The Price Effect of Temporary Short-selling Bans: Theory and Evidence

Haoshu Tian, Xuemin (Sterling) Yan, and Lingling Zheng*

Abstract

We develop a model of temporary short-selling bans by extending the infinite-horizon speculative bubble model of Scheinkman and Xiong (2003). Our model predicts that a temporary short-selling ban leads to a price inflation that is the highest at the beginning of the ban and gradually converges to zero at the end of the ban. Examining the 2008 short-selling ban of financial stocks in the U.S., we find evidence consistent with this prediction. The innovation of our empirical design is to use the financial segments of non-banned stocks as a control group for the banned financial stocks.

November 2021

*Haoshu Tian is affiliated with Renmin University of China, Email: tianhaoshu@foxmail.com. Xuemin (Sterling) Yan is at College of Business, Lehigh University, Email: xuy219@lehigh.edu. Lingling Zheng is at Renmin University of China, Email: zhenglingling@rmbs.ruc.edu.cn. We thank Chris Sims, Harrison Hong, Matthew Ringgenburg, Weinan E, Jose Scheinkman, Wei Xiong, Yingzi Zhu, Zhigang Qiu, and seminar participants at Beijing University, Cheung Kong Graduate School of Business, Princeton University, and Renmin University of China for helpful comments.

1. Introduction

Regulators around the world have increasingly resorted to short-selling bans as a market-stabilizing tool in times of crises. On September 18, 2008, the U.S. Securities and Exchange Commission (SEC) issued an emergency order to temporarily ban the short selling of nearly all financial stocks. Many countries followed suit and imposed similar bans during the global financial crisis (Beber and Pagano (2013)). In August 2011, at the height of Eurozone sovereign debt crisis, four European countries (i.e., Belgium, France, Italy, and Spain) imposed short-selling bans on financial stocks. More recently, in the wake of the COVID-19 pandemic, market regulators in many European and Asian countries once again restricted short sales.

Short selling in general, and short-sale restrictions in particular, have been the subject of extensive academic research both before and after 2008. While there is general consensus that short-selling bans are detrimental to market liquidity and price discovery,[1] there is considerable disagreement about their impact on stock prices. Miller (1977) argues that, in the presence of divergence of opinion, short-sale constraints will lead to overvaluation because pessimistic investors are kept out of the market and prices only reflect the views of optimistic investors. Diamond and Verrecchia (1987) develop a rational expectations model and show that short-sale constraints will not lead to overvaluation because rational agents take into account the fact that investors with negative information are sidelined by short-sale constraints. Bai, Chang, and Wang (2006) show that, when investors are risk averse, short-sale constraints increase the risk perceived (and the expected return required) by uninformed investors, which in turn leads to lower prices. Therefore, in theory, short-selling bans could have a positive, negative, or no effect on prices.

The empirical evidence on the price effect of short-selling bans is equally mixed. Autore, Billingsley, and Kovacs (2011) examine the 2008 short-selling ban of financial stocks in the U.S. and find that, consistent with Miller (1977), banned stocks exhibit positive abnormal

[1] See, e.g., Battalio and Schultz (2011), Beber and Pagano (2013), and Boehmer, Jones, and Zhang (2013).

returns at the initiation of the ban. Battalio, Mehran, and Schultz (2012), however, conclude that the 2008 ban failed to support the price of financial stocks. Beber and Pagano (2013) show that the 2008-2009 short selling bans are not associated with better stock price performance in most countries except the U.S. Boehmer et al. (2013) find that the 2008 ban in the U.S. leads to higher prices for banned financial stocks relative to a matching sample of unbanned stocks, but only for large stocks and not for small stocks. Moreover, the price effect is non-existent for stocks added to the ban list after the initial announcement. Both Beber and Pagano (2013) and Boehmer et al. (2013) attribute the price inflation of the banned stocks in the U.S. to the confounding news about the Troubled Asset Relief Program (TARP) instead of the short-selling ban. Finally, Harris, Namvar, and Phillips (2013) use a factor-model approach and estimate that the 2008 short-selling ban in the U.S. leads to a price inflation of 10-12% in the banned stocks after controlling for a TARP factor.

In this paper, we study the price effect of temporary short-selling bans both theoretically and empirically. We first build a continuous-time asset pricing model with heterogenous beliefs and temporary short-selling bans by extending the work of Scheinkman and Xiong (2003). The key prediction of our model is that a temporary short-selling ban leads to a price inflation that is the highest at the beginning of the ban and gradually converges to zero at the end of the ban. We test this prediction by revisiting the 2008 short-selling ban of financial stocks in the U.S. The innovation of our empirical design is to use the financial segments of non-banned non-financial stocks as a control group for the banned financial stocks. We find evidence consistent with our model's prediction.

The setting of our model is similar to that of Scheinkman and Xiong (2003). There is a single risky asset with limited supply. The dividend for the risky asset is driven by a fundamental variable, which is not observable. There are two noisy signals about the fundamental variable and there are two groups of risk-neutral agents. The signals are public information. However, the two groups of agents interpret the signals differently. Specifically, each group of agents overestimate the informativeness of a different signal. This overconfidence leads to heterogenous beliefs and valuations.

As in Scheinkman and Xiong (2003), owning a share of the risky asset today provides an opportunity to profit from future overvaluation by the other group of agents, i.e., by selling the risky asset to the other group of agents. This resale option has value because short selling is prohibited. As a consequence, the demand price of the agents depends not only on their assessment of the fundamentals but also on the value of this resale option. Scheinkman and Xiong (2003) call the resale option value a "bubble". In this paper, we also refer to it as overpricing, overvaluation, or price inflation.

Scheinkman and Xiong (2003) study an infinite-horizon problem where short-selling is banned permanently. As explained by Scheinkman and Xiong (2003, p.1195), "the infinite horizon problem is stationary, greatly reducing the mathematical difficulty". In practice, short-selling bans are almost always temporary in nature—the duration of the 2008-2009 short-selling bans ranges from 19 days in Canada and U.S. to 277 days in Ireland and Luxembourg (Beber and Pagano (2013)). Departing from Scheinkman and Xiong (2003), we study a finite-horizon problem, which allows us to examine the effect of temporary short-selling bans. In this case, the resale option value, i.e., the magnitude of the overvaluation, also depends on the remaining life of the ban. While the analytical solutions to an infinite-horizon problem can be obtained by solving an ordinary differential equation, one has to solve a partial differential equation for the finite-horizon problem, which is more challenging.

We obtain a closed-form solution for the resale option value by using the Laplace transform and Kummer functions. This solution enables us to study the dynamics of the overpricing caused by a temporary short-selling ban for the first time. Our model predicts that the size of the initial overpricing is increasing in the length of the ban and the degree of overconfidence, and decreasing in the level of interest rate and the speed of the mean reversion in the fundamentals. More importantly, our model predicts that a temporary short-selling ban leads to an overpricing that is the highest at the beginning of the ban and gradually shrinks to zero at the end of the ban.

We test the above prediction by examining the 2008 short-selling ban of financial stocks in the U.S. As stated earlier, an extensive literature has examined the price effect of the 2008 ban with mixed findings. The key challenge in this literature, as argued by Battalio and Schultz (2012 p.3), is that "since almost all financial stocks were targeted by the ban, it is difficult (if not impossible) to find an appropriate benchmark against which to evaluate their returns." Most prior studies employ benchmarks consisting mainly of non-financial stocks. For example, Boehmer et al. (2013) construct a matched control sample based on listing exchange, market capitalization, trading volume, and whether or not listing options exist. Battalio et al. (2012) and Harris et al. (2013) compare the return of banned stocks with an index extracted from all non-banned stocks, the vast majority of which are non-financial. Beber and Pagano (2013) attempt to use both of these methods. Autore et al. (2011) calculate the abnormal return for banned stocks relative to the entire market or the Fama and French (1993) three-factor model.

Comparing financial stocks with a matched sample of non-financial stocks is likely sufficient for studying the ban's effect on liquidity and price discovery, but is not ideal for examining the price effect. Abnormal returns estimated using this approach may be attributed to differences between financial and non-financial stocks instead of the effects of the ban. In particular, the fundamentals of financial stocks are likely to be quite different from those of non-financial stocks during the 2008 financial crisis. Beber and Pagano (2013, p. 355) recognize this limitation and acknowledge that their approach "has the drawback of leaving only financials in the treated group and only non-financials in the control group."

We overcome this challenge by using the financial segments of the non-banned non-financial stocks as a control group for the banned financial stocks. We observe that many non-banned non-financial firms have financial segments. We identify these financial segments by using the Compustat Business Segment data and decompose the daily stock returns of these non-financial firms into a portion that is driven by non-financial segments and a portion driven by financial segments. We then use these implied returns driven by the financial segments as the benchmark returns to the banned financial stocks.

When we use all non-financial stocks as the control group, which is similar to what prior literature has done, we find that banned financial stocks experience a significant price inflation at the beginning of the short-selling ban, and this price inflation persists well after the ban is lifted. Based on this evidence, one might conclude that it is not the short-selling ban, but rather the contemporaneous news about TARP, that causes the price inflation in financial stocks. In contrast, when using the financial segments of the non-financial stocks as the control group, we find significant evidence of price inflation at the onset of the ban, and this initial overvaluation reverts at the end of the ban. Namely, both the ban and the price inflation are temporary. This convergence result is new in the literature, contrary to the existing findings, and consistent with the prediction of our model.

We make two contributions to the short selling literature. The first is to develop a model of temporary short-selling bans. Prior studies have studied either static models or dynamic models with permanent short-selling bans. These models offer great insights into the price effect of short-selling bans in general. However, existing theories do not explicitly model temporary short-selling bans; as such, they cannot predict how prices evolve from the start of the ban to the end of the ban. Despite the technical challenges involved in solving a finite-horizon dynamic model, we are able to obtain a closed-form solution for the overpricing caused by the short-selling ban. This explicit solution allows us to study the dynamics of the price effect associated with temporary short-selling bans for the first time.

Our second contribution is to offer an alternative empirical design to study the price effect of the 2008 short-selling ban. In contrast to previous studies which use all non-financial stocks or a matching sample of these stocks as the control group for the banned financial stocks, we use the financial segments of non-financial stocks as the control group. These financial segments share the same features as financial firms except that they are not subject to the ban. Although our approach is not perfect—one might argue that the financial segments of non-financial stocks are different from the banned financial stocks—our control is new, and arguably better than those in the existing literature for the purpose of examining the price effect of the 2008 ban.

Our paper is related to an extensive literature showing that short selling activities and short-sale constraints predict the cross-section of stock returns. Desai, Ramesh, Thiagarajan, and Balachandran (2002) and Asquith, Pathak, and Ritter (2005) present evidence that monthly short interest is negatively related to future stock returns. Boehmer, Jones, and Zhang (2008) and Diether, Lee, and Werner (2009) examine daily shorting flows and show that heavily shorted stocks underperform lightly shorted stocks. Engelberg, Reed, and Ringgenberg (2012), Jones, Reed, and Waller (2016), Kelley and Tetlock (2016), McLean, Pontiff, and Reilly (2020), and Wang, Yan, and Zheng (2020) also present evidence that short sellers are informed. Beneish, Lee, and Nichols (2015) and Drechsler and Drechsler (2016) document that loan fees and short-sale supply are strong predictors of the cross-section of stock returns. Engelberg, Reed, and Ringgenberg (2018) show that the dynamic risk of short selling is priced in the cross-section of stock returns. In particular, stocks with higher short selling risk earn lower returns.

The rest of the paper proceeds as follows. Section 2 presents our model of temporary short-selling bans. Section 3 describes our data and presents the empirical results. Section 4 concludes.

2. The Model

In this section, we develop a continuous-time asset pricing model with heterogenous beliefs and temporary short-selling bans. Our setting is similar to that of Scheinkman and Xiong (2003) except that they study an infinite-horizon problem whereas we analyze a finite-horizon problem. The equilibrium price becomes time dependent in the finite period setting. Section 2.1 presents the information structure and beliefs. Section 2.2 solves for the equilibrium price. Section 2.3 characterizes the dynamics of the price.

2.1. *Information Structure and Beliefs*

There is a single risky asset with limited supply. The stock pays dividends during $[0, T]$. The dividend process is driven by a fundamental variable f_t and noise dZ_t^D.

$$dD_t = f_t dt + \sigma_D dZ_t^D \tag{1}$$

where Z^D is a standard Brownian motion and σ_D is a constant volatility parameter.

The fundamental variable f_t is not observable, and it follows a mean-reverting process.

$$df_t = -\lambda(f_t - \bar{f})dt + \sigma_f dZ_t^f \tag{2}$$

where λ is the mean reversion parameter, $\bar{f}$ is the long-run mean of f.

The risky asset is in finite supply and we normalize the total supply to one. There are two groups of risk-neutral agents (Group A and Group B) trading the risky asset during $[0, T]$. Following Scheinkman and Xiong (2003), we assume that the number of agents in each group is infinitely large and that short selling of the risky asset is prohibited. To value future dividends, we assume that agents discount all future cash flows using the interest rate r and that each group has infinite total wealth.

All agents observe two noisy signals about f.

$$ds_t^A = f_t dt + \sigma_s dZ_t^A \tag{3}$$

and

$$ds_t^B = f_t dt + \sigma_s dZ_t^B \tag{4}$$

We assume that all four processes Z^D, Z^f, Z^A, and Z^B are independent with each other. All agents know the structure (the stochastic differential equations and the parameters) but cannot observe f_t. They try to infer f from the dividend D and the two signals s_A and s_B.

The agents are not perfectly rational when interpreting the signals. Specifically, agents in Group A (B) think of s_A (s_B) as their own signal and overestimate its informativeness. The behavioral finance literature characterizes such agents as being overconfident. In the meantime, agents in Group A (B) correctly estimate the informativeness of the other signal s_B (s_A).

We operationalize overconfidence in the same way as in Scheinkman and Xiong (2003). Specifically, agents in Group A (B) believe that dZ^A (dZ^B) is correlated with dZ^f with a correlation parameter ϕ ($0<\phi<1$). The agents in Group A therefore believe,

$$\begin{cases} ds_t^A = f_t dt + \sigma_s \phi dZ_t^f + \sigma_s \sqrt{1-\phi^2} dZ_t^A \\ ds_t^B = f_t dt + \sigma_s dZ_t^B \end{cases} \tag{5}$$

The agents in Group A optimally filter the signals D_t, s_t^A, and s_t^B to form their beliefs about the fundamentals f_t. Based on the optimal filtering theory, the conditional belief about f_t by investors in Group A is Gaussian with mean $\hat{f}_t^A$ and variance γ, where γ is a constant

$$\gamma \equiv \frac{-[\lambda + (\phi\sigma_f/\sigma_s)] + \sqrt{[\lambda + (\phi\sigma_f/\sigma_s)]^2 + (1-\phi^2)[(\sigma_f^2/\sigma_D^2) + (2\sigma_f^2/\sigma_s^2)]}}{(1/\sigma_D^2) + (2/\sigma_s^2)} \tag{6}$$

It is easy to show that γ is decreasing in ϕ. This justifies associating the parameter ϕ to overconfidence. Following Scheinkman and Xiong (2003), we refer to $\hat{f}_t^A$ as the belief of Group A investors. This conditional mean satisfies

$$d\hat{f}^A = -\lambda(\hat{f}^A - \bar{f})dt + \frac{\phi\sigma_s\sigma_f + \gamma}{\sigma_s^2}(ds^A - \hat{f}^A dt) + \frac{\gamma}{\sigma_s^2}(ds^B - \hat{f}^A dt) + \frac{\gamma}{\sigma_D^2}(dD - \hat{f}^A dt) \tag{7}$$

The conditional belief is mean-reverting, as shown in the first term. The other three terms correspond to surprises due to the dividend process and the two signals. Comparing the two terms associated with signal s^A and signal s^B, we can see that Group A investors respond more to surprises in signal s^A than in signal s^B due to overconfidence.

Investors in Group B are overconfident about signal s_t^B and form their belief $\hat{f}_t^B$ in a similar way. Investors in Group B believe that

$$\begin{cases} ds_t^A = f_t dt + \sigma_s dZ_t^A \\ ds_t^B = f_t dt + \sigma_s \phi dZ_t^f + \sigma_s \sqrt{1-\phi^2} dZ_t^B \end{cases} \tag{8}$$

Based on this, they form their belief $\hat{f}_t^B$ on the factor f_t.

$$d\hat{f}^B = -\lambda(\hat{f}^B - \bar{f})dt + \frac{\gamma}{\sigma_s^2}(ds^A - \hat{f}^B dt) + \frac{\phi\sigma_s\sigma_f + \gamma}{\sigma_s^2}(ds^B - \hat{f}^B dt) + \frac{\gamma}{\sigma_D^2}(dD - \hat{f}^B dt) \tag{9}$$

The difference of beliefs $g^A = \hat{f}^B - \hat{f}^A$ satisfies a simple stochastic differential equation (SDE) based on Proposition 1 in Schenkman and Xiong (2003).

$$dg^A = -\rho g^A dt + \sigma dW_g^A \tag{10}$$

where ρ and σ are constant

$$\rho = \sqrt{(\lambda + \phi\frac{\sigma_f}{\sigma_s})^2 + (1-\phi^2)\sigma_f^2(\frac{2}{\sigma_s^2} + \frac{1}{\sigma_D^2})} \tag{11}$$

$$\sigma = \sqrt{2}\phi\sigma_f. \tag{12}$$

We refer to σ as the volatility of the difference in beliefs. It is increasing in ϕ (the degree of overconfidence) and σ_f (the volatility of the fundamentals).

In a similar fashion, we can show that, for agents in Group B, $g^B = \hat{f}^A - \hat{f}^B$ satisfies

$$dg^B = -\rho g^B dt + \sigma dW_g^B \tag{13}$$

where ρ and σ are identical to those shown above.

2.2. Equilibrium Price

The agents who are more optimistic about the future payoffs will be willing to pay a higher price for the risky asset and consequently hold the entire supply. For tractability, we assume there is no trading cost.[2] In this case, trading occurs when the beliefs of the two groups of agents differ. The short-sale ban, a finite supply of shares, and the infinite number of agents in each group guarantee that the equilibrium price will be equal to the reservation price of the more optimistic group of agents.

At time $t \in [0, T]$, Group A investors' expectation of the discounted future dividends is[3]

$$\mathbf{E}_t^A \left[\int_t^T e^{-r(s-t)} dD_s \right] = \frac{\bar{f}}{r}(1 - e^{-r(T-t)}) + \frac{\hat{f}_t^A - \bar{f}}{r + \lambda}(1 - e^{-(r+\lambda)(T-t)}) \tag{14}$$

If there are only Group A investors in the market, then the price of the stock will be equal to this expectation. However, there are also Group B investors in the market with a different expectation of future dividends

$$\mathbf{E}_t^B \left[\int_t^T e^{-r(s-t)} dD_s \right] = \frac{\bar{f}}{r}(1 - e^{-r(T-t)}) + \frac{\hat{f}_t^B - \bar{f}}{r + \lambda}(1 - e^{-(r+\lambda)(T-t)}) \tag{15}$$

Because short-selling is banned, pessimistic investors with a lower expectation will sit out of the market. Optimistic investors with a higher expectation will be the owners of the stock and set the price. It is possible that the pessimistic investors will become optimistic in the future and want to buy the stock from the current owner. Thus when setting the price, the current owner will also consider the value of the resale option. So the price is the current owner's expectation of future dividends plus the value of a resale option.

[2] Scheinkman and Xiong (2003) assume positive trading costs in their model because of their desire to study trading volume and volatility. We focus more on the price effect of the short-selling ban. Scheinkman and Xiong (2003) show that the speculative bubble still exists when there is no trading cost.

[3] The method to calculate the expected future dividends is as follows: since the expectations of all the stochastic parts are zero, we ignore the stochastic terms when calculating the expectation. Thus by equation (7), one has $d\hat{f}^A = -\lambda(\hat{f}^A - \bar{f})dt$; the solution to this ODE is $\hat{f}_s^A = \bar{f} + Ce^{-\lambda(s-t)}$. The constant C can be determined by letting $s = t$. Thus $C = \hat{f}_t^A - \bar{f}$. Thus $\hat{f}_s^A = \bar{f} + (\hat{f}_t^A - \bar{f})e^{-\lambda(s-t)}$. Ignoring the stochastic part, based on equation (1), one has $dD_s = f_s ds$. Thus $\mathbf{E}_t^A[\int_t^T e^{-r(s-t)} dD_s] = \int_t^T e^{-r(s-t)}[\bar{f} + (\hat{f}_t^A - \bar{f})e^{-\lambda(s-t)}]ds = \frac{\bar{f}}{r}(1 - e^{-r(T-t)}) + \frac{\hat{f}_t^A - \bar{f}}{r+\lambda}(1 - e^{-(r+\lambda)(T-t)})$

$$p_t^o = p(\hat{f}_t^o, g_t^o, t) = \frac{\bar{f}}{r}(1 - e^{-r(T-t)}) + \frac{\hat{f}_t^o - \bar{f}}{r + \lambda}(1 - e^{-(r+\lambda)(T-t)}) + q(g_t^o, t) \tag{16}$$

where the superscript "o" denotes the group of agents who currently own the asset and "$\bar{o}$" denotes the other group of agents who do not own the stock at this moment t. $o = A$ and $\bar{o} = B$ if $\hat{f}_t^A > \hat{f}_t^B$, and vice versa. The value of the resale option $q(g_t^o, t)$ depends on the difference of beliefs between the two groups, i.e., $g^o = \hat{f}^{\bar{o}} - \hat{f}^o$.

The owner determines an optimal time τ to sell the stock to maximize the dividends collected and the capital gain.

$$p_t^o = \sup_{t \leqslant \tau \leqslant T} \mathbf{E}_t^o \left[\int_t^\tau e^{-r(s-t)} dD_s + e^{-r(\tau-t)} p_\tau^{\bar{o}} \right] \tag{17}$$

At the end of the trading period, the price is $p_T^o = 0$ because there are no more dividends to be collected and there are no more chances to profit from re-selling the stock. Further, the price is required to be smooth (continuously differentiable). This technical requirement guarantees the uniqueness of the solution.

Expressing p_t^o and $p_\tau^{\bar{o}}$ in the optimal decision equation (17) by the candidate equilibrium price equation (16), one can derive the equation for the resale option. Together with the boundary condition and technical requirement, one has

$$\begin{cases} q(g_t^o, t) = \sup\limits_{t \leqslant \tau \leqslant T} \mathbf{E}_t^o \{ e^{-r(\tau-t)} [\frac{g_t^o}{r+\lambda}(1 - e^{-(r+\lambda)(T-\tau)}) + q(g_\tau^{\bar{o}}, \tau)] \} \\ q(g_t^o, T) = 0 \\ q \ \ is \ \ smooth \end{cases} \tag{18}$$

This is an optimal execution problem. Let $x = g_t^o = \hat{f}_t^{\bar{o}} - \hat{f}_t^o$ be the current belief difference between the two groups of investors.

Lemma 1. The owner will wait if $x < 0$ and sell if $x \geqslant 0$.

When $x < 0$, the owner is more optimistic and chooses not to sell. If there is a positive trading cost, investors will choose to wait if x<k, where k is a positive number that depends on the trading cost and remaining time until T. However, if there is no trading cost, as is assumed here, trading will occur whenever the beliefs cross. Thus the owner will wait if $x < 0$ and sell if $x \geqslant 0$. We provide the proof of Lemma 1 in Appendix A.1.

The discounted value of this resale option $e^{-rt}q(g_t^o, t)$ is a martingale to the current holder. Based on Ito's lemma and the evolution equation for g_t^o, one has

$$\frac{1}{2}\sigma_g^2 q_{xx} - \rho x q_x + q_t - rq = 0, \qquad x < 0 \tag{19}$$

If the current belief difference is non-negative, $x \geqslant 0$, the owner is more pessimistic and chooses to sell right now. Then the optimal execution time is equal to current time $\tau = t$ and the value of the resale option is

$$q(x,t) = \frac{x}{r+\lambda}(1 - e^{-(r+\lambda)(T-t)}) + q(-x,t), \qquad x \geqslant 0 \tag{20}$$

The requirement of $q(x,t)$ being smooth means $q(x,t)$ and $q_x(x,t)$ are continuous when x crosses 0.

Since

$$\lim_{x \to 0+} q(x,t) = \lim_{x \to 0+} [\frac{x}{r+\lambda}(1 - e^{-(r+\lambda)(T-t)}) + q(-x,t)] = q(0,t) \tag{21}$$

and

$$\lim_{x \to 0-} q(x,t) = q(0,t) \tag{22}$$

the continuation of $q(x,t)$ is satisfied automatically.

Since

$$\lim_{x \to 0+} q_x(x,t) = \lim_{x \to 0+} [\frac{1}{r+\lambda}(1 - e^{-(r+\lambda)(T-t)}) - q_x(-x,t)] = \frac{1}{r+\lambda}(1 - e^{-(r+\lambda)(T-t)}) - q_x(0,t) \tag{23}$$

and

$$\lim_{x \to 0-} q_x(x,t) = q_x(0,t) \tag{24}$$

the continuation of $q_x(x,t)$ means

$$q_x(0,t) = \frac{1}{2(r+\lambda)}(1 - e^{-(r+\lambda)(T-t)}) \tag{25}$$

Thus the optimal execution problem is transformed to a PDE for $x < 0$ and $0 \leqslant t \leqslant T$.

$$\begin{cases} \frac{1}{2}\sigma_g^2 q_{xx} - \rho x q_x + q_t - rq = 0 \\ q(x,T) = 0 \\ q_x(0,t) = \frac{1}{2(r+\lambda)}(1 - e^{-(r+\lambda)(T-t)}) \end{cases} \tag{26}$$

If the second order derivative term were $\frac{1}{2}\sigma_g^2 x^2 q_{xx}$ instead of $\frac{1}{2}\sigma_g^2 q_{xx}$, it would be the Black–Scholes equation and can be solved by transforming it into a diffusion equation. Without the x^2 term, the problem is much more challenging to solve. Also, if the boundary condition were the Dirac delta function, physicists can solve it by performing a Fourier transform. The more complicated boundary condition further increases the difficulty to solve this PDE.

By performing a Laplace transform on the equation and using the integral form of the Kummer function, we are able to obtain a solution to the above PDE.

Theorem 1. The solution to equation (26) is

$$u(x,t) = \frac{\sigma}{4(r+\lambda)\sqrt{\pi\rho}} \int_{N(t)}^{\infty} e^{-\frac{\rho}{\sigma^2}x^2 y} y^{\frac{r-2\rho}{2\rho}} (1+y)^{-\frac{r+\rho}{2\rho}} [1 - e^{-(r+\lambda)(T-t-M(y))}] dy \tag{27}$$

where $N(t) = \frac{e^{-2\rho(T-t)}}{1-e^{-2\rho(T-t)}}$, $M(y) = \frac{1}{2\rho} ln(1 + \frac{1}{y})$, $x < 0$ and $0 \leqslant t \leqslant T$. The speculative bubble, i.e., the value of the resale option is then

$$q(x,t) = \begin{cases} \frac{x}{r+\lambda}(1 - e^{-(r+\lambda)(T-t)}) + u(-x,t) & x \geqslant 0 \\ u(x,t) & x < 0 \end{cases} \tag{28}$$

We provide a detailed proof of Theorem 1 in the Appendix A.2.

2.3. Dynamics of the Price during the Short-selling Ban

The main difference between our model and Scheinkman and Xiong (2003) is that they study an infinite-horizon problem, while we analyze a finite-horizon problem. We do so because we want to study how prices evolve during the life of the short-selling ban. Based on Theorem 1, we show that the speculative bubble shrinks during the ban and vanishes at the end of the ban.

Theorem 2. The speculative bubble $q(x,t)$ is decreasing in t and equals 0 when $t = T$.

This is the main testable prediction of our model. This is also where our model differs from Scheinkman and Xiong (2003). In Scheinkman and Xiong (2003), the speculative bubble is not dependent on t. The detailed proof of Theorem 2 is contained in the Appendix A.3.

We note that in Theorem 1 the magnitude of the speculative bubble $q(x,t)$ is determined by time t and the current difference of beliefs from the owners' perspective, x. Averaging the stochastic term x in $q(x,t)$ can lead to the mean magnitude of the bubble $\bar{q}(t)$, which only depends on time.

The difference of beliefs from Group A investors' perspective $g^A = \hat{f}^B - \hat{f}^A$ satisfies a simple SDE based on equation (13)

$$dg^A = -\rho g^A dt + \sigma dW_g \tag{29}$$

The stationary distribution of this process is normal, $N(0, \eta^2)$, where $\eta^2 = \frac{\sigma^2}{2\rho}$. The same distribution applies to g^B since $g^B = \hat{f}^A - \hat{f}^B = -g^A$. When Group A investors hold the stock, $\hat{f}^B < \hat{f}^A$, thus $x = g^A < 0$. When Group B investors hold the stock , $\hat{f}^A < \hat{f}^B$, thus $x = g^B = -g^A < 0$. So x is always less than or equal to zero with a stationary distribution that doubles the left half of the normal distribution $N(0, \eta^2)$:

$$w(x)dx = 2 \cdot \frac{1}{\eta\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2\eta^2}} dx \qquad x < 0 \tag{30}$$

Theorem 3. Averaging the speculative bubble $q(x, t)$ in equation (28) with the stationary distribution $w(x)dx$ in equation (30) leads to the mean magnitude of the bubble

$$\bar{q}(t) = \frac{\sigma\sqrt{\rho}}{2\lambda\sqrt{\pi}} \left[\frac{1}{r}(1 - e^{-r(T-t)}) - \frac{1}{r+\lambda}(1 - e^{-(r+\lambda)(T-t)}) \right]. \tag{31}$$

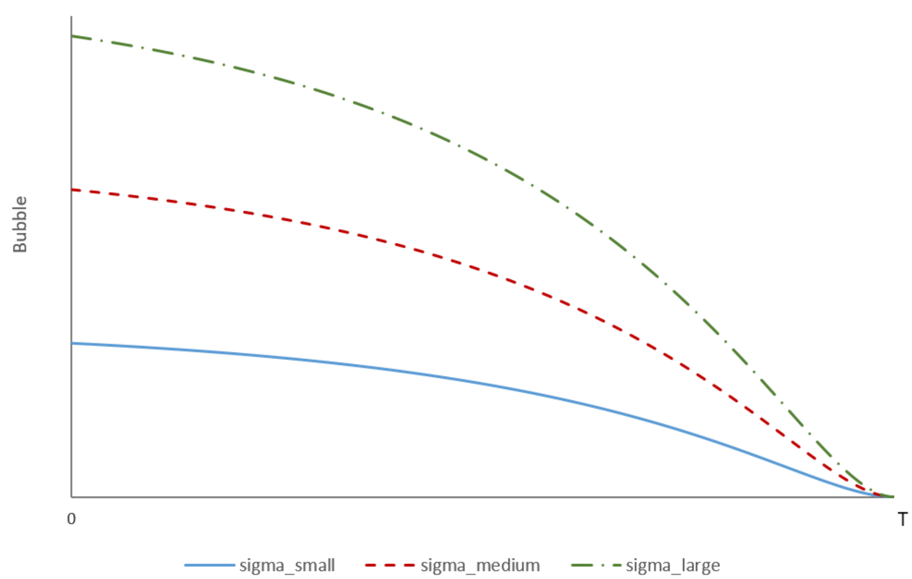

Fig. 1. **Dynamics of the Bubble and Volatility of the Disagreement.**

Based on Theorem 3, we can see that the size of the average bubble increases linearly in σ, the volatility of the disagreement process. It also increases linearly in $\sqrt{\rho}$, the square root of the mean reversion coefficient of the disagreement process, and decreases in the mean reversion coefficient of the fundamentals. It is also clear that the average bubble size is increasing in T, and decreasing in t and r.

In Figure 1, we plot the evolution of the bubble over time for different values of the volatility of the disagreement. The x-axis shows the time starting at 0 and ending at T. The y-axis shows the average overvaluation caused by the ban. When the disagreement process is more volatile, it's more likely to observe extreme disagreement and trading occurs more frequently, which leads to a higher speculative bubble.

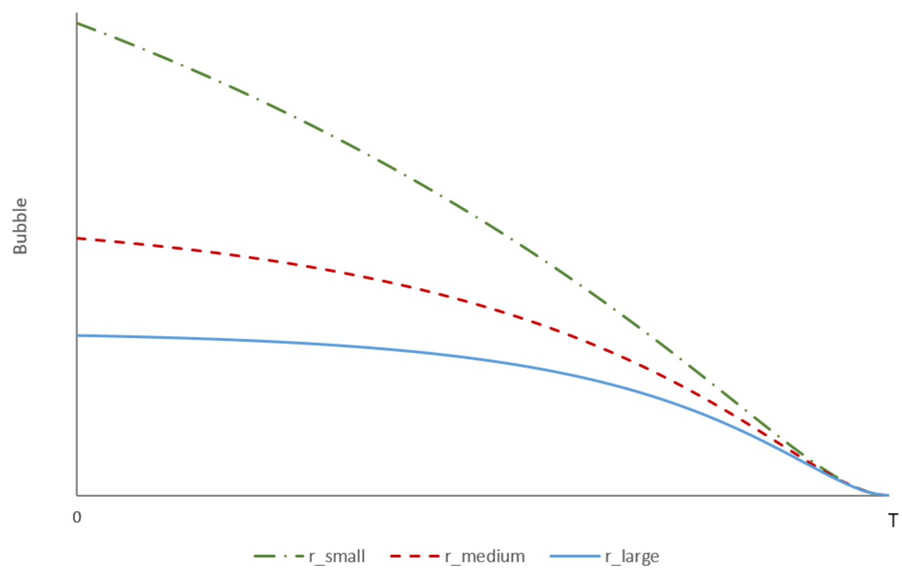

Fig. 2. **Dynamics of the Bubble and Interest Rate.**

In Figure 2, we plot the evolution of the bubble over time for different interest rates. Generally, increasing interest rate r will make the size of the bubble smaller because it decreases the present value of the future resale price, which makes investors less willing to speculate.

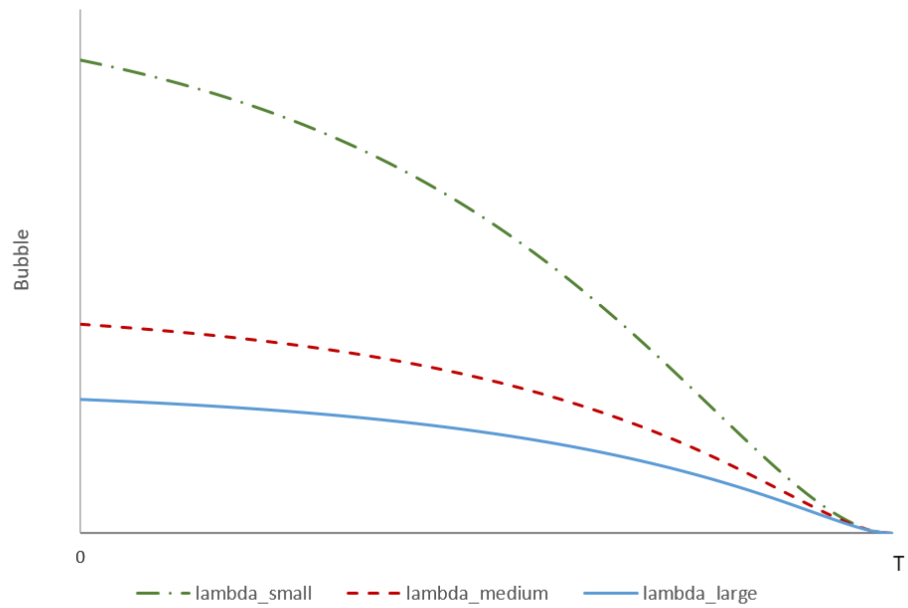

Fig. 3. **Dynamics of the Bubble and Mean-reversion Coefficient of Fundamental.**

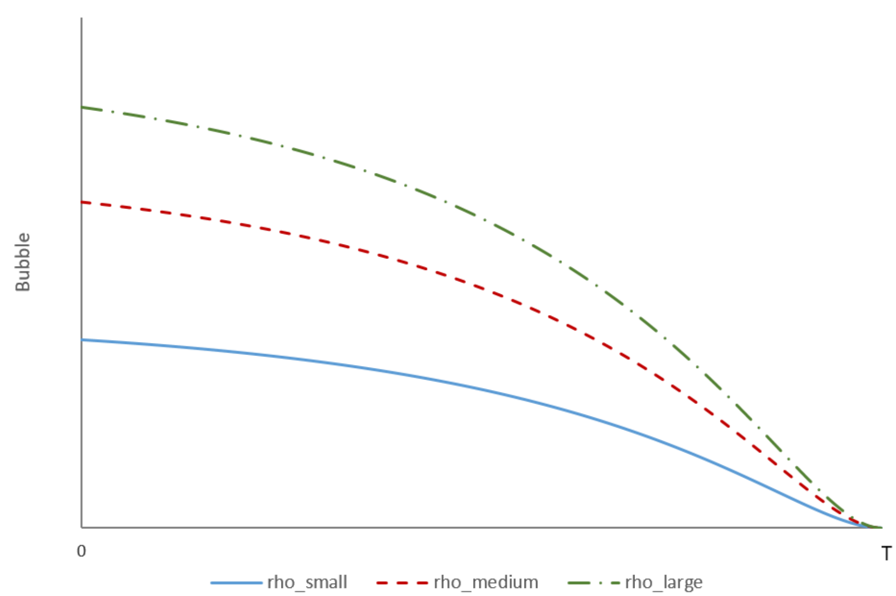

Fig. 4. **Dynamics of the Bubble and Mean-reversion Coefficient of the Disagreement.**

In Figure 3, we plot the evolution of the bubble over time for different speed of mean-reversion in the fundamentals. Increasing λ , the mean-reversion coefficient of the true fundamental of the stock, will decrease the size of the bubble because it makes the fundamental converge to the long-term mean faster, which decreases the excess dividends that optimistic investors expect to collect.

In Figure 4, we plot the evolution of the bubble over time for different speed of mean-reversion in the disagreement between the two groups of investors. Increasing ρ will increase the size of the bubble because it makes the investor beliefs to cross more likely, and thus increases the value of the resale option.

Theorem 4. The average bubble $\bar{q}(t)$ is decreasing in t and vanishes at the end of the ban. Let $\tau^* = \frac{1}{\lambda} ln(1 + \frac{\lambda}{r})$; then the inflection point of $\bar{q}(t)$ is at $t^* = T - \tau^*$. When $t < t^*$, $\bar{q}(t)$ is concave; when $t^* < t < T$, $\bar{q}(t)$ is convex. At the inflection point, the bubble has the highest decreasing speed $\frac{d\bar{p}(t)}{dt}|_{t=t^*} = \frac{-\sigma\sqrt{\rho}}{2(r+\lambda)\sqrt{\pi}}(\frac{r}{r+\lambda})^{\frac{r}{\lambda}}$.

Based on Theorem 4, the remaining time when the bubble crosses an inflection point is $\tau^* = \frac{1}{\lambda} ln(1 + \frac{\lambda}{r})$. When the remaining time $T - t$ is bigger than τ^*, the curve of the bubble is concave. When it's very close to the end of the ban (i.e., $T - t$ is smaller than τ^*), the curve is convex. As $r \to +\infty$, $\tau^* \to 0$. This means the curve is always concave, the highest speed of decreasing occurs at the end of the ban. As $r \to 0$, $\tau^* \to +\infty$. The whole curve will be convex. As $\lambda \to +\infty$, $\tau^* \to 0$ and as $\lambda \to 0$, $\tau^* \to \frac{1}{r}$. r and λ together determine the position of the inflection point and thus the shape of the bubble curve.

Fixing the other parameters, increasing T can be realized by moving the $t = 0$ point to the left (extend the curve to the left such that it takes time T to vanish). Bigger T leads to a higher initial bubble; however, the bubble won't go to infinity. As $T \to +\infty$, $\bar{p}(t) \equiv \bar{p}_{T=\infty} = \frac{\sigma\sqrt{\rho}}{2(r+\lambda)\sqrt{\pi}}$. An infinite ban leads to a constant average bubble.

3. Empirical Analysis

The central prediction of our model is that a temporary short-selling ban leads to a price inflation that is the highest at the beginning of the ban and shrinks to zero at the end of the ban. We test this prediction by examining the 2008 short-selling ban of financial stocks in the U.S. Section 3.1 discusses the timeline of the 2008 ban. Section 3.2 describes our data, sample, and methodology. Section 3.3 presents the empirical results.

3.1. Timeline of the 2008 Ban

Beginning in the summer of 2008, the SEC implemented a series of short-selling restrictions. In July 2008, the SEC issued an emergency order restricting naked short selling in 19 large financial stocks. After the emergency order expired in mid-August, the SEC announced on September 17 a permanent ban on naked shorting in all U.S. stocks effective the next day. On September 18, the SEC issued another emergency order, temporarily banning short sales in 797 financial stocks. The initial order covered ten business days, and was set to expire on October 2, 2008. On September 21, the SEC announced that all decisions about banned stocks are delegated to the individual exchanges. From September 22 and October 7, an additional 148 stocks were added to the ban list by the exchanges. On October 2, 2008, at the end of the initial 10-day ban period, the SEC extended the ban to October 17, 2008 or three days after the passage of the TARP bill, whichever is earlier. The TARP bill was passed on October 3 and the short-sale ban ended at mid-night on October 8. As of October 9, short selling was again permitted (except that naked shorting is still banned).

The 2008 short-selling ban is contemporaneous with news about TARP, which was conceived in response to the market turmoil following the collapse of Lehman Brothers. On September 18, Treasury secretary Henry Paulson announced the intention to work on a market stabilization plan, i.e., to create TARP. On September 29, the plan was rejected by the Congress, sending the Dow Jones Industrial Average down 778 points. Two days later on

October 1, 2008, the Senate passed the Emergency Economic Stabilization Act (EESA), the bill that authorized TARP by a vote of 74 to 25. On October 3, 2008, the House of Representatives passed the EESA by a vote of 263 to 171 and president Bush signed the EESA into law[4]. On October 13 and 14, the U.S. Treasury announced a standardized program to purchase equity in a broad array of financial institutions and the Capital Purchase Program (CPP), respectively.

3.2. Data, Sample, and Methods

3.2.1. Data and Sample

As stated earlier, the SEC initially banned short sales in 797 financial stocks. The individual exchanges then added an additional 148 stocks to the ban list from September 22 and October 7 of 2008. Our empirical analysis includes only stocks on the initial ban list.[5] This choice is motivated by two factors. First, the initial list is arguably a cleaner sample because the short-selling ban announced on September 18, 2008 was truly a surprise. In comparison, the additions to the ban list during subsequent days may be partially anticipated. Second, the main testable prediction of our model is that a short-selling ban leads to a price inflation that shrinks during the life of the ban period. Examining banned stocks on the initial list allows us to study a more uniform window and hence more cleanly test the prediction of our model.

We obtain stock price and return data from the Center for Research in Security Prices (CRSP). We restrict our sample to common stocks (with a sharecode of 10 or 11) traded on the three major U.S. exchanges, i.e., NYSE, NASDAQ, and AMEX. Our sample period is from August 19 (one month before the ban starts) to November 7, 2008 (one month after the ban ends). We require that all stocks have full price and return data during these 58 trading

[4] http://www.treasury.gov/initiatives/financial-stability/reports/Pages/TARP-Tracker.aspx

[5] The 148 stocks added to the ban list after September 18 are completely removed from our analysis. They serve neither as banned stocks nor as control stocks.

days. We obtain the business segments and the corresponding SIC codes from Compustat. We merge the CRSP and Compustat data by using the CRSP-Compustat link file. We also require that all stocks have business segment data. Our final sample includes a total of 3,330 stocks.

3.2.2. Methodology

We use the financial segments of non-financial firms as the benchmark to evaluate the price effect of the ban. Each stock in the 2008 Compustat business segment database has an SIC code to identify its principal business. In addition, each stock has one or more business segments. For each segment, there is a segment-level SIC code and a record of segment sales and assets. We identify financial segment based on this segment-level SIC code (those with an SIC code between 6000 and 6999). We remove a number of segments that are created to record eliminations due to accounting issues. These segments usually have names such as "Elimination", "Other", "Adjustment". The assets or sales are sometimes negative for these segments.

Table 1 illustrates the business segments of two example companies, namely Cohen and Steers Inc. and CASS Information Systems. Cohen and Steers's principal SIC code is 6282, which is investment banking. The company has only one business segment, which not surprisingly also has an SIC code of 6282. The segment, as well as the company as a whole, generated a total sale of $185.83 million in 2008. Cohen and Steers is a financial firm and was on the initial short-selling ban list.

In comparison, CASS Information Systems has two business segments. The first segment is in banking services with an SIC code of 6022. This segment generated a sale of $15.16 million in 2008. The second segment is in information systems and has an SIC code of 7389. This segment generated a sale of $76.57 million. Because the principal business of CASS Information Systems is in information systems, its principal SIC code is 7389. As a result, the company was not banned from short selling in 2008.

Table 1: Business Segments and Financial Factor

Company Name	SIC	Banned	Segment	Segment SIC	Sales ($ million)
Cohen & Steers INC	6282	**Yes**	Investment Banking	6282	185.83
CASS Information Systems	7389	**No**	Banking Services	6022	15.16
			Information Services	7389	76.57

For each company in our sample, we calculate its financial factor f as the financial segment's contribution to total sales

$$f = \frac{\sum\limits_{fin} Sales}{\sum\limits_{all} Sales} \tag{32}$$

where fin is the index of all financial segments.[6]

For CASS Information Systems in Table 1, we have

$$f = \frac{15.16}{15.16 + 76.57} = 16.53\% \tag{33}$$

A stock with financial factor $f = 1$ is a purely financial firm, like Cohen and Steers. If a company has no financial segment, then $f = 0$. A company with $0 < f < 1$ has both financial and non-financial segments.

During the entire 58 trading days, which includes 22 trading days before the ban, 14 days during the ban, and 22 days after the ban, the banned stocks have a ban indicator $b_{it} = 1$ and the unbanned stocks have $b_{it} = 0$. The stocks with $b_{it} = 1$ before or after the ban serve as the placebo group: the banned and non-banned financial stocks should behave more similarly before the ban than during the ban.

On each trading day t, we assume that the return of each stock can be potentially explained by four factor returns: the factor return of Banned Financial stocks R_t^{BF}, the

[6] In robustness tests, we calculate alternative f based on free cash flows, operating profits, and total assets instead of sales.

factor return of Non-banned Financial stocks R_t^{NF}, the factor return of Banned Non-financial stocks R_t^{BN}, and the factor return of Non-banned Non-financial stocks R_t^{NN}. Specifically, we decompose the return of each stock on each day into these factor returns based on the ban indicator and the financial factor.

$$r_{it} = f_i(b_{it}R_t^{BF} + (1 - b_{it})R_t^{NF}) + (1 - f_i)(b_{it}R_t^{BN} + (1 - b_{it})R_t^{NN}) + \epsilon_{it} \quad (34)$$

In the example shown in Table 1, Cohen and Steers is a pure financial company. Thus its financial factor is $f_i = 1$. Since the stock is banned from short-selling, its ban indicator is $b_{it} = 1$, for all t. Substitute these into Equation (34), we have

$$r_{it} = R_t^{BF} + \epsilon_{it} \quad (35)$$

As a purely financial company that is banned from short-selling, Cohen and Steers' daily return is assumed to be driven by the factor return of banned financial stocks.

Because 16.53% business of CASS Information Systems is financial, its financial factor is $f_j = 0.1653$. The stock is not banned from short-selling, so its ban indicator is $b_{jt} = 0$, for all t. Substitute these into Equation (34), we have

$$r_{jt} = 0.1653 \cdot R_t^{NF} + 0.8347 \cdot R_t^{NN} + \epsilon_{jt} \quad (36)$$

As a non-banned stock with 16.53% of its business being financial and 83.47% of its business being non-financial, its daily return is assumed to be driven by the factor return of non-banned financial stocks and the factor return of the non-banned non-financial stocks.

We estimate the four factors, i.e., R_t^{BF}, R_t^{NF}, R_t^{BN}, and R_t^{NN}, from Equation (34) by running a cross-sectional regression of stock returns on the coefficients of these four factors each day. For example, the four coefficients are 1, 0, 0, and 0 for Cohen and Steers, and 0, 0.1653, 0, and, 0.8347 for CASS Information Systems. To measure the price effect of the ban, we take the difference between R_t^{BF} and R_t^{NF}.

3.3. Empirical Results

Before presenting the results for our methodology, i.e., using the financial segments of non-financial stocks as a control group, we first show what happens when we use all nonbanned stocks as a control group. This approach is similar to those employed by prior studies. Doing so establishes a baseline to which we can compare the result from our approach.

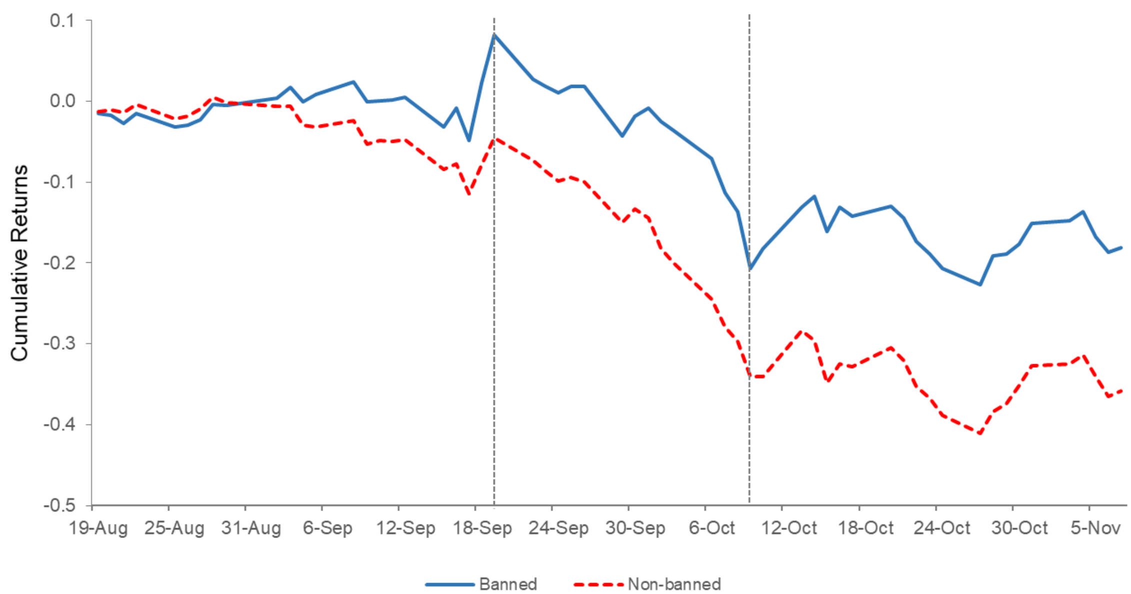

Fig. 5. **Cumulative Returns of Banned vs. Non-banned Stocks.** This figure plots the cumulative returns of banned stocks and the cumulative returns of the non-banned stocks. The vertical lines indicate the first day and the last day of the ban.

We begin by constructing equal-weighted portfolios of all banned stocks and non-banned stocks, respectively. We then compute the cumulative returns of each portfolio for the period from August 19 to November 7. We plot the time-series of these cumulative returns in Figure 5. We can see that the performance of banned stocks began diverging from non-banned stocks in early September and this divergence widened significantly when the short-selling ban was announced, indicating a significant price inflation of banned stocks when compared to the non-banned stocks. This overvaluation, however, persisted not only during the short-selling ban but also well after the ban was lifted. Based on this result, one would likely reach the same conclusion as Beber and Pagano (2013) and Boehmer, Jones, and Zhang (2013), i.e., the observed price inflation of the banned stocks is unlikely to be driven by the short-selling ban and more likely caused by the contemporaneous news about TARP.

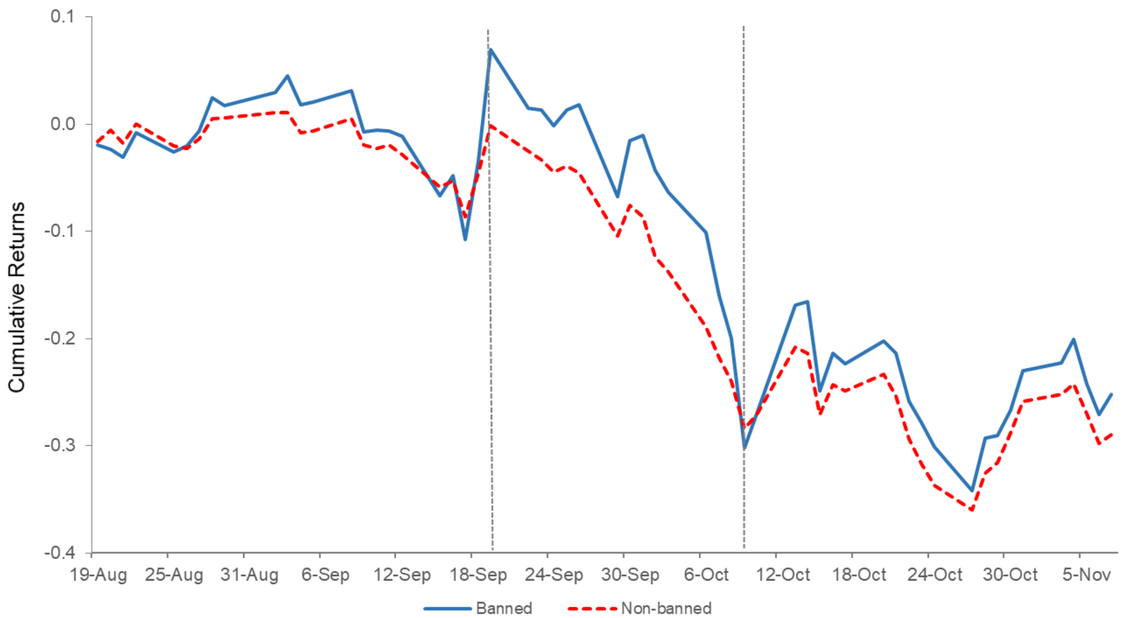

Fig. 6. **The Price Effect of the 2008 Short-selling Ban on Financial Stocks.** This figure plots the cumulative returns of R^{BF} and the cumulative returns of R^{NF} estimated from Equation (34). The vertical lines indicate the first and the last day of the ban.

We now present the results from our approach. As stated earlier, we estimate the four factors, i.e., R_t^{BF}, R_t^{NF}, R_t^{BN}, and R_t^{NN}, from Equation (34). We then use the difference between R_t^{BF} and R_t^{NF} to measure the price effect of the ban. Figure 6 plots the cumulative returns of R^{BF} and the cumulative returns of R^{NF}. The two returns are reasonably close to each other before the ban. On the announcement day and the subsequent day, the two returns significantly diverged. Specifically, R_t^{BF} significantly outperformed R_t^{NF} on those two days. This divergence, however, shrank during the ban period and then disappeared at the end of the ban. This finding is in stark contrast to what we find in Figure 5.

Table 2: **The Price Effect of the 2008 Short Selling Ban on Financial Stocks**

This table presents the daily returns of R^{BF} and R^{NF} estimated from Equation (34). R^{BF} is the factor return for banned financial stocks. R^{NF} is the factor return for non-banned financial stocks.

Date	R^{BF}	R^{NF}	R^{BF}-R^{NF}	Stderr	t-value
20080915	-0.0559	-0.0313	-0.0246	0.0068	-3.62
20080916	0.0197	0.0061	0.0136	0.0082	1.66
20080917	-0.0625	-0.0350	-0.0275	0.0073	-3.79
20080918	0.0826	0.0442	0.0384	0.0107	3.58
20080919	0.1076	0.0462	0.0614	0.0103	5.95
20080922	-0.0512	-0.0235	-0.0276	0.0078	-3.53
20080923	-0.0015	-0.0087	0.0072	0.0068	1.06
20080924	-0.0145	-0.0113	-0.0032	0.0061	-0.52
20080925	0.0145	0.0056	0.0089	0.0063	1.41
20080926	0.0047	-0.0070	0.0117	0.0066	1.78
20080929	-0.0841	-0.0614	-0.0227	0.0106	-2.13
20080930	0.0564	0.0319	0.0245	0.0108	2.28
20081001	0.0044	-0.0113	0.0157	0.0076	2.05
20081002	-0.0327	-0.0407	0.0080	0.0077	1.04
20081003	-0.0217	-0.0160	-0.0058	0.0073	-0.79
20081006	-0.0396	-0.0595	0.0200	0.0093	2.15
20081007	-0.0659	-0.0360	-0.0298	0.0094	-3.17
20081008	-0.0477	-0.0282	-0.0195	0.0095	-2.05
20081009	-0.1267	-0.0570	-0.0697	0.0112	-6.25
20081010	0.0445	0.0180	0.0265	0.0152	1.75
20081013	0.1389	0.0858	0.0531	0.0165	3.21
20081014	0.0042	-0.0075	0.0117	0.0115	1.02
20081015	-0.0998	-0.0723	-0.0275	0.0123	-2.24

Table 2 presents the detailed day-by-day results of R_t^{BF} and R_t^{NF}. We find that R_t^{BF} significantly outperformed R_t^{NF} by 3.84% on September 18, 2008, the announcement day of the short-selling ban, and by 6.14% on September 19, 2008, the first effective day of the ban. Both of these return differences are statistically significant. R_t^{BF} then significantly underperformed R_t^{NF} by 1.95% on October 8, 2008, the last day of the short-selling ban, and by 6.97% on October 9, 2008, the first day without the ban. Overall, the initial price inflation is completely reversed by the end of the ban, supporting the prediction of our model.

In our main analysis, we calculate the financial factor as the fraction of sales attributed to financial segments. We consider three alternative definitions of the financial factor. Specifically, instead of sales, we use free cash flows, operating profits, and total assets to construct the financial factor and repeat our analysis. We present the results in Figures 7, 8, and 9. Overall, we obtain very similar results to those presented in Figure 6. We continue to find that R_t^{BF} significantly outperformed R_t^{NF} at the beginning of the ban, and this divergence disappeared toward the end of the ban.

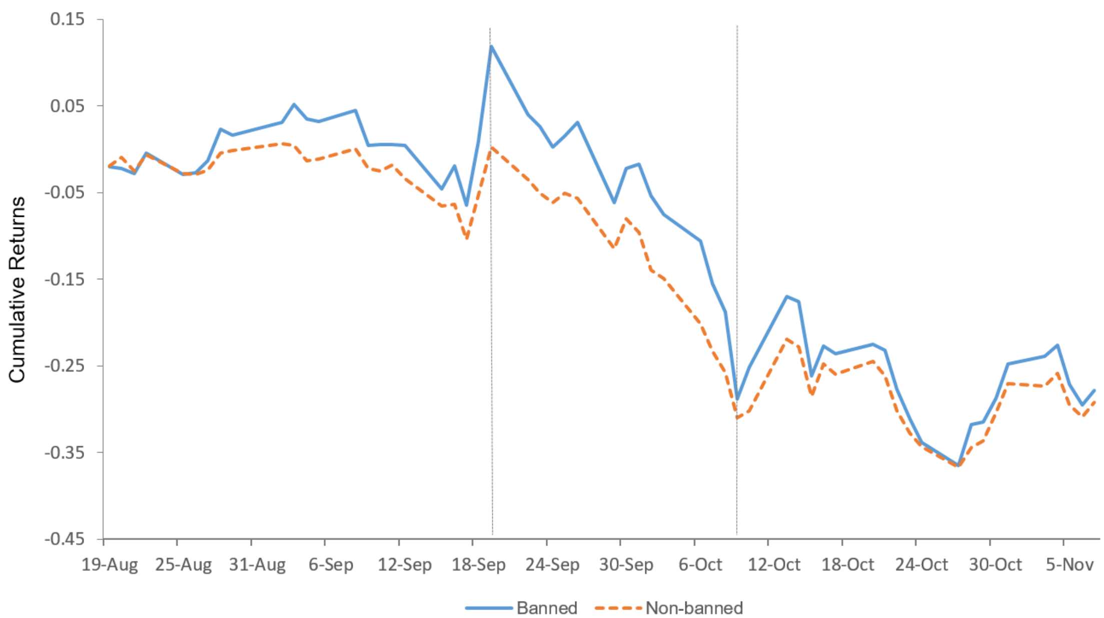

Fig. 7. **The Price Effect of the 2008 Short Selling Ban on Financial Stocks: Alternative Definition of f Based on Free Cashflows.** This figure plots the cumulative returns of R^{BF} and the cumulative returns of R^{NF} estimated from Equation (34). The vertical lines indicate the first and the last day of the ban.

Any test of the price effect of the 2008 short-selling ban is potentially confounded by the contemporaneous news about TARP. In our setting, we acknowledge that the TARP program potentially benefits the banned financial stocks more than the financial segments of the non-financial stocks. However, TARP is not characterized by a single announcement, but rather represented by a series of news, some of which are negative. As stated earlier, Treasury secretary Henry Paulson announced the intention to work on a market stabilization plan, i.e.,

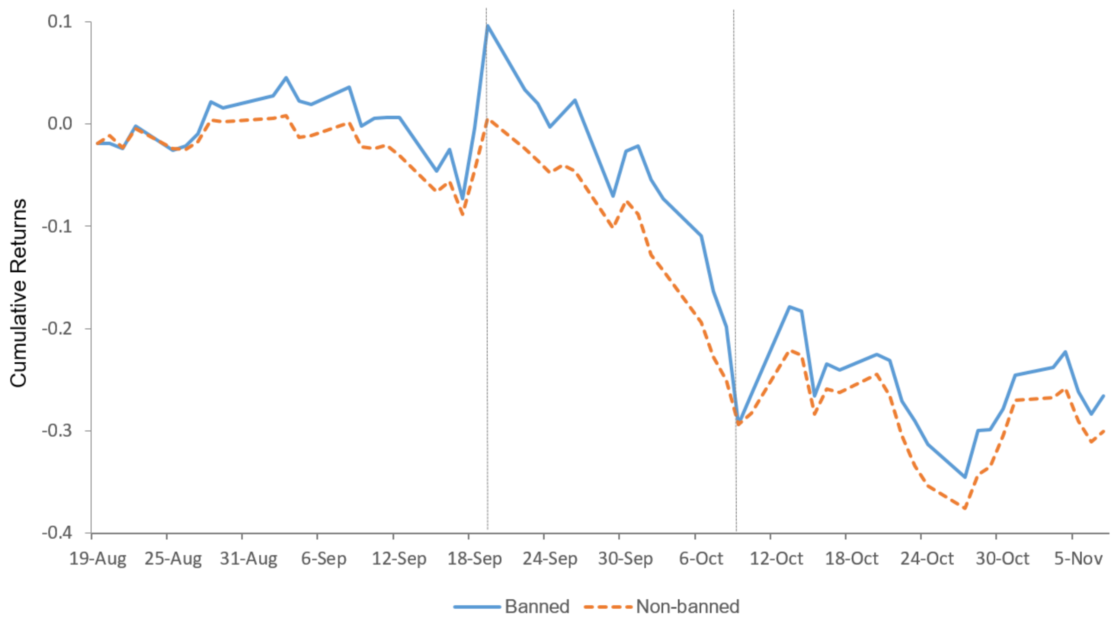

Fig. 8. **The Price Effect of the 2008 Short Selling Ban on Financial Stocks: Alternative Definition of f Based on Operating Profits.** This figure plots the cumulative returns of R^{BF} and the cumulative returns of R^{NF} estimated from Equation (34). The vertical lines indicate the first and the last day of the ban.

to create TARP on September 18. On September 29, however, the plan was rejected by the Congress. The Congress did pass the Emergency Economic Stabilization Act (EESA) and president Bush signed it into law a few days later. The detailed Capital Purchase Program was not announced, however, until October 14, after the short-selling ban was lifted. There is some ambiguity about the valuation effect of the TARP program (Bayaitova and Shivdasani (2012)), especially during the short-selling ban period, before the details of the program are spelled out. Indeed, Bayaitova and Shivdasani (2012) suggest that the TARP relief announcement on October 14 was largely unanticipated by the market. Although we cannot completely rule out the impact of TARP, our convergence result strongly suggests that the price inflation of the banned financial stocks at the beginning of the ban cannot be entirely due to TARP. If the initially price inflation is driven by TARP and this effect is permanent, then we should observe such overvaluation to persist after the ban. In other words, the fact

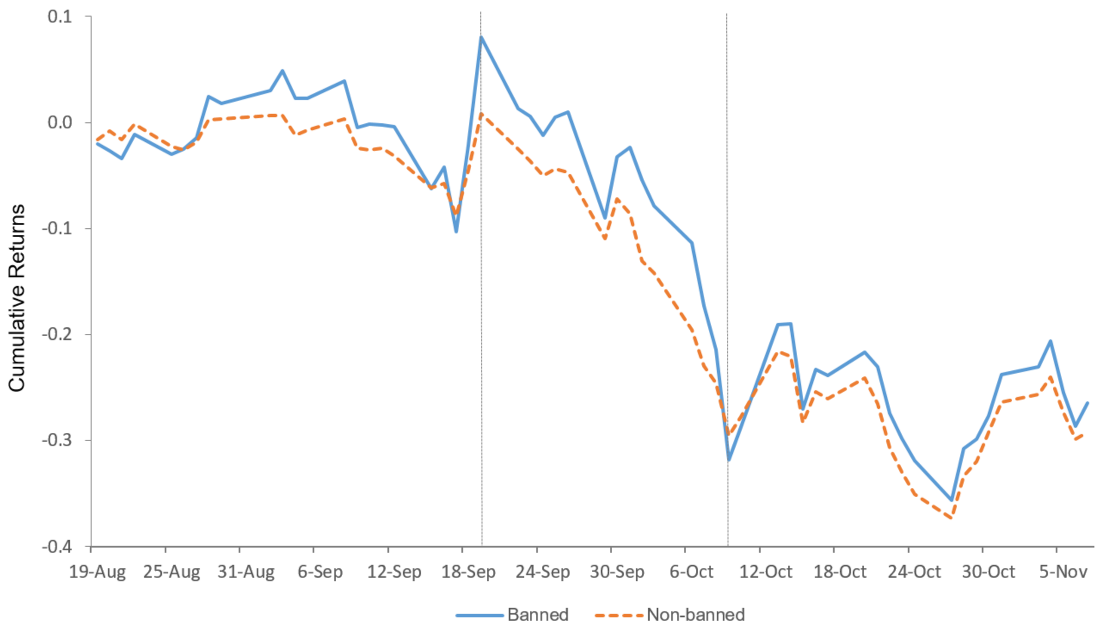

Fig. 9. **The Price Effect of the 2008 Short Selling Ban on Financial Stocks: Alternative Definition of f Based on Total Assets.** This figure plots the cumulative returns of R^{BF} and the cumulative returns of R^{NF} estimated from Equation (34). The vertical lines indicate the first and the last day of the ban.

that the initial overpricing shrinks to zero at the end of the ban suggests that our results are more likely due to the ban itself rather than the TARP.

4. Conclusion

Short-selling bans have become a regular and predictable response from regulators when facing market crises. In this paper, we study the price effect of temporary short-selling bans both theoretically and empirically. We first build a continuous-time asset pricing model with heterogenous beliefs and temporary short-selling bans by extending the work of Scheinkman and Xiong (2003). We obtain a closed-form solution for the overpricing caused by the short-selling ban. This solution enables us to study the dynamics of the overpricing caused by a temporary short-selling ban for the first time. Our model predicts that the size of the initial overpricing is increasing in the length of the ban and the degree of overconfidence, and decreasing in the level of interest rate and the speed of the mean reversion in the fundamentals. More importantly, our model predicts that a temporary short-selling ban leads to an overpricing that is the highest at the beginning of the ban and gradually converges to zero at the end of the ban.

We test this prediction by revisiting the 2008 short-selling ban of financial stocks in the U.S. Previous studies compare banned financial stocks to non-banned non-financial stocks. Such comparison cannot account for the fundamental difference between financial and non-financial stocks during the financial crisis. The innovation of our empirical design is to use the financial segments of non-banned stocks as a control group for the banned financial stocks. We find evidence consistent with our model's prediction. Because the initial overpricing shrinks to zero at the end of the ban, our results are unlikely to be driven entirely by the TARP effect.

References

Asquith, P., Pathak, P. A., Ritter, J. R., 2005. Short interest, institutional ownership, and stock returns. Journal of Financial Economics 78, 243–276.

Autore, D., Billingsley, R., Kovacs, T., 2011. The 2008 short sale ban: Liquidity, dispersion of opinion, and the cross-section of returns of us financial stocks. Journal of Banking & Finance 35, 2252–2266.

Bai, Y., Chang, E. C., Wang, J., 2006. Asset prices under short-sale constraints. Working Paper MIT .

Battalio, R., Mehran, H., Schultz, P., 2012. Market declines: What is accomplished by banning short-selling? Current Issues in Economics and Finance 18, 1–7.

Battalio, R., Schultz, P., 2011. Regulatory uncertainty and market liquidity: The 2008 short sale ban's impact on equity option markets. The Journal of Finance 66, 2013–2053.

Beber, A., Pagano, M., 2013. Short-selling bans around the world: Evidence from the 2007–09 crisis. The Journal of Finance 68, 343–381.

Beneish, M., Lee, C., Nichols, D., 2015. In short supply: Short-sellers and stock returns. Journal of Accounting and Economics 60, 33–57.

Boehmer, E., Jones, C., Zhang, X., 2013. Shackling Short Sellers: The 2008 Shorting Ban. Review of Financial Studies 26, 1363–1400.

Boehmer, E., Jones, C. M., Zhang, X., 2008. Which shorts are informed? The Journal of Finance 63, 491–527.

Desai, H., Ramesh, K., Thiagarajan, S. R., Balachandran, B. V., 2002. An investigation of the informational role of short interest in the nasdaq market. The Journal of Finance 57, 2263–2287.

Diamond, D. W., Verrecchia, R. E., 1987. Constraints on short-selling and asset price adjustment to private information. Journal of Financial Economics 18, 277–311.

Diether, K. B., Lee, K.-H., Werner, I. M., 2009. Short-Sale Strategies and Return Predictability. The Review of Financial Studies 22, 575–607.

Drechsler, I., Drechsler, Q. F., 2016. The shorting premium and asset pricing anomalies. Working Paper New York University .

Engelberg, J. E., Reed, A. V., Ringgenberg, M. C., 2012. How are shorts informed?: Short sellers, news, and information processing. Journal of Financial Economics 105, 260–278.

Engelberg, J. E., Reed, A. V., Ringgenberg, M. C., 2018. Short-selling risk. The Journal of Finance 73, 755–786.

Fama, E. F., French, K. R., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 3–56.

Harris, L., Namvar, E., Phillips, B., 2013. Price inflation and wealth transfer during the 2008 sec short-sale ban. The Journal of Investment Management, Second Quarter 11, 1–23.

Jones, C. M., Reed, A. V., Waller, W., 2016. Revealing Shorts An Examination of Large Short Position Disclosures. The Review of Financial Studies 29, 3278–3320.

Kelley, E. K., Tetlock, P. C., 2016. Retail Short Selling and Stock Prices. The Review of Financial Studies 30, 801–834.

McLean, R. D., Pontiff, J., Reilly, C., 2020. Taking sides on return predictability. Working Paper Georgetown University .

Miller, E. M., 1977. Risk, uncertainty, and divergence of opinion. The Journal of Finance 32, 1151–1168.

Scheinkman, J. A., Xiong, W., 2003. Overconfidence and speculative bubbles. Journal of Political Economy 111, 1183–1220.

Wang, X., Yan, X. S., Zheng, L., 2020. Shorting flows, public disclosure, and market efficiency. Journal of Financial Economics 135, 191–212.

Appendix A.

A.1. Proof of Lemma 1

Let $k \geqslant 0$ be the minimum difference of opinions to generate a trade.

If the current belief difference is greater than k, the owner chooses to sell right now. Then the optimal execution time is equal to current time $\tau = t$ and the optimal value of the right hand side is realized in the first equation of (18) right now.

$$q(x,t) = \frac{x}{r+\lambda}(1 - e^{-(r+\lambda)(T-t)}) + q(-x,t), \qquad x \geqslant k \tag{37}$$

The requirement of $q(x,t)$ being smooth in equation (18) means $q(x,t)$ and $q_x(x,t)$ are continuous when x crosses k.

Since $\underset{x\to k+}{lim} q(x,t) = \underset{x\to k+}{lim}[\frac{x}{r+\lambda}(1-e^{-(r+\lambda)(T-t)})+q(-x,t)] = \frac{k}{r+\lambda}(1-e^{-(r+\lambda)(T-t)})+q(-k,t)$ and $\underset{x\to k-}{lim} q(x,t) = q(k,t)$, the continuation of $q(x,t)$ implies:

$$k \cdot \frac{1}{r+\lambda}(1 - e^{-(r+\lambda)(T-t)}) = q(k,t) - q(-k,t) \tag{38}$$

Since $\underset{x\to k+}{lim} q_x(x,t) = \underset{x\to k+}{lim}[\frac{1}{r+\lambda}(1 - e^{-(r+\lambda)(T-t)}) - q_x(-x,t)] = \frac{1}{r+\lambda}(1 - e^{-(r+\lambda)(T-t)}) - q_x(-k,t)$ and $\underset{x\to 0-}{lim} q_x(x,t) = q_x(k,t)$, the continuation of $q_x(x,t)$ implies:

$$\frac{1}{(r+\lambda)}(1 - e^{-(r+\lambda)(T-t)}) = q_x(k,t) + q_x(-k,t) \tag{39}$$

substitute (39) into (38), we know that k must satisfy:

$$k \cdot [q_x(k,t) + q_x(-k,t)] - q(k,t) + q(-k,t) = 0 \tag{40}$$

Let $l(k) = k \cdot [q_x(k,t) + q_x(-k,t)] - q(k,t) + q(-k,t)$, we have $l(0) = 0$. Thus 0 is one solution of (40).

we also have $l'(k) = k \cdot [q_{xx}(k,t) - q_{xx}(-k,t)]$. Similar to Lemma 2 in Scheinkman and

Xiong (2003), we know that $q_{xx}(x,t)$ is increasing. When $k > 0$, we have $k > -k$. This means $q_{xx}(k,t) - q_{xx}(-k,t) > 0$. Thus $l'(k) > 0$. When $k < 0$, we have $k < -k$. This means $q_{xx}(k,t) - q_{xx}(-k,t) < 0$. Thus we also have $l'(k) > 0$. Thus we know that $l'(k) > 0$ for all $k \neq 0$. This means 0 is the only solution of (40). We have proved that the owner will wait if $x < 0$ and sell if $x \geqslant 0$.

A.2. Proof of Theorem 1

The PDE to be solved is

$$\begin{cases} \frac{1}{2}\sigma^2 q_{xx} - \rho x q_x + q_t - rq = 0 \\ q(x,T) = 0 \\ q_x(0,t) = \frac{1}{2(r+\lambda)}(1 - e^{-(r+\lambda)(T-t)}) \end{cases} \tag{41}$$

where $x < 0$ and $0 \leqslant t \leqslant T$. The boundary condition is at the end of the ban when $t = T$. By the transform $\tau = T - t$, one can make the boundary condition lie at the beginning of the new time τ. Let $l(x,\tau) = q(x, T-\tau)$, then $l(x,\tau)$ satisfies

$$\begin{cases} \frac{1}{2}\sigma^2 l_{xx} - \rho x l_x - l_\tau - rl = 0 & \textit{PDE} \\ l(x,0) = 0 & \textit{Initial} \quad \textit{condition} \\ l_x(0,\tau) = \frac{1}{2(r+\lambda)}(1 - e^{-(r+\lambda)\tau}) & \textit{Boundary} \quad \textit{condition} \end{cases} \tag{42}$$

where $x < 0$ and $0 \leqslant \tau$. The PDE and the boundary condition is well-defined for all $\tau > T$, thus we can solve it for all $\tau \geqslant 0$ and then pick the $0 \leqslant \tau \leqslant T$ part.

Consider performing a Laplace transform on this PDE. The Laplace transform of a function $f(\tau)$ is $F(s) = \mathfrak{L}[f(\tau)] = \int_0^\infty e^{-s\tau} f(\tau) d\tau$. Let $L(x,s) = \mathfrak{L}[l(x,\tau)]$. Performing a Laplace transform on the PDE of $l(x,\tau)$ and using the following properties of the Laplace transform

$$\mathfrak{L}[af(\tau) + bg(\tau)] = aF(s) + bG(s) \tag{43}$$

$$\mathfrak{L}[f'(\tau)] = sF(s) - f(0) \tag{44}$$

one gets

$$\frac{1}{2}\sigma_g^2 L_{xx} - \rho x L_x - [sL(x,s) - l(x,0)] - rL = 0 \tag{45}$$

Using the initial condition $l(x, 0) = 0$, one has

$$\frac{1}{2}\sigma_g^2 L_{xx} - \rho x L_x - (s + r)L = 0 \tag{46}$$

For a fixed s, this is an ODE of x. This ODE has a class of solutions of the form

$$L(x, s) = \beta(s)h(x, s) \tag{47}$$

where $\beta(s)$ is an arbitrary function of s and

$$h(x, s) = U(\frac{r + s}{2\rho}, \frac{1}{2}, \frac{\rho}{\sigma^2}x^2) \tag{48}$$

where $U(\cdot)$ is the Kummer U function[7]

$$U(a, b, z) = \frac{1}{\Gamma(a)}\int_0^\infty e^{-zy}y^{a-1}(1 + y)^{b-a-1}dy. \tag{49}$$

We can use the boundary condition to pin down the function $\beta(s)$. Using linearity and the following property of Laplace transform

$$\mathfrak{L}[e^{a\tau}] = \frac{1}{s - a} \tag{50}$$

one has the Laplace transform of the boundary condition

$$L_x(0, s) = \mathfrak{L}[l_x(0, \tau)] = \mathfrak{L}[\frac{1}{2(r + \lambda)}(1 - e^{-(r+\lambda)\tau})] = \frac{1}{2(r + \lambda)}(\frac{1}{s} - \frac{1}{s + r + \lambda}). \tag{51}$$

On the other hand, we know $L_x(0, s) = \beta(s)h_x(0, s)$ and $h_x(0, s) = \frac{\pi\sqrt{\rho}}{\sigma\Gamma(\frac{r+s}{2\rho})\Gamma(\frac{3}{2})}$. Thus we can solve $\beta(s)$

[7] Scheinkman and Xiong (2003) use this Kummer U function to express the solution of a similar ODE. They use the sum of series for the U function. We use the integration form of the Kummer U function because it makes the inverse Laplace transform possible.

$$\beta(s) = \frac{\sigma\Gamma(\frac{r+s}{2\rho})\Gamma(\frac{3}{2})}{2(r+\lambda)\pi\sqrt{\rho}}(\frac{1}{s} - \frac{1}{s+r+\lambda}). \tag{52}$$

Based on equations (47), (48), (49) and (52) we have

$$L(x,s) = \int_0^\infty A(x,y)\left(\frac{1}{s} - \frac{1}{s+r+\lambda}\right)e^{-M(y)s}dy \tag{53}$$

where $A(x,y) = \frac{\sigma}{4(r+\lambda)\sqrt{\pi\rho}}e^{-\frac{\rho}{\sigma^2}x^2y}y^{\frac{r-2\rho}{2\rho}}(1+y)^{-\frac{r+\rho}{2\rho}}$ and $M(y) = \frac{1}{2\rho}ln(1+\frac{1}{y})$

With the following property of the inverse Laplace transform, we can get $l(x,\tau)$ from $L(x,s)$:

$$\mathfrak{L}^{-1}[\frac{1}{s+a}e^{-bs}] = u(\tau-b)e^{-a(\tau-b)} \tag{54}$$

where $u(x)$ is the unit step function: if $x > 0$, $u(x) = 1$; if $x < 0$, $u(x) = 0$.

$$l(x,\tau) = \mathfrak{L}^{-1}[L(x,s)] = \int_{\tilde{N}(\tau)}^\infty A(x,y)[1 - e^{-(r+\lambda)(\tau - M(y))}]dy \tag{55}$$

where $\tilde{N}(\tau) = \frac{e^{-2\rho\tau}}{1-e^{-2\rho\tau}}$.

Thus, the solution to our original PDE (41) is $u(x,t) = l(x, T-t)$. So

$$u(x,t) = \frac{\sigma}{4(r+\lambda)\sqrt{\pi\rho}}\int_{N(t)}^\infty e^{-\frac{\rho}{\sigma^2}x^2y}y^{\frac{r-2\rho}{2\rho}}(1+y)^{-\frac{r+\rho}{2\rho}}[1 - e^{-(r+\lambda)(T-t-M(y))}]dy \tag{56}$$

where $x < 0, 0 \leqslant t \leqslant T$, $N(t) = \frac{e^{-2\rho(T-t)}}{1-e^{-2\rho(T-t)}}$ and $M(y) = \frac{1}{2\rho}ln(1+\frac{1}{y})$.

A.3. *Proof of Theorem 2*

First, we prove $u_t(x,t) < 0$.

Since

$$u(x,t) = \int_{N(t)}^{\infty} A(x,y)[1 - e^{-(r+\lambda)(T-t-M(y))}]dy \tag{57}$$

where $A(x,y) = \frac{\sigma}{4(r+\lambda)\sqrt{\pi\rho}} e^{-\frac{\rho}{\sigma^2}x^2 y} y^{\frac{r-2\rho}{2\rho}} (1+y)^{-\frac{r+\rho}{2\rho}}$, $N(t) = \frac{e^{-2\rho(T-t)}}{1-e^{-2\rho(T-t)}}$ and $M(y) = \frac{1}{2\rho} ln(1+\frac{1}{y})$, we have

$$\begin{aligned} u_t(x,t) = \quad & -A(x,N(t))[1 - e^{-(r+\lambda)\{T-t-M[N(t)]\}}]\frac{dN(t)}{dt} \\ & - \textstyle\int_{N(t)}^{\infty} A(x,y)(r+\lambda)e^{-(r+\lambda)(T-t-M(y))}dy \end{aligned} \tag{58}$$

where $A(x,N(t))[1-e^{-(r+\lambda)\{T-t-M[N(t)]\}}]\frac{dN(t)}{dt} = 0$ because $M[N(t)] = T-t$. The integral $\int_{N(t)}^{\infty} A(x,y)(r+\lambda)e^{-(r+\lambda)(T-t-M(y))}dy$ is positive, because

$$(r+\lambda)e^{-(r+\lambda)(T-t-M(y))} > 0 \tag{59}$$

by definition and when $y > N(t)$, we have $A(x,y) > 0$. Thus $u_t(x,t) < 0$.

Since

$$q(x,t) = \begin{cases} \frac{x}{r+\lambda}(1 - e^{-(r+\lambda)(T-t)}) + u(-x,t) & x \geqslant 0 \\ u(x,t) & x < 0. \end{cases} \tag{60}$$

we know

$$q_t(x,t) = \begin{cases} -xe^{-(r+\lambda)(T-t)} + u_t(-x,t) & x \geqslant 0 \\ u_t(x,t) & x < 0. \end{cases} \tag{61}$$

Thus $q_t(x,t) < 0$.

On the other hand,

$$q(x,T) = \begin{cases} u(-x,T) & x \geqslant 0 \\ \\ u(x,T) & x < 0 \end{cases} \tag{62}$$

Thus $q(x,T) = 0$ because $u(x,T) = 0$.

A.4. Proof of Theorem 3

Since $x \leqslant 0$,

$$\bar{q}(t) = \int_{-\infty}^{0} u(x,t) \cdot w(x) dx \tag{63}$$

where, according to equation (30),

$$w(x)dx = 2 \cdot \frac{1}{\eta\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2\eta^2}} dx \tag{64}$$

with $\eta^2 = \frac{\sigma^2}{2\rho}$ and

$$u(x,t) = \frac{\sigma}{4(r+\lambda)\sqrt{\pi\rho}} \int_{N(t)}^{\infty} e^{-\frac{\rho}{\sigma^2}x^2 y} y^{\frac{r-2\rho}{2\rho}} (1+y)^{-\frac{r+\rho}{2\rho}} [1 - e^{-(r+\lambda)(T-t-M(y))}] dy \tag{65}$$

where $N(t) = \frac{e^{-2\rho(T-t)}}{1-e^{-2\rho(T-t)}}$, $M(y) = \frac{1}{2\rho} ln(1+\frac{1}{y})$, $x < 0$ and $0 \leqslant t \leqslant T$ according to Theorem 1.

In $u(x,t)$, the only term containing x is $e^{-\frac{\rho}{\sigma^2}x^2 y}$. Let $\phi = \frac{\eta}{\sqrt{1+y}}$, then

$$\int_{-\infty}^{0} e^{-\frac{\rho}{\sigma^2}x^2 y} \cdot w(x) dx = (1+y)^{-\frac{1}{2}} \cdot \frac{2}{\phi\sqrt{2\pi}} \int_{-\infty}^{0} e^{-\frac{x^2}{2\phi^2}} dx = (1+y)^{-\frac{1}{2}}. \tag{66}$$

With the term containing x integrated,

$$\bar{p}(t) = \frac{\sigma}{4(r+\lambda)\sqrt{\pi\rho}} \int_{N(t)}^{\infty} (1+y)^{-\frac{1}{2}} y^{\frac{r-2\rho}{2\rho}} (1+y)^{-\frac{r+\rho}{2\rho}} [1 - e^{-(r+\lambda)(T-t-M(y))}] dy. \tag{67}$$

Plugging $M(y) = \frac{1}{2\rho} ln(1+\frac{1}{y})$ into this equation and reorganizing the terms,

$$\bar{p}(t) = \frac{\sigma}{4(r+\lambda)\sqrt{\pi\rho}}[\int_{N(t)}^{\infty} y^{\frac{r}{2\rho}-1}(1+y)^{-\frac{r}{2\rho}-1}dy - e^{-(r+\lambda)(T-t)}\int_{N(t)}^{\infty} y^{-\frac{\lambda}{2\rho}-1}(1+y)^{\frac{\lambda}{2\rho}-1}dy] \quad (68)$$

With the integration formula $\int y^{a-1}(1+y)^{-a-1}dy = \frac{1}{a}y^a(1-y)^{-a}$, $\bar{p}(t)$ can be simplified as

$$\bar{p}(t) = \frac{\sigma}{4(r+\lambda)\sqrt{\pi\rho}}[\frac{2\rho}{r}y^{\frac{r}{2\rho}}(1+y)^{-\frac{r}{2\rho}} + e^{-(r+\lambda)(T-t)}\frac{2\rho}{\lambda}y^{-\frac{\lambda}{2\rho}}(1+y)^{\frac{\lambda}{2\rho}}]|_{N(t)}^{\infty}. \quad (69)$$

Plugging $N(t) = \frac{e^{-2\rho(T-t)}}{1-e^{-2\rho(T-t)}}$ into this equation and reorganizing the terms,

$$\bar{q}(t) = \frac{\sigma\sqrt{\rho}}{2\lambda\sqrt{\pi}}[\frac{1}{r}(1-e^{-r(T-t)}) - \frac{1}{r+\lambda}(1-e^{-(r+\lambda)(T-t)})]. \quad (70)$$

A.5. Proof of Theorem 4

Based on the equation of $\bar{q}(t)$

$$\frac{d\bar{p}(t)}{dt} = -\frac{\sigma\sqrt{\rho}}{2\lambda\sqrt{\pi}} e^{-r(T-t)}[1 - e^{-\lambda(T-t)}] < 0 \tag{71}$$

since all the parameters are positive and $0 \leqslant t \leqslant T$.

It is also easy to verify that $\bar{q}(T) = 0$. Thus the bubble shrinks and vanishes at the end of the ban.

Since

$$\frac{d^2\bar{p}(t)}{dt^2} = -\frac{\sigma\sqrt{\rho}}{2\lambda\sqrt{\pi}}[r - (r+\lambda)e^{-\lambda(T-t)}]e^{-r(T-t)} \tag{72}$$

Letting $\frac{d^2\bar{p}(t)}{dt^2} = 0$, one can solve for the inflection point $t^* = T - \tau^*$, where $\tau^* = \frac{1}{\lambda} ln(1 + \frac{\lambda}{r})$.

The size of the bubble at the inflection point is $\bar{p}(t^*) = \frac{\sigma\sqrt{\rho}}{2r(r+\lambda)\sqrt{\pi}}[1 - (\frac{r}{r+\lambda})^{\frac{r}{\lambda}} \frac{2r+\lambda}{r+\lambda}]$.

The decreasing speed at the inflection point is $\frac{d\bar{p}(t)}{dt}|_{t=t^*} = \frac{-\sigma\sqrt{\rho}}{2(r+\lambda)\sqrt{\pi}}(\frac{r}{r+\lambda})^{\frac{r}{\lambda}}$.