A Note on Market Structure with Transaction Costs

著者	Yamazaki Akira, Yoshikawa Daisuke
出版者	法政大学イノベーション・マネジメント研究センター
journal or publication title	法政大学イノベーション・マネジメント研究センター ワーキングペーパーシリーズ
volume	191
page range	1-33
year	2018-04-02
URL	http://hdl.handle.net/10114/13776

WORKING PAPER SERIES

Akira Yamazaki and Daisuke Yoshikawa

A Note on Market Structure with Transaction Costs

April 2, 2018

No. 191

The Research Institute for Innovation Management, HOSEI UNIVERSITY

A note on market structure with transaction costs

Akira Yamazaki and Daisuke Yoshikawa*

March 22, 2018

Abstract

Transaction costs play a significant role in financial markets, and many studies have been conducted on this topic to date. Research on this topic may be divided into two categories. The first category of studies examines the optimal trading strategy of the investor who has to pay transaction costs. The second group investigates the optimal transaction costs that ensure the market operates as smoothly as possible, while retaining the profits of the market maker. We consider simultaneous optimization by the investors and the market maker, and analyse the impact of market parameters on the optimized transaction costs.

1 Introduction

Transaction costs play a significant role in financial markets, in the sense that they give agents, such as investment bankers, incentives to make a market. The activities of such agents may improve market liquidity. At the same time, the existence of transaction costs makes it difficult for investors to determine a unique, optimal trading strategy because of the lack of a unique no-arbitrage price; i.e., the difference in the sell and buy price resulting from transaction costs breaks the law of one price.

*Akira Yamazaki is at the Graduate School of Business Administration and the Research Institute for Innovation Management, Hosei University. Daisuke Yoshikawa is at the Department of Business Administration, Hokkai-Gakuen University. Yamazaki is supported by the financial support of JSPS KAKENHI Grant Number 26380402 and the Research Institute for Innovation Management at Hosei University. Yoshikawa is supported by a grant-in-aid from JSPS KAKENHI Grant Number 15K03546, Seimeikai Foundation and Kampo Foundation.

This paper focuses on deriving optimal transaction costs, in the sense of being optimal not only for a market maker, but also for an investor. To address this significant issue, we have separated our analysis into three parts.

First, we propose a new method to derive optimal transaction costs. Second, we provide an explicit form of the investor trading strategy in the model with transaction costs. Third, we illustrate the intuitive features of our model using numerical examples. We consider that this approach facilitates a comprehensive understanding of the role of transaction costs in the market. In this introductory section, we discuss the three components of our approach in more detail.

Let us consider the first point. To make a larger profit, the market maker's preference is that the transaction costs paid for each transaction are larger. However, if the transaction costs are too large, this decreases the amount of transactions, which may lead to a reduction in the market maker's profit. Further, this creates a lack of liquidity in the market. With insufficient liquidity, the market maker's inventory may not be cleared for a long time, which is an unfavourable situation for market makers. Therefore, the main purpose of the market maker is to maximize his/her profit, while also keeping the market running smoothly (providing liquidity) and minimizing inventory costs.

Discussions of this problem have a long history in the economic and financial literature; for example, we can date the earliest discussions back to Amihud and Mendelson (1980) and Bradfield (1982). Recently, the discussion has been reignited in the context of the progress of high-frequency trading (see Pham (2014)).

However, many studies establish an investors' order that is too simplistic. For instance, Bradfield (1982) defined the investors' order as the linear function of transaction costs. Chevalier et al. (2017) and Mildenstein and Schleef (1983) considered the investors' order as given by the point process, where the jump intensity depends on the bid–ask spread. Although these set-ups are interesting, they do not lead to interaction between the investors' order and the market maker's quote.

Thus, in this paper, we derive the optimal transaction costs that maximize the market maker's profit, taking into account the optimization of investors. This is our first contribution.

To derive the optimal transaction costs, we first need to derive the investor's trading strategy in the presence of transaction costs. There is a great deal of informative research on the problem of deriving the investor's optimal trading strategy. The earliest papers discussing this topic include Davis and Norman (1990) and Jouini (1995). Since these early papers, there has been a vigorous discussion of transaction costs conducted by many researchers. One of the turning points in the literature was the study by Kabanov and Last (2002), who suggested the concept of the transaction matrix. A longstanding issue in this field has been how to clearly describe transactions if investors hold several assets in markets with transaction costs, as these costs make the status of investors' wealth unclear. The transaction matrix makes it possible to describe transactions of multiple assets even if transaction costs are charged for each transaction.

Further important progress in the research on models with transaction costs was made when Bouchard (2002, 2000); Bouchard et al. (2001) developed a liquidity function that makes it simpler and easier to derive an optimal trading strategy for each investor even if the model includes transaction costs.

The results of both Kabanov and Bouchard have been extended by others. For instance, Schachermayer (2004) considered a model where transaction costs stochastically change and showed how the optimal trading strategy should be derived. This problem was later positively solved by Campi and Owen (2011) and Benedetti et al. (2013).

Another method for dealing with the optimization problem in models with transaction costs is to introduce the idea of shadow price, as developed by Kallsen and Muhle-Karbe (2010), Kallsen and Muhle-Karbe (2011), Guasoni and Muhle-Karbe (2013), Kallsen and Muhle-Karbe (2017), Gerhold et al. (2014) and others. The shadow price is a hypothetical price process that maintains consistency with the utility maximization problem of the investor. By regarding the shadow price process as the underlying asset process, we can deal with the optimization problem in models with transaction costs as a frictionless optimization problem. To derive the optimal trading strategy, we apply the results of Kallsen and Muhle-Karbe (2010) and Gerhold et al. (2014) to our problem. Under the assumption of the log-utility type of investor, Kallsen and Muhle-Karbe (2010) showed the existence of the shadow price process, that is, the existence of a solution to the utility maximization problem. We further manipulate the underlying asset as a discount price process. Then, we can derive the explicit form of the value function of the investor, which makes it very simple to deal with the investor's activity. This is our second contribution in this paper.

In addition, we illustrate the intuitive features of the model using numerical examples.

First, we show the consistency of our model with transaction costs with the frictionless model. That is, the model with infinitely small transaction costs is equivalent to the smooth market model. In addition to theoretically proving this consistency, we show it numerically.

Further, we show that a larger appreciation rate and lower volatility decreases the optimal transaction cost. This is intuitive because the investor's risky weight may imply their willingness to undertake transactions. Here, note that the risky weight of investors increases roughly with the larger appreciation rate and smaller volatility.

We also demonstrate the stability of the market maker's profit. If the market maker performs the optimization well, his/her profit should remain stable in the face of the market fluctuations, in this case, a change in the appreciation rate and volatility. Indeed, using the numerical examples, we can demonstrate that the market maker's profit is stable when the appreciation rate and the volatility change.

Finally, building on the existing empirical research, we demonstrate how to examine the validity of the model. Although using the numerical examples indicates that models such as ours with log-utility types of investors and a risk-neutral market maker have limitations, this examination clearly points to directions for making the model setting more sophisticated. This is our third contribution.

The remainder of the paper is structured as follows. We establish the market model and define the problem in Section 2. In Section 3, we provide the main results, including the optimal strategies of the market maker and the investor. Section 4 demonstrates the model's implications using numerical examples. The detailed procedures for deriving the main results are described in the Appendices.

2 Model

Let us define the behaviour of a single market maker, who quotes the bid and ask prices of a single stock.[1] For the market maker, the time for conducting business is finite. That is, let $(\Omega, \mathcal{F}, \mathbb{F}, P)$ be a filtered probability space, where we consider the time interval $[0, T]$ and fix $\mathcal{F}_T = \mathcal{F}$ and $\mathcal{F}_0 = \{\emptyset, \Omega\}$.

The prices offered to the investor are given as follows:

$$S_t^a := (1 + \epsilon) S_t$$
$$S_t^b := (1 - \epsilon) S_t,$$

where $S = \{S_t;\ t \in [0, T]\}$ is given as follows:[2]

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW_t,$$

[1]The issue of the bid and ask prices when there are several market makers is considered elsewhere; for instance, see Ho and Stoll (1980).

[2]Ho and Stoll (1981) took into account the possibility that the market maker does not know the true price of the stock. In this case, the quotation by the market maker makes the price deviate from the equilibrium price of the stock, which is considered to be the 'true' price. Then, the true price will be revealed by transactions.

with constant $\mu, \sigma > 0$, $\epsilon \in [0, 1)$. Further, we assume that the risk-free asset B is constant; i.e., $B_t \equiv 1$ for $t \in [0, T]$. In other words, we regard the process S as the discounted price process.

For these assets, we summarize the trading code as follows: the market maker and the investor do not need to pay transaction costs on risk-free assets, but the investor is required to pay the transaction costs for trading risky assets. More precisely, the ask price of risky assets is given by $(1 + \epsilon)S_t$ and the bid price is $(1-\epsilon)S_t$; i.e., the market maker buys one unit of the risky asset by paying $(1-\epsilon)S_t$ and sells one unit of the risky asset by receiving $(1+\epsilon)S_t$. Conversely, the investor can buy one unit of the risky asset by paying $(1+\epsilon)S_t$ and sell one unit of the risky asset by receiving $(1 - \epsilon)S_t$ at time t.

Using the trading code described above, we summarize the evolution of the investor's and the market maker's wealth as follows:

Cash position Market cash is accumulated when the agent sells the stock and paid out when the agent purchases the stock. This logic is common to any type of agent, that is, it applies to both the market maker and the investor. Any balance in the cash amount earns (or pays) the risk-free rate of the bond B. However, in this paper, we assume that B is constant. Thus, the value of the cash amount X^M (where the superscript M denotes the market maker) is given as follows:

$$dX_t^M := (1 + \epsilon)S_t d\varphi_t^a - (1 - \epsilon)S_t d\varphi_t^b, \tag{1}$$

where $d\varphi_t^a$ is the market maker's sales (the investor's purchases) and $d\varphi_t^b$ is the market maker's purchases (the investor's sales) at t. Thus, the position of the market maker is given by $\varphi_t^M = \varphi_0^M - \int_0^t d\varphi_t^a + \int_0^t d\varphi_t^b$, where φ_0^M is the initial risky position of the market maker. Then, the cash amount X^I (where the superscript I denotes the investor) is given as follows:

$$dX_t^I := -(1 + \epsilon)S_t d\varphi_t^a + (1 - \epsilon)S_t d\varphi_t^b,$$

and the position of the investor is given by $\varphi_t^I = \varphi_0^I + \int_0^t d\varphi_t^a - \int_0^t d\varphi_t^b$, where φ_0^I is the initial risky position of the investor. Hereafter, for simplicity, we denote $\varphi = \varphi^I$. Further, note that $\varphi_t^M = \varphi_0^M - (\varphi_t^I - \varphi_0^I)$.

Inventory The market maker's inventory (risky position) consists of shares of the one stock in which he/she makes a market. The change in the value of the inventory account, Y^M, is:

$$\begin{aligned} dY_t^M &:= -S_t d\varphi_t^a + S_t d\varphi_t^b + Y_t^M \frac{dS_t}{S_t} \\ &= -S_t d\varphi_t^a + S_t d\varphi_t^b + Y_t^M \mu dt + Y_t^M \sigma dW_t, \end{aligned} \tag{2}$$

where the first and second terms reflects the transaction with the investor and the remaining terms reflect the fluctuation of the current risky position.

Likewise, the investor's inventory Y^I is defined as follows:

$$\begin{aligned} dY_t^I :&= S_t d\varphi_t^a - S_t d\varphi_t^b + Y_t^I \frac{dS_t}{S_t} \\ &= S_t d\varphi_t^a - S_t d\varphi_t^b + Y_t^I \mu dt + Y_t^I \sigma dW_t. \end{aligned}$$

The total wealth of the market maker is $A^M = X^M + Y^M$ and that of the investor is $A^I = X^I + Y^I$.

For the market maker, the world ends at $\bar{T} \le T$ when the market maker is assumed to liquidate his/her inventory and base wealth at their market values without transaction costs.[3] We assume that the end of the world for the investor is T, which is longer than $\bar{T}$; i.e., the investor can continue his/her trading even if the market maker leaves the market, assuming that another market maker will appear. We have two optimization problems: one is of the market maker, and the other one is of the investor. We seek the optimal strategy of the market maker for choosing the optimal transaction costs ϵ to maximize the market maker's profit. The benefit arises from the fee the market maker is able to charge for the service of immediate trading. As the model assumes a single market maker in the stock market, that market maker has the ability to earn monopoly profits. Conversely, the optimal strategy of the investor for choosing the values φ^a, φ^b that maximize the investor's preference function. The benefit is defined by his/her terminal wealth, $A_{\bar{T}}^I$, where $\bar{T} \le T \le \infty$.

Essentially, this paper assumes that the market maker places the limit orders and the investor places the market orders. In such a model, we consider the market making of the market maker.

Now, let us describe our optimization problems. We consider the optimization problem of the market maker:

$$u^M(x, y) := \max_{\epsilon > 0} \mathbb{E}[U^M(X_{\bar{T}}^M + Y_{\bar{T}}^M)], \tag{3}$$

where we assume a profit-maximizing market maker. This implies a linear utility function, $U^M(x) = x$. We assume that the trading prices controlled by the market maker, i.e., the bid–ask spread ϵ, are only determined at the initial time $t = 0$. This is a reasonable assumption as it avoids creating confusion in regard to the investor's transactions. Thus, in the above utility maximization problem, the argument does not include t and the expected value on the right hand side is

[3]This assumption may be relaxed, as in Bichuch and Sircar (2015), but we adopt it for simplicity.

evaluated at $\mathcal{F}_0$. That is, we do not address dynamic programming and we only need to consider the optimization problem at $t = 0$.

Next, we set the optimization problem of the investor. Let $U^I(z) := \ln z$ be the utility function of the investor. Consider the value function of the investor's preference such that, for a given $x = X_t^I, y = Y_t^I$, we have:

$$u^I(t, x, y) := \sup_{\varphi^a, \varphi^b} \left\{ \mathbb{E}[U^I(X_T^I + Y_T^I)|\mathcal{F}_t] \right\}. \tag{4}$$

Addressing this issue is not straightforward. Before attempting to address it, we begin by discussing the utility maximization problem in the frictionless case.

Frictionless case In the frictionless case, the transaction cost ϵ is zero. Note that this case is consistent with the Black–Scholes model.

When $\epsilon = 0$, the dynamics of the cash amount X^I and the risky inventory Y^I is given as follows:

$$\begin{aligned} dX_t^I &= -S_t d\varphi_t \\ dY_t^I &= S_t d\varphi_t + Y_t^I(\mu dt + \sigma dW_t). \end{aligned}$$

From the above, the evolution of the total wealth A^I is given by:

$$dA_t^I = dX_t^I + dY_t^I = Y_t^I(\mu dt + \sigma dW_t).$$

Consider the following utility maximization problem:

$$u(t, x, y) \equiv u(t, a) = \sup_{A^I \in \mathcal{A}} \mathbb{E}\left[U^I(A_T^I)|\mathcal{F}_t\right],$$

where $\mathcal{A}$ is the set of total assets with all admissible strategies.

Then, the value function is explicitly given as follows:

$$u(t, x, y) = \ln(x + y) + \frac{1}{2}\frac{\mu^2}{\sigma^2}(T - t), \tag{5}$$

and the optimal risky weight $\pi_t = Y_t^I / A_t^I$ is given by:

$$\pi_t = -\frac{\mu u_a}{\sigma^2 u_{aa} A_t^I} = \frac{\mu}{\sigma^2}.$$

The deduction of the above results is not excessively complicated (see, for example, Chapter 14.5 of Back (2017)), and they form the benchmark of the frictional model.

3 Main results

Here, the main purpose is to derive the solution of the market maker's optimization problem, taking into account the optimal strategy of the investor. Thus, first, we need to clarify the investor's activity in the frictional case, that is, we need to clarify the value function and the trading volume of the investor.

Optimal investment strategy Here, we assume that:

$$0 < \mu/\sigma^2 < 1.$$

This assumption excludes short selling or leveraging the risky asset.[4]

Then, we can derive the value function (4); i.e., it holds that the value function is given by:

$$u^I(t, x, y) = \ln(x + y) + \frac{1}{2}\frac{\mu^2}{\sigma^2}\left(\frac{\sigma^2}{\mu} - 2\right)^2 \int_t^T \left(1 + 2\lambda\,(1 - \tilde{\pi}_s)^{-1}\right)^2 ds, \tag{6}$$

The first term (6) is the intrinsic value and the second term is the time value, as in the frictionless case. Note that the intrinsic value is the same as in the frictionless case and that only the time value is different. Thus, we need to discuss the time value.

As described later, λ drives the effect of the transaction costs on the investor's trading strategy. The process $\tilde{\pi}_t$ is interpreted as the risky weight in light of the shadow price, which is a hypothetical price process that corrects the original risky asset price S and maintains the consistency of the utility maximization problem in the frictionless and frictional cases. A more rigorous definition of the shadow price is given later (see Definition A.1). The form of $\tilde{\pi}$ is given by:

$$d\tilde{\pi}_t = -\,(1 - \tilde{\pi}_t + 2\lambda)\left(1 - \frac{2\mu}{\sigma^2}\right)\sigma dW_t + \tilde{\pi}_t(1 - \tilde{\pi}_t)d\Phi_t - \tilde{\pi}_t(1 - \tilde{\pi}_t)d\Psi_t,$$

where Φ and Ψ increase only on $\{\tilde{\pi}_t = \tilde{\pi}_+\}$ and $\{\tilde{\pi}_t = \tilde{\pi}_-\}$, respectively. The

[4]According to the Merton problem, $(\mu - r)/\gamma\sigma^2$ is the optimal ratio for the risky asset. As we set $r = 0$ and adopt the log utility ($\gamma = 1$), this ratio is μ/σ^2. Hence, the assumption that $0 < \mu/\sigma^2 < 1$ means that short selling and leveraging are excluded. Even if the Merton problem is discussed in the context of a smooth market, according to Davis and Norman (1990), see Theorem 5.1 and Remark 5.2, the condition $0 < \mu/\sigma^2 < 1$ ensures that investors cannot short sell or leverage in models with transaction costs.

definitions of $\tilde{\pi}_+$ and $\tilde{\pi}_-$ are given as follows:

$$\tilde{\pi}_- = \frac{\mu}{\sigma^2} + \frac{1}{2}\sqrt{\left(\left(1-\frac{2\mu}{\sigma^2}\right)+1\right)^2 + 8\lambda\left(1-\frac{2\mu}{\sigma^2}\right)}, \tag{7}$$

$$\tilde{\pi}_+ = \frac{\mu}{\sigma^2} - \frac{1}{2}\sqrt{\left(\left(1-\frac{2\mu}{\sigma^2}\right)+1\right)^2 + 8\lambda\left(1-\frac{2\mu}{\sigma^2}\right)}. \tag{8}$$

Recall that μ/σ^2 is the risky weight in the frictionless case. The second term in both (7) and (8) can be considered the regulation term resulting from the transaction costs. Thus, the process $\tilde{\pi}_t$ is, indeed, a similar process to the risky weight in the frictionless case.

Finally, we need to discuss λ, which is given a constant such that $\lambda \in \left[-\frac{(1-\mu/\sigma^2)^2}{2(1-2\mu/\sigma^2)}, -\frac{1}{2}\right]$ for $\mu/\sigma^2 < 1/2$ and $\lambda \in \left[0, -\frac{(1-\mu/\sigma^2)^2}{2(1-2\mu/\sigma^2)}\right]$ for $\mu/\sigma^2 \geq 1/2$. The precise form of λ satisfies the following:

$$\begin{aligned}\frac{1-\epsilon}{1+\epsilon} = &\left(\frac{4\lambda\left(1-\frac{2\mu}{\sigma^2}\right)+\left(1-\frac{2\mu}{\sigma^2}\right)+1+\sqrt{\left(\left(1-\frac{2\mu}{\sigma^2}\right)+1\right)^2+8\lambda\left(1-\frac{2\mu}{\sigma^2}\right)}}{4\lambda\left(1-\frac{2\mu}{\sigma^2}\right)+\left(1-\frac{2\mu}{\sigma^2}\right)+1-\sqrt{\left(\left(1-\frac{2\mu}{\sigma^2}\right)+1\right)^2+8\lambda\left(1-\frac{2\mu}{\sigma^2}\right)}}\right)\\ &\times\left(\frac{4\lambda+\left(1-\frac{2\mu}{\sigma^2}\right)+1+\sqrt{\left(\left(1-\frac{2\mu}{\sigma^2}\right)+1\right)^2+8\lambda\left(1-\frac{2\mu}{\sigma^2}\right)}}{4\lambda+\left(1-\frac{2\mu}{\sigma^2}\right)+1-\sqrt{\left(\left(1-\frac{2\mu}{\sigma^2}\right)+1\right)^2+8\lambda\left(1-\frac{2\mu}{\sigma^2}\right)}}\right)^{\frac{1}{1-\frac{2\mu}{\sigma^2}}}.\end{aligned} \tag{9}$$

This shows that when ϵ is zero, $\lambda = -\frac{(1-\mu/\sigma^2)^2}{2(1-2\mu/\sigma^2)}$. However, according to (7) and (8), this implies that $\tilde{\pi}_- = \tilde{\pi}_+ = \mu/\sigma^2$. It confirms the fact that $\tilde{\pi}$ is the corrected risky weight of the frictionless risky weight π, such that $\tilde{\pi}$ converges to π when ϵ approaches zero. Indeed, we can confirm that the case $\epsilon = 0$ reduces to $u^I(t,x,y) = u(t,x,y)$ by (6) and this confirms the consistency with the frictionless case.

Further, let us consider that ϵ approaches one, which is the maximum level of transaction costs. Then, λ tends towards $-1/2$ when $\mu/\sigma^2 < 1/2$. This implies that $\tilde{\pi}_- = 0$ and $\tilde{\pi}_+ = 2\mu/\sigma^2$. That is, if ϵ is approaching one, the boundary that the risky weight $\tilde{\pi}$ can take, this widens to $(0, 2\mu/\sigma^2)$ when $\mu/\sigma^2 < 1/2$. Conversely, if ϵ approaches one, then λ tends to zero when $\mu/\sigma^2 \geq 1/2$. This implies that $\tilde{\pi}_- = 1$ and $\tilde{\pi}_+ = 2\mu/\sigma^2 - 1$. That is, if ϵ is approaching one, the boundary that the risky weight $\tilde{\pi}$ can take widens to $(1, 2\mu/\sigma^2 - 1)$. The former case implies that the frictionless risky weight $\pi = \mu/\sigma^2 < 0.5$ and the latter case implies that the frictionless risky weight $\pi > 0.5$. Thus, the range of widened boundaries in the former case is, roughly, less than the range in the latter case. In summary, when the transaction costs ϵ become close to one, the boundary $(\tilde{\pi}_+, \tilde{\pi}_-)$ widens because λ works to translate the effect of the transaction costs into the optimal trading strategy.

Further, the dynamics of the trading strategy are as follows:

$$\frac{d\varphi_t}{\varphi_t} = (1 - \tilde{\pi}_-)d\Phi_t - (1 - \tilde{\pi}_+)d\Psi_t. \tag{10}$$

This implies that the trading volume is given by:

$$\frac{d||\varphi_t||}{\varphi_t} = (1 - \tilde{\pi}_-)d\Phi_t + (1 - \tilde{\pi}_+)d\Psi_t. \tag{11}$$

For $d\varphi/\varphi$ and $d||\varphi||/\varphi$, we define the average position of risky assets and the average trading volume such that:

$$\bar{r} := \lim_{t\to\infty} \frac{1}{t} \int_0^t \frac{d\varphi_t}{\varphi_t},$$
$$\bar{R} := \lim_{t\to\infty} \frac{1}{t} \int_0^t \frac{d||\varphi_t||}{\varphi_t}.$$

The details for deriving the above are discussed in Appendices A.1 and A.2.

Optimal market making strategy Next, we discuss the optimization of the market maker. To maximize profit, the market maker takes into account the strategy of the investor. From this game theoretic point of view, we replace (x, y) from the value function (3) with (x^M, y^M, y^I), where x^M and y^M denote the initial safe and risky positions of the market maker, respectively, and y^I denotes the initial risky position of the investor. Then, we can solve (3); i.e., the value function of the market maker is given by:

$$u^M(x^M, y^M, y^I) = x^M + y^M e^{\mu\bar{T}} + y^I \left(e^{\mu\bar{T}} - 1 - \frac{\epsilon R^* - \mu}{r^* + \mu} \left(e^{(r^*+\mu)\bar{T}} - 1 \right) \right), \tag{12}$$

where ϵ^* maximizes:

$$\left(-\mu + \epsilon\bar{R} \right) \frac{1}{\bar{r} + \mu} \left(e^{(\bar{r}+\mu)T} - 1 \right), \tag{13}$$

and R^* and r^* are given by substituting ϵ^* into the average rate of the trading volume $\bar{R}$ and the average rate of the trading strategy $\bar{r}$. A comparison the investor's value function (6) indicates that the value function (12) is not clearly divided into the time and the intrinsic value. Instead, this value function shows the strong dependency of the investor's activity, through terms such as y^I, R^* and r^*. It also shows the intersection of the investor and the market maker.

Further, we note that $\bar{R}$ and $\bar{r}$ give the approximate solutions for (10) and (11), as follows:

$$\varphi_t \approx \varphi_0 e^{\bar{r}t},$$

where:

$$\bar{r} = -\lambda\sigma^2\left(1 - 2\frac{\mu}{\sigma^2}\right),$$

and

$$||\varphi_t|| \approx \varphi_0 e^{\bar{R}t},$$

where:

$$\bar{R} = -\frac{\lambda\mu\left(1 - 2\frac{\mu}{\sigma^2}\right)}{2\sqrt{\left(1 - \frac{\mu}{\sigma^2}\right)^2 + 2\lambda\left(1 - 2\frac{\mu}{\sigma^2}\right)}}.$$

Note that the average trading volume $\bar{R}$ and the average position $\bar{r}$ are the strategies of the investor, but the market maker affects them by regulating the transaction costs ϵ via λ. In other words, the market maker cannot directly maximize his/her profit, but he/she can indirectly maximize it by affecting the investor's strategy. This is the game theoretic solution and the transaction cost ϵ^* is the optimizer of the interaction between the investor's and the market maker's strategies.

The interaction between the market maker and the investor may not be sufficiently clear from the mathematical formulation. Therefore, we illustrate its implications using the numerical examples in the following section.

4 Numerical Examples and Model Implications

In this section, we provide numerical examples to demonstrate the implications of our model. First, we need to confirm the consistency of our model with the standard model that excludes transaction costs. If the consistency with the standard model holds, the value function of our model given by (6) converges to the standard one given by (5). We calculate (6) by using the upper and lower bounds of $\tilde{\pi} \in [\tilde{\pi}_-, \tilde{\pi}_+]$ and show the results in Figure 1. The common parameters used are: $S_0 = 100$, $T = 1$, $x + y = 100$, $\sigma = 0.16$. In the left panel, the parameter is $\mu = 0.01$ and in the right panel, it is $\mu = 0.02$. In the figures, the dashed line represents the value function of the standard model given by (5). This implies that the value functions will coincide with each other when the transaction costs approach zero.

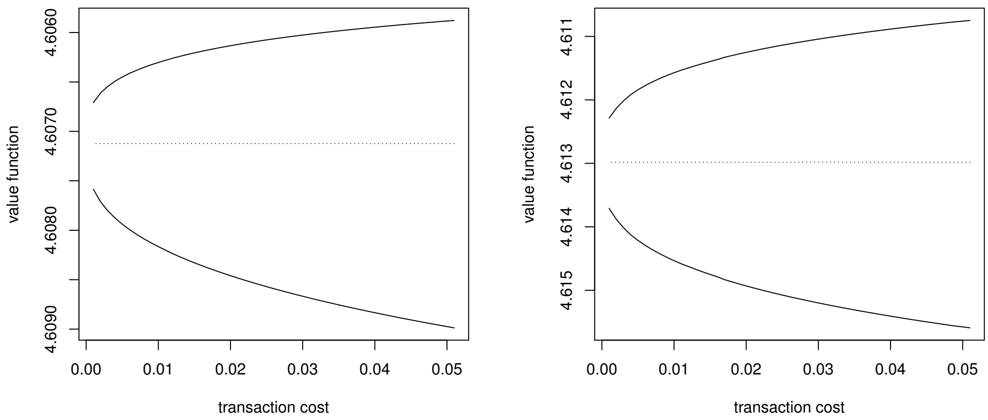

Figure 1: Convergence of the value function of the model with transaction costs to the value function of the standard model. The bold lines represent the upper and lower bounds of the value functions of the model with transaction costs. The dashed line is the value function of the standard model. The left panel is calculated on the condition that $\mu/\sigma^2 < 0.5$ and the right panel on the condition that $\mu/\sigma^2 \geq 0.5$.

The reason that we divided Figure 1 into two panels is that we need to consider two cases, $\mu/\sigma^2 < 0.5$ and $\mu/\sigma^2 \geq 0.5$. As we discussed in the previous section, the region of the parameter λ is given by $\lambda \in \left[-\frac{(1-\mu/\sigma^2)^2}{2(1-2\mu/\sigma^2)}, -\frac{1}{2}\right]$ when $\mu/\sigma^2 < 1/2$ and $\lambda \in \left[0, -\frac{(1-\mu/\sigma^2)^2}{2(1-2\mu/\sigma^2)}\right]$ when $\mu/\sigma^2 \geq 1/2$, although the form of λ is commonly given by (9). Thus, the region of $\tilde{\pi}$ is divided into two. We consider two cases of convergence of the value function.

As discussed in the previous section, we obtain the optimal transaction costs by numerically solving (13). The left panel of Figure 2 shows that when the appreciation rate μ increases, the optimal transaction costs decrease, for a fixed stock price $S_0 = 100$ (dollars), maturity $T = 1$ (Year), and volatility level $\sigma = 0.16$ (=16%). In contrast, as the right panel of Figure 2 shows, if there is increasing volatility σ, the optimal transaction costs increase, for a fixed appreciation rate $\mu = 0.02$ (=2%), when the stock price and maturity are also fixed at $S_0 = 100$ and $T = 1$, respectively.

There are two key points to discuss in relation to the response of the optimal transaction costs to changes in the appreciation rate and volatility. First, it is important to consider why the optimal transaction costs respond in the opposite direction to changes in the appreciation rate compared with changes in volatility. Second, we need to discuss the amount of transaction costs, shown in Figure 2.

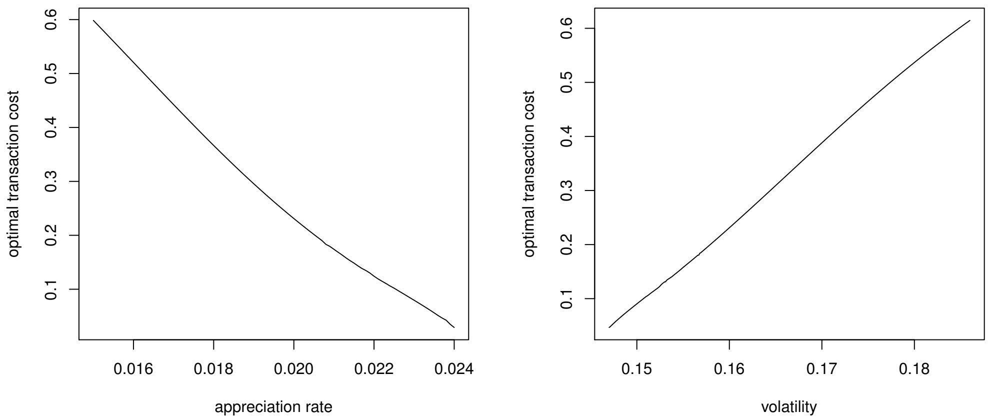

Figure 2: Increasing appreciation rate decreases the optimal transaction costs (left panel). Conversely, increasing volatility increases the optimal transaction costs (right panel).

First, we discuss why the response of the optimal transaction costs is reversed for a change of the appreciation rate compared with a change of volatility. The first explanation is simple: A large appreciation rate and a low level of volatility imply a large risky weight (see the first term of both (7) and (8)). This implies that the investor is confident in holding much amounts of the risky asset and may be induced to willingly trade the risky assets. Hence, low transaction costs may be sufficient to ensure the market maker's profit. This means the optimal transaction costs will be small when there is a higher appreciation rate and lower volatility.

From another perspective, the above conjecture is confirmed by a numerical example. Figure 3 shows that when the frictionless risky weight π increases, the spread between boundaries $\tilde{\pi}_+ - \tilde{\pi}_-$ decreases. As $\pi = \mu/\sigma^2$ increases when the appreciation rate μ increases and the volatility σ decreases, increasing μ and decreasing σ leads to a trading opportunity via the narrowing spread $\tilde{\pi}_+ - \tilde{\pi}_-$. Further, by changing μ and σ, the transaction cost ϵ is optimized. Thus, the reduction of the width between boundaries $\tilde{\pi}_+$ and $\tilde{\pi}_-$ is accelerated (see the right panel of Figure 3). This shows that transactions may easily occur. Indeed, because of the increase of the frictional risky position $\tilde{\pi}_t$, the possibility of $\tilde{\pi}_t$ touching the boundaries $\tilde{\pi}_+$ and $\tilde{\pi}_-$ is increased. This allows for the low transaction costs.

Although transactions increase, this has no benefit to the market maker if his/her total profit is reduced by the small transaction costs. Thus, we need to confirm that the market maker is able to regulate the transactions costs to ensure that his/her total profit is stable for an increasing appreciation rate and decreasing

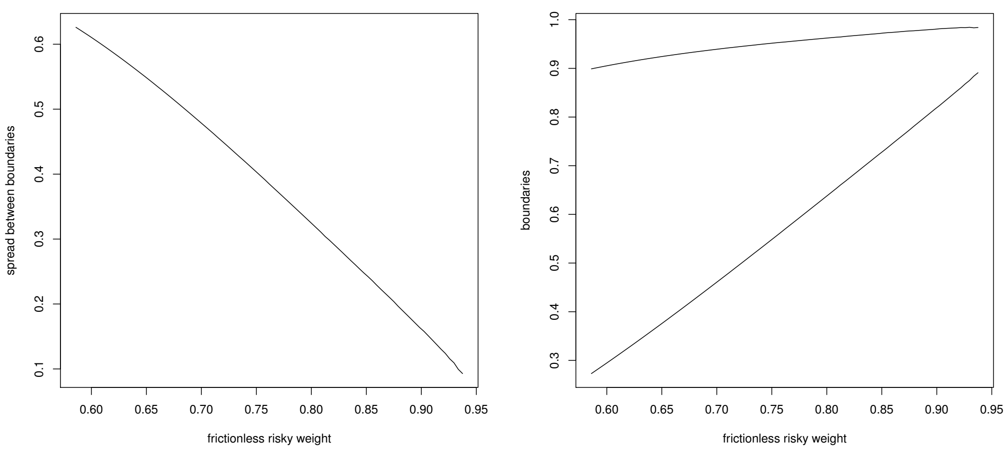

Figure 3: When the frictionless risky weight π increases, the spread between boundaries $\tilde{\pi}_+ - \tilde{\pi}_-$ decreases. The left panel shows the difference of boundaries $\tilde{\pi}_+$ and $\tilde{\pi}_-$ and the right panel shows each of the boundaries.

volatility. That is, we calculate the total profit $\int_0^T \varphi_t^M dS_t + \epsilon \int_0^T S_t d\|\varphi\|_t$. We set $y^M = y^I = 100$ and $\bar{T} = 1$ and show the result in Figure 4.

Indeed, as Figure 4 shows, even if the appreciation rate and volatility increase, the total profit of the market maker appears to be stable. This implies that the regulation of the transaction costs is optimally conducted by the market maker, although the transaction costs decrease or increase.

Now, we proceed to the second point, the amount of transaction costs. It is well known that transaction costs are not necessarily observed in the real market. One reason for this is that there are several market makers, so that one market maker cannot determine transaction costs in isolation. Another reason is that the transaction costs (represented by the bid–ask spread) are not perfectly controlled by a market maker, but experience fluctuations owing to random limit orders. This feature provides additional support for our idea that the transaction costs are game theoretically determined.

Based on these features of transaction costs, several methods have been suggested to empirically measure the real transaction costs; see, for instance, Roll (1984), Lesmond et al. (1999), Corwin and Schultz (2012) and Abdi and Ranaldo (2017). The estimated transaction costs vary depending on the methods used and the time periods in which they are measured. In general, however, transaction costs on small stocks are larger than those on large stocks. For instance, Lesmond et al. (1999) found that transaction costs varied within a range from 1.2% 10.3%, with the former applying to large stocks and the latter to small stocks. Further,

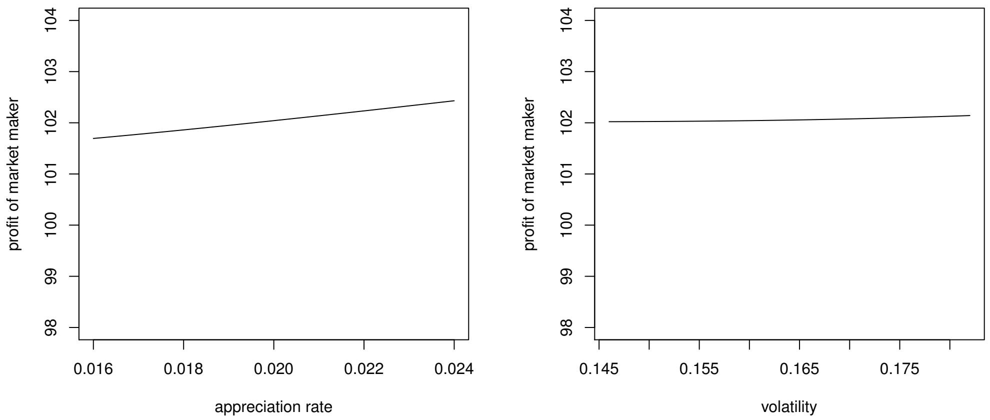

Figure 4: Total profit of the market maker when the appreciation rate and volatility increase.

Corwin and Schultz (2012) reported that the transaction costs on small stocks jumped to over 55% during the Great Depression. [5]

This phenomena makes intuitive sense because strong recessions, such as the Great Depression and the Lehman crisis, are characterized by very low appreciation rates and very high volatility. In such periods, even market makers may prefer to alleviate their inventory risks and thus may set excessively high transaction costs to avoid trading orders. From this viewpoint, excessively high transaction costs may be optimal in such situations, although it should be noted that even very high transaction costs could not prevent the fire sales during the Lehman crisis.

We can observe the corresponding parameters for these estimations of transaction costs using Figure 2. For instance, transaction costs of 10%, which apply to small stocks, correspond to appreciation rates of around 0.225 and volatility rates of 0.15. This implies a risky weight of around 85%. Further, transaction costs of 50%, which occurred during the Great Depression, correspond to an appreciation rate of around 0.15 and volatility of around 0.185. This implies that the risky weight moved to around 60% at this time. Although the corresponding risky weight may appear too high, our model simply gives the corresponding values of μ and σ for the empirically possible transaction costs. The reason that such high risky weights occur in our model may be the assumption of log utility, under which risk tolerance is fixed to one. If we relax this assumption and allow investors to adopt more risk-tolerant preferences, we may determine more reasonable risky weights for market estimates of transaction costs. This is a project for

[5]This report may be excessive as Abdi and Ranaldo (2017) reported more moderate estimates.

future research. The same procedure that we have employed in this paper could be applied to such a future project. This will enable us to test the efficiency of the extended model and provide economic interpretations.

In this regard, we note the range of μ and σ in Figure 2 and Figure 4. Compared with volatility of around 15%, the appreciation is very small (around 2%). This is because, to avoid leverage, we made the assumption that $0 < \pi = \mu/\sigma^2 < 1$. To examine the cases where μ and σ take values within a wider range, we have two options: First, we could relax the leverage condition. Second, we could introduce a power utility investor with a risk aversion $\gamma > 1$, which would relax the assumption such that $0 < \mu/\sigma^2 < \gamma$, or we could introduce a more risk-tolerant market maker.

A Results in the frictional case

The structure of this section is as follows: First, we discuss the value function (4). Second, we derive the trading volume (11) of the investor by using the result of the first step. Finally, we show the value function of the market maker; i.e., the solution of (3), based on the first and second steps.

A.1 The value function in the frictional case

For deriving the value function (4), it is convenient to use the shadow price approach. The shadow price is given by Definition A.1.

Definition A.1. *(Gerhold et al. (2014)) A shadow price is a frictionless price process $\tilde{S}_t$ evolving within the bid-ask spread, such that there is an optimal strategy for $\tilde{S}_t$ which is of finite variation generating same expected value with the frictional case and entails buying only when the shadow price $\tilde{S}_t$ equals the ask price $(1+\epsilon)S_t$, and selling only when $\tilde{S}_t$ equals the bid price $(1-\epsilon)S_t$.*

By definition of shadow price, the cash and risky position in terms of shadow price is given as follows:

$$d\tilde{X}^I_t = -\tilde{S}_t d\varphi_t,$$
$$d\tilde{Y}^I_t = \tilde{S}_t d\varphi_t + \tilde{Y}^I_t \frac{d\tilde{S}_t}{\tilde{S}_t}.$$

Note that we don't need to consider the effect of transaction costs in view of shadow price processes. Thus, the wealth process $\tilde{A}^I_t = \tilde{X}^I_t + \tilde{Y}^I_t$ in terms of shadow price is given by

$$d\tilde{A}^I_t = \tilde{Y}^I \frac{d\tilde{S}_t}{\tilde{S}_t}.$$

According to Theorem 2.10, 2.12, and 3.2 of Czichowsky (2017), there exists the above shadow prices. Further, Theorem 2.5 of Czichowsky et al. (2018) implies that the shadow price process is given by

$$\frac{d\tilde{S}_t}{\tilde{S}_t} = \tilde{\mu}_t dt + \tilde{\sigma}_t dW_t, \tag{14}$$

where $(\tilde{\mu}_t)_{t\in[0,T]}$ and $(\tilde{\sigma}_t)_{t\in[0,T]}$ are predictable processes such that the solution to (14) is well-defined in the sense of Itô integration. Note that Czichowsky et al. (2018) discusses the fractional Brownian motion drives the risky asset S. However, the standard Brownian motion is in a subset of fractional Brownian motion. Thus, the result is consistent with our case.

Let the value function in terms of shadow price be $\tilde{u}^I$ such that

$$u^I(t,x,y) = \tilde{u}^I(t,a) = \sup_{\tilde{A}^I\in\mathcal{A}} \mathbb{E}\left[U^I(\tilde{A}_T^I)|\mathcal{F}_t\right].$$

Since $\tilde{\mu}$ and $\tilde{\sigma}$ are well-defined, via the similar procedure to derive the value function discussed in the frictionless case of Section 2, we may approach the following value function;

$$\tilde{u}^I(t,a) = \ln(x+y) + \frac{1}{2}\int_t^T \frac{\tilde{\mu}_s^2}{\tilde{\sigma}_s^2} ds.$$

Further, the optimal risky weight $\tilde{\pi}_t$ is given by

$$\tilde{\pi}_t = -\frac{\tilde{\mu}_t \tilde{u}_a^I}{\tilde{\sigma}_t^2 \tilde{u}_{aa}^I \tilde{A}_t^I} \tag{15}$$

However, we have not yet known the explicit form of $\tilde{\mu}_t$ and $\tilde{\sigma}_t$. To clarify them, we need to explicitly derive the shadow price processes (14).

By the definition of shadow price, if the investor changes the position of risky asset by ν, then the value of the investor's safe position changes from X_t^I to $X_t^I - \nu\tilde{S}_t$, while her risky position changes from Y_t^I to $Y_t^I + \nu S_t$. Note that ν can be positive and negative, i.e., we consider selling and buying of risky asset, here. Thus, the following inequality holds,

$$u^I(t, x-\nu\tilde{S}_t, y+\nu S_t) \le u^I(t,x,y).$$

By a Taylor expansion on ν, the left hand side is $u^I(t, x-\nu\tilde{S}_t, y+\nu S_t) = u^I(t,x,y) - \left(u_x^I\tilde{S}_t - u_y^I S_t\right)\nu + \left(\frac{1}{2}u_{xx}^I\tilde{S}_t^2 - u_{xy}^I\tilde{S}_t S_t + \frac{1}{2}u_{yy}^I S_t^2\right)\nu^2 + \cdots = u^I(t,x,y) - \left(u_x^I\tilde{S}_t - u_y^I S_t\right)\nu + o(\nu)$. Thus, for sufficiently small ν, it follows

$$u_y^I \nu S_t \le u_x^I \nu \tilde{S}_t.$$

This holds independently of the sign of ν, we have the equality of the above:

$$\tilde{S}_t = \frac{u^I_y}{u^I_x} S_t.$$

This implies that the shadow price can be parametrized by a process C which is valued in $[C_-, C_+]$ such that

$$\tilde{S}_t = S_t \exp(C), \tag{16}$$

where

$$C_- := \ln(1-\epsilon),\ C_+ := \ln(1+\epsilon)$$

due to the definition of the shadow price; i.e., $\tilde{S} \in [(1-\epsilon)S, (1+\epsilon)S]$. Formally, we write the process C as

$$dC_t = \tilde{\mu}^C_t dt + \tilde{\sigma}^C_t dW_t,$$

with adapted processes $\tilde{\mu}^C, \tilde{\sigma}^C$. By (16), it holds that

$$\begin{aligned}
&\frac{d\tilde{S}_t}{\tilde{S}_t} - \frac{1}{2}\frac{d\langle \tilde{S}\rangle}{\tilde{S}^2} = \frac{dS_t}{S_t} - \frac{1}{2}\frac{d\langle S\rangle}{S^2} + dC \\
&\quad \leftrightarrow \left(\tilde{\mu}_t - \frac{1}{2}\tilde{\sigma}^2_t\right)dt + \tilde{\sigma}_t dW_t = \left(\mu - \frac{1}{2}\sigma^2\right)dt + \sigma dW_t + \tilde{\mu}^C_t dt + \tilde{\sigma}^C_t dW_t.
\end{aligned}$$

That is, it follows that

$$\begin{aligned}
\tilde{\mu}_t =& \mu + \tilde{\mu}^C_t - \frac{1}{2}\left(\sigma^2 - (\tilde{\sigma}_t)^2\right) = \mu + \tilde{\mu}^C_t + \frac{1}{2}\left(\sigma + \tilde{\sigma}^C_t\right)^2 - \frac{1}{2}\sigma^2 \\
\tilde{\sigma}_t =& \sigma + \tilde{\sigma}^C_t.
\end{aligned}$$

According to (15), we have to note that the optimal risky weight $\tilde{\pi}$ is given as follows;

$$\tilde{\pi} = \frac{\tilde{\mu}_t / \tilde{A}^I_t}{\tilde{\sigma}^2_t \tilde{A}^I_t / (\tilde{A}^I_t)^2} = \frac{\tilde{\mu}_t}{\tilde{\sigma}^2_t} = \frac{\mu + \tilde{\mu}^C - \frac{1}{2}\sigma^2}{(\sigma + \tilde{\sigma}^C)^2} + \frac{1}{2}. \tag{17}$$

By assumption of short selling constraint, $\tilde{\pi} > 0$. We define a process

$$\beta := \ln\left(\frac{\tilde{\pi}}{1-\tilde{\pi}}\right). \tag{18}$$

Note that the optimal strategy trades the shadow price process only when it coincides with bid or ask price. Thus, risky position φ_t and safe position φ^0_t must be constant on $(0, \tau)$ with $\tau := \inf\{t > 0 : C_t \in [C_-, C_+]\}$. Therefore, it follows that, on $[0, \tau]$,[6]

$$d\beta_t = d\ln\varphi_t + d\ln\tilde{S}_t - d\ln\varphi_t^0 \tag{19}$$

$$= d\ln\tilde{S}_t = \left(\mu - \frac{\sigma^2}{2} + \tilde{\mu}_t^C\right)dt + \left(\sigma + \tilde{\sigma}_t^C\right)dW_t \tag{20}$$

On the other hand, since $\tilde{\pi}$ is dependent on the process C, β is also dependent on the process C. That is, for some function f, it holds $\beta = f(C)$. By Itô's lemma, this implies that

$$d\beta_t = \left(f'(C_t)\tilde{\mu}_t^C + f''(C_t)\frac{(\tilde{\sigma}_t^C)^2}{2}\right)dt + f'(C_t)\tilde{\sigma}_t^C dW_t. \tag{21}$$

From (17), (20) and (21), it holds that

$$\frac{1}{1+e^{-f}} = \frac{\mu + \tilde{\mu}^C - \sigma^2/2}{(\sigma + \tilde{\sigma}^C)^2} + \frac{1}{2}, \tag{22}$$

$$\mu - \frac{\sigma^2}{2} + \tilde{\mu}^C = f'\tilde{\mu}^C + f''\frac{(\tilde{\sigma}^C)^2}{2}, \tag{23}$$

$$\sigma + \tilde{\sigma}^C = f'\tilde{\sigma}^C. \tag{24}$$

(22) and (24) yields

$$\tilde{\sigma}^C = \frac{\sigma}{f' - 1}, \tag{25}$$

$$\tilde{\mu}^C = -\mu + \frac{1}{2}\sigma^2 + \frac{1}{2}\sigma^2\left(\frac{f'}{f'-1}\right)^2\frac{1-e^{-f}}{1+e^{-f}}. \tag{26}$$

Inserting them into (23), we obtain the following ODE;

$$f''(x) = \left(\frac{2\mu}{\sigma^2} - 1\right)f'(x) + \left(2 - \frac{4\mu}{\sigma^2} + \frac{1-e^{-f(x)}}{1+e^{-f(x)}}\right)(f'(x))^2 + \left(2\frac{\mu}{\sigma^2} - 1 - \frac{1-e^{-f(x)}}{1+e^{-f(x)}}\right)(f'(x))^3. \tag{27}$$

The boundary condition is given by the boundaries $C_- = \ln(1-\epsilon)$ and $C_+ = \ln(1+\epsilon)$.

---

[6] Of course, if we consider the time interval $[0, T]$ in general, then φ_t fluctuates and (20) may be written as

$$d\beta_t = d\ln\tilde{S}_t + d\Phi_t - d\Psi_t$$

for processes Φ and Ψ which only increase when β touches $\beta_+ = f(C_-)$ and touches $\beta_- = f(C_+)$, as later discussed.

According to the logic discussed in pp.1347 of Kallsen and Muhle-Karbe (2010), for keeping the arbitrage free condition, it is necessary that Itô process representation of C to hold even when C reaches the boundaries C_- and C_+. This requires that f cannot be a C^2-function on the closed interval $[C_-, C_+]$; i.e., it follows that $f'(C_-) = f'(C_+) = -\infty$. This is boundary conditions of the ODE on f, but this is not easy to treat. Then, we consider the inverse function $g := f^{-1}$.[7] Then, since $f' = 1/g'$ and $f'' = -g''/(g')^3$, the ODE (27) is reduced to

$$
\begin{aligned}
-\frac{g''(y)}{(g'(y))^3} =&\left(\frac{2\mu}{\sigma^2} - 1\right)\frac{1}{g'(y)} + \left(2 - \frac{4\mu}{\sigma^2} + \frac{1-e^{-y}}{1+e^{-y}}\right)\frac{1}{(g'(y))^2} + \left(2\frac{\mu}{\sigma^2} - 1 - \frac{1-e^{-y}}{1+e^{-y}}\right)\frac{1}{(g'(y))^3} \\
\leftrightarrow g''(y) =&\left(1 - \frac{2\mu}{\sigma^2} + \frac{1-e^{-y}}{1+e^{-y}}\right) + \left(\frac{4\mu}{\sigma^2} - 2 - \frac{1-e^{-y}}{1+e^{-y}}\right)g'(y) + \left(1 - \frac{2\mu}{\sigma^2}\right)(g'(y))^2 \\
=&(1-g'(y))\left(\left(1 - \frac{2\mu}{\sigma^2}\right)(1-g'(y)) + \frac{1-e^{-y}}{1+e^{-y}}\right),
\end{aligned}
\tag{28}
$$

on the a priori unknown interval $[\beta_-, \beta_+] := [f(C_+), f(C_-)]$, with boundary (and smooth pasting) conditions such that

$$g(\beta_-) = C_+,\ g(\beta_+) = C_-,\ g'(\beta_-) = 0,\ g'(\beta_+) = 0. \tag{29}$$

According to Proposition 4.2 of Kallsen and Muhle-Karbe (2010),[8] there exist $\beta_- < \beta_+$ and a strictly decreasing mapping $g : [\beta_-, \beta_+] \to [C_-, C_+]$ satisfying the free boundary problem (28) and (29). Thus, the problem is how to derive the explicit form of the function $g(\cdot)$. Indeed, we do it as follows.

By (28), it holds that

$$
\begin{aligned}
\frac{g''(y)}{(1-g'(y))} &= \left(1 - \frac{2\mu}{\sigma^2}\right)(1-g'(y)) + \frac{1-e^{-y}}{1+e^{-y}} \\
\leftrightarrow -\frac{d\ln(1-g'(y))}{dx} &= \left(1 - \frac{2\mu}{\sigma^2}\right)(1-g'(y)) + \frac{1-e^{-y}}{1+e^{-y}}.
\end{aligned}
$$

Define $h(y) := \ln(1-g'(y))$ and $A := \left(1 - \frac{2\mu}{\sigma^2}\right)$ and rewrite the above equation; i.e.,

$$-h'(y) = Ae^{h(y)} + \frac{1-e^{-y}}{1+e^{-y}}.$$

This has the general solution such that

$$h(y) = \ln\left(-\frac{e^y}{A(1+e^y)(1+2\lambda+2\lambda e^y)}\right),$$

[7] According to Proposition 4.4 of Kallsen and Muhle-Karbe (2010), the monotonicity of g and f is assured.

[8] Although coefficients are slightly different from the problem discussed in Kallsen and Muhle-Karbe (2010), the logic is essentially same to this paper. Thus, the result of Proposition 4.2 of Kallsen and Muhle-Karbe (2010) is applicable for our discussion.

where λ is a constant. Hence, the general form of $g'(y)$ and $g(y)$ are given as follows:

$$g'(y) = 1 + \frac{e^y}{A(1+e^y)(1+2\lambda+2\lambda e^y)}, \tag{30}$$

$$g(y) = y + \frac{\ln(1+e^y) - \ln(1+2\lambda+2\lambda e^y)}{A} + \Lambda, \tag{31}$$

where Λ is an integral constant. From the above and the boundary conditions, we can specify λ, Λ, β_- and β_+. Indeed, by (30) and $g'(\beta_-) = g'(\beta_+) = 0$ and noting that $\beta_- < \beta_+$, it follows that, if $\lambda A > 0$, then

$$\beta_- = \ln\left(-1 - \frac{A+1+\sqrt{(A+1)^2+8\lambda A}}{4\lambda A}\right), \tag{32}$$

$$\beta_+ = \ln\left(-1 - \frac{A+1-\sqrt{(A+1)^2+8\lambda A}}{4\lambda A}\right), \tag{33}$$

else if $\lambda A \leq 0$, then

$$\beta_- = \ln\left(-1 - \frac{A+1-\sqrt{(A+1)^2+8\lambda A}}{4\lambda A}\right), \beta_+ = \ln\left(-1 - \frac{A+1+\sqrt{(A+1)^2+8\lambda A}}{4\lambda A}\right).$$

Since $g(\beta_-) = C_+ = \ln(1+\epsilon)$ and $g(\beta_+) = C_- = \ln(1-\epsilon)$ and (31), it also follows that

$$\ln\left(\frac{1-\epsilon}{1+\epsilon}\right) = \beta_+ - \beta_- + \frac{1}{A}\left(\ln\left(\frac{1+e^{\beta_+}}{1+e^{\beta_-}}\right) - \ln\left(\frac{1+2\lambda+2\lambda e^{\beta_+}}{1+2\lambda+2\lambda e^{\beta_-}}\right)\right).$$

Substituting (32) and (33) into the above, if $\lambda A > 0$, then we have

$$\ln\left(\frac{1-\epsilon}{1+\epsilon}\right) = \ln\left(\frac{-4\lambda A - A - 1 + \sqrt{(A+1)^2+8\lambda A}}{-4\lambda A - A - 1 - \sqrt{(A+1)^2+8\lambda A}}\right)$$
$$+ \frac{1}{A}\left(\ln\left(\frac{-(A+1)+\sqrt{(A+1)^2+8\lambda A}}{-(A+1)-\sqrt{(A+1)^2+8\lambda A}}\right) - \ln\left(\frac{A-1+\sqrt{(A+1)^2+8\lambda A}}{A-1-\sqrt{(A+1)^2+8\lambda A}}\right)\right).$$

This can be simplified as follows:

$$\begin{aligned}
\frac{1-\epsilon}{1+\epsilon} = &\left(\frac{-4\lambda A - A - 1 + \sqrt{(A+1)^2+8\lambda A}}{-4\lambda A - A - 1 - \sqrt{(A+1)^2+8\lambda A}}\right) \\
&\times \left(\frac{\left(-(A+1)+\sqrt{(A+1)^2+8\lambda A}\right)\left(A-1-\sqrt{(A+1)^2+8\lambda A}\right)}{\left(-(A+1)-\sqrt{(A+1)^2+8\lambda A}\right)\left(A-1+\sqrt{(A+1)^2+8\lambda A}\right)}\right)^{1/A} \\
= &\left(\frac{4\lambda A + A + 1 - \sqrt{(A+1)^2+8\lambda A}}{4\lambda A + A + 1 + \sqrt{(A+1)^2+8\lambda A}}\right)\left(\frac{4\lambda + A + 1 - \sqrt{(A+1)^2+8\lambda A}}{4\lambda + A + 1 + \sqrt{(A+1)^2+8\lambda A}}\right)^{1/A}.
\end{aligned} \tag{34}$$

Similarly, if $\lambda A \leq 0$, then if follows that

$$\frac{1-\epsilon}{1+\epsilon} = \left(\frac{4\lambda A + A + 1 + \sqrt{(A+1)^2 + 8\lambda A}}{4\lambda A + A + 1 - \sqrt{(A+1)^2 + 8\lambda A}}\right)\left(\frac{4\lambda + A + 1 + \sqrt{(A+1)^2 + 8\lambda A}}{4\lambda + A + 1 - \sqrt{(A+1)^2 + 8\lambda A}}\right)^{1/A}. \tag{35}$$

Solving the above, λ is determined and substituting λ into (32) and (33), we have β_- and β_+. We also have Λ by using $g(\beta_-) = C_+$ or $g(\beta_+) = C_-$.

Using (25) and (26), the process β is given on $(0, T)$ by non-decreasing adapted processes Φ, Ψ which increase only on the set $\{\beta = \beta_-\}$ and $\{\beta = \beta_+\}$, respectively[9], as follows:

$$d\beta_t = a(\beta_t)dt + b(\beta_t)dW_t + d\Phi_t - d\Psi_t, \tag{36}$$

$$\beta_0 \in [\beta_-, \beta_+], \tag{37}$$

for functions a, b defined by

$$a(y) := \frac{1}{2}\sigma^2 \left(\frac{1}{1-g'(y)}\right)^2 \frac{1-e^{-y}}{1+e^{-y}}, \; b(y) := \frac{\sigma}{1-g'(y)}. \tag{38}$$

Since $\beta \in [\beta_-, \beta_+]$, it holds that

$$\beta_t = \beta_0 + \int_0^t a(\beta_s)ds + \int_0^t b(\beta_s)dW_s + \Phi_t - \Psi_t. \tag{39}$$

Note that the definition $\beta = f(C)$; i.e., it follows $C = g(\beta)$. And now, we have β. Thus, we attain the explicit form of the process C, as follows:

$$\begin{aligned} dC_t =& \left(g'(\beta_t)a(\beta_t) + \frac{1}{2}g''(\beta_t)b(\beta_t)^2\right)dt + g'(\beta_t)b(\beta_t)dW_t \\ =& \left(-\mu + \frac{1}{2}\sigma^2 + \frac{1}{2}\sigma^2 \frac{1-e^{-y}}{1+e^{-y}} \frac{1}{(1-g'(y))^2}\right)dt + \sigma\frac{g'(\beta_t)}{1-g'(\beta_t)}dW_t, \end{aligned}$$

where we used that

$$\begin{aligned} & g'(\beta_t)a(\beta_t) + \frac{1}{2}g''(\beta_t)b(\beta_t)^2 \\ =& g'(\beta_t)\left(\frac{1}{2}\sigma^2\left(\frac{1}{1-g'(y)}\right)^2 \frac{1-e^{-y}}{1+e^{-y}}\right) \\ &+ \frac{1}{2}(1-g'(y))\left(\left(1-\frac{2\mu}{\sigma^2}\right)(1-g'(y)) + \frac{1-e^{-y}}{1+e^{-y}}\right)\left(\frac{\sigma}{1-g'(y)}\right)^2 \\ =& -\mu + \frac{1}{2}\sigma^2 + \frac{1}{2}\sigma^2\frac{1-e^{-y}}{1+e^{-y}}\frac{1}{(1-g'(y))^2}. \end{aligned}$$

[9] See Lemma 4.3 of Kallsen and Muhle-Karbe (2010) and Skorokhod (1961) for more details.

The above is derived by using (38) and (28).

This is consistent with the definition of C_t. Indeed, by using (25) and (26), it follows that

$$\begin{aligned}\tilde{\sigma}^C =&\frac{g'}{1-g'}\sigma,\\ \tilde{\mu}^C =&-\mu+\frac{1}{2}\sigma^2+\frac{1}{2}\sigma^2\left(\frac{1}{1-g'}\right)^2\frac{1-e^{-\beta_t}}{1+e^{-\beta_t}}.\end{aligned}$$

Further, we specify the coefficients of the shadow price $\tilde{S}_t$; i.e., it follows that

$$\begin{aligned}\tilde{\sigma}_t =&\sigma+\tilde{\sigma}^C_t=\frac{1}{1-g'}\sigma,\\ \tilde{\mu} =&\mu+\tilde{\mu}^C-\frac{1}{2}(\sigma^2-\tilde{\sigma}^2)=\frac{1}{(1-g')^2}\frac{\sigma^2}{1+e^{-\beta_t}}.\end{aligned}$$

Summarizing the above, the value function is given by

$$\begin{aligned}u^I(t,x,y)&=\ln(x+y)+\frac{1}{2}\int_t^T\frac{\tilde{\mu}_s^2}{\tilde{\sigma}_s^2}ds\\ &=\ln(x+y)+\frac{\sigma^2}{2}\int_t^T\frac{1}{(1-g'(\beta_s))^2(1+e^{-\beta_s})^2}ds\\ &=\ln(x+y)+\frac{\sigma^2}{2}\int_t^T A^2(1+2\lambda+2\lambda e^{\beta_s})^2 ds\end{aligned}$$

Substituting $g'(\cdot)$ of (30) into the above, (6) is derived.

Remark A.1. *The definition of the shadow price also implies that the shadow price process $\tilde{S}$ coincide with the original risky asset price process S when $\epsilon\to 0$; i.e., it means that $\tilde{\mu}\to\mu$ and $\tilde{\sigma}\to\sigma$.*

Indeed, the above holds, because the monotonicity of $g(\cdot)$ and $\beta_- < \beta_+$ implies that $g'(\beta_t)=0$ for all β_t, when $\epsilon\to 0$. Further, when the transaction cost $\epsilon\to 0$, the risky weight $\tilde{\pi}$ coincides with the frictionless case, i.e., $\tilde{\pi}=\mu/\sigma^2$. This means that $\beta_t\to-\ln(-1+\sigma^2/\mu)$ (see (18)). By using them, it follows that, when $\epsilon\to 0$,

$$\begin{aligned}\tilde{\mu}_t =&\frac{1}{(1-g')^2}\frac{\sigma^2}{1+e^{-\beta_t}}\to\mu,\\ \tilde{\sigma}_t =&\frac{1}{1-g'}\sigma\to\sigma.\end{aligned}$$

This implies that the consistency of the shadow price $\tilde{S}$ with S and the consistency of the value function of the frictional case with standard Merton problem such that, when $\epsilon\to 0$,

$$u^I(t,x,y)\to u(t,x,y).$$

Remark A.2. *Another way to confirm the consistency with standard Merton problem is to derive β_- and β_+ when $\epsilon \to 0$. When $\lambda A > 0$ and $\epsilon \to 0$ on (34), then the left hand side is 1. Hence, it holds that $\lambda = -(A+1)^2/8A$.*

Substituting this into (32) and (33), it follows that

$$\beta_- \to \ln\left(-1 + \frac{2}{A+1}\right),$$
$$\beta_+ \to \ln\left(-1 + \frac{2}{A+1}\right).$$

That is, $\beta_t = \beta_- = \beta_+ \to \ln\left(-1 + \frac{2}{A+1}\right)$.

Therefore, the value function converges such that

$$u^I(t,x,y) \to \ln(x+y) + \frac{1}{2}\frac{\mu^2}{\sigma^2}(T-t) = u(t,x,y).$$

A.2 Trading volume

By (18), it holds that $\tilde{\pi}_t = \frac{1}{1+e^{-\beta_t}}$. Since the process β is given by (39), the process of the risky weight $\tilde{\pi}$ is also determined. Using $\tilde{\pi}$, the wealth $\tilde{A}$ in terms of $\tilde{S}$ is described as follows;

$$\tilde{A}^I_t = (X^I_0 + \tilde{Y}^I_0)\mathcal{E}\left(\int_0^{\cdot} \tilde{\pi}_s \frac{d\tilde{S}_s}{\tilde{S}_s}\right)_t = (X^I_0 + \tilde{Y}^I_0)\mathcal{E}\left(\int_0^{\cdot} \frac{1}{1+e^{-\beta_s}} \frac{d\tilde{S}_s}{\tilde{S}_s}\right)_t \qquad (40)$$

For given X^I_0, Y^I_0, the wealth process $\tilde{A}$ is completely defined. Further, the risky position φ is also defined such that

$$\varphi_t = \tilde{\pi}_t \frac{\tilde{A}^I_t}{\tilde{S}_t}.$$

Using Itô's lemma, it follows that

$$\begin{aligned} d\varphi_t =& d\tilde{\pi}_t \frac{\tilde{A}^I_t}{\tilde{S}_t} + \tilde{\pi}_t \frac{d\tilde{A}^I_t}{\tilde{S}_t} - \tilde{\pi}_t \frac{\tilde{A}^I_t}{\tilde{S}_t}\frac{d\tilde{S}_t}{\tilde{S}_t} + \tilde{\pi}_t \frac{\tilde{A}^I_t}{\tilde{S}^3_t} d\langle \tilde{S}\rangle_t + \frac{d\langle \tilde{\pi}, \tilde{A}^I\rangle_t}{\tilde{S}_t} - \frac{d\langle \tilde{\pi}, \tilde{S}\rangle_t}{\tilde{S}^2_t}\tilde{A}^I_t - \tilde{\pi}_t \frac{d\langle \tilde{S}, \tilde{A}^I\rangle_t}{\tilde{S}^2_t} \\ =& \varphi_t \frac{d\tilde{\pi}_t}{\tilde{\pi}_t} + \varphi_t \frac{d\tilde{A}^I_t}{\tilde{A}^I_t} - \varphi_t \frac{d\tilde{S}_t}{\tilde{S}_t} + \varphi_t \frac{d\langle \tilde{S}\rangle_t}{\tilde{S}^2_t} + \varphi_t \frac{d\langle \tilde{\pi}, \tilde{A}^I\rangle_t}{\tilde{\pi}_t \tilde{A}^I_t} - \varphi_t \frac{d\langle \tilde{\pi}, \tilde{S}\rangle_t}{\tilde{\pi}_t \tilde{S}_t} - \varphi_t \frac{d\langle \tilde{A}^I, \tilde{S}\rangle_t}{\tilde{A}^I_t \tilde{S}_t}. \end{aligned} \qquad (41)$$

According to (40), it holds that

$$\frac{d\tilde{A}^I_t}{\tilde{A}^I_t} = \tilde{\pi}_t \frac{d\tilde{S}_t}{\tilde{S}_t}.$$

Since $d\beta_t = d\ln\tilde{S}_t + d\Phi_t - d\Psi_t$ by definition, it holds that

$$\begin{aligned}
d\beta_t =& \frac{d\tilde{S}_t}{\tilde{S}_t} - \frac{1}{2}\frac{d\langle\tilde{S}\rangle_t}{\tilde{S}_t^2} + d\Phi_t - d\Psi_t = \frac{d\tilde{S}_t}{\tilde{S}_t} - \frac{1}{2}\tilde{\sigma}_t^2 dt + d\Phi_t - d\Psi_t,\\
d\langle\beta\rangle_t =& \frac{d\langle\tilde{S}\rangle_t}{\tilde{S}_t^2}.
\end{aligned}$$

Further, note that $1-\tilde{\pi}_t = \frac{e^{-\beta_t}}{1+e^{-\beta_t}}$. Using them, it follows that

$$\begin{aligned}
d\tilde{\pi}_t =& \frac{e^{-\beta_t}}{(1+e^{-\beta_t})^2}d\beta_t + \frac{1}{2}\left(-\frac{e^{-\beta_t}}{(1+e^{-\beta_t})^2} + 2\frac{e^{-2\beta_t}}{(1+e^{-\beta_t})^3}\right)d\langle\beta\rangle_t\\
=& \tilde{\pi}_t(1-\tilde{\pi}_t)d\beta_t - \frac{1}{2}\tilde{\pi}_t(1-\tilde{\pi}_t)\left(1-2(1-\tilde{\pi}_t)\right)d\langle\beta\rangle_t\\
=& \tilde{\pi}_t(1-\tilde{\pi}_t)\frac{d\tilde{S}_t}{\tilde{S}_t} - \frac{1}{2}\tilde{\pi}_t(1-\tilde{\pi}_t)\frac{d\langle\tilde{S}\rangle_t}{\tilde{S}_t^2} - \frac{1}{2}\tilde{\pi}_t(1-\tilde{\pi}_t)\left(1-2(1-\tilde{\pi}_t)\right)\frac{d\langle\tilde{S}\rangle_t}{\tilde{S}_t^2}\\
&+ \tilde{\pi}_t(1-\tilde{\pi}_t)d\Phi_t - \tilde{\pi}_t(1-\tilde{\pi}_t)d\Psi_t\\
=& \tilde{\pi}_t(1-\tilde{\pi}_t)\frac{d\tilde{S}_t}{\tilde{S}_t} - \tilde{\pi}_t^2(1-\tilde{\pi}_t)\frac{d\langle\tilde{S}\rangle_t}{\tilde{S}_t^2} + \tilde{\pi}_t(1-\tilde{\pi}_t)d\Phi_t - \tilde{\pi}_t(1-\tilde{\pi}_t)d\Psi_t.
\end{aligned}$$

Thus, it holds that

$$\begin{aligned}
\frac{d\tilde{\pi}_t}{\tilde{\pi}_t} =& (1-\tilde{\pi}_t)\frac{d\tilde{S}_t}{\tilde{S}_t} - \tilde{\pi}_t(1-\tilde{\pi}_t)\frac{d\langle\tilde{S}\rangle_t}{\tilde{S}_t^2} + (1-\tilde{\pi}_t)d\Phi_t - (1-\tilde{\pi}_t)d\Psi_t\\
=& -A\frac{1-\tilde{\pi}_t}{\tilde{\pi}_t}\left(1+\frac{2\lambda}{1-\tilde{\pi}_t}\right)\sigma dW_t + (1-\tilde{\pi}_t)d\Phi_t - (1-\tilde{\pi}_t)d\Psi_t.
\end{aligned}$$

Note that by definition of $\tilde{\pi}$ and β, it holds that $d\Phi$ and $d\Psi$ increase only on $\{\tilde{\pi}_t = \tilde{\pi}_-\}$ and $\{\tilde{\pi}_t = \tilde{\pi}_+\}$, respectively, where $\pi_- := \frac{1}{1+e^{-\beta_-}}$ and $\pi_+ := \frac{1}{1+e^{-\beta_+}}$. More precisely, we have, if $\lambda\left(1-\frac{2\mu}{\sigma^2}\right) > 0$,

$$\begin{aligned}
\tilde{\pi}_- =& \frac{\mu}{\sigma^2} + \frac{1}{2}\sqrt{\left(\left(1-\frac{2\mu}{\sigma^2}\right)+1\right)^2 + 8\lambda\left(1-\frac{2\mu}{\sigma^2}\right)},\\
\tilde{\pi}_+ =& \frac{\mu}{\sigma^2} - \frac{1}{2}\sqrt{\left(\left(1-\frac{2\mu}{\sigma^2}\right)+1\right)^2 + 8\lambda\left(1-\frac{2\mu}{\sigma^2}\right)},
\end{aligned}$$

else if $\lambda\left(1-\frac{2\mu}{\sigma^2}\right) \leq 0$, given by

$$\tilde{\pi}_- = \frac{\mu}{\sigma^2} - \frac{1}{2}\sqrt{\left(\left(1-\frac{2\mu}{\sigma^2}\right)+1\right)^2 + 8\lambda\left(1-\frac{2\mu}{\sigma^2}\right)},$$
$$\tilde{\pi}_+ = \frac{\mu}{\sigma^2} + \frac{1}{2}\sqrt{\left(\left(1-\frac{2\mu}{\sigma^2}\right)+1\right)^2 + 8\lambda\left(1-\frac{2\mu}{\sigma^2}\right)}.$$

Note that the former case implies that $\tilde{\pi}_+ < \tilde{\pi}_-$. However, this means that when the optimal risky weight $\tilde{\pi}$ touches boundaries $\tilde{\pi}_-$ and $\tilde{\pi}_+$, $\tilde{\pi}$ exits from the region covered by these boundaries. This contradicts the optimality of $\tilde{\pi}$. Indeed, the model with transaction costs require to keep the position in fixed region to save the transaction costs. Thus, we can exclude the case $\lambda\left(1-\frac{2\mu}{\sigma^2}\right) > 0$, hereafter. That is, we focus on the case $\lambda\left(1-\frac{2\mu}{\sigma^2}\right) \leq 0$

Next thing we do is to refine the available range of λ. For this, we write (35) again here:

$$\frac{1-\epsilon}{1+\epsilon} = \left(\frac{4\lambda A + A + 1 + \sqrt{(A+1)^2 + 8\lambda A}}{4\lambda A + A + 1 - \sqrt{(A+1)^2 + 8\lambda A}}\right)\left(\frac{4\lambda + A + 1 + \sqrt{(A+1)^2 + 8\lambda A}}{4\lambda + A + 1 - \sqrt{(A+1)^2 + 8\lambda A}}\right)^{1/A}. \tag{42}$$

Since the left hand side of (42) is less than 1, it follows that

$$4\lambda A + A + 1 - \sqrt{(A+1)^2 + 8\lambda A} \leq 4\lambda A + A + 1 + \sqrt{(A+1)^2 + 8\lambda} A \leq 0,$$
$$4\lambda + A + 1 - \sqrt{(A+1)^2 + 8\lambda A} \leq 4\lambda + A + 1 + \sqrt{(A+1)^2 + 8\lambda} A \leq 0.$$

This implies that $\lambda(1+2\lambda) \geq 0$. Thus, it is necessary that

$$\lambda \leq -\frac{1}{2},\ 0 \leq \lambda.$$

Further, for $\sqrt{(A+1)^2 + 8\lambda A}$ to be real number,

$$\lambda \geq -\frac{(A+1)^2}{8A}, \text{ if } A \geq 0,$$
$$\lambda \leq -\frac{(A+1)^2}{8A}, \text{ otherwise }.$$

Assume $A = \left(1-\frac{2\mu}{\sigma^2}\right) \geq 0$. Then, $\lambda \leq -1/2$, since $\lambda(1-2\mu/\sigma^2) \leq 0$. By elementary but tedious calculations, we can show that the right hand side of (42) is monotone decreasing on λ. Further, it holds that the right hand side of (42) is 0 for $\lambda = -1/2$ which is corresponding to the case of $\epsilon = 1$ and the right hand side of (42) is 1 for $\lambda = -\frac{(A+1)^2}{8A}$ which is corresponding to $\epsilon = 0$.

Next, we consider the case $A = \left(1 - \frac{2\mu}{\sigma^2}\right) \leq 0$. Then, it follows that $\lambda \geq 0$. We can also show that the right hand side of (42) is monotone increasing. When $\lambda = -\frac{(A+1)^2}{8A}$, the right hand side of (42) is 1 and when $\lambda = 0$, the right hand side of (42) is $-\infty$. In summary, λ is given as a mapping such that $\epsilon : [0,1] \to \lambda : [-(A+1)^2/8A, -1/2]$ when $A = \left(1 - \frac{2\mu}{\sigma^2}\right) \geq 0$ and $\epsilon : [0,1] \to \lambda : [0, -(A+1)^2/8A]$ when $A = \left(1 - \frac{2\mu}{\sigma^2}\right) \leq 0$.

Using the above results, we can calculate

$$\begin{aligned}
\frac{d\langle \tilde{\pi}, \tilde{A}^I \rangle_t}{\tilde{\pi}_t \tilde{A}^I_t} =& (1 - \tilde{\pi}_t)\tilde{\pi}_t \frac{d\langle \tilde{S} \rangle_t}{\tilde{S}^2_t} \\
\frac{d\langle \tilde{\pi}, \tilde{S} \rangle_t}{\tilde{\pi}_t \tilde{S}_t} =& (1 - \tilde{\pi}_t) \frac{d\langle \tilde{S} \rangle_t}{\tilde{S}^2_t} \\
\frac{d\langle \tilde{A}^I, \tilde{S} \rangle_t}{\tilde{A}^I_t \tilde{S}_t} =& \tilde{\pi}_t \frac{d\langle \tilde{S} \rangle_t}{\tilde{S}^2_t}.
\end{aligned}$$

Substituting them into (41), it follows that

$$\begin{aligned}
\frac{d\varphi_t}{\varphi_t} =& (1 - \tilde{\pi}_t)\frac{d\tilde{S}_t}{\tilde{S}_t} - \tilde{\pi}_t(1 - \tilde{\pi}_t)\frac{d\langle \tilde{S} \rangle_t}{\tilde{S}^2_t} + (1 - \tilde{\pi}_t)d\Phi_t - (1 - \tilde{\pi}_t)d\Psi_t + \tilde{\pi}_t \frac{d\tilde{S}_t}{\tilde{S}_t} \\
& - \frac{d\tilde{S}_t}{\tilde{S}_t} + \frac{d\langle \tilde{S} \rangle_t}{\tilde{S}^2_t} + (1 - \tilde{\pi}_t)\tilde{\pi}_t \frac{d\langle \tilde{S} \rangle_t}{\tilde{S}^2_t} - (1 - \tilde{\pi}_t)\frac{d\langle \tilde{S} \rangle_t}{\tilde{S}^2_t} - \tilde{\pi}_t \frac{d\langle \tilde{S} \rangle_t}{\tilde{S}^2_t} \\
=& (1 - \tilde{\pi}_t)d\Phi_t - (1 - \tilde{\pi}_t)d\Psi_t.
\end{aligned} \tag{43}$$

This implies that the absolute change of the trading strategy φ_t is given by

$$\frac{d\|\varphi_t\|}{\varphi_t} = (1 - \tilde{\pi}_t)d\Phi_t + (1 - \tilde{\pi}_t)d\Psi_t. \tag{44}$$

Note that Φ and Ψ increase only on $\{\beta = \beta_-\}$ and $\{\beta = \beta_+\}$, respectively, it follows that

$$\frac{d\|\varphi_t\|}{\varphi_t} = \frac{e^{-\beta_-}}{1 + e^{-\beta_-}} d\Phi_t + \frac{e^{-\beta_+}}{1 + e^{-\beta_+}} d\Psi_t.$$

Similarly, it holds that

$$\frac{d\varphi_t}{\varphi_t} = \frac{e^{-\beta_-}}{1 + e^{-\beta_-}} d\Phi_t - \frac{e^{-\beta_+}}{1 + e^{-\beta_+}} d\Psi_t.$$

A.3 Optimal transaction costs

The trading volume given by (44) is important for deriving the optimal transaction costs of the market maker. However, the local time representation is not good for tractability. Thus, we consider the long term average of the growth rate of the trading volume; i.e., define $\bar{R}$ as follows[10],

$$\bar{R} := \lim_{t\to\infty} \frac{1}{t} \int_0^t \frac{d\|\varphi_t\|}{\varphi_t}.$$

Lemma C.1. of Gerhold et al. (2014) and (36) and (37) imply that

$$\lim_{t\to\infty} \frac{\int_0^t d\Phi_t}{t} = \frac{b(\beta_-)^2\, v(\beta_-)}{2}, \quad \lim_{t\to\infty} \frac{\int_0^t d\Psi_t}{t} = \frac{b(\beta_+)^2\, v(\beta_+)}{2},$$

where $v(\cdot)$ is the invariant density of β_t.

To derive invariant distribution, we first derive the scale function $s(x)$ (see (5.42) in Chapter 5 of Karatzas and Shreve (1991)) which is given by

$$\begin{aligned} s(x) &= \int_c^x \exp\left(-2\int_c^y \frac{a(z)}{b(z)^2} dz\right) dy \\ &= \frac{(1+e^c)^2}{e^c} \int_c^x \frac{e^y}{(1+e^y)^2} dy, \end{aligned}$$

where we can arbitrary choose a constant $c \in \mathbb{R}$ (see Problem 5.12 in Chapter 5 of Karatzas and Shreve (1991)).

By using the scale function $s(x)$, the speed measure $m(x)$ is given by (see (5.51) of Karatzas and Shreve (1991))

$$m(dx) = 1_{[\beta_-,\beta_+]} \frac{2dx}{b(x)^2 s'(x)} = 1_{[\beta_-,\beta_+]} \frac{2(1-g'(x))^2 e^c (1+e^x)^2 dx}{\sigma^2 (1+e^c)^2 e^x}$$

Since we can arbitrary choose c, we fix $c = \beta_-$ due to the lower bound of y. Thus, it holds that

$$m([\beta_-,\beta_+]) = \int_{\beta_-}^{\beta_+} m(dx) = \frac{2e^{\beta_-}}{\sigma^2(1+e^{\beta_-})^2} \int_{\beta_-}^{\beta_+} \frac{(1-g'(x))^2(1+e^x)^2}{e^x} dx$$

Since the invariant distribution $v(dx)$ is given by the normalizing the speed measure (see Remark 6.19 in Chapter 5 of Karatzas and Shreve (1991)), it follows that

$$v(dx) = \frac{m(dx)}{m([\beta_-,\beta_+])} = \frac{(1-g'(x))^2(1+e^x)^2}{e^x \int_{\beta_-}^{\beta_+} \frac{(1-g'(y))^2(1+e^y)^2}{e^y} dy} dx.$$

[10] Kallsen and Muhle-Karbe (2017) suggest another approximation of the trading volume. This may be a good candidate for the application to our model.

From this, the long term average of the absolute change of the trading volume is given by

$$
\begin{aligned}
\bar{R} =& \frac{e^{-\beta_-}}{1+e^{-\beta_-}} \frac{\sigma^2}{2} \frac{(1+e^{\beta_-})^2}{e^{\beta_-} \int_{\beta_-}^{\beta_+} \frac{(1-g'(y))^2(1+e^y)^2}{e^y} dy} + \frac{e^{-\beta_+}}{1+e^{-\beta_+}} \frac{\sigma^2}{2} \frac{(1+e^{\beta_+})^2}{e^{\beta_+} \int_{\beta_-}^{\beta_+} \frac{(1-g'(y))^2(1+e^y)^2}{e^y} dy} \\
=& \frac{\sigma^2(1+e^{-\beta_-})}{2\int_{\beta_-}^{\beta_+} \frac{(1-g'(y))^2(1+e^y)^2}{e^y} dy} + \frac{\sigma^2(1+e^{-\beta_+})}{2\int_{\beta_-}^{\beta_+} \frac{(1-g'(y))^2(1+e^y)^2}{e^y} dy}. \qquad (45)
\end{aligned}
$$

Note that

$$
\begin{aligned}
\int_{\beta_-}^{\beta_+} \frac{(1-g'(y))^2(1+e^y)^2}{e^y} dy = \int_{\beta_-}^{\beta_+} \frac{e^y}{A^2(1+2\lambda+2\lambda e^y)^2} dy &= -\frac{1}{A^2} \frac{1}{2\lambda(1+2\lambda+2\lambda e^y)} \Bigg|_{\beta_-}^{\beta_+} \\
&= -\frac{1}{A^2} \frac{1}{2\lambda(1+2\lambda+2\lambda e^{\beta_+})} + \frac{1}{A^2} \frac{1}{2\lambda(1+2\lambda+2\lambda e^{\beta_-})} \\
&= \frac{\sqrt{(A+1)^2+8\lambda A}}{2A^2\lambda(1+2\lambda)}.
\end{aligned}
$$

Substituting the above into (45), we have the explicit form of $\bar{R}$; i.e.,

$$
\bar{R} = \frac{\sigma^2 A(A-1)\lambda}{\sqrt{(A+1)^2+8\lambda A}} \geq 0.
$$

By definition of $\bar{R}$, we have the following approximated formula:

$$
d\|\varphi_t\| \approx \varphi_t \bar{R} dt.
$$

Similarly, we can consider the average position of risky assets $\bar{r}$ such that

$$
\bar{r} := \lim_{t\to\infty} \frac{1}{t} \int_0^t \frac{d\varphi_t}{\varphi_t}.
$$

Noting that (43), it follows that

$$
\bar{r} = -\lambda A \sigma^2 \geq 0.
$$

Using $\bar{r}$, the average position of risky assets up to time t is approximately given by

$$
\varphi_t \approx \varphi_0 e^{\bar{r}t}.
$$

We can now derive the solution of (3). According to (1) and (2), safe and risky position of market maker is described as follows:

$$X_T^M = X_0^M + (1+\epsilon)\int_0^{\bar{T}} S_t d\varphi_t^a - (1-\epsilon)\int_0^{\bar{T}} S_t d\varphi_t^b,$$
$$Y_T^M = Y_0^M - \int_0^{\bar{T}} S_t d\varphi_t^a + \int_0^{\bar{T}} S_t d\varphi_t^b + \int_0^{\bar{T}} Y_t^M \frac{dS_t}{S_t}.$$

Thus, the wealth of the market maker is given by

$$\begin{aligned} X_T^M + Y_T^M &= X_0^M + Y_0^M + \int_0^{\bar{T}} Y_t^M \frac{dS_t}{S_t} + \epsilon\int_0^{\bar{T}} S_t d\varphi_t^a + \epsilon\int_0^{\bar{T}} S_t d\varphi_t^b \\ &= X_0^M + Y_0^M + \int_0^{\bar{T}} \varphi_t^M dS_t + \epsilon\int_0^{\bar{T}} S_t d\|\varphi^I\|_t \\ &= X_0^M + Y_0^M + (\varphi_0^M + \varphi_0^I)\int_0^{\bar{T}} dS_t - \int_0^{\bar{T}} \varphi_t^I dS_t + \epsilon\int_0^{\bar{T}} S_t d\|\varphi^I\|_t \\ &= X_0^M + Y_0^M + \varphi_0^M S_{\bar{T}} - \varphi_0^M S_{\bar{T}} - Y_0^M - Y_0^I - \int_0^{\bar{T}} \varphi_t^I dS_t + \epsilon\int_0^{\bar{T}} S_t d\|\varphi^I\|_t \end{aligned}$$

Thus, (3) is described as follows:

$$\begin{aligned} u^M(x^M, y^M, x^I, y^I) &= \max_{\epsilon>0}\mathbb{E}\left[x^M + y^M + (\varphi_0^M + \varphi_0^I)S_{\bar{T}} - y^M - y^I - \int_0^{\bar{T}} \varphi_t^I dS_t + \epsilon\int_0^{\bar{T}} S_t d\|\varphi^I\|_t\right] \\ &= \max_{\epsilon>0}\left(x^M + y^M + (\varphi_0^M + \varphi_0^I)S_0 e^{\mu\bar{T}} - y^M - y^I - \mathbb{E}\left[-\int_0^{\bar{T}} \varphi_t^I dS_t + \epsilon\int_0^{\bar{T}} S_t d\|\varphi^I\|_t\right]\right) \\ &= \max_{\epsilon>0}\left(x^M + (y^M + y^I)e^{\mu\bar{T}} - y^I - \mathbb{E}\left[-\int_0^{\bar{T}} \varphi_t^I dS_t + \epsilon\int_0^{\bar{T}} S_t d\|\varphi^I\|_t\right]\right). \end{aligned}$$

Extracting the argument dependent on ϵ, we need to consider

$$\begin{aligned} \mathbb{E}\left[-\int_0^{\bar{T}} \varphi_t dS_t + \epsilon\int_0^{\bar{T}} S_t d\|\varphi^I\|_t\right] &\approx -\int_0^{\bar{T}} \varphi_0^I \mu S_0 e^{(\bar{r}+\mu)t}dt + \epsilon\int_0^{\bar{T}} S_0 e^{(\bar{r}+\mu)t}\varphi_0^I \bar{R}dt \\ &= \varphi_0^I S_0\left(-\mu + \epsilon\bar{R}\right)\frac{1}{\bar{r}+\mu}\left(e^{(\bar{r}+\mu)\bar{T}} - 1\right). \end{aligned}$$

In summary, our focus is to find the transaction costs maximizing the above argument. Let ϵ^* be such transaction costs and R^*, r^* are corresponding to the ϵ^*. Then, we have the value function of the market maker as follows:

$$u^M(x^M, y^M, y^I) = x^M + y^M e^{\mu\bar{T}} + y^I\left(e^{\mu\bar{T}} - 1 - \frac{\epsilon R^* - \mu}{r^* + \mu}\left(e^{(r^*+\mu)\bar{T}} - 1\right)\right),$$

where we omit the investor's initial safe wealth x^I in arguments of $u^M(\cdot)$.

References

Abdi, F. and A. Ranaldo (2017). A simple estimation of bid-ask spreads from daily close, high and low prices. *The Review of Financial Studies 30*, 4437–4480.

Amihud, Y. and H. Mendelson (1980). Dealership market: Market making with inventory. *Journal of Finance 42*, 533–553.

Back, K. (2017). *Asset Pricing and Portfolio Choice Theory (2nd edition)*. New York: Oxford University Press.

Benedetti, G., L. Campi, and J. Kallsen (2013). On the existence of shadow prices. *Finance and Stochastics 17*(4), 801–818.

Bichuch, M. and R. Sircar (2015). Optimal Investment with Transaction Costs and Stochastic Volatility Part II: Finite Horizon. Preprint.

Bouchard, B. (2000). *Stochastic control and applications in mathematical finance*. Ph.D. dissertation, Université Paris IX.

Bouchard, B. (2002). Utility maximization on the real line under proportional transaction costs. *Finance and Stochastics 6*, 495–516.

Bouchard, B., Y. Kabanov, and N. Touzi (2001). Option pricing by large risk aversion utility under transaction costs. *Decisions in Economics and Finance 24*, 127–136.

Bradfield, J. (1982). Optimal dynamic behavior of a market maker in choosing bid and asked prices. *Journal of Economics and Business 34*, 303–315.

Campi, L. and M. Owen (2011). Multivariate utility maximization with proportional transaction costs. *Finance and Stochastics 15*, 461–499.

Chevalier, E., M. Gaïgi, V. Vath, and M. Mnif (2017). Optimal market dealing under constraints. *Journal of Optimization Theory and Applications 173*, 313–335.

Corwin, S. and P. Schultz (2012). A simple way to estimate bid-ask spreads from daily high and low prices. *The Journal of Finance 67*, 719–759.

Czichowsky, C. (2017). Shadow prices for continuous processes. *Mathematical Finance 27*, 623–658.

Czichowsky, C., R. Peyre, W. Schachermayer, and J. Yang (2018). Shadow prices, fractional Brownian motion, and portfolio optimization under transaction costs. *Finance and Stochastics 22*, 161–180.

Davis, M. and A. Norman (1990). Portfolio selection with transaction costs. *Mathematics of Operations Research 15*, 679–713.

Gerhold, S., P. Guasoni, J. Muhle-Karbe, and W. Schachermayer (2014). Transaction costs, trading volume, and the liquidity premium. *Finance and Stochastics 18*(1), 1–37.

Guasoni, P. and J. Muhle-Karbe (2013). Portfolio choice with transaction costs: a user's guide. In *Paris-Princeton Lectures in Mathematical Finance*.

Ho, T. and H. Stoll (1980). On dealer markets under competition. *Journal of Finance 35*(May), 259–267.

Ho, T. and H. Stoll (1981). Optimal dealer pricing under transactions and return uncertainty. *Journal of Financial Economics 9*(1), 47–73.

Jouini, E. (1995). Martingales and arbitrage in securities markets with transaction costs. *Journal of Economic Theory 66*, 178–197.

Kabanov, Y. and G. Last (2002). Hedging under transaction costs in currency markets: a continuous-time model. *Mathematical Finance 12*, 63–70.

Kallsen, J. and J. Muhle-Karbe (2010). On using shadow prices in portfolio optimization with transaction costs. *The Annals of Applied Probability 20*(4), 1341–1358.

Kallsen, J. and J. Muhle-Karbe (2011). Existence of shadow prices in finite probability spaces. *Mathematical Methods of Operations Research 73*(2), 251–262.

Kallsen, J. and J. Muhle-Karbe (2017). The general structure of optimal investment and consumption with small transaction costs. *Mathematical Finance 27*, 659–703.

Karatzas, I. and S. Shreve (1991). *Brownian Motion and Stochastic Calculus (Second Edition)*. New York: Springer.

Lesmond, D., J. Ogden, and C. Trzcinka (1999). A new estimate of transaction costs. *The Review of Financial Studies 12*, 1113–1141.

Mildenstein, E. and H. Schleef (1983). The optimal pricing policy of a monopolistic market maker in the equity market. *The Journal of Finance* (1), 218–231.

Pham, H. (2014). Optimal high-frequency trading with limit and market orders. *Quantitative Finance 13*(1), 79–94.

Roll, R. (1984). A simple implicit measure of the effective bid-ask spread in an efficient market. *The Journal of Finance 39*, 1127–1139.

Schachermayer, W. (2004). The fundamental theorem of asset pricing under proportional transaction costs in finite discrete time. *Mathematical Finance 14*, 19–48.

Skorokhod, A. (1961). Stochastic equations for diffusion processes in a bounded region. *Theory of Probability and its applications 6*(3), 264 – 274.

イノベーション・マネジメント研究センター
The Research Institute for Innovation Management

本ワーキングペーパーの掲載内容については、著編者が責任を負うものとします。

法政大学イノベーション・マネジメント研究センター
The Research Institute for Innovation Management, HOSEI UNIVERSITY

〒102-8160 東京都千代田区富士見 2-17-1
TEL: 03(3264)9420　FAX: 03(3264)4690
URL: http://riim.ws.hosei.ac.jp
E-mail: cbir@adm.hosei.ac.jp

（非売品）