Option Pricing with an Illiquid Underlying Asset Market*

Hong Liu†
Olin School of Business
Washington University
St. Louis, MO 63130

Jiongmin Yong ‡
Department of Mathematics
University of Central Florida
Orlando, FL 32816

First Draft: October 2000
Last revised: April 26, 2004

Abstract

We examine how price impact in the underlying asset market affects the replication of a European contingent claim. We obtain a generalized Black-Scholes pricing PDE and establish the existence and uniqueness of a classical solution to this PDE. We show that unlike the case with transaction costs, replication in the presence of price impact is always cheaper than superreplication. This model implies *endogenous* stochastic volatility. Compared to the Black-Scholes case, a trader generally buys more stock and borrows more (shorts and lends more) to replicate a call (put). The excess replicating cost over the Black-Scholes price is significant. Furthermore, price impact implies that an out-of-money option has lower implied volatility than an in-the-money option. This finding has important implications for empirical analysis on "volatility smile."

Journal of Economic Literature Classification Numbers: D11, D91, G11, C61.
Keywords: Price Impact, Option Pricing, Illiquidity

---

*We thank Kerry Back, Domenico Cuoco, Yuriy Fedyk, and two anonymous referees for extremely helpful suggestions. We are also grateful to Xavier Gabaix, Cars Hommes (the Editor), Leonid Kogan, Jun Pan, Anna Pavlova, Steve Ross, Dimitri Vayanos, Jiang Wang, and seminar participants at MIT and the 2001 International Mathematical Finance Conference for useful comments. Of course, any remaining errors are ours.

†Corresponding author. Tel.: 314-935-5883; fax: 314-935-6359. *E-mail address*: liuh@wustl.edu.

‡This work is supported in part by the NSFC under grant 10131030, the Chinese Education Ministry Science Foundation under grant 2000024605, the Cheung Kong Scholars Programme, and Shanghai Commission of Science and Technology Grant 02DJ14063.

I. Introduction

Most of the option pricing models assume that an option trader cannot affect the underlying asset price in trading the underlying asset to replicate the option payoff, regardless of her trading size. This is reasonable only in a perfectly liquid market. In a market with imperfect liquidity, however, trading does affect the underlying asset price. Indeed, the presence of price impact of investors' trading has been widely documented and extensively analyzed in the literature (see, for example, Chan and Lakonishok (1995), Keim and Madhavan (1996), Sharpe, Alexander, and Bailey (1999), and Jorion (2000)). Even for a very liquid market, trading beyond the quoted depth usually results in a worse price for at least part of the trade.

Consistent with the above discussion, in this paper we take this imperfect liquidity as given and examine how it affects the replication of a European option by a typical option trader. In particular, we assume that as the trader buys the underlying asset (stock) to hedge her position in the option, the stock price goes up and as she sells, the price goes down.

Several issues are critical for understanding how this price impact affects the replication of a European option. First, since in the presence of this adverse price impact trading in the stock to replicate the option affects the stock price, it is not clear whether the option is still perfectly replicable or not. Second, the presence of adverse price impact increases the replicating costs. It is well known that in the presence of transaction cost, superreplication of an option (e.g., buying a share to superreplicate a call) costs less than exact replication. Therefore a natural question is whether it is also cheaper to superreplicate than to exactly replicate in the presence of price impact. Third, if it is cheaper to exactly replicate, what is the extra replication cost over the Black-Scholes price that a trader has to incur? In addition, how should the trader trade the underlying asset to replicate? Finally, what are the implications of the price impact on the well-known "volatility smile"?

To answer these questions, we use the idea of a four-step scheme for forward and backward stochastic differential equations (FBSDEs) (see Ma, Protter, and Yong (1994), Yong (1999), Yong and Zhou (1999)) to derive a generalized, nonlinear Black-Scholes partial differential equation (PDE) for computing the replicating cost of a European option. We provide sufficient conditions under which the option is perfectly replicable. This pricing PDE shows that the effect of the price impact on the replicating cost is only through the impact of the trader's trading on the stock return volatility. We then show that, unlike the transaction cost case, superreplication is more costly than exact replication. Furthermore, like the Black-Scholes case, the replicating strategy involves an initial block trade followed by continuous trading. In addition, we show that the excess replicating cost over the Black-Scholes price is significant, even with a small price impact in the underlying asset market.

We find that a trader generally buys more stock and borrows more (shorts and lends more) to replicate a call (put). In the special case in which option payoffs are linear in the stock price (e.g., forwards, futures, or shares), the trader adopts the same strategy as in the case without price impact. However, the cost is higher due to the adverse price impact.

The presence of price impact implies that although a special form of put-call parity still holds, the implied volatility for a put is *different* from the implied volatility for the otherwise identical call. We find that out-of-money options have lower implied volatility than in-the-money options. This pattern is consistent with the well-known volatility smile found for calls, but is different from the volatility smile observed for puts (see Dumas, Fleming, and Whaley (1998), for example). Intuitively, as a trader trades the price moves against her, so she incurs higher replicating costs. When an option is in the money, she needs to trade more in the stock. So the extra replicating cost over the Black-Scholes price is greater. On the other hand, option price is not sensitive to volatility in the Black-Scholes world for an in-the-money option. This implies that a large implied volatility is required to generate the higher replicating costs that resulted from the price impact. When an option is out of the money, however, she needs to trade less in the stock to replicate the option. So the extra cost is smaller and thus the implied volatility is also smaller. Although the existence of price impact implies opposite patterns to the volatility smile documented for puts, it does have important implications for empirical analysis on the volatility smile. For example, the opposite implied volatility patterns produced by the price impact for calls versus puts suggest that if this liquidity imperfectness were taken into account, empirical studies such as Dumas, Fleming, and Whaley (1998) would have to explain a flatter skew for calls and a steeper skew for puts.

A number of option valuation models in the literature attempt to explain the volatility smile. The stochastic volatility models of Heston (1993) and Hull and White (1987), for example, can potentially explain the smile to some extent when the asset price and the volatility are negatively correlated. Similarly, the jump model of Bates (1996) is also consistent with the smile when the mean jump is negative. The deterministic volatility model examined by Dumas, Fleming, and Whaley (1998) can also generate a similar smile pattern. However, all these models assume an exogenous volatility process. In contrast, in our model the volatility process is endogenous and is affected by the trading of the trader.

In the presence of price impact, the no-arbitrage price of an option for a trader is no longer unique. Rather, it consists of a continuum of prices within an interval. We find that this no-arbitrage interval expands as the price impact increases. In addition, the price impact also introduces nonlinearity into the dynamics of the replicating portfolio value. We show that the excess replicating cost is approximately quadratic in the number of units of an option.

There is an extensive literature on the effect of price impact. In the presence of asymmetric information, Kyle (1985) and Back (1993) use an equilibrium approach to investigate how informed traders reveal information and affect the market price through trading. As shown by Kyle (1985) and Back (1993), equilibrium asset prices are directly affected by the informed trader's trades. Vayanos (2001) studies a dynamic model of a financial market with a large trader who does not have any private information on the asset value but trades only to share risk. He shows that the equilibrium stock price is linear in the investor's order size. These models provide theoretic justifications for the existence, the form, and the direction of the price impact a trader can have on stock prices. In particular, the price impact form used in this paper, which is linear in the trading size, is consistent with the equilibrium price impact forms derived in these models.

Cvitanić and Ma (1996) and Ma and Yong (1999) examine the hedging costs of options for a trader in the presence of price impact. Cuoco and Cvitanić (1998) consider the effect of the price impact on the optimal consumption and investment policy. In these papers, it was assumed that price impact depends only on the total wealth and the position of a trader but not on how she trades.

Our model includes Frey (2000) as a special case. Frey (2000) ignores the initial block trade (and thus the initial extra cost) that is necessary for any replication. We show that ignoring the initial block trade would significantly understate the replicating costs and produce qualitatively misleading conclusions. In addition, he does not show the uniqueness or the existence of a solution to the pricing PDE he derived and thus does not show the replicability of an option. Moreover, economic analysis in Frey (2000) is also very limited. Sircar and Papanicolaou (1998) assume an exogenous demand function for the reference traders and derive a different nonlinear pricing PDE that depends on the exogenous income process of the reference traders and the relative size of the program traders.

To focus on our main objective of understanding the replicating strategy and the replicating cost for an option written on an illiquid asset, we use a partial equilibrium approach in this paper. In particular, we take the price impact function as given. As will be shown, this model provides an economically sensible characterization of the replicating strategy and a reasonable estimate of the replicating cost for a typical trader in a European option market. In addition, it can be justified by many equilibrium models such as the following example, which is similar to the model in Back (1993). Consider an economy in which there are risk-neutral and competitive stock market makers, risk-averse and competitive option market makers, a risk-neutral informed trader, and liquidity investors. The stock market makers trade only in the stock whereas the option market makers trade only in the options. The insider and liquidity traders can trade both the stock and the options. At a future time T, there is to be a public information release on the stock value that will fix the stock price at an exogenous level $\hat{S}(T)$, which is known in advance only by the informed trader. All options expire immediately after the release of the public information at T. Due to competition and risk aversion, the option market makers trade options at the replicating costs and perfectly hedge in the stock market. In equilibrium, the insider will trade in such a way that the stock price at T will be exactly equal to $\hat{S}(T)$ given the option market makers' hedging trades and liquidity traders' liquidity trades. The price impact function assumed in the model can be interpreted as the equilibrium price response function to the option market maker's hedging trades, given the liquidity traders' and the insider's equilibrium trades. In this framework of price manipulation in the stock market, such as that suggested by Jarrow (1992), Allen and Gale (1992), Vila (1989), Bagnoli and Lipman (1990), and Schönbucher and Wilmott (2000), to affect the payoff of an option is impossible because the payoff only depends on $\hat{S}(T)$ and the option's strike price.

The rest of the paper is organized as follows. In Section 2, we introduce our model. In Section 3, we derive the generalized Black-Scholes pricing PDE in the presence of price impact and provide sufficient conditions under which a European option can be replicated. We also show that superreplication is more expensive than exact replication. In Section 4, we provide a numerical analysis of the model for European calls and puts. Section 5 contains the concluding remarks. In the Appendix, we provide proofs of the theorems.

II. The Model

Throughout this paper we fix a complete filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbf{P})$ on which a standard one dimensional Brownian motion $B(t)$ is defined with $\{\mathcal{F}_t\}_{t\geq 0}$ being its natural filtration augmented by all the $\mathbf{P}$-null sets. All the stochastic processes in this paper are assumed to be $\{\mathcal{F}_t\}_{t\geq 0}$-adapted.

There are two assets being continuously traded in the primary market. The first asset is a money market account. The second is a risky asset, which we will call a stock. Let $S(t)$ be the ex-dividend stock price and $\delta(t, S(t))$ be the dividend yield of the stock. We assume that the risk-free asset price $S_0(t)$ satisfies

$$\frac{dS_0(t)}{S_0(t)} = r(t, S(t))dt, \qquad t \geq 0, \tag{1}$$

where $r(t, S(t))$ is the interest rate and we allow it to depend on the current stock price directly. In addition, there is a derivative market in which a trader can also trade options on the stock. In contrast to the standard framework and consistent with a market with imperfect liquidity, we assume that the option trader's trading in the stock market has a direct impact on the stock price. In particular, when the trader buys, the stock price goes up and when she sells, the stock price goes down. Specifically, let $N(t)$ be the number of shares that the trader has in the stock at time t. Then the stock price $S(t)$ is assumed to evolve as follows,

$$\frac{dS(t)}{S(t)} = \mu(t, S(t))dt + \sigma(t, S(t))dB(t) + \lambda(t, S(t))dN(t), \quad t \geq 0, \tag{2}$$

where $\lambda(t, S(t)) \geq 0$ is the price impact function of the trader, and $\mu(t, S(t))$ and $\sigma(t, S(t))$ are the expected return and the volatility respectively in the absence of any trading by the trader. The term $\lambda(t, S(t))dN(t)$ represents the price impact of the investor's trading. We note that the classical Black-Scholes model is a special case of this model where $\lambda(t, S(t)) \equiv 0$.

The wealth process $W(\cdot)$ for the trader then satisfies the following budget equation:

$$\begin{aligned} dW(t) = r(t, S(t))W(t)dt + N(t)S(t)[\mu(t, S(t)) + \delta(t, S(t)) - r(t, S(t))]dt \\ + N(t)S(t)\sigma(t, S(t))dB(t) + N(t)S(t)\lambda(t, S(t))dN(t), \quad t \geq 0. \end{aligned} \tag{3}$$

The price impact term in (2) leads to the last quadratic term in the budget equation. This quadratic term is the only difference from the wealth equation for a small trader who has no price impact. The presence of this term implies that unlike the standard case, the wealth dynamics for a trader is no longer linear in her trading strategy N. Implications of this nonlinearity will be explored in later parts of this paper. In the absence of price impact (i.e., the Black-Scholes case), the replicating strategy involves trading a discrete number of shares at time 0 and then trading continuously according to an Itô process. Accordingly, throughout this paper we only consider trading strategies that are Itô processes except possibly a discrete jump at time 0; i.e., we assume that $N(\cdot)$ satisfies

$$\begin{cases} dN(t) = \eta(t)dt + \zeta(t)dB(t), \quad t \geq 0, \\ N(0) = N_0, \end{cases} \tag{4}$$

for some processes $\eta(\cdot)$ and $\zeta(\cdot)$ (to be endogenously determined), where N_0 is the number of shares in the stock after possibly an initial jump from $N(0_-)$. Thus, by (4) and (2), we have

$$\frac{dS(t)}{S(t)} = [\mu(t, S(t)) + \lambda(t, S(t))\eta(t)]dt + [\sigma(t, S(t)) + \lambda(t, S(t))\zeta(t)]dB(t), \quad t \geq 0. \tag{5}$$

Consequently, the wealth process $W(\cdot)$ satisfies the following stochastic differential equation (SDE):

$$\begin{aligned} dW(t) \quad = \quad & \Big\{ r(t, S(t))W(t) + [\mu(t, S(t)) + \delta(t, S(t)) - r(t, S(t)) + \lambda(t, S(t))\eta(t)] \\ & N(t)S(t) \Big\} dt + [\sigma(t, S(t)) + \lambda(t, S(t))\zeta(t)]N(t)S(t)dB(t), \quad t \geq 0. \end{aligned} \tag{6}$$

III. Replication of a European Option

Let $h(S(T))$ be the payoff of a European contingent claim maturing at time T, where $h : \mathbb{R}_+ \to \mathbb{R}$ is a piecewise linear function and $S(T)$ is the price of the stock at time T. Hereafter, for convenience, we simply call $h(S(T))$ an option. We assume that the option trader's objective is to replicate the option for a perfect hedge.[1] Then replicating such an option amounts to solving (4), (5), and (6) subject to the terminal condition $W(T) = h(S(T))$.[2] For clarity, we collect these SDEs together to form the following system:

$$
\begin{cases}
dN(t) = \eta(t)dt + \zeta(t)dB(t), \\
\frac{dS(t)}{S(t)} = [\mu(t, S(t)) + \lambda(t, S(t))\eta(t)]dt \\
\qquad\qquad + [\sigma(t, S(t)) + \lambda(t, S(t))\zeta(t)]dB(t), \\
dW(t) = \Big\{ r(t, S(t))W(t) + [\mu(t, S(t)) + \delta(t, S(t)) - r(t, S(t)) \\
\qquad\qquad + \lambda(t, S(t))\eta(t)]N(t)S(t) \Big\} dt \\
\qquad\qquad + [\sigma(t, S(t)) + \lambda(t, S(t))\zeta(t)]N(t)S(t)dB(t), \\
N(0) = N_0, \quad S(0) = S_0 > 0, \quad W(T) = h(S(T)).
\end{cases}
\tag{7}
$$

This system of SDEs is called a forward-backward stochastic differential equation (FBSDE) system, since it involves solving forward for $N(t)$ and $S(t)$ and solving backward for $W(t)$.

In the presence of price impact, as an option trader trades the underlying to hedge, the stock price is directly affected and the potential payoff of the option may also be changed. Therefore, one of the interesting questions is whether the trader can still replicate the option in the presence of this price impact. We show next that under some regularity conditions the answer is positive and we provide a generalized nonlinear Black-Scholes pricing PDE required to compute the replicating cost.

First, we introduce some additional notations. Let $x = \ln S$ be the log stock price. Let

$$
\begin{cases}
\widetilde{\mu}(t, x) = \mu(t, e^x), \qquad \widetilde{\sigma}(t, x) = \sigma(t, e^x), \\
\widetilde{r}(t, x) = r(t, e^x), \qquad \widetilde{\delta}(t, x) = \delta(t, e^x), \\
\widetilde{h}(x) = h(e^x), \qquad \widetilde{\lambda}(t, x) = \lambda(t, e^x)e^{-x}, \\
\widetilde{f}(t, x) = f(t, e^x).
\end{cases}
\tag{8}
$$

For any set G in a Euclidean space, let $C(G)$ be the set of all continuous functions

[1] Implicitly, we assume that the option trader cannot manipulate the stock price to reduce her hedging cost, which is consistent with the equilibrium setup discussed in the Introduction. This assumption may also be consistent with the presence of portfolio constraints (e.g., short sale constraint, see Cuoco and Liu (2000) for example), capital shortage, or position limits.

[2] We assume that the European contingent claim is settled by the physical delivery of the underlying asset at maturity. Therefore we do not include the liquidation cost at maturity of the replicating portfolio in these replicating conditions. This assumption is consistent with the common contract specification for most exchange-traded European options on stocks.

$\varphi : G \to \mathbb{R}$. Now, for any $\alpha \in (0,1)$, and $u \in C([0,T] \times \mathbb{R})$, let

$$\begin{cases} \|u\|_0 \equiv |u|_0 \stackrel{\Delta}{=} \sup_{(t,x)\in[0,T]\times\mathbf{R}} |u(t,x)|, \\ |u|_\alpha \stackrel{\Delta}{=} \sup\limits_{(t,x)\neq(s,y),|x-y|\leq 1} \dfrac{|u(t,x)-u(s,y)|}{|t-s|^{\frac{\alpha}{2}} + |x-y|^\alpha}, \\ \|u\|_{2m+\alpha} \stackrel{\Delta}{=} \sum\limits_{2i+j\leq 2m} |\partial_t^i \partial_x^j u|_0 + \sum\limits_{2i+j=2m} |\partial_t^i \partial_x^j u|_\alpha. \end{cases} \tag{9}$$

Similarly, for any $\varphi \in C(\mathbb{R})$

$$\begin{cases} |\varphi|_0 \stackrel{\Delta}{=} \sup_{x\in\mathbf{R}} |\varphi(x)|, \\ |\varphi|_\alpha \stackrel{\Delta}{=} \sup\limits_{0<|x-y|\leq 1} \dfrac{|\varphi(x)-\varphi(y)|}{|x-y|^\alpha}, \\ \|\varphi\|_{m+\alpha} \stackrel{\Delta}{=} \sum\limits_{i\leq m} |\partial_x^i \varphi|_0 + |\partial_x^m \varphi|_\alpha. \end{cases} \tag{10}$$

Let

$$\begin{cases} C^{\frac{m+\alpha}{2},2m+\alpha}([0,T]\times\mathbb{R}) \stackrel{\Delta}{=} \{u \in C([0,T]\times\mathbb{R}) \mid \ \|u\|_{2m+\alpha} < \infty\}, \\ C^{m+\alpha}(\mathbb{R}) \stackrel{\Delta}{=} \{\varphi \in C(\mathbb{R}) \mid \ \|\varphi\|_{m+\alpha} < \infty\}. \end{cases} \tag{11}$$

In what follows, we adopt the following assumptions.

(A1) Functions $\widetilde{r}, \widetilde{\mu}, \widetilde{\sigma}, \widetilde{\delta} : [0,T]\times\mathbb{R} \to \mathbb{R}$ are in $C^{1+\frac{\alpha}{2},2+\alpha}([0,T]\times\mathbb{R})$ for some $\alpha \in (0,1)$, and there exists a constant $\delta_0 > 0$, such that

$$\widetilde{\sigma}(t,x)^2 \geq \delta_0 > 0, \qquad \forall (t,x) \in [0,T]\times\mathbb{R}. \tag{12}$$

(A2) Function $\widetilde{h} : \mathbb{R} \to \mathbb{R}$ is Lipschitz continuous and $e^{-\beta\langle\,\cdot\,\rangle}\widetilde{h}(\cdot)$ is bounded for some $\beta \geq 0$, where $\langle\, x \,\rangle = \sqrt{1+x^2}$.

The following theorem shows that under certain conditions an option is still replicable even in the presence of price impact and provides a generalized Black-Scholes pricing PDE for computing the replicating costs.

Theorem 1 *Suppose (A1) and (A2) hold for some $\alpha \in (0,1)$ and $\beta \geq 0$. Then there exists a constant $\varepsilon_0 > 0$ such that for any $\widetilde{\lambda} \in C^{\frac{\alpha}{2},\alpha}([0,T]\times\mathbb{R})$ with*

$$\|\widetilde{\lambda}(\cdot\,,\cdot)e^{\beta\langle\,\cdot\,\rangle}\|_\alpha \leq \varepsilon_0 \tag{13}$$

there exists a unique classical solution $f(\cdot\,,\cdot)$ of the following generalized Black-Scholes pricing PDE:

$$\begin{cases} f_t + \dfrac{\sigma(t,S)^2 S^2 f_{SS}}{2[1-\lambda(t,S)Sf_{SS}]^2} + (r(t,S)-\delta(t,S))Sf_S - r(t,S)f = 0, \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad (t,S) \in [0,T)\times(0,\infty), \\ f(T,S) = h(S). \end{cases} \tag{14}$$

Moreover, for any $\widetilde{\lambda} \in C^{\frac{1+\alpha}{2}, 2+\alpha}([0,T] \times \mathbb{R})$ *with*

$$\|\widetilde{\lambda}(\cdot\,,\cdot)e^{\beta\langle\,\cdot\,\rangle}\|_{2+\alpha} \leq \varepsilon_0, \tag{15}$$

FBSDE (7) admits a unique adapted solution $(S(\cdot), W(\cdot), N(\cdot))$ *such that*

$$\begin{cases} W(t) = f(t,S), \\ N(t) = f_S(t,S), \end{cases} \tag{16}$$

and $S(\cdot)$ *satisfies*

$$\frac{dS(t)}{S(t)} = \widehat{\mu}(t,S(t))dt + \widehat{\sigma}(t,S(t))dB(t), \qquad t \geq 0, \tag{17}$$

where

$$\begin{cases} \widehat{\mu}(t,S) \equiv \dfrac{\mu(t,S) + \lambda(t,S)f_{St}}{1-\lambda(t,S)Sf_{SS}} + \dfrac{\lambda(t,S)\sigma(t,S)^2S^2f_{SSS}}{2\Big(1-\lambda(t,S)Sf_{SS}\Big)^3}, \\ \widehat{\sigma}(t,S) \equiv \dfrac{\sigma(t,S)}{1-\lambda(t,S)Sf_{SS}}. \end{cases} \tag{18}$$

Proof. Here we provide a heuristic derivation of the generalized Black-Scholes PDE (14) and we leave the rest of the proof to the Appendix.

Suppose $(S(\cdot), W(\cdot), N(\cdot))$ is an adapted solution of FBSDE (7) and

$$W(t) = f(t,S(t)), \qquad t \in [0,T], \text{ a.s. }, \tag{19}$$

for some smooth function $f(\cdot\,,\cdot)$. Applying Itô's formula to (19) and using (7), we obtain (suppressing the arguments (S,t) for simplicity)

$$\begin{aligned} &[rW + (\mu+\delta-r+\lambda\eta)NS]dt + [\sigma+\lambda\zeta]NSdB = dW \\ =\ &\{f_t + Sf_S(\mu+\lambda\eta) + \tfrac{1}{2}S^2f_{SS}(\sigma+\lambda\zeta)^2\}dt + Sf_S(\sigma+\lambda\zeta)dB. \end{aligned} \tag{20}$$

Comparing the diffusion terms in the above equation, we see that one should choose

$$N(t) = f_S(S(t),t), \quad t \in [0,T] \text{ a.s.} \tag{21}$$

Then comparing the drift terms in (20) and using (19) and (21), one has

$$\begin{aligned} 0 &= f_t + Sf_S(\mu+\lambda\eta) + \tfrac{1}{2}S^2f_{SS}(\sigma+\lambda\zeta)^2 - [rf + (\mu+\delta-r+\lambda\eta)Sf_S] \\ &= f_t + \tfrac{1}{2}S^2f_{SS}(\sigma+\lambda\zeta)^2 + (r-\delta)Sf_S - rf. \end{aligned} \tag{22}$$

We hope to obtain an equation in $f(\cdot\,,\cdot)$. Thus, we need to eliminate ζ in the above equation. To this end, let us first note that

$$\begin{aligned} \eta dt + \zeta dB = dN = d[f_S] = [&f_{St} + Sf_{SS}(\mu+\lambda\eta) + \tfrac{1}{2}S^2f_{SSS}(\sigma+\lambda\zeta)^2]dt \\ &+Sf_{SS}(\sigma+\lambda\zeta)dB. \end{aligned} \tag{23}$$

Hence, comparing the diffusion terms, we obtain

$$\zeta = (\sigma + \lambda\zeta) S f_{SS}, \tag{24}$$

which implies (assuming that $\lambda S f_{SS} \neq 1$)

$$\zeta = \frac{\sigma S f_{SS}}{1 - \lambda S f_{SS}}. \tag{25}$$

Thus, the volatility of the stock (noting (7)) is given by

$$\widehat{\sigma} \equiv \sigma + \lambda\zeta = \frac{\sigma}{1 - \lambda S f_{SS}}. \tag{26}$$

Combining (22) and (26), we see that one should choose $f(\cdot\,,\cdot)$ to be a solution of the PDE (14) in Theorem 1.

On the other hand, comparing the drift terms in (23) we have (suppressing arguments (t, S))

$$\eta = f_{St} + S f_{SS}(\mu + \lambda\eta) + \tfrac{1}{2} S^2 f_{SSS}(\sigma + \lambda\zeta)^2, \tag{27}$$

which implies

$$\eta = \frac{1}{1 - \lambda S f_{SS}} \Big\{ f_{St} + \mu S f_{SS} + \frac{\sigma^2 S^2 f_{SSS}}{2(1 - \lambda S f_{SS})^2} \Big\}. \tag{28}$$

Hence, the instantaneous expected return of the stock is given by (noting (7) and (28))

$$\widehat{\mu} \equiv \mu + \lambda\eta = \frac{\mu + \lambda f_{St}}{1 - \lambda S f_{SS}} + \frac{\lambda \sigma^2 S^2 f_{SSS}}{2(1 - \lambda S f_{SS})^3}. \tag{29}$$

□

This theorem suggests that, to replicate an option, one has to first trade a discrete $f_S(0, S(0))$ shares of the stock and then follow a continuous trading strategy prescribed by $N(t)$. Given the stock price $S(0_-)$ before this discrete trade and (2), the stock price $S(0)$ the initial trade will drive to can be calculated as follows. Let $N(0_-) = 0$ and $N(0) = f_S(0, S(0))$. Assuming the trader can still work the initial order even when she trades a discrete amount,[3] by (2), we have

$$dS = \lambda(t, S) S dN(t), \tag{30}$$

for a discrete trade. This implies that $S(0)$ solves

$$\int_{S(0_-)}^{S(0)} \frac{dS}{\lambda(0, S) S} = f_S(0, S(0)) \tag{31}$$

[3]Alternatively, one can assume that all $f_S(0, S(0))$ shares are traded at $S(0)$. This would only increase the replicating cost, magnify the effect of price impact, and thus strengthen the main results in this paper.

and the initial cost of acquiring $f_S(0, S(0))$ shares is thus

$$c = \int_{S(0_-)}^{S(0)} \frac{dS}{\lambda(0, S)}. \tag{32}$$

The initial cost of replicating the option $h(S(T))$ is therefore

$$f^h(S(0_-)) = f(0, S(0)) - S(0) f_S(0, S(0)) + c. \tag{33}$$

The pricing PDE (14) implies that the effect of the price impact on the replication after the initial trade is only through a trade's impact on the stock volatility. This suggests, in particular, that if a trader can only affect the expected return of a stock but not the volatility or the interest rate, then the replicating cost for the trader will be the same as that for the case without price impact.

In the presence of price impact, a special form of put-call parity still holds when the primary market is not perfectly liquid. To see this, suppose the stock does not pay any dividend (i.e., $\delta(t, S) = 0$) for simplicity. Let $f_p(S)$ solve the pricing PDE (14) for a European put with strike price K and $f_c(S)$ solve (14) for the otherwise identical call. Then it is straightforward to verify that $f_p(S) + S - Ke^{-r(T-t)}$ also solves (14) for the otherwise identical call. Therefore, by the uniqueness of the solution, we must have

$$f_c(S) = f_p(S) + S - Ke^{-r(T-t)}. \tag{34}$$

Since in addition we also have that the delta of a call minus the delta of a put is equal to one, the put-call parity still holds even after taking into consideration the initial block trade price impact. This is because the number of shares a trader needs to buy to replicate a position consisting of a share and a put (i.e., a protective put position) is the same as the number of shares she needs to buy for replicating a call. In other words, using the notation in (33) let f^c be the replicating cost of a call and f^{p+S} be the replicating cost of a position consisting of a put and a share after incorporating the initial extra cost. Then we have the following special form of the put-call parity:

$$f^c(S(0_-)) = f^{p+S}(S(0_-)) - Ke^{-r(T-t)}. \tag{35}$$

Since $f^{p+S}(S(0_-)) \neq f^p(S(0_-)) + S(0_-)$ because of the price impact of the initial trades, we have that in the presence of price impact,

$$f^c(S(0_-)) \neq f^p(S(0_-)) + S(0_-) - Ke^{-r(T-t)},$$

i.e., at time $t = 0_-$,

$$\textit{Call price} \neq \textit{Put price} + \textit{Stock price} - Ke^{-r(T-t)}.$$

Therefore the implied volatility (using Black-Scholes formula) from the call price $f^c(S(0_-))$ is *different* from the implied volatility from the put price $f^p(S(0_-))$ (see Figures 9 and 10).[4]

As is well known, in the presence of transaction costs, superreplicating an option (buying a share and never trading again to super-replicate a call, for example) is cheaper than replicating. This is because exact replication involves continuous trading and thus incurs infinite costs (Liu and Loewenstein (2002), Liu (2004)). Similar to the transaction cost case, with price impact for every round trip trade an option trader also incurs additional costs. This may suggest that superreplicating might also be less expensive than exact replicating in the presence of price impact. However, the following theorem shows that this is not the case.

Theorem 2 *Let $h(\cdot)$ and $\overline{h}(\cdot)$ be such that*

$$h(S) < \overline{h}(S), \qquad S \in (0, \infty). \tag{36}$$

Suppose $f(\cdot,\cdot)$ and $\overline{f}(\cdot,\cdot)$ satisfy the following:

$$\begin{cases} f_t + \dfrac{\sigma(t,S)^2 S^2 f_{SS}}{2[1-\lambda(t,S)Sf_{SS}]^2} + (r(t,S)-\delta(t,S))Sf_S - r(t,S)f = 0, \\ \qquad\qquad\qquad\qquad\qquad\qquad (t,S) \in [0,T) \times (0,\infty), \\ f(T,S) = h(S) \end{cases} \tag{37}$$

and

$$\begin{cases} \overline{f}_t + \dfrac{\sigma(t,S)^2 S^2 \overline{f}_{SS}}{2[1-\lambda(t,S)S\overline{f}_{SS}]^2} + (r(t,S)-\delta(t,S))S\overline{f}_S - r(t,S)\overline{f} = 0, \\ \qquad\qquad\qquad\qquad\qquad\qquad (t,S) \in [0,T) \times (0,\infty), \\ \overline{f}(T,S) = \overline{h}(S). \end{cases} \tag{38}$$

Then under the same conditions as in Theorem 1, the cost of the superreplicating strategy $\bar{f}(t,S)$ is greater than the cost of the replicating strategy $f(t,S)$. In addition, ignoring the cost from the initial trading necessary for replication would understate the replication cost.

Proof. See Appendix. □

This theorem shows that the replicating strategy described above is indeed the cheapest way to hedge a European option for the trader. This result suggests that the excess cost incurred from the adverse price impact is of a lower order than that from transaction cost. Furthermore, the existing literature (e.g., Frey (2000)) ignores the extra cost from the initial trading for acquiring the initial position N_0 that is necessary for replication. By Theorem 2, one would then understate the replication cost from this omission.

[4] The reason we call equation (35) a put-call parity is that as the usual put-call parity in the absence of price impact, equation (35) states that *Call price* = *Protective Put price* − $Ke^{-r(T-t)}$.

IV. Analysis of the Replication of a European Option

In this section, we provide a numerical analysis of the effect of the illiquidity in the underlying asset market on the replication of a European option.

In all the subsequent analysis, we focus on the Black-Scholes economy; i.e., all the price coefficients are constant. Unless otherwise stated, we assume the default parameter values to be as follows: the current stock price $S = \$50$, the strike price $K = \$50$, the interest rate $r = 6\%$, the dividend yield $\delta = 0$, the volatility $\sigma = 40\%$, and the time to maturity $T = 0.25$ (3 months). Consistent with the price impact form obtained in Kyle (1985), Back (1993), and Vayanos (2001), and the one used in Bertsimas and Lo (1998), we assume

$$\lambda(t, S) = \begin{cases} \frac{\gamma}{S}(1 - e^{-\beta(T-t)}) & \text{if } \underline{S} \leq S \leq \bar{S} \\ 0 & \text{otherwise,} \end{cases}$$

where the constant price impact coefficient $\gamma > 0$ measures the price impact per traded share, and $\underline{S}$ and $\bar{S}$ represent respectively the lower and upper limit of the stock price within which there is a price impact. This form assumes that as a trader buys, the stock price goes up and as she sells, the stock price goes down. In addition, the magnitude of the impact is proportional to the number of shares traded within a certain range. Outside this range, the price impact is zero. The linearity is required by the absence of arbitrage as shown by Huberman and Stanzl (2001). Moreover, we assume that as time passes, the private information about the asset value is gradually revealed so that the price impact gradually decreases to zero at maturity, which also prevents any stock price manipulation at maturity. It can be easily shown that one can find a smooth function that approximates this price impact function arbitrarily well, also satisfies the regularity conditions in Theorem 1, and thus guarantees the existence of a unique solution of the pricing PDE (14). According to Sharpe, Alexander, and Bailey (1999), a trade size of 2000 shares (the smallest block trade size listed in the book) results in \$0.04 price impact per share for a stock with an average price of \$48. We thus set the default value of γ to be 0.04 accordingly. In addition, we set $\underline{S} = 20$ and $\bar{S} = 80$, and $\beta = 100$ for the subsequent numerical analysis.[5]

A. Replicating strategy

It is easily verified that if $h(S) = S$ then $f(t, S) = S$. Therefore, in the presence of price impact, to replicate the payoff of a share at a future time, one also has to purchase a share at time 0 and hold it, the same way as in the case of no price impact.

[5]Choosing a different value for $\gamma, \beta, \bar{S}$, or $\underline{S}$ will change the magnitude of the subsequent results (as suggested by Figures 5 and 6). However, the main qualitative results remain valid.

However, the cost will be higher. Specifically, the cost is c as defined in (31) and (32) instead of $S(0_-)$. A similar result applies to other options whose payoffs are linear in the stock price, e.g., forward and futures contracts. From now on we will focus on the exchange traded European options.

The first question we would like to address is how an option trader trades the underlying asset to replicate an option in the presence of price impact. To this end, we plot the time 0 difference (after taking into account the initial price impact) between the delta of her call replicating portfolio and the corresponding Black-Scholes delta against the stock price $S(0_-)$ in Figure 1 and show the corresponding put case in Figure 2. These figures show that the trader generally buys (or shorts) more stock compared to the Black-Scholes strategy. In addition, the difference in the deltas peaks when an option is close to the money and gradually decreases as the option becomes more away from the money. Intuitively, if an option ends in the money at maturity, to replicate the option one needs to long or short one share of the stock at maturity. If the option ends out of the money at maturity, then one should not have any stock when the option matures. Thus, when an option is in the money, an increase in the volatility would decrease the absolute value of the delta because this increases the chance of the option ending out of the money. When an option is out of the money, an increase in the volatility would instead increase the absolute value of the delta because this increases the chance of the option maturing in the money. This is called the volatility effect on delta. As suggested by Theorem 1, one of the main effects of the price impact is to increase the volatility of the stock return. Therefore the large investor effectively faces a larger volatility than in the Black-Scholes world. The volatility effect then implies that she trades more to replicate an out-of-the-money option and trades less to replicate an in-the-money option. As the time to horizon decreases, it becomes more and more certain how much she needs to have in the stock in order to replicate the option. Therefore for away-from-the-money options, her trading strategy is less and less different from the case without price impact. On the other hand, as stock price increases, the call delta increases and as stock price decreases, the absolute value of the put delta increases. This is called the price effect on delta. As the trader buys, the stock price increases and as she sells, the stock price decreases. Therefore, the price effect always makes the trader buy or short more compared to the Black-Scholes strategy. This implies that for out-of-the-money options both the volatility effect and the price effect increase the absolute value of the delta. For in-the-money options, however, these two effects act in opposite directions. The net effect depends on which one dominates. Figures 1 and 2 show that the price effect always dominates the volatility effect. This makes her always buy or short more. A more detailed analysis shows that the price effect is mainly from the price impact of the initial block trade. Ignoring this initial block trade, one would conclude that a large trader trades less for the in-the-money options; i.e., the volatility effect dominates the price effect. When an option is very deep in the money or very deep out of the money, the delta is relatively insensitive to the change in the volatility and the change in the stock price. So both the volatility effect and the price effect are small for these options, which implies that she will trade about the same way as in the case without any price impact.

Figure 3 plots the difference between the amount borrowed by the trader to replicate a call and the corresponding amount in the Black-Scholes strategy at time 0. Figure 4 plots the difference in the amount lent for replicating a put. Consistent with Figure 1, these figures show that the trader generally also borrows more (for a call) or lends more (for a put) due to the adverse price impact.

B. Replicating costs

Figure 5 plots the replicating costs against the price impact coefficient γ. The middle thin line is the Black-Scholes price. As the price impact coefficient increases, the replicating cost of a long call increases almost linearly and the replicating revenue of a short call decreases almost linearly. Therefore, there is a spread around the Black-Scholes price and the spread increases as the price impact coefficient increases. The linearity of the replicating cost in the price impact coefficient reflects the linear price impact form assumed. The effect of the price impact on the replicating cost (revenue) is significant even when the price impact is very small. For example, with $\gamma = 0.04$, the trader has about 0.08% price impact on the stock price. In contrast, the replicating cost is about \$0.014 higher than the Black-Scholes price, which amounts to about 0.32% of the Black-Scholes price, almost four times the impact on the stock price in the percentage term.

Figure 6 plots the replicating cost of a put against the price impact coefficient γ. The effect is similar to that for a call except that the wedge is now narrower. This is due to the fact that the payoff of a put is bounded above by the strike price.

Without the price impact, there is a unique no-arbitrage price for an option.[6] In contrast, in the presence of the adverse price impact, there is a continuum of no-arbitrage prices. In particular, we have the following straightforward implication of the replicability of an option and the adverse price impact. Let $f^{\alpha h} \geq 0$ be the replicating cost of $\alpha \geq 0$ units of an option $h(S(T)) \geq 0$ and $-f^{-\alpha h} \geq 0$ be the revenue from replicating α units of the option $-h(S(T)) \leq 0$. Then there is no arbitrage for trading α units of the option $h(S(T))$ if and only if the bid price for α units of the option is smaller than $f^{\alpha h}$ and the ask price for α units is greater than $-f^{-\alpha h}$. In particular, in the absence of bid-ask spread, this implies that a price between $f^{\alpha h}$ and $-f^{-\alpha h}$ is arbitrage free for trading α units of the option. The wedges depicted in Figures 5 and 6 represent the case with $\alpha = 1$. For example, with $\gamma = 0.04$, the bid price for one call must be below \$4.350 and the ask price for one call must be

[6] Here an arbitrage means that one can make a positive profit without any risk or initial investment. The profit made is not required to be infinite.

above \$4.324. This implies any price between \$4.324 and \$4.350 prevents arbitrage from trading one call in the presence of price impact. In contrast, in the absence of price impact and bid-ask spread, the call price has to be exactly \$4.336.

Because of the nonlinearity in the dynamics of the replicating portfolio value as shown in (3), the cost of replicating α units of an option is not equal to α times the cost of replicating one unit of the option. To help us understand this nonlinearity, we plot the average replicating cost of a long position and the average revenue from replicating a short position as functions of the units of options traded (α) in Figure 7 for calls.[7] This figure shows that the *average* replicating cost and revenue are almost linear in the number of units to be replicated, which implies that the excess replicating cost is approximately *quadratic* in the number of units α. In addition, as the number of units decreases, the average replicating cost and revenue decrease and converge to the Black-Scholes price of the option. This suggests that in the presence of the nonlinearity, the absence of arbitrage for trading a certain number of units of an option does not imply the absence of arbitrage for trading a smaller number of units of the option. Therefore, in the presence of price impact, to prevent arbitrage when trading any number of units is allowed, the option price must be set at the Black-Scholes price in the absence of the bid-ask spread.

C. Excess cost and the smile

Figure 8 plots the excess replicating costs above the corresponding Black-Scholes price for a call as a function of the strike price. As the option becomes more and more out of the money, the excess cost decreases and converges monotonically to zero. As the option becomes more and more in the money, the excess cost converges to the excess cost of buying one share. Intuitively, as the option gets more and more in the money, the investor needs to buy more and more of the underlying asset and eventually, when the option is far in the money, the trader has to buy one share almost surely to replicate the call. On the other hand, as the option gets more and more out of the money, the trader needs to buy less and less stock and eventually, when the option is far out of the money, the investor does not need to buy any share almost surely.

Next, similar to Grossman and Zhou (1996) and Platen and Schweizer (1998), we consider the relationship between illiquidity and the volatility smile. The excess cost pattern shown in the above figures implies that the implied volatility would not be constant across options with different moneyness (K/S). To this end, we assume that an option trader sells an option at the replicating cost of a long position in the option. We can then plug the replicating cost of one unit of the long option computed from our model with price impact into the Black-Scholes formula to back out the implied volatility of the option. Figure 9 plots the implied volatility for

[7] The corresponding results of Figures 7 and 8 for puts are very similar and thus omitted.

calls against moneyness K/S. It shows that even though the volatility of the stock return without price impact is constant, the implied volatility for a call is no longer constant and changes with moneyness in the presence of price impact. In particular, the implied volatility increases as the option gets more and more in the money. This is consistent with the volatility smile for calls widely documented in the literature (see Dumas, Fleming, and Whaley (1998), for example). Intuitively, as shown in Figure 8, as the call gets more and more in the money, the excess cost increases and therefore the implied volatility also increases. Similarly, as shown in Figure 10, the implied volatility increases as the put becomes more and more in the money. However, compared to the implied volatility for a call, the implied volatility for the otherwise identical put is much smaller for the same range of the moneyness. This is in contrast to the implied volatility pattern documented for puts (see Dumas, Fleming, and Whaley (1998), for example).[8] These findings have important empirical implications for explaining the volatility smile. In particular, it suggests that the negative correlation between the stock price and the volatility (as assumed in Heston (1993) and Bates (1996), for example) that was required to generate the smile would have to be weaker for calls, but stronger for puts, if the price impact were taken into account.

V. Concluding Remarks

In this paper, we investigate how the imperfect liquidity in the underlying asset market affects the replication of a European option. In a market with imperfect liquidity, trading affects stock price. We obtain a generalized nonlinear Black-Scholes pricing partial differential equation. We derive sufficient conditions for the existence and uniqueness of a classical solution. These are also sufficient conditions for the replicability of the option. We also show that in contrast to the case with transaction costs, superreplication in the presence of adverse price impact is more costly than exact replication.

We find that compared to the Black-Scholes strategy, a trader in an illiquid underlying asset market generally needs to buy (short) more stock and to borrow (lend) more to replicate a call (put). The excess replicating cost is approximately quadratic in the number of units of options to be replicated. The excess cost a trader incurs is found to be significant even with small price impact. We also show that although a special form of put-call parity (equation (1)) still holds, the implied volatility for a put is different from the implied volatility for the otherwise identical call.

In contrast to most of the existing literature on large traders, this model allows a direct price impact. This provides a reasonable model also for the pricing of a block

[8]The different implied volatility for a call and the otherwise identical put is due to the presence of price impact for the initial block trade. See equation (35) and the discussions therein.

order. An interesting problem would be to estimate the price impact functions for illiquid assets. This way one can then empirically test the implications of this model, such as comparing the model's implied excess costs over the Black-Scholes prices to the observed excess costs. An equilibrium model with informed traders trading in both options and stocks would shed light on the form and magnitude of the price impact. Another interesting issue is to analyze the optimal liquidation strategy for a fund that has significant adverse price impact.

Appendix

In this appendix, we present detailed proofs of Theorems 1 and 2.

Proof of Theorem 1. Suppose $f(\cdot\,,\cdot)$ is a classical solution of (14) satisfying the following regularity condition:

Condition 1. $f(t,S)$ is $C^{1,4}$ in (t,S) and $S \mapsto S\widehat{\mu}(t,S)$ and $S \mapsto S\widehat{\sigma}(t,S)$ are Lipschitz continuous in S, uniform in $t \in [0,T]$, and

$$\begin{aligned} 1 - \lambda(t,S) S f_{SS}(t,S) \geq \delta_0 > 0, \\ \forall (t,S) \in [0,T] \times [0,\infty). \end{aligned} \tag{39}$$

Let $S(\cdot)$ be the (unique) strong solution of (17), and let $(W(\cdot), N(\cdot), \eta(\cdot), \zeta(\cdot))$ be defined by (16), (27), and (25). Applying Itô's formula to $W(\cdot)$ and $N(\cdot)$, using (20)–(29), we see that $(W(\cdot), S(\cdot), N(\cdot))$ is an adapted solution to (7).

Next we want to show that under some regularity conditions, Condition 1 is satisfied. The proof of this part of Theorem 1 is much more technical. Given (A2), one can find a sequence of smooth functions $h_k(\cdot)$ that converge to $h(\cdot)$ in $H^{1,\infty}_{loc}(\mathbb{R})$ (i.e., the set of all locally Lipschitz continuous functions). Now, for each $h_k(\cdot)$ the corresponding PDE (14) admits a unique classical solution $f_k(\cdot)$. By a priori estimate, one can prove that $f_k(\cdot)$ is equicontinuous. Thus, one can assume that $f_k(\cdot)$ is convergent to $f(\cdot)$ (in $H^{1,p}_{loc}(\mathbb{R})$ for any $p \in [1,\infty)$), which satisfies the equation (14) in $[0,T) \times (0,\infty)$ and the terminal condition in the sense that

$$\lim_{t \to T} f(t,S) = h(S).$$

Therefore without loss of generality, we can assume $h(\cdot)$ satisfies

$$e^{-\beta \langle \cdot \rangle} \widetilde{h}(\cdot) \in C^{4+\alpha}(\mathbb{R}), \tag{40}$$

for some $\beta \geq 0$, and the same $\alpha \in (0,1)$ as in (A1), where $\langle x \rangle = \sqrt{1+x^2}$.

By (8), (14) can be written as

$$\begin{cases} \widetilde{f}_t + \dfrac{\widetilde{\sigma}^2[\widetilde{f}_{xx} - \widetilde{f}_x]}{2[1 - \widetilde{\lambda}(\widetilde{f}_{xx} - \widetilde{f}_x)]^2} + (\widetilde{r} - \widetilde{\delta})\widetilde{f}_x - \widetilde{r}\widetilde{f} = 0, \quad (t,x) \in [0,T) \times \mathbb{R}, \\ \widetilde{u}\big|_{t=T} = \widetilde{h}. \end{cases} \tag{41}$$

It is clear that the solvability of (14) is equivalent to that of (41).

We will now show the following lemma.

Lemma A.1. *Suppose (A1) and (A2) hold for some $\alpha \in (0,1)$ and $\beta \geq 0$. Then there exists a constant $\varepsilon_0 > 0$ such that for any $\widetilde{\lambda} \in C^{\frac{\alpha}{2},\alpha}([0,T] \times \mathbb{R})$ with*

$$\|\widetilde{\lambda}(\cdot\,,\cdot) e^{\beta \langle \cdot \rangle}\|_\alpha \leq \varepsilon_0, \tag{42}$$

the generalized Black-Scholes pricing PDE (41) admits a unique classical solution $\widetilde{f}(t,x)$ satisfying

$$e^{-\beta\langle\cdot\rangle}\widetilde{f}(\cdot\,,\cdot)\in C^{1+\frac{\alpha}{2},2+\alpha}([0,T]\times\mathbb{R}). \tag{43}$$

Moreover, for any $\widetilde{\lambda}\in C^{\frac{1+\alpha}{2},2+\alpha}([0,T]\times\mathbb{R})$ with

$$\|\widetilde{\lambda}(\cdot\,,\cdot)e^{\beta\langle\cdot\rangle}\|_{2+\alpha}\le\varepsilon_0, \tag{44}$$

the solution $\widetilde{f}(\cdot\,,\cdot)$ has all the properties required in Theorem 1.

First, we recall some classical results for linear second order parabolic PDEs. For later convenience, we modify the statements to fit our framework. Consider the following terminal value problem:

$$\begin{cases} v_t+a(t,x)v_{xx}+b(t,x)v_x+c(t,x)v=\varphi(t,x), \\ v|_{t=T}=\psi(x). \end{cases} \tag{45}$$

We introduce the following assumptions.

(P) Functions $a(\cdot\,,\cdot)$, $b(\cdot\,,\cdot)$, and $c(\cdot\,,\cdot)$ are all in $C^{\frac{\alpha}{2},\alpha}([0,T]\times\mathbb{R})$ for some $\alpha\in(0,1)$. There exists a constant $\delta_0>0$ such that

$$a(t,x)\ge\delta_0,\qquad\forall(t,x)\in[0,T]\times\mathbb{R}. \tag{46}$$

According to Ladyženskaja et al. (1968), we have the following result.

Proposition A.1. *Let (P) hold. Then for any $\varphi\in C^{\frac{\alpha}{2},\alpha}([0,T]\times\mathbb{R})$ and $\psi\in C^{2+\alpha}(\mathbb{R})$, Cauchy problem (4.4) admits a unique solution $v\in C^{1,\frac{\alpha}{2},2+\alpha}([0,T]\times\mathbb{R})$. Moreover, there exists a constant $K>0$ only depending on δ_0, $\|a\|_\alpha$, $\|b\|_\alpha$, and $\|c\|_\alpha$ such that*

$$\|v\|_{2+\alpha}\le K(\|\varphi\|_\alpha+\|\psi\|_{2+\alpha}). \tag{47}$$

If we write $K\equiv K(\|a\|_\alpha+\|b\|_\alpha+\|c\|_\alpha,\delta_0^{-1})$, then we can assume that $K(\cdot\,,\cdot)$ is nondecreasing in both arguments. Now, we will apply the above result to our equation.

Suppose $\widetilde{f}(\cdot\,,\cdot)$ is a solution of the following:

$$\begin{cases} \widetilde{f}_t(t,x)+\dfrac{\widetilde{\sigma}(t,x)^2[\widetilde{f}_{xx}(t,x)-\widetilde{f}_x(t,x)]}{2\{1-\widetilde{\lambda}(t,x)[\widetilde{f}_{xx}(t,x)-\widetilde{f}_x(t,x)]\}^2} \\ \qquad+[\widetilde{r}(t,x)-\widetilde{\delta}(t,x)]\widetilde{f}_x(t,x)-\widetilde{r}(t,x)\widetilde{f}(t,x)=0. \\ \widetilde{f}(T,x)=\widetilde{h}(x). \end{cases} \tag{48}$$

Define

$$v(t,x)=e^{-\beta\langle x\rangle}\widetilde{f}(t,x),\qquad(t,x)\in[0,T]\times\mathbb{R}. \tag{49}$$

Then

$$
\begin{cases}
\widetilde{f}_t(t,x) = e^{\beta\langle x\rangle} v_t(t,x), \\
\widetilde{f}_x(t,x) = e^{\beta\langle x\rangle}\Big[\dfrac{\beta x}{\langle x\rangle} v(t,x) + v_x(t,x)\Big] \equiv e^{\beta\langle x\rangle}[Bv(t,x)+v_x(t,x)], \\
\widetilde{f}_{xx}(t,x) = e^{\beta\langle x\rangle}\Big[\big(\dfrac{\beta^2x^2}{\langle x\rangle^2} + \dfrac{\beta}{\langle x\rangle^3}\big)v(t,x) + \dfrac{2\beta x}{\langle x\rangle} v_x(t,x) + v_{xx}(t,x)\Big] \\
\qquad\qquad = e^{\beta\langle x\rangle}[Av(t,x) + 2Bv_x(t,x) + v_{xx}(t,x)],
\end{cases}
\tag{50}
$$

where

$$
A \triangleq \frac{\beta^2x^2}{\langle x\rangle^2} + \frac{\beta}{\langle x\rangle^3}, \quad B \triangleq \frac{\beta x}{\langle x\rangle}. \tag{51}
$$

Let

$$
F(x,v,v_x,v_{xx}) \triangleq (A-B)v + (2B-1)v_x + v_{xx}. \tag{52}
$$

Then by (48)–(52), we see that $v(\cdot\,,\cdot)$ satisfies the following equation (for notational simplicity, we suppress (t,x) in $\widetilde{r}(t,x)$, $\widetilde{\sigma}(t,x)$, $\widetilde{\lambda}(t,x)$, $v(t,x)$, $v_x(t,x)$, and $v_{xx}(t,x)$ and suppress (x,v,v_x,v_{xx}) in $F(x,v,v_x,v_{xx})$):

$$
\begin{cases}
v_t + \dfrac{\widetilde{\sigma}^2}{2(1-\widetilde{\lambda}e^{\beta\langle x\rangle}F)^2} v_{xx} + \Big[\dfrac{(2B-1)\widetilde{\sigma}^2}{2(1-\widetilde{\lambda}e^{\beta\langle x\rangle}F)^2} + \widetilde{r} - \widetilde{\delta}\Big]v_x \\
\quad + \Big[\dfrac{(A-B)\widetilde{\sigma}^2}{2(1-\widetilde{\lambda}e^{\beta\langle x\rangle}F)^2} + B(\widetilde{r}-\widetilde{\delta}) - \widetilde{r}\Big]v = 0, \\
\\
v(T,x) = e^{-\beta\langle x\rangle}\widetilde{h}(x).
\end{cases}
\tag{53}
$$

We now want to establish the existence and uniqueness of a classical solution to (53). To this end, let us present some simple lemmas first.

Lemma A.2. *Let φ and ψ be proper functions of $(t,x) \in [0,T]\times\mathbb{R}$. Then the following hold provided the involved expressions make sense.*

$$
\begin{cases}
\|\varphi+\psi\|_\alpha \le \|\varphi\|_\alpha + \|\psi\|_\alpha, \\
\|\varphi\psi\|_\alpha \le \|\varphi\|_\alpha \|\psi\|_\alpha, \\
\Big\|\dfrac{1}{\varphi}\Big\|_\alpha \le \dfrac{\|\varphi\|_\alpha}{\min\limits_{(t,x)\in[0,T]\times\mathbf{R}} |\varphi(x)|^2}.
\end{cases}
\tag{54}
$$

By Lemma A.2 and (52), we have (provided all the terms involved are meaningful)

$$
\|F\|_\alpha \le \|(A-B)v\|_\alpha + \|(2B-1)v_x\|_\alpha + \|v_{xx}\|_\alpha \le C_0\|v\|_{2+\alpha}. \tag{55}
$$

Here, C_0 is a constant only depending on β. Consequently, we have

$$
\begin{aligned}
\Big\|\frac{\widetilde{\sigma}^2}{2(1-\widetilde{\lambda}e^{\beta\langle x\rangle}F)^2}\Big\|_\alpha &\le \frac{\|\widetilde{\sigma}\|_\alpha^2}{2}\,\frac{1+\|\widetilde{\lambda}e^{\beta\langle x\rangle}\|_\alpha^2\|F\|_\alpha^2}{[1-\|\widetilde{\lambda}e^{\beta\langle x\rangle}\|_\alpha^2\|F\|_\alpha^2]^2} \\
&\le \frac{\|\widetilde{\sigma}\|_\alpha^2}{2}\,\frac{1+C_0^2\|\widetilde{\lambda}e^{\beta\langle x\rangle}\|_\alpha^2\|v\|_{2+\alpha}^2}{[1-C_0^2\|\widetilde{\lambda}e^{\beta\langle x\rangle}\|_\alpha^2\|v\|_{2+\alpha}^2]^2},
\end{aligned}
\tag{56}
$$

$$\Big\|\frac{(2B-1)\widetilde{\sigma}^2}{2(1-\widetilde{\lambda}e^{\beta\langle x\rangle}F)^2}+\widetilde{r}-\widetilde{\delta}\Big\|_\alpha \le C\Big\|\frac{\widetilde{\sigma}^2}{2(1-\widetilde{\lambda}e^{\beta\langle x\rangle}F)^2}\Big\|_\alpha+\|\widetilde{r}\|_\alpha+\|\widetilde{\delta}\|_\alpha$$
$$\le C\|\widetilde{\sigma}\|_\alpha^2\frac{1+C_0^2\|\widetilde{\lambda}e^{\beta\langle x\rangle}\|_\alpha^2\|v\|_{2+\alpha}^2}{[1-C_0^2\|\widetilde{\lambda}e^{\beta\langle x\rangle}\|_\alpha^2\|v\|_{2+\alpha}^2]^2}+\|\widetilde{r}\|_\alpha+\|\widetilde{\delta}\|_\alpha, \tag{57}$$

$$\Big\|\frac{(A-B)\widetilde{\sigma}^2}{2(1-\widetilde{\lambda}e^{\beta\langle x\rangle}F)^2}+B(\widetilde{r}-\widetilde{\delta})-\widetilde{r}\Big\|_\alpha$$
$$\le C\Big\{\|\widetilde{\sigma}\|_\alpha^2\frac{1+C_0^2\|\widetilde{\lambda}e^{\beta\langle x\rangle}\|_\alpha^2\|v\|_{2+\alpha}^2}{[1-C_0^2\|\widetilde{\lambda}e^{\beta\langle x\rangle}\|_\alpha^2\|v\|_{2+\alpha}^2]^2}+\|\widetilde{r}\|_\alpha+\|\widetilde{\delta}\|_\alpha\Big\}. \tag{58}$$

Hereafter, C represents a generic constant that can be different at different places. Also, we have

$$\frac{\widetilde{\sigma}^2}{2(1-\widetilde{\lambda}e^{\beta\langle x\rangle}F)^2}\ge\frac{\delta_0^2}{2(1-|\widetilde{\lambda}e^{\beta\langle x\rangle}F|)^2}$$
$$\ge\frac{\delta_0^2}{2(1-C_0\|\widetilde{\lambda}e^{\beta\langle x\rangle}\|_\alpha\|v\|_{2+\alpha})^2}. \tag{59}$$

Now, we fix any $v\in C^{\frac{2+\alpha}{2},2+\alpha}([0,T]\times\mathbb{R})$ with

$$C_0\|\widetilde{\lambda}e^{\beta\langle x\rangle}\|_\alpha\|v\|_{2+\alpha}\le\frac{1}{2}. \tag{60}$$

With this v in $F\equiv F(t,x,v,v_x,v_{xx})$, it follows from (59) that

$$\frac{\widetilde{\sigma}^2}{2(1-\widetilde{\lambda}e^{\beta\langle x\rangle}F)^2}\ge\frac{\delta_0^2}{2(1-C_0\|\widetilde{\lambda}e^{\beta\langle x\rangle}\|_\alpha\|v\|_{2+\alpha})^2}\ge 2\delta_0^2\stackrel{\Delta}{=}\delta_1>0, \tag{61}$$

and it follows from (56)–(58) and (60) that

$$\Big\|\frac{\widetilde{\sigma}^2}{2(1-\widetilde{\lambda}e^{\beta\langle x\rangle}F)^2}\Big\|_\alpha+\Big\|\frac{(2B-1)\widetilde{\sigma}^2}{2(1-\widetilde{\lambda}e^{\beta\langle x\rangle}F)^2}+\widetilde{r}-\widetilde{\delta}\Big\|_\alpha$$
$$+\Big\|\frac{(A-B)\widetilde{\sigma}^2}{2(1-\widetilde{\lambda}e^{\beta\langle x\rangle}F)^2}+B(\widetilde{r}-\widetilde{\delta})-\widetilde{r}\Big\|_\alpha$$
$$\le C\Big\{\|\widetilde{\sigma}\|_\alpha^2\frac{1+C_0^2\|\widetilde{\lambda}e^{\beta\langle x\rangle}\|_\alpha^2\|v\|_{2+\alpha}^2}{[1-C_0^2\|\widetilde{\lambda}e^{\beta\langle x\rangle}\|_\alpha^2\|v\|_{2+\alpha}^2]^2}+\|\widetilde{r}\|_\alpha+\|\widetilde{\delta}\|_\alpha\Big\}$$
$$\le C(\|\widetilde{\sigma}\|_\alpha^2+\|\widetilde{r}\|_\alpha+\|\widetilde{\delta}\|_\alpha)\stackrel{\Delta}{=}A_0. \tag{62}$$

We see that A_0 and δ_1 are independent of v satisfying (60).

Next, by Proposition A.1, for such a v, there exists a unique classical solution, denoted by $\overline{v}$, to the following linear PDE:

$$\left\{\begin{array}{l} \overline{v}_t + \dfrac{\widetilde{\sigma}^2}{2(1-\widetilde{\lambda}e^{\beta\langle x\rangle}F)^2}\overline{v}_{xx} + \Big[\dfrac{(2B-1)\widetilde{\sigma}^2}{2(1-\widetilde{\lambda}e^{\beta\langle x\rangle}F)^2} + \widetilde{r} - \widetilde{\delta}\Big]\overline{v}_x \\ \quad +\Big[\dfrac{(A-B)\widetilde{\sigma}^2}{2(1-\widetilde{\lambda}e^{\beta\langle x\rangle}F)^2} + B(\widetilde{r}-\widetilde{\delta}) - \widetilde{r}\Big]\overline{v} = 0, \\ \overline{v}(T,x) = e^{-\beta\langle x\rangle}\widetilde{h}(x), \end{array}\right. \tag{63}$$

where $F = F(t,x,v,v_x,v_{xx})$ is given by (52) with the fixed v. Moreover, the following estimate holds:

$$\|\overline{v}\|_{2+\alpha} \le K(A_0,\delta_1^{-1})\|e^{-\beta\langle x\rangle}\widetilde{h}\|_{2+\alpha}. \tag{64}$$

Thus, we obtain a map $v \mapsto \overline{v} \triangleq S(v)$ from $C^{\frac{2+\alpha}{2},2+\alpha}([0,T]\times\mathbb{R})$ to itself. Next, we would like to show that for a suitable $\widetilde{\lambda}$, map S admits a unique fixed point. To this end, we take any $v,w \in C^{\frac{2+\alpha}{2},2+\alpha}([0,T]\times\mathbb{R})$ satisfying (60). Let $\overline{v} = S(v)$ and $\overline{w} = S(w)$. Denote $\overline{\xi} = \overline{v} - \overline{w}$, $\xi = v - w$, and $G \triangleq F(t,x,w,w_x,w_{xx})$. Then $\overline{\xi}$ satisfies the following PDE:

$$\left\{\begin{array}{l} \overline{\xi}_t + \dfrac{\widetilde{\sigma}^2}{2(1-\widetilde{\lambda}e^{\beta\langle x\rangle}F)^2}\overline{\xi}_{xx} + \Big[\dfrac{(2B-1)\widetilde{\sigma}^2}{2(1-\widetilde{\lambda}e^{\beta\langle x\rangle}F)^2} + \widetilde{r} - \widetilde{\delta}\Big]\overline{\xi}_x \\ \quad +\Big[\dfrac{(A-B)\widetilde{\sigma}^2}{2(1-\widetilde{\lambda}e^{\beta\langle x\rangle}F)^2} + B(\widetilde{r}-\widetilde{\delta}) - \widetilde{r}\Big]\overline{\xi} \\ \quad +\Big[\dfrac{\widetilde{\sigma}^2}{2(1-\widetilde{\lambda}e^{\beta\langle x\rangle}F)^2} - \dfrac{\widetilde{\sigma}^2}{2(1-\widetilde{\lambda}e^{\beta\langle x\rangle}G)^2}\Big] \\ \quad\quad \cdot\Big[\overline{w}_{xx} + (2B-1)\overline{w}_x + (A-B)\overline{w}\Big] = 0, \\ \overline{\xi}(T,x) = 0. \end{array}\right. \tag{65}$$

By Lemma A.2, we have

$$\begin{array}{l} \Big\|\dfrac{1}{(1-\widetilde{\lambda}e^{\beta\langle x\rangle}F)^2} - \dfrac{1}{(1-\widetilde{\lambda}e^{\beta\langle x\rangle}G)^2}\Big\|_\alpha \\ = \Big\|\dfrac{(1-\widetilde{\lambda}e^{\beta\langle x\rangle}G)^2 - (1-\widetilde{\lambda}e^{\beta\langle x\rangle}F)^2}{(1-\widetilde{\lambda}e^{\beta\langle x\rangle}F)^2(1-\widetilde{\lambda}e^{\beta\langle x\rangle}G)^2}\Big\|_\alpha \\ \le \dfrac{\|2-\widetilde{\lambda}e^{\beta\langle x\rangle}(F+G)\|_\alpha\|\widetilde{\lambda}e^{\beta\langle x\rangle}(F-G)\|_\alpha\|(1-\widetilde{\lambda}e^{\beta\langle x\rangle}F)^2(1-\widetilde{\lambda}e^{\beta\langle x\rangle}G)^2\|_\alpha}{(1-\|\widetilde{\lambda}e^{\beta\langle x\rangle}F\|_\alpha)^4(1-\|\widetilde{\lambda}e^{\beta\langle x\rangle}G\|_\alpha)^4}. \end{array} \tag{66}$$

Note that

$$\begin{array}{l} \|2-\widetilde{\lambda}e^{\beta\langle x\rangle}(F+G)\|_\alpha\|(1-\widetilde{\lambda}e^{\beta\langle x\rangle}F)^2(1-\widetilde{\lambda}e^{\beta\langle x\rangle}G)^2\|_\alpha \\ \le [2+\|\widetilde{\lambda}e^{\beta\langle x\rangle}\|_\alpha C_0(\|v\|_{2+\alpha}+\|w\|_{2+\alpha})] \\ \quad \cdot(1+\|\widetilde{\lambda}e^{\beta\langle x\rangle}\|_\alpha C_0\|v\|_{2+\alpha})^2(1+\|\widetilde{\lambda}e^{\beta\langle x\rangle}\|_\alpha C_0\|w\|_{2+\alpha})^2 \le C. \end{array} \tag{67}$$

On the other hand,

$$(1-\|\widetilde{\lambda}e^{\beta\langle x\rangle}F\|_\alpha)^4(1-\|\widetilde{\lambda}e^{\beta\langle x\rangle}G\|_\alpha)^4\geq\frac{1}{256}.$$

Hence (comparing (55)),

$$\begin{aligned}\Big\|\frac{1}{(1-\widetilde{\lambda}e^{\beta\langle x\rangle}F)^2}-\frac{1}{(1-\widetilde{\lambda}e^{\beta\langle x\rangle}G)^2}\Big\|_\alpha&\leq C\|\widetilde{\lambda}e^{\beta\langle x\rangle}\|_\alpha\|F-G\|_\alpha\\&\leq C\|\widetilde{\lambda}e^{\beta\langle x\rangle}\|_\alpha\|\xi\|_{2+\alpha}.\end{aligned}\tag{68}$$

Now, applying (47) to the solution ξ of (65), we obtain

$$\begin{aligned}\|\overline{\xi}\|_{2+\alpha}\leq K(A_0,\delta_1^{-1})\Big\|&\Big[\frac{\widetilde{\sigma}^2}{2(1-\widetilde{\lambda}e^{\beta\langle x\rangle}F)^2}-\frac{\widetilde{\sigma}^2}{2(1-\widetilde{\lambda}e^{\beta\langle x\rangle}G)^2}\Big]\\&\cdot\Big[\overline{w}_{xx}+(2B-1)\overline{w}_x+(A-B)\overline{w}\Big]\Big\|_{2+\alpha}\\\leq C\|\overline{w}\|_{2+\alpha}&\|\widetilde{\lambda}e^{\beta\langle x\rangle}\|_\alpha\|\xi\|_{2+\alpha}\\\leq C\|e^{-\beta\langle x\rangle}&\widetilde{h}\|_{2+\alpha}\|\widetilde{\lambda}e^{\beta\langle x\rangle}\|_{2+\alpha}\|\xi\|_{2+\alpha}\leq C\|\widetilde{\lambda}e^{\beta\langle x\rangle}\|_{2+\alpha}\|\xi\|_{2+\alpha}.\end{aligned}\tag{69}$$

Here, $C>0$ is an absolute constant. By choosing, say, $\varepsilon_0=\frac{1}{2(C+1)}$, we see that when (13) is satisfied, we have from (69) that

$$\|S(v)-S(w)\|_{2+\alpha}\leq\frac{1}{2}\|v-w\|_{2+\alpha}.\tag{70}$$

Therefore, S admits a unique fixed point in the set of all functions $v(\cdot\,,\cdot)$ satisfying (60). This means that (53) admits a classical solution v. Also, by equation (53), one sees that

$$\|v_t\|_\alpha<\infty.\tag{71}$$

Next, we would like to establish some further regularity of solution v to (53). To this end, for notational simplicity, we define

$$\Gamma\triangleq\frac{\widetilde{\sigma}^2}{2(1-\widetilde{\lambda}e^{\beta\langle x\rangle}F)^2}.\tag{72}$$

Then (53) can be written as (recall (51) for definitions of A and B)

$$\begin{cases}v_t+\Gamma v_{xx}+\Big[(2B-1)\Gamma+\widetilde{r}-\widetilde{\delta}\Big]v_x+\Big[(A-B)\Gamma+B(\widetilde{r}-\widetilde{\delta})-\widetilde{r}\Big]v=0,\\v(T,x)=e^{-\beta\langle x\rangle}\widetilde{h}(x).\end{cases}\tag{73}$$

Differentiating the above in x once, we have

$$\begin{cases}(v_x)_t+\Gamma(v_x)_{xx}+\{\Gamma_x+(2B-1)\Gamma+\widetilde{r}-\widetilde{\delta})\}(v_x)_x\\\qquad+\{(2B-1)\Gamma_x+(2B_x+A-B)\Gamma+\widetilde{r}_x-\widetilde{\delta}_x+B(\widetilde{r}-\widetilde{\delta})-\widetilde{r}\}v_x\\\qquad+\{(A-B)\Gamma_x+(A_x-B_x)\Gamma+[B(\widetilde{r}-\widetilde{\delta})-\widetilde{r}]_x\}v=0,\\v_x(T,x)=[e^{-\beta\langle x\rangle}\widetilde{h}(x)]_x.\end{cases}\tag{74}$$

We note that

$$\Gamma_x = \frac{\widetilde{\sigma}\widetilde{\sigma}_x}{(1-\widetilde{\lambda}e^{\beta\langle x\rangle}F)^2} + \frac{\widetilde{\sigma}^2[(\widetilde{\lambda}e^{\beta\langle x\rangle})_x F + \widetilde{\lambda}e^{\beta\langle x\rangle}F_x]}{(1-\widetilde{\lambda}e^{\beta\langle x\rangle}F)^3}, \tag{75}$$

$$F_x = v_{xxx} + (2B-1)v_{xx} + (2B_x + A - B)v_x + (A_x - B_x)v. \tag{76}$$

Thus, only the following in (74) contains the term v_{xxx}:

$$\begin{aligned} &\Gamma v_{xxx} + \Gamma_x[v_{xx} + (2B-1)v_x + (A-B)v] = \Gamma v_{xxx} + \Gamma_x F \\ &= \Big\{\frac{\widetilde{\sigma}^2}{2(1-\widetilde{\lambda}e^{\beta\langle x\rangle}F)^2} + \frac{\widetilde{\sigma}^2\widetilde{\lambda}e^{\beta\langle x\rangle}F}{(1-\widetilde{\lambda}e^{\beta\langle x\rangle}F)^3}\Big\}v_{xxx} + R(v,v_x,v_{xx}) \\ &= \frac{\widetilde{\sigma}^2[1+\widetilde{\lambda}e^{\beta\langle x\rangle}F]}{2(1-\widetilde{\lambda}e^{\beta\langle x\rangle}F)^3}v_{xxx} + R(v,v_x,v_{xx}). \end{aligned} \tag{77}$$

Then similar to the above proof, with a more careful analysis, we obtain that when ε_0 is small enough, (74) admits a unique classical solution that is nothing but v_x in the space $C^{1+\frac{\alpha}{2},2+\alpha}([0,T]\times\mathbb{R})$, such that

$$(\|v_{xt}\|_\alpha + \|v_x\|_{2+\alpha}) < \infty. \tag{78}$$

Further, by differentiating (74) once more, we obtain

$$\begin{cases} (v_{xx})_t + \Gamma(v_{xx})_{xx} + \Big\{2\Gamma_x + (2B-1)\Gamma + \widetilde{r} - \widetilde{\delta}\Big\}(v_{xx})_x \\ \qquad +\Big\{\Gamma_{xx} + 2[(2B-1)\Gamma + \widetilde{r} - \widetilde{\delta}]_x + (A-B)\Gamma + B(\widetilde{r}-\widetilde{\delta}) - \widetilde{r}\Big\}v_{xx} \\ \qquad +\Big\{[(2B-1)\Gamma + \widetilde{r} - \widetilde{\delta}]_{xx} + 2[(A-B)\Gamma + B(\widetilde{r}-\widetilde{\delta}) - \widetilde{r}]_x\Big\}v_x \\ \qquad +[(A-B)\Gamma + B(\widetilde{r}-\widetilde{\delta}) - \widetilde{r}]_{xx}v = 0, \\ v_{xx}(T,x) = [e^{-\beta\langle x\rangle}\widetilde{h}(x)]_{xx}. \end{cases} \tag{79}$$

By a similar argument as above (with much longer computation and much careful analysis), we are able to prove that the above has a classical solution, provided ε_0 in the statement of Theorem 2 is small enough, and the solution coincides with v_{xx}. Therefore, we have

$$\|v_{xxt}\|_\alpha + \|v_{xx}\|_{2+\alpha} < \infty. \tag{80}$$

The above, in turn, implies that (41) admits a classical solution $\widetilde{f}(\cdot\,,\cdot)$ such that the following estimate holds (with some simple computation):

$$\|e^{-\beta\langle\cdot\rangle}\{|\widetilde{f}| + |\widetilde{f}_x| + |\widetilde{f}_t| + |\widetilde{f}_{xx}| + |\widetilde{f}_{xt}| + |\widetilde{f}_{xxx}| + |\widetilde{f}_{xxt}| + |\widetilde{f}_{xxxx}|\}\|_\alpha \le C. \tag{81}$$

Moreover, by (15), we have

$$\|(\widetilde{\lambda} + |\widetilde{\lambda}_x|)\{|\widetilde{f}_{xx}| + |\widetilde{f}_{xt}| + |\widetilde{f}_{xxx}| + |\widetilde{f}_{xxt}| + |\widetilde{f}_{xxxx}|\}\|_\alpha \le C. \tag{82}$$

Now, if we define $\widehat{\mu}(t,S)$ and $\widehat{\sigma}(t,t)$ by (18), then with $S=e^x$, we have

$$\left\{\begin{aligned} \widehat{\mu}(t,S) &= \frac{\widetilde{\mu}(t,x)+\widetilde{\lambda}(t,x)\widetilde{f}_{xt}(t,x)}{1-\widetilde{\lambda}(t,x)[\widetilde{f}_{xx}(t,x)-\widetilde{f}_x(t,x)]} \\ &\quad + \frac{\widetilde{\lambda}(t,x)\widetilde{\sigma}(t,x)^2[\widetilde{f}_{xxx}(t,x)-3\widetilde{f}_{xx}(t,x)+2\widetilde{f}_x(t,x)]}{\{1-\widetilde{\lambda}(t,x)[\widetilde{f}_{xx}(t,x)-\widetilde{f}_x(t,x)]\}^3} \\ \widehat{\sigma}(t,S) &= \frac{\widetilde{\sigma}(t,x)}{1-\widetilde{\lambda}(t,x)[\widetilde{f}_{xx}(t,x)-\widetilde{f}_x(t,x)]}. \end{aligned}\right. \tag{83}$$

Hence, noting

$$\left\{\begin{array}{l} [S\widehat{\mu}(t,S)]_S = \widehat{\mu}(t,S)+S\widehat{\mu}_S(t,S) = \widehat{\mu}(t,e^x)+[\widehat{\mu}(t,e^x)]_x, \\ [S\widehat{\sigma}(t,S)]_S = \widehat{\sigma}(t,S)+S\widehat{\sigma}_S(t,S) = \widehat{\sigma}(t,e^x)+[\widehat{\sigma}(t,e^x)]_x, \end{array}\right. \tag{84}$$

and (82), together with the conditions for $\widetilde{\mu}(\cdot)$ and $\widetilde{\sigma}(\cdot)$, we see that the maps $S \mapsto S\widehat{\mu}(t,S)$ and $S \mapsto S\widehat{\sigma}(t,S)$ are Lipschitz continuous. This completes the proof of Theorem 1. $\square$

We now present a proof of a more general version of Theorem 2.

Proof of Theorem 2. According to the proof of Theorem 1, we see under the regularity condition that if

1. (A1) and (15) hold,
2. $h(\cdot)$ and $\overline{h}(\cdot)$ satisfy (A2),

we have unique solutions v and $\overline{v}$ of (53) corresponding to $h(\cdot)$ and $\overline{h}(\cdot)$, respectively. Moreover, if we let

$$\left\{\begin{array}{l} F = v_x x + (2B-1)v_x + (A-B)v \\ \overline{F} = \overline{v}_{xx} + (2B-1)\overline{v}_x + (A-B)\overline{v}, \end{array}\right. \tag{85}$$

then

$$\|\widetilde{\lambda}e^{\beta\langle\cdot\rangle}F\|_\alpha,\ \|\widetilde{\lambda}e^{\beta\langle\cdot\rangle}\overline{F}\|_\alpha < 1. \tag{86}$$

Now, $\xi \overset{\Delta}{=} v-\overline{v}$ satisfies the following equation:

$$\left\{\begin{aligned} &\xi_t + \frac{\widetilde{\sigma}^2}{2(1-\widetilde{\lambda}e^{\beta\langle x\rangle}F)^2}\xi_{xx} + \Big[\frac{(2B-1)\widetilde{\sigma}^2}{2(1-\widetilde{\lambda}e^{\beta\langle x\rangle}F)^2}+\widetilde{r}-\widetilde{\delta}\Big]\xi_x \\ &\quad +\Big[\frac{(A-B)\widetilde{\sigma}^2}{2(1-\lambda e^{\beta\langle x\rangle}F)^2}+B(\widetilde{r}-\widetilde{\delta})-\widetilde{r}\Big]\xi \\ &\quad +\Big[\frac{\widetilde{\sigma}^2}{2(1-\widetilde{\lambda}e^{\beta\langle x\rangle}F)^2}-\frac{\widetilde{\sigma}^2}{2(1-\widetilde{\lambda}e^{\beta\langle x\rangle}\overline{F})^2}\Big] \\ &\quad\ \cdot\Big[\overline{v}_{xx}+(2B-1)\overline{v}_x+(A-B)\overline{v}\Big] \geq 0, \\ &\xi(T,x) = [\widetilde{h}(x)-\widetilde{\overline{h}}(x)]e^{-\beta\langle x\rangle} \leq 0, \end{aligned}\right. \tag{87}$$

Note that

$$
\begin{aligned}
&(1-\widetilde{\lambda}e^{\beta\langle x\rangle}\overline{F})^2-(1-\widetilde{\lambda}e^{\beta\langle x\rangle}F)^2\\
&=[2-\widetilde{\lambda}e^{\beta\langle x\rangle}(F+\overline{F})]\widetilde{\lambda}e^{\beta\langle x\rangle}(F-\overline{F})\\
&=[2-\widetilde{\lambda}e^{\beta\langle x\rangle}(F+\overline{F})]\widetilde{\lambda}e^{\beta\langle x\rangle}[\xi_{xx}+(2B-1)\xi_x+(A-B)\xi].
\end{aligned}
\tag{88}
$$

Hence, we can write (87) as follows.

$$
\left\{
\begin{aligned}
&\xi_t+\frac{\widetilde{\sigma}^2\{(1-\widetilde{\lambda}e^{\beta\langle x\rangle}\overline{F})^2+\widetilde{\lambda}e^{\beta\langle x\rangle}[2-\widetilde{\lambda}e^{\beta\langle x\rangle}(F+\overline{F})]\}}{2(1-\widetilde{\lambda}e^{\beta\langle x\rangle}F)^2(1-\widetilde{\lambda}e^{\beta\langle x\rangle}\overline{F})^2}\xi_{xx}\\
&\quad+\Big\{\frac{(2B-1)\widetilde{\sigma}^2}{2(1-\widetilde{\lambda}e^{\beta\langle x\rangle}F)^2}+\widetilde{r}-\widetilde{\delta}\\
&\quad+\frac{\widetilde{\sigma}^2B\widetilde{\lambda}e^{\beta\langle x\rangle}[2-\widetilde{\lambda}e^{\beta\langle x\rangle}(F+\overline{F})]}{(1-\widetilde{\lambda}e^{\beta\langle x\rangle}F)^2(1-\widetilde{\lambda}e^{\beta\langle x\rangle}\overline{F})^2}\overline{F}\Big\}\xi_x\\
&\quad+\Big\{\frac{(A-B)\widetilde{\sigma}^2}{2(1-\widetilde{\lambda}e^{\beta\langle x\rangle}F)^2}+B(\widetilde{r}-\widetilde{\delta})-\widetilde{r}\\
&\quad+\frac{\widetilde{\sigma}^2A\widetilde{\lambda}e^{\beta\langle x\rangle}[2-\widetilde{\lambda}e^{\beta\langle x\rangle}(F+\overline{F})]}{2(1-\widetilde{\lambda}e^{\beta\langle x\rangle}F)^2(1-\widetilde{\lambda}e^{\beta\langle x\rangle}\overline{F})^2}\overline{F}\Big\}\xi\geq 0,\\
&\xi(T,x)=[\widetilde{h}(x)-\widetilde{\overline{h}}(x)]e^{-\beta\langle x\rangle}\leq 0,
\end{aligned}
\right.
\tag{89}
$$

Clearly, the coefficient of ξ_{xx} is positive. Thus, by the maximum principle for parabolic partial differential equations, we obtain

$$
f(t,S)\leq\overline{f}(t,S),\qquad (t,S)\in[0,T]\times(0,\infty).
\tag{90}
$$

If a strict equality holds in (36), it also holds in (90). The above is the case if, in particular, f and $\overline{f}$ are solutions to (14) corresponding to h and $\overline{h}$.

Let $S(0_-)$ be the stock price before the initial trade to replicate or superreplicate. Now motivated by (33) we define

$$
\bar{C}(z)=\bar{f}(0,z)-z\int_{S(0_-)}^{z}\frac{dS}{\lambda(0,S)S}+\int_{S(0_-)}^{z}\frac{dS}{\lambda(0,S)}
$$

and let $\bar{S}(0)$ be the stock price after buying (shorting) the initial number of shares as required by the superreplicating strategy; i.e., $\bar{S}(0)$ solves

$$
\int_{S(0_-)}^{\bar{S}(0)}\frac{dS}{\lambda(0,S)S}=\bar{f}_S(0,\bar{S}(0)).
$$

Then $\bar{C}(\bar{S}(0))$ is the time 0 cost function for the superreplicating strategy $\bar{f}(t,S)$. It is easy to verify that $\bar{C}'(\bar{S}(0))=0$ and $C(\cdot)$ is strictly concave by (39), which we have already proven before. Therefore

$$
\bar{C}(\bar{S}(0))=\max_z\bar{C}(z).
\tag{91}
$$

Next for the replicating strategy define

$$C(z) = f(0,z) - z\int_{S(0_-)}^{z} \frac{dS}{\lambda(0,S)S} + \int_{S(0_-)}^{z} \frac{dS}{\lambda(0,S)}$$

and let $S(0)$ be the stock price after buying (shorting) the initial number of shares as required by the replicating strategy; i.e., $S(0)$ solves (31). Then $C(S(0))$ is the time 0 cost function for the replicating strategy $f(t,S)$. By a similar argument to the one for the superreplicating strategy, we have that

$$C(S(0)) = \max_z C(z),$$

which in particular implies that $C(S(0)) > C(S(0_-))$; i.e., if one ignores the extra cost that resulted from the initial trade necessary for replication, then one would understate the replicating cost of an option.

Finally, by (90) and (91), we have that

$$C(S(0)) \leq \bar{C}(S(0)) < \bar{C}(\bar{S}(0));$$

i.e., the cost of the superreplicating strategy $\bar{f}(t,S)$ is greater than the cost of the replicating strategy $f(t,S)$. □

References

Allen, F., Gale D., 1992. Stock-price manipulation. Review of Financial Studies 5, 503-529.

Back, K., 1993. Asymmetric information and options. Review of Financial Studies 6, 435-472.

Bagnoli, M., Lipman, B.L., 1990. Stock price manipulation through takeover bids. The Rand Journal of Economics 27, 124-147.

Bates, D., 1996. Jumps and stochastic volatility: Exchange rate process implicit in Deutsche mark options. Review of Financial Studies 9, 69-107.

Bertsimas, D., Lo, A.W., 1998. Optimal control of execution costs. Journal of Financial Markets 1, 1-50.

Chan, L., Lakonishok, J., 1995. The behavior of stock prices around institutional trades. Journal of Finance 50, 1147-1174.

Cuoco, D., Cvitanić, J., 1998. Optimal consumption choices for a 'large' investor. Journal of Economic Dynamics and Control 22, 401-436.

Cuoco, D., Liu, H., 2000. A martingale characterization of consumption choices and hedging costs with margin requirements. Mathematical Finance 10, 355-385.

Cvitanić, J., Ma, J., 1996. Hedging options for a large investor and forward-backward SDEs. Annals of Applied Probability 6, 370-398.

Dumas, B., Fleming, J., Whaley, R.E., 1998. Implied volatility functions: Empirical tests. Journal of Finance 53, 2059-2106.

Frey, R., 2000. Market illiquidity as a source of model risk in dynamic hedging. Working paper.

Grossman, S., Zhou, Z., 1996. Equilibrium analysis of portfolio insurance. Journal of Finance 51, 1379-1403.

Heston, S., 1993. A closed-form solution for options with stochastic volatility with applications to bond and currency options. Review of Financial Studies 6, 327-343.

Huberman, G., Stanzl, W., 2001. Arbitrage-free price-update and price-impact functions. Working paper, Yale University.

Hull, J., White, A., 1987. The pricing of options on assets with stochastic volatility. Journal of Finance 42, 281-300.

Jarrow, R.A., 1992. Market manipulation, bubbles corners and short squeezes. Journal of Financial and Quantitative Analysis 27, 311-336.

Jorion, P., 2000. Value at Risk. McGraw-Hill, New York.

Keim, D., Madhavan, A., 1996. The upstairs market for large-block transactions: Analysis and measurement of price effects. Review of Financial Studies 9, 1-36.

Kyle, A., 1985. Continuous auctions and insider trading. Econometrica 53, 1315-1335.

Ladyženskaja, O.A., Solonnikov, V.A., Ural'ceva, N.N., 1968. Linear and Quasi-Linear Equations of Parabolic Type. AMS, Providence, RI.

Liu, H., 2004. Optimal consumption and investment with transaction costs and multiple risky assets. Journal of Finance 59, 269-318.

Liu, H., Loewenstein M., 2002. Optimal portfolio selection with transaction costs and finite horizons. Review of Financial Studies 15, 805-835.

Ma, J., Protter, P., Yong, J., 1994. Solving forward-backward stochastic differential equations explicitly – A four step scheme. Probability Theory and Related Fields 98, 339-359.

Ma, J., Yong, J., 1999, Forward Backward Stochastic Differential Equations and Their Applications. Springer-Verlag, New York.

Platen, E., Schweizer, W., 1998. On feedback effects from hedging derivatives. Mathematical Finance 8, 67-84.

Schönbucher, P.J., Wilmott, P., 2000. The feedback effect of hedging in illiquid markets. SIAM Journal of Applied Mathematics 61, 232-272.

Sharpe, W.F., Alexander, G.J., Bailey, J.V., 1999. Investments. Prentice Hall, New Jersey.

Sircar, K. R., Papanicolaou, G., 1998. General Black-Scholes models accounting for increased market volatility from hedging strategies. Applied Mathematical Finance 5, 45-82.

Vayanos, D., 2001. Strategic trading in a dynamic noisy market. Journal of Finance 56, 131-171.

Vila, J.L., 1989. Simple games of market manipulation. Economic Letters 29, 21-26.

Yong, J., 1999. European type contingent claims in an incomplete market with constrained wealth and portfolio. Mathematical Finance 9, 387-412.

Yong, J., Zhou X.Y., 1999. Stochastic Controls: Hamiltonian Systems and HJB Equations. Springer-Verlag, New York.

Figure Legends

Figure 1. The difference in call delta from Black-Scholes as a function of stock price at time 0.

The graph plots the difference in the delta of the call replicating portfolio from Black-Scholes against S for parameters $\gamma = 0.04$, $K = 50$, $r = 6\%$, $\delta = 0$, and $\sigma = 40\%$. The thickest line is for $T = 1$, the thick one is for $T = 0.0833$, and the thin one is for $T = 0.01$.

Figure 2. The difference in put delta from Black-Scholes as a function of stock price at time 0.

The graph plots the difference in the delta of the put replicating portfolio from Black-Scholes against S for parameters $\gamma = 0.04$, $K = 50$, $r = 6\%$, $\delta = 0$, and $\sigma = 40\%$. The thickest line is for $T = 1$, the thick one is for $T = 0.0833$, and the thin one is for $T = 0.01$.

Figure 3. The difference in amount borrowed from Black-Scholes as a function of stock price at time 0.

The graph plots the difference in the borrowed amount in the call replicating portfolio from Black-Scholes against S for parameters $\gamma = 0.04$, $K = 50$, $r = 6\%$, $\delta = 0$, and $\sigma = 40\%$. The thickest line is for $T = 1$, the thick one is for $T = 0.0833$, and the thin one is for $T = 0.01$.

Figure 4. The difference in amount lent from Black-Scholes as a function of stock price at time 0.

The graph plots the difference in the amount lent in the put replicating portfolio from Black-Scholes against S for parameters $\gamma = 0.04$, $K = 50$, $r = 6\%$, $\delta = 0$, and $\sigma = 40\%$. The thickest line is for $T = 1$, the thick one is for $T = 0.0833$, and the thin one is for $T = 0.01$.

Figure 5. The call replicating cash flow as functions of the price impact coefficient.

The graph plots the replicating cash flow of a call against γ for parameters $S = 50$, $K = 50$, $r = 6\%$, $\delta = 0$, $\sigma = 40\%$, and $T = 0.25$. The middle thin line is the Black-Scholes call price.

Figure 6. The put replicating cash flow as functions of the price impact coefficient.

The graph plots the replicating cash flow of a put against γ for parameters $S = 50$, $K = 50$, $r = 6\%$, $\delta = 0$, $\sigma = 40\%$, and $T = 0.25$. The middle thin line is the Black-Scholes put price.

Figure 7. The average call replicating cash flows as functions of the number of calls traded.

The graph plots the average call replicating cash flows as functions of the number of calls traded for parameters $\gamma = 0.04$, $S = 50$, $K = 50$, $r = 6\%$, $\delta = 0$, $\sigma = 40\%$, and $T = 0.25$. The middle thin line is the Black-Scholes call price.

Figure 8. Replicating cost of a call above Black-Scholes price as a function of the strike price.

The graph plots the replicating cost above Black-Scholes price for a call against the strike price K for parameters $\gamma = 0.04$, $S = 50$, $r = 6\%$, $\delta = 0$, $\sigma = 40\%$, and $T = 0.25$.

Figure 9. Implied volatility for a call as a function of moneyness.

The graph plots the implied volatility against K/S for parameters $\gamma = 0.04$, $S = 50$, $r = 6\%$, $\delta = 0$, $\sigma = 40\%$, and $T = 0.25$.

Figure 10. Implied volatility for a put as a function of moneyness.

The graph plots the implied volatility against K/S for parameters $\gamma = 0.04$, $S = 50$, $r = 6\%$, $\delta = 0$, $\sigma = 40\%$, and $T = 0.25$.

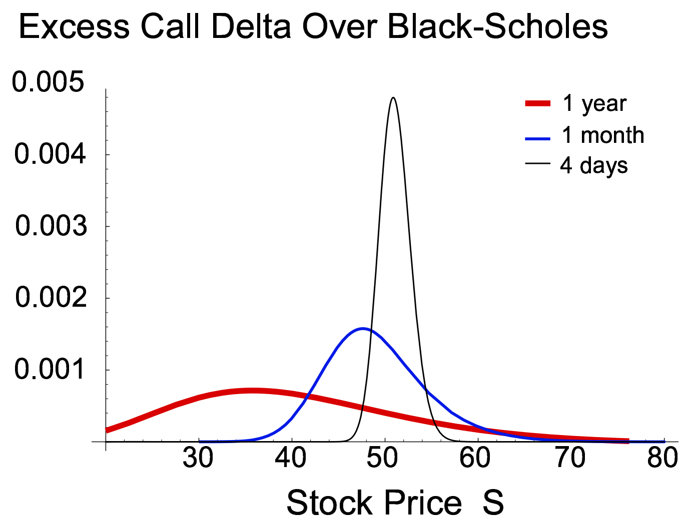

Figure 1: The difference in call delta from Black-Scholes as a function of stock price at time 0.

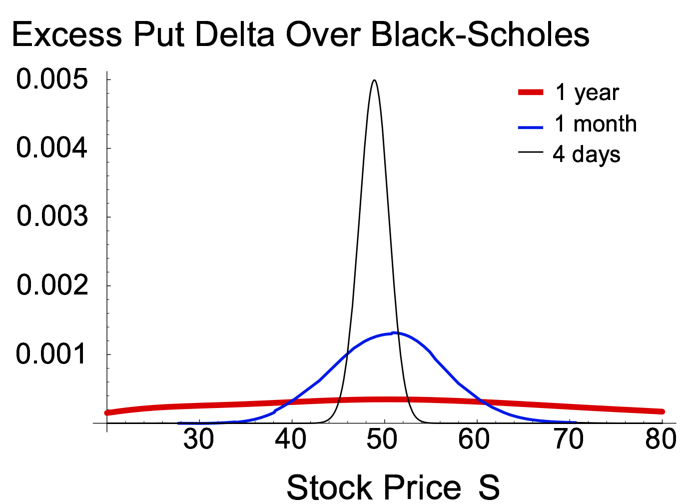

Figure 2: The difference in put delta from Black-Scholes as a function of stock price at time 0.

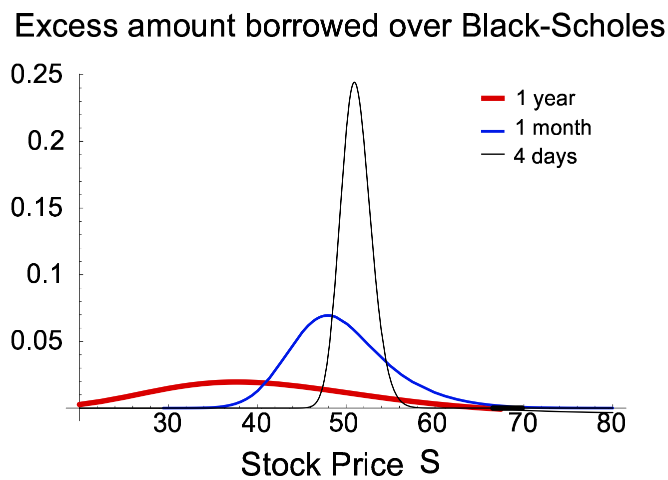

Figure 3: The difference in amount borrowed from Black-Scholes as a function of stock price at time 0.

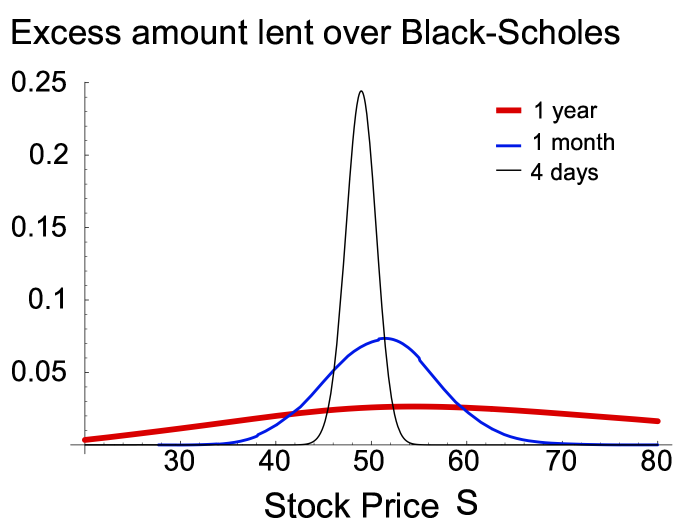

Figure 4: The difference in amount lent from Black-Scholes as a function of stock price at time 0.

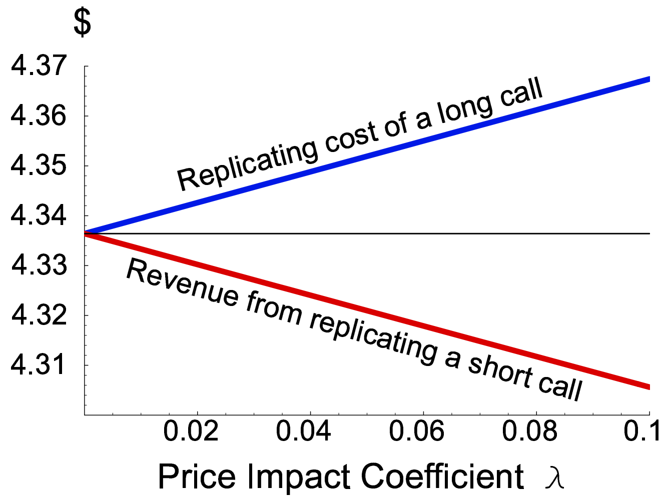

Figure 5: The call replicating cashflow as functions of the price impact coefficient.

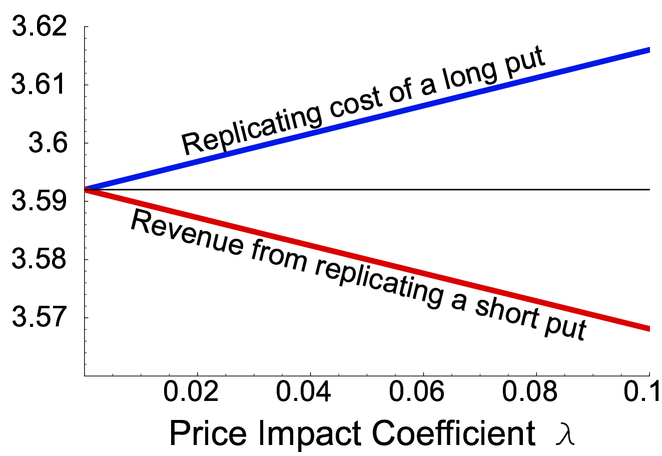

Figure 6: The put replicating cashflow as functions of the price impact coefficient.

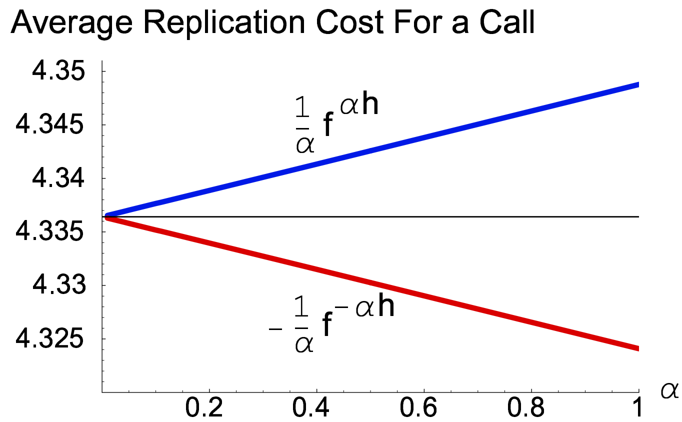

Figure 7: The average call replicating cashflows as functions of the number of calls traded.

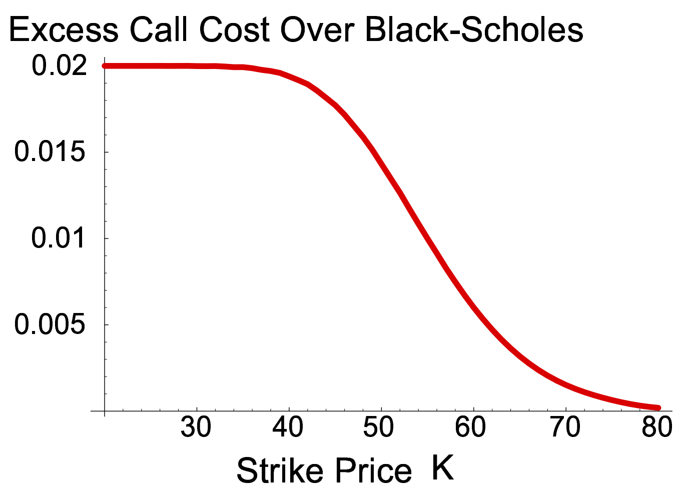

Figure 8: Replicating cost of a call above Black-Scholes price as a function of the strike price.

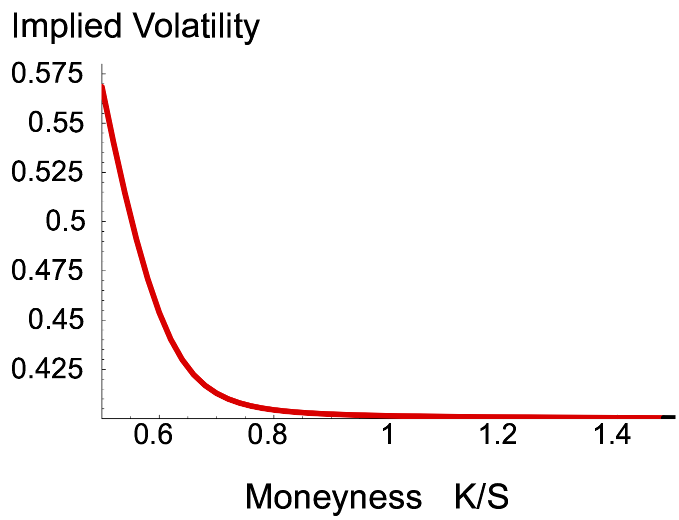

Figure 9: Implied volatility for a call as a function of moneyness.

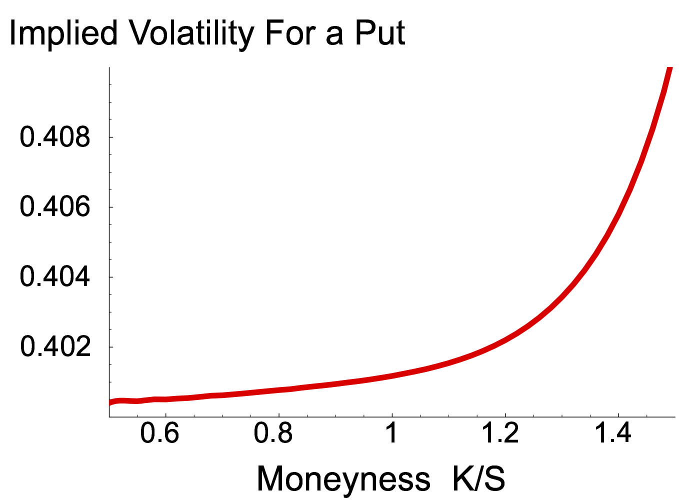

Figure 10: Implied volatility for a put as a function of moneyness.