

WEALTH OR STEALTH? THE CAMOUFLAGE EFFECT IN INSIDER TRADING

Jin Ma^{*†}Weixuan Xia^{*‡}Jianfeng Zhang^{*§}

This Version: December 9, 2025

Abstract

We consider a Kyle-type model where insider trading takes place among a potentially large population of liquidity traders and is subject to legal penalties. Insiders exploit the liquidity provided by the trading masses to “camouflage” their actions and balance expected wealth with the necessary stealth to avoid detection. Under a diverse spectrum of prosecution schemes, we establish the existence of equilibria for arbitrary population sizes and a unique limiting equilibrium. A convergence analysis determines the scale of insider trading by a stealth index γ , revealing that the equilibrium can be closely approximated by a simple limit due to diminished price informativeness. Empirical aspects are derived from two calibration experiments using non-overlapping data sets spanning from 1980 to 2018, which underline the indispensable role of a large population in insider trading models with legal risk, along with important implications for the incidence of stealth trading and the deterrent effect of legal enforcement.

JEL CLASSIFICATIONS: C73; D82; G14

KEYWORDS: Insider trading; stealth trading; large population; civil and criminal penalties; equilibrium convergence; stealth index γ

1 Introduction

The main objective of the present paper is to conduct an in-depth analysis of the intrinsic link between an insider’s trade size choices and the number of liquidity traders (or noise traders) during concurrent trading episodes, in uncovering and quantifying the level of stealth that insiders exercise in illicit trading activity. The central methodology is based on a newly-constructed game-theoretic framework (following Kyle (1985)) for insider trading that adapts to a wide variety of detection mechanisms and penalty functions while incorporating the population size of liquidity traders. In this general setting, the assumption of normally distributed liquidity trades is well supported by the large size of the liquidity trading crowd and the independence among individual trades, and we shall show that an

^{*}Department of Mathematics, University of Southern California

[†]Email: jinma@usc.edu

[‡]Email: weixuanx@usc.edu

[§]Email: jianfenz@usc.edu

equilibrium exists for any finite population of liquidity traders – a result that is in itself highly non-trivial. As the population size tends to infinity, we also establish the uniqueness of the corresponding limiting equilibrium – a critical outcome of considering the entire trading crowd. The convergence properties lead to the discovery of a stealth index γ that sets up a rigorous yet simple link between insider trading and stealth trading. Such quantification sheds light upon the prevalence of an often overlooked moderate stealth level in illicit insider trading, revealing an incentive whose dual dependence on both regulatory scrutiny and potential legal penalties illuminates practical interconnections between market behavior and the legal enforcement landscape.

The canonical model of Kyle (1985) provides a foundational framework for understanding informed trading in financial markets by exploring how an insider (informed trader) with private information about an asset's value trades strategically to maximize profit. It was demonstrated that the insider's trades are gradual and proportional to market liquidity, balancing profitability with informational concealment,¹ and the equilibrium reached highlights a tradeoff between information efficiency and risks from market manipulation.

While the Kyle model does not explicitly address illicit insider trading – focusing instead on informational concealment for profit maximization rather than avoidance of legal penalties – many recent studies have advanced the framework to incorporate the modeling of legal risk associated with illicit trading for a comprehensive exploration of the price impact and strategic behavior influenced by the *dual* objectives of wealth expansion and penalty mitigation. For instance, Carré, Collin-Dufresne, and Gabriel (2022) explicitly extended Kyle's one-period model to a setting subject to legal penalties and analyzed insider trading regulations striking a balance between market liquidity and price informativeness, and Kacperczyk and Pagnotta (2024), using data sourced from the U.S. Securities and Exchange Commission (SEC) case files across 1995 to 2018, argued that (illegal) insiders adjust trading strategies as they internalize legal risk from regulators, and that legal enforcement is indeed effective in deterring insider trading aggression and containing price informativeness.²

A notable feature in much of the literature along these lines, whether or not legal risks are considered, is the treatment of non-insider trades collectively as a single entity. In essence, these trades can be thought of as being placed by a single (representative) liquidity (or noise) trader.³ Indeed, without legal risks, the number of liquidity traders is a relatively minor concern because the only potential harm from insider trading – from an insider's viewpoint – lies in reduced profits caused by their price impact; as a result, it is optimal to trade at the same scale as the representative liquidity trader, or equivalently, all liquidity traders combined. As legal risks arise, however, (illegal) insiders should suppress their trade sizes further, as trading at comparable levels to the entirety of liquidity traders is likely to expose them to legal repercussions. More precisely, it increases the likelihood of drawing regulatory attention and triggering red flags (see, e.g., Picardo (2022)). This raises the following question: *How do insiders exploit the presence of a large crowd of liquidity traders (representing normal trading activity) to moderate their trade sizes, thereby reducing detection risk while simultaneously*

¹Developments in continuous time further highlight a gradual reduction in the tendency to conceal information towards the time when the asset's value becomes publicly known; see Back (1992), Back and Baruch (2004), and Caldentey and Stacchetti (2010), among others.

²Additionally, Çetin (2025), adopting a continuous-time approach, found that legal risks can impede the optimality for insiders to bring prices to the true valuation at the end of the trading episode.

³For example, in the original Kyle model, liquidity (noise) trading quantities are assumed to be normally distributed; we also refer to Bagnoli, Viswanathan, and Holden (2001) and Boulatov, Kyle, and Livdan (2013) regarding the existence and uniqueness properties of a (linear) equilibrium.

mitigating price impact to maximize profits? In particular, with a sizeable population of liquidity traders, it can be reasonably suspected that in equilibrium, while insiders still internalize legal risk when devising their trading strategies, they no longer account for the risks through the profit channel, as their consideration of price impact diminishes under the deterrent effect of legal consequences.

Although intuitive and seemingly natural, the understanding that the number of liquidity traders should far exceed that of insider traders – on insider trading days – can be justified by various legal and societal factors. First, laws against insider trading create a high barrier for engaging in it, and criminal and civil penalties, including fines, disgorgement, and prison sentences, deter most traders from engaging in insider trading (see, e.g., Patel and Putnins (2021) Sect. 1 and Kacperczyk and Pagnotta (2024) Tab. I).⁴ In addition, insider trading is contingent on access to material, non-public information, which is generally restricted to a small group (such as executives and directors) within firms, and only a minority of individuals within companies, or those heavily tied, have access to such sensitive information. Present legal risk, the large pool of liquidity traders can effectively provide insiders with the “necessary” cover to engage in illicit trading, while simultaneously reducing the likelihood of immediate detection, and contrarily, absence of a substantial trading crowd would readily render such illegal activities apparent and swiftly prosecuted, e.g. due to the relatively low costs associated with investigations, including legal and data analysis efforts (Picardo (2022)).

In psychological terms, this phenomenon is commonly known as the “camouflage effect,” which describes situations where individuals blend into a crowd to avoid detection while engaging in wrongful or harmful acts; see, e.g., Griskevicius et al. (2006). In the insider trading context, this “simple” act of hiding was originally hinted at in the model of Kyle (1985), with the implication that when maximizing profit, (illicitly) informed traders tend to camouflage their information by splitting up their trades over time, while Admati and Pfeiderer (1988) argued that the same can also be achieved by purposefully engaging in trading amid high liquidity volume.

Noteworthily, the camouflage effect is closely associated with the well-studied concept of “stealth trading,” where informed traders strategically concentrate their trades in medium sizes. Intuitively, faced with legal risk, informed traders tend to execute volumes smaller than the total volume during periods of heavy trading activity to obscure their identities and reduce the likelihood of detection and prosecution. Meanwhile, driven by a clear profit motive, they are also inclined to trade in significantly larger volumes compared to the average liquidity trader. This phenomenon was early examined in Meulbroek (1992) in studying the price impact of informed trades using data sourced from the SEC case files over the 1980s. In particular, a notable finding is that the (daily) median ratio of the insider trading volume to the target firms’ total trading volume is about 11%. A later study by Del Guercio, Odders-White, and Ready (2017), utilizing SEC case files data from 2003 to 2011, has shown a significantly reduced price impact in contrast with the finding of Meulbroek (1992).⁵

This consideration also leads to what is known as the “stealth trading hypothesis” in the literature, originally proposed by Barclay and Warner (1993) in addressing the issue of informed traders’ choices of trade sizes. The hypothesis predicts that medium-size (precisely defined as 500 to 9,900 shares) trades come with disproportionately large cumulative price changes. Apart from their own empirical evidence from a sample of NYSE firms between 1981 and 1984, there have been numerous studies

⁴ According to recent reports, the SEC prosecuted a total of 583 enforcement actions in 2024, obtaining orders for \$8.2 billion in financial remedies. See <https://www.sec.gov/newsroom/press-releases/2024-186>.

⁵ This is ascribed to gradually increased prudence among insiders after 2000, likely in response to enhanced regulatory measures (such as increased regulatory budgets and the introduction of the SEC Whistleblower Program).

over the years to confirm the predictions of this hypothesis. To name a few, using audit trail data from NYSE firms, Chakravarty (2001) highlighted the disproportionately large role played by informed traders utilizing medium-sized trades in driving price movements. Anand and Chakravarty (2007) confirmed this preference for medium-sized trades in the options markets as well, particularly for high-leverage options. However, it is worth noting that this tendency toward moderate trade sizes is not consistently observed in non-U.S. markets, likely due to elevated levels of price manipulation associated with liquidity trading (see Cai, Cai, and Keasey (2006)). Another empirical study by Frino *et al.* (2013) shows that insiders' trade sizes are largely affected by both the probability of detection and expected penalty, and that insider volumes surpass liquidity trades at the individual level for a specific expected return while falling below the aggregate level.

A natural implication from the last line is that if detection of insider trading is triggered with abnormal order flow imbalances exceeding certain watermarks (see, e.g., DeMarzo, Fishman, and Hagerty (1998) and Kacperczyk and Pagnotta (2024) Sect. II), then it would be optimal for insiders to execute trades at medium levels where normal trades are conspicuous to reduce the likelihood of detection. Alexander and Peterson (2007) referred to such occurrence as trade-size clustering, documenting that NYSE and NASDAQ trades typically cluster around multiples of 500, 1,000, or 5,000 shares, associated with the greatest price impact, and the clustering strengthens with aggregate trading activity; see also Chen (2019) for similar considerations for cryptocurrency trading. On the other hand, using futures trading data, Chang, Pinegar, and Schachter (1997) showed that when large speculative trades are singled out, a significantly stronger price–volume relationship is observed; see also Blau (2017). The concentration of medium-sized trades has the econometric implication of enhanced sparsity in trade size distributions despite significant overall trading volumes, as modeled and demonstrated by Fei and Xia (2024). From different angles, these studies consistently support the presence of stealth trading, confirming that informed traders camouflage their activities by favoring medium-sized trades during periods of heavy market activity.

With the above considerations, we aim to develop a Kyle-type model (Section 2) with legal penalties (compare Carré, Collin-Dufresne, and Gabriel (2022) Sect. 2 and Çetin (2025) Sect. 2) that incorporates the liquidity trading population for a formal study of the camouflage effect in insider trading. We quantify the camouflage effect by introducing a measure termed “stealth index,” gauging the level of stealth (or caution) that insiders exercise in illicit trading (Section 3). Determination of this stealth index relies in large part on the convergence of the corresponding equilibria as the liquidity trading population grows, which deciphers how insiders' trade sizes vary with the number of liquidity traders in addition to the severity of legal penalties. We would like to highlight that the resulting limiting equilibrium can effectively reproduce a significantly reduced price impact which is adequate for practical purposes when justified by the presence of a large trading crowd.

It is worth noting that such a study would not be possible with only a scale parameter tracking the size of all liquidity trades (as adopted in Kyle (1985) Sect. 2). The reason is that with abnormal order flow imbalance-based detection of insider trading (DeMarzo, Fishman, and Hagerty (1998)), market liquidity is well factored by insiders, whereas regulatory investigations are conducted on a case-by-case basis, i.e., on the individual trader level rather than the volume of trades.⁶ Also, given that each liquidity trader places trades independently responding to idiosyncratic needs, the total liquidity order flow is in proportion to their average trade size but *not* their population size. This means that the

⁶What complicates the investigation process is not the total orders from liquidity traders, but their sheer number, thus giving insiders the opportunity to “camouflage.”

particular size of liquidity trades does not directly contribute to the camouflage effect, which requires a sufficiently large population to manifest.⁷

From a modeling perspective, the present paper also runs parallel to research on regulatory approaches to insider trading, including investigation schemes and the imposition of penalties. We consider a flexible detection mechanism (Section 2.1) that is adaptive to the detection mechanism postulated by DeMarzo, Fishman, and Hagerty (1998), who were among the first to consider the optimal design of insider trading regulation. Besides, our choice of the penalty function covers combinations of civil penalties and criminal penalties, with civil penalties determined based on insiders' illicit trading profit and criminal penalties depending primarily on their trading strategies. Aside from reflecting practical scenarios encompassing both civil and criminal cases (detailed discussion in Section 2.2), this consideration is broad enough to include most forms of penalty functions discussed in the literature, such as linear penalties (as in Carré, Collin-Dufresne, and Gabriel (2022) Sect. 3.4.2), quadratic penalties (as in Shin (1996), Carré, Collin-Dufresne, and Gabriel (2022) Sect. 3.4.1, and Çetin (2025)), and multiples of illicit profit (as in Kacperczyk and Pagnotta (2024) Sect. II).⁸ In Section 3, we shall also demonstrate that such a penalty combination is materially significant by both accounting for the nuances of realized illicit gains and assessing the pre-trade intent in determining the appropriate legal charges.

In a theoretical construct, this paper focuses on the following main contributions:

- Construct a Kyle-type model with legal risk incorporating a flexible detection mechanism as well as both civil and criminal penalties, in which the number of liquidity traders can be large (Section 2).
- Propose a stealth index to quantify the camouflage effect in insider trading, implying that in equilibrium, insiders prefer a trading intensity that lies strictly between that of an average liquidity trader and the combined activity of all liquidity traders (Section 3 and Section 6).
- Conduct a thorough convergence analysis to show that using the limiting equilibrium (with diminished price impact) as an approximate equilibrium offers great ease in computation while not altering the equilibrium implications (Section 3 and Section 6).

Our model is also amenable to calibration to market data on insider trading volumes, while also aligning with unresolved empirical findings from the literature. With the calibration experiments in Section 5 based on two data sets, we confirm insiders' internalization of legal risk from regulators as well as the effectiveness of legal enforcement in deterring insider trading, along with reduced price informativeness (Del Guercio, Odders-White, and Ready (2017) and Kacperczyk and Pagnotta (2024)). On the other hand, insiders adjust their trade sizes according to the size of liquidity trades (e.g., Barclay and Warner (1993) and Chakravarty (2001)) in attempting to disguise their trades for legal risk avoidance, to such extent that their price impact asymptotically vanishes in equilibrium. All main proofs are provided in Appendices A, B, and C, while the proofs of additional results and other auxiliary details are presented in Supplemental Appendices A and B.

⁷This perspective will be further clarified by the formal mathematical treatment in Section 2.

⁸Notably, while criminal penalties are naturally nondecreasing in the (insider) trading strategy, in the case of civil penalties, the expected penalty function is generally not symmetric or nondecreasing (see Section 4), and so it operates outside the framework of Carré, Collin-Dufresne, and Gabriel (2022) and Çetin (2025).

2 A Kyle model with legal risk

We begin with a description of our model framework. Following the standard one-period model of Kyle (1985), the market features three types of agents trading a single risky asset with fundamental value V , which is revealed, at time 1, to be 1 with probability $0 < p < 1$ and 0 otherwise.

There are a total of N **liquidity traders** who will trade the asset non-strategically for exogenous needs. At time 0, the liquidity traders have no prior access to the asset value V and submit independent orders with sizes being i.i.d. random variables with zero mean and variance σ^2 . Thus, with the central limit theorem in place, we assume that the total (net) order flow from liquidity trading can be written as $\sqrt{N}W$, where $W \stackrel{d}{=} \text{Normal}(0, \sigma^2)$ is independent of V . There is a single (risk-neutral) **insider trader** who observes the value V at time 0 at no cost and in a strategic manner submits his order $Z(V)$ as a function of V , knowing that placing the order is likely to introduce a contemporary price impact on the risky asset. At time 0, a competitive **market maker** sees the total (net) order flow $\sqrt{N}W + Z(V) = Y$ and executes orders at some price $P(Y)$ as a function of Y .⁹

One major distinction from the standard Kyle model (and its variations with legal risk) is in the **population size** of liquidity traders N ,¹⁰ which we expect to be “very large.” Beyond the practical reasons for a large number of liquidity traders discussed in Section 1, a modeling perspective also supports choosing a normal distribution for W . Indeed, if the population size were small, any specific distributional assumption (whether binomial, uniform, normal, etc.) would risk being overly restrictive; on the other hand, with a sufficiently large population size, independence among the liquidity traders sets the ground for the normal approximation.¹¹

2.1 Prosecution mechanism

The insider trader is also aware of legal risk that he faces in the presence of a **regulator**, who initiates investigations into his trading behavior by various means, such as detections of abnormal trading activity (DeMarzo, Fishman, and Hagerty (1998)), market surveillance associated with corporate events (Patel and Putninsk (2021)), or informant reports (such as the SEC Whistleblower Program). Once an investigation starts, prosecution of the insider will ensue depending on the actual trading behavior. A general description of the prosecution mechanism event could be based on an independent Bernoulli random variable B_N with strategy-dependent parameter, ρ_N ; that is, given his order $Z(V) = z$ placed at time 0, the insider will be charged with illicit trading with probability $\rho_N(z)$ at time 1, where the dependence on z is implied by the regulator gaining access to the insider’s trading account. More precisely, we consider this probability parameter, or the probability of successful prosecution, in the following form:

$$\rho_N(z) = 1 - e^{-\lambda_N(z)}, \quad z \in \mathbb{R}, \tag{2.1.1}$$

where $\lambda_N(z) \geq 0$ is interpreted as the **hazard rate** and is a *strictly increasing* function in the magnitude of z ; contrarily, $1 - \rho_N(z) = e^{-\lambda_N(z)}$ is the probability of no investigation or a dismissal from court

⁹While taking account of his own price impact, the insider cannot design his trading strategy based on the orders from the liquidity traders; in other words, the insider’s order is submitted no later than the realization of W .

¹⁰If the one period (from time 0 to time 1) represents a generic insider trading episode, N can also be regarded as the total number of liquidity trades, assuming that each trade is placed by a different (liquidity) trader.

¹¹We may use 30 as a practical threshold – following conventional statistical guidelines – to show that the population size is large enough for this approximation to take hold. In our convergence analysis later, however, a larger value of N may be required to ensure effectiveness of the convergence.

– or broadly, the corresponding “survival” probability. The structure (2.1.1) allows us to focus on the prosecution intensity, directly related to the likelihood, and it immediately implies that successful prosecution is never guaranteed, which is also practical given the complex circumstances of each insider trading case. Meanwhile, it is important to impose that $\lambda_N(0) = 0$, for, if the insider does not trade at all, then there is no room for prosecution, while the increasing feature of λ_N signifies that aggressive (insider) trading strategies necessarily result in a heightened probability of prosecution.

It is also crucial that the hazard rate (and thus the prosecution probability) explicitly depends on the population size N , which is to incorporate the idea that the probability of identifying an insider trader should decrease as the total number of normal traders increases. Indeed, a large number of liquidity traders can create an elevated noise level, making it harder to distinguish abnormal trading patterns from normal trading activity. Since the insider’s trades are usually detected as they represent outliers in terms of order flow imbalances (DeMarzo, Fishman, and Hagerty (1998)), with an increasing number of liquidity trades, the threshold for what counts as a statistical outlier also rises, in turn making it more challenging to detect his trades. The insider can then “camouflage” more easily, with his suspicious behavior becoming less conspicuous. To examine the scale effect of N on the insider’s strategies, we impose a general power-type structure,

$$\lambda_N(z) = \lambda(N^{-\beta}z),$$

where $\lambda \equiv \lambda_1$ is a *size-modulated* hazard rate and $\beta \geq 0$ is a power coefficient governing the scale.

As an important example, the probability of prosecution may be conceptualized as arising from the sequential processes of abnormal order flow imbalance-triggered investigation and subsequent successful legal actions, i.e.,

$$\rho_N(z) = D(N^{-\frac{1}{2}}z)\mathbb{P}\{|W + N^{-\frac{1}{2}}z| \geq \bar{y}\}, \quad (2.1.2)$$

where $D : \mathbb{R} \mapsto [0, 1]$ is a size-modulated function satisfying that $D(0) = 0$, representing the probability of prosecution conditional on investigation, and $\bar{y} > 0$ is a custom detection threshold. Then, we have the following functional form for the hazard rate:

$$\begin{aligned} \lambda_N(z) &= -\log(1 - \rho_N(z)) \\ &= -\log\left(1 - \frac{D(N^{-\frac{1}{2}}z)}{2}\left(\operatorname{erfc}\left(\frac{\bar{y} + N^{-\frac{1}{2}}z}{\sqrt{2\sigma^2}}\right) + \operatorname{erfc}\left(\frac{\bar{y} - N^{-\frac{1}{2}}z}{\sqrt{2\sigma^2}}\right)\right)\right), \end{aligned} \quad (2.1.3)$$

where $\operatorname{erfc}(\cdot)$ is the complementary Gauss error function; in this case, $\beta = 1/2$ exactly. Detailed derivations are found in Supplemental Appendix B. In Kacperczyk and Pagnotta (2024), the prosecution probability is precisely ρ_1 from (2.1.2), except with $D \in [0, 1]$ being a constant probability.

In general, the power coefficient β allows us to consider a flexible dependence structure of the prosecution probability with respect to the aggregate noise level, which is crucial to examining its impact on the behavior of the insider’s strategy Z when the population size becomes large.

2.2 Composition of legal penalties

Once successfully prosecuted, the insider trader is subject to legal penalties depending on the jurisdiction and the specific circumstances of the case. First, a common component of insider trading

penalties is *disgorgement* of illicit profits, which involves the return of any profits gained from the illicit trading activity, ensuring unprofitability. The amount of disgorgement is usually calculated based on the total profits earned by the insider, equaling $(V - P)Z(V)$ upon revelation of the asset value.

Another notable aspect is the adoption of so-called “penalty multipliers,” which are imposed in some cases as an enhancement of the deterrent effect of the penalties, based on factors such as the egregiousness of the violation or the level of harm caused to investors or the market. For example, in the U.S., the maximum civil penalty for insider trading is up to three times the profit gained or loss avoided, a.k.a. the “treble damages” provision under THE INSIDER TRADING AND SECURITIES FRAUD ENFORCEMENT ACT OF 1988 (ITSFEA). Similarly, as amended by the ITSFEA, depending on evidence of the insider’s level of intent and trading behavior, significant penalties can be triggered in criminal cases to deter future misconduct; this includes prison sentences and steep fines – e.g., up to 20 years and up to \$5 million for individuals in the U.S.

With these factors in mind, we believe that a comprehensive model for legal penalties should allow for explicit dependence on *both* the insider trader’s illicit profit (as a civil penalty) and his trading strategy (as a criminal penalty). Thus, it is desirable that a penalty function should be a bivariate function of the insider’s trading strategy (z) as well as his profit gained $((v - P(y))z)$, taking the general form $C(z, (v - P(y))z)$. In the present paper, we focus on the below composition:

$$C(z, (v - P(y))z) = C_0(z) + \chi((v - P(y))z)^+, \quad v \in \{0, 1\}, \quad (2.2.1)$$

where $(\cdot)^+$ denotes the positive part. On the right side of (2.2.1), the first component with C_0 incorporates **criminal penalties**, including imprisonment and fines,¹² that directly depends on the insider’s trading strategy, while the second stands for **civil penalties** (with regard to the profit) controlled by a **penalty multiplier** χ . For a meaningful consideration, we require $C_0(z)$ to be increasing (not necessarily strictly) in the magnitude of z , and it makes sense to force $C_0(0) = 0$ in the case of no trading. Additionally, we require that $\chi \geq 1$, which implies disgorgement of the insider’s illicit profit following successful prosecution.¹³ Clearly, when $\chi = 1$, only C_0 will remain on top of disgorgement, and penalization is largely strategy-based, while by taking $C_0 \equiv 0$, the second component dominates, leading to profit-based (civil) penalties.

2.3 Equilibrium definition

With the aforementioned legal risk specifications, we now formulate the insider trader’s optimization problem and give a formal definition of the equilibrium, for any population size N of liquidity traders.

Accessing the value V at time 0, the insider trader takes the price function P from the market maker as given and designs a trading strategy to maximize his expected net profit. His objective function is given by

$$J_N(P; z, v) := \mathbb{E}[(v - P(\sqrt{N}W + z))z - \rho_N(z)C(z, (v - P(\sqrt{N}W + z))z)], \quad (2.3.1)$$

for $z \in \mathbb{R}$, $v \in \{0, 1\}$, which he seeks to maximize over z for each value of v , namely $z = Z(v)$. At the same time, given the insider’s trading strategy $Z = (Z(0), Z(1))$, the market maker observes the

¹²In the case of imprisonment, C_0 can be considered equivalent to a capital penalty due to its severe, lifelong impact on one’s economic, social, and personal “capital.” Such an impact could manifest as opportunity costs, social and economic disconnection, lasting stigma, etc.

¹³This feature is also in keeping with the institutional parametric condition made in Kacperczyk and Pagnotta (2024) Sect. II, ensuring a penalty no less than the profit gained.

total order flow Y and sets a rational price function for break-even,¹⁴ namely setting $P(y) = P_N(Z; y)$, where

$$P_N(Z; y) := \mathbb{E}[V | \sqrt{N}W + Z(V) = y], \quad y \in \mathbb{R}. \quad (2.3.2)$$

Since V only takes values 0 and 1, by applying Bayes' rule, we easily obtain

$$\begin{aligned} P_N(Z; y) &= \mathbb{P}\{V = 1 | \sqrt{N}W + Z(V) = y\} = \frac{pe^{-\frac{(y-Z(1))^2}{2N\sigma^2}}}{pe^{-\frac{(y-Z(1))^2}{2N\sigma^2}} + (1-p)e^{-\frac{(y-Z(0))^2}{2N\sigma^2}}} \\ &= \frac{1}{1 + qe^{\frac{(2y-Z(0)-Z(1))(Z(0)-Z(1))}{2N\sigma^2}}}, \quad \text{with } q := \frac{1-p}{p}. \end{aligned} \quad (2.3.3)$$

With the population size N fixed, the (original) hazard rate λ_N and the liquidity order flow $\sqrt{N}W \stackrel{d}{=} \text{Normal}(0, N\sigma^2)$ are both well-defined and finite, and the notion of equilibrium can be stated as in the standard Kyle model setting.

Definition 2.1. For any fixed $N \geq 1$, an equilibrium is a couple (Z_N^*, P_N^*) of trading strategy and price function such that:

(i) Given P_N^* , the insider trader maximizes his expected net profit from trade (2.3.1), i.e.,

$$J_N(P_N^*; Z_N^*(v), v) = \sup_{z \in \mathbb{R}} J_N(P_N^*; z, v), \quad v \in \{0, 1\}. \quad (2.3.4)$$

(ii) Given Z_N^* , the market maker sets a rational price function for break-even, namely

$$P_N^*(y) = P_N(Z_N^*; y), \quad y \in \mathbb{R}.$$

According to condition (ii) above, for equilibrium analysis it is useful to rewrite the insider's objective function as

$$\hat{J}_N(Z; z, v) := J_N(P_N(Z; \cdot); z, v),$$

where Z can be viewed as the market maker's target strategy. Since V is supported on $\{0, 1\}$, from (2.3.2) it is clear that $0 < P_N < 1$. Then, in light of (2.2.1) and (2.3.1), regarding the profit maximization (2.3.4), the insider would rationally concentrate on sell strategies with $z \leq 0$ when $v = 0$ and buy strategies with $z \geq 0$ when $v = 1$; indeed, by abstaining from trading entirely ($z = 0$), the investor guarantees a nonnegative net profit. However, a zero trading strategy is also trivial on account of legal risk and is demonstrably suboptimal for the insider.¹⁵ With this consideration in mind, we may assume without loss of generality that the insider's strategy satisfies the constraint $Z = (Z(0), Z(1)) \in (-\infty, 0) \times (0, \infty)$. As a result, the objective (2.3.4) is equivalent to

$$\hat{J}_N(Z_N^*; Z_N^*(v), v) = \sup_{z \in \mathbb{R}_v} \hat{J}_N(Z_N^*; z, v), \quad \text{with } \mathbb{R}_v := \begin{cases} (-\infty, 0), & \text{if } v = 0, \\ (0, \infty), & \text{if } v = 1. \end{cases} \quad (2.3.5)$$

¹⁴This can be justified by a Bertrand competition argument; see e.g. Kyle (1985). An alternative interpretation is that the market maker chooses P to minimize the squared error $\mathbb{E}[(V - P(\sqrt{N}W + Z(V)))^2]$.

¹⁵Under the key Assumption 4.1, it can be shown that a zero strategy cannot be optimal in equilibrium; see Step 1 in the proof of Theorem 4.1 in Appendix B.

Besides, with $v \in \{0, 1\}$ and $0 < P_N < 1$, we have in (2.2.1) that

$$((v - P(y))z)^+ = (v - P(y))z \text{ for all } z \in \mathbb{R}_v, \quad \text{and } P = P_N(Z; \cdot). \quad (2.3.6)$$

Then, recalling (2.3.1) and (2.2.1), we define the insider's expected price, expected (gross) profit, and expected additional penalties respectively as

$$\begin{aligned} \Phi_N(Z; z) &:= \mathbb{E}[P_N(Z; \sqrt{N}W + z)], \quad Q_N(Z; z, v) := (v - \Phi_N(Z; z))z, \\ \Psi_N(Z; z, v) &:= C_0(z) + (\chi - 1)Q_N(Z; z, v), \end{aligned} \quad (2.3.7)$$

which allow us to recast the objective function (2.3.1) into the equivalent form

$$\tilde{J}_N(Z; z, v) = e^{-\lambda_N(z)}Q_N(Z; z, v) - (1 - e^{-\lambda_N(z)})\Psi_N(Z; z, v), \quad z \in \mathbb{R}_v. \quad (2.3.8)$$

In particular, (2.3.3) shows that the function Φ_N is precisely the expected value of a Gaussian random variable under a sigmoid-type transformation.¹⁶

3 Limiting equilibrium and convergence

In exercising stealth trading, the insider trader has the tendency to trade in larger quantities compared to a single liquidity trader, while his trade size cannot be “too large” at the same time in view of the risk of detection and prosecution. We again refer to Frino et al. (2013) Sect. V.B, Meulbroek (1992) Sect. E and Chakravarty (2001) for related empirical evidence. This suggests that an insider's optimal trading strategy should, ideally, scale with the total population size to adequately capture its asymptotic behavior.

From a mathematical viewpoint, in the limit as N goes to infinity, the optimal trading strategy Z_N^* from Definition 2.1 can well be infinity. Since we consider power-law growth in the hazard rate, it is reasonable to introduce the scaled strategy $\tilde{Z}_N^* = N^{-\gamma}Z_N^*$ for some (yet-to-be-determined) coefficient $\gamma \geq 0$, subject to the requirement that the limiting scaled strategy $\tilde{Z}_\gamma^* := \lim_{N \rightarrow \infty} \tilde{Z}_N^*$ exist with nonzero values in $\mathbb{R}_0 \times \mathbb{R}_1$. In other words, we suspect the insider's (original) strategy to have the property that $Z_N^* \sim N^\gamma \tilde{Z}_\gamma^*$, i.e., $\lim_{N \rightarrow \infty} Z_N^*(v)/(N^\gamma \tilde{Z}_\gamma^*(v)) = 1$ for $v \in \{0, 1\}$, as $N \rightarrow \infty$. The coefficient γ bears direct connections to the camouflage effect as it directly reflects the insider's desire to balance his wealth expected from illicit trading and his level of stealth to avoid detection and prosecution. A smaller value is associated with a smaller chance of detection, or equivalently, a higher stealth level. In particular, we formally name it the **stealth index** and, considering the aforementioned moderateness of insider trade sizes, a desirable – but yet-to-be-verified – value range for this index will be $(0, 1/2)$.¹⁷ Nevertheless, in what follows we shall still consider the general value range $[0, \infty) \ni \gamma$.

To formalize the notion of a limiting equilibrium based on the scaled trading strategy, $\tilde{Z} = N^{-\gamma}Z$, we need to consider the limit of the price function (2.3.3) as $N \rightarrow \infty$. Since for z at the scale of N^γ , $y = \sqrt{N}W + z$ is at the scale of $N^{\max\{\gamma, 1/2\}}$, we can introduce the scaled order flow $\tilde{y} = N^{-\max\{\gamma, 1/2\}}y$ and re-parameterize the equilibrium price function in (2.3.2) and (2.3.3) as

$$\tilde{P}_N^\gamma(\tilde{Z}; \tilde{y}) = P_N(N^\gamma \tilde{Z}; N^{\max\{\gamma, 1/2\}}\tilde{y}), \quad \tilde{Z} = (\tilde{Z}(0), \tilde{Z}(1)) \in \mathbb{R}_0 \times \mathbb{R}_1, \quad \tilde{y} \in \mathbb{R}. \quad (3.1)$$

¹⁶Despite no known closed form, it can be expressed in terms of an infinite series containing Gaussian moments and explicitly computable coefficients by adopting an expansion argument.

¹⁷Clearly, $\gamma = 0$ would imply that the insider's trade size is comparable to that of a single liquidity trader, while if $\gamma = 1/2$, he would trade at the same scale as all the liquidity traders combined.

Throughout the paper, we use the tilde notation ($\tilde{\cdot}$) to indicate the scaled quantities. The next proposition shows that the corresponding limit generally depends on γ .

Proposition 3.1. *For \tilde{P}_N^γ as in (3.1), $\tilde{Z} \in \mathbb{R}_0 \times \mathbb{R}_1$, and $\tilde{y} \in \mathbb{R}$, it holds that*

$$\lim_{N \rightarrow \infty} \tilde{P}_N^\gamma(\tilde{Z}; \tilde{y}) = \tilde{P}_\infty^\gamma(\tilde{Z}; \tilde{y}) := \begin{cases} p, & \text{if } \gamma \in [0, \frac{1}{2}), \\ P_1(\tilde{Z}; \tilde{y}), & \text{if } \gamma = \frac{1}{2}, \\ 1_{\{2\tilde{y} > \tilde{Z}(0) + \tilde{Z}(1)\}} + p 1_{\{2\tilde{y} = \tilde{Z}(0) + \tilde{Z}(1)\}}, & \text{if } \gamma > \frac{1}{2}. \end{cases} \quad (3.2)$$

Proposition 3.1 demonstrates that for the interesting range $\gamma \in (0, 1/2)$ of the stealth index, the limiting price function no longer depends on the insider's trading strategy and is, in particular, constant, equal to the expected value of V . In the case $\gamma = 1/2$, as the insider trades at the scale of all liquidity trades, the limit is, unsurprisingly, invariant to the population size N , which can be normalized to 1, as considered in the literature. The case $\gamma > 1/2$ points to an audacious insider trading strategy, with volumes far exceeding typical aggregate levels, and the asset price is set in the usual way (equal to p) when the market maker perceives no insider involvement, and is otherwise adjusted to the extremal values (0 or 1), making detection and prosecution nearly inevitable (no stealth). This observation suggests $[0, 1/2] \ni \gamma$ as a plausible range and is to be verified through equilibrium analysis.

The price limits in Proposition 3.1 allow to examine the corresponding limiting behaviors of the expected price, profit, and penalties, Φ_N , Q_N , and Ψ_N in equilibrium. However, simply analyzing these limits appears to offer little insight into the correct choice of the stealth index γ from the insider's perspective. Instead, we shall re-scale the insider's objective function (or expected net profit): Given a price function \tilde{P} on the scaled order flow \tilde{y} , we consider the following limiting scaled objective function:

$$\tilde{J}_\infty^\gamma(\tilde{P}; \tilde{z}, v) := \lim_{N \rightarrow \infty} N^{-\gamma} J_N(P_N^\gamma; N^\gamma \tilde{z}, v), \quad \tilde{z} \in \mathbb{R}_v, \quad v \in \{0, 1\}, \quad (3.3)$$

where J_N is the original objective function in (2.3.8), and $P_N^\gamma(y) := \tilde{P}(N^{-\max\{\gamma, 1/2\}}y)$ is the recovered price function. On the right side of (3.3), J_N is scaled exactly by $N^{-\gamma}$ because based on (2.3.8), when $P_N^\gamma(y) = P_N(N^\gamma \tilde{Z}; y)$, the expected price Φ_N is bounded and the expected additional penalty Ψ_N must be controlled not to blow up either. Hence, the limiting equilibrium paired with Definition 2.1 can be defined with γ chosen such that the limiting objective function (3.3) is well-defined, and to ensure that the effect of γ is meaningful, it is important to restrict attention to nonzero limiting strategies as well as nonzero limiting objective function values.

Definition 3.1. *A γ -limiting equilibrium is a couple $(\tilde{Z}_\gamma^*, \tilde{P}_\gamma^*)$ of scaled trading strategy and price function such that:*

(i) *The insider trader maximizes his limiting scaled expected net profit (3.3), i.e.,*

$$\tilde{J}_\infty^\gamma(\tilde{P}_\gamma^*; \tilde{Z}_\gamma^*(v), v) = \sup_{\tilde{z} \in \mathbb{R}_v} \tilde{J}_\infty^\gamma(\tilde{P}_\gamma^*; \tilde{z}, v) \neq 0, \quad \tilde{Z}_\gamma^*(v) \neq 0, \quad v \in \{0, 1\}.$$

(ii) *The market maker sets a rational price function according to (3.2), $\tilde{P}_\gamma^* = \tilde{P}_\infty^\gamma(\tilde{Z}_\gamma^*; \cdot)$.*

On paper, from the objective function forms (2.3.8) and (3.3), a sophisticated choice of the stealth index γ should exhibit reliance on the growth rate of the size-modulated hazard rate λ , controlled

by $\beta \geq 0$, as well as that of the criminal penalty component C_0 , with respect to the insider's trading strategy.¹⁸ Intuitively, with light-to-moderate criminal penalties, the expected price Φ_N tends to overshadow potential legal risk, prompting the insider to trade at larger scales and worry less about being detected, thereby using a larger γ , while severe criminal penalties are likely to place a deterrent effect on the trading behavior and yields a relatively small γ . Condition (ii) in Definition 3.1 also reveals that if the insider trader exercises stealth trading with $\gamma < 1/2$, then in the limit of the population size N , the equilibrium price function becomes constant and *decoupled* from the trading strategy due to diminished price informativeness (see again Chakravarty (2001) and Kacperczyk and Pagnotta (2024)). This decoupling feature significantly simplifies subsequent equilibrium analysis.

The above suggests that the exact structure of the penalty function C plays an important role in determining the correct stealth index with an increasing population size of liquidity traders. To provide a thorough understanding of convergence towards the limiting equilibria, instead of providing a balanced treatment of all penalty types, the following analysis concentrates on the predominant scenario of civil penalties, which represents the most common enforcement outcome in insider trading cases and matches the empirical nature of the data sets to be presented in Section 5. The examination of scenarios involving criminal and mixed penalties is reserved for further discussions in Section 6, also helping align with various cases considered in the literature (Carré, Collin-Dufresne, and Gabriel (2022), Kacperczyk and Pagnotta (2024), and Çetin (2025)), apart from highlighting the technical differences.

4 Predominant scenario: Civil penalties

In the predominant scenario, upon successful prosecution, the insider trader faces civil penalties only (as considered in Kacperczyk and Pagnotta (2024) Sect. 2). The penalties are imposed on his illicit profit $((V - P_N)Z(V))$ through the penalty multiplier $\chi \geq 1$. In the absence of criminal charges, we shall take $C_0 = 0$ in (2.2.1).

Given a population size $N \geq 1$, the equilibrium objective function (2.3.8) reduces to

$$\begin{aligned}\hat{J}_N(Z; z, v) &= e^{-\lambda_N(z)} Q_N(Z; z, v) - (1 - e^{-\lambda_N(z)}) \Psi_N(Z; z, v) \\ &= e^{-\lambda_N(z)} Q_N(Z; z, v) - \chi_0(1 - e^{-\lambda_N(z)}) Q_N(Z; z, v), \quad z \in \mathbb{R}_v,\end{aligned}\tag{4.1}$$

where $\chi_0 := \chi - 1 \geq 0$ is the *disgorgement-adjusted* penalty multiplier. Based on the expression of Q_N in (2.3.7), a notable feature of the civil penalties is their explicit dependence on the asset value $v \in \{0, 1\}$, and the insider designs his trading strategies knowing that the total order flow ($Y = \sqrt{NW} + Z(V)$) observed by the market maker will automatically absorb uncertainty embedded in such strategies.

On a closer look, the expected (additional) penalty Ψ_N in (4.1) can increase at most linearly with the insider's trade size ($|z|$), since the expected price Φ_N is bounded, and more importantly, it is not a monotone function of z ,¹⁹ which arises from a reduction in price informativeness that a significantly larger number of trades placed by the insider inevitably shifts the total order flow towards revelation

¹⁸On the other hand, note that from (2.2.1), the civil penalty component is always of linear growth as $|z| \rightarrow \infty$ due to the boundedness of the price function P_N .

¹⁹This non-monotonicity property is what makes the structure of civil penalties inherently different from that of criminal penalties – hence not regardable as a special instance of the latter; see Section 6.

of the fundamental asset value (V), thereby pushing the market maker to adjust the price closer to this true value. This adjustment can then substantially reduce or even wipe out the insider's profit, leading to a smaller penalty (in addition to disgorgement). From another viewpoint, considering that civil penalties are typically less severe than criminal penalties (reserved for serious violations), if only civil penalties are in force, the insider is likely to trade large amounts, knowing that the penalty might decrease as the profits diminish.²⁰ For this reason, sole reliance on profit-based civil penalties may seem "overly idealistic" in that it fails to adequately address and penalize extremely violent insider trading actions, as the absence of significant illicit profit should not absolve the insider of his malicious behavior from the outset. This key observation further motivates a more general consideration combining civil and criminal penalties, to be addressed in Section 6.

To precisely determine the stealth index γ , we make the following technical assumption on the (size-modulated) hazard rate, which is also essential for establishing existence and uniqueness.

Assumption 4.1. *For $v \in \{0, 1\}$, λ is continuously differentiable and convex (not necessarily strictly) on \mathbb{R}_v , with $\lambda' < 0$ on $(-\infty, 0)$ and $\lambda' > 0$ on $(0, \infty)$.*

The requirement that λ be convex (apart from its strict increase) is nonrestrictive, which signifies a growing inclination to detect and prosecute the insider trader as his trade size increases. This is, for example, the case for the order flow imbalance-based detection mechanism (DeMarzo, Fishman, and Hagerty (1998)): For (2.1.3) one can show that whenever $D(z) = K_D|z|^{\theta_D}$ for some $K_D > 0$ and $\theta_D \geq 1$, then λ' is guaranteed to be increasing, and $|\lambda'(z)| \geq K'_\theta|z|^{\theta_D}$ for some $K'_\theta > 0$. Details are given in Supplemental Appendix B.

The following theorem, serving as the main result of this section, details the existence and uniqueness of the finite- N equilibrium and the limiting equilibrium, along with the desired convergence.

Theorem 4.1. *Consider the setting of (4.1), and let Assumption 4.1 hold true. Then, we have the following three assertions.*

- (i) *For every $N \geq 1$, there exists an equilibrium (Z_N^*, P_N^*) in the sense of Definition 2.1.*
- (ii) *There exists a γ -limiting equilibrium in the sense of Definition 3.1 if and only if*

$$\gamma = \min \left\{ \beta, \frac{1}{2} \right\}. \quad (4.2)$$

Moreover, when $\gamma < 1/2$, the γ -limiting equilibrium $(\tilde{Z}_\gamma^*, \tilde{P}_\gamma^*)$ is unique with $\tilde{P}_\gamma^* = p$.

(iii) *Let $\gamma < 1/2$ and $(\tilde{Z}_\gamma^*, \tilde{P}_\gamma^* \equiv p)$ be the unique γ -limiting equilibrium from assertion (ii). Then, there exists a constant $K > 0$, depending only on the model parameters but not on N , such that, for any equilibrium (Z_N^*, P_N^*) from assertion (i),*

$$|N^{-\gamma} Z_N^*(v) - \tilde{Z}_\gamma^*(v)| \leq KN^{2\gamma-1}, \quad v \in \{0, 1\}, \quad (4.3)$$

and, with $\tilde{P}_\gamma^* = p$,

$$\begin{aligned} |P_N^*(y) - p| &\leq K(|y|N^{\gamma-1} + N^{2\gamma-1}), \quad y \in \mathbb{R}, \\ |P_N^*(\sqrt{N}W + Z_N^*(V)) - p| &\leq K(|W|N^{\gamma-1/2} + N^{2\gamma-1}) \leq K(|W| + 1)N^{\gamma-\frac{1}{2}}, \quad \mathbb{P}\text{-a.s.} \end{aligned} \quad (4.4)$$

²⁰From the proof of Theorem 4.1 we shall see that in the present scenario with only civil penalties, for every $N \geq 1$ and both $v \in \{0, 1\}$, $\Psi_N(Z; z, v) \rightarrow 0$ as $|z| \rightarrow \infty$, meaning that the insider can technically reduce the penalty to none by increasing his trade size indefinitely.

An immediate implication from the first two assertions is that Assumption 4.1 provides sufficient conditions to ensure the existence of a finite- N equilibrium and a limiting equilibrium, the latter also being unique when $\gamma < 1/2$ – as is the interesting case with stealth trading. Technically, analyzing the finite- N equilibrium is significantly more involved than the limiting equilibrium and relies on a key lemma (Lemma B.1 in Appendix B), which elucidates a fundamental property of the function governing the insider’s expected (gross) profit.²¹ Besides, according to (4.2), the stealth index γ for the trading strategy is capped at 1/2, meaning that the insider deliberately refrains from overly aggressive trading – at quantities exceeding the total orders – inevitably exposing himself.

Assertion (iii) from Theorem 4.1 expounds the convergence of any finite- N equilibrium towards the (unique) γ -limiting equilibrium, in terms of both the insider’s optimal strategy and the market maker’s price function, on the assumption that $\gamma < 1/2$. For the price function, the rate of convergence, identified as the inverse power of N , is $\gamma - 1/2 < 0$, proportionate to the stealth index γ . The intuition behind this is clear: As $\gamma \searrow 0$, the insider’s trade size, scaled to match that of a single liquidity trader, becomes negligible within the total order flow, leaving the population size N as the dominant factor; conversely, as $\gamma \nearrow 1/2$, the trade size effectively matches the entire population of liquidity traders, nullifying the size impact, and convergence is out of question.

The convergence rate $(2\gamma - 1)$ for the insider’s optimal strategies is proportionally related to how the population size obscures investigation and prosecution, as measured by the coefficient β , and stems directly from the convergence of the price function. An important regulatory insight from this result is that the intensity of investigation (linked to β) can significantly impact how closely insiders’ trade sizes approach their limit associated with diminished price impact. More specifically, when investigations are highly effective ($\beta = 0$) – nearly unaffected by the trading population – informed traders’ strategies tend to converge rapidly for any given stealth level and are driven by the price function’s behavior.

Pertaining to Definition 3.1, Theorem 4.1 further demonstrates that if γ happens to be less than 1/2, the limiting equilibrium price function becomes constant, decoupled from insider trading information, hence making itself much easier to analyze than that with a finite population, which is even explicitly solvable in specific settings. We give the following illustrative example, which considers a quadratic hazard rate while adhering to the “treble damages” provision. The proof is given in Supplemental Appendix B.

Example 4.1. Let $p \in (0, 1)$ be arbitrary, and let $\lambda(z) = z^2$ (satisfying Assumption 4.1) and $\chi = 3$. Suppose that $\gamma = \beta \in [0, 1/2]$. Then, the γ -limiting equilibrium is uniquely determined as

$$\tilde{Z}_\gamma^* = \left(-\sqrt{\frac{1}{2} - W_0(\frac{\sqrt{e}}{3})}, \sqrt{\frac{1}{2} - W_0(\frac{\sqrt{e}}{3})} \right) \approx (-0.350753, 0.350753), \quad \tilde{P}_\gamma^* = p, \quad (4.5)$$

where $W_0(\cdot)$ denotes the Lambert W function (a.k.a. the product logarithm).

An interesting observation in Example 4.1 is that regardless of the probability p of revealing the asset value $V = 1$, at equilibrium, the insider sticks to trading the same amount in both states ($|\tilde{Z}_\gamma^*(0)| = \tilde{Z}_\gamma^*(1)$). An explanation is that as the civil penalties are tied to the insider’s illicit profit, they can effectively remove his incentive to compare and exploit the price differences in the two states, even though the contingent expected profits may still be different.

²¹Although this property inherently follows from the normal distribution of total liquidity order flow, it is in fact shared by a broad class of log-concave distributions (including many infinitely divisible distributions); see Saumard and Wellner (2014) and Yamazato (1978).

Following Theorem 4.1, the approximation power of the limiting equilibrium towards any finite-population equilibrium can be further illuminated with the concept of ϵ -equilibria, connected to Definition 3.1, which we show in Definition 4.1.²²

Definition 4.1. *For any fixed $N \geq 1$ and $\epsilon \geq 0$, an ϵ -equilibrium is a couple $(Z_N^{*,\epsilon}, P_N^{*,\epsilon})$ of trading strategy and price function such that, based on (2.3.3), (3.1), and (3.3),*

$$N^{-\gamma} J_N(P_N^{*,\epsilon}; Z_N^{*,\epsilon}(v), v) \geq \sup_{z \in \mathbb{R}_v} N^{-\gamma} J_N(P_N^{*,\epsilon}; z, v) - \epsilon, \quad v \in \{0, 1\}, \quad (4.6)$$

and

$$|P_N^{*,\epsilon}(N^{\max\{\gamma, \frac{1}{2}\}} \tilde{y}) - P_N(Z_N^{*,\epsilon}; N^{\max\{\gamma, \frac{1}{2}\}} \tilde{y})| \leq \epsilon(1 + |\tilde{y}|), \quad \tilde{y} \in \mathbb{R}. \quad (4.7)$$

We have the next important result about the approximation of any finite-population equilibrium via an arbitrary limiting equilibrium, with no uniqueness required.

Theorem 4.2. *Consider the setting of Theorem 4.1 assertion (iii), with $\gamma < 1/2$ and (\tilde{Z}_γ^*, p) being the unique γ -limiting equilibrium. Then, for every $N \geq 1$, $(N^\gamma \tilde{Z}_\gamma^*, p)$ is an ϵ_N -equilibrium in the sense of Definition 4.1, where for some constant K depending only on the model parameters,*

$$\epsilon_N = KN^{\gamma - \frac{1}{2}}. \quad (4.8)$$

Theorem 4.2 asserts that any limiting equilibrium defined under Definition 3.1 automatically qualifies as an ϵ_N -equilibrium within the actual, finitely populated market, for which the magnitude of ϵ_N , as a function of the population size N , carries the same economic implications in terms of the stealth index γ as the equilibrium strategies. The statement highlights the central idea that the limiting equilibrium is a practical and robust approximation across a wide range of detection schemes, whenever the population of liquidity traders is justifiably large enough.

5 Empirical perspectives

In this section, we address the issue of estimating the population size of liquidity traders, N , and the stealth index γ , using available data on insider trading cases. The main idea of this empirical analysis, as mentioned in Section 1, is to verify that the population of liquidity traders is, as expected, considerably large for any generic risky asset and – more importantly – that insiders favor trading with medium intensity (compared to all liquidity traders present in the same trading episodes) in the presence of legal risk by adopting a moderate stealth level, therefore confirming the deterrent effect of legal risk on insiders' trade size choices, associated with diminished price informativeness (as implied from the convergence properties in Section 3); see, again, Barclay and Warner (1993), Chakravarty (2001), and Kacperczyk and Pagnotta (2024). We conduct the analysis with two calibration experiments, which rely on different calibration conditions and yet confirm the same phenomena of interest.

²²Generally speaking, the convergence of non-unique equilibria in stronger senses (e.g., with respect to the Hausdorff distance) cannot be established without imposing additional technical conditions. A thorough exploration of this convergence issue in the insider trading context is left for further research.

5.1 Calibration experiment I

Our first calibration experiment makes use of insider trading volume data collected and analyzed in one of the earliest financial studies of its kind, conducted by Meulbroek (1992). Specifically, the data set consists of a list of 320 defendants formally charged with insider trading by the SEC in civil or administrative cases from 1980 to 1989, with the (daily) trading volumes for these defendants sourced from (both public and non-public) SEC documents, along with the target firms' total trading volumes on the days of insider trading (sourced from Iterative Data Services' Investment Statistical Listing Tapes and from S&P's Daily Stock Price Record).²³

Table 1 summarizes useful statistics for this data set, extracted from Meulbreuk (1992) Sect. III.E. It is worth mentioning that the data set only covers detected insider trading violations (as reported to the SEC), and the volumes are precisely those traded by the defendants (who have been detected and prosecuted), while much insider trading remains undetected over the observation period. On average, the (detected) insider trading volume constitutes about 8% of the total volume, both in terms of shares traded and dollar value, and so we shall utilize the share volume statistics exclusively.

Table 1: Statistics for (daily) insider trading volume data (1980–1989)
(Source: Meulbroek (1992))

Average insider share volume	9,819
Average insider dollar volume	\$300,023
Average total share volume	113,909
Standard error for average total share volume	10,246
Average total dollar volume	\$4,121,533
Standard error for average total dollar volume	\$594,327
Median insider-to-total volume ratio	11.3%

Under the equilibrium model framework, we adopt a *method-of-moment*-type estimation approach, which starts by deriving the theoretical counterparts of the statistics in Table 1 as functions of N and γ , and then setting them to be equal to the sample statistics, respectively, on the assumption that the underlying economy is in equilibrium. It is hence understood that the (single) insider trader in our model acts as a *representative agent* of insider trading during any insider trading episode, with N tracking the relative size of liquidity (non-inside or normal) traders compared to insiders. Also, as only defendants formally charged with insider trading are included in the data set, the statistics must be interpreted as conditional on prosecution – the reported average insider volumes exclude undetected or unprosecuted cases and unequivocally underestimate the actual insider trading volumes.

The theoretical counterparts of the above statistics are tractable to compute within the model by appealing to the approximate normality in the effect of large N , as discussed below; detailed derivations are in Supplemental Appendix B. Recall that B_N is the (independent) Bernoulli random variable representing the prosecution mechanism (see (2.1.1)). Then, given the insider's strategy Z , the estimated prosecution probability is

$$\mathbb{P}\{B_N(Z(V)) = 1\} = \sum_{v=0}^1 p(v)(1 - e^{-\lambda_N(Z(v))}), \quad \text{with } p(v) := \begin{cases} 1 - p, & \text{if } v = 0, \\ p, & \text{if } v = 1. \end{cases} \quad (5.1.1)$$

²³A more detailed breakdown of the data set, including a yearly analysis of reported insider trading volumes, can be found in Meulbroek (1992) Tab. I.

The average insider trading volume (in shares) corresponds to the conditional expectation of the (representative) insider's trade size upon successful prosecution, namely

$$\mathbb{E}[|Z(V)| \mid B_N(Z(V)) = 1] = \frac{\sum_{v=0}^1 p(v)|Z(v)|(1 - e^{-\lambda_N(Z(v))})}{\sum_{v=0}^1 p(v)(1 - e^{-\lambda_N(Z(v))})}. \quad (5.1.2)$$

To construct a meaningful match for the average total volume (in shares), let us recall that according to the model framework (Section 2), all the liquidity traders place independent orders, each with zero mean and variance σ^2 . Assuming that such orders are successfully filled, a reasonable proxy for the total liquidity volume is then the absolute sum of these independent random variables, and by the central limit theorem again, this sum can be approximated by a normal random variable $X_N \stackrel{d}{=} \text{Normal}(N\mu, N(\sigma^2 - \mu^2))$ for some parameter $\mu > 0$ (again, under the large- N assumption).²⁴ This coefficient, which depends on the exact trade distribution of an average liquidity trader, is yet to be determined, while for now we take it as given. Hence, the average total volume (in shares) should be matched to

$$\mathbb{E}[X_N + |Z(V)| \mid B_N(Z(V)) = 1] = N\mu + \frac{\sum_{v=0}^1 p(v)|Z(v)|(1 - e^{-\lambda_N(Z(v))})}{\sum_{v=0}^1 p(v)(1 - e^{-\lambda_N(Z(v))})}. \quad (5.1.3)$$

For the insider-to-total volume ratio, it is more convenient to consider its reciprocal (“total-to-insider volume ratio”), whose conditional tail distribution function on prosecution is given by

$$\mathbb{P}\left\{\frac{X_N + |Z(V)|}{|Z(V)|} > x \mid B_N(Z(V)) = 1\right\} = \frac{\sum_{v=0}^1 p(v)(1 - e^{-\lambda_N(Z(v))}) \operatorname{erfc}\frac{|Z(v)|(x-1)-N\mu}{\sqrt{2N\sigma^2}}}{2 \sum_{v=0}^1 p(v)(1 - e^{-\lambda_N(Z(v))})}, \quad x \geq 1. \quad (5.1.4)$$

Since the data set focuses on civil cases with pecuniary charges only, we follow the setting of Example 4.1 to design our calibration experiment (with $\gamma = \beta$). Assuming market efficiency, we fix the probability of revealing the upstate ($V = 1$) at $p = 1/2$. The consideration of a quadratic (size-modulated) hazard rate $\lambda(z) = Kz^2$, $z \in \mathbb{R}$, for some $K > 0$, agrees with the tail behaviors of (2.1.3) for abnormal order flow imbalance-based detection and can be seen as a reasonable approximation of the latter, and we have further $K = 1/(2\sigma^2)$; see Supplemental Appendix B for details. For civil penalties, we emphasize the “treble damages” provision again by testing three values of the penalty multiplier: $\chi \in \{1, 2, 3\}$ (or $\chi_0 \in \{0, 1, 2\}$).

If N happens to be large, then according to assertion (iii) in Theorem 4.1 and Theorem 4.2, we can take the re-scaled limiting equilibrium $(N^\gamma \tilde{Z}_\gamma^*, \tilde{P}_\gamma^*)$ as a reasonable approximation for the actual (finite- N) equilibrium (Z_N^*, P_N^*) . This implicit assumption can be verified posteriorly after the calibration. In the limiting case, with the above specified parameters, by following Example 4.1 we obtain the following insider trading strategy at equilibrium:

$$(\tilde{Z}_\gamma^*(0), \tilde{Z}_\gamma^*(1)) = (-\sigma \times \alpha, \sigma \times \alpha), \quad \text{with } \alpha = \sqrt{1 - 2W_0\left(\frac{\sqrt{e}\chi_0}{2\chi}\right)}, \quad (5.1.5)$$

which is proportional to σ , i.e., the trade size of an average liquidity trader, with a constant limiting price function $\tilde{P}_\gamma^* = 1/2$.

²⁴As each liquidity trader places an order ξ_i , $i = 1, \dots, N$, i.i.d. with zero mean and variance σ^2 , the total liquidity order flow is $\sum_{i=1}^N \xi_i \approx \sqrt{N}\bar{W}$ (assuming that N is large), while the total liquidity volume is approximated as $\sum_{i=1}^N |\xi_i| = X_N$. Also, with $\mathbb{E}[|\xi_1|] = \mu$, we have $\operatorname{Var}(|\xi_1|) = \mathbb{E}[|\xi_1|^2] - \mathbb{E}[|\xi_1|]^2 = \sigma^2 - \mu^2$.

Next, we make the substitution $Z = N^\gamma \tilde{Z}_\gamma^*$ from (5.1.5) in (5.1.2), (5.1.3), and (5.1.4), and then establish three calibration conditions by equating (5.1.2) and (5.1.3) to the numbers 9819 and 113909, respectively, and setting (5.1.4) to be equal to 1/2 for $x = 1/11.3\%$. However, it turns out that under (5.1.5), these conditions are not able to determine the values of N and σ simultaneously, because the latter is inherently a scaling factor for the total order flow. Instead, since N and σ clearly exhibit an inverse relationship, it is always possible to obtain a lower bound estimate for N by imposing an upper bound on σ . In accordance with the standard 100 round lot in the U.S. as well as the phenomenon of trade-size clustering (Alexander and Peterson (2007)), we set $\sigma = 1000$, which is a conservative enough estimate in this context given the (daily) average price around \$36.18 traded during the observation period – amounting to \$36,180 worth of trades placed per trader each day. Using smaller values of σ will necessarily lead to N increasing, to which the value of γ (as a power coefficient) is also resistant.

In addition, while the additional parameter $\mu > 0$ may be determined in several ways, we prefer to take on a direct approach leveraging the standard deviation of the total volume in the same data set, estimated to be 248,452; again, see Supplemental Appendix B for details. By equating the standard deviation of X_N , or $\sqrt{N(\sigma^2 - \mu^2)}$, to this value and using the calibration condition from (5.1.3), we arrive at an estimate $\hat{\mu} \approx 1.68625$.²⁵ This estimate is in keeping with the significant asymmetric and leptokurtic feature of trading volumes, which is a well-documented phenomenon in market microstructure; see, e.g., Mike and Farmer (2008) and Çetin and Waelbroeck (2024).²⁶ In particular, at the individual level, the average trading volume is exceeded substantially by its standard deviation (σ) in value, the relative smallness persisting with smaller values of σ . Then, by setting $\mu = 1.68625$ and $\sigma = 1000$, we continue to solve exactly two calibration conditions at a time and compare the parameter estimates $(\hat{N}, \hat{\gamma})$. Results are reported in Table 2.

Table 2: Results on calibration experiment I

Equations used		(5.1.2) and (5.1.3)	(5.1.2) and (5.1.4)	(5.1.3) and (5.1.4)
$(\hat{N}, \hat{\gamma})$	$\chi = 1$	(61729, 0.207091)	(45708, 0.21289)	(59918, 0.23226)
	$\chi = 2$	(61729, 0.249565)	(45708, 0.256553)	(59918, 0.274849)
	$\chi = 3$	(61729, 0.270651)	(45708, 0.27823)	(59918, 0.295992)

Table 2 shows that estimates for the stealth index γ are quite robust when using different calibration conditions, standing around 0.25, strictly lying between 0 and 0.5 as expected. These values represent a moderate stealth level and suggests that over the observation period, an insider tends to trade at significantly smaller scales than the entire population of liquidity traders while transcending an average liquidity trader. In the case $\chi = 2$, for instance, by noting that $15 < \hat{N}^\gamma < 21$, while $213 < \sqrt{\hat{N}} < 249$, the insider trade size is roughly 17 times larger than a liquidity trade size, albeit 13 times smaller than the size of all trades combined. This observation is in line with a key conclusion of Meulbroek (1992) that in spite of abnormal volumes on insider trading days, the insider trading vol-

²⁵The estimate is numerically stable – using the more complex conditions (5.1.2) or (5.1.4) instead of (5.1.3) produces similar values in the range [1, 2], and so for succinctness we stick with this estimate.

²⁶This phenomenon indicates that while many trading episodes carry small or no trades, a few episodes contain very large trades, which also closely reflects the aforementioned trade size clustering (e.g., Chang, Pinegar, and Schachter (1997), Alexander and Peterson (2007), and Fei and Xia (2024)). Notably, the heavy-tailed behavior does not conflict with the approximate-normality treatment for the present analysis, as the stated statistics are not tail-dependent and thus robust; indeed, replacing normality with a heavy-tailed distribution for X_N yields virtually no change to the results to follow.

ume only makes up a small portion of the total. The (lower bound) estimates for the population size N exhibit slightly larger variations across conditions but stand at a scale of $10^4 \gg 30$ (see Footnote 11), and the equilibrium model estimates that, on average, one insider trade is present for (at least) approximately every 50,000 non-insider trades over the same trading episodes.

One seemingly counterintuitive observation is that the estimate for γ exhibits an increasing trend with respect to the penalty multiplier χ . While one might well expect heightened penalties to deter insider activity and thus reduce γ , this result is actually consistent with the calibration approach. As the calibration works by fixing the sample statistics and deriving implied model parameters, achieving the same insider trade size under stricter penalties necessitates more aggressive insider trading behavior. As such, Table 2 should not be interpreted as a guide for designing optimal penalty multipliers.

We proceed to solve the finite-population equilibrium (existent) with the values of \hat{N} in Table 2 to check its closeness to the limiting equilibrium employed for the calibration, for which purpose we employ a *fixed-point algorithm* developed according to the equilibrium conditions in (B.11) and (B.14) in Appendix B. For a concise presentation, we concentrate on the case $\chi = 3$, with the resulting equilibrium objects presented in Table 3 and Figure 1 below – results in the cases $\chi = 1$ and $\chi = 2$ are substantially no different. Clearly, the finite- \hat{N} equilibrium objects are all very close to their limiting counterparts, verifying the validity of the approximation $(Z_N^*, P_N^*) \approx (N^\gamma \tilde{Z}_\gamma^*, \tilde{P}_\gamma^*)$ in the outset. From Theorem 4.1, the rate of equilibrium convergence is proportionate to the stealth index γ .

Table 3: Comparison of equilibrium strategies in calibration experiment I ($\chi = 3$)

Equations used	(5.1.2) and (5.1.3)	(5.1.2) and (5.1.4)	(5.1.3) and (5.1.4)
$(Z_{\hat{N}}^*(0), Z_{\hat{N}}^*(1))$	(-9813, 9813)	(-9811, 9811)	(-12862, 12862)
$(\hat{N}^\gamma \tilde{Z}_\gamma^*(0), \hat{N}^\gamma \tilde{Z}_\gamma^*(1))$	(-9819, 9819)	(-9819, 9819)	(-12872, 12872)

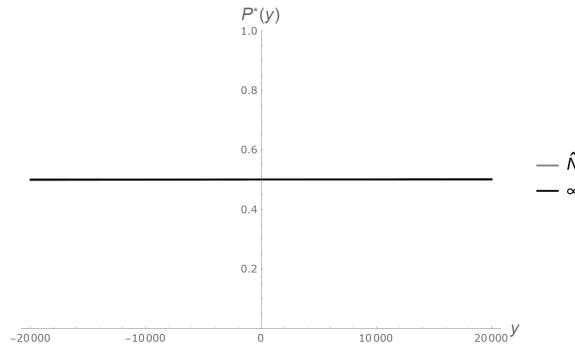


Figure 1: Comparison of equilibrium price function in calibration experiment I ($\chi = 3$)

5.2 Calibration experiment II

In the second calibration experiment, we leverage available data from two recent studies, Kacperczyk and Pagnotta (2024) and Patel and Putnins (2021), on legal risk in insider trading. First, using a data set of 530 insider trading cases prosecuted by the SEC between 1995 and 2018, Kacperczyk and Pagnotta (2024) constructed two key empirical proxies: the total dollar volume traded by the

insider over a trading episode (in units of days) (denoted as $\widehat{\text{Bet}}$ therein) and the corresponding (daily) insider-to-total volume ratio ($\widehat{\text{BetNorm}}$), the latter serving to provide a clearer indication of the market impact of insider trades. For these calculations, the total dollar volume represents the average daily dollar volume traded for the corresponding assets over the prior calendar year. Second, Patel and Putnins (2021) constructed a data set spanning the period from 1996 to 2016 and utilized a two-stage detection-controlled estimation method to derive contemporaneous estimates for the probability of detection and prosecution of insider trading across different corporate events. Thus, these two data sets cover observation periods that are considerably close in range.²⁷

We summarize some useful statistics for the calibration in Table 4.²⁸ The two statistics are gleaned from Kacperczyk and Pagnotta (2024) Tab. III & Fig. IA.5.

The idea of the calibration is in principle identical to that of experiment I. In the setting of Example 4.1, we set $p = 1/2$ and consider three penalty multiplier values $\chi \in \{1, 2, 3\}$ for experimentation, along with a quadratic hazard rate $\lambda(z) = z^2/(2\sigma^2)$.²⁹ In this case, the limiting equilibrium strategy \tilde{Z}_γ^* is as given in (5.1.5), with $\tilde{P}_\gamma^* = 1/2$.

Table 4: Statistics for insider trading volume data (1995–2018)
(Source: Kacperczyk and Pagnotta (2024))

Median insider volume	4,900
Median insider-to-total volume ratio	2.6%

On the assumption that $N \gg 1$, according to Table 4, we can match (5.1.2) to the number 4900 and equate (5.1.4) to 1/2 for $x = 1/2.6\%$ in order to form two calibration conditions. Indeed, in Table 4, we use the median insider volume instead of the average because the insider volume distribution over these observation periods is severely right-skewed and fat-tailed; the theoretical counterpart can still be represented by the conditional expectation (5.1.2), as the sample median converges to this value asymptotically given that $V \stackrel{d}{=} \text{Bernoulli}(1/2)$. For the trade size of an average liquidity trader, we stick to the previous values, namely $\mu = 1.68625$ and $\sigma = 1000$, in order to obtain a comparable conservative lower bound estimate for N . As before, the insider volume data over the observation periods should be interpreted as being conditional on prosecution. The theoretical counterpart of the prosecution probability is given by (5.1.1). However, upon applying $Z = N^\gamma \tilde{Z}_\gamma^*$ (with $\gamma = \beta$), the expression in (5.1.1) is not a function of the variables N and γ but can be used for later verifying the functional form of the hazard rate. Table 5 summarizes the calibration results in this experiment.

Table 5: Results on calibration experiment II

$(\hat{N}, \hat{\gamma})$	$\chi = 1$	(108858, 0.137029)
	$\chi = 2$	(108858, 0.177425)
	$\chi = 3$	(108858, 0.19748)

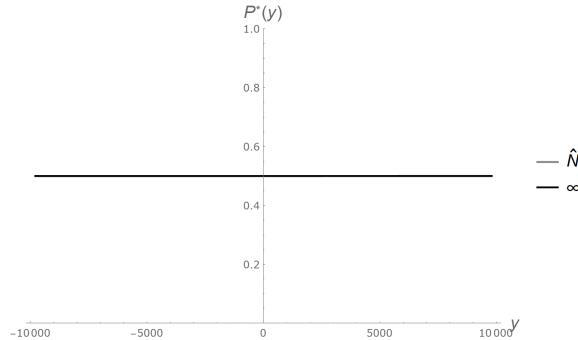
²⁷For more details regarding the two data sets we refer to Kacperczyk and Pagnotta (2024) Tab. I and Patel and Putnins (2021) Sect. 2, respectively.

²⁸Although Kacperczyk and Pagnotta (2024) primarily focuses on dollar volume, for adequate comparison (with calibration experiment I) we still use share volume statistics.

²⁹The equilibrium simulation in Kacperczyk and Pagnotta (2024) is conducted in a very similar setting, with a binary asset value V with range equal to 1 and $p = 1/2$. The (size-modulated) hazard rate λ_1 from (2.1.3) can be reasonably approximated by the quadratic hazard rate.

Table 6: Comparison of equilibrium strategies in calibration experiment II ($\chi = 3$)

	Strategy	implied prosecution probability
$(Z_{\hat{N}}^*(0), Z_{\hat{N}}^*(1))$	$(-4900, 4900)$	11.572%
$(\hat{N}^{\gamma} \tilde{Z}_{\hat{\gamma}}^*(0), \hat{N}^{\gamma} \tilde{Z}_{\hat{\gamma}}^*(1))$	$(-4900, 4900)$	11.576%

Figure 2: Comparison of equilibrium price function in calibration experiment II ($\chi = 3$)

We see that based on statistics from this non-overlapping new observation period (1995–2018), the stealth index γ is again estimated to strictly lie within the interval $(0, 0.5)$, as expected, though the estimates suggest a somewhat higher stealth level compared to the estimates from the first experiment covering a pre-1990 period. This could indicate gradually increased prudence among insiders after 2000, likely in response to enhanced regulatory measures (such as increased regulatory budgets and the introduction of the SEC Whistleblower Program), which also agrees with the increasing trends of the prosecution probability as discussed in Patel and Putnins (2021) Sect. 4.3. In addition, the implied lower bound of the population size, \hat{N} , turns out to be larger than those from the first experiment (Table 2), standing at the 10^5 scale.

As before, we compute the finite-population equilibrium (existent) using the estimates $(\hat{N}, \hat{\gamma})$ from Table 5 to justify the validity of employing the limiting equilibrium; again, we only illustrate the case $\chi = 3$. Table 6 and Figure 2 verify the closeness between the equilibrium objects, respectively. In the last column, the model-implied prosecution probabilities are computed by plugging the equilibrium strategy values into (5.1.1), and notably, are at a comparable level to the average detection rate of 15% reported by Patel and Putnins (2021) Sect. 2, a key conclusion of their study (with data range 1996–2016).³⁰ This observation also empirically validates the quadratic form of the hazard rate.

6 Further discussions: Criminal and mixed penalties

In this section, we discuss the general scenario where criminal charges are within consideration. Suppose that on successful prosecution, besides losing all of his illicit profit, the insider trader faces a mixture of both civil and criminal penalties. While civil penalties are calculated with penalty multipliers applied to the insider's illicit profit as before, criminal penalties are determined based on his

³⁰The implied values both fall within the 95% confidence interval for the average detection rate for insider trading ahead of earnings announcements; see Patel and Putnins (2021) Sect. 4.2.

realized trading strategy (Z). Therefore, we are looking at the general form (2.2.1) of the penalty function.

For any population size $N \geq 1$, the insider's objective function is given by (2.3.8) in equilibrium, which is detailed as, for $v \in \{0, 1\}$ and $z \in \mathbb{R}_v$,

$$\hat{J}_N(Z; z, v) = e^{-\lambda_N(z)} Q_N(Z; z, v) - (1 - e^{-\lambda_N(z)}) (C_0(z) + \chi_0 Q_N(Z; z, v)). \quad (6.1)$$

Recall that $\chi_0 = \chi - 1$ is the adjusted penalty multiplier, and Q_N denotes the insider's expected profit in (2.3.7). The last term in (6.1) shows the total penalty besides disgorgement, reflecting a balanced mechanism that accounts for both the insider's malicious intent in engaging in illicit trading and his realized profit, without either directly offsetting the other. We emphasize that under the general form (2.2.1), any leniency or discretion exercised by the regulator can be incorporated through the functional form of the criminal penalty function (C_0) or the penalty multiplier (χ_0) governing civil penalties, and so the approach by adding both types of penalties is sufficiently general from a practical standpoint as well.

By setting the penalty multiplier $\chi = 1$ in the structure (2.2.1), we come to a situation where the insider faces criminal penalties exclusively on top of disgorgement. Then, with $\chi_0 = 0$, (6.1) becomes

$$\hat{J}_N(Z; z, v) = e^{-\lambda_N(z)} Q_N(Z; z, v) - (1 - e^{-\lambda_N(z)}) C_0(z), \quad z \in \mathbb{R}_v, v \in \{0, 1\}. \quad (6.2)$$

A comparison between (6.2) and (4.1) shows that, different from the predominant scenario involving civil penalties only, the penalty function C_0 has no explicit dependence on the asset value $v \in \{0, 1\}$, focusing instead on the insider's trading behavior for a specific asset value – and not that in the alternative situation, e.g., through the price function P_N (or the general form of Ψ_N). This is reasonable, as in criminal cases, insider trading behavior is assessed based on the realized asset value (to which the insider has prior access), rather than evaluated in hypothetical situations (Picardo (2022)). The insider is also aware that penalties necessarily increase with his actual trading activity, governed by the monotonicity of the penalty function C_0 .

As noted earlier, the asymptotic growth of C_0 should be decisive of the asymptotic behaviors of the trading strategies. This leads to the following conditions in addition to Assumption 4.1.

Assumption 6.1. *The following three conditions hold.*

- (i) For $v \in \{0, 1\}$, both λ and C_0 are continuously differentiable and convex (not necessarily strictly) on \mathbb{R}_v , with $\lambda' < 0$ and $C'_0 \leq 0$ on $(-\infty, 0)$, while $\lambda' > 0$ and $C'_0 \geq 0$ on $(0, \infty)$.
- (ii) There exist constants $\theta \geq 1$, $\theta' > 0$, and $K_\theta > 0$, such that³¹

$$\lambda(z) = K_\theta |z|^\theta (1 + o(1)), \quad \lambda'(z) = K_\theta \theta z |z|^{\theta-2} + O(|z|^{\theta(1+\theta')-1}), \quad \text{as } |z| \searrow 0. \quad (6.3)$$

- (iii) There exist constants $\alpha \geq 1$ and $K_\alpha > 0$ and, if $\alpha > 1$, an additional constant $\alpha' > 0$, such that

$$C_0(z) = K_\alpha |z|^\alpha (1 + o(1)), \quad C'_0(z) = K_\alpha \alpha z |z|^{\alpha-2} + O(|z|^{\alpha-1-\alpha'}), \quad \text{as } |z| \rightarrow \infty. \quad (6.4)$$

³¹Given the convexity of λ , an application of the monotone density theorem (see, e.g., Bingham, Goldie, and Teugels (1989)) ensures that $\lambda'(z) = K_\theta \theta z |z|^{\theta-2}$ if $\lambda(z) \sim K_\theta |z|^\theta$, both as $|z| \searrow 0$; however, there are no direct implications for the asymptotic behavior of the remainder ($\lambda'(z) - K_\theta \theta z |z|^{\theta-2}$), hence the imposition of (6.3).

Since the choice of the penalty function is entirely up to the regulator, it is customizable and can be adjusted flexibly on a case-by-case basis. By contrast, the small argument behavior assumption regarding the hazard rate is natural and covers abnormal order flow imbalance-based detection (DeMarzo, Fishman, and Hagerty (1998)). For example, in (2.1.3), since the sum of the two complementary error functions is asymptotically equivalent to z^2 as $|z| \searrow 0$, we have that $\theta = \min\{\theta_D, 2\}$ if $D(z) \sim K_D|z|^{\theta_D}$ as $|z| \searrow 0$ for $K_D > 0$ and $\theta_D \geq 1$; details are in Supplemental Appendix B. The value of α is likely inversely related to γ , and in fact, the assumption on λ is only dominant when α becomes “sufficiently large,” as the insider will judiciously reduce the prosecution probability so as to offset the fast-growing (additional) penalty term in (6.2). Thus, Assumption 6.1 includes small arguments because the scaled quantity $N^{-\beta}z$ can still be small (as $N \rightarrow \infty$) even if z is large.

Under Assumption 6.1, we are able to extrapolate the main results in the predominant scenario (Section 4 Theorem 4.1) to the present general scenario concerning both civil and criminal penalties. In particular, a finite-population equilibrium (Z_N^*, P_N^*) in the sense of Definition 2.1 continues to exist; see Supplemental Appendix A for details. Most striking is the existence of a unique γ -limiting equilibrium $(\tilde{Z}_\gamma^*, \tilde{P}_\gamma^*)$ in the sense of Definition 3.1, which is also unique when $\gamma < 1/2$; in particular, (4.2) is to be replaced by the more elaborate condition

$$\gamma = \min\left\{\frac{\beta\theta}{\theta + \alpha - 1}, \frac{1}{2}\right\}. \quad (6.5)$$

Based on (6.5), the stealth index still has the upper bound $1/2$, while it further exhibits a clear inverse relationship with α . If $\alpha = 1$, the insider trader views the criminal penalties as relatively minor and simply chooses $\gamma = \min\{\beta, 1/2\}$, based entirely on the scale effect from λ (equivalently, the prosecution probability), as in (4.2); in this case, the small argument behavior of λ (namely θ) plays no role. Conversely, as α exceeds 1 , indicating severe criminal penalties, the insider begins to place greater weight on the consequences of prosecution and, accordingly, trades less aggressively with a smaller value $\gamma < \beta$. These results are formally stated in Theorem SA.1 in Supplemental Appendix A, along with the corresponding convergence rates that are jointly determined by all the coefficients $\beta, \theta, \alpha, \theta'$, and α' .

We give the following illustrative example to highlight the analytical simplicity of the limiting equilibrium.

Example 6.1. Let $p = 1/3$, $\lambda(z) = C_0(z) = |z|$, and $\chi = 1$, which satisfy Assumption 6.1, with $\theta = \alpha = 1$. Suppose that $\beta \in [0, 1/2)$, so that $\gamma = \beta < 1/2$. Then, with the Lambert W function W_0 , the γ -limiting equilibrium is uniquely determined as

$$\tilde{Z}_\gamma^* \equiv (\tilde{Z}_\gamma^*(0), \tilde{Z}_\gamma^*(1)) = \left(W_0\left(\frac{3e}{4}\right) - 1, 1 - W_0\left(\frac{3e}{5}\right)\right) \approx (-0.138547, 0.23844), \quad \tilde{P}_\gamma^* = \frac{1}{3}.$$

In Example 6.1, the (size-modulated) hazard rate and the criminal penalty function are both perfectly proportional to the insider’s trading quantities, and the stealth index γ does not alter the form of the limiting equilibrium (Definition 3.1), while only affecting the convergence rate of the finite-population equilibrium (Theorem SA.1 assertion (iii)) – through the scaled equilibrium strategy \tilde{Z}_γ^* . With the asset value V more likely revealed to be 0 than 1 ($p < 1/2$), at equilibrium the insider also tends to trade more if $V = 1$ (with $\tilde{Z}_\gamma^*(1) > |\tilde{Z}_\gamma^*(0)|$) to exploit the price difference, which behavior differs fundamentally from what we have seen in Example 4.1 involving only civil penalties. From the regulator’s viewpoint, this fundamental difference speaks to a crucial benefit of retaining criminal penalties: to prosecute based on the intent of illicit trading – a factor that can exhibit significant asymmetry across different realizations (see, e.g., Öberg (2014)).

The next example shows that if the convexity conditions in Assumption 6.1 are violated, then the uniqueness of the limiting equilibrium is not guaranteed.

Example 6.2. Let $p = 1/3$, $\beta = 0$, and $\chi = 1$. Consider $\lambda(z) = \log(|z| + 1)$, which is concave (violating Assumption 6.1), and the piecewise penalty function

$$C_0(z) = \begin{cases} \frac{1}{4z} + \frac{1}{12}, & \text{if } z \leq -6, \\ -\frac{z}{144}, & \text{if } -6 < z \leq 0, \\ \frac{25z}{144}, & \text{if } 0 < z \leq \frac{6}{5}, \\ \frac{5}{12} - \frac{1}{4z}, & \text{if } z > \frac{6}{5}, \end{cases}$$

which is strictly increasing and continuously differentiable but again concave in $|z|$ on \mathbb{R}_v , $v \in \{0, 1\}$. Then, there is a continuum of 0-limiting equilibria:

$$\tilde{Z}_0^* \in (-\infty, -6] \times [\frac{6}{5}, \infty), \quad \tilde{P}_0^* = \frac{1}{3}. \quad (6.6)$$

Example 6.2 illustrates an extreme case where the prosecution mechanism is considerably weak and ineffective, with the maximum penalty capped, leading to little-to-zero deterrent effect on insider trading. Consequently, the insider's optimal strategies can involve unboundedly large trade sizes as the population of liquidity traders grows.

7 Conclusions

This paper has introduced a Kyle-type model for insider trading with legal risk featuring a flexible scheme of legal detection and penalty and the presence of a large liquidity trading crowd. Besides justifying the normal distribution of liquidity trade sizes, the population size N of liquidity traders has spawned a series of limiting equilibria that provide a profound understanding of the camouflage effect in insider trading, particularly uncovering insiders' choices of trading scales when taking advantage of the surrounding trading masses in avoidance of detection. These trading scales are quantified by the stealth index γ , which is heavily associated with the population size over concurrent trading episodes.

The equilibrium analysis covering various types of penalty functions has shown that the stealth index implied by insiders' trading behavior is largely dependent on the mechanisms of detection as well as the imposition of penalties from regulators' side (Picardo (2022)). In particular, if only civil penalties are imposed – based on illicit profit from trade – then in equilibrium, insiders choose the stealth index solely based on the scale effect from the perceived probability of prosecution (see (4.2)), as long as the resulting trade size does not reach the level of the combined activity of all liquidity traders. On the other hand, when severe criminal penalties are devised (with $\alpha > 1$), insiders are prone to adopting an elevated stealth level (see (6.5)) compared to what the underlying detection mechanism implies. This striking disparity highlights the importance of reserving criminal sanctions in insider trading cases, due to their effectiveness in deterring illicit trading driven by a highly strategic intent such as market abuse (Öberg (2014) and Dalko and Wang (2016)).

On a technical level, the proven equilibrium convergence properties (Theorem 4.2 and assertions (iii) in Theorems 4.1 and SA.1) convey a key message that employing a limiting equilibrium significantly enhances the ease of analysis without altering the equilibrium's fundamental implications. The

empirical perspectives in Section 5 further demonstrate the ease of equilibrium calibration to insider trading volume data and confirm the prevalence of stealth trading amid sizeable normal trading activity, revealing a moderate stealth level in insider trading (Barclay and Warner (1993) and Chakravarty (2001)) accompanied by a significant reduction in price informativeness (Kacperczyk and Pagnotta (2024)) – hence conforming to the camouflage effect.

Acknowledgements

The authors are grateful to Marcin Kacperczyk for kindly sharing supplementary statistics for their data set in Kacperczyk and Pagnotta (2024). Jin Ma and Jianfeng Zhang are supported in part by U.S. NSF grant #DMS-2510403.

References

- Abramowitz, M., & Stegun, I.A. (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 10th printing. U.S. National Bureau of Standards, Washington, D.C.
- Admati, A.R., & Pfleiderer, P. (1988). A theory of intraday patterns: Volume and price variability. *The Review of Financial Studies*, 1(1), 3–40.
- Alexander, G.J., & Peterson, M.A. (2007). An analysis of trade-size clustering and its relation to stealth trading. *Journal of Financial Economics*, 84(2), 435–471.
- Anand, A., & Chakravarty, S. (2007). Stealth trading in options markets. *Journal of Financial and Quantitative Analysis*, 42(1), 167–187.
- Back, K. (1992). Insider trading in continuous time. *The Review of Financial Studies*, 5(3), 387–409.
- Back, K., & Baruch, S. (2004). Information in securities markets: Kyle meets Glosten and Milgrom. *Econometrica*, 72(2), 433–465.
- Bagnoli, M., Viswanathan, S., & Holden, C. (2001). On the existence of linear equilibria in models of market making. *Mathematical Finance*, 11(1), 1–31.
- Barclay, M.J., & Warner, J.B. (1993). Stealth trading and volatility: Which trades move prices? *Journal of Financial Economics*, 34(3), 281–305.
- Bingham, N.H., Goldie, C.M., & Teugels, J. L. (1989). *Regular Variation* (Vol. 27). Cambridge university press.
- Blau, B. M. (2017). Price dynamics and speculative trading in bitcoin. *Research in International Business and Finance*, 41, 493–499.
- Boulatov, A., Kyle, A.S., Livdan, D. (2013). Uniqueness of equilibrium in the single period Kyle'85 model. *Working Paper*. 35 pages.
- Brascamp, H.J., & Lieb, E.H. (1976). On extensions of the Brunn–Minkowski and Prékopa–Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation. *Journal of functional analysis*, 22(4): 366–389.
- Cai, B.M., Cai, C.X., & Keasey, K. (2006). Which trades move prices in emerging markets?: Evidence from China's stock market. *Pacific-Basin Finance Journal*, 14(5), 453–466.

- Caldentey, R., & Stacchetti, E. (2010). Insider trading with a random deadline. *Econometrica*, 78(1): 245–283.
- Carré, S., Collin-Dufresne, P., & Gabriel, F. (2022). Insider trading with penalties. *Journal of Economic Theory*, 203, 105461.
- Çetin, U. (2025). Insider trading with penalties in continuous time. *Journal of Economic Theory*, 228: 106061.
- Çetin, U., & Waelbroeck, H. (2024). Power laws in market microstructure. In: *Peter Carr Gedenkschrift: Research Advances in Mathematical Finance* (pp. 753–819).
- Chakravarty, S. (2001). Stealth-trading: Which traders' trades move stock prices? *Journal of Financial Economics*, 61(2), 289–307.
- Chang, E.C., Pinegar, J.M., & Schachter, B. (1997). Interday variations in volume, variance and participation of large speculators. *Journal of Banking and Finance*, 21(6), 797–810.
- Chen, T. (2019). The price impact of trade-size clustering: Evidence from an intraday analysis. *Journal of Business Research*, 101, 300–314.
- Dalko, V., & Wang, M.H. (2016). Why is insider trading law ineffective? Three antitrust suggestions. *Studies in Economics and Finance*, 33(4), 704–715.
- Del Guercio, D., Odders-White, E.R., & Ready, M.J. (2017). The deterrent effect of the Securities and Exchange Commission's enforcement intensity on illegal insider trading: Evidence from run-up before news events. *The Journal of Law and Economics*, 60(2), 269–307.
- DeMarzo, P.M., Fishman, M.J., & Hagerty, K.M. (1998). The optimal enforcement of insider trading regulations. *Journal of Political Economy*, 106(3), 602–632.
- Fei, Z., & Xia, W. (2024). Regulating stochastic clocks §. *Quantitative Finance*, 24(7), 921–953.
- Frino, A., Satchell, S., Wong, B., & Zheng, H. (2013). How much does an illegal insider trade? *International Review of Finance*, 13(2), 241–263.
- Griskevicius, V., Goldstein, N.J., Mortensen, C.R., Cialdini, R.B., & Kenrick, D.T. (2006). Going along versus going alone: When fundamental motives facilitate strategic (non) conformity. *Journal of Personality and Social Psychology*, 91(2), 281–294.
- Kacperczyk, M., & Pagnotta, E.S. (2024). Legal risk and insider trading. *The Journal of Finance*, 79(1), 305–355.
- Kyle, A.S. (1985). Continuous auctions and insider trading. *Econometrica*, 53, 1315–1335.
- Meulbroek, L.K. (1992). An empirical analysis of illegal insider trading. *The Journal of Finance*, 47(5), 1661–1699.
- Mike, S., & Farmer, J.D. (2008). An empirical behavioral model of liquidity and volatility. *Journal of Economic Dynamics and Control*, 32(1), 200–234.
- Öberg, J. (2014). Is it “essential” to imprison insider dealers to enforce insider dealing laws?. *Journal of Corporate Law Studies*, 14(1), 111–138.
- Patel, V., & Putnins, T.J. (2021). How much insider trading happens in stock markets? *SSRN Working Paper*. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3764192.
- Picardo, E. (2022). How the SEC tracks insider trading. *Investopedia*. Available at <https://www.investopedia.com/articles/investing/021815/how-sec-tracks-insider-trading.asp>.
- Saumard, A., & Wellner, J.A. (2014). Log-concavity and strong log-concavity: A review. *Statistics Surveys*, 8: p45.
- Shin, J. (1996). The optimal regulation of insider trading. *Journal of Financial Intermediation*, 5(1), 49–73.
- Yamazato, M. (1978). Unimodality of infinitely divisible distribution functions of class L. *The Annals of Probability*, 523–531.

Appendix A Proof of Proposition 3.1

Based on (2.3.3), we can write

$$\begin{aligned}\tilde{P}_N^\gamma(\tilde{Z}; \tilde{y}) &= P_N(N^\gamma \tilde{Z}; N^{\max\{\gamma, \frac{1}{2}\}} \tilde{y}) = \frac{1}{1 + q e^{\tilde{\kappa}_N(\tilde{y})}}, \\ \text{with } \tilde{\kappa}_N(\tilde{y}) &:= \frac{(2N^{\frac{1}{2}-\gamma, 0}) \tilde{y} - \tilde{z}_0 - \tilde{z}_1)(\tilde{z}_0 - \tilde{z}_1)}{2N^{1-2\gamma} \sigma^2}, \quad \tilde{z}_v := \tilde{Z}(v).\end{aligned}\tag{A.1}$$

First, if $\gamma < 1/2$, then the exponent $\tilde{\kappa}_N(\tilde{y})$ above is

$$\tilde{\kappa}_N(\tilde{y}) = \frac{(2N^{\frac{1}{2}-\gamma} \tilde{y} - \tilde{z}_0 - \tilde{z}_1)(\tilde{z}_0 - \tilde{z}_1)}{2N^{1-2\gamma} \sigma^2} = \frac{\tilde{z}_0 - \tilde{z}_1}{2\sigma^2} \left(\frac{2\tilde{y}}{N^{\frac{1}{2}-\gamma}} - \frac{\tilde{z}_0 + \tilde{z}_1}{N^{1-2\gamma}} \right),\tag{A.2}$$

which tends to 0 as $N \rightarrow \infty$. Thus, (A.1) becomes $\lim_{N \rightarrow \infty} \tilde{P}_N^\gamma(\tilde{Z}; \tilde{y}) = 1/(1 + q) = p$.

Second, if $\gamma = 1/2$, the terms involving N cancel out, and we have for (A.1) that $\tilde{P}_N^\gamma(\tilde{Z}; \tilde{y}) = 1/(1 + q e^{(2\tilde{y} - \tilde{z}_0 - \tilde{z}_1)(\tilde{z}_0 - \tilde{z}_1)/(2\sigma^2)}) = P_1(\tilde{Z}; \tilde{y})$.

Third, let $\gamma > 1/2$. Then $\tilde{\kappa}_N(\tilde{y}) = (2\tilde{y} - \tilde{z}_0 - \tilde{z}_1)(\tilde{z}_0 - \tilde{z}_1)/(2N^{1-2\gamma} \sigma^2)$. If $\tilde{z}_0 = \tilde{z}_1 = 0$ or $2\tilde{y} - \tilde{z}_0 - \tilde{z}_1 = 0$, it is clear that $\tilde{\kappa}_N(\tilde{y}) = 0$, and thus $\tilde{P}_N^\gamma(\tilde{Z}; \tilde{y}) = p$. Otherwise, we have $\tilde{z}_0 - \tilde{z}_1 < 0$, and so if $2\tilde{y} - \tilde{z}_0 - \tilde{z}_1 > 0$, then with the exponent $\tilde{\kappa}_N(\tilde{y})$ tending to 0, we have $\tilde{P}_N^\gamma(\tilde{Z}; \tilde{y}) \rightarrow 1$, as $N \rightarrow \infty$, while if $2\tilde{y} - \tilde{z}_0 - \tilde{z}_1 < 0$, the exponent $\tilde{\kappa}_N(\tilde{y})$ goes to ∞ , yielding that $\tilde{P}_N^\gamma(\tilde{Z}; \tilde{y}) \rightarrow 0$.

Summarizing the three cases, we have verified the limit expressions in (3.2). ■

Appendix B Proof of Theorem 4.1

We start with the following lemma that governs the log-concavity property of the insider's expected profit function.

Lemma B.1. *For any given $N \geq 1$, $Z = (Z(0), Z(1)) \in \mathbb{R}_0 \times \mathbb{R}_1$, and $v \in \{0, 1\}$, the function $Q_N(Z; z, v)$ in (2.3.7) is strictly log-concave in $z \in \mathbb{R}_v$, i.e., $\log Q_N(Z; \cdot, v)$ is a strictly concave function on \mathbb{R}_v .*

Proof. Let $N \geq 1$ and $Z = (Z(0), Z(1)) \in \mathbb{R}_0 \times \mathbb{R}_1$ be fixed. We first consider the case $v = 0$, with $z < 0$. By (2.3.3), we write succinctly

$$\begin{aligned}P_N(Z; y) &= \frac{1}{1 + e^{ay+b}}, \\ \text{where } a &:= \frac{Z(0)-Z(1)}{N\sigma^2} < 0, \quad b := \log q + \frac{Z(1)^2-Z(0)^2}{2N\sigma^2}.\end{aligned}\tag{B.1}$$

Then, $\log P_N(Z; y) = -\log(1 + e^{ay+b})$, and so

$$\partial_y \log P_N(Z; y) = -\frac{ae^{ay+b}}{1 + e^{ay+b}} = -a + \frac{a}{1 + e^{ay+b}}, \quad \partial_{yy} \log P_N(Z; y) = -\frac{a^2 e^{ay+b}}{(1 + e^{ay+b})^2} < 0,$$

which show that $P_N(Z; y)$ is strictly log-concave in y . Also, note that

$$\Phi_N(Z; z) = \int_{\mathbb{R}} P_N(Z; \sqrt{N}x + z) \frac{e^{-\frac{x^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}} dx.\tag{B.2}$$

Since it is well-known that the Gaussian distribution is log-concave, in the sense that the density function $e^{-x^2/(2\sigma^2)}/\sqrt{2\pi\sigma^2}$ is log-concave in $x \in \mathbb{R}$, an application of the Prékopa–Leindler theorem (see, e.g., Brascamp and Lieb (1976)) yields that $\Phi_N(Z; z)$ in (B.2) is log-concave in z . Note further that, for $v = 0$ and $z < 0$, we have

$$\log Q_N(Z; z, 0) = \log(-z) + \log \Phi_N(Z; z).$$

Then, by the (strict) concavity of the logarithm, we see that $Q_N(Z; z, 0)$ is (strictly) log-concave in $z < 0$.

Next, consider the case $v = 1$ and $z > 0$, with

$$\log Q_N(Z; z, 1) = \log z + \log(1 - \Phi_N(Z; z)), \quad (\text{B.3})$$

and $1 - P_N(Z; y) = e^{ay+b}/(1 + e^{ay+b}) = e^{ay+b}P_N(Z; y)$, and so $\log(1 - P_N(Z; y)) = ay + b + \log P_N(Z; y)$. The log-concavity of $1 - P_N(Z; y)$ in y , along with the Prékopa–Leindler theorem again, implies that $1 - \Phi_N(Z; z)$ is log-concave in z . It then follows from (B.3) that $Q_N(Z; z, 1)$ is (strictly) log-concave in $z > 0$ as well. ■

Proof of Theorem 4.1 assertion (i). Consider the finite- N equilibrium in Definition 2.1. We proceed in two steps.

Step 1 This step is concerned with the insider's optimization problem – that is, given $N \geq 1$, $Z = \overline{(Z(0), Z(1))} \in \mathbb{R}_0 \times \mathbb{R}_1$, and $v \in \{0, 1\}$, we shall apply Lemma B.1 to show that the objective function $\tilde{J}_N(Z; z, v)$ has a unique maximum point $z_v^* \in \mathbb{R}_v$. We assume for simplicity that $\lambda(z) \equiv \lambda_1(z)$ (and hence $\lambda_N(z)$) is twice-differentiable for $z \neq 0$.³² By (4.1) we have

$$\tilde{J}_N(Z; z, v) = (\chi e^{-\lambda_N(z)} - \chi_0)Q_N(Z; z, v), \quad z \in \mathbb{R}_v. \quad (\text{B.4})$$

We shall only prove the case $v = 0$. The case $v = 1$ can be proved analogously.

We now restrict to $z < 0$, and denote $\Lambda_N(z) := \chi e^{-\lambda_N(z)} - \chi_0$. Recall that $Q_N > 0$. By Assumption 4.1, since $\lambda'_N(z) < 0$, the (continuous) function Λ_N is strictly increasing, with $\lim_{z \rightarrow -\infty} \Lambda_N(z) = -\chi_0 \leq 0$ and $\lim_{z \nearrow 0} \Lambda_N(z) = 1$. Thus, if $\chi_0 > 0$, then Λ_N has a unique zero $z_0^* < 0$; if $\chi_0 = 0$, $\Lambda_N > 0$ on $(-\infty, 0)$, and we can simply set $z_0^* = -\infty$. In either case, on $(z_0^*, 0)$, we have $\Lambda_N > 0$, and so $\tilde{J}_N(Z; z, 0) > 0$. Note that $\lim_{z \rightarrow z_0^+} \tilde{J}_N(Z; z, 0) = \lim_{z \nearrow 0} \tilde{J}_N(Z; z, 0) = 0$, and then $\tilde{J}_N(Z; z, 0)$ admits a local maximum point $z_0^* \in (z_0^*, 0)$.

To show that z_0^* is the unique maximum point of $\tilde{J}_N(Z; \cdot, 0)$ on $\mathbb{R}_0 = (-\infty, 0)$, let us note further that $\tilde{J}_N(Z; \cdot, 0) \leq 0$ on $(-\infty, z_0^*]$ if $\chi_0 < 0$, and so it suffices to consider $z \in (z_0^*, 0)$. By Lemma B.1, the function $Q_N(Z; \cdot, 0) > 0$ is strictly log-concave. Since $\Lambda_N > 0$ and $\lambda''_N \geq 0$, then

$$\frac{d^2}{dz^2} \log \Lambda_N(z) = -\frac{\chi \chi_0 (\lambda'_N(z))^2 e^{-\lambda_N(z)}}{(\Lambda_N(z))^2} - \frac{\chi \lambda''_N(z) e^{-\lambda_N(z)}}{\Lambda_N(z)} \leq 0; \quad (\text{B.5})$$

that is, Λ_N is log-concave on $(z_0^*, 0)$, and by (B.4) and Lemma B.1, $\tilde{J}_N(Z; \cdot, 0)$ is strictly log-concave. This implies that $\tilde{J}_N(Z; \cdot, 0)$ admits at most one maximum point on $(z_0^*, 0)$, which is z_0^* .

³²The twice-differentiability of λ is mainly to ease verification of the log-concavity of Λ_N in (B.5). The same property can be verified by slightly more involved arguments if λ is only convex.

Step 2 In this step, we show the existence of a finite- N equilibrium (Definition 2.1). By continuous differentiability from Assumption 4.1, we take the first-order conditions for (B.4). Then, following the analysis in Step 1, any equilibrium strategy $Z_N^* = (Z_N^*(0), Z_N^*(1)) \in \mathbb{R}_0 \times \mathbb{R}_1$ must satisfy the equilibrium conditions

$$\check{J}_N(Z_N^*; Z_N^*(v), v) := [e^{\lambda_N(z)} \partial_z \check{J}_N(Z_N^*; z, v)]|_{z=Z_N^*(v)} = 0, \quad v \in \{0, 1\}. \quad (\text{B.6})$$

By (2.3.7), we have

$$\check{J}_N(Z; z, v) = -\chi \lambda'_N(z)(v - \Phi_N(Z; z))z - (\chi - \chi_0 e^{\lambda_N(z)})(\partial_z \Phi_N(Z; z)z - v + \Phi_N(Z; z)). \quad (\text{B.7})$$

Step 2.1 Consider the first-order conditions in (B.6) for $v = 0$ first. Denote $\zeta_N^* := Z_N^*(1) - Z_N^*(0) > 0$. Note that, for fixed $\zeta > 0$ and $z < 0$, by (2.3.3) and (2.3.7) we have

$$\begin{aligned} \Phi_N((z, \zeta + z); z) &= \bar{\Phi}_N(\zeta), \quad (\partial_z \Phi_N)((z, \zeta + z); z) = \bar{\Phi}'_N(\zeta), \quad \text{where} \\ \bar{\Phi}_N(\zeta) &:= \mathbb{E}\left[\frac{1}{1 + qe^{\bar{\eta}_N(\zeta)}}\right] > 0, \quad \bar{\Phi}'_N(\zeta) := \frac{\zeta}{N\sigma^2} \mathbb{E}\left[\frac{qe^{\bar{\eta}_N(\zeta)}}{(1 + qe^{\bar{\eta}_N(\zeta)})^2}\right] > 0, \\ \bar{\eta}_N(\zeta) &:= \frac{\zeta(\zeta - 2\sqrt{NW})}{2N\sigma^2}. \end{aligned} \quad (\text{B.8})$$

Based on (B.7), we define for any fixed $\zeta > 0$

$$\begin{aligned} F_0^N(\zeta; z) &:= \check{J}_N((z, \zeta + z); z, 0) = g_{1,N}(z)\bar{\Phi}_N(\zeta) + g_{2,N}(z)\bar{\Phi}'_N(\zeta), \quad z < 0, \\ \text{where } g_{1,N}(z) &:= \chi z \lambda'_N(z) - \chi + \chi_0 e^{\lambda_N(z)}, \quad g_{2,N}(z) := (\chi_0 e^{\lambda_N(z)} - \chi)z. \end{aligned} \quad (\text{B.9})$$

Since $Z_N^*(0) \leq -\zeta_N^*$, the effective domain of F_0^N in (B.9) should be $(-\infty, -\zeta]$; however, for technical reasons, we extend F_0^N to the whole $(-\infty, 0)$.

By Assumption 4.1, since $g'_{1,N}(z) = (\chi_0 e^{\lambda_N(z)} + \chi)\lambda'_N(z) + \chi z \lambda''_N(z) < 0$, the function $g_{1,N}$ is strictly decreasing on $(-\infty, 0)$. Further, noting that $\lim_{z \rightarrow -\infty} g_{1,N}(z) = \infty$ and $\lim_{z \nearrow 0} g_{1,N}(z) = -1$, then $g_{1,N}$ has a unique zero $z_0^\diamond < 0$, and thus $g_{1,N}(z) < 0$ for $z \in (z_0^\diamond, 0)$. We now focus on the interval $(z_0^\diamond, 0)$. Note that $g'_{2,N}(z) = \chi_0 e^{\lambda_N(z)}(z \lambda'_N(z) + 1) - \chi$ and $\lambda_N(z) > 0 > \lambda'_N(z)$ for $z < 0$, then for $z \in (z_0^\diamond, 0)$,

$$g_{1,N}(z) - g'_{2,N}(z) = z \lambda'_N(z)(\chi - \chi_0 e^{\lambda_N(z)}) = z \lambda'_N(z)(z \lambda'_N(z) - g_{1,N}(z)) > 0; \quad (\text{B.10})$$

thus $g'_{2,N}(z) < g_{1,N}(z) < 0$. Since $\bar{\Phi}_N, \bar{\Phi}'_N > 0$ on $(0, \infty)$, this implies that $F_0^N(\zeta; \cdot)$, along with $g_{2,N}$, is strictly decreasing on $(z_0^\diamond, 0)$. Moreover, since $\lim_{z \nearrow 0} g_{2,N}(z) = 0$, then $g_{2,N}(z_0^\diamond) > 0$, and $F_0^N(\zeta; z^\diamond) = g_{2,N}(z_0^\diamond)\bar{\Phi}'_N(\zeta) > 0$, with $\lim_{z \nearrow 0} F_0^N(\zeta; z) = -\bar{\Phi}_N(\zeta) < 0$. Thus, $F_0^N(\zeta; \cdot)$ has a unique zero in $(z_0^\diamond, 0)$, denoted as $z_0(\zeta)$.

Denote $z_1(\zeta) := \zeta + z_0(\zeta)$ and $\mathcal{Z}^\zeta := (z_0(\zeta), z_1(\zeta))$. Recall the function Λ_N and the point z_0^\diamond from Step 1. From (B.10) we also see that $\Lambda_N(z) = e^{-\lambda_N(z)}(\chi - \chi_0 e^{\lambda_N(z)}) > 0$ for $z \in (z_0^\diamond, 0)$, which implies that $(z_0^\diamond, 0) \subset (z_0^\diamond, 0)$, and thus, from Step 1, $\check{J}_N(\mathcal{Z}^\zeta; \cdot, 0)$ is strictly log-concave on $(z_0^\diamond, 0)$. It then follows from the first-order condition (B.6) and Step 1 that $z_0(\zeta)$ is the (unique) maximum point of $\check{J}_N(\mathcal{Z}^\zeta; \cdot, 0)$ in \mathbb{R}_0 . We remark that the above analysis does not require $z_1(\zeta) > 0$, which will be verified later for the equilibrium however.

For the purpose of the next sub-step, let us also note that

$$F_0^N(\zeta; z_0(\zeta)) = g_{1,N}(z_0(\zeta))\bar{\Phi}_N(\zeta) + g_{2,N}(z_0(\zeta))\bar{\Phi}'_N(\zeta) = 0. \quad (\text{B.11})$$

Since $g_{1,N}, g_{2,N}$ are continuous and strictly decreasing on $(z_0^\diamond, 0)$, and $\bar{\Phi}_N, \bar{\Phi}'_N$ are obviously continuous in ζ , it is clear that $\varkappa_0(\zeta)$ is continuous in ζ . Moreover, from (B.8) one can easily see that $\bar{\Phi}_N(0) = 1, \bar{\Phi}'_N(0) = 0$. Hence, by (B.11), we have that $\lim_{\zeta \searrow 0} g_{1,N}(\varkappa_0(\zeta)) = 0$, and so

$$\lim_{\zeta \searrow 0} \varkappa_0(\zeta) = z_0^\diamond. \quad (\text{B.12})$$

Step 2.2 Now we turn to (B.6) for $v = 1$, restricting to $\mathcal{Z}^\zeta := (\varkappa_0(\zeta), \varkappa_1(\zeta))$ based on Step 2.1. We stress again that at this point, we do not require $\varkappa_1(\zeta) > 0$. Note that

$$\begin{aligned} \Phi_N((z - \zeta, z); z) &= \hat{\Phi}_N(\zeta), \quad (\partial_z \Phi_N)((z - \zeta, z); z) = \hat{\Phi}'_N(\zeta), \quad \text{where} \\ \hat{\Phi}_N(\zeta) &:= \mathbb{E}\left[\frac{1}{1 + qe^{\hat{\eta}_N(\zeta)}}\right] > 0, \quad \hat{\Phi}'_N(\zeta) := \frac{\zeta}{N\sigma^2} \mathbb{E}\left[\frac{qe^{\hat{\eta}_N(\zeta)}}{(1 + qe^{\hat{\eta}_N(\zeta)})^2}\right] > 0, \\ \hat{\eta}_N(\zeta) &:= -\frac{\zeta(\zeta + 2\sqrt{N}W)}{2N\sigma^2}. \end{aligned} \quad (\text{B.13})$$

Recalling (B.7) and (B.9), we define

$$G_1^N(\zeta) := \check{J}_N(\mathcal{Z}^\zeta; \varkappa_1(\zeta), 1) = g_{1,N}(\varkappa_1(\zeta))(\hat{\Phi}_N(\zeta) - 1) + g_{2,N}(\varkappa_1(\zeta))\hat{\Phi}'_N(\zeta). \quad (\text{B.14})$$

Note that $\varkappa_1(\zeta) = \zeta + \varkappa_0(\zeta)$ is continuous for $\zeta > 0$ and $\varkappa_0(\zeta) > z_0^\diamond$. By (B.12) we have $\lim_{\zeta \searrow 0} \varkappa_1(\zeta) = z_0^\diamond < 0$ and $\lim_{\zeta \rightarrow \infty} \varkappa_1(\zeta) = \infty$. It is implied that there exists $\check{\zeta} > 0$ such that $\varkappa_1(\check{\zeta}) = 0$, and we choose the largest such $\check{\zeta}$, i.e., such that $\varkappa_1(\zeta) > 0$ for all $\zeta > \check{\zeta}$. Since $\hat{\Phi}_N(\check{\zeta}) < 1$, then by (B.9) and (B.14) we have

$$\lim_{\zeta \searrow \check{\zeta}} G_1^N(\zeta) = (\chi - \chi_0 e^{\lambda_N(0)})(1 - \hat{\Phi}_N(\check{\zeta})) = 1 - \hat{\Phi}_N(\check{\zeta}) > 0. \quad (\text{B.15})$$

On the other hand, since $g_{1,N}(0) = -\chi + \chi_0 = -1 < 0$, and clearly $\lim_{z \rightarrow \infty} g_{1,N}(z) = \infty$, then there exists $\hat{z} \in (0, \infty)$ such that $g_{1,N}(\hat{z}) = 0$. In particular, this implies that $\chi_0 e^{\lambda_N(\hat{z})} - \chi = -\chi \hat{z} \lambda'_N(\hat{z}) < 0$, and so $g_{2,N}(\hat{z}) = (\chi_0 e^{\lambda_N(\hat{z})} - \chi)\hat{z} < 0$. Moreover, recalling again that $\varkappa_1(\zeta)$ is continuous with $\varkappa_1(\check{\zeta}) = 0$ and $\lim_{\zeta \rightarrow \infty} \varkappa_1(\zeta) = \infty$, then there exists $\hat{\zeta} > \check{\zeta}$ such that $\varkappa_1(\hat{\zeta}) = \hat{z}$. Therefore,

$$G_1^N(\hat{\zeta}) = g_{1,N}(\hat{z})(\hat{\Phi}_N(\hat{\zeta}) - 1) + g_{2,N}(\hat{z})\hat{\Phi}'_N(\hat{\zeta}) = g_{2,N}(\hat{z})\hat{\Phi}'_N(\hat{\zeta}) < 0.$$

This, together with (B.15), implies that there exists $\zeta^* \in (\check{\zeta}, \hat{\zeta})$ such that $G_1^N(\zeta^*) = 0$.

Finally, since $\check{\zeta}$ is the largest zero of \varkappa_1 and $\zeta^* > \check{\zeta}$, we have $\varkappa_1(\zeta^*) > 0$. Then, by the reasoning in Step 2.1 and Step 1 again, it follows that $\varkappa_1(\zeta^*)$ is also the unique maximum point of $\check{J}_N(\mathcal{Z}^{\zeta^*}; \cdot, 1)$ on \mathbb{R}_1 . This, together with Step 2.1, implies that $(\mathcal{Z}^{\zeta^*}, P_N(\mathcal{Z}^{\zeta^*}; \cdot))$ is an equilibrium in the sense of Definition 2.1, as required. ■

Proof of Theorem 4.1 Assertion (ii). We shall prove the “if” direction and the “only if” direction separately. Let us recall Definition 3.1 condition (ii), (2.3.1), (2.2.1) with $C_0 = 0$, and (2.3.6). We see that, for any γ -limiting equilibrium $(\tilde{Z}_\gamma^*, \tilde{P}_\gamma^*)$ ($\tilde{Z}_\gamma^* \equiv (\tilde{Z}_\gamma^*(0), \tilde{Z}_\gamma^*(1))$),

$$\begin{aligned} \tilde{J}_\infty^\gamma(\tilde{P}_\gamma^*; \tilde{z}, v) &= \lim_{N \rightarrow \infty} N^{-\gamma} J_N(\tilde{P}_\infty^\gamma(\tilde{Z}_\gamma^*; N^{-\max\{\gamma, \frac{1}{2}\}}); N^\gamma \tilde{z}, v) \\ &= \lim_{N \rightarrow \infty} N^{-\gamma} (N^\gamma \tilde{z}) \left(e^{-\lambda_N(N^\gamma \tilde{z})} - \chi_0 (1 - e^{-\lambda_N(N^\gamma \tilde{z})}) \right) (v - \tilde{\Phi}_N^\gamma(\tilde{Z}_\gamma^*; \tilde{z})) \\ &= \lim_{N \rightarrow \infty} \tilde{z} (\chi e^{-\lambda(N^\gamma \tilde{z})} - \chi_0) (v - \tilde{\Phi}_N^\gamma(\tilde{Z}_\gamma^*; \tilde{z})), \quad \tilde{z} \in \mathbb{R}_v, \quad v \in \{0, 1\}, \end{aligned}$$

$$\text{where } \tilde{\Phi}_N^\gamma(\tilde{Z}_\gamma^*; \tilde{z}) := \mathbb{E}[\tilde{P}_\infty^\gamma(\tilde{Z}_\gamma^*; N^{-\max\{\gamma, \frac{1}{2}\}}(\sqrt{N}W + N^\gamma \tilde{z}))], \quad (\text{B.16})$$

with $\lambda \equiv \lambda_1$. Under Assumption 6.1, we can easily verify the following limits: For $\tilde{z} \neq 0$, with $z_\gamma^* := (\tilde{Z}_\gamma^*(0) + \tilde{Z}_\gamma^*(1))/2$,

$$\lim_{N \rightarrow \infty} e^{-\lambda(N^\gamma - \beta \tilde{z})} = \begin{cases} 1, & \text{if } \gamma < \beta, \\ \lambda(\tilde{z}), & \text{if } \gamma = \beta, \\ 0, & \text{if } \gamma > \beta; \end{cases} \quad \lim_{N \rightarrow \infty} \tilde{\Phi}_N^\gamma(\tilde{Z}_\gamma^*; \tilde{z}) = \begin{cases} p, & \text{if } \gamma < \frac{1}{2}, \\ \Phi_1(\tilde{Z}_\gamma^*; \tilde{z}), & \text{if } \gamma = \frac{1}{2}, \\ 1\mathbb{1}_{\{\tilde{z} > z_\gamma^*\}} + \frac{1}{2}1\mathbb{1}_{\{\tilde{z} = z_\gamma^*\}}, & \text{if } \gamma > \frac{1}{2}; \end{cases} \quad (\text{B.17})$$

in particular, the case $\gamma > 1/2$ follows by using (3.2) and (B.16) so that

$$\tilde{\Phi}_N^\gamma(\tilde{Z}_\gamma^*; \tilde{z}) = \mathbb{E}[1\mathbb{1}_{\{N^{1/2-\gamma}W + \tilde{z} > z_\gamma^*\}} + p1\mathbb{1}_{\{N^{1/2-\gamma}W + \tilde{z} = z_\gamma^*\}}] \rightarrow 1\mathbb{1}_{\{\tilde{z} > z_\gamma^*\}} + \frac{1}{2}1\mathbb{1}_{\{\tilde{z} = z_\gamma^*\}}.$$

Step 1 In this step we prove the “if” direction. Assume $\gamma = \min\{\beta, 1/2\}$. We naturally examine (B.16) in two cases.

Case 1 $\gamma = \beta < 1/2$. In this case, by Definition 3.1 (ii) and (3.3) from Proposition 3.1, we must set $\tilde{P}_\gamma^* = p$, and it is unique. Further, (B.16) gives that

$$\tilde{J}_\infty^\gamma(\tilde{P}_\gamma^*; \tilde{z}, v) = (v - p)\tilde{z}(\chi e^{-\lambda(\tilde{z})} - \chi_0), \quad \tilde{z} \in \mathbb{R}_v. \quad (\text{B.18})$$

Similar to (B.6), we define for $\tilde{z} \in \mathbb{R}_v$

$$\begin{aligned} \tilde{J}_\infty^\gamma(\tilde{z}, v) &:= e^{\lambda(\tilde{z})} \partial_{\tilde{z}} \tilde{J}_\infty^\gamma(\tilde{P}_\gamma^*; \tilde{z}, v) = (v - p)(\chi - \chi\tilde{z}\lambda'(\tilde{z}) - \chi_0 e^{\lambda(\tilde{z})}); \\ \partial_{\tilde{z}} \tilde{J}_\infty^\gamma(\tilde{z}, v) &= -(v - p)(\chi\lambda'(\tilde{z}) + \chi\tilde{z}\lambda''(\tilde{z}) + \chi_0 e^{\lambda(\tilde{z})}\lambda'(\tilde{z})). \end{aligned} \quad (\text{B.19})$$

When $v = 0$, by Assumption 4.1, one can easily see that

$$\begin{aligned} \lim_{\tilde{z} \nearrow 0} \tilde{J}_\infty^\gamma(\tilde{P}_\gamma^*; \tilde{z}, 0) &= 0, \quad \lim_{\tilde{z} \rightarrow -\infty} \tilde{J}_\infty^\gamma(\tilde{P}_\gamma^*; \tilde{z}, 0) = -\infty; \\ \lim_{\tilde{z} \nearrow 0} \tilde{J}_\infty^\gamma(\tilde{z}, 0) &= -p < 0, \quad \lim_{\tilde{z} \rightarrow -\infty} \tilde{J}_\infty^\gamma(\tilde{z}, 0) = \infty; \quad \partial_{\tilde{z}} \tilde{J}_\infty^\gamma(\tilde{z}, 0) < 0, \quad \tilde{z} < 0. \end{aligned} \quad (\text{B.20})$$

Based on the first and third limits in (B.20), we have that $\tilde{J}_\infty^\gamma(\tilde{P}_\gamma^*; \tilde{z}, 0) > 0$ when $\tilde{z} < 0$ is close to 0. Along with the second limit, this implies that the function $\tilde{J}_\infty^\gamma(\tilde{P}_\gamma^*; \tilde{z}, 0)$ has a maximum point $\tilde{Z}_\gamma^*(0) \in (-\infty, 0)$, with $\tilde{J}_\infty^\gamma(\tilde{P}_\gamma^*; \tilde{Z}_\gamma^*(0), 0) > 0$. Clearly, $\partial_{\tilde{z}} \tilde{J}_\infty^\gamma(\tilde{P}_\gamma^*; \tilde{Z}_\gamma^*(0), 0) = 0$. Also, from the second line of (B.20), we see that $\tilde{J}_\infty^\gamma(\tilde{z}, 0)$, and hence $\partial_{\tilde{z}} \tilde{J}_\infty^\gamma(\tilde{P}_\gamma^*; \tilde{z}, 0)$, has a unique zero. Thus, $\tilde{Z}_\gamma^*(0)$ is the unique maximum point of $\tilde{J}_\infty^\gamma(\tilde{P}_\gamma^*; \cdot, 0)$ in $(-\infty, 0)$.

Similarly, when $v = 1$, we have

$$\begin{aligned} \lim_{\tilde{z} \searrow 0} \tilde{J}_\infty^\gamma(\tilde{P}_\gamma^*; \tilde{z}, 1) &= 0, \quad \lim_{\tilde{z} \rightarrow \infty} \tilde{J}_\infty^\gamma(\tilde{P}_\gamma^*; \tilde{z}, 1) = -\infty; \\ \lim_{\tilde{z} \searrow 0} \tilde{J}_\infty^\gamma(\tilde{z}, 1) &= 1 - p > 0, \quad \lim_{\tilde{z} \rightarrow \infty} \tilde{J}_\infty^\gamma(\tilde{z}, 1) = -\infty; \quad \partial_{\tilde{z}} \tilde{J}_\infty^\gamma(\tilde{z}, 1) < 0, \quad \tilde{z} > 0. \end{aligned}$$

By similar observations it follows that $\tilde{J}_\infty^\gamma(\tilde{P}_\gamma^*; \tilde{z}, 1)$ has a unique maximum point $\tilde{Z}_\gamma^*(1) \in (0, \infty)$, with $\tilde{J}_\infty^\gamma(\tilde{P}_\gamma^*; \tilde{Z}_\gamma^*(1), 1) > 0$. Therefore, $\tilde{Z}_\gamma^* = (\tilde{Z}_\gamma^*(0), \tilde{Z}_\gamma^*(1))$ and $\tilde{P}_\gamma^* \equiv p$ together constitute the unique γ -limiting equilibrium.

Case 2 $\gamma = 1/2 \leq \beta$. In this case, by (3.2), (B.16), and (2.3.7) (with $N = 1$), we have

$$\tilde{J}_\infty^\gamma(\tilde{P}_\gamma^*; \tilde{z}, v) = \lim_{N \rightarrow \infty} (\chi e^{-\lambda(N^\gamma - \beta \tilde{z})} - \chi_0) Q_1(\tilde{Z}_\gamma^*; \tilde{z}, v), \quad \tilde{z} \in \mathbb{R}_v. \quad (\text{B.21})$$

When $\beta = 1/2$, this is the same as (B.4) with $N = 1$, and by assertion (i) of Theorem 4.1 we know that there exists a γ -limiting equilibrium.

Now consider the case $\beta > 1/2 = \gamma$. Then $\tilde{J}_\infty^\gamma(\tilde{P}_\gamma^*; \tilde{z}, v) = Q_1(\tilde{Z}_\gamma^*; \tilde{z}, v)$, $\tilde{z} \in \mathbb{R}_v$, which is in the form of (B.4) with $N = 1$ and $\lambda = 0$ (hence violating Assumption 4.1). In particular, (B.9) becomes (using z instead of \tilde{z} and fixed $\zeta > 0$) $F_0^1(\zeta; z) = -\bar{\Phi}_1(\zeta) - z\bar{\Phi}'_1(\zeta)$, $z < 0$, with a unique zero $z_0(\zeta) := -\bar{\Phi}_1(\zeta)/\bar{\Phi}'_1(\zeta) < 0$. We claim that

$$\lim_{\zeta \searrow 0} z_0(\zeta) = -\infty, \quad \lim_{\zeta \rightarrow \infty} z_0(\zeta) = 0. \quad (\text{B.22})$$

From (B.8), the first limit is obvious. To see the second limit, consider an arbitrary $c > 0$ and the function

$$\bar{\Phi}_1^c(\zeta) := \bar{\Phi}_1(\zeta) - c\bar{\Phi}'_1(\zeta) = \mathbb{E}\left[\frac{1 + q(1 - \frac{c\zeta}{\sigma^2})e^{\bar{\eta}_1(\zeta)}}{(1 + qe^{\bar{\eta}_1(\zeta)})^2}\right].$$

Recall that $W \stackrel{\text{d}}{=} \text{Normal}(0, \sigma^2)$. By (B.8), $\bar{\eta}_1(\zeta) = \zeta(\zeta - 2W)/(2\sigma^2) \stackrel{\text{d}}{=} \text{Normal}(\zeta^2/(2\sigma^2), \zeta^2/\sigma^2)$. After some rearrangement, we obtain

$$\begin{aligned} \bar{\Phi}_1^c(\zeta) &= \int_{\mathbb{R}} \frac{1 + q(1 - \frac{c\zeta}{\sigma^2})e^x}{(1 + qe^x)^2} \frac{\sigma}{\sqrt{2\pi}\zeta} \exp\left(-\frac{\sigma^2(x - \frac{\zeta^2}{2\sigma^2})^2}{2\zeta^2}\right) dx \\ &= \frac{\sigma e^{-\frac{\zeta^2}{8\sigma^2}}}{\sqrt{2\pi}\zeta} (\bar{A}_1(\zeta) + q(1 - \frac{c\zeta}{\sigma^2})\bar{A}_2(\zeta)), \\ \text{where } \bar{A}_1(\zeta) &:= \int_{\mathbb{R}} \frac{e^{-\frac{\sigma^2}{2\zeta^2}x^2 + \frac{1}{2}x}}{(1 + qe^x)^2} dx, \quad \bar{A}_2(\zeta) := \int_{\mathbb{R}} \frac{e^{-\frac{\sigma^2}{2\zeta^2}x^2 + \frac{3}{2}x}}{(1 + qe^x)^2} dx. \end{aligned}$$

Then, note that

$$\lim_{\zeta \rightarrow \infty} \bar{A}_1(\zeta) = \int_{\mathbb{R}} \frac{e^{\frac{1}{2}x}}{(1 + qe^x)^2} dx \in (0, \infty), \quad \lim_{\zeta \rightarrow \infty} \bar{A}_2(\zeta) = \int_{\mathbb{R}} \frac{e^{\frac{3}{2}x}}{(1 + qe^x)^2} dx \in (0, \infty).$$

Thus, it is clear that $\lim_{\zeta \rightarrow \infty} (\bar{A}_1(\zeta) + q(1 - c\zeta/\sigma^2)\bar{A}_2(\zeta)) = -\infty$, which implies that $\bar{\Phi}_1^c(\zeta) < 0$ for $\zeta > 0$ sufficiently large. Thus, $\bar{\Phi}_1(\zeta) - c\bar{\Phi}'_1(\zeta) < 0$ and $z_0(\zeta) = -\bar{\Phi}_1(\zeta)/\bar{\Phi}'_1(\zeta) \geq -c$ for all sufficiently large ζ . Since $c > 0$ is arbitrary, we must have $\lim_{\zeta \rightarrow \infty} z_0(\zeta) = 0$.

Next we analyze $G_1^1(\zeta)$. In particular, (B.14) becomes

$$G_1^1(\zeta) = -(\hat{\Phi}_1(\zeta) - 1) - z_1(\zeta)\hat{\Phi}'_1(\zeta), \quad z_1(\zeta) = \zeta + z_0(\zeta). \quad (\text{B.23})$$

By (B.22), we have that $\lim_{\zeta \searrow 0} z_1(\zeta) = -\infty$ and $\lim_{\zeta \rightarrow \infty} z_1(\zeta) = \infty$. Thus, there exists $\check{\zeta} > 0$ such that $z_1(\check{\zeta}) = 0$ and $z_1(\zeta) > 0$ for all $\zeta > \check{\zeta}$. From (B.23), clearly $G_1^1(\check{\zeta}) > 0$. Also, recalling (B.13) and noting that $\hat{\eta}_1(\zeta) \stackrel{\text{d}}{=} \text{Normal}(-\zeta^2/(2\sigma^2), \zeta^2/\sigma^2)$, in a similar fashion as above,

$$G_1^1(\zeta) = \mathbb{E}\left[\frac{qe^{\hat{\eta}_1(\zeta)}}{1 + qe^{\hat{\eta}_1(\zeta)}} - \frac{qe^{\hat{\eta}_1(\zeta)}\zeta z_1(\zeta)}{\sigma^2(1 + qe^{\hat{\eta}_1(\zeta)})^2}\right] = \frac{\sigma e^{-\frac{\zeta^2}{8\sigma^2}}}{\sqrt{2\pi}\zeta} (\hat{A}_1(\zeta) - \frac{1}{q}(\frac{\zeta z_1(\zeta)}{\sigma^2} - 1)\hat{A}_2(\zeta)),$$

$$\text{where } \hat{A}_1(\zeta) := \int_{\mathbb{R}} \frac{e^{-\frac{\sigma^2}{2\zeta^2}x^2 + \frac{1}{2}x}}{(1 + \frac{e^x}{q})^2} dx, \quad \hat{A}_2(\zeta) := \int_{\mathbb{R}} \frac{e^{-\frac{\sigma^2}{2\zeta^2}x^2 + \frac{3}{2}x}}{(1 + \frac{e^x}{q})^2} dx.$$

Noting that $\lim_{\zeta \rightarrow \infty} \hat{A}_1(\zeta), \lim_{\zeta \rightarrow \infty} \hat{A}_2(\zeta) \in (0, \infty)$, and $\lim_{\zeta \rightarrow \infty} z_1(\zeta) = \infty$, we have $\lim_{\zeta \rightarrow \infty} (\hat{A}_1(\zeta) - (\zeta z_1(\zeta)/\sigma^2)\hat{A}_2(\zeta)) = -\infty$, and so $G_1^1(\zeta) < 0$ for ζ sufficiently large. Since $G_1^1(\check{\zeta}) > 0$, then there exists $\zeta^* > \check{\zeta}$ such that $G_1^1(\zeta^*) = 0$. Thus, following the same arguments as in Step 2 of the proof of assertion (i), $(\mathcal{Z}^{\zeta^*}, P_1(\mathcal{Z}^{\zeta^*}; \cdot))$ is a γ -limiting equilibrium in the sense of Definition 3.1.

Step 2 We now prove the “only if” direction. Assume that there exists a γ -limiting equilibrium.

Case 3 $\gamma < 1/2$. Then, from (3.2), $\tilde{P}_\gamma^* = p$, and so by (B.16) and (B.17),

$$\tilde{J}_\infty^\gamma(\tilde{P}_\gamma^*; \tilde{z}, v) = \begin{cases} (v - p)\tilde{z}, & \text{if } \gamma < \beta, \\ (v - p)\tilde{z}(\chi e^{-\lambda(z)} - \chi_0), & \text{if } \gamma = \beta, \quad \tilde{z} \in \mathbb{R}_v, \\ -\chi_0(v - p)\tilde{z}, & \text{if } \gamma > \beta, \end{cases}$$

Clearly, in the cases $\gamma < \beta$ and $\gamma > \beta$ with $\chi_0 > 0$, $\tilde{J}_\infty^\gamma(\tilde{P}_\gamma^*; \cdot, v)$ has no maximum point in \mathbb{R}_v . In the case $\gamma > \beta$ with $\chi_0 = 0$, $\tilde{J}_\infty^\gamma(\tilde{P}_\gamma^*; \tilde{z}, v) = 0$ for all $\tilde{z} \in \mathbb{R}_v$, but since in Definition 3.1 condition (i) we require $\tilde{J}_\infty^\gamma(\tilde{P}_\gamma^*; \tilde{Z}_\gamma^*(v), v) \neq 0$, there is no such $\tilde{Z}_\gamma^*(v)$. Thus, we must have $\gamma = \beta$, and since $\gamma < 1/2$, we obtain $\gamma = \min\{\beta, 1/2\}$.

Case 4 $\gamma = 1/2$. If $\beta < 1/2$, then by (B.21) and (B.17), $\tilde{J}_\infty^\gamma(\tilde{P}_\gamma^*; \tilde{z}, v) = -\chi_0 Q_1(\tilde{Z}_\gamma^*; \tilde{z}, v)$, $\tilde{z} \in \mathbb{R}_v$. By Lemma B.1, $Q_1(\tilde{Z}_\gamma^*; \tilde{z}, v)$ is strictly log-concave in $z \in \mathbb{R}_v$; hence, if $\chi_0 > 0$, $\tilde{J}_\infty^\gamma(\tilde{P}_\gamma^*; \cdot, v)$ has no maximum point in \mathbb{R}_v , while if $\chi_0 = 0$, then $\tilde{J}_\infty^\gamma(\tilde{P}_\gamma^*; \cdot, v) = 0$, also a violation, and it must be that $\beta \geq 1/2$; thus, $\gamma = \min\{\beta, 1/2\}$.

Case 5 $\gamma > 1/2$. Then, (B.16) and (B.17) imply that

$$\tilde{J}_\infty^\gamma(\tilde{P}_\gamma^*; \tilde{z}, v) = \tilde{z}(v - 1\mathbf{1}_{\{\tilde{z} > z_\gamma^*\}} - \frac{1}{2}\mathbf{1}_{\{\tilde{z} = z_\gamma^*\}}) \lim_{N \rightarrow \infty} (\chi e^{-\lambda(N^{\gamma-\beta}\tilde{z})} - \chi_0), \quad \tilde{z} \in \mathbb{R}_v,$$

with the limit being finite. When $z_\gamma^* \geq 0$, we have $\tilde{J}_\infty^\gamma(\tilde{P}_\gamma^*; \tilde{z}, 0) = 0$ for all $\tilde{z} < 0$, which again violates the requirement $\tilde{J}_\infty^\gamma(\tilde{P}_\gamma^*; \tilde{Z}_\gamma^*(0), 0) \neq 0$. Similarly, when $z_\gamma^* < 0$, we have $\tilde{J}_\infty^\gamma(\tilde{P}_\gamma^*; \tilde{z}, 1) = 0$ for all $\tilde{z} > 0$. Therefore, there is no γ -limiting equilibrium. ■

Proof of Theorem 4.1 assertion (iii). Let Assumption 4.1 still hold. We now prove the convergence of a finite- N equilibrium (Z_N^*, P_N^*) (Definition 2.1) towards the unique γ -limiting equilibrium $(\tilde{Z}_\gamma^*, \tilde{P}_\gamma^* = p)$, in the case $\gamma = \beta < 1/2$. We proceed in three steps.

Step 1 In this step, we show that $\tilde{z}_v^N := N^{-\gamma} Z_N^*(v)$ is bounded in N . Consider $v = 0$ first. From Step 2.1 in the proof of Assertion (i), we see that $z_0^{*,N} < Z_N^*(0) < 0$, where $z_0^{*,N} = z_0^*$ is the unique zero of $g_{1,N}$ defined in (B.9). Noting that

$$g_{1,N}(z) = \chi z N^{-\beta} \lambda'(N^{-\beta} z) - \chi + \chi_0 e^{\lambda(N^{-\beta} z)} = g_{1,1}(N^{-\beta} z), \quad z < 0, \quad (\text{B.24})$$

then $z_0^{*,N} = N^\beta z_0^{*,1}$, where $z_0^{*,1}$ is the unique zero of $g_{1,1}$, depending only on the model parameters. Since $\gamma = \beta$, we have $z_0^{*,1} \leq \tilde{z}_0^N < 0$, and so \tilde{z}_0^N is bounded. We shall put $|\tilde{z}_0^N| \leq K_0$, for some constant $K_0 > 0$.

In the same fashion we can prove that $|\tilde{z}_1^N| \leq K_0$ for possibly larger K_0 . Note that one may not use the function G_1^N from (B.14) for this purpose – instead, the following function that is the counterpart of F_0^N in (B.9) for $v = 1$ and satisfies $G_1^N(\zeta) = F_1^N(\zeta; z_1(\zeta))$ should be used:

$$F_1^N(\zeta; z) := \check{J}_N((z - \zeta, z); z, 1) = g_{1,N}(z)(\hat{\Phi}_N(\zeta) - 1) + g_{2,N}(z)\hat{\Phi}'_N(\zeta), \quad z > 0. \quad (\text{B.25})$$

Step 2 In this step, we prove the desired uniform convergence of $F_v^N(N^\gamma \tilde{\zeta}; N^\gamma \tilde{z})$. Recall $\check{J}_\infty^\gamma(\tilde{z}, v)$ in (B.19). For $v = 0$, $\tilde{\zeta} > 0$, and $\tilde{z} < 0$, using (B.9) further we have

$$\begin{aligned} |F_0^N(N^\gamma \tilde{\zeta}; N^\gamma \tilde{z}) - \check{J}_\infty^\gamma(\tilde{z}, 0)| &= \left| g_{1,1}(\tilde{z})\bar{\Phi}_N(N^\gamma \tilde{\zeta}) + g_{2,1}(\tilde{z})N^\gamma \bar{\Phi}'_N(N^\gamma \tilde{\zeta}) - pg_{1,1}(\tilde{z}) \right| \\ &\leq |g_{1,1}(\tilde{z})|\|\bar{\Phi}_N(N^\gamma \tilde{\zeta}) - p\| + |g_{2,1}(\tilde{z})|N^\gamma |\bar{\Phi}'_N(N^\gamma \tilde{\zeta})|. \end{aligned}$$

From Step 1, we know $|\tilde{z}_v^N| \leq K_0$ and $\tilde{\zeta}^N := \tilde{z}_1^N - \tilde{z}_0^N \leq 2K_0$, so from now on we shall restrict to $-K_0 \leq \tilde{z} < 0$ and $0 < \tilde{\zeta} \leq 2K_0$. In particular, this implies that $|g_{1,1}(\tilde{z})|, |g_{2,1}(\tilde{z})| \leq K$; here, we use $K > 0$ to denote a generic constant that depends only on the model parameters, and its value may vary from line to line.

Since $x/(1+x)^2 \leq 1/4$, by (B.8) we have for $\tilde{\zeta} \in (0, 2K_0]$,

$$N^\gamma \bar{\Phi}'_N(N^\gamma \tilde{\zeta}) = \frac{N^{2\gamma} \tilde{\zeta}}{N\sigma^2} \mathbb{E}\left[\frac{qe^{\tilde{\eta}_N(\zeta)}}{(1+qe^{\tilde{\eta}_N(\zeta)})^2}\right] \leq \frac{N^{2\gamma} \tilde{\zeta}}{4N\sigma^2} \leq KN^{2\gamma-1}.$$

Moreover, one can easily verify that $\bar{\Phi}_N(N^\gamma \tilde{\zeta}) = \bar{\Phi}_1(N^{\gamma-1/2} \tilde{\zeta})$, $\bar{\Phi}_1(0) = p$, as well as $\bar{\Phi}'_1(0) = -\mathbb{E}[e^{\tilde{\eta}_1(\zeta)} q(\zeta - W)/\sigma^2 / (1 + qe^{\tilde{\eta}_1(\zeta)})^2] \Big|_{\zeta=0} = 0$. Then, again, for $\tilde{\zeta} \in (0, 2K_0]$, and hence $N^{\gamma-1/2} \tilde{\zeta} \in (0, 2K_0]$, we have

$$|\bar{\Phi}_N(N^\gamma \tilde{\zeta}) - p| = |\bar{\Phi}_1(N^{\gamma-1/2} \tilde{\zeta}) - \bar{\Phi}_1(0)| \leq K|N^{\gamma-1/2} \tilde{\zeta}|^2 \leq KN^{2\gamma-1}.$$

Putting things together, we have shown that

$$\sup_{\tilde{\zeta} \in (0, 2K_0], \tilde{z} \in [-K_0, 0)} |F_0^N(N^\gamma \tilde{\zeta}; N^\gamma \tilde{z}) - \check{J}_\infty^\gamma(\tilde{z}, 0)| \leq KN^{2\gamma-1}. \quad (\text{B.26})$$

It is similar to establish that $\sup_{\tilde{\zeta} \in (0, 2K_0], \tilde{z} \in (0, K_0]} |F_1^N(N^\gamma \tilde{\zeta}; N^\gamma \tilde{z}) - \check{J}_\infty^\gamma(\tilde{z}, 1)| \leq KN^{2\gamma-1}$.

Step 3 We now prove (4.3). Since $F_0^N(N^\gamma \tilde{\zeta}^N; N^\gamma \tilde{z}_0^N) = \check{J}_\infty^\gamma(\tilde{Z}_\gamma^*(0), 0) = 0$, (B.26) implies

$$|\check{J}_\infty^\gamma(\tilde{z}_0^N, 0) - \check{J}_\infty^\gamma(\tilde{Z}_\gamma^*(0), 0)| = |\check{J}_\infty^\gamma(\tilde{z}_0^N, 0) - F_0^N(N^\gamma \tilde{\zeta}^N; N^\gamma \tilde{z}_0^N)| \leq KN^{2\gamma-1}. \quad (\text{B.27})$$

Note that by (B.19),

$$\partial_{\tilde{z}} \check{J}_\infty^\gamma(\tilde{z}, 0) = p(\chi \lambda'(\tilde{z}) + \chi \tilde{z} \lambda''(\tilde{z}) + \chi_0 e^{\lambda(\tilde{z})} \lambda'(\tilde{z})) \leq p\chi \lambda'(\tilde{z}) < 0, \quad \tilde{z} < 0.$$

As λ' is increasing (by Assumption 4.1), $\partial_{\tilde{z}} \check{J}_\infty^\gamma(\tilde{z}, 0) \leq p\chi \lambda'(\tilde{Z}_\gamma^*(0)/2) < 0$ for all $\tilde{z} \leq \tilde{Z}_\gamma^*(0)/2$. Then, denoting $c_0 := p\chi |\lambda'(\tilde{Z}_\gamma^*(0)/2)|$, by the monotonicity of $\check{J}_\infty^\gamma(\tilde{z}, 0)$,

$$|\check{J}_\infty^\gamma(\tilde{z}, 0) - \check{J}_\infty^\gamma(\tilde{Z}_\gamma^*(0), 0)| \geq \frac{1}{2} c_0 |\tilde{Z}_\gamma^*(0)| > 0, \quad \tilde{z} \in [\frac{1}{2} \tilde{Z}_\gamma^*(0), 0].$$

By (B.27), this further implies that for N large such that $KN^{2\gamma-1} < c_0|\tilde{Z}_\gamma^*(0)|/2$, we must have $\tilde{z}_0^N \leq \tilde{Z}_\gamma^*(0)/2$. Thus, with $c_0 \leq |\partial_{\tilde{z}} \tilde{J}_\infty^\gamma(\tilde{z}, 0)|$ for all $\tilde{z} \leq \tilde{Z}_\gamma^*(0)/2$,

$$KN^{2\gamma-1} \geq |\tilde{J}_\infty^\gamma(\tilde{z}_0^N, 0) - \tilde{J}_\infty^\gamma(\tilde{Z}_\gamma^*(0), 0)| \geq c_0|\tilde{z}_0^N - \tilde{Z}_\gamma^*(0)|,$$

which immediately implies that

$$|\tilde{z}_0^N - \tilde{Z}_\gamma^*(0)| \leq \frac{K}{c_0} N^{2\gamma-1} \leq KN^{2\gamma-1}.$$

It is similar to show that $|\tilde{z}_1^N - \tilde{Z}_\gamma^*(1)| \leq KN^{2\gamma-1}$, justifying (4.3).

For (4.4), note that

$$P_N^*(y) = P_N(Z_N^*; y) = \tilde{P}_N^\gamma(N^{-\gamma} Z_N^*; N^{-\max\{\gamma, \frac{1}{2}\}} y) = \frac{1}{1 + qe^{\tilde{\kappa}_N(\tilde{y})}},$$

where $\tilde{\kappa}_N$ is as in (A.1) corresponding to $\tilde{Z} = N^{-\gamma} Z_N^*$ and $\tilde{y} = N^{-1/2} y$. By Step 1, $|\tilde{Z}| \leq K_0$. Since $\gamma < 1/2$, then it follows from (A.2) that

$$|\tilde{\kappa}_N(\tilde{y})| \leq K(|\tilde{y}|N^{\gamma-\frac{1}{2}} + N^{2\gamma-1}) = K(|y|N^{\gamma-1} + N^{2\gamma-1}).$$

Therefore, recalling $q = (1-p)/p$, we have

$$|P_N^*(y) - p| = \left| \frac{1}{1 + qe^{\tilde{\kappa}_N(\tilde{y})}} - p \right| = (1-p) \frac{|1 - e^{\tilde{\kappa}_N(\tilde{y})}|}{1 + qe^{\tilde{\kappa}_N(\tilde{y})}} \leq K(|y|N^{\gamma-1} + N^{2\gamma-1}),$$

proving the first estimate in (4.4). Finally, the last bound, along with the boundedness of $N^{-\gamma} Z_N^*(v)$ established in Step 1, immediately leads to the second estimate in (4.4), completing the proof. ■

Appendix C Proof of Theorem 4.2

We have $\gamma = \beta < 1/2$. First, (4.1) and (B.18) together imply that, for $v \in \{0, 1\}$ and $z \in \mathbb{R}_v$,

$$N^{-\gamma} J_N(p; z, v) = (v-p)N^{-\gamma} z(\chi e^{-\lambda(N^{-\gamma} z)} - \chi_0) = \tilde{J}_\infty^\gamma(p; N^{-\gamma} z, v).$$

Then, based on Definition 3.1, since

$$N^{-\gamma} J_N(p; N^\gamma \tilde{Z}_\gamma^*(v), v) = \tilde{J}_\infty^\gamma(p; \tilde{Z}_\gamma^*(v), v) = \sup_{z \in \mathbb{R}_v} \tilde{J}_\infty^\gamma(p; N^{-\gamma} z, v) = \sup_{z \in \mathbb{R}_v} N^{-\gamma} J_N(p; z, v),$$

we have that (4.6) in Definition 4.1 holds with $\epsilon = 0$.

Next, since $\max\{\gamma, 1/2\} = 1/2$, by (A.1) we have $P_N(N^\gamma \tilde{Z}_\gamma^*; N^{1/2} \tilde{y}) = 1/(1 + qe^{\tilde{\kappa}_N(\tilde{y})})$, where $\tilde{\kappa}_N$ corresponds to $\tilde{Z} = \tilde{Z}_\gamma^*$. As \tilde{Z}_γ^* is independent of N , it follows from the exact same arguments for (4.3) that

$$|P_N(N^\gamma \tilde{Z}_\gamma^*; N^{\frac{1}{2}} \tilde{y}) - p| \leq K(|N^{\frac{1}{2}} \tilde{y}|N^{\gamma-1} + N^{2\gamma-1}) \leq KN^{\gamma-\frac{1}{2}}(|\tilde{y}| + 1),$$

which verifies (4.7) with $\epsilon \equiv \epsilon_N = KN^{\gamma-1/2}$, hence (4.8). The proof is complete. ■

Supplemental Appendix A Proofs in further discussions

This supplemental appendix presents formal statements and their detailed proofs in the further discussions on the general scenario involving criminal penalties in Section 6. The proofs are structurally similar to that of Theorem 4.1 and Theorem 4.2 as presented in Appendix B and Appendix C, but with notable technical differences.

Theorem SA.1. Consider the setting of (6.1), and let Assumption 6.1 hold. We have the following four assertions.

- (i) For every $N \geq 1$, there exists an equilibrium (Z_N^*, P_N^*) in the sense of Definition 2.1.
- (ii) There exists a γ -limiting equilibrium $(\tilde{Z}_\gamma^*, \tilde{P}_\gamma^*)$ in the sense of Definition 3.1 if and only if (6.5) holds. Moreover, when $\gamma < 1/2$, the γ -limiting equilibrium $(\tilde{Z}_\gamma^*, \tilde{P}_\gamma^*)$ is unique with $\tilde{P}_\gamma^* = p$.
- (iii) Let $\gamma < 1/2$ and $(\tilde{Z}_\gamma^*, \tilde{P}_\gamma^* \equiv p)$ be the unique γ -limiting equilibrium from assertion (ii). Then, there exists a constant $K > 0$, depending only on the model parameters but not on N , such that, for any equilibrium (Z_N^*, P_N^*) from assertion (i),

$$|N^{-\gamma} Z_N^*(v) - \tilde{Z}_\gamma^*(v)| \leq \begin{cases} KN^{2\gamma-1}, & \text{if } \beta = 0 \text{ or } \alpha = 1, \\ KN^{\max\{2\gamma-1, -\gamma(\alpha-1)\theta', -\gamma\alpha'\}}, & \text{if } \beta > 0 \text{ and } \alpha > 1, \end{cases} \quad v \in \{0, 1\}, \quad (\text{SA.1})$$

and, with $\tilde{P}_\gamma^* = p$, the estimates in (4.4) stand.

- (iv) As in assertion (iii), let $\gamma < 1/2$ and let (\tilde{Z}_γ^*, p) be the unique γ -limiting equilibrium. For every $N \geq 1$, $(N^\gamma \tilde{Z}_\gamma^*, p)$ is an ϵ_N -equilibrium in the sense of Definition 4.1, where, for a constant K depending only on the model parameters,

$$\epsilon_N = \begin{cases} KN^{\gamma-\frac{1}{2}}, & \text{if } \beta = 0 \text{ or } \alpha = 1, \\ KN^{\max\{\gamma-\frac{1}{2}, -\gamma(\alpha-1), -\gamma(\alpha-1)\theta', -\gamma\alpha'\}}, & \text{if } \beta > 0 \text{ and } \alpha > 1. \end{cases} \quad (\text{SA.2})$$

Before proving Theorem SA.1, we make a few additional comments on the convergence implications. Compared with Theorem 4.1, assertion (iii) in Theorem SA.1 indicates the joint dependence of the convergence rate for the optimal strategies on both the intensity of investigation (β) and the severity of criminal penalties (α), signaling a tradeoff between small and large values of γ . Within the maximum brackets in (SA.1), the first component $(2\gamma - 1)$ stems from the convergence of the price function, whereas the second and third components $(-\gamma(\alpha-1)\theta' \text{ or } -\gamma\alpha')$ is a result of the large argument behavior for C_0 (as stated in Assumption 6.1). Intuitively, when investigations are sensitive to the population size ($\beta > 0$), the imposition of criminal penalties upon prosecution can substantially alter the convergence rate depending on their specific form. In this case, the jump in the convergence rate as $\alpha \searrow 1$ in (SA.1) results from the at-most-linear growth of the insider's expected profit with his trade size (see (6.2)), which can be coalesced into the penalty component if and only if $\alpha = 1$. Further details can be found in the proof of Theorem SA.1.

To prepare for the proof of Theorem SA.1, we establish a technical lemma.

Lemma SA.1. Consider the setting of Lemma B.1, and define

$$R_N(Z; z, v) := e^{\lambda_N(z)} \partial_z ((\chi e^{-\lambda_N(z)} - \chi_0) Q_N(Z; z, v)), \quad z \in \mathbb{R}_v, \quad v \in \{0, 1\}. \quad (\text{SA.3})$$

Then, the following two assertions hold.

- (i) $R_N(Z; z, 0)$, $z < 0$, admits a unique zero $\check{z}_0 < 0$. Also, $R_N(Z; z, 0) > 0$ for $z < \check{z}_0$, and $R_N(Z; z, 0)$ is negative and strictly decreasing for $z \in (\check{z}_0, 0)$.
- (ii) $R_N(Z; z, 1)$, $z > 0$, admits a unique zero $\check{z}_1 > 0$. Also, $R_N(Z; z, 1) < 0$ for $z > \check{z}_1$, and $R_N(Z; z, 1)$ is positive and strictly decreasing for $z \in (0, \check{z}_1)$.

Proof. We only consider the case $v = 0$, with $z < 0$; the case $v = 1$ (with $z > 0$) can be proved in the same way. As before, without loss of generality, we assume in the proof that $\lambda(z)$ is twice-differentiable for $z \neq 0$, and for notational simplicity we suppress the symbol Z in all the functions.

By Step 1 in the proof of Theorem 4.1 assertion (i), the function $(\chi e^{-\lambda_N(z)} - \chi_0)Q_N(z, 0)$, $z < 0$, admits a unique maximum point $\check{z}_0 < 0$, and it satisfies $R_N(\check{z}_0, 0) = 0$, $R_N(z, 0) > 0$ for $z < \check{z}_0$, and $R_N(z, 0) < 0$ for $z \in (\check{z}_0, 0)$. Moreover, since $Q_N > 0$, it is clear that $\chi e^{-\lambda_N(\check{z}_0)} - \chi_0 > 0$, and then by the monotonicity of λ_N , we see that $\chi e^{-\lambda_N(z)} - \chi_0 > 0$ for all $z \in (\check{z}_0, 0)$.

It remains to show that $R_N(z, 0)$ is strictly decreasing in $(\check{z}_0, 0)$. Let us write

$$H_N(z) := (\chi - \chi_0 e^{\lambda_N(z)})\Phi_N(z) > 0, \quad z \in (\check{z}_0, 0). \quad (\text{SA.4})$$

Then, $(\chi e^{-\lambda_N(z)} - \chi_0)Q_N(z, 0) = (-z)e^{-\lambda_N(z)}H_N(z)$, and thus, for $z \in (\check{z}_0, 0)$,

$$R_N(z, 0) = (z\lambda'_N(z) - 1)H_N(z) - zH'_N(z) < 0; \quad (\text{SA.5})$$

$$\partial_z R_N(z, 0) = (\lambda'_N(z) + z\lambda''_N(z))H_N(z) + (z\lambda'_N(z) - 2)H'_N(z) - zH''_N(z). \quad (\text{SA.6})$$

Recall that $\chi - \chi_0 e^{\lambda_N(z)} > 0$ for $z \in (\check{z}_0, 0)$, and similar to (B.5), we have

$$\frac{d^2}{dz^2} \log(\chi - \chi_0 e^{\lambda_N(z)}) = -\frac{\chi_0 e^{\lambda_N(z)}((\lambda'_N(z))^2 + \lambda''_N(z))}{\chi - \chi_0 e^{\lambda_N(z)}} - \frac{(\chi_0 \lambda'_N(z) e^{\lambda_N(z)})^2}{(\chi - \chi_0 e^{\lambda_N(z)})^2} \leq 0,$$

implying that $\chi - \chi_0 e^{\lambda_N(z)}$ is log-concave. From the proof of Lemma B.1, we know that Φ_N is also log-concave, and so H_N is log-concave in $(\check{z}_0, 0)$. This implies that $H_N(z)H''_N(z) - (H'_N(z))^2 \leq 0$. Hence, from (SA.6) we obtain

$$\begin{aligned} \partial_z R_N(z, 0) &\leq (\lambda'_N(z) + z\lambda''_N(z))H_N(z) + (z\lambda'_N(z) - 2)H'_N(z) - z \frac{(H'_N(z))^2}{H_N(z)} \\ &= (\lambda'_N(z) + z\lambda''_N(z))H_N(z) + \frac{H'_N(z)}{H_N(z)}((z\lambda'_N(z) - 2)H_N(z) - zH'_N(z)) \\ &= (\lambda'_N(z) + z\lambda''_N(z))H_N(z) + \frac{H'_N(z)}{H_N(z)}(R_N(z, 0) - H_N(z)), \end{aligned}$$

where the last line is due to (SA.5). Note further that for $z \in (\check{z}_0, 0)$,

$$H'_N(z) := -\chi_0 e^{\lambda_N(z)}\lambda'_N(z)\Phi_N(z) + (\chi - \chi_0 e^{\lambda_N(z)})\Phi'_N(z) > 0,$$

and thus, by (SA.5) again and Assumption 6.1, $\partial_z R_N(z, 0) < 0$ for $z \in (\check{z}_0, 0)$. ■

Proof of Theorem SA.1 assertion (i). Consider the equilibrium in Definition 2.1 for fixed $N \geq 1$. Similar to the proof of Theorem 4.1 assertion (i), we proceed in two steps.

Step 1 In this step we show that, given $Z = (Z(0), Z(1)) \in \mathbb{R}_0 \times \mathbb{R}_1$, the insider's objective function $\check{J}_N(Z; z, v)$ from (2.3.8) has a unique maximum point $z_v^* \in \mathbb{R}_v$. Without loss of generality, assume as before that λ and C_0 are twice-differentiable for $z \neq 0$, and again we shall only consider the case $v = 0$.

First, by (6.1) and (B.4) we have that, for $z < 0$,

$$\check{J}_N(Z; z, 0) = (\chi e^{-\lambda_N(z)} - \chi_0) Q_N(Z; z, v) - (1 - e^{-\lambda_N(z)}) C_0(z).$$

Then, recalling (SA.3), we have

$$\check{J}_N(Z; z, 0) := e^{\lambda_N(z)} \partial_z \check{J}_N(Z; z, 0) = R_N(Z; z, 0) + A_N(z), \quad (\text{SA.7})$$

$$\text{where } A_N(z) := -\lambda'_N(z) C_0(z) - (e^{\lambda_N(z)} - 1) C'_0(z). \quad (\text{SA.8})$$

By Assumption 6.1, it is clear that $A_N(z) \geq 0$ for $z < 0$, and

$$A'_N(z) = -\lambda''_N(z) C_0(z) - (e^{\lambda_N(z)} + 1) \lambda'_N(z) C'_0(z) - (e^{\lambda_N(z)} - 1) C''_0(z) \leq 0. \quad (\text{SA.9})$$

Let \check{z}_0 be as in Lemma SA.1. Then, it follows from (SA.7) that

$$\check{J}_N(Z; z, 0) > 0, \quad z \leq \check{z}_0 \quad \text{and} \quad \partial_z \check{J}_N(Z; z, 0) < 0, \quad z \in (\check{z}_0, 0). \quad (\text{SA.10})$$

Note further that, by (SA.5) and (SA.4),

$$\lim_{z \nearrow 0} \check{J}_N(Z; z, 0) = \lim_{z \nearrow 0} R_N(Z; z, 0) = -\Phi_N(Z; 0) < 0, \quad (\text{SA.11})$$

which means that $\check{J}_N(Z; \cdot, 0)$ has a unique zero $z_0^* \in (\check{z}_0, 0)$, and by (SA.10), $\check{J}_N(Z; z, 0) > 0$ for $z < z_0^*$ and $\check{J}_N(Z; z, 0) < 0$ for $z \in (z_0^*, 0)$. By continuity, this implies that z_0^* is the unique maximum point of $\check{J}_N(Z; \cdot, 0)$ on $\mathbb{R}_0 = (-\infty, 0)$.

Step 2 This step is concerned with the existence of a finite- N equilibrium (Definition 2.1). As established in Step 1, $Z_N^* = (Z_N^*(0), Z_N^*(1)) \in \mathbb{R}_0 \times \mathbb{R}_1$ is an equilibrium strategy, i.e., $(Z_N^*, P_N(Z_N^*; \cdot))$ is an equilibrium, if and only if it satisfies the equilibrium conditions

$$\check{J}_N(Z_N^*; Z_N^*(v), v) = R_N(Z_N^*; Z_N^*(v), v) + A_N(Z_N^*(v)) = 0, \quad v \in \{0, 1\}, \quad (\text{SA.12})$$

again with $Z_N^*(0), Z_N^*(1) \neq 0$.

Consider (SA.12) for $v = 0$. With $\zeta_N^* = Z_N^*(1) - Z_N^*(0) \geq 0$, fixing arbitrary $\zeta > 0$ as a parameter, we introduce the following function on $(-\infty, 0)$:

$$F_0^N(\zeta; z) := \check{J}_N((z, \zeta + z); z, 0) = R_N((z, \zeta + z); z, 0) + A_N(z), \quad z < 0. \quad (\text{SA.13})$$

Recalling (B.8) and (B.9), we have

$$F_0^N(\zeta; z) = g_{1,N}(z) \bar{\Phi}_N(\zeta) + g_{2,N}(z) \bar{\Phi}'_N(\zeta) + A_N(z), \quad z < 0. \quad (\text{SA.14})$$

By Step 2.1 in the proof of Theorem 4.1 (i), $g_{1,N}$ has a unique zero $z_0^\diamond < 0$ such that $g'_{1,N}(z) < 0$ and $g'_{2,N}(z) < g_{1,N}(z) < 0 < g_{2,N}(z)$ for $z \in (z_0^\diamond, 0)$. Since $\bar{\Phi}_N(\zeta) > 0$, $\bar{\Phi}'_N(\zeta) > 0$, and $A_N(z) \geq 0$, $A'_N(z) \leq 0$, then $F_0^N(\zeta; z_0^\diamond) > 0$ and $\partial_z F_0^N(\zeta; z) < 0$ for $z \in (z_0^\diamond, 0)$. Similarly as in (SA.11), we have

$\lim_{z \nearrow 0} F_0^N(\zeta; z) = -\bar{\Phi}_N(\zeta) < 0$, and so $F_0^N(\zeta; \cdot)$ has a unique zero in $(z_0^\diamond, 0)$, denoted as $x_0(\zeta)$, which is a continuous function of ζ . Moreover, since $\bar{\Phi}_N(0) = 1$ and $\bar{\Phi}'_N(0) = 0$, we have

$$0 = \lim_{\zeta \searrow 0} F_0^N(\zeta; x_0(\zeta)) = \lim_{\zeta \searrow 0} (g_{1,N}(x_0(\zeta)) + A_N(x_0(\zeta))).$$

Again, noting that $g'_{1,N} < 0$ and $A'_N \leq 0$, then by (B.9) and (SA.8), $\lim_{z \nearrow 0} (g_{1,N}(z) + A_N(z)) = -1 < 0$, and so we have $\lim_{\zeta \searrow 0} x_0(\zeta) < 0$.

Next, consider (SA.12) for $v = 1$. As in the proof of Theorem 4.1 assertion (i), we denote $x_1(\zeta) := \zeta + x_0(\zeta)$ and $\mathcal{Z}(\zeta) := (x_0(\zeta), x_1(\zeta))$. Recalling (B.13), by (B.14),

$$G_1^N(\zeta) := \tilde{J}_N(\mathcal{Z}^\zeta; x_1(\zeta), 1) = g_{1,N}(x_1(\zeta))(\hat{\Phi}_N(\zeta) - 1) + g_{2,N}(x_1(\zeta))\hat{\Phi}'_N(\zeta) + A_N(x_1(\zeta)),$$

where $\hat{\Phi}_N$ and $\hat{\Phi}'_N$ are given in (B.13). Since $\lim_{\zeta \searrow 0} x_0(\zeta) < 0$ and $x_0(\zeta) \in (z_0^\diamond, 0)$, $\lim_{\zeta \searrow 0} x_1(\zeta) < 0$ and $\lim_{\zeta \rightarrow \infty} x_1(\zeta) = \infty$. Hence, there exists $\check{\zeta} > 0$ such that $x_1(\check{\zeta}) = 0$. Let $\check{\zeta}$ be the largest one having this property, so that $x_1(\zeta) > 0$ for all $\zeta > \check{\zeta}$. Note that $\lim_{z \nearrow 0} g_{1,N}(z) = -1$, $\lim_{z \nearrow 0} g_{2,N}(z) = 0$, and $\lim_{z \nearrow 0} A_N(z) = 0$, and so $G_1^N(\check{\zeta}) = 1 - \hat{\Phi}_N(\check{\zeta}) > 0$. Moreover, similarly to Step 2.2 in the proof of Theorem 4.1 assertion (i), there exists $\hat{\zeta} > \check{\zeta}$ such that $g_{1,N}(x_1(\hat{\zeta})) = 0$ and $g_{2,N}(x_1(\hat{\zeta})) < 0$. Note that $\hat{\Phi}'_N(\hat{\zeta}) > 0$ and, by (SA.8), $A_N(x_1(\hat{\zeta})) \leq 0$ (as $x_1(\hat{\zeta}) > 0$), and then $G_1^N(\hat{\zeta}) < 0$. Thus, there exists $\zeta^* \in (\check{\zeta}, \hat{\zeta})$ such that $G_1^N(\zeta^*) = 0$. Therefore, $(\mathcal{Z}^{\zeta^*}, P_N(\mathcal{Z}^{\zeta^*}; \cdot))$, with $\mathcal{Z}^{\zeta^*} = (x_0(\zeta^*), x_1(\zeta^*))$, is a finite- N equilibrium in the sense of Definition 2.1. ■

Proof of Theorem SA.1 assertion (ii). As before, we prove the “if” direction and the “only if” direction separately. In a similar fashion as in (B.16), especially for $\tilde{\Phi}_N^\gamma(\tilde{Z}_\gamma^*; \tilde{z})$, given any γ -limiting equilibrium $(\tilde{Z}_\gamma^*, \tilde{P}_\gamma^*)$, we have that for $\tilde{z} \in \mathbb{R}_v$, $v \in \{0, 1\}$,

$$\tilde{J}_\infty^\gamma(\tilde{P}_\gamma^*; \tilde{z}, v) = \lim_{N \rightarrow \infty} \left(\tilde{z} (\chi e^{-\lambda(N^\gamma - \beta \tilde{z})} - \chi_0)(v - \tilde{\Phi}_N^\gamma(\tilde{Z}_\gamma^*; \tilde{z})) - (1 - e^{-\lambda(N^\gamma - \beta \tilde{z})})N^{-\gamma}C_0(N^\gamma \tilde{z}) \right). \quad (\text{SA.15})$$

Let us recall (B.16) and the limits in (B.17). Also, under Assumption 6.1, the following additional limits are easily justified: For $\tilde{z} \neq 0$,

$$\lim_{N \rightarrow \infty} N^{-\gamma}C_0(N^\gamma \tilde{z}) = \begin{cases} C_0(\tilde{z}), & \text{if } \gamma = 0, \\ K_\alpha |\tilde{z}|, & \text{if } \gamma > 0, \alpha = 1, \\ \infty, & \text{if } \gamma > 0, \alpha > 1; \end{cases} \quad (\text{SA.16})$$

$$\lim_{N \rightarrow \infty} (1 - e^{-\lambda(N^\gamma - \beta \tilde{z})})N^{-\gamma}C_0(N^\gamma \tilde{z}) = \begin{cases} 0, & \text{if } 0 < \gamma < \frac{\beta\theta}{\theta+\alpha-1}, \\ K_\theta K_\alpha |\tilde{z}|^{\theta+\alpha}, & \text{if } \gamma = \frac{\beta\theta}{\theta+\alpha-1} < \beta, \\ \infty, & \text{if } \frac{\beta\theta}{\theta+\alpha-1} < \gamma < \beta. \end{cases} \quad (\text{SA.17})$$

Step 1 In this step we prove the “if” direction in two cases. Assume (6.5) holds, and note that $\beta\theta/(\theta + \alpha - 1) \leq \beta$, with equality holding if and only if $\beta = 0$ or $\alpha = 1$.

Case 1 $\gamma = \beta\theta/(\theta + \alpha - 1) < 1/2$. We must set $\tilde{P}_\gamma^* = p$, and it is unique.

Case 1.1 $\gamma = \beta = 0$. In this case, (SA.15) becomes

$$\tilde{J}_\infty^\gamma(\tilde{P}_\gamma^*; \tilde{z}, v) = (v - p)\tilde{z} (\chi e^{-\lambda(\tilde{z})} - \chi_0) - (1 - e^{-\lambda(\tilde{z})})C_0(\tilde{z}), \quad \tilde{z} \in \mathbb{R}_v, v \in \{0, 1\}. \quad (\text{SA.18})$$

Analogous to (B.19), for $A_N(\tilde{z})$ (from (SA.8)) with $N = 1$,

$$\tilde{J}_\infty^\gamma(\tilde{z}, v) := e^{\lambda(\tilde{z})} \partial_{\tilde{z}} \tilde{J}_\infty^\gamma(\tilde{P}_\gamma^*; \tilde{z}, v) = (v - p)(\chi - \chi \tilde{z} \lambda'(\tilde{z}) - \chi_0 e^{\lambda(\tilde{z})}) + A_1(\tilde{z});$$

$$\partial_{\tilde{z}} \tilde{J}'_\infty(\tilde{z}, v) = -(v - p)(\chi \lambda'(\tilde{z}) + \chi \tilde{z} \lambda''(\tilde{z}) + \chi_0 e^{\lambda(\tilde{z})} \lambda'(\tilde{z})) + A'_1(\tilde{z}), \quad z < 0. \quad (\text{SA.19})$$

For $v = 0$, under Assumption 6.1, it is clear that all the statements in (B.20) remain true. Thus, by following the same arguments after (B.20) we see that $\tilde{J}'_\infty(\tilde{P}_\gamma^*; \cdot, 0)$ has a unique maximum point $\tilde{Z}_\gamma^*(0) < 0$. For $v = 1$, it is similar that $\tilde{J}'_\infty(\tilde{P}_\gamma^*; \cdot, 1)$ has a unique maximum point $\tilde{Z}_\gamma^*(1) > 0$. Therefore, $\tilde{Z}_\gamma^* = (\tilde{Z}_\gamma^*(0), \tilde{Z}_\gamma^*(1))$, together with $\tilde{P}_\gamma^* \equiv p$, forms the unique γ -limiting equilibrium.

Case 1.2 $\beta > 0$ and $\alpha = 1$. Then, $\gamma = \beta > 0$, and (SA.15) becomes

$$\tilde{J}'_\infty(\tilde{P}_\gamma^*; \tilde{z}, v) = (v - p)\tilde{z}(\chi e^{-\lambda(\tilde{z})} - \chi_0) - (1 - e^{-\lambda(\tilde{z})})K_\alpha |\tilde{z}|, \quad \tilde{z} \in \mathbb{R}_v, \quad (\text{SA.20})$$

which is the same as (SA.18) with $C_0(\tilde{z}) = K_\alpha |\tilde{z}|$. Thus, we claim again that there exists a unique γ -limiting equilibrium.

Case 1.3 $\beta > 0$ and $\alpha > 1$. Then, $\gamma = \beta\theta/(\theta + \alpha - 1) < \beta$, and (SA.15) becomes

$$\tilde{J}'_\infty(\tilde{P}_\gamma^*; \tilde{z}, v) = (v - p)\tilde{z} - K_\theta K_\alpha |\tilde{z}|^{\theta+\alpha}, \quad \tilde{z} \in \mathbb{R}_v. \quad (\text{SA.21})$$

Since $\theta + \alpha > 1$, it is clear that $\tilde{J}'_\infty(\tilde{P}_\gamma^*; \tilde{z}, v)$ above has a unique maximum point $\tilde{Z}_\gamma^*(v) \in \mathbb{R}_v$, and so there exists a unique γ -limiting equilibrium.

Case 2 $\gamma = 1/2 \leq \beta\theta/(\theta + \alpha - 1) \leq \beta$.

Case 2.1 $\beta = \gamma = 1/2$. In this case, we have $\alpha = 1$, and based on Assumption 6.1 conditions (i) and (iii), observe that for some constant $K_\alpha^\gamma \in \mathbb{R}$,

$$C_0(z) \sim K_\alpha |z| + K_\alpha^\gamma, \quad \text{as } |z| \rightarrow \infty. \quad (\text{SA.22})$$

Indeed, when $v = 1$, $C_0(z) - K_\alpha z$ is still convex and $C_0(z) - K_\alpha z = o(z)$ as $z \rightarrow \infty$, and so the only possibility is that $C_0(z) - K_\alpha z$ is constant when $z > 0$ is large. The same idea goes for $v = 0$.

With this observation, (SA.15) specializes to

$$\tilde{J}'_\infty(\tilde{P}_\gamma^*; \tilde{z}, v) = (\chi e^{-\lambda(\tilde{z})} - \chi_0)Q_1(\tilde{Z}_\gamma^*; \tilde{z}, v) - (1 - e^{-\lambda(\tilde{z})})K_\alpha |\tilde{z}|, \quad \tilde{z} \in \mathbb{R}_v, \quad (\text{SA.23})$$

which is a special case of (6.1) with $N = 1$ and $C_0(z) = K_\alpha |z|$. Thus, by assertion (i), a γ -limiting equilibrium exists.

Case 2.2 $\gamma = 1/2 = \beta\theta/(\theta + \alpha - 1) < \beta$. Then, (SA.15) becomes

$$\tilde{J}'_\infty(\tilde{P}_\gamma^*; \tilde{z}, v) = Q_1(\tilde{Z}_\gamma^*; \tilde{z}, v) - K_\theta K_\alpha |\tilde{z}|^{\theta+\alpha}, \quad \tilde{z} \in \mathbb{R}_v.$$

Adapting the analysis in the proof of assertion (i), one can easily check that in this case, the condition (B.9) becomes (using z instead of \tilde{z}):

$$F_0^1(\zeta; z) = -\bar{\Phi}_1(\zeta) - z\bar{\Phi}'_1(\zeta) + K_\theta K_\alpha(\theta + \alpha)|z|^{\theta+\alpha-1}, \quad z < 0.$$

Since $\theta + \alpha > 1$, then clearly,

$$\lim_{z \nearrow 0} F_0^1(\zeta; z) = -\bar{\Phi}_1(\zeta) < 0, \quad \lim_{z \rightarrow -\infty} F_0^1(\zeta; z) = \infty, \quad \partial_z F_0^1(\zeta; z) < 0.$$

Thus, $F_0^1(\zeta; z)$ has a unique zero $z_0(\zeta) < 0$. Also, it is clear that $z_0(\zeta) \geq -\bar{\Phi}_1(\zeta)/\bar{\Phi}'_1(\zeta)$. By (B.22), we have $\lim_{\zeta \rightarrow \infty} z_0(\zeta) = 0$. Moreover, with $\lim_{\zeta \searrow 0} \bar{\Phi}_1(\zeta) = 1$ and $\lim_{\zeta \searrow 0} \bar{\Phi}'_1(\zeta) = 0$, we easily see that $\lim_{\zeta \searrow 0} z_0(\zeta) = -(K_\theta K_\alpha(\theta + \alpha))^{1/(\theta+\alpha-1)} < 0$. In particular, this implies that, for $z_1(\zeta) = \zeta + z_0(\zeta)$,

$$\lim_{\zeta \searrow 0} z_1(\zeta) < 0, \quad \lim_{\zeta \rightarrow \infty} z_1(\zeta) = \infty. \quad (\text{SA.24})$$

On the other hand, the condition (B.23) becomes

$$G_1^1(\zeta) = -(\hat{\Phi}_1(\zeta) - 1) - \alpha_1(\zeta)\hat{\Phi}'_1(\zeta) - K_\theta K_\alpha(\theta + \alpha)|\alpha_1(\zeta)|^{\theta+\alpha-1}. \quad (\text{SA.25})$$

By (SA.24), there exists $\check{\zeta} > 0$ such that $\alpha_1(\check{\zeta}) = 0$, and $\alpha_1(\zeta) > 0$ for all $\zeta > \check{\zeta}$. By (SA.25) and (SA.24), it is also clear that $G_1^1(\check{\zeta}) > 0$ and $\lim_{\zeta \rightarrow \infty} G_1^1(\zeta) = -\infty$, and so there exists $\zeta^* > \check{\zeta}$ such that $G_1^1(\zeta^*) = 0$. Therefore, it follows from the same arguments as in the proof of Theorem 4.1 assertion (i) that, upon setting $\tilde{Z}_\gamma^* := (\alpha_0(\zeta^*), \alpha_1(\zeta^*), (\tilde{Z}_\gamma^*, P_1(\tilde{Z}_\gamma^*; \cdot)))$ is a γ -limiting equilibrium.

Case 2.3 $\gamma = 1/2 < \beta\theta/(\theta + \alpha - 1) \leq \beta$. Then, (SA.15) reduces to $\tilde{J}_\infty^\gamma(\tilde{P}_\gamma^*; \tilde{z}, v) = Q_1(\tilde{Z}_\gamma^*; \tilde{z}, v)$ for $\tilde{z} \in \mathbb{R}_v$, which has been analyzed in Case 2 in the proof of Theorem 4.1 assertion (ii), so there exists a γ -limiting equilibrium.

Step 2 We now prove the “only if” direction. Assuming that there is a γ -limiting equilibrium $(\tilde{Z}_\gamma^*, \tilde{P}_\gamma^*)$, our goal is to prove that γ satisfies (6.5). Notably, the limit (SA.16) is used when $\gamma \geq \beta$, while (SA.17) is used when $\gamma < \beta$, and so we may exclude the cases ($\gamma \geq \beta, \gamma > 0, \alpha > 1$) and $\beta\theta/(\theta + \alpha - 1) < \gamma < \beta$. Now, if $\beta\theta/(\theta + \alpha - 1) < \gamma$, then by the second excluded case, we have $\gamma \geq \beta$, which by the first excluded case further implies $\alpha = 1$. Hence, we can focus on the following two cases:

$$\gamma \leq \frac{\beta\theta}{\theta + \alpha - 1} \leq \beta \quad \text{or} \quad \alpha = 1, \gamma > \beta = \frac{\beta\theta}{\theta + \alpha - 1}. \quad (\text{SA.26})$$

Case 3 $\gamma < 1/2$. In the second case of (SA.26), with $\tilde{P}_\gamma^* = p$, (SA.15) gives

$$\tilde{J}_\infty^\gamma(\tilde{P}_\gamma^*; \tilde{z}, 1) = -\chi_0 \tilde{z} (1 - p) - K_\alpha |\tilde{z}| = -(\chi_0(1 - p) + K_\alpha) \tilde{z}, \quad \tilde{z} > 0,$$

which has no maximum point in $(0, \infty)$, and so we must have the first case of (SA.26), but if $\gamma < \beta\theta/(\theta + \alpha - 1) \leq \beta$, then (SA.15) becomes $\tilde{J}_\infty^\gamma(\tilde{P}_\gamma^*; \tilde{z}, v) = \tilde{z}(v - p)$, with no maximum point in \mathbb{R}_v either. Thus, it can only be that $\gamma = \beta\theta/(\theta + \alpha - 1)$, which verifies (6.5) because $\gamma < 1/2$ here.

Case 4 $\gamma = 1/2$. Suppose $\beta\theta/(\theta + \alpha - 1) < 1/2 = \gamma$. Then, we have the second case of (SA.26), and (SA.15) becomes

$$\tilde{J}_\infty^\gamma(\tilde{P}_\gamma^*; \tilde{z}, v) = -\chi_0 Q_1(\tilde{Z}_\gamma^*; \tilde{z}, v) - K_\alpha |\tilde{z}|, \quad \tilde{z} \in \mathbb{R}_v.$$

Note that $Q_1(\tilde{Z}_\gamma^*; \tilde{z}, v) > 0$ for $\tilde{z} \neq 0$. Then, $\tilde{J}_\infty^\gamma(\tilde{P}_\gamma^*; \tilde{z}, v) < 0$ for $\tilde{z} \neq 0$, while $\lim_{\tilde{z} \rightarrow 0} \tilde{J}_\infty^\gamma(\tilde{P}_\gamma^*; \tilde{z}, v) = 0$ clearly, which means that $\tilde{J}_\infty^\gamma(\tilde{P}_\gamma^*; \cdot, v)$ has no maximum point in $\mathbb{R} \setminus \{0\}$. Therefore, we must have $\beta\theta/(\theta + \alpha - 1) \geq 1/2 = \gamma$, verifying (6.5).

Case 5 $\gamma > 1/2$. We consider two subcases. Again, we write $z_\gamma^* := (\tilde{Z}_\gamma^*(0) + \tilde{Z}_\gamma^*(1))/2$.

Case 5.1 $z_\gamma^* \geq 0$. Note that in this case, for $v = 0$, $-(\mathbb{1}_{\{\tilde{z} > z_\gamma^*\}} + \mathbb{1}_{\{\tilde{z} = z_\gamma^*\}})/2 = 0$ for all $\tilde{z} < 0$. Then, based on (SA.26), by (SA.15) we have that for $\tilde{z} < 0$,

$$\tilde{J}_\infty^\gamma(\tilde{P}_\gamma^*; \tilde{z}, 0) = \begin{cases} 0, & \text{if } \gamma < \frac{\beta\theta}{\theta+\alpha-1} \leq \beta, \\ -K_\theta K_\alpha |\tilde{z}|^{\theta+\alpha}, & \text{if } \gamma = \frac{\beta\theta}{\theta+\alpha-1} < \beta, \\ -(1 - e^{-\lambda(\tilde{z})}) K_\alpha |\tilde{z}|, & \text{if } \gamma = \frac{\beta\theta}{\theta+\alpha-1} < \beta, \\ -K_\alpha |\tilde{z}|, & \text{if } \alpha = 1, \gamma > \beta = \frac{\beta\theta}{\theta+\alpha-1}, \end{cases} \quad (\text{SA.27})$$

and in all these cases, $\tilde{J}_\infty^\gamma(\tilde{P}_\gamma^*; \tilde{Z}_\gamma^*(0), 0) = \sup_{\tilde{z} < 0} \tilde{J}_\infty^\gamma(\tilde{P}_\gamma^*; \tilde{z}, 0) = 0$. This violates the requirement $\tilde{J}_\infty^\gamma(\tilde{P}_\gamma^*; \tilde{Z}_\gamma^*(0), 0) \neq 0$ in Definition 3.1.

Case 5.2 $z_\gamma^* < 0$. In this case, for $v = 1$, $1 - (\mathbb{1}_{\{\tilde{z} > z_\gamma^*\}} + \mathbb{1}_{\{\tilde{z} = z_\gamma^*\}})/2 = \mathbb{1}_{\{\tilde{z} < z_\gamma^*\}} + \mathbb{1}_{\{\tilde{z} = z_\gamma^*\}}/2 = 0$ for all $\tilde{z} > 0$. Then, $\tilde{J}_\infty^\gamma(\tilde{P}_\gamma^*; \tilde{z}, 1)$ has the same expression as in (SA.27) except with $\tilde{z} > 0$, which results in the same violation. Thus, we conclude that no γ -limiting equilibrium exists in the case $\gamma > 1/2$. ■

Proof of Theorem SA.1 assertion (iii). Let Assumption 6.1 hold; by (6.5), let $\gamma = \beta\theta/(\theta + \alpha - 1) < 1/2$. We proceed as in the proof of Theorem 4.1 assertion (iii), particularly focusing on the case $v = 0$.

Step 1 Let us note that in (SA.14), the zero $Z_N^*(0)$ of the function $F_0^N(Z_N^*(1) - Z_N^*(0); \cdot)$ satisfies $z^\diamond < Z_N^*(0) < 0$, where $z^\diamond \equiv z_0^\diamond$ is the unique zero of $g_{1,N}$ defined in (B.9). Write $\tilde{z}_0^N := N^{-\gamma}Z_N^*(0)$ and $\hat{z}_0^N := N^{-\beta}Z_N^*(0)$. By (B.24) and the subsequent arguments, we have that $|\hat{z}_0^N| \leq K_1$ for some $K_1 > 0$; in particular, when $\gamma = \beta$, $\tilde{z}_0^N = \hat{z}_0^N$ is also bounded in N by K_1 .

Next, we justify the boundedness of \tilde{z}_0^N in the case $\gamma < \beta$, which is recalled to imply $\alpha > 1$ and $\gamma > 0$. Denote as before $\zeta_N^* := Z_N^*(1) - Z_N^*(0)$, and note that $F_0^N(\zeta_N^*, Z_N^*(0)) = 0$, where F_0^N is defined in (SA.14). From the arguments after (B.10), we see that $g_{2,N}$ is decreasing in $(z^\diamond, 0)$ with $\lim_{z \nearrow 0} g_{2,N}(z) = 0$, so $g_{2,N}(z) > 0$ for $z \in (z^\diamond, 0)$, and in particular $g_{2,N}(Z_N^*(0)) > 0$. Recall (SA.8) and note that $-\lambda'_N(Z_N^*(0))C_0(Z_N^*(0)) > 0$, $-(e^{\lambda_N(Z_N^*(0))} - 1)C_0'(Z_N^*(0)) > 0$. Also note that $\lambda_N(Z_N^*(0)) = \lambda(\hat{z}_0^N)$ and $\lambda'_N(Z_N^*(0)) = N^{-\beta}\lambda'(\hat{z}_0^N)$. Since for $\zeta > 0$, $0 < \bar{\Phi}_N(\zeta) < 1$ and $\bar{\Phi}'_N(\zeta) > 0$, by (SA.14) and (B.9),

$$\begin{aligned} N^{-\beta}|\lambda'(\hat{z}_0^N)C_0(Z_N^*(0))| &= |\lambda'_N(Z_N^*(0))C_0(Z_N^*(0))| \leq |g_{1,N}(Z_N^*(0))\bar{\Phi}_N(\zeta_N^*)| \\ &\leq |\chi Z_N^*(0)\lambda'_N(Z_N^*(0)) - \chi + \chi_0 e^{\lambda_N(Z_N^*(0))}| = |\chi \hat{z}_0^N \lambda'(\hat{z}_0^N) - \chi + \chi_0 e^{\lambda(\hat{z}_0^N)}| \leq K_2, \end{aligned} \quad (\text{SA.28})$$

for some constant $K_2 > 0$, where we have used the crucial fact that $|\hat{z}_0^N| \leq K_1$. Moreover, by Assumption 6.1 condition (ii), we have $K_3 := \sup_{z \in [-K_1, 0]} |z|^{\theta-1}/|\lambda'(z)| < \infty$, and so $|\lambda'(\hat{z}_0^N)| \geq |\hat{z}_0^N|^{\theta-1}/K_3$. Further note that by assuming without loss of generality that $|\hat{z}_0^N| \geq 1$ (as otherwise we have established the desired boundedness of \hat{z}_0^N), then $|Z_N^*(0)| \geq N^\gamma$, and thus by Assumption 6.1 (iii) we have $C_0(Z_N^*(0)) \geq K_\alpha |Z_N^*(0)|^\alpha/2$ for N large. Therefore, we obtain

$$\begin{aligned} \frac{2K_2 K_3}{K_\alpha} &\geq \frac{2K_2}{K_\alpha} N^{-\beta} |\lambda'(\hat{z}_0^N)C_0(Z_N^*(0))| \geq N^{-\beta} |\hat{z}_0^N|^{\theta-1} |Z_N^*(0)|^\alpha \\ &= N^{-\beta} |N^{\gamma-\beta} \hat{z}_0^N|^{\theta-1} |N^\gamma \hat{z}_0^N|^\alpha = N^{-\beta+(\gamma-\beta)(\theta-1)+\gamma\alpha} |\hat{z}_0^N|^{\theta+\alpha-1} = |\hat{z}_0^N|^{\theta+\alpha-1}, \end{aligned}$$

where the last equality is due to $\gamma = \beta\theta/(\theta + \alpha - 1)$. This immediately implies that $|\hat{z}_0^N| \leq K_0$ for some $K_0 > 0$.

Furthermore, in this case (B.25) generalizes into

$$F_1^N(\zeta; z) := g_{1,N}(z)(\hat{\Phi}_N(\zeta) - 1) + g_{2,N}(z)\hat{\Phi}'_N(\zeta) + A_N(z), \quad z > 0, \quad (\text{SA.29})$$

and in the same manner we can show that $\tilde{z}_1^N := N^{-\gamma}Z_N^*(1)$ is also bounded by K_0 .

Step 2 In this step we prove the uniform convergence of $F_v^N(N^\gamma \tilde{\zeta}; N^\gamma \tilde{z})$, for $0 < \tilde{\zeta} \leq 2K_0$ and $\tilde{z} \in \mathbb{R}_v$ with $|\tilde{z}| \leq K_0$, and we shall only do it for $v = 0$. The case $v = 1$ can be proved analogously. We consider three cases, corresponding to, respectively, Case 1.1, Case 1.2, and Case 1.3 in the proof of assertion (ii).

Case 1 $\gamma = \beta = 0$. Recalling (SA.14) and (SA.19), we have

$$\begin{aligned} |F_0^N(N^\gamma \tilde{\zeta}; N^\gamma \tilde{z}) - \tilde{J}_\infty^\gamma(\tilde{z}, 0)| &\leq I_N + |A_N(N^\gamma \tilde{z}) - A_1(\tilde{z})|, \\ \text{where } I_N &:= |g_{1,N}(N^\gamma) \bar{\Phi}_N(N^\gamma \tilde{\zeta}) + g_{2,N}(N^\gamma \tilde{z}) \bar{\Phi}'_N(N^\gamma \tilde{\zeta}) - p g_{1,1}(\tilde{z})|, \end{aligned}$$

where I_N is exactly the (absolute) difference indicated in (B.26), with $|I_N| \leq KN^{2\gamma-1}$. Moreover, from (SA.8) we see that $A_N(N^\gamma \tilde{z}) = A_1(\tilde{z})$ when $\gamma = \beta = 0$ or $\gamma = \beta > 0$ and $C_0(z) = K_\alpha |z|$. Thus,

$$|F_0^N(N^\gamma \tilde{\zeta}; N^\gamma \tilde{z}) - \tilde{J}_\infty^\gamma(\tilde{z}, 0)| \leq KN^{2\gamma-1}. \quad (\text{SA.30})$$

Case 2 $\gamma = \beta > 0$ and $\alpha = 1$. This is the same as Case 1, with the specialization $C_0(z) = K_\alpha |z|$, and thus (SA.30) remains true.

Case 3. $\beta > 0$ and $\alpha > 1$, so that $\gamma = \beta\theta/(\theta + \alpha - 1) < \beta$. In this case, recalling (SA.21),

$$\tilde{J}_\infty^\gamma(\tilde{z}, 0) := \partial_{\tilde{z}} \tilde{J}_\infty^\gamma(\tilde{P}_\gamma^*; \tilde{z}, v) = -p + K_\theta K_\alpha (\theta + \alpha) |\tilde{z}|^{\theta+\alpha-1}. \quad (\text{SA.31})$$

Then, by (SA.14), (B.9), and (SA.8), we have

$$\begin{aligned} & |F_0^N(N^\gamma \tilde{\zeta}; N^\gamma \tilde{z}) - \tilde{J}_\infty^\gamma(\tilde{z}, 0)| \\ & \leq |g_{1,N}(N^\gamma \tilde{z}) \bar{\Phi}_N(N^\gamma \tilde{\zeta}) + p| + |g_{2,N}(N^\gamma \tilde{z}) \bar{\Phi}'_N(N^\gamma \tilde{\zeta})| \\ & \quad + |- \lambda'_N(N^\gamma \tilde{z}) C_0(N^\gamma \tilde{z}) - (e^{\lambda_N(N^\gamma \tilde{z})} - 1) C'_0(N^\gamma \tilde{z}) - K_\theta K_\alpha (\theta + \alpha) |\tilde{z}|^{\theta+\alpha-1}|. \end{aligned}$$

Recall that $N^\gamma \bar{\Phi}'_N(N^\gamma \tilde{\zeta}) + |\bar{\Phi}_N(N^\gamma \tilde{\zeta}) - p| \leq KN^{2\gamma-1}$ for $0 < \tilde{\zeta} \leq 2K_0$, and in the following K again denotes a generic positive constant that can vary. Moreover, by Assumption 6.1 conditions (ii) and (iii), defining $\tilde{\theta}' := \theta(1 + \theta')$, $\tilde{\alpha}' := \alpha - \alpha'$, for $\tilde{z} \in [-K_0, 0]$,

$$\begin{aligned} g_{1,N}(N^\gamma \tilde{z}) &= \chi N^\gamma \tilde{z} N^{-\beta} \lambda'(N^{\gamma-\beta} \tilde{z}) - \chi + \chi_0 e^{\lambda(N^{\gamma-\beta} \tilde{z})} \\ &= \chi \tilde{z} N^{\gamma-\beta} (-K_\theta \theta |N^{\gamma-\beta} \tilde{z}|^{\theta-1} + O(|N^{\gamma-\beta}|^{\tilde{\theta}'-1})) - \chi + \chi_0 (1 + K_\theta |N^{\gamma-\beta} \tilde{z}|^\theta + O(|N^{\gamma-\beta}|^{\tilde{\theta}'})) \\ &= -1 + O(N^{(\gamma-\beta)\theta(1+\theta')}); \\ g_{2,N}(N^\gamma \tilde{z}) &= (\chi_0 e^{\lambda(N^{\gamma-\beta} \tilde{z})} - \chi) N^\gamma \tilde{z} = N^\gamma \tilde{z} (-1 + o(1)); \\ \lambda'_N(N^\gamma \tilde{z}) C_0(N^\gamma \tilde{z}) &= N^{-\beta} \lambda'(N^{\gamma-\beta} \tilde{z}) C_0(N^\gamma \tilde{z}) \\ &= N^{-\beta} (-K_\theta \theta |N^{\gamma-\beta} \tilde{z}|^{\theta-1} + O(|N^{\gamma-\beta}|^{\tilde{\theta}'-1})) (K_\alpha |N^\gamma \tilde{z}|^\alpha + O(|N^\gamma|^{|\tilde{\alpha}'|})) \\ &= -K_\theta K_\alpha \theta |\tilde{z}|^{\theta+\alpha-1} N^{-\beta+(\gamma-\beta)(\theta-1)+\gamma\alpha} + O(N^{-\beta+\max\{(\gamma-\beta)(\tilde{\theta}'-1)+\gamma\alpha, (\gamma-\beta)(\theta-1)+\gamma\tilde{\alpha}'\}}) \\ &= -K_\theta K_\alpha \theta |\tilde{z}|^{\theta+\alpha-1} + O(N^{\max\{(\gamma-\beta)\theta\theta', -\gamma\alpha'\}}); \\ (e^{\lambda_N(N^\gamma \tilde{z})} - 1) C'_0(N^\gamma \tilde{z}) &= (e^{\lambda(N^{\gamma-\beta} \tilde{z})} - 1) C'_0(N^\gamma \tilde{z}) \\ &= (-K_\theta |N^{\gamma-\beta} \tilde{z}|^\theta + O(N^{(\gamma-\beta)\tilde{\theta}'})) (K_\alpha \alpha |N^\gamma \tilde{z}|^{\alpha-1} + O(N^{\gamma(\tilde{\alpha}'-1)})) \\ &= -K_\theta K_\alpha \alpha |\tilde{z}|^{\theta+\alpha-1} N^{(\gamma-\beta)\theta+\gamma(\alpha-1)} + O(N^{\max\{(\gamma-\beta)\theta+\gamma(\tilde{\alpha}'-1), (\gamma-\beta)\tilde{\theta}'+\gamma(\alpha-1)\}}) \\ &= -K_\theta K_\alpha \theta |\tilde{z}|^{\theta+\alpha-1} + O(N^{\max\{(\gamma-\beta)\theta\theta', -\gamma\alpha'\}}), \end{aligned}$$

where the notations O and o are understood as being uniform in \tilde{z} . Then, since by $\gamma = \beta\theta/(\theta + \alpha - 1)$, $\beta = \gamma(1 + (\alpha - 1)/\theta)$,

$$\begin{aligned} & |F_0^N(N^\gamma \tilde{\zeta}; N^\gamma \tilde{z}) - \tilde{J}_\infty^\gamma(\tilde{z}, 0)| \\ & \leq |(-1 + O(N^{(\gamma-\beta)\theta(1+\theta')}))(p + O(N^{2\gamma-1})) + p| + |O(N^{2\gamma-1})| + |O(N^{\max\{(\gamma-\beta)\theta\theta', -\gamma\alpha'\}})| \\ & \leq K(N^{2\gamma-1} + N^{\max\{(\gamma-\beta)\theta\theta', -\gamma\alpha'\}}) \leq KN^{\max\{2\gamma-1, -\gamma(\alpha-1)\theta', -\gamma\alpha'\}}. \end{aligned} \quad (\text{SA.32})$$

Step 3 We now prove (SA.1). First, in Case 1 above, by (SA.9) we see $A'_1(\tilde{z}) \leq 0$, $\tilde{z} < 0$. Then, based on Step 3 in the proof of Theorem 4.1 assertion (iii) as well as (SA.19), we have $\partial_{\tilde{z}} \tilde{J}'_\infty(\tilde{z}, 0) \leq p\chi\lambda'(\tilde{z}) \leq -p\chi|\lambda'(\tilde{Z}_\gamma^*(0)/2)|$, for all $\tilde{z} \leq \tilde{Z}_\gamma^*(0)/2$. Thus, by (SA.30) and exactly the same arguments as in Step 3 in the proof of Theorem 4.1 assertion (iii), we obtain $|\tilde{z}_0^N - \tilde{Z}_\gamma^*(0)| \leq KN^{2\gamma-1}$. Similarly, the same estimate holds for Case 2 above.

For Case 3 above, by (SA.31),

$$\begin{aligned}\partial_{\tilde{z}} \tilde{J}'_\infty(\tilde{z}, 0) &= \partial_{\tilde{z}} \left(-p + K_\theta K_\alpha (\theta + \alpha) |\tilde{z}|^{\theta+\alpha-1} \right) = -K_\theta K_\alpha (\theta + \alpha)(\theta + \alpha - 1) |\tilde{z}|^{\theta+\alpha-2} \\ &\leq -K_\theta K_\alpha (\theta + \alpha)(\theta + \alpha - 1) \left| \frac{1}{2} \tilde{Z}_\gamma^*(0) \right|^{\theta+\alpha-2}, \quad \tilde{z} \leq \frac{1}{2} \tilde{Z}_\gamma^*(0),\end{aligned}$$

from where similar arguments and (SA.32) yield the last estimate in (SA.1).

Finally, the case $v = 1$ can be proved similarly. Additionally, the same estimate in (4.4) in the setting of Theorem 4.1 stands – by the exact same arguments in its proof. ■

Proof of Theorem SA.1 assertion (iv). We follow the proof of Theorem 4.2 presented in Appendix C. In Definition 4.1, first, (4.7) is justified based on Theorem 4.2 with $\epsilon = \epsilon_N = KN^{\gamma-1/2}$, so it suffices to verify (4.6) with some appropriate ϵ . To this end, note that by (6.1),

$$\begin{aligned}N^{-\gamma} J_N(p; N^\gamma \tilde{z}, v) &= N^{-\gamma} e^{-\lambda(N^\gamma \tilde{z})} (v - p) N^\gamma \tilde{z} - N^{-\gamma} (1 - e^{-\lambda(N^\gamma \tilde{z})}) C_0(N^\gamma \tilde{z}) \\ &= (v - p) \tilde{z} (\chi e^{-\lambda(N^\gamma \tilde{z})} - \chi_0) - (1 - e^{-\lambda(N^\gamma \tilde{z})}) N^{-\gamma} C_0(N^\gamma \tilde{z}), \quad z \in \mathbb{R}_v.\end{aligned}\quad (\text{SA.33})$$

Since, again, $\gamma = \beta\theta/(\theta + \alpha - 1) < 1/2$ and $\tilde{P}_\gamma^* = p$, we turn to the two cases in Step 2 in the above proof of assertion (iii).

In Case 1, namely $\gamma = \beta = 0$, (SA.18) gives that

$$N^{-\gamma} J_N(p; N^\gamma \tilde{z}, v) = (v - p) \tilde{z} (\chi e^{-\lambda(\tilde{z})} - \chi_0) - (1 - e^{-\lambda(\tilde{z})}) C_0(\tilde{z}) = \tilde{J}'_\infty(p; \tilde{z}, v).$$

Then, by Definition 3.1,

$$N^{-\gamma} J_N(p; N^\gamma \tilde{Z}_\gamma^*(v), v) = \tilde{J}'_\infty(p; \tilde{Z}_\gamma^*(v), v) = \sup_{\tilde{z} \in \mathbb{R}_v} \tilde{J}'_\infty(p; \tilde{z}, v) = \sup_{z \in \mathbb{R}_v} N^{-\gamma} J_N(p; z, v),$$

which implies (4.6) with $\epsilon = 0$. Therefore, combined with (4.7), $\epsilon = \epsilon_N = KN^{\gamma-1/2}$. The same holds in Case 2: $\gamma = \beta > 0$ and $\alpha = 1$.

It remains to consider Case 3, namely $\beta > 0$ and $\alpha > 1$. We will verify (4.6) only for $v = 0$; the case $v = 1$ can be considered similarly. Recall (SA.33) and denote

$$\begin{aligned}\check{J}_N(p; \tilde{z}, 0) &:= e^{-\lambda(N^\gamma \tilde{z})} \partial_{\tilde{z}} \left(N^{-\gamma} J_N(p; N^\gamma \tilde{z}, 0) \right) = -p(\chi - \chi_0 e^{\lambda(N^\gamma \tilde{z})}) \\ &+ p\chi N^{\gamma-\beta} \tilde{z} \lambda'(N^\gamma \tilde{z}) - N^{-\beta} \lambda'(N^\gamma \tilde{z}) C_0(N^\gamma \tilde{z}) - (e^{\lambda(N^\gamma \tilde{z})} - 1) C'_0(N^\gamma \tilde{z}), \quad \tilde{z} < 0,\end{aligned}\quad (\text{SA.34})$$

with

$$\lim_{\tilde{z} \searrow 0} J_N(p; N^\gamma \tilde{z}, 0) = 0, \quad \lim_{\tilde{z} \rightarrow -\infty} J_N(p; N^\gamma \tilde{z}, 0) = -\infty, \quad \lim_{\tilde{z} \searrow 0} \check{J}_N(p; \tilde{z}, 0) = -p < 0.$$

By the first and third limits, it is clear that $\sup_{\tilde{z} < 0} J_N(p; N^\gamma \tilde{z}, 0) > 0$. Then, by the first two limits above, we see that $J_N(p; \cdot, 0)$ has a maximum point $\tilde{z}_N^* < 0$, which satisfies $\check{J}_N(p; \tilde{z}_N^*, 0) = 0$.

We claim that $|\tilde{z}_N^*| \leq K_0$ for some $K_0 > 0$. Indeed, note that all three terms in the second line of (SA.34) are positive. Then, denoting $\hat{z}_N^* := N^{\gamma-\beta}\tilde{z}_N^*$, we must have

$$|p\chi\hat{z}_N^*\lambda'(\hat{z}_N^*)| + |N^{-\beta}\lambda'(\hat{z}_N^*)C_0(N^\gamma\tilde{z}_N^*)| \leq p(\chi - \chi_0 e^{\lambda(\hat{z}_N^*)}) \leq p\chi.$$

In particular, $|\hat{z}_N^*\lambda'(\hat{z}_N^*)| \leq 1$. Then, clearly $|\hat{z}_N^*| \leq K_1$ for some $K_1 > 0$, and

$$|N^{-\beta}\lambda'(\hat{z}_N^*)C_0(N^\gamma\tilde{z}_N^*)| \leq p\chi,$$

which is (SA.28) with $K_2 = p\chi$. Then, employing the arguments following (SA.28), we obtain the desired boundedness of \tilde{z}_N^* .

Now, by (SA.21), we have

$$\tilde{J}_\infty^\gamma(\tilde{P}_\gamma^*; \tilde{z}, 0) = -p\tilde{z} - K_\theta K_\alpha |\tilde{z}|^{\theta+\alpha}, \quad \tilde{z} < 0,$$

from which using that $|\tilde{z}| \leq K_0$, by (SA.33) we obtain that as $N \rightarrow \infty$,

$$\begin{aligned} & |N^{-\gamma} J_N(p; N^\gamma \tilde{z}, 0) - \tilde{J}_\infty^\gamma(\tilde{P}_\gamma^*; \tilde{z}, 0)| \\ &= |-p\tilde{z}(\chi e^{-\lambda(N^{\gamma-\beta}\tilde{z})} - \chi_0) - (1 - e^{-\lambda(N^{\gamma-\beta}\tilde{z})})N^{-\gamma} C_0(N^\gamma \tilde{z}) + p\tilde{z} + K_\theta K_\alpha |\tilde{z}|^{\theta+\alpha}| \\ &\leq K|e^{-\lambda(N^{\gamma-\beta}\tilde{z})} - 1| + |(1 - e^{-\lambda(N^{\gamma-\beta}\tilde{z})})N^{-\gamma} C_0(N^\gamma \tilde{z}) - K_\theta K_\alpha |\tilde{z}|^{\theta+\alpha}| \\ &\leq K|\lambda(N^{\gamma-\beta}\tilde{z})| + |(\lambda(N^{\gamma-\beta}\tilde{z}) + O(\lambda^2(N^{\gamma-\beta}\tilde{z})))N^{-\gamma} C_0(N^\gamma \tilde{z}) - K_\theta K_\alpha |\tilde{z}|^{\theta+\alpha}| \\ &\leq K|N^{\gamma-\beta}\tilde{z}|^\theta + |(K_\theta|N^{\gamma-\beta}\tilde{z}|^\theta + O(|N^{\gamma-\beta}\tilde{z}|^{\theta(1+\theta')}) + O(|N^{\gamma-\beta}|^{2\theta})) \\ &\quad \times N^{-\gamma}(K_\alpha|N^\gamma \tilde{z}|^\alpha + O(|N^\gamma|^{|\alpha-\alpha'|})) - K_\theta K_\alpha |\tilde{z}|^{\theta+\alpha}| \\ &\leq K(N^{(\gamma-\beta)\theta} + N^{(\gamma-\beta)\theta-\gamma+\gamma(\alpha-\alpha')} + N^{(\gamma-\beta)\theta(1+\theta')-\gamma+\gamma\alpha} + N^{(\gamma-\beta)2\theta-\gamma+\gamma\alpha}) \\ &\leq K(N^{-\gamma(\alpha-1)} + N^{-\gamma\alpha'} + N^{-\gamma(\alpha-1)\theta'}) \leq KN^{\max\{-\gamma(\alpha-1), -\gamma\alpha', -\gamma(\alpha-1)\theta'\}} =: \frac{1}{2}\epsilon'_N. \end{aligned}$$

Therefore,

$$\begin{aligned} \sup_{z<0} N^{-\gamma} J_N(p; z, 0) &= N^{-\gamma} J_N(p; N^\gamma \tilde{z}_N^*, 0) \leq \tilde{J}_\infty^\gamma(\tilde{P}_\gamma^*; \tilde{z}_N^*, 0) + \frac{1}{2}\epsilon'_N \\ &\leq \tilde{J}_\infty^\gamma(\tilde{P}_\gamma^*, \tilde{Z}_\gamma^*(0), 0) + \frac{1}{2}\epsilon'_N \leq N^{-\gamma} J_N(p; N^\gamma \tilde{Z}_\gamma^*(0), 0) + \epsilon'_N, \end{aligned}$$

which verifies (4.6) with $\epsilon = \epsilon'_N$ and whose combination with (4.7) yields that in this case, $\epsilon = \max\{\epsilon_N, \epsilon'_N\} = KN^{\max\{\gamma-1/2, -\gamma(\alpha-1), -\gamma\alpha', -\gamma(\alpha-1)\theta'\}}$. This completes the proof. ■

Supplemental Appendix B Auxiliary details

This supplemental appendix provides auxiliary details for minor illustrative or computational arguments made throughout the paper, including all those not covered in the preceding appendices.

Details for hazard rate (2.1.3)

Let $N = \sigma = 1$ without loss of generality. Then, for (2.1.2), it is easy to see that

$$\rho_1(z) = D(z)\mathbb{P}\{|W+z| \geq \bar{y}\} = \frac{D(z)}{2} \left(\operatorname{erfc}\frac{\bar{y}+z}{\sqrt{2}} + \operatorname{erfc}\frac{\bar{y}-z}{\sqrt{2}} \right), \quad z \in \mathbb{R},$$

where $\bar{y} > 0$. Matching this to (2.1.1) and using that λ_1 is positive-valued gives (2.1.3).

Further, for the complementary error function, it is familiar that (see Abramowitz and Stegun (1972) Chap. 7)

$$\operatorname{erfc} x = 1 - \frac{2x}{\sqrt{\pi}} + O(x^2), \quad \text{as } x \rightarrow 0.$$

Thus,

$$\begin{aligned}\lambda_1(z) \equiv \lambda(z) &= -\log \left(1 - \frac{D(z)}{2} \left(\operatorname{erfc} \frac{\bar{y}+z}{\sqrt{2}} + \operatorname{erfc} \frac{\bar{y}-z}{\sqrt{2}} \right) \right) \\ &= -\log(1 - D(z)) + O(z^2) = O(|z|^{\min\{\theta_D, 2\}}) \quad \text{as } z \rightarrow 0,\end{aligned}\tag{SB.1}$$

assuming that $D(z) \sim K_D |z|^{\theta_D}$ as $|z| \searrow 0$, with $K_D > 0$ and $\theta_D \geq 1$. Note that since $|\operatorname{erfc} x| < 1$, $\forall x \in \mathbb{R}$, the function D is required to be bounded by 1, or else (SB.1) is not real-valued. Moreover, if D is twice continuously differentiable, then so is λ in (SB.1). Using that $\operatorname{erfc}' x = -2e^{x^2}/\sqrt{\pi}$, for $z > 0$ we obtain

$$\lambda'(z) = \frac{\frac{D(z)}{\sqrt{2\pi}}(e^{-(\bar{y}+z)^2/2} + e^{-(\bar{y}-z)^2/2}) + \frac{D'(z)}{2}(\operatorname{erfc} \frac{\bar{y}+z}{\sqrt{2}} + \operatorname{erfc} \frac{\bar{y}-z}{\sqrt{2}})}{1 - \frac{D(z)}{2}(\operatorname{erfc} \frac{\bar{y}+z}{\sqrt{2}} + \operatorname{erfc} \frac{\bar{y}-z}{\sqrt{2}})}.\tag{SB.2}$$

Since the function $\operatorname{erfc}((\bar{y}+z)/\sqrt{2}) + \operatorname{erfc}((\bar{y}-z)/\sqrt{2})$, $z > 0$, is strictly increasing, (SB.2) is clearly increasing provided that D' is also increasing; in particular,

$$\lambda'(z) \geq \frac{1}{\sqrt{2\pi}} e^{-(\bar{y}+z)^2/2} (e^{2\bar{y}z} - 1) D(z) \geq \sqrt{\frac{2}{\pi}} \bar{y} e^{-\bar{y}^2/2} D(z), \quad z > 0.$$

It follows from the oddity of λ' that $|\lambda'(z)| \geq bD(z)$, $z \neq 0$, with $b = 2\bar{y}e^{-\bar{y}^2/2}/\sqrt{2\pi}$. Therefore, for $D(z) = K_D |z|^{\theta_D}$, we have $|\lambda'(z)| \geq bD(z) = bK_D |z|^{\theta_D}$, $z \neq 0$. Additionally, with $D'(0)$ and $D''(0)$ both finite, a further expansion (SB.2) around $z = 0$ directly shows that $\lambda'(z) = \lambda'(0) + O(|z|)$ as $z \rightarrow 0$. ■

Details for Example 4.1. Using Theorem 4.1, since $\gamma = \beta < 1/2$, the equilibrium price function $\tilde{P}_\gamma^* = p \in (0, 1)$, and, as in (B.6), the limiting equilibrium conditions are the decoupled equations $\tilde{J}_\infty^\gamma(\tilde{z}, v) = 0$, $v \in \{0, 1\}$, for $\tilde{J}_\infty^\gamma(\tilde{z}, v)$ defined in (B.19). With $\lambda(\tilde{z}) = \tilde{z}^2$ and $\chi_0 = \chi - 1 \geq 0$, they specialize to

$$\tilde{J}_\infty^\gamma(\tilde{z}, v) = (v - p)(\chi - 2\chi\tilde{z}^2 - \chi_0 e^{\tilde{z}^2}) = 0,\tag{SB.3}$$

to be solved for $\tilde{z} \in \mathbb{R}_v$. Note that $\tilde{J}_\infty^\gamma(\tilde{z}, 1) = -(1-p)/p \tilde{J}_\infty^\gamma(-\tilde{z}, 0)$ for $\tilde{z} > 0$, which explains the symmetry (about 0) of their solutions. By letting $x = 1/2 - \tilde{z}^2$, (SB.3) is equivalent to

$$2\chi x - \chi_0 e^{\frac{1}{2}-x} = 0 \iff x e^x = \frac{\sqrt{e}\chi_0}{2\chi}.$$

whose unique nonnegative solution can be expressed in terms of the Lambert W function as $x = W_0(\sqrt{e}\chi_0/(2\chi))$, and $\tilde{z} = \pm\sqrt{1/2-x}$. Thus, setting $\chi = 3$ (and $\chi_0 = 2$) gives us the γ -limiting equilibrium strategy specified in (4.5). ♦

Details for Example 6.1. From the proof of Theorem SA.1 assertion (ii), since $\gamma = \beta < 1/2$, we clearly have $\tilde{P}_\gamma^* = p = 1/3$, while the limiting equilibrium conditions are again the decoupled system $\tilde{J}_\infty^\gamma(\tilde{z}, v) = 0$, $v \in \{0, 1\}$, for $\tilde{J}_\infty^\gamma(\tilde{z}, v)$ in (SA.19), both for $\beta = 0$ and for $\beta > 0$, the latter case being

due to $C_0(z) = |z|$ as used in (SA.20) (with $K_\alpha = 1$). With $\lambda(z) = |\tilde{z}|$ further, and recalling $\chi = 1$, by (SA.19) and (SA.8) we find that

$$\tilde{J}_\infty^\gamma(\tilde{z}, v) = (v - \frac{1}{3})(1 - |\tilde{z}|) + (-1)^v(|\tilde{z}| + e^{|\tilde{z}|} - 1) = 0, \quad (\text{SB.4})$$

to be solved for $\tilde{z} \in \mathbb{R}_v$, $v \in \{0, 1\}$. Let $\tilde{z}_v \in \mathbb{R}_v$ denote the solution, and set $x_v = 1 - |\tilde{z}_v|$. Then, for $v = 0$, (SB.4) is equivalent to

$$-\frac{1}{3}x_0 + (-x_0 + e^{1-x_0}) = 0 \iff x_0 e^{x_0} = \frac{3e}{4}.$$

Then, $x_0 = W_0(3e/4)$, and so $\tilde{z}_0 = W_0(3e/4) - 1$. Similarly, for $v = 1$, (SB.4) is equivalent to

$$\frac{2}{3}x_1 - (-x_1 + e^{1-x_1}) = 0 \iff x_1 e^{x_1} = \frac{3e}{5}.$$

Then, $x_1 = W_0(3e/5)$, and so $\tilde{z}_1 = 1 - W_0(3e/5)$. ■

Details for Example 6.2. With $\gamma = \beta = 0$, we have $\tilde{P}_0^* = p = 1/3$, and the limiting equilibrium strategy amounts to maximizing the limiting objective functions in (SA.18), with $\chi = 1$ and $\chi_0 = 0$. By plugging in the given expressions of λ and C_0 and by straightforward calculations, we obtain

$$\tilde{J}_\infty^0(p; \tilde{z}, 0) = \begin{cases} \frac{1}{4}, & \text{if } \tilde{z} \leq -6, \\ \frac{\tilde{z}(\tilde{z}+48)}{144(\tilde{z}-1)}, & \text{if } -6 < \tilde{z} < 0, \end{cases} \quad \tilde{J}_\infty^0(p; \tilde{z}, 1) = \begin{cases} \frac{\tilde{z}(96-25\tilde{z})}{144(\tilde{z}+1)}, & \text{if } 0 < \tilde{z} < \frac{6}{5}, \\ \frac{1}{4}, & \text{if } \tilde{z} \geq \frac{6}{5}. \end{cases}$$

Since $\tilde{z}(\tilde{z}+48)/(144(\tilde{z}-1)) < 1/4$ for $-6 < \tilde{z} < 0$ and $\tilde{z}(96-25\tilde{z})/(144(\tilde{z}+1)) < 1/4$ for $0 < \tilde{z} < 6/5$, the continuum of (limiting) equilibrium strategies in (6.6) is justified. ■

Details for calibration experiments in Section 5.

Suppose that there exists $\bar{z} > 0$ such that $D(z) = 1$ for $|z| \geq \bar{z}$, meaning that as the insider increases his trade size indefinitely, the probability of prosecution conditional on investigation can reach 1. Then, using that (see, again, Abramowitz and Stegun (1972) Chap. 7)

$$\text{erfc}x = \begin{cases} 2 + e^{-x^2} \left(\frac{1}{\sqrt{\pi}x} + O(x^{-2}) \right), & \text{as } x \rightarrow -\infty, \\ e^{-x^2} \left(\frac{1}{\sqrt{\pi}x} + O(x^{-2}) \right), & \text{as } x \rightarrow \infty, \end{cases}$$

we further see that the hazard rate (2.1.3) satisfies

$$\lambda_1(z) \equiv \lambda(z) \sim \frac{z^2}{2\sigma^2}, \quad \text{as } |z| \rightarrow \infty,$$

justifying the choice of the approximation $\lambda(z) = Kz^2$ with $K = 1/(2\sigma^2)$. For example, Figure 3 below gives an illustration with $\sigma = 1000$, $\beta = 0$, $\bar{y} = 3$, and $D(z) = \min\{z^2, 1\}$ (with $\bar{z} = 1$).

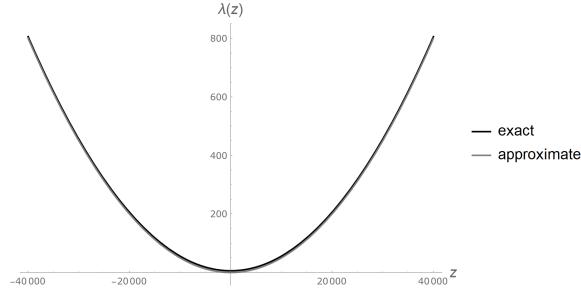


Figure 3: Hazard rate approximation

The formula (5.1.2) follows straight from the law of total probability:

$$\mathbb{E}[|Z(V)| \mid B_N(Z(V)) = 1] = \frac{\mathbb{E}[|Z(V)| \mathbf{1}(B_N(Z(V)) = 1)]}{\mathbb{P}\{B_N(Z(V)) = 1\}} = \frac{\mathbb{E}[|Z(V)|(1 - e^{-\lambda_N(Z(V))})]}{\mathbb{E}[1 - e^{-\lambda_N(Z(V))}]},$$

recalling that $V \stackrel{d}{=} \text{Bernoulli}(p)$. Similarly, for (5.1.3) and (5.1.4), note that $X_N \stackrel{d}{=} \text{Normal}(N\mu, N(\sigma^2 - \mu^2))$, independent of V , and then

$$\begin{aligned} \mathbb{E}[X_N + |Z(V)| \mid B_N(Z(V)) = 1] &= \mathbb{E}[X_N] + \mathbb{E}[|Z(V)| \mid B_N(Z(V)) = 1] \\ &= N\mu + \frac{\sum_{v=0}^1 p(v)|Z(v)|(1 - e^{-\lambda_N(Z(v))})}{\sum_{v=0}^1 p(v)(1 - e^{-\lambda_N(Z(v))})}, \\ \mathbb{P}\left\{\frac{X_N + |Z(V)|}{|Z(V)|} > x \mid B_N(Z(V)) = 1\right\} &= \mathbb{P}\left\{X_N > (x - 1)|Z(V)| \mid B_N(Z(V)) = 1\right\} \\ &= \frac{\sum_{v=0}^1 p(v)(1 - e^{-\lambda_N(Z(v))}) \frac{1}{2} \operatorname{erfc} \frac{|Z(v)|(x-1) - N\mu}{\sqrt{2N(\sigma^2 - \mu^2)}}}{\sum_{v=0}^1 p(v)(1 - e^{-\lambda_N(Z(v))})}. \end{aligned}$$

For the calibration, by writing $Z = N^\gamma \tilde{Z}_\gamma^*$ in terms of (5.1.5) and using the symmetry $|\tilde{Z}_\gamma^*(0)| = \tilde{Z}_\gamma^*(1) > 0$, the equations (5.1.2), (5.1.3), and (5.1.4) can be significantly simplified, permitting explicit solutions. In particular,

$$(5.1.2) = N^\gamma \sigma \times \alpha, \quad (5.1.3) = N\mu + N^\gamma \sigma \times \alpha, \quad (5.1.4) = \frac{1}{2} \operatorname{erfc} \frac{(N^\gamma \sigma \times \alpha)(x-1) - N\mu}{\sqrt{2N(\sigma^2 - \mu^2)}}.$$

For conciseness, let us denote the average insider share volume, the average total share volume, and the median insider-to-total volume ratio from data by i , v , and r , respectively, and recall that $\alpha = \sqrt{1 - 2W_0(\sqrt{e}\chi_0/(2\chi))}$, with $\chi_0 = \chi - 1$.

The first calibration condition is due to (5.1.2) and (5.1.5),

$$N^\gamma \sigma \times \alpha = i. \tag{SB.5}$$

Similarly, from (5.1.3) we have the second calibration condition,

$$N\mu + N^\gamma \sigma \times \alpha = v. \tag{SB.6}$$

The third calibration condition is obtained in the same manner by evaluating (5.1.4). In particular, since $\operatorname{erfc} - 1$ is an odd function, we obtain

$$N^\gamma \sigma \times \alpha \times (r^{-1} - 1) - N\mu = 0. \tag{SB.7}$$

Notably, in calibration experiment I, the average total trading volume is associated with the standard error 10,246, respectively, as reported by Meulbroek (1992). Since there are a total of 588 insider trading episodes included in the data set, a reasonable estimate for the standard deviation of the trading volume, given independence between insider trades and normal trades, is

$$s = \sqrt{10246^2 \times 588} \approx 248452. \tag{SB.8}$$

Thus, the additional condition $N(\sigma^2 - \mu^2) = \mathfrak{s}^2$ can be used along with (SB.5) and (SB.6) to obtain an estimate for the parameter $\mu > 0$, namely

$$\hat{\mu} = \frac{\sqrt{\mathfrak{s}^4 + 4\sigma^2 \times (\mathfrak{v} - \mathfrak{i})^2} - \mathfrak{s}^2}{2 \times (\mathfrak{v} - \mathfrak{i})}, \quad (\text{SB.9})$$

which we simply take as μ .

Taking (SB.9) as given, the preceding calibration conditions are easily solvable by elementary means. In particular, from (SB.5) and (SB.6) we obtain

$$\hat{N} = \frac{\mathfrak{v} - \mathfrak{i}}{\mu}, \quad \hat{\gamma} = \frac{\log \frac{\mathfrak{i}}{\sigma \times \mathfrak{a}}}{\log \hat{N}}, \quad (\text{SB.10})$$

if (SB.7) is used instead of (SB.6),

$$\hat{N} = \frac{\mathfrak{i} \times (1 - \mathfrak{r})}{\mu \times \mathfrak{r}}, \quad \hat{\gamma} = \frac{\log \frac{\mathfrak{i}}{\sigma \times \mathfrak{a}}}{\log \hat{N}}, \quad (\text{SB.11})$$

when only (SB.6) and (SB.7) are used, we have

$$\hat{N} = \frac{\mathfrak{v} \times (1 - \mathfrak{r})}{\mu}, \quad \hat{\gamma} = \frac{\log \frac{\mathfrak{v} \times \mathfrak{r}}{\sigma \times \mathfrak{a}}}{\log \hat{N}}. \quad (\text{SB.12})$$

According to the statistics shown in Table 1, setting $\mathfrak{i} = 9819$, $\mathfrak{v} = 113909$, $\mathfrak{r} = 0.113$, $\sigma = 1000$, and $\chi \in \{1, 2, 3\}$ and evaluating (SB.9) verifies $\mu = 1.68625$, and subsequently evaluating (SB.9), (SB.10), (SB.11), and (SB.12) gives rise to Table 2 for calibration experiment I.

In the same vein, for calibration experiment II, Table 5 is a result of evaluating (SB.11) with $\mathfrak{i} = 4900$, $\mathfrak{r} = 0.026$, $\mu = 1.68625$, $\sigma = 1000$, and $\chi \in \{1, 2, 3\}$, based on the information in Table 4.

Moreover, the formula (5.1.1) follows directly from the law of total probability, $\mathbb{E}[B(Z(V))] = \mathbb{E}[1 - e^{-\lambda_N(Z(V))}]$. Then, using that $|\tilde{Z}_\gamma^*(0)| = \tilde{Z}_\gamma^*(1) > 0$, the prosecution probability implied by the equilibrium is

$$1 - e^{-\lambda_N(N^\gamma \tilde{Z}_\gamma^*(1))} = 1 - e^{-\lambda(\tilde{Z}_\gamma^*(1))}, \quad (\text{SB.13})$$

which upon setting $\gamma = \beta$ becomes independent of N and γ . Putting the solved finite-population and limiting strategy values from Table 6 into (5.1.1) (with $Z = Z_N^*$) and (B.7), respectively, gives the last column of Table 6. ■