Short-Sales Constraints and Price Bubbles

Bryan Y. Lim*
Department of Finance, University of Melbourne
August 9, 2010

Abstract

Miller (1977) demonstrated that if investors have heterogeneous beliefs and short sales are restricted, trade of a security will disproportionately reflect positive information, generating a price bubble. As this intuition applies most relevantly to short intervals of trade, a question arises as to the longevity of such a bubble. In this paper, I argue that a bubble effected by short-sales constraints persists only if agents cannot distinguish between order flow caused by positive information or order flow caused by the constraints. If the constraint is common knowledge, it should have no effect on the long-term pricing of the stock. If, however, the constraint is random and unknown, a price bubble may form.

*E-mail: blim@unimelb.edu.au. I am indebted to my committee members Stephen F. LeRoy (chair), Gary Charness, and Doug Steigerwald as well as Will Goetzmann, Eric Matolo, Cheng Zhong Qin, Joao Rosario, Jeffrey Stopple, and seminar participants at the University of Melbourne and the 2007 Financial Management Association annual meeting for their helpful comments. I particularly thankful to an anonymous referee for guidance with this paper.
Empirical literature has demonstrated positive correlation between short-sales constraints and equity prices.\(^1\) This phenomenon has typically been cited as validation of Miller (1977), who argued that when investors have heterogeneous beliefs and there are constraints on short sales, trade will disproportionately reflect the most optimistic investors, pushing prices higher than they would be in the absence of such constraints. This result is generated from a simple one-period model, with a downward-sloping demand curve and inelastic supply of a risky asset. Short-sales constraints shift the supply inward (to the left), increasing the equilibrium price and reducing the equilibrium quantity traded.

If we call this increase in the equilibrium price a “bubble”, how durable should we expect this bubble to be? Miller’s (1977) intuition is obviously appealing, but if one were to consider a subsequent round of trading, the likelihood of the bubble persisting drops. Traders in the second round should condition their beliefs on the information revealed in the first round. If, as rational expectations would imply, they recognize this apparent positive information is skewed by constraints on short selling, for the overpricing to continue traders in the second round must be even more optimistic than those in the first round (assuming supply does not change). Absent such increasing optimism, it is not immediately clear how the bubble described by Miller (1977) survives over successive rounds of trade. Indeed, Diamond and Verrecchia (1987) convincingly show in a microstructure model with rational expectations that there is no upward price pressure from short-sales constraints. In their model, agents infer negative information from gaps in order flow resulting from constrained would-be traders unable to locate shares to sell short. This information is incorporated into bid and ask prices, nullifying any potential upward bias.

In light of the empirical literature, then, the disconnect between the single-period result of Miller (1977) and the multi-period result of Diamond and Verrecchia (1987) begs the question: In a dynamic market, can short-sales constraints persistently bias prices upward? If investors induct information from earlier trades, can short-sales constraints generate persistent upward price pressure? I argue in this paper that the answer is yes, short-sales constraints can bias prices upward in a dynamic market, but to do so, there must be some uncertainty about the magnitude of the constraint. If known with adequate precision, short-sales constraints will have no significant long-run effect on prices. As constraints on short sales effectively truncate the trading

---

\(^1\) See, for example, Jones and Lamont (2002) and Chang, Cheng, and Yu (2007).
population, if agents know \textit{a priori} the magnitude of the constraint, prices will account for the truncated population, nullifying the potential upward bias. As in Diamond and Verrecchia (1987), the information revelation process slows down as constraints reduce the informational efficiency of trades, but in the limit, the market price converges to the asset’s fundamental value.

For short-sales constraints to generate price bubbles in a dynamic setting, traders must be unable to distinguish whether a relative inflow of buy orders is caused by positive information or by the inability of some traders to place sell orders. Recall that Miller (1977) specified two requirements to generate a bubble: 1) short-sales constraints and 2) heterogeneous beliefs. My paper adds a third: 3) uncertainty about the short-sales constraint. The conditions under which short-sales constraints generate a price bubble require that agents sufficiently underestimate the magnitude of the constraint. The threshold for “sufficient” underestimation decreases both as the quality of agents’ information decreases and as the constraint becomes more binding. The implication is that the noisier their private information is or the more constrained the market is, the more accurate agents’ estimates of the short-sales constraint must be in order to prevent a bubble.

As an illustrative if unrealistic example, imagine a market for a risky security where short selling is not possible yet all traders condition their beliefs as though it were. The security has zero value, but short-selling constraints may cause an excess of buy orders, which traders interpret as an excess of optimism and bid up the market price. In essence, they would have confused the truncated trading population with positive information. A second and arguably more realistic example would be the same market except with the extent of the constraint unknown. Suppose instead that the distribution of the constraint is common knowledge: there is a 50% probability the short-sales constraints are binding for a significant portion of the population and a 50% probability the constraints are not binding. In this case, agents are unable to determine with certainty whether the observed buy orders are the results of general optimism across all investors or of pessimistic investors’ being unable to place (short) sell orders. Prices will reflect a weighted average of both possibilities and will consequently be higher than if a) the extent of the short-sales constraint were known or b) the constraint didn’t exist at all.

In terms of a contribution to the theoretical literature, this paper may be considered a bridge between Miller (1977) and Diamond and Verrecchia (1987). It is distinguished from the former in its multi-period structure. The single-period model of Miller (1977)
would seem most appropriate for explaining the effects of shorting constraints over brief intervals, like IPOs\(^2\), while my model is intended to examine trade over longer horizons. It is distinguished from the latter by its allowance for the constraint to be random and unknown rather than certain. This uncertainty is the means by which the upward price pressure is possible in my paper.

More generally, this paper contributes to the class of theory papers examining the pricing effects of short-sales constraints. Models like Jarrow (1980), Bai, Chang, and Wang (2006) and Cao, Zhang, and Zhou (2007) find short sales constraints have mixed effects on prices. On the one hand, supply is reduced, which puts upward pressure on prices; on the other hand, it is known that some investors with negative information are unable to trade, which puts downward pressure on prices. Duffie, Garleanu, and Pedersen (2002) demonstrate that if short sales are difficult to execute, prices will rise initially and be expected to decline subsequently. Gallmeyer and Hollifield (2008) examine short-sales constraints in a dynamic general equilibrium setting and argue that the constraints reduce expected returns, a result which has both income and substitution effects. Which effect dominates is determined by the investor’s intertemporal elasticity of substitution.

My motivation is a parsimonious explanation for the documented relation between short-sales constraints and price inflation, beyond what may be explained by Miller (1977). Much of the empirical literature measures persistent distortionary effects of short-sales constraints. Ofek and Richardson (2003) posit that share lockups amongst insiders following the waves of 1990’s tech IPOs limited the number of shares to available to short sell, thus inflating prices which collapsed when the share lockups expired.\(^3\) Chang, Cheng, and Yu (2007) find that newly-shortable equities in the Hong Kong market have lower risk-adjusted returns than their non-shortable counterparts. Jones and Lamont (2002) use shorting data from the 1926 to 1933 to show that short-sales constrained stocks have higher prices and lower returns. Chen, Hong and Stein (2002) use breadth of ownership – defined roughly as the number of investors with long positions in a particular stock – as a proxy for short-sales constraints. They document high prices associated stocks whose breadth is relatively low. Diether, Malloy and Scherbina (2002), using the dispersion of analysts forecasts as a proxy for heterogeneous beliefs,

\(^2\)Curiously, Edwards and Hanley (2008) find no evidence that short-sales constraints are responsible for the well-documented underpricing phenomenon in IPOs.

\(^3\)Battalio and Schultz (2006) dispute this result, observing that investors could cheaply short synthetically using options.
show that wide dispersion for a stock is negatively correlated with future returns, which
they interpret as support for the basic Miller (1977) hypothesis. Cohen, Diether and
Malloy (2007) find that increases in shorting demand leads to negative abnormal re-

This paper is organized as follows. Section 1 introduces the basic model with the
assumption that the short-sales constraint is known to all participants. Section 2 relaxes
this assumption, allowing for the constraint to be random and for participants to know
only its distribution. Section 3 discusses qualitative implications of the model. Section
4 concludes.

1 The Model

The model is based on Avery and Zemsky (1998). There is a single risky asset which is
supplied and demanded infinitely by a market specialist. The asset takes value \( V = 0 \)
or 1 with equal probability.

A sequence of risk-neutral investors trades with the specialist. The sequence is
assumed random, implying that traders are unable to time the market and that the
time at which they transact is not strategically motivated. Investors are indexed by
\( n = 1, 2, ..., N \), the order in which they trade.

Each investor may buy or sell a maximum of one unit of the asset. Though this
simplification may seem particularly unrealistic, it mimics the fact that in practice a
market specialist’s quoted prices are valid up to a fixed number of shares (typically
1,000 or 10,000).

Before trading, \( \lambda \) percent of investors receive a private signal regarding \( V \): \( x_n \in
\{High, Low\} \). For all \( n \), \( x_n \) is accurate with probability \( q \in (1/2, 1) \). The conditional
probabilities of \( x_n \) are given by

\[
\begin{array}{|c|c|c|}
\hline
& V = 0 & V = 1 \\
\hline
P[x = Low|V] & q & 1 - q \\
\hline
P[x = High|V] & 1 - q & q \\
\hline
\end{array}
\]

\( q \) is both uniform across all informed agents and bounded below 1. As such, there
are no agents who know \( V \) with certainty. Echoing Owens and Steigerwald (2005),

---

4The term “specialist” is used somewhat loosely here, as it is meant to include market makers (e.g.,
dealers in NASDAQ markets) as well.
this specification represents a departure from much of the information-based pricing literature, particularly those derived from Glosten and Milgrom (1985). Typically in such models, informed agents are assumed to know with probability 1 the terminal value of the asset. In contrast, in my model, no matter how informative the signal is, buying or selling the asset necessitates bearing some degree of risk.

To motivate this information structure, one might think of private signals instead as private forecasts, a perhaps less controversial characterization with respect to financial markets. One may assume that some traders may have invested either the time or money necessary to create a forecast on the asset in question and, having done so, enjoy an informational advantage over other less informed agents in the market. In the present context, quantifying these costs is nebulous and ultimately not germane to the information revelation process, so it will be left out of the model. Note that regardless of whichever term one prefers—“signal” or “forecast”—the difference is entirely superficial as the mathematics remain unchanged.

While \( \lambda \) percent of investors receive private signals, the remaining \( 1 - \lambda \) percent of investors are “noise” traders, agents whose actions are purely random and therefore contain no information. Noise traders buy and sell with equal probability, the latter insofar as they are not constrained to do so. (This symmetry between buys and sells is not critical to the model and is used only for expositional clarity.) By assumption, the short-sales constraint does not affect noise traders. This rules out the possibility that an uninformed agent would choose to short sell the asset.

An agent may sell the asset with probability \( \theta \in (0, 1] \). The parameter \( \theta \) captures the percentage of agents who either already own the asset or are not short sales constrained. It is the latter case that is of principal interest. In practice, there are any number of variables—observable and unobservable—that may affect the short-sales component of \( \theta \), including but not exclusive to the number of tradable shares (asset float) and the size of the trading population.\(^5\)

**Assumption 1.** If an agent chooses not to trade\(^6\), her action is not recorded as part of the market history and subsequent agents cannot infer her signal. She drops out of the

\(^5\)The reader may notice that in my specification of \( \theta \), I implicitly abstract the actual mechanics of locating and financing short sales. This is by design: this paper is concerned with the interaction between constraints and informational quality. For an analysis of the implications of short sales costs, see Duffie, Garleanu, and Pedersen (2002).

\(^6\)As implied by Proposition 1 later in this section, an informed agent chooses not to trade when she receives a *Low* signal but is constrained from doing so. A noise trader chooses not to trade when she (randomly) would like to sell but is constrained from doing so.
queue, and the line moves forward instantaneously. The queue represents a truncated trading population: only those agents who will execute a trade are considered.

A functionally equivalent assumption is that there is no information to be inferred by the interval between trades. This is a departure from both Diamond and Verrecchia (1987) and Easley and O’Hara (1992), both of which argue that such intervals may convey information. Easley and O’Hara (1992), for example, posit that the specialist is trying to determine whether an information event has occurred. The probability of an agent choosing not to trade is correlated with the probability of an informational event occurring, and thus an agent’s not trading is informative. In contrast, my model assumes an information event has occurred with probability 1 and that this occurrence is common knowledge. The market specialist is tasked with processing the only trades that occur as a result of this information event. Consequently, there is no sense in which time itself reveals information to him.

Assuming rational expectations, I first consider the case where the magnitude of the sales constraint is known to all agents. Formally, I define Information Structure I (IS I): The value of \( \theta \) is common knowledge prior to trading. In reality, the plausibility of agents knowing the value of \( \theta \) with certainty is suspect, at best, given that \( \theta \) is fundamentally unobservable, ex ante or ex post. Nevertheless, some interesting results obtain from analyzing the model under this information structure. These results will provide a basis for comparison to those in the subsequent section, which considers the case where only the distribution of or an estimate of \( \theta \) is common knowledge.

Denoting the \( n \)th trade as \( t_n \) and the trader’s type as \( \tau_n \in \{ \text{Informed, Noise} \} \), I define \( \pi_{I}^{t_n} \) as the probability the \( n \)th trade is a Buy, conditional on \( V \) and the trader’s being informed:

\[
\pi_{I}^{I} = P[t_n = \text{Buy}|V = 1, \tau_n = I] \\
\pi_{I}^{0} = P[t_n = \text{Buy}|V = 0, \tau_n = I]
\]

Then the probability of a Buy conditional on \( V \) is

\[
\pi_{I} = P[t_n = \text{Buy}|V = 1] = \lambda \pi_{I}^{I} + (1 - \lambda) \frac{1}{1 + \theta} \\
\pi_{0} = P[t_n = \text{Buy}|V = 0] = \lambda \pi_{I}^{0} + (1 - \lambda) \frac{1}{1 + \theta}
\]
Figure 1: Game tree given sales constraint $\theta$.

implying the probability of a Sell, conditional on $V$ is

$$1 - \pi_1 = P[t_n = Sell | V = 1]$$
$$1 - \pi_0 = P[t_n = Sell | V = 0]$$

1.1 The Market Specialist’s Problem

The market specialist processes a single order at a time, adjusting his bid and ask quotes after each trade. Though I have specified that traders act sequentially, whether or not they literally form a queue is irrelevant. As given by Assumption 1, there is no sense in which the interval between trades is informative, neither to the specialist nor any other observer. In effect, the queue represents order flow. The specialist is only concerned with a) posting bid and ask prices based on the trade history and b) processing the next submitted trade. From his perspective, each trade represents the outcome of the stationary process illustrated in Figure 1.
The specialist operates in a competitive environment and thus sets bid and ask prices such that his expected profit on any trade is zero. More specifically, he posts prices equal to the expected value of the security conditional on the history of publicly observable trades prior to the \( n \)th trade \( h_n \) and the agent’s trade \( t_n \):

\[
\begin{align*}
\text{Bid}_n &= E[V|t_n = \text{Sell}, h_n] \\
&= P[V = 1|t_n = \text{Sell}, h_n] \\
\text{Ask}_n &= E[V|t_n = \text{Buy}, h_n] \\
&= P[V = 1|t_n = \text{Buy}, h_n]
\end{align*}
\]

As in Glosten and Milgrom (1985), the bid-ask spread represents the specialist’s effort to protect himself against the potentially superior information of the agents with whom he is trading.

At the start of trading, the specialist sets the following bid and ask prices:

\[
\begin{align*}
\text{Bid}_1 &= P[V = 1|t_1 = \text{Sell}] \\
&= \frac{1 - \pi_1}{(1 - \pi_1) + (1 - \pi_0)} \\
\text{Ask}_1 &= P[V = 1|t_1 = \text{Buy}] \\
&= \frac{\pi_1}{\pi_1 + \pi_0}
\end{align*}
\]

Defining \( Price_n \) as the price of the \( n \)th transaction, \( Price_1 = \text{Bid}_1 \) if the first trade was a Sell and \( Price_1 = \text{Ask}_1 \) if the first trade was a Buy. Note that whatever the first trade was, I can write

\[
Price_1 = \frac{y}{y + z}
\]

for some \( y \) and \( z \). Given \( Price_1 \), one can easily confirm the specialist sets the following bid and ask prices for the next trader:

\[
\begin{align*}
\text{Bid}_2 &= \frac{(1 - \pi_1)y}{(1 - \pi_1)y + (1 - \pi_0)z} \\
\text{Ask}_2 &= \frac{\pi_1y}{\pi_1y + \pi_0z}
\end{align*}
\]

Iterating over \( n \), I find equations (1), (2), and (3) hold for all \( n \): \( Price_n, \text{Bid}_{n+1}, \),
and \( \text{Ask}_{n+1} \). For any \( k \) Buys in \( n \) trades, I can write

\[
\text{Price}_n(k) = \frac{\pi^k_1(1 - \pi_1)^{n-k}}{\pi^k_1(1 - \pi_1)^{n-k} + \pi^k_0(1 - \pi_0)^{n-k}}
\]

(4)

\[
\text{Bid}_{n+1}(k) = \frac{\pi^{k+1}_1(1 - \pi_1)^{n-k+1}}{\pi^{k+1}_1(1 - \pi_1)^{n-k+1} + \pi^{k+1}_0(1 - \pi_0)^{n-k+1}}
\]

(5)

\[
\text{Ask}_{n+1}(k) = \frac{\pi^{k+1}_1(1 - \pi_1)^{n-k}}{\pi^{k+1}_1(1 - \pi_1)^{n-k} + \pi^{k+1}_0(1 - \pi_0)^{n-k}}
\]

(6)

Since \( V \) is normalized to being either 1 or 0, equation (4) implies \( \text{Price}_n \) is equivalent to agents’ belief that \( V = 1 \), conditional on the public history of trades after \( n \) trades, \( H_n \):

\[
\text{Price}_n = E[V|H_n] = P[V = 1|H_n]
\]

1.2 The Informed Trader’s Problem

If the \( n \)th trader is informed, her information set includes her private signal \( x_n \) and the history of prior trades \( H_{n-1} \). Risk neutrality ensures informed agents follow the simple trading rule:

- sell if \( E[V|x_n, H_{n-1}] < \text{Bid}_n \)
- buy if \( E[V|x_n, H_{n-1}] > \text{Ask}_n \)

where selling is contingent on either already owning the asset or not being short sales constrained.

**Proposition 1.** Informed agents follow their signal, buying when \( x_n = \text{High} \) and selling when \( x_n = \text{Low} \), conditional on the sales constraint \( \theta \).

**Proof.** All proofs are in the appendix. \( \square \)

Given Proposition 1, I can explicitly calculate \( \pi^I_1 \), the probability the \( n \)th trade is a Buy conditional on \( V \) and the trader’s being informed:

\[
\pi^I_1 = \frac{P[t_n = \text{Buy}|V = 1, \tau_n = I]}{P[t_n = \text{Buy}, V = 1, \tau_n = I]}
= \frac{P[t_n = \text{Buy}, V = 1, \tau_n = I] + P[t_n = \text{Sell}, V = 1, \tau_n = I]}{q + (1 - q)\theta}
\]
\[
\pi_0^t = P[t_n = \text{Buy}|V = 0, \tau_n = I] \\
= \frac{P[t_n = \text{Buy}, V = 0, \tau_n = I]}{P[t_n = \text{Buy}, V = 0, \tau_n = I] + P[t_n = \text{Sell}, V = 0, \tau_n = I]} \\
= \frac{1 - q}{1 - q + q\theta}
\]

As in Avery and Zemsky (1998), in the absence of noise traders, the specialist’s zero-profit bid and ask prices would exactly equal each agent’s expected value of the asset, conditional on seeing a Low or High signal, respectively. No trade would take place, as implied by Milgrom and Stokey (1982). The inclusion of noise traders reduces the informational content of any given trade, shrinking the bid-ask spread and creating gains from trade for informed agents.

1.3 The Price Process

Given \( q, \lambda, \) and \( \theta, \) I can solve for \( \pi_1 \) and \( \pi_0 \) in order to derive equation (4):

\[
\text{Price}_n(k) = \frac{\pi_1^k(1 - \pi_1)^{n-k}}{\pi_1^k(1 - \pi_1)^{n-k} + \pi_0^k(1 - \pi_0)^{n-k}}
\]

**Proposition 2.** Unconditional on \( V, \) \( \text{Price}_n(k) \) forms a martingale with respect to the public trade history \( H_n. \)

Proposition 2 establishes that for agents in this market, next-period prices are not predictable conditional on the current public information set. Thus, even in the presence of both a bid-ask spread and a short-sales constraint, the market remains (semi-strong form) efficient. An immediate implication of Proposition 2 is that uninformed agents cannot expect to trade profitably strictly on the observed drift.

**Definition 1.** A price bubble occurs when \( V = 0 \) and \( \text{Price}_n \) converges almost surely to \( P^* > 0 \) as \( n \to \infty. \)

By this definition, I rule out bubbles as finite-lived price inflation generated by chance, as may be the case with a binomial series. This is not to suggest, however, that a bubble continues indefinitely. The trading period in this model implicitly corresponds to the lifetime of an informational event, however defined. Thus, a bubble by Definition 1 lasts as long as the information remains private (e.g., until an earnings announcement at some point in the future).
**Proposition 3.** Under IS I, $\text{Price}_n$ converges almost surely to $V$ as $n \to \infty$.

Proposition 3 is true for all possible values of $q$, $\lambda$, $\theta$, and $V$. The implication is that no matter how noisy the signal is ($q$) or how few informed agents there are ($\lambda$) or how few informed agents are able to sell ($\theta$), as long as all agents correctly identify these parameters, the price of the asset will converge to its fundamental value. No matter how constrained the market is, if agents know the magnitude of the constraint, after a sufficient number of trades the price will approach $V$, and no bubble will occur.

**Proposition 4.** Under IS I, conditional on $V$, expected price converges monotonically to $V$ as $n \to \infty$.

Proposition 4 may be interpreted as a stronger version of Proposition 3. Taken together, the two propositions imply that short-sales constraints, no matter how binding, do not by themselves necessarily generate (persistent) price bubbles. If agents are fully aware of the constraints, both price and expected price converge to the fundamental value. Echoing Diamond and Verrecchia (1987), the propositions expose a limit to Miller (1977): if agents know the extent of the restriction, no bubble will form.

### 1.4 Comparative Statics

Though the market remains efficient despite being short-sales constrained under IS I, the constraint nevertheless affects the distribution of prices. In this subsection, I present first analytic then numerical solutions to determine the effect on prices of increasing the constraint.

#### 1.4.1 Bid and Ask Prices

**Proposition 5.** The bid and ask are increasing with respect to $\theta$.

Proposition 5 implies that as the market becomes more constrained ($\theta$ decreases), the market specialist reacts by reducing both the bid and the ask prices. When fewer agents are able to sell, the probability of a buy increases while the information contained in each buy correspondingly decreases. In a constrained market, a buy order conveys comparatively less (positive) information than a buy order in an equivalent but unconstrained market, so the ask price drops. Similarly, in a constrained market a sell order conveys comparatively more (negative) information than a sell order in
an unconstrained market. The bid and ask prices reflect these relative changes in the informational content of a trade.

The result for the ask price may initially seem counterintuitive. As the market becomes more constrained, one might expect the market maker to push up his selling price, as implied by Miller (1977). This dynamic does not occur in the present context because market maker operates in a competitive environment, and any price above \( \text{Ask}_n \) would get bid down by other dealers.

### 1.4.2 Price Convergence

Propositions 3 and 4 state that under IS I, \( \text{Price}_n \) will converge to the fundamental value \( V \) regardless of the parametrization. However, as it affects both the informational content of a given trade and the order flow of executed trades, the short-sales constraint will affect the observed price path.

**Proposition 6.** Under IS I, the speed at which \( \text{Price}_n \) converges to \( V \) is increasing with respect to \( \theta \).

While Proposition 3 ensures that \( \text{Price}_n \) will converge to \( V \), Proposition 6 indicates rate at which \( \text{Price}_n \) converges decreases as the probability an informed investor is short sales constrained increases. The dual dynamic of more buys but less information in each buy offset each other, with the net effect being comparatively smaller movements in price over time. The information contained in a Buy order decreases to offset the higher probability of a buy. By assumption, agents recognize that the prevalence of Buys may be due not so much to positive signals but to the inability of agents with negative signals to sell. The net effect is that price changes after a given number of trades will be less dramatic. As \( \theta \) goes to zero, there is an increasingly disproportionate percentage of Buys in the trading queue, with each of those Buys revealing decreasingly less information.

Proposition 6 echoes the main result from Diamond and Verrecchia (1987), that the price revelation process slows down as short-sales constraints become more binding. In their model, the decrease in speed is due to agents' “sitting out” their turn in the queue. By contrast, in my model, the decrease in speed is due to the decreased informational content of the remaining trades. In either case, the end result is equivalent.
2 Constraint Uncertainty

As defined in this paper, the short-sales constraint is in practice impossible to measure. Though agents may make informed estimates, the constraint $\theta$ is categorically unobservable. The assumption, then, that agents know the value of $\theta$ with certainty is acknowledged here as an extreme version of rational expectations which, though analytically convenient, is functionally impractical. To this end, I relax the condition in IS I and analyze arguably more realistic scenarios regarding agents’ beliefs about the sales constraint.

Rational expectations does not require agents necessarily know the realization of a random parameter, only its distribution. With this in mind, let us consider the model under Information Structure II (IS II): the short-sales constraint $\theta$ is a random variable and agents know only its distribution $f(\theta)$. The distribution is assumed independent of $V$, which is roughly equivalent to the condition that the factors which determine the probability an agent is short sales constrained are independent of those that determine the firms’ future cash flows.

Let $\theta^*$ represent the realized value of the constraint, with $\pi_i^*$ equal to the corresponding (true) probability of a buy conditional on $V = i$. Under the assumption that agents only know the distribution but not realization of $\theta$, $\pi_i^*$ is unobservable to agents. Instead, agents must calculate the probability of a buy for each possible value of $\theta$:

$$\pi_1^I(\theta) = \frac{q}{q + (1-q)\theta}$$
$$\pi_0^I(\theta) = \frac{1-q}{1 - q + q\theta}$$

and

$$\pi_1(\theta) = P[Buy|V = 1, \theta] = \lambda \pi_1^I(\theta) + (1 - \lambda) \frac{1}{1 + \theta}$$
$$\pi_0(\theta) = P[Buy|V = 0, \theta] = \lambda \pi_0^I(\theta) + (1 - \lambda) \frac{1}{1 + \theta}$$

To solve for prices, note that I can rewrite
\[ \text{Price}_n = \int P[V = 1|\theta, H_n]f(\theta)d\theta \]
\[ = \int \text{Price}_{n|\theta}f(\theta)d\theta \]

where \( \text{Price}_{n|\theta} \) is the price after \( n \) trades if agents believed the constraint were \( \theta \). The limit of \( \text{Price}_n \) is simply a weighted average of the limits of \( \text{Price}_{n,\theta} \) for each \( \theta \).

Integrating over all values of \( \theta \), I have

\[ \text{Price}_n = \int \frac{[\pi_1(\theta)]^k[1 - \pi_1(\theta)]^{n-k} + [\pi_0(\theta)]^k[1 - \pi_0(\theta)]^{n-k-1}f(\theta)d\theta} {[\pi_1(\theta)]^k[1 - \pi_1(\theta)]^{n-k+1} + [\pi_0(\theta)]^k[1 - \pi_0(\theta)]^{n-k-1}} \quad (7) \]

\[ \text{Bid}_{n+1} = \int \frac{[\pi_1(\theta)]^k[1 - \pi_1(\theta)]^{n-k+1} + [\pi_0(\theta)]^k[1 - \pi_0(\theta)]^{n-k}f(\theta)d\theta} {[\pi_1(\theta)]^k[1 - \pi_1(\theta)]^{n-k+1} + [\pi_0(\theta)]^k[1 - \pi_0(\theta)]^{n-k+1}} \quad (8) \]

\[ \text{Ask}_{n+1} = \int \frac{[\pi_1(\theta)]^{k+1}[1 - \pi_1(\theta)]^{n-k} + [\pi_0(\theta)]^{k+1}[1 - \pi_0(\theta)]^{n-k}f(\theta)d\theta} {[\pi_1(\theta)]^{k+1}[1 - \pi_1(\theta)]^{n-k+1} + [\pi_0(\theta)]^{k+1}[1 - \pi_0(\theta)]^{n-k+1}} \quad (9) \]

Since \( \pi_1^*(\theta) > \pi_1(\theta) \) and \( \pi_0(\theta) > \pi_0^*(\theta) \) for all \( \theta \), it must be the case that

\[ E[V|x_n = \text{High}, \text{Price}_{n-1}] > \text{Ask}_n \]

\[ E[V|x_n = \text{Low}, \text{Price}_{n-1}] < \text{Bid}_n \]

Proposition 1 still holds: informed agents follow their signal. Additionally, the martingale property of Proposition 2 holds. The market is efficient with respect to the public information set.

Of primary interest is the limit of the market price: \( \lim_{n \to \infty} \text{Price}_{n|\theta} \). While the price reflects agents’ aggregated beliefs over the distribution of \( \theta \), the underlying stochastic process generating the order flow depends on \( \theta^* \). The true probability of a Buy is either \( \pi_0^* \) or \( \pi_1^* \), both functions of \( \theta^* \). The buy-sell ratio is consequently governed by \( \pi_0^* \) or \( \pi_1^* \), but prices are determined by solving for \( \text{Price}_{n|\theta} \) for each \( \theta \). This discrepancy yields

**Proposition 7.** Given some \( \theta \),

1. Conditional on \( V = 0 \), there exists some \( \theta_0 \in (\theta^*, \theta^* q^2/(1-q)^2) \) such that \( \text{Price}_{n|\theta} \) converges almost surely to 0 if \( \theta < \theta_0 \) and 1 if \( \theta > \theta_0 \).

2. Conditional on \( V = 1 \), there exists some \( \theta_1 \in (\theta^*(1-q)^2/q^2, \theta^*) \) such that \( \text{Price}_{n|\theta} \) converges almost surely to 1 if \( \theta > \theta_1 \) and 0 if \( \theta < \theta_1 \).
Proposition 7 states that for each \( \theta \), \( Price_{n|\theta} \) converges to either 0 or 1. This implies the limit of the price is

\[
\lim_{n \to \infty} \int Price_{n|\theta} f(\theta) d\theta
\]

The market price converges to the weighted average of the limit of the price for each \( \theta \). For any parameters \( q, \theta^* \), and \( \lambda \), in order for \( Price_{n|\theta} \) to converge to \( V \), \( \theta \) must lie in the interval between \( \theta_1 \) and \( \theta_0 \), with \( \theta_1 < \theta_0 \). Estimating \( \theta \) presents a dilemma for agents in this market: if the short-sales constraint is sufficiently underestimated (\( \theta \) is too high), the price may converge to 1 even if the fundamental value is 0. Similarly, if the constraint is sufficiently overestimated (\( \theta \) is too low), the price may converge to 0 even if the fundamental value is 1.\(^7\) The condition necessary for price to converge to \( V \) is

\[
\theta^*(1-q)^2/q^2 < \theta < \theta_0 < \theta^*q^2/(1-q)^2
\]

Figure 2 plots the numerical solutions for \( \theta_1 \) and \( \theta_0 \) over all \((q, \theta^*)\) combinations for \( \lambda = 0.9 \). (As noted in the proof, there are no closed form solutions for \( \theta_0 \) or \( \theta_1 \).) The upper surface represents \( \theta_0 \) while the lower surface represents \( \theta_1 \). Values of \( \theta_0 \) greater than unity are truncated to 1, a functionally equivalent measure given that \( \theta \in (0,1] \).

To interpret the plots: given any ordered pair \((q, \theta^*)\), locate the corresponding point on the \( q-\theta^* \) plane and imagine a line orthogonal to the plane at that point. \( \theta \) will lie somewhere on this line. For \( Price_{n|\theta} \) to converge almost surely to \( V \), \( \theta \) must fall on the interval between the two surfaces. Examining Figure 2, one observes that the interval \((\theta_1, \theta_0)\) is smallest when either \( q \) or \( \theta^* \) is at a minimum. When the signal is very noisy (\( q \) close to 0.5) or the market is very constrained (\( \theta^* \) close to 0), even a slight miscalculation in estimating \( \theta \) sends prices in the wrong direction.

For each \( \theta \), the associated price converges either to 0 or to 1, so the market price of the asset is a weighted average of zeroes and ones. This result, together with Definition 1, implies

Remark 1. If \( V = 0 \) and there exists a \( \theta \) such that \( f(\theta) > 0 \) and \( \theta > \theta_0 \), a price bubble occurs.\(^8\) The degree of mispricing is governed by the distribution \( f(\theta) \).

\(^7\)One might argue, albeit completely subjectively, that this latter case is not particularly likely.
\(^8\)The opposite result is true as well: if \( V = 1 \) and there exists a \( \theta \) such that \( f(\theta) > 0 \) and \( \theta < \theta_1 \), a price crash occurs.
The upper surface represents $\theta_0$ while the lower surface represents $\theta_1$

Figure 2: Numerical solutions for $\theta_0$ and $\theta_1$ over all possible combinations of $q$ and $\theta^*$ ($\lambda$ fixed at 0.9).

The intuition of Remark 1 is simple. The market price is a weighted average of the prices associated with possible values of $\theta$, where the weights are probabilities associated with each $\theta$. If the price conditional on some $\theta$ converges to 1 when the asset is fundamentally valueless, the fair price will be a convex combination of zeroes and ones.

Remark 1 is the central idea of the paper: with sequential trade, a price bubble generated by the imposition of short-sales constraints reflects not only the disproportionate trade of optimistic investors but also the inability of investors to distinguish whether that optimism is a result of positive information or short-sales constraints. Absent any
uncertainty about the short-sales constraint, each successive investor doesn’t interpret prior buy orders as positive information so much as the inability of negative information to reach the market. There is a reduction in the informational efficiency of trades but no systematic long-term bias in prices. What constraint uncertainty adds is the possibility, from the investor’s point of view, that the prior buy orders are in fact the result of positive information. More specifically, the investor is unable to distinguish whether the observed order flow is due to positive information or short-sales constraints, and consequently her beliefs are a weighted average across both possibilities.

Notice that the greater $\theta^*$ is (the less the market is short-sales constrained), the greater the upper bond on $\theta_0$ and, consequently, the less likely a given $\theta$ is to exceed $\theta_0$. This is perhaps intuitively trivial: the less the market is constrained from short-selling, the less likely a is a constraint-induced bubble.

Revisiting the example in the introduction more formally, suppose $V = 0$, $\theta^* = 0.5$, and $q = 0.51$: the asset is has no value, the short-sales constraint is binding, and the signal is very noisy. $\theta$ is known to be either 0.5 or 1 with equal probability. To solve for $\lim_{n \to \infty} Price_n$, we must first solve for the limits of $Price_n|\theta$ for each $\theta$. For $\theta = 0.5$, $Price_n|\theta=0.5$ converges to 0, as implied by Proposition 3. If investors calculate prices with the realized value of the constraint, no long-term mispricing occurs. For $\theta = 1$, however, $Price_n|\theta=1$ converges to 1, as implied by Proposition 7. (According to Equation 10, the upper bound on $\theta_0$ is $0.5 \cdot 0.51^2 / 0.49^2 \approx 0.542$.) $Price_n|\theta=1$ represents the market price if agents assume there are no short-sales constraints (i.e., any pessimistic seller can trade). Order flow, however, is governed by $\theta^* = 0.5$, which along with $q = 0.51$ implies that an executed order is more likely to be a buy than a sell. In the limit, the majority of trades will be buys, and by construction $Price_n|\theta=1$ does not account for this bias towards buys, so $Price_n|\theta=1$ converges to 1. Since the fair price is given by $\int Price_n|\theta f(\theta) d\theta$, $\lim Price_n = 0.5(0) + 0.5(1) = 0.5$: a bubble.

3 Discussion

3.1 Price Paths

To give some intuition to the analysis in Section 2, I present expected price paths, conditional on $V$. Figure 3 plots expected price paths under $IS I$ for $V \in \{0, 1\}$ and $\theta \in \{0.5, 1\}$, for $q = 0.55$ and $\lambda = 0.75$. The short-sales constraint reduces the speed at which the expected price is converging to $V$. With fewer agents are able to sell, executed
trades from the informed population will increasingly reflect agents with High signals wanting to buy. The probability of a Buy increases while the informational content of those Buys decreases; i.e., \( P[Buy|V] \) increases as \( \theta \) decreases. In the limit, as \( \theta \to 0 \), no informed agent can sell, and no information can be inferred from a trade. The specialist would always set the bid and ask prices equal to \( 1/2 \), the unconditional expected value of \( V \).

Of greater interest are price paths under IS II. In Figure 4, I plot price paths under the distribution for \( \theta \) presented in the preceding example: \( \theta \in \{0.5, 1\} \) with equal probability. Again, I fix the parameters \( q = 0.55 \) and \( \lambda = 0.75 \) and plot expected price paths conditional on \( V \) and \( \theta^* \). The resulting plots correspond to

\[
E[Price_n|V] = \sum_{k=0}^{n} Bin(n, k, \pi_V) \frac{1}{2} [Price_{n|\theta=0.5} + Price_{n|\theta=1}]
\]

where \( Bin(\cdot) \) is the binomial probability distribution function. The plot for \( V = 0, \theta^* = 0.5 \) represents the short-sales constraint-generated bubble. The initial upward trajectory is generated as the
3.2 Durability of a Bubble

If short-sales constraints are both unobservable and underestimated, the bubble is “rational”. Unlike the irrational bubbles literature, informed agents’ following their private signals generate the mispricing. The only irrational agents in this model are noise traders, but these agents serve no purpose other than to facilitate trade. There is no sense in which noise traders are responsible for the price inflation. At the same time, unlike the existing rational bubbles literature, the bubble cannot be ruled out via backwards induction, since at no point does an agent know with certainty a bubble is occurring.

In comparison to the Avery and Zemsky (1998) paper on which my model is based, both can be classified under the “noisy rational expectations” heading. There are multiple sources of uncertainty, and prices cannot account for both sources, potentially generating a price bubble. In contrast, however, there are no perfectly informed agents in my model, implying arbitrage is not possible. Additionally, there is no information cascade or herd behavior generating this bubble (or its collapse). Though the price paths generated under Proposition 7 may appear to be herd behavior to an observer.
of the market, prices will in fact reflect informed agents’ following their private signals insofar as they are capable. As such, the bubble in my model is not necessarily fragile with respect to increases in \( q \), the accuracy of private signals. As long as agents sufficiently underestimate the short-sales constraint, the price inflation can persist.

The durability of the bubble warrants further examination. Proposition 7 states that under certain conditions, a price bubble occurs almost surely. Suppose these conditions are met (agents sufficiently underestimate the short-sales constraint). If, by whatever means, some agent becomes informed of the true value of \( \theta \), she may have no recourse by which to profit. Assuming she can find the shares to short, she knows the bubble will only increase over the foreseeable horizon. According to the theory, prices will decrease only when either the signal accuracy or the asset float increases, neither of which is guaranteed. In the interim, she must cover her short position which, depending on the range of \( V \), could become prohibitively expensive.

3.3 Potential Caveats

Given the assumptions required to remain tractable, the model unfortunately does not circumvent the full range of caveats associated with the models derived from the Glosten and Milgrom (1985) paradigm. In this section, I address potential areas of concern.

3.3.1 Volume is not relevant

The dual assumptions that agents are restricted to unit trades and that the specialist demands and supplies the asset infinitely are problematic in that they render the concept of volume irrelevant at both the aggregate and individual trading levels. By “aggregate” I mean that there is no sense in which the dealer considers inventory when setting bid and ask prices. This is undoubtedly an unrealistic simplification. Consider, for example, Comerton-Forde, Hendershott, Jones, Moulton, and Seasholes (2010), who find evidence that when specialists hold large positions – long or short – effective bid and ask spreads tend to increase. By “individual” I mean that there is no differentiation between small and large trades. As noted earlier, to the extent that specialists’ quoted prices are valid up to some predefined range, the model may be considered accurate. While empirical evidence shows high volume trades transact at less favorable prices (e.g., Dann, Mayers, and Raab, 1977; Holthausen, Leftwich, and Mayers, 1987), these trades generally take place via syndication (O’Hara, 1995), outside the market, and for this reason, they are not directly relevant to the model.
While there is no parsimonious solution to the aggregate problem, the individual problem could be addressed by correlating trade size with informational content, a la Easley and O'Hara (1987). For example, there could be two sizes of trades, large and small, and two levels of private signal accuracy, high and moderate. Each trade size would be correlated with a trader's type: highly informed, moderately informed, and uninformed. Unfortunately, including yet another source of uncertainty would add tremendous complexity to the model, at the cost of expository clarity.

A potential alternative would be to regard the model as pertaining strictly to block trades, with the smaller trades simply represented by a white noise process that doesn't significantly affect prices. If one makes the assumption that each block trade contains a fixed amount of information $q$, the model still holds, and the volume critique may be attenuated.

3.3.2 The sequence of agents is random

It would seem unlikely that in practice the arrival of informed agents is random. Even in this stylized model, the largest expected profits are made by the earliest traders. From the point of view of the specialist, however, if informed agents receive a private signal with uniform accuracy $q$, the order in which they trade may as well be random since each informed trade adds the same amount to the public information set.

That said, the model could easily be modified to account for a non-random sequence. In particular, I could replace $q$ with $q_n$, the $n$th informed trader's signal accuracy, and make an assumption about the law of motion for $q_n$. Earlier agents could have superior information, so $q_n$ would decrease over time. Alternatively, earlier agents could have inferior information, and $q_n$ would increase over time.

3.3.3 The inclusion of noise traders

In general, information-based game theory models must somehow circumvent the so-called “no-trade” theorem of Milgrom and Stokey (1982). Various ad-hoc tactics have been employed to facilitate trade, from noise traders to gains from trade (certain agents have higher utility from holding the asset than others; e.g., Glosten and Milgrom, 1985) to investor overconfidence (e.g., Scheinkman and Xiong, 2003), though one could argue that none of these solutions is entirely satisfying.

In the context of my model, including agents whose motivations are unmodeled is certainly not ideal. At the same time, there are a myriad of such motivations for which
to account. An individual may need to liquidate assets for consumption purposes. A mutual fund manager may need to purchase (sell) shares to rebalance the portfolio after a cash inflow (outflow). A technical analyst may buy or sell shares based on the recent price trend. The inclusion of noise traders allows us to account for such contingencies without sacrificing pedagogical simplicity. Importantly, the noise traders in this model do not generate the abnormal returns. They simply provide the mathematical lubrication for trade with informed agents to be possible.

4 Conclusion

In this paper, I examine the effects of short-sales constraints on the market for a risky security. When the magnitude of the constraint is common knowledge, the market price will converge to the asset’s fundamental value regardless of the parameters. The constraint becoming more binding affects only the speed at which the price converges, a result consistent with Diamond and Verrecchia (1987).

That agents know the extent of the short-sales constraint is ultimately a straw man proposition, given that the constraint is not observable, even in hindsight. More realistically, agents must estimate it, presumably with error. If there is sufficient uncertainty about the constraint, a multi-period version of the Miller (1977) argument will hold: short-sales constraints lead to too much positive information reaching the market, generating a price bubble. As in Miller (1977), I find that bubbles are more likely when

1. signals about the asset’s value are noisy (heterogeneous beliefs) and

2. short-sales constraints bind

This paper specifies an additional requirement for a bubble:

3. signals about the constraints are noisy

A bubble is predicated on the information set of the representative investor, whose qualities vary across markets and time. An investor with superior information than the representative investor’s may not be able to trade profitably on the basis of that ostensible advantage, given that the price path depends on the representative investor’s beliefs. In contrast to the behavioral finance literature, a bubble generated under these conditions is not the result of irrational investors. Rational investors following their signals create the discrepancy between the asset’s price and its fundamental value.

---

This is particularly true for index funds.


Appendix

Proof of Proposition 1. As implied by equations (1) - (3), given any observed \( Price_{n-1} = y/(y+z) \), the subsequent bid and ask prices will be

\[
Bid_n = \frac{(1-\pi_1)y}{(1-\pi_1)y + (1-\pi_0)z}
\]

\[
Ask_n = \frac{\pi_1 y}{\pi_1 y + \pi_0 z}
\]

Note that \( Price_{n-1} \) represents the public posterior after \( n-1 \) trades. Based on her private signal \( x_n \) and \( Price_{n-1} \), an informed agent calculates:

\[
E[V|x_n = High, Price_{n-1}] = P[V = 1|x_n = High, Price_{n-1}]
\]

\[
= \frac{qy}{qy + (1-q)z}
\]

\[
E[V|x_n = Low, Price_{n-1}] = P[V = 1|x_n = Low, Price_{n-1}]
\]

\[
= \frac{(1-q)y}{(1-q)y + qz}
\]

The agent buys if

\[
E[V|x_n = High, Price_{n-1}] > Ask_n
\]

\[
= \frac{qy}{qy + (1-q)z} > \frac{\pi_1 y}{\pi_1 y + \pi_0 z}
\]

\[
= q(\pi_1 y + \pi_0 z) > \pi_1(qy + (1-q)z)
\]

\[
q \pi_0 > \pi_1(1-q)
\]

which holds, given \( q > \pi_1 \) and \( \pi_0 > 1-q \). Similarly, the agent sells if

\[
E[V|x_n = Low, Price_{n-1}] < Bid_n
\]

\[
= \frac{(1-q)y}{(1-q)y + qz} < \frac{(1-\pi_1)y}{(1-\pi_1)y + (1-\pi_0)z}
\]

\[
= \frac{(1-q)(1-\pi_0)}{(1-\pi_1)} < (1-\pi_1)q
\]

which holds, given \( 1-q < 1-\pi_1 \) and \( 1-\pi_0 < q \). □

Proof of Proposition 2. (Cipriani and Guarino, 2003) Recall that \( Price_n = E[V|H_n] \).
This implies

$$E[\text{Price}_{n+1}|H_n] = E[E[V|n+1]|H_n] = E[V|H_n] = \text{Price}_n$$

where $E[E[V|H_{n+1}]|H_n] = E[V|H_n]$ is given by the law of iterated expectations.

Proof of Proposition 3. (See Appendix to Chapter 3, O’Hara, 1995) Without loss of generality, let us consider the case where $V = 1$ (a symmetric proof holds for $V = 0$). Given that $\text{Price}_n = P[V = 1|(n,k)]$, it is enough to show that the agents’ posterior $P[V = 1|(n,k)]$ converges almost surely to 1. As such, I can examine the ratio

$$\frac{P[V = 0|(n,k)]}{P[V = 1|(n,k)]} = \frac{P[V = 0]_0}{P[V = 1]_1} (1 - \pi_0)^{n-k}$$

Taking logs,

$$\log \left( \frac{P[V = 0|(n,k)]}{P[V = 1|(n,k)]} \right) = \log \left( \frac{P[V = 0]}{P[V = 1]} \right) + k \log \pi_0 + (n - k) \log (1 - \pi_0) - k \log \pi_1 - (n - k) \log (1 - \pi_1)$$

$$= \log \left( \frac{P[V = 0]}{P[V = 1]} \right) + k \log \left( \frac{\pi_0}{\pi_1} \right) + (n - k) \log \left( \frac{1 - \pi_0}{1 - \pi_1} \right)$$

Divide both sides by $n$:

$$\frac{1}{n} \log \left( \frac{P[V = 0|(n,k)]}{P[V = 1|(n,k)]} \right) = \frac{1}{n} \log \left( \frac{P[V = 0]}{P[V = 1]} \right) \text{ goes to } 0 \text{ as } n \to \infty$$

$$+ \frac{n}{n} \log \left( \frac{\pi_0}{\pi_1} \right) \xrightarrow{a.s.} \pi_1$$

$$+ \frac{n - k}{n} \log \left( \frac{1 - \pi_0}{1 - \pi_1} \right) \xrightarrow{a.s.} (1 - \pi_1)$$

Taking limits,

$$\frac{1}{n} \log \left( \frac{P[V = 0|(n,k)]}{P[V = 1|(n,k)]} \right) \xrightarrow{a.s.} \pi_1 \log \left( \frac{\pi_0}{\pi_1} \right) + (1 - \pi_1) \log \left( \frac{1 - \pi_0}{1 - \pi_1} \right)$$

25
If the right hand side of (12) is negative, it must be the case that \( \log \left( \frac{P[V=0|(n,k)]}{P[V=1|(n,k)]} \right) \) is going to negative infinity, which implies that \( P[V = 1|(n,k)] \xrightarrow{a.s.} 1 \), our desired result.

For any two probabilities \( q \) and \( p \), relative entropy (Kullback-Leibler divergence) is defined as

\[
I_q(p) \equiv q \log \left( \frac{q}{p} \right) + (1 - q) \log \left( \frac{1 - q}{1 - p} \right)
\]

where \( I_q(p) > 0 \) for all \( q \neq p \). So I know

\[
I \equiv \pi_1 \log \left( \frac{\pi_1}{\pi_0} \right) + (1 - \pi_1) \log \left( \frac{1 - \pi_1}{1 - \pi_0} \right) > 0
\]

More specifically, I have shown

\[
-I \equiv \pi_1 \log \left( \frac{\pi_0}{\pi_1} \right) + (1 - \pi_1) \log \left( \frac{1 - \pi_0}{1 - \pi_1} \right) < 0
\]  

(13)

Given (13), the right hand side of (12) is negative, implying \( P[V = 1|(n,k)] \xrightarrow{a.s.} 1 \).

\[\square\]

**Proof of Proposition 4.** With \( V \) normalized to be either 0 or 1, the market price is equal to the probability that the true value is 1 given \( k \) buys in \( n \) trades: \( \text{Price}_n(k) = P[V = 1|(n,k)] \). Prices represent probabilities and consequently are bounded between 0 and 1, inclusive. Given Proposition 3 (\( \text{Price}_n \) converges almost surely to \( V \)), to demonstrate the expected price converges monotonically to \( V \) one needs only to show

1. \( E[\text{Price}_n|V = 0] \) is strictly decreasing with respect to \( n \) and
2. \( E[\text{Price}_n|V = 1] \) is strictly increasing with respect to \( n \).

Note that for any \( (n,k) \), \( \text{Price}_n(k) \) can be rewritten as

\[
\text{Price}_n(k) = \frac{y}{y + z} \tag{14}
\]

where \( y = \pi_1^k(1 - \pi_1)^{n-k} \) and \( z = \pi_0^k(1 - \pi_0)^{n-k} \).

Without loss of generality, let us consider the case where \( V = 0 \) (a symmetric argument holds for \( V = 1 \)). Given any \( (n,k) \), I can calculate the expected price after
In order to show $E[\text{Price}_n | V = 0]$ is decreasing over $n$, I need to find the conditions under which

$$\text{Price}_n(k) > E[\text{Price}_{n+1} | V = 0, (n,k)] \tag{17}$$

Plugging (14) and (15) into (17),

$$\frac{y}{y+z} - \pi_0 \frac{\pi_1 y}{\pi_1 y + \pi_0 z} - (1-\pi_0) \frac{(1-\pi_1) y}{(1-\pi_1) y + (1-\pi_0) z} > 0 \tag{18}$$

Reducing (18), I find $\text{Price}_n(k) \geq E[\text{Price}_{n+1} | V = 0, (n,k)]$ when

$$\pi_1^2 - 2\pi_1\pi_0 + \pi_0^2 > 0 \tag{19}$$

Given the constraints on $q, \theta$, and $\lambda$, it must be the case that $0.5 < \pi_1 < 1$ and $0 < \pi_0 < \pi_1$. Thus, (17) is satisfied for all possible values of $q, \theta$, and $\lambda$.

Satisfaction of (17) implies

$$\sum_{k=0}^{n} P[(n,k)|V = 0] \text{Price}_n(k) > \sum_{k=0}^{n} P[(n,k)|V = 0] E[\text{Price}_{n+1} | V = 0, (n,k)] \tag{20}$$

As the left-hand side is equal to $E[\text{Price}_n | V = 0]$ and the right-hand side is equal to $E[\text{Price}_{n+1} | V = 0]$, (20) can be rewritten as

$$E[\text{Price}_n | V = 0] > E[\text{Price}_{n+1} | V = 0]$$

Proof of Proposition 5. For any $\text{Price}_{n-1} = y/(y+z)$, bid and ask prices for the $n$th
trade are

\[ Ask_n = \frac{\pi_1 y}{\pi_1 y + \pi_0 z} \]

\[ Bid_n = \frac{(1 - \pi_1) y}{(1 - \pi_1) y + (1 - \pi_0) z} \]

The corresponding first derivatives are

\[
\frac{\partial Ask_n}{\partial \theta} = \frac{\frac{\partial \pi_1}{\partial \theta} y (\pi_1 y + \pi_0 z) - \pi_1 y \left( \frac{\partial \pi_1}{\partial \theta} y + \frac{\partial \pi_0}{\partial \theta} z \right)}{(\pi_1 y + \pi_0 z)^2}
= \frac{[\frac{\partial \pi_1}{\partial \theta} \pi_0 - \frac{\partial \pi_0}{\partial \theta} \pi_1] yz}{(\pi_1 y + \pi_0 z)^2}
\]

\[
\frac{\partial Bid_n}{\partial \theta} = \frac{-\frac{\partial \pi_1}{\partial \theta} y ((1 - \pi_1) y + (1 - \pi_0) z) - (1 - \pi_1) y \left( -\frac{\partial \pi_1}{\partial \theta} y - \frac{\partial \pi_0}{\partial \theta} z \right)}{((1 - \pi_1) y + (1 - \pi_0) z)^2}
= \frac{\left[-\frac{\partial \pi_1}{\partial \theta} (1 - \pi_1) + \frac{\partial \pi_0}{\partial \theta} (1 - \pi_1) \right] yz}{((1 - \pi_1) y + (1 - \pi_0) z)^2}
\]  \hspace{1cm} (21)

Since both \( \pi_0 \) and \( \pi_1 \) are between 0 and 1, it follows that

\[
\frac{\partial Ask_n}{\partial \theta} > 0 \iff \frac{\partial \pi_1}{\partial \theta} \pi_0 - \frac{\partial \pi_0}{\partial \theta} \pi_1 > 0 \]  \hspace{1cm} (23)

\[
\frac{\partial Bid_n}{\partial \theta} > 0 \iff -\frac{\partial \pi_1}{\partial \theta} (1 - \pi_0) + \frac{\partial \pi_0}{\partial \theta} (1 - \pi_1) > 0 \]  \hspace{1cm} (24)

Solving for the derivative of \( \pi_V \) with respect to \( \theta \), I have

\[
\frac{\partial \pi_1}{\partial \theta} = -\lambda \frac{q (1 - q)}{(q + (1 - q) \theta)^2} - (1 - \lambda) \frac{1}{(1 + \theta)^2} < 0
\]

\[
\frac{\partial \pi_0}{\partial \theta} = -\lambda \frac{(1 - q) q}{(1 - q + q \theta)^2} - (1 - \lambda) \frac{1}{(1 + \theta)^2} < 0
\]

I can substitute into Equation (23) and find \( \frac{\partial \pi_1}{\partial \theta} \pi_0 - (\frac{\partial \pi_1}{\partial \theta} \pi_0) > 0 \), implying \( \frac{\partial Ask_n}{\partial \theta} > 0 \).

Solving for the sign of the partial derivative of the Bid is substantially more complicated. I can rewrite Equation 24 as

\[
\frac{\partial \pi_0}{\partial \theta} (1 - \pi_1) > \frac{\partial \pi_1}{\partial \theta} (1 - \pi_0)
\]

After some lengthy algebra, one may verify that the inequality holds, implying \( \frac{\partial Bid_n}{\partial \theta} > 0 \).
Proof of Proposition 6. The proof follows directly from the proof of Proposition 3. For $V = 1$, I found
\[
\frac{1}{n} \log \left( \frac{P[V = 0|(n, k)]}{P[V = 1|(n, k)]} \right) \xrightarrow{a.s.} -I
\]
where $-I = \pi_1 \log \left( \frac{\pi_0}{\pi_1} \right) + (1 - \pi_1) \log \left( \frac{1 - \pi_0}{1 - \pi_1} \right)$. Applying the exponential function to (25), I have
\[
\left( \frac{P[V = 0|(n, k)]}{P[V = 1|(n, k)]} \right)^{n-1} \xrightarrow{a.s.} e^{-I}
\]
\[
\frac{P[V = 0|(n, k)]}{P[V = 1|(n, k)]} \xrightarrow{a.s.} e^{-ln}
\]
The ratio is converging at rate $-I$.

To complete the proof, I need only show that the derivative of $I$ is increasing with respect to $\theta$. Recall that relative entropy measures the distance between two probabilities and is therefore increasing in absolute value as the difference between the probabilities increases. One may verify that $\partial \pi_1 / \partial \theta - \partial \pi_0 / \partial \theta$ is positive, implying $\partial I / \partial \theta$ is positive as well. A symmetric argument holds for $V = 0$. \hfill \Box

Proof of Proposition 7. For each $\theta$ with $f(\theta) > 0$, agents calculate the “possible” probabilities $\pi_0(\theta)$ and $\pi_1(\theta)$ which, except for $\theta = \theta^*$, are not equal to the corresponding “true” probabilities $\pi_0^*$ and $\pi_1^*$. The possible probabilities govern the price process, while the true probabilities govern the distribution of trades.

Without loss of generality, let us consider the case where $V = 0$ (a symmetric argument holds for $V = 1$). I (again) examine the ratio:
\[
\frac{P[V = 0|(n, k), \theta]}{P[V = 1|(n, k), \theta]} = \frac{P[V = 0](\pi_0(\theta))^k(1 - \pi_0(\theta))^{n-k}}{P[V = 1](\pi_1(\theta))^k(1 - \pi_1(\theta))^{n-k}}
\]
Taking logs and dividing both sides by $n$, in the limit I have the IS II equivalent of equation (12):
\[
\frac{1}{n} \log \left( \frac{P[V = 0|(n, k), \theta]}{P[V = 1|(n, k), \theta]} \right) \xrightarrow{a.s.} \pi_0^* \log \left( \frac{\pi_0(\theta)}{\pi_1(\theta)} \right) + (1 - \pi_0^*) \log \left( \frac{1 - \pi_0(\theta)}{1 - \pi_1(\theta)} \right)
\]
(26)
Since $\pi_0(\theta) < \pi_1(\theta)$, it must be the case that $\log(\pi_0(\theta)/\pi_1(\theta)) < 0$ and $\log((1 - \pi_0(\theta))/(1 - \pi_1(\theta))) > 0$. Rewrite the right hand side of (26) as

$$I(\theta) \equiv \pi_0^* \log\left(\frac{\pi_0(\theta)}{\pi_1(\theta)}\right) + (1 - \pi_0^*) \log\left(\frac{1 - \pi_0(\theta)}{1 - \pi_1(\theta)}\right)$$  \hspace{0.5cm} (27)

$I(\theta)$ is not relative entropy. But recall that relative entropy represents a convex combination of two components – one positive, one negative. $I(\theta)$ is a reweighted convex combination of those components and is significant in that it gives us a basis to determine the sign of (26):

If $I(\theta) < 0$, it must be the case that $\log\left(\frac{P[V=0|n,k,\theta]}{P[V=1|n,k,\theta]}\right)$ is going to negative infinity, which implies that $P[V = 1|(n,k),\theta] \xrightarrow{a.s.} 1$.

If $I(\theta) > 0$, it must be the case that $\log\left(\frac{P[V=0|n,k,\theta]}{P[V=1|n,k,\theta]}\right)$ is going to positive infinity, which implies that $P[V = 1|(n,k),\theta] \xrightarrow{a.s.} 0$.

To begin, suppose $\theta > \theta^* q^2/(1 - q)^2$. This is equivalent to the condition that $\pi_0^* > \pi_1(\theta)$, so I can write $\pi_0^* = \pi_1(\theta) + \varepsilon$ for some $\varepsilon > 0$. Bearing in mind the properties of relative entropy, it must be the case that

$$\pi_1(\theta) \log\left(\frac{\pi_0(\theta)}{\pi_1(\theta)}\right) + (1 - \pi_1(\theta)) \log\left(\frac{1 - \pi_0(\theta)}{1 - \pi_1(\theta)}\right) < 0$$  \hspace{0.5cm} (28)

Rewriting (27):

$$I(\theta) = (\pi_1(\theta) + \varepsilon) \log\left(\frac{\pi_0(\theta)}{\pi_1(\theta)}\right) + (1 - \pi_1(\theta) - \varepsilon) \log\left(\frac{1 - \pi_0(\theta)}{1 - \pi_1(\theta)}\right) < 0$$  \hspace{0.5cm} (29)

Comparing (29) to (28), note that the weight on the negative component increases while the weight on the positive component decreases, implying the left hand side of (29) is less than that of (28). Having shown the right hand side of (26) is negative, one concludes that $Price_{n|\theta}$ converges almost surely to 1.

Alternatively, suppose $\theta < \theta^*$. This is equivalent to the condition that $\pi_0^* < \pi_0(\theta)$, so I can write $\pi_0^* = \pi_0(\theta) - \varepsilon$ for some $\varepsilon > 0$. Again, bearing in mind the properties of relative entropy, it must be the case that

$$\pi_0(\theta) \log\left(\frac{\pi_0(\theta)}{\pi_1(\theta)}\right) + (1 - \pi_0(\theta)) \log\left(\frac{1 - \pi_0(\theta)}{1 - \pi_1(\theta)}\right) > 0$$  \hspace{0.5cm} (30)
Rewriting (27), I have

\[ I(\theta) = (\pi_0(\theta) - \varepsilon) \log \left( \frac{\pi_0(\theta)}{\pi_1(\theta)} \right) + (1 - \pi_0(\theta) + \varepsilon) \log \left( \frac{1 - \pi_0(\theta)}{1 - \pi_1(\theta)} \right) > 0 \]  

(31)

Comparing (31) to (30), I note that the weight on the negative component decreases while the weight on the positive component increases, implying the left hand side of (31) is greater than that of (30). Having shown the right hand side of (26) is positive, I conclude that \( Price_{n,\theta} \) converges almost surely to 0.

To complete the proof, suppose \( \theta \) is equal to exactly \( \theta^* q^2 / (1 - q)^2 \), in which case I know \( I(\theta) \) is negative. Consider the effect on \( I(\theta) \) of decreasing \( \theta \): ceteris paribus, the negative component \( \log(\pi_0(\theta) / \pi_1(\theta)) \) decreases in magnitude while the positive component \( \log((1 - \pi_0(\theta)) / (1 - \pi_1(\theta))) \) increases in magnitude, implying \( I(\theta) \) increases overall. If I continue to decrease \( \theta \) until \( \theta = \theta^* \), at some point \( I(\theta) \) must become positive since \( I(\theta^*) > 0 \). As \( \pi_0(\theta) / \pi_1(\theta) \) is continuous in \( \theta \) and \( \partial \log(\pi_0(\theta) / \pi_1(\theta)) / \partial \theta < 0 \) for all \( \theta \) (the reader may verify these results), there must be some \( \theta_0 \) between \( \theta^* \) and \( \theta^* q^2 / (1 - q)^2 \) such that \( I(\theta_0) \) equals 0. For \( \theta > \theta_0 \), \( I(\theta) \) is negative (\( Price_{n,\theta} \rightarrow 1 \)); for \( \theta < \theta_0 \), \( I(\theta) \) is positive (\( Price_{n,\theta} \rightarrow 0 \)).

\( \theta_0 \) satisfies the equation:

\[ \pi_0^* \log \left( \frac{\pi_0(\theta_0)}{\pi_1(\theta_0)} \right) + (1 - \pi_0^*) \log \left( \frac{1 - \pi_0(\theta_0)}{1 - \pi_1(\theta_0)} \right) = 0 \]  

(32)

Unfortunately, there is no closed-form solution for \( \theta_0 \) given equation (32). As noted in the text, numerical solutions for \( \lambda = 0.9 \), \( q \in (0.5, 1) \), and \( \theta \in (0, 1) \) are plotted in Figure 2.
References


