bounded first and second derivatives.

[17] That is, the sensitivity of prices to changes in demand, which coincides with the derivative of H.

[18] Recall once more that $V = f(\eta)$, for some non-decreasing f and standard Normal η.

[19] That is, $|h(y)| \le C_1 \exp(C_2|y|)$, for all $y \in \mathbb{R}$, for some constant C_is.

$\mathcal{P}(t, Y_t^*; V)$ at all times. When the additional transaction costs are interpreted as potential legal fines, this constant is proportional to the success probability of investigations and inversely proportional to the rate of pecuniary penalties.[20]

Recall that although, to a certain extent, the insider controls the evolution of the total demand by trading, the terminal value of the demand in equilibrium is no longer given by a deterministic function of her private signal V in the presence of penalties. Thus, the market price at time-1 is random even for the insider, and, consequently, the insider makes significant trades only if she believes that the terminal market price will end up to be far away from the fundamental value as a result of large liquidity shocks. It is also important to observe that her expectation of terminal price, that is, $E^{0,v}[h(Y_1^*)|\mathcal{F}_t^I] = \mathcal{P}(t, Y_t^*; v)$, is different from the market price $H(t, Y_t^*)$, since the latter is not a martingale for the insider, and the former depends on the realization of V.

An interesting connection between the expected loss of noise traders and the price efficiency of equilibrium is given in the next corollary. Recall that in Back (1992), the equilibrium price is efficient in the sense that all the private information of the informed trader is fully disseminated to the market by the end of the trading period. That is, the conditional variance of V, given the market's information at time-1 is zero. This is no longer the case in the present model since the market price does not converge to V as $t \to 1$. Indeed, following a strategy to ensure that the final price coincides with the fundamental value is too costly for the insider under quadratic transaction costs on the traded volume. The following defines a measure of *price inefficiency* of equilibrium as in Carré et al. (2022), and gives its magnitude.

Corollary 5.1. *Consider the equilibrium of Theorem 5.1. Then the* price inefficiency *of the equilibrium, denoted by δ, is given by*

$$\delta := \mathbb{E}[Var(V|\mathcal{F}_1^{Y^*})] = \hat{c}\mathbb{E}\Big[\frac{d^2\phi^*}{dy^2}(Y_1^*)\Big]. \tag{5.45}$$

Similarly,

$$Var(V|\mathcal{F}_t^{Y^*}) = \mathbb{E}\Big[\hat{c}\frac{d^2\phi^*}{dy^2}(Y_1^*) + (h^*(Y_1^*))^2|\mathcal{F}_t^{Y^*}\Big] - (H^*(t, Y_t^*))^2. \tag{5.46}$$

Proof. It follows from (4.28) that $\frac{d^2\phi^*}{dy^2}(y) = \hat{c}^{-1}\mathrm{Var}(V|\mathcal{F}_1^{Y^*}, Y_1^* = y)$, which yields the first claim.

To show the second claim, it suffices to show that $\mathbb{E}[V^2|\mathcal{F}_t^{Y^*}] = \mathbb{E}[\hat{c}\frac{d^2\phi^*}{dy^2}(Y_1^*) + (h^*(Y_1^*))^2|\mathcal{F}_t^{Y^*}]$. However, this follows from the fact that $\hat{c}\frac{d^2\phi^*}{dy^2}(Y_1^*) = \mathrm{Var}(V|\mathcal{F}_1^{Y^*})$ and $\mathbb{E}[V|\mathcal{F}_1^{Y^*}] = h^*(Y_1^*)$. $\square$

Comparing the aforementioned to the loss of noise traders from Theorem 5.1 reveals the interesting observation that, independent of the distribution of the fundamental value of the asset, the expected loss of noise traders equals $\frac{\delta}{c}$, where δ is the price inefficiency. Since the noise traders lose in equilibrium in the Kyle model with $c = 0$, one can easily deduce from this relationship that the price inefficiency persists and does not become negligible quickly even if c decreases.

Remark 4. That the insider does not aggregate all the information by the end of trading period is robust to different specifications of the penalty function. Indeed, suppose that the cumulative transaction costs C is given by

$$C_t = \int_0^t c(\alpha_s)ds,$$

where $c(\alpha) \sim c|\alpha|^p$ for some $c > 0$ and $p > 1$ for large values of α. Then, the optimal strategy in Back (1992) will yield infinite cumulative transaction costs since

$$\int_0^1 \frac{|\sigma\eta - Y_t|^p}{(1-t)^p} dt = \infty$$

for $p > 1$. In particular, any strictly convex cost function will prohibit the fundamental value from being fully revealed by the end of the trading period.

On the other hand, if transaction costs are at most linear for large orders, it may be possible for the insider to aggregate all her private information. However, this requires a further analysis as the optimization problem of the insider is no longer strictly convex.

5.1. Gaussian case

Since a Gaussian random variable admits all exponential moments, Theorem 4.1 establishes the existence of a fixed point for T when Π is a Gaussian distribution. This subsection studies the resulting equilibrium under the following assumption.

[20] Recall that c reflects the probability of success for investigations as well as the level of legal fines.

Assumption 5.2. The distribution of V is Gaussian; that is, $V \sim N(\mu, \gamma^2)$.

The next theorem directly follows from Example 4.2.

Theorem 5.2. *Suppose that Assumption 5.2 holds. Then* $T_0\phi^* = \phi^*$ *with* $\phi^*(y) = \frac{\lambda^* y^2}{2} + \mu y$, *for*

$$\lambda^* = \frac{-c + \sqrt{c^2 + 4\frac{\gamma^2}{\sigma^2}}}{2}, \tag{5.47}$$

which is the positive root of $\sigma^2\lambda^2 + \hat{c}\lambda - \gamma^2 = 0$. *Consequently,*

$$\Psi^*(v) = \hat{c}\log T_1\phi^*(v) = \frac{\hat{c}}{2}\log\frac{c}{c+\lambda^*} + \frac{(\mu - v)^2}{2(c+\lambda^*)}. \tag{5.48}$$

The formula for λ^* suggests a parametrization for the penalty rate c that will be useful when considering the equilibrium with Gaussian fundamental value. The easy proof of the following corollary is left to the reader.

Corollary 5.2. *Let* $c = \kappa\frac{\gamma}{\sigma}$, *for some* $\kappa > 0$. *Then* $\lambda^* = \Lambda(\kappa)\frac{\gamma}{\sigma}$, *where* $\Lambda : (0,\infty) \to (0,1)$ *is a decreasing function given by*

$$\Lambda(\kappa) = \frac{\sqrt{\kappa^2+4} - \kappa}{2}. \tag{5.49}$$

In particular, $\Lambda(0+) = 1 = 1 - \Lambda(\infty)$, *and* Λ *also satisfies*

$$\Lambda^2(\kappa) + \Lambda(\kappa)\kappa = 1, \qquad \kappa \geq 0. \tag{5.50}$$

Recall that Kyle's lambda equals $\frac{\gamma}{\sigma}$ when $c = 0$. Therefore, an immediate consequence of the corollary above is that the equilibrium parameters converge to the levels in Back (1992) as the rate of penalties vanishes.

Note that the pair (ϕ^*, Ψ^*) from Theorem 5.2 satisfy all the conditions of Theorem 5.1. Thus, an equilibrium exists and has the following dynamics as a straightforward corollary. To understand the dependency of the equilibrium on the amount of adverse selection, given by $\frac{\gamma}{\sigma}$, the above parametrization will be employed.

Corollary 5.3. *Suppose that Assumption 5.2 holds and* $c = \kappa\frac{\gamma}{\sigma}$, *for some* $\kappa > 0$. *Then* (h^*, θ^*) *is an equilibrium, where* $h^*(y) = \lambda^* y + \mu$, *with* $\lambda^* = \Lambda(\kappa)\frac{\gamma}{\sigma}$, *and*

$$d\theta^*_t = \Lambda(\kappa)\frac{\sigma}{\gamma}\frac{V - \mu - \lambda^* Y^*_t}{1 - t\Lambda^2(\kappa)}dt, \tag{5.51}$$

where Λ *is the function defined in Corollary 5.2. The following statements are also valid:*

(1) *The conditional distribution of* V *given* $\mathcal{F}^{Y^*}_t$ *is Gaussian with mean* $\lambda^* Y^*_t + \mu$ *and variance* $\gamma^2(1 - t\Lambda^2(\kappa))$.
(2) *Equilibrium demand* Y^* *follows*

$$dY^*_t = \sigma dB_t + \Lambda(\kappa)\sigma\frac{\frac{V-\mu}{\gamma} - \frac{\Lambda(\kappa)Y^*_t}{\sigma}}{1 - t\Lambda^2(\kappa)}dt.$$

(3) *The insider's expected profit is given by*

$$E^{0,v}[W^{\theta^*}_1] = \frac{\hat{c}}{2}\log\frac{c}{c+\lambda^*} + \frac{(\mu - v)^2}{2(c+\lambda^*)} + \frac{\lambda^*\sigma^2}{2}. \tag{5.52}$$

(4) *The expected loss of noise traders is given by* $\lambda^*\sigma^2 = \Lambda(\kappa)\gamma\sigma$.
(5) *The price inefficiency is given by* $\delta = \hat{c}\lambda^* = \kappa\Lambda(\kappa)\gamma^2$, *and*

$$Var(V|\mathcal{F}^{Y^*}_t) = \kappa\Lambda(\kappa)\gamma^2 + \Lambda^2(\kappa)\gamma^2(1-t) = \gamma^2(1-t) + t\kappa\Lambda(\kappa)\gamma^2.$$

The statements above show that the insider buys when the market valuation is lower than the fundamental value and sells otherwise, as in the Kyle model without penalties.[21] Thus, from the point of view of the insider, prices still mean-revert around the

[21] This does not contradict the conclusion of Proposition 5.2. Straightforward calculations show that $V \geq H(t, Y^*_t)$ if and only if $V \geq \mathcal{P}(t, Y^*_t; V)$. A key difference is that while the insider trades the same constant multiple (independent of time) of the difference between V and $\mathcal{P}(t, Y^*_t; V)$, the ratio of the trading rate and the difference between V and the market price is not constant in time.

insider's valuation. However, the speed of mean reversion is much smaller. Indeed, simple computations imply that the price process follows

$$dS_t = \lambda^* \sigma dB_t + \Lambda^2(\kappa)\frac{V - S_t}{1 - t\Lambda^2(\kappa)}dt.$$

Thus, the mean reversion speed as measured by

$$r(t) := \frac{\Lambda^2(\kappa)}{1 - t\Lambda^2(\kappa)} \tag{5.53}$$

is considerably smaller than what it would have been in the model without penalties, where the corresponding speed equals $\frac{1}{1-t}$. Thus, the insider is less keen to keep the price close to the fundamental value when it is moved away by large liquidity shocks. This less aggressive trading results in prices not fully revealing the insider's private information at the end of the trading period, differently from the Kyle model without penalties.

The risk of large penalties also implies that the insider buys and sells at a constant rate. Direct calculations show that the distribution of $(\frac{V-\mu}{\gamma} - \frac{\Lambda(\kappa)Y_t^*}{\sigma})(1 - t\Lambda^2(\kappa))^{-1}$ conditional on V is Gaussian with mean $\frac{V-\mu}{\gamma}$ and variance $\frac{t\Lambda^2(\kappa)}{1-t\Lambda^2(\kappa)}$. This immediately leads to the following.

Corollary 5.4. *Let $\alpha_t^* = \frac{d\theta_t^*}{dt}$, where θ^* is the strategy of the insider in equilibrium from Corollary 5.3. Then*

$$\alpha_t^* \sim N\Big(\frac{V-\mu}{\gamma}\Lambda(\kappa)\sigma, \frac{t\sigma^2\Lambda^4(\kappa)}{1 - t\Lambda^2(\kappa)}\Big).$$

Note that Λ decreases to 0 as κ increases. (See also Fig. 2.) Thus, even after normalizing the insider's trades by $\Lambda(\kappa)$, the standard deviation of the trading rate remains small, that is, order of $\Lambda(\kappa)$, for large κ, which in turn implies that the deviations from the mean level of trades is negligible for large penalties. As a result, the insider buys (sells) at the constant rate $\frac{|V-\mu|}{\gamma}\Lambda(\kappa)\sigma$ if her private value is greater (smaller) than the initial price, μ, of the asset.

Recall that the Kyle model without penalties corresponds to the special case of $\Lambda(\kappa) = 1$. Thus, although the mean level of trading rate remains $\frac{|V-\mu|}{\gamma}\sigma$, its variance becomes infinite as the trader approaches to the liquidation date, which is consistent with the efforts of an insider bringing the market price to V at time-1.

As expected, the market liquidity is positively affected by the presence of penalties. Since Kyle's lambda equals $\frac{\gamma}{\sigma}$ when there is no additional penalty, one can define

$$L(\kappa) := \frac{1}{\Lambda(\kappa)} - 1 = \frac{1 - \Lambda(\kappa)}{\Lambda(\kappa)} \tag{5.54}$$

to represent the gain in liquidity when the regulators impose a quadratic penalty with rate given by $\kappa\frac{\gamma}{\sigma}$. This gain is clearly decreasing in Λ, which is a decreasing function of κ. The market thus becomes perfectly liquid in the limit, but it will also be completely inefficient.

Another interesting consequence of the reparametrization manifests itself in the representation of insider's ex-ante profit.

Corollary 5.5. *Consider the equilibrium in Corollary 5.3. Then the following statements are valid:*

(1) *The ex-ante profit of the insider equals*

$$\mathcal{W}(\kappa) := \mathbb{E}[W_1^{\theta^*}] = \frac{\gamma\sigma}{2}\big(\kappa\log(\kappa\Lambda(\kappa)) + 2\Lambda(\kappa)\big).$$

The function $\mathcal{W}$ is convex and decreasing in κ and vanishes in the limit as $\kappa \to \infty$.

(2) *The total welfare that measures the total expected profit of all traders is given by*

$$\frac{\gamma\sigma}{2}\kappa\log(\kappa\Lambda(\kappa)) = \frac{\gamma\sigma}{2}\kappa\log(1 - \Lambda^2(\kappa)). \tag{5.55}$$

As $\Lambda(K) \in (0, 1)$, the statements above show that the welfare is negative. This is not surprising since the market makers are competitive, and, therefore, the total welfare equals in magnitude to the expected penalty to the insider.

Similarly, one also expects the insider's total wealth to vanish as the rate of penalties increases without bound. On the other hand, the following result shows that the expected penalty in equilibrium is bounded in κ. Thus, the vanishing wealth in the limit is not a consequence of large expected costs but rather results from the insider trading small quantities. This is consistent with the distribution of the trading rate via Corollary 5.4.

The expected penalty is also non-monotone, as can be deduced from the next corollary or Fig. 1.

Corollary 5.6. *Consider the equilibrium in Corollary 5.3. The insider's expected penalties in equilibrium equals*

$$P(\kappa) = -\frac{\gamma\sigma}{2}\kappa\log(\kappa\Lambda(\kappa)).$$

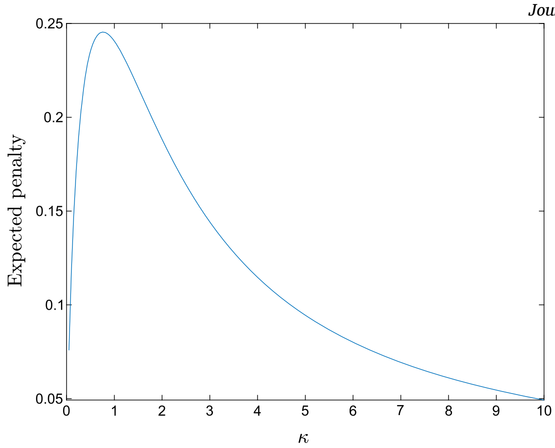

Fig. 1. Expected penalty paid by the insider in equilibrium as a function of κ. The values on the y-axis are normalized by $\gamma\sigma$.

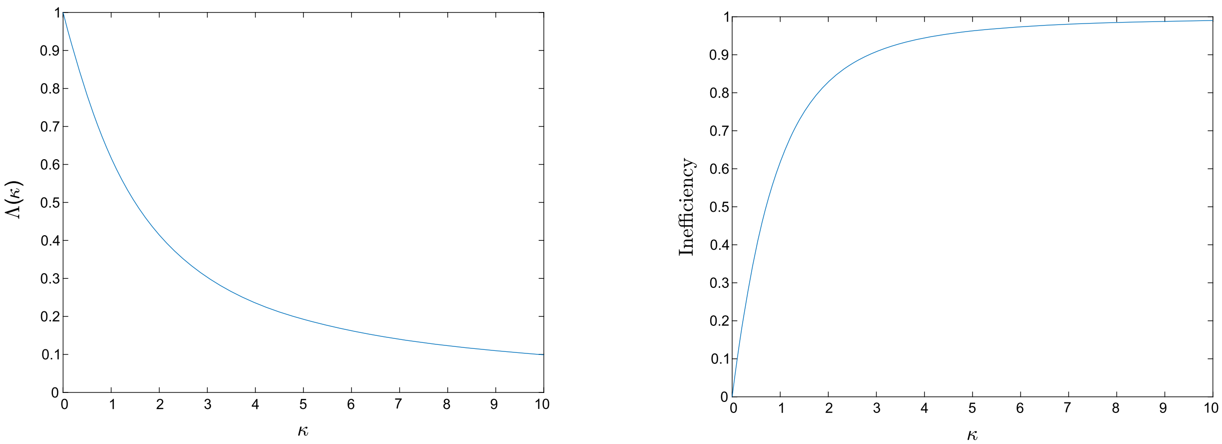

Fig. 2. The left pane shows the expected loss of noise traders normalized by $\gamma\sigma$, whereas the right pane is the price inefficiency of equilibrium normalized by γ^2.

P is bounded by $\frac{\gamma\sigma}{2}$, and there exists a κ_0 such that it is concave on $(0,\kappa_0)$ and convex otherwise. Furthermore, $P(0+)=P(\infty)=0$.

Although the expected penalties are non-monotone, the loss of the noise traders and price efficiency decrease as the rate of penalties, κ, increases. These can be verified by direct differentiation of the relevant expressions and are illustrated in Fig. 2.

6. Optimal insider trading regulation

Regulatory policies toward insider trading violations typically aim to limit the losses of uninformed traders. On the other hand, too much restriction on informed trading harms the price informativeness. This section will address the concerns of a regulator seeking the optimal trade-off between the price efficiency and the losses of uninformed (noise) traders given a budget constraint.

It is assumed that the fundamental value has a Gaussian distribution, as in Section 5.1, and the regulator has the following simple objective:

$$\begin{aligned} &\min_{\kappa} \Lambda(\kappa)\gamma\sigma + R\kappa\Lambda(\kappa)\gamma^2, \\ &\text{subject to } -\frac{\gamma\sigma}{2}\kappa\log(\kappa\Lambda(\kappa)) \geq b, \end{aligned} \tag{6.56}$$

for some $R>0$ and $b>0$. Recall that the first term is the expected loss of the noise traders and the second is the expected post-trade variance adjusted by the factor R, which measures the sensitivity of the policy toward price efficiency. The budget constraint stipulates that the expected penalty must exceed the cost of investigation.

The expected penalties in equilibrium are bounded in penalty rate κ, as shown in Corollary 5.6. More precisely,

$$\bar{P} := \sup_{\kappa} \frac{P(\kappa)}{\gamma\sigma} < \infty. \tag{6.57}$$

This motivates considering the normalized cost $b_0 := \frac{b}{\gamma\sigma}$.

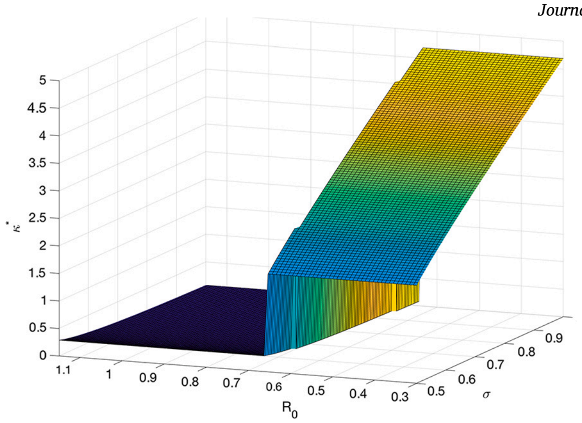

Fig. 3. The optimal penalty rate κ^* as a function of noise volatility σ, and the regulator's sensitivity toward price efficiency, R_0. The figure assumes $\gamma = 1$ and $b = 0.1$.

If $b_0 > \bar{P}$, the constraint cannot be satisfied. Note that $\sigma\gamma$ is a measurement of the risk that regulator faces: σ is roughly the amount of uninformed trade, and γ captures the adverse selection risk. Thus, normalizing the cost of investigation by *the regulatory risk* $\gamma\sigma$ allows the regulators to assess whether the cost of beginning an investigation is justified by the benefit it provides.

If the costs outweigh the benefits, that is, $b_0 > \bar{P}$, the regulator either sets the penalty at null and does not investigate insider trading, or sets the penalty at prohibitively high levels; that is, $\kappa = \infty$. Under the former scenario, the insider trades as Kyle's informed trader. In this case, post-trade variance is 0, and the expected loss of noise traders is $\gamma\sigma$. On the other hand, if the regulator applies the extreme penalty of the alternative scenario, Λ becomes 0, yielding zero loss to the noise traders but, at the same time, the price inefficiency will be at the worst level of γ^2. More precisely, the value of this policy to the regulator will be $R\gamma^2$. Comparing the two alternatives, the regulator will choose a no-penalty policy if $R > \frac{\sigma}{\gamma}$. That is, there will be no insider trading investigation if the regulator's sensitivity toward price efficiency is higher than the amount of noise trading per adverse selection risk.

In practice, particularly in the case of an illegal insider, it is reasonable to assume that γ is not small since no one will be willing to take the risk associated with obtaining information illegally unless its value, γ, is large. Thus, if the amount of noise trading is low, that is, small σ, R is likely to be higher than $\frac{\sigma}{\gamma}$ and the regulator will choose not to investigate the insider trading allegations.

If $b_0 \le \bar{P}$, the budget constraint is binding. Indeed, writing the objective function in terms of Λ implies that the regulator minimizes

$$\gamma\sigma\Lambda + R\gamma^2(1 - \Lambda^2).$$

As this objective is concave in Λ, its minimizer is at the boundary values of Λ, which belong to $\{0, 1\}$. However, at these boundary values, the expected penalty on the insider is 0, and, hence, cannot meet the constraint.

The budget constraint entails that the candidate optimal κ must belong to the set $\mathcal{K}(b_0)$, where

$$\mathcal{K}(a) := \{\kappa : -\kappa \log(\kappa\Lambda(\kappa)) \ge a\}. \tag{6.58}$$

It follows from Corollary 5.6 (see also Fig. 1) that $\mathcal{K}(a) = [-\kappa_0(a), \kappa_1(a)]$, where $\kappa_i(a)$ solves

$$-\kappa \log(\kappa\Lambda(\kappa)) = a.$$

It is easy to see that the equation above has exactly two solutions for $a < \bar{P}$, and the solution is unique for $a = \bar{P}$. Moreover, by rewriting R as $R_0\frac{\sigma}{\gamma}$ for some $R_0 \ge 0$, the following characterization of the optimal policy is obtained.

Theorem 6.1. *Consider the equilibrium given in Corollary 5.3, and denote by κ^* the optimal penalty rate policy determined by (6.56). Then*

$$\kappa^* = \underset{\kappa \in \mathcal{K}(b_0)}{\arg\min}\, \Lambda(\kappa) + R_0(1 - \Lambda^2(\kappa)), \tag{6.59}$$

where $\mathcal{K}(b_0)$ is defined in (6.58).

Observe that although the new parametrization makes the objective function independent of γ and σ, the optimizer continues to depend on those parameters since the budget constraint does.

The set $\mathcal{K}(b_0)$ for $b_0 < \bar{P}$ contains two elements corresponding to a high liquidity (small Λ) and low liquidity (high Λ) regimes. Therefore, if $b_0 < \bar{P}$, regulators choose κ, leading to lower liquidity, when they are more concerned with price informativeness, that is, big R. Otherwise, they choose the penalty rate associated with higher liquidity, that is, smaller loss for noise traders.

As one can see from Fig. 3, once the regulator's normalized sensitivity toward price efficiency, R_0, is above a certain level, the optimal penalty level is rather low, leading to a less liquid but more efficient market. On the other hand, when the sensitivity is relatively low, the optimal penalty level is rapidly increasing with the amount of noise trading measured by σ.

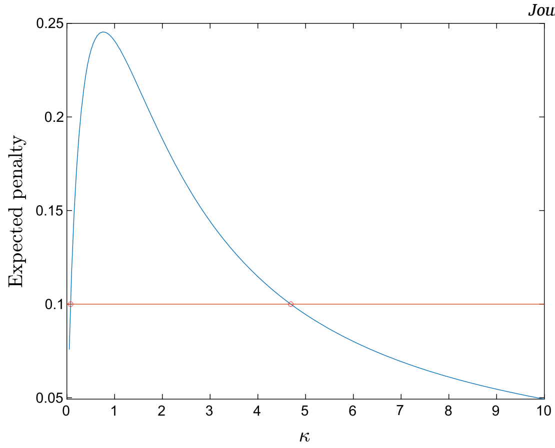

Fig. 4. The expected penalty normalized by $\gamma\sigma$ vs. the budget constraint at $b_0 = 0.1$. Intersection on the left occurs roughly at $\kappa = 0.0791$, the right intersection occurs around $\kappa = 4.6899$.

An interesting future of the shape of the optimal penalty surface is that while the optimal rate is increasing with σ for small enough R_0, it is decreasing in σ otherwise. The reason for this seemingly contradictory outcome lies in the budget constraint: When σ is larger, the budget constraint is less restrictive (small b_0) and the optimal rate is closer to the boundary values of 0 and ∞. Subsequently, if the regulators care less about the price efficiency, they would choose a rather high penalty, which is responsible for the optimal penalty rate increasing in noise volatility for small R_0. On the other hand, if the regulators care more about the price efficiency, they will choose small penalty (closer to the boundary 0), which explains the optimal rate being small and decreasing in σ for high R_0. Also, observe from Fig. 1 that the graph of expected penalties as a function of κ is heavily skewed to the left. Thus, when the regulators consider (6.59) for a given (normalized) penalty level b_0, the optimal κ is always to the left of 1 when they care more about price efficiency, whereas the optimal level is unbounded from above when they are chiefly concerned with the losses of the uninformed. Fig. 4 illustrates this point when $b_0 = 0.1$.

7. Conclusion

When there are quadratic penalties, the equilibrium in the continuous time version of the Kyle model changes drastically. The insider always trades a constant multiple of the discrepancy between her own valuation and what she believes to be the market price right before her private information becomes public. As a result, the insider's trading strategy explicitly depends on the distribution of the fundamental value of the asset.

The noise traders lose less on average as the potential penalties on the insider increase. However, the total expected penalties in equilibrium is non-monotone in the rate of penalties when the fundamental value is normally distributed. Interestingly, the maximum expected penalty that is incurred by the insider occurs when the rate of penalties is sufficiently small.

A regulator with budget constraints runs an investigation only if the benefits of the investigation is sufficiently larger than what it costs. Furthermore, the optimal penalty policy reduces to a choice of one of two extremal penalty levels that correspond to high and low liquidity regimes. The regulator's choice is determined by the amount of noise trading and the relative importance of price informativeness.

Existence of equilibrium is established via the theory of entropy regularized optimal transport. The interesting extension to equilibrium with general convex costs that will possibly require a different type of regularization to optimal transportation remains for future research. Such an extension will be valuable to regulators searching for the optimal penalty structure beyond quadratic costs to deter illegal insider trading, as discussed by Carré et al. (2022) in a one-period setting.

Declaration of competing interest

None.

Appendix A. Proofs and other technical considerations

A.1. *Constructing the insider's filtration*

Augmenting the insider's filtration with the $P^{0,v}$-null sets will create an extra dependence on the value of V purely for technical reasons. To avoid this, we assume that the informed trader's filtration, denoted by $\mathcal{F}^I$, is the right continuous augmentation[22] of the filtration generated by S and V with the sets of

[22] See Section 3 of Sharpe (1988) for the precise procedure.

$$\mathcal{N}^I := \{E \subset \mathcal{F} : P^{0,v}(E) = 0, \forall v \in f(\mathbb{R})\}. \tag{A.60}$$

Similarly, $\mathcal{F}^{B,V}$ will denote the right continuous augmentation of the filtration generated by B and V with the sets of $\mathcal{N}^I$. Note that $\mathcal{F}^I = \mathcal{F}^{B,V}$ if H is strictly increasing in total order, which will be the case when H is an *admissible pricing rule* per Definition 2.1.

A.2. Existence and uniqueness of a fixed point for T

Proposition A.1. *Consider $\phi \in \mathcal{D}(T_1)$ and let $\Psi = \hat{c}\log T_1\phi$. Then the following statements are valid:*

(1) *If $\mathbb{E}[|\phi(\sigma B_1)|] < \infty$, Ψ is bounded from below as follows:*

$$\Psi \geq -\int_{\mathbb{R}} \phi(y)\frac{1}{\sqrt{2\pi\sigma^2}}\exp\Big(-\frac{y^2}{2\sigma^2}\Big)dy.$$

(2) *Suppose that Ψ is bounded from below and all exponential moments of V is finite. Then $\phi \in \mathcal{D}(T)$, and $T\phi$ is infinitely differentiable and strictly convex. In particular,*

$$\frac{d}{dy}T\phi(y) = \frac{\int_{f(\mathbb{R})} v\exp\Big(\frac{yv-\Psi(v)}{\hat{c}}\Big)\Pi(dv)}{\int_{f(\mathbb{R})} \exp\Big(\frac{yv-\Psi(v)}{\hat{c}}\Big)\Pi(dv)} \quad \textit{and} \tag{A.61}$$

$$\frac{d^2}{dy^2}T\phi(y) = \frac{1}{\hat{c}}\frac{\int_{f(\mathbb{R})} v^2\exp\Big(\frac{yv-\Psi(v)}{\hat{c}}\Big)\Pi(dv)}{\int_{f(\mathbb{R})} \exp\Big(\frac{yv-\Psi(v)}{\hat{c}}\Big)\Pi(dv)} - \frac{1}{\hat{c}}\left(\frac{\int_{f(\mathbb{R})} v\exp\Big(\frac{yv-\Psi(v)}{\hat{c}}\Big)\Pi(dv)}{\int_{f(\mathbb{R})} \exp\Big(\frac{yv-\Psi(v)}{\hat{c}}\Big)\Pi(dv)}\right)^2. \tag{A.62}$$

Proof. (1) This is a straightforward consequence of Jensen's inequality applied to the natural logarithm and definition of T_1 in (4.23).

(2) Since $\phi \in \mathcal{D}(T_1)$, Ψ is finite. Thus, $T_2\Psi$ never vanishes. In addition, the lower bound on Ψ yields, for some constant K, that

$$T_2\Psi(y) \leq K\int_{f(\mathbb{R})} \exp\Big(\frac{yv}{\hat{c}}\Big)\Pi(dv) < \infty.$$

Thus, $T\phi$ is finite everywhere. This yields the claim.
Note that since the moment-generating function is finite everywhere, it is smooth and in particular (see (21.24) in Section 21 and the preceding discussion in Billingsley (1995)),

$$M_V^{(n)}(r) = \int_{f(\mathbb{R})} v^n e^{rv}\Pi(dv) < \infty, \qquad \forall r \in \mathbb{R} \text{ and } n \geq 0.$$

Thus, the integrability property above, the lower bound on Ψ, and the dominated convergence theorem immediately yield (A.61). By iterating this operation one obtains that the numerator in (4.27) is infinitely differentiable by the smoothness of $M_V^{(n)}$. Similarly, the second derivative is given by (A.62), which is positive being a positive multiple of the variance of a random variable $\tilde{v}$ with the probability density

$$P(\tilde{v} \in dx) = \frac{\exp\Big(\frac{yx-\Psi(x)}{\hat{c}}\Big)\Pi(dx)}{\int_{f(\mathbb{R})} \exp\Big(\frac{yv-\Psi(v)}{\hat{c}}\Big)\Pi(dv)}, \qquad x \in f(\mathbb{R}).$$

Since $\tilde{v}$ is not constant, the second derivative is strictly positive; hence, the strict convexity follows. □

An analogous result, whose proof is omitted, also holds for Ψ. To present its statement, define

$$\tilde{T}\Psi = \hat{c}\log T_1(\hat{c}\log T_2(\Psi)). \tag{A.63}$$

Proposition A.2. *Consider $\Psi \in \mathcal{D}(T_2)$ and let $\phi = \hat{c}\log T_2\Psi$. Then the following statements are valid:*

(1) *If $\mathbb{E}[|\Psi(V)|] < \infty$, ϕ is bounded from below as follows:*

$$\phi(y) \geq y\mathbb{E}[V] - \mathbb{E}[\Psi(V)].$$

(2) *Suppose that ϕ is bounded from below by an affine function. Then $\Psi \in \mathcal{D}(\tilde{T})$, and $\tilde{T}\Psi$ is infinitely differentiable and strictly convex. In particular,*

$$\frac{d}{dv}\tilde{T}\Psi(v) = \frac{\int_{\mathbb{R}} y \exp\left(\frac{yv-\phi(y)}{\hat{c}}\right)p(\sigma,y)dy}{\int_{\mathbb{R}} \exp\left(\frac{yv-\phi(y)}{\hat{c}}\right)p(\sigma,y)dy} \text{ and}$$

$$\frac{d^2}{dv^2}\tilde{T}\Psi(v) = \frac{\int_{\mathbb{R}} y^2 \exp\left(\frac{yv-\phi(y)}{\hat{c}}\right)p(\sigma,y)dy}{\int_{\mathbb{R}} \exp\left(\frac{yv-\phi(y)}{\hat{c}}\right)p(\sigma,y)dy} - \left(\frac{\int_{\mathbb{R}} y \exp\left(\frac{yv-\phi(y)}{\hat{c}}\right)p(\sigma,y)dy}{\int_{\mathbb{R}} \exp\left(\frac{yv-\phi(y)}{\hat{c}}\right)p(\sigma,y)dy}\right)^2,$$

where $p(\sigma,\cdot)$ is the density of a normal distribution with variance σ^2.

Remark 5. The differentiability results above seemingly do not make sense when the support of V contains singletons. However, since the object of interest is the fixed point of $\tilde{T}$, one can limit attention to Ψ of the form $\Psi(v) = \hat{c}\log T_1(\tilde{\phi})$. Then one can directly extend the definition of Ψ to the interval $I = \{(\ell, r) : \ell = \inf V, r = \sup V\}$ since $\hat{c}\log T_1(\tilde{\phi})$ is well-defined over the entire I.

The results above indicate that the fixed point of T is likely to be a smooth convex function.

Proof of Theorem 4.1. The existence of (ϕ,Ψ) solving the system (4.21)-(4.22) follows from Theorem 3 in Rüschendorf and Thomsen (1993). Indeed, existence of such a solution is equivalent to solving the following system:

$$\begin{aligned} \int_{f(\mathbb{R})} \exp\left(\frac{-(y-v)^2/2 - \Psi^0(v)}{\hat{c}}\right)\Pi(dv) &= \exp\left(\frac{\phi^0(y)}{\hat{c}}\right), \qquad y \in \mathbb{R}; \\ \int_{\mathbb{R}} \exp\left(\frac{-(y-v)^2/2 - \phi^0(y)}{\hat{c}}\right)\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{y^2}{2\sigma^2}\right)dy &= \exp\left(\frac{\Psi^0(v)}{\hat{c}}\right), \qquad v \in f(\mathbb{R}). \end{aligned} \tag{A.64}$$

Clearly, (ϕ,Ψ) solve the system (4.21)-(4.22) if and only if (ϕ^0,Ψ^0) solve (A.64), where $\phi^0(u) = \phi(u) - u^2/2$ and $\Psi^0(v) = \Psi(v) - v^2/2$. In the notation of Rüschendorf and Thomsen (1993), consider the measure Q on the product Borel σ-algebra of $\mathbb{R}\times f(\mathbb{R})$ given by $Q(dy,dv) = C\exp(-(y-v)^2/(2\hat{c}))p(\sigma,y)dy\Pi(dv)$, where C is a normalizing constant to turn Q into a probability measure. Set $P(dy,dv) = p(\sigma,y)dy\Pi(dv)$. Since the relative entropy $H(P||Q) = \int \frac{(y-v)^2}{2\hat{c}}p(\sigma,y)dy\Pi(dv) - \log C < \infty$, Theorem 3 therein yields the existence of a solution to (A.64) as well as uniqueness up to an additive constant. This proves the existence of a unique fixed point ϕ of T with $\phi(0) = 0$.

Moreover, Remark 1 following the same theorem in Rüschendorf and Thomsen (1993) also yields that

$$\int |\phi^0(y)| + |\Psi^0(v)|\Pi(dv)p(\sigma,y)dy < \infty$$

since the condition of Proposition 1 therein is satisfied due to the product form of P. This establishes the claimed integrability conditions of $\phi(\sigma B_1)$ and $\Psi(V)$ in view of the relationship between (ϕ,Ψ) and (ϕ^0,Ψ^0) since the normal distribution and V have finite second moments.

The remaining assertions follow from Propositions A.1 and A.2. □

A.2.1. Proofs for Section 5

Lemma A.1. *The following bounds apply:*

$$E^{0,v}\left[\exp\left(-\frac{j^*(\sigma B_1,v)}{\hat{c}}\right)\Big|\sigma B_t = x\right] \leq \exp\left(-\frac{\Psi^*(v)}{\hat{c}} + \frac{(1-t)(v-h^*(0))^2}{2c^2\sigma^2} + (v-h^*(0)x\right)$$

$$E^{0,v}\left[\exp\left(-\frac{j^*(\sigma B_1,v)}{\hat{c}}\right)\Big|\sigma B_t = x\right] \geq \exp\left(\frac{vx - \Psi^*(v) - E^{0,v}[\phi^*(\sigma B_1)|\sigma B_t = x]}{\hat{c}}\right).$$

Proof. Since ϕ^* is convex by Proposition A.1 and vanishes at 0, $\phi^*(y) \geq yh^*(0)$. Thus,

$$\begin{aligned} E^{0,v}\left[\exp\left(-\frac{j^*(\sigma B_1,v)}{\hat{c}}\right)\Big|\sigma B_t = x\right] &= \exp\left(-\frac{\Psi^*(v)}{\hat{c}}\right)E^{0,v}\left[\exp\left(\frac{v\sigma B_1 - \phi^*(\sigma B_1)}{\hat{c}}\right)\Big|\sigma B_t = x\right] \\ &\leq \exp\left(-\frac{\Psi^*(v)}{\hat{c}}\right)E^{0,v}\left[\exp\left(\frac{\sigma B_1(v-h^*(0))}{\hat{c}}\right)\Big|\sigma B_t = x\right], \end{aligned}$$

which yields the first claim in view of the moment-generating function of normal distribution. The second claim is a direct consequence of Jensen's inequality. □

Proposition A.3. *Suppose that Assumption 5.1 holds. For each v, there exists a unique strong solution to (5.39).*

Let $\mathbb{Q}^*$ be the law induced by Y^* on the space of continuous functions on $[0,1]$ vanishing at 0; that is, $C_0([0,1])$. Let $(\mathcal{B}_t)_{t\in[0,1]}$ be the right continuous augmentation of the natural filtration of the coordinate process X and $\mathbb{W}$ be the Wiener measure. Then $\mathbb{Q}^*$ is given by the following h-transform of Brownian motion:

$$\mathbb{E}^{\mathbb{Q}^*}[F] = \frac{\mathbb{E}^{\mathbb{W}}\left[F \exp\left(\frac{v\sigma X_1 - \phi^*(\sigma X_1)}{\hat{c}}\right)\right]}{\mathbb{E}^{\mathbb{W}}\left[\exp\left(\frac{v\sigma X_1 - \phi^*(\sigma X_1)}{\hat{c}}\right)\right]}, \quad F \in \mathcal{B}_1. \tag{A.65}$$

The relative entropy of this h-transform is given by

$$\frac{1}{2}E^{0,v}\left[\sigma^2 \int_0^1 \left(\frac{\rho^*_y(s, Y^*_s, v)}{\rho^*(s, Y_s, v)} ds\right)^2\right] = H(\mathbb{Q}^*||\mathbb{W}) := \mathbb{E}^{\mathbb{W}}\left[\frac{d\mathbb{Q}^*}{d\mathbb{W}} \log \frac{d\mathbb{Q}^*}{d\mathbb{W}}\right] \tag{A.66}$$
$$= \frac{v(\Psi^*)'(v) - \Psi^*(v) - E^{0,v}[\phi^*(Y^*_1)]}{\hat{c}} < \infty.$$

Proof. Existence of a unique strong solution to (5.39) follows from Theorem 2.2.7 in Çetin and Danilova (2018) since for each v, $\rho^*(\cdot,\cdot,v) \in C^{1,2}([0,1],\mathbb{R})$ and ρ^* never vanishes. Note that the claimed smoothness is valid even at time-1 since ϕ^* is smooth by Assumption 5.1. Then (A.65) follows from Girsanov's theorem after noticing that $\rho(1,y,v) = \exp(-\frac{\Psi(v)}{\hat{c}})\exp(\frac{yv-\phi(y)}{\hat{c}})$ so that $\exp(-\frac{\Psi(v)}{\hat{c}})$ drops out in change of measure calculations.

Thus, the equality in (A.66) is a consequence of (A.65) and the definition of relative entropy. To compute its desired expression, note that

$$\frac{d\mathbb{Q}^*}{d\mathbb{W}} = \exp\left(\frac{v\sigma X_1 - \phi^*(\sigma X_1) - \Psi^*(v)}{\hat{c}}\right).$$

Thus,

$$\begin{aligned}
H(\mathbb{Q}^*||\mathbb{W}) &= \mathbb{E}^{\mathbb{W}}\left[\frac{v\sigma X_1 - \phi^*(\sigma X_1) - \Psi^*(v)}{\hat{c}} \exp\left(\frac{v\sigma X_1 - \phi^*(\sigma X_1) - \Psi^*(v)}{\hat{c}}\right)\right] \\
&= \frac{v(\Psi^*)'(v) - \Psi^*(v)}{\hat{c}} - \mathbb{E}^{\mathbb{W}}\left[\frac{\phi^*(\sigma X_1)}{\hat{c}} \exp\left(\frac{v\sigma X_1 - \phi^*(\sigma X_1) - \Psi^*(v)}{\hat{c}}\right)\right] \\
&= \frac{v(\Psi^*)'(v) - \Psi^*(v) - E^{0,v}[\phi^*(Y^*_1)]}{\hat{c}},
\end{aligned}$$

where the first equality follows from (4.25). The above is finite by the hypothesis of this section that j^* satisfies (3.17). □

Proof of Theorem 5.1. (1) *Market efficiency:* Note that given θ^*, Y^* is the unique strong solution of (5.39). Then it follows from Proposition 3.1 and Theorem 4.1 that $\mathbb{P}(V \in dv|\mathcal{F}^{Y^*}_t) = \rho^*(t, Y^*_t, v)\Pi(dv)$, and (3.16) holds with ρ replaced by ρ^* since $\rho^*(0,0,\cdot) \equiv 1$ by construction. This implies that $\frac{Y^*}{\sigma}$ is a Brownian motion in its own filtration by considering the optional projection of the drift. Moreover,

$$H^*(t, Y^*_t) = \mathbb{E}[V|\mathcal{F}^{Y^*}_t]$$

by (5.33). That is, the market efficiency condition of equilibrium is satisfied.

(2) *Insider optimality and value function:* Suppose H^* is the pricing rule of the market makers, and consider an admissible strategy $d\theta_t = \alpha_t dt$. Then Ito's formula and the PDE (5.37) yield

$$\begin{aligned}
J(u, Y_u) &= J(0,0) + \int_0^u J_y(t, Y_t)\{\sigma dB_t + \alpha_t dt\} - \int_0^u \frac{J_y(t,Y_t) + v - H^*(t,Y_t)}{2c} dt \\
&= J(0,0) + \int_0^u J_y(t, Y_t)\sigma dB_t - \int_0^u \frac{\left(J_y(t,Y_t) + v - H^*(t,Y_t) - \alpha c\right)^2}{2c} dt \\
&+ \frac{c}{2}\int_0^u \alpha_t^2 dt - \int_0^u (v - H^*(t, Y_t))\alpha_t dt.
\end{aligned} \tag{A.67}$$

First, observe that $J \geq 0$ by a simple application of Jensen's inequality. Furthermore, admissibility condition on θ implies α and H^* satisfies (2.6). Thus, the local martingale in (A.67) is bounded from below by an integrable random variable. That is, $(\int_0^u J_y(t, Y_t) dB_t)$ is a $P^{0,v}$ supermartingale. Using the boundary condition of J, one obtains

$$E^{0,v}W^\theta_1 \leq J(0,0) - E^{0,v}\int_0^u \frac{\left(J_y(t, Y_t + v) - H^*(t,Y_t) - \alpha c\right)^2}{2c} dt \leq J(0,0).$$

Note that the first inequality becomes equality if $(\int_0^u J_y(t,Y_t)dB_t)$ is a martingale, and the second term will disappear when $\theta=\theta^*$. Thus, θ^* is optimal provided it is admissible and the stochastic integral above is a martingale. In such a case, the expected wealth is given by

$$J(0,0)=J^0(0,0)+\hat{c}\log\rho^*(0,0,v)=E^{0,v}[j^*(\sigma B_1,v)],$$

which coincides with (5.40).

(3) *Optimality of θ^*:* Proposition A.3 yields that

$$E^{0,v}\Big[\int_0^1\Big(\frac{\rho_y(s,Y_s,v)}{\rho(s,Y_s,v)}\Big)^2 ds\Big]<\infty, \tag{A.68}$$

for all v in the support of V. Thus, θ^* is admissible when $E^{0,v}\int_0^1 H^2(t,Y_t^*)dt<\infty$.

To this end, note that since $H(t,Y_t^*)=\mathbb{E}[V|\mathcal{F}_t^{Y^*}]$, $\mathbb{E}[H^2(t,Y_t^*)]\le\mathbb{E}[V^2]<\infty$. Thus, $E^{0,v}\int_0^1 H^2(t,Y_t^*)dt$ is finite for almost all v. On the other hand, it follows from (A.65) that

$$E^{0,v}\Big[\int_0^1 H^2(t,Y_t^*)dt\Big]=\frac{\mathbb{E}^{\mathbb{W}}\Big[\int_0^1\exp\Big(\frac{v\sigma X_1-\phi^*(\sigma X_1)}{\hat{c}}\Big)H^2(t,\sigma X_t)dt\Big]}{\mathbb{E}^{\mathbb{W}}\Big[\exp\Big(\frac{v\sigma X_1-\phi^*(\sigma X_1)}{\hat{c}}\Big)\Big]}.$$

Note that the denominator is finite for each v by Lemma A.1, convex in v, and larger than $\exp(-\hat{c}^{-1}\mathbb{E}^{\mathbb{W}}[\phi^*(\sigma X_1)])>0$. Thus, the numerator is finite for almost all v. Since it is also convex in v, its finiteness for all v follows. Therefore, $E^{0,v}\int_0^1 H^2(t,Y_t^*)dt<\infty$ for all v.

Finally, it remains to show that $(\int_0^u J_y(t,Y_t)dB_t)$ is a $P^{0,v}$-martingale. However, $J_y(t,y,v)=\hat{c}\frac{\rho_y(t,y,v)}{\rho(t,y,v)}+H^*(t,y)-v$. Thus, $(\int_0^u J_y(t,Y_t)dB_t)$ is a $P^{0,v}$-martingale since $E^{0,v}\int_0^1 H^2(t,Y_t^*)dt<\infty$, and (A.68) holds.

(4) *Noise traders loss:* It follows from integration by parts that the wealth of noise traders is given by

$$\sigma B_1(V-H^*(1,Y_1))-\sigma\int_0^1 B_t dH^*(t,Y_t^*)=\sigma B_1 V-\sigma\int_0^1 H^*(t,Y_t^*)dB_t-\sigma^2\int_0^1 H_y^*(t,Y_t^*)dt.$$

The admissibility of θ^* entails that the stochastic integral with respect to B is a martingale. Also note that $H_y^*(t,Y_t^*)$ is an $\mathcal{F}^{Y^*}$-martingale. Thus, taking expectations leads to the following expected wealth:

$$-\sigma^2\mathbb{E}[H_y^*(1,Y_1^*)]=-\sigma^2\mathbb{E}\Big[\frac{d^2\phi^*}{dy^2}(Y_1^*)\Big].$$

(5) *Expected penalty in equilibrium:* This is a direct consequence of Proposition A.3 since

$$\alpha_t^*=\sigma^2\frac{\rho_y^*(t,Y_t,V)}{\rho^*(t,Y_t,V)}.\quad\square$$

Proof of Proposition 5.2. First, observe that $\frac{\rho_y(t,Y_t^*,v)}{\rho(t,Y_t^*,v)}$ is a $P^{0,v}$-martingale. Indeed, it follows from Proposition A.3 that

$$\begin{aligned}E^{0,v}\Big[\frac{\rho_y(1,Y_1^*,v)}{\rho(1,Y_1^*,v)}\Big|Y_t^*=y\Big]&=\frac{E^{0,v}\big[\rho_y(1,\sigma B_1,v)\big|\sigma B_t=y\big]}{E^{0,v}\big[\rho(1,\sigma B_1,v)\big|B_t=y\big]}\\&=\frac{E^{0,v}\big[\rho_y(1,\sigma B_1,v)\big|\sigma B_t=y\big]}{\rho(t,y,v)},\end{aligned}$$

where the last equality follows from the fact that $\rho(t,\sigma B_t,v)$ is a $P^{0,v}$-martingale by the definition of ρ. Thus, $\frac{\rho_y(t,Y_t^*,v)}{\rho(t,Y_t^*,v)}$ will be a $P^{0,v}$-martingale upon $\rho_y(t,\sigma B_t,v)$ becoming a $P^{0,v}$-martingale. Note that this will follow when one can differentiate under the expectation in (5.34) due to the fact that $p_y(t,y,z)=-p_z(t,y,z)$, where $p(t,y,z)$ represents the transition density of Brownian motion. Then the required differentiability readily follows from the assumption that h^* has at most exponential growth.

Thus,

$$\alpha_t^*=\sigma^2 E^{0,v}\Big[\frac{\rho_y(1,Y_1^*,v)}{\rho(1,Y_1^*,v)}\Big|Y_t^*=y\Big],$$

which coincides with the expression in the statement by direct manipulation. $\square$

Proof of Corollary 5.3. The only claim that is not explicitly given by Theorems 5.1 and 5.2 is the conditional distribution of V. Note that the conditional distribution at time-1 is given by $\exp(-j^*(y,v)/c)\Pi(dv)$, where j^* is given by (5.32). Straightforward computations show that this distribution is Gaussian with density proportional to

$$\exp\Big(-(\mu-v)^2 \frac{\gamma^2+\hat{c}(c+\lambda^*)}{2\gamma^2\hat{c}(c+\lambda^*)} + \frac{vy}{\hat{c}}\Big).$$

Since $\frac{\gamma^2(c+\lambda^*)}{\gamma^2+\hat{c}(c+\lambda^*)} = \lambda^*\hat{c}$ by (4.30), it follows that the conditional distribution at time-1 is Gaussian with mean $\lambda^* Y_1^* + \mu$ and variance $\hat{c}\lambda^*$. Thus, the conditional density at time t given $Y_t^* = y$ equals

$$\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma^2(1-t)}} \exp\Big(-\frac{(y-z)^2}{2\sigma^2(1-t)}\Big) \frac{1}{\sqrt{2\pi\hat{c}\lambda^*}} \exp\Big(-\frac{(v-\mu-\lambda^* z)^2}{2\hat{c}\lambda^*}\Big) dz dv = \rho^*(t,y,v)\Pi(dv).$$

This will be a Gaussian density as well. Indeed, by considering its characteristic function

$$\begin{aligned}\int_{\mathbb{R}} e^{irv}\rho^*(t,y,v)\Pi(dv) &= e^{-\frac{r^2\hat{c}\lambda^*}{2}} \int_{\mathbb{R}} \frac{e^{ir(\mu+\lambda^* z)}}{\sqrt{2\pi\sigma^2(1-t)}} \exp\Big(-\frac{(y-z)^2}{2\sigma^2(1-t)}\Big) dz \\ &= e^{-\frac{r^2}{2}(\hat{c}\lambda^*+(\lambda^*)^2\sigma^2(1-t))+ir(\mu+\lambda^* y)}.\end{aligned}$$

Therefore, the conditional distribution is Gaussian with mean $\mu + \lambda^* Y_t^*$ and variance $\hat{c}\lambda^* + (\lambda^*)^2\sigma^2(1-t)$. Using the quadratic equation that λ^* solves from Theorem 5.2, the variance, in fact, equals $\gamma^2 - t(\lambda^*\sigma)^2$. Using the Gaussian density, this readily yields the equilibrium θ^*, using Corollary 5.2. □

Proof of Corollary 5.5. The first and last claims follow from Corollary 5.3 upon noticing that $(c+\lambda^*)\frac{\sigma}{\gamma} = \kappa + \Lambda(\kappa) = \frac{1}{\Lambda(\kappa)}$ and $\kappa\Lambda(\kappa) = 1-\Lambda^2(\kappa)$, in view of (5.50).

Differentiating $\mathcal{W}$ with respect to κ and noting that $1-\Lambda^2(\kappa) = \Lambda(\kappa)\kappa$ yield

$$\mathcal{W}'(\kappa) = \frac{\gamma\sigma}{2}\log(1-\Lambda^2(\kappa)), \quad \mathcal{W}''(\kappa) = -\gamma\sigma\Lambda'(\kappa)\kappa^{-1}. \tag{A.69}$$

The identity in (5.50) also yields

$$\Lambda'(\kappa) = -\frac{\Lambda(\kappa)}{2\Lambda+\kappa} = -\frac{\Lambda(\kappa)}{\Lambda(\kappa)+\frac{1}{\Lambda(\kappa)}} = -\frac{\Lambda^2(\kappa)}{1+\Lambda^2(\kappa)} < 0.$$

This completes the proof of the claims on derivatives. To show that the limiting wealth is 0, first recall from Corollary 5.2 that $\Lambda(\infty) = 0$. On the other hand,

$$\kappa\log(\kappa\Lambda(\kappa)) = \frac{1-\Lambda^2(\kappa)}{\Lambda(\kappa)}\log(1-\Lambda^2(\kappa)).$$

Thus, the proof will be complete if $\lim_{x\to 0}\frac{\log(1-x^2)}{x} = 0$, which follows by a quick application of the L'Hospital rule. □

Proof of Corollary 5.6. Note that for $x \in (0,1)$, $\log(1-x) \geq -\frac{x}{1-x}$ in view of the mean value theorem. Thus,

$$\frac{1-\Lambda^2(\kappa)}{\Lambda(\kappa)}\log(1-\Lambda^2(\kappa)) \geq -\Lambda(\kappa) \geq -1.$$

This shows the bound on P. Moreover, (A.69) implies that

$$P''(\kappa) = \gamma\sigma(\Lambda'(\kappa)\kappa^{-1} + 2\Lambda''(\kappa)) = \gamma\sigma\Big(-\frac{\kappa^{-1}\Lambda^2(\kappa)}{1+\Lambda^2(\kappa)} + 4\frac{\Lambda^3(K)}{(1+\Lambda^2(\kappa))^3}\Big),$$

which is positive if and only if

$$\frac{(1+\Lambda^2(\kappa))^2}{\kappa} \leq 4.$$

Note that the left side is decreasing from ∞ to 0 as κ increases to ∞. Thus, there exists a κ_0 such that P is concave on $(0,\kappa_0)$ and convex otherwise.

The limiting values at 0 and ∞ have already been observed in computations leading to the respective limits for $\mathcal{W}$ in the proof of the previous corollary. □

Data availability

No data was used for the research described in the article.

References

Back, Kerry, 1992. Insider trading in continuous time. Rev. Financ. Stud. 5 (3), 387–409.
Back, Kerry, Baruch, Shmuel, 2004. Information in securities markets: Kyle meets Glosten and Milgrom. Econometrica 72 (2), 433–465.
Back, Kerry, Pedersen, Hal, 1998. Long-lived information and intraday patterns. J. Financ. Mark. 1 (3–4), 385–402.
Back, Kerry, Cao, C. Henry, Willard, Gregory A., 2000. Imperfect competition among informed traders. J. Finance 55 (5), 2117–2155.
Back, Kerry, Cocquemas, Francois, Ekren, Ibrahim, Lioui, Abraham, 2020. Optimal transport and risk aversion in Kyle's model of informed trading. ArXiv preprint arXiv:2006.09518.
Barger, Weston, Donnelly, Ryan, 2021. Insider trading with temporary price impact. Int. J. Theor. Appl. Finance 24 (02), 2150006.
Bauer, Heinz, 1996. Probability Theory. De Gruyter Studies in Mathematics, vol. 23. Walter de Gruyter & Co., Berlin.
Billingsley, Patrick, 1995. Probability and Measure, 3rd edition. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York.
Campi, Luciano, Çetin, Umut, Danilova, Albina, 2011. Dynamic Markov bridges motivated by models of insider trading. Stoch. Process. Appl. 121 (3), 534–567.
Campi, Luciano, Çetin, Umut, Danilova, Albina, 2013. Equilibrium model with default and dynamic insider information. Finance Stoch. 17 (3), 565–585.
Carré, Sylvain, Collin-Dufresne, Pierre, Gabriel, Franck, 2022. Insider trading with penalties. J. Econ. Theory 203, 105461.
Çetin, Umut, 2018. Financial equilibrium with asymmetric information and random horizon. Finance Stoch. 22 (1), 97–126.
Çetin, Umut, Danilova, Albina, 2018. Dynamic Markov Bridges and Market Microstructure: Theory and Applications, vol. 90. Springer.
Çetin, Umut, Danilova, Albina, 2021. On pricing rules and optimal strategies in general kyle–back models. SIAM J. Control Optim. 59 (5), 3973–3998.
Cho, Kyung-Ha, 2003. Continuous auctions and insider trading: uniqueness and risk aversion. Finance Stoch. 7 (1), 47–71.
Daskalakis, Constantinos, Deckelbaum, Alan, Tzamos, Christos, 2013. Mechanism design via optimal transport. In: Proceedings of the Fourteenth ACM Conference on Electronic Commerce, pp. 269–286.
Daskalakis, Constantinos, Deckelbaum, Alan, Tzamos, Christos, 2017. Strong duality for a multiple-good monopolist. Econometrica 85 (3).
DeMarzo, Peter M., Fishman, Michael J., Hagerty, Kathleen M., 1998. The optimal enforcement of insider trading regulations. J. Polit. Econ. 106 (3), 602–632.
Fishe, Raymond PH, Robe, Michel A., 2004. The impact of illegal insider trading in dealer and specialist markets: evidence from a natural experiment. J. Financ. Econ. 71 (3), 461–488.
Fishman, Michael J., Hagerty, Kathleen M., 1992. Insider trading and the efficiency of stock prices. Rand J. Econ., 106–122.
Frino, Alex, Satchell, Stephen, Wong, Brad, Zheng, Hui, 2013. How much does an illegal insider trade? Int. Rev. Finance 13 (2), 241–263.
Galichon, Alfred, 2016. Optimal Transport Methods in Economics. Princeton University Press.
Kacperczyk, Marcin T., Pagnotta, Emiliano, 2019. Becker meets Kyle: inside insider trading. Available at SSRN: https://ssrn.com/abstract=3464447.
Kacperczyk, Marcin T., Pagnotta, Emiliano, 2023. Legal risk and insider trading. Available at SSRN: https://ssrn.com/abstract=3142006.
Karatzas, Ioannis, Shreve, Steven E., 1991. Brownian Motion and Stochastic Calculus, 2nd edition. Graduate Texts in Mathematics, vol. 113. Springer-Verlag, New York.
Kolotilin, Anton, Corrao, Roberto, Wolitzky, Alexander, 2025. Persuasion and matching: optimal productive transport. J. Polit. Econ. 133 (4), 1334–1381.
Kramkov, Dmitry, Xu, Yan, 2022. An optimal transport problem with backward martingale constraints motivated by insider trading. Ann. Appl. Probab. 32 (1), 294–326.
Kyle, Albert S., 1985. Continuous auctions and insider trading. Econometrica, 1315–1335.
Rochet, Jean-Charles, Vila, Jean-Luc, 1994. Insider trading without normality. Rev. Econ. Stud. 61 (1), 131–152.
Rüschendorf, Ludger, Thomsen, Wolfgang, 1993. Note on the Schrödinger equation and I-projections. Stat. Probab. Lett. 17 (5), 369–375.
Sharpe, Michael, 1988. General Theory of Markov Processes. Pure and Applied Mathematics, vol. 133. Academic Press, Inc., Boston, MA.
Shin, Jhinyoung, 1996. The optimal regulation of insider trading. J. Financ. Intermed. 5 (1), 49–73.