Essays in Identification and Estimation of Duration Models and Varying Coefficient Models

by

Xiaoqi He

B.S., Wuhan University, 2006
B.A., Wuhan University, 2006

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

The Faculty of Graduate and Postdoctoral Studies

(Economics)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

February 2016

Abstract

Chapter 1 studies the identification of a preemption game where the timing decisions are expressed as mixed hitting time (MHT). It considers a preemption game with private information, where agents choose optimal time to invest, with payoffs driven by Geometric Brownian Motion. The game delivers the optimal timing of investment based on a threshold rule that depends on both the observed covariates and the unobserved heterogeneity. The timing decision rules specify durations before the irreversible investment as the first time the Geometric Brownian Motion hits a heterogeneous threshold, which fits the MHT framework. As a result, identification strategies for MHT can be used for a first stage identification analysis of the model primitives. Estimation results are performed in a Monte Carlo simulation study.

Chapter 2 studies the identification of a real options game similar to chapter 1, but with complete information. Because of the multiple equilibria problems associated with the complete information game, the point identification is only achieved for a duopoly case. This simple complete information game delivers two possible different kinds of equilibria, and we can separate the parameter space of unobserved investment cost accordingly for different equilibria. We also show the non-identification result for a three-player case in appendix B.4.

Chapter 3 studies the estimation of a varying coefficient model without a complete data set. We use a nearest-matching method to combine two incomplete samples to get a complete data set. We demonstrate that the simple local linear estimator of the varying coefficient model using the combined sample is inconsistent and in general the convergence rate is slower than the parametric rate to its probability limit. We propose the bias-corrected estimator and investigate the asymptotic properties. In particular, the bias-corrected estimator attains the parametric convergence rate if the number of matching variables is one. Monte Carlo simulation results are consistent with our findings.

Preface

Chapter 3 is based on work collaborated with Zhu,Yajing, a Ph.D student from Concordia University. I am very grateful to Zhu for allowing me to put this joint work into my thesis. All errors are my own.

Table of Contents

Appendices

List of Tables

List of Figures

Acknowledgements

I am especially grateful to my supervisor Kevin Song for his invaluable advice, unfailing encouragement and extreme patience over my years of study at UBC. I also would like to thank my thesis committee Hiro Kasahara, Vadim Marmer and Paul Schrimpf for their insightful comments and feedback which improved my thesis greatly.

Dedication

To my parents and my husband.

Chapter 1

Identification of Mixed Hitting Time in a Preemption Game

Real options theory tells us that an investment opportunity can be seen as an option to invest and the value of this option can be very large in the case of irreversible investment. For example, to make a real estate investment, a company may wait until the market price is very high to seize this investment opportunity. Therefore, irreversible investments under uncertainty will be delayed compared with the traditional net present value rule. The empirical industrial organization literature also suggests that there is often first mover's advantage in an entry game. For example, the investor may enjoy a period of monopoly benefit by introducing a new product first into the market. However the first investor may also want to delay his investment due to the market uncertainty. Therefore, there is a tradeoff between pre-emption and delay in a real options game. This paper studies such situations theoretically and provides a methodology to identify the game primitives nonparametrically.

Let us consider a game of timing in continuous time. Suppose that there is an investment opportunity available to all the investors, and the investment is irreversible. The payoffs before the investment is normalized to zero while the payoffs after the investment are driven by Geometric Brownian Motion. Each investor observes the value of the payoffs at each instant time and knows his own investment cost as private information at the beginning of the game, and then chooses his optimal time to invest. The optimal decision rule to this game involves a threshold: each player will invest as soon as the Geometric Brownian Motion hits the threshold. In other words, the optimal investment timing is a duration determined by the first time a Geometric Brownian Motion hits a threshold. Hitting time models have been used in the statistic literature (Lee and Whitmore (2006)) to investigate durations determined by optimal threshold decision rules. To study the interactive investment decisions under uncertainty, we will further explore the interactive durations in a preemption game.

This paper studies the identification of a preemption game where the timing decisions are expressed as mixed hitting time (MHT). The MHT models (Abbring (2012)) are mixture duration models where durations are specified as the first time a latent stochastic process crosses a heterogeneous threshold. Our model builds on both the literature of MHT models and the literature investigating preemption games with incomplete information for irreversible investments with real options values. In this context, a symmetric Bayesian-Nash equilibrium is derived and the solution characterizes the optimal timing of investment based on a threshold rule that depends on both the observed covariates and the unobserved heterogeneity. The timing decision rules specify durations before the investment as the first time the Geometric Brownian Motion hitting a heterogeneous threshold, which fits the MHT framework. Therefore, the identification strategies for the MHT model (Abbring (2012)) can be used for the identification of the last player's payoff function in the first stage. In the second stage analysis, we utilize the unique symmetric Bayesian-Nash equilibrium solution to this game and achieve point identification of the payoff function for the next to the last player. Repeating the procedure and we will achieve point identification of all the model primitives. We also present some parametric estimation results from a Monte Carlo simulation. The results are related to the literature on the threshold regression for first-hitting time models, in which the threshold is dependent on covariates. Because the duration before investment in our game follows a mixed inverse gaussian distribution, given the observed covariates and the distribution of the unobserved covariates, the estimation analysis is also connected to the research on the generalized linear mixed regression models.

The paper contributes to the joint framework of real options and preemption games. Unlike most of the real options game in the literature, investors in our model have incomplete information about their competitors' investment costs and we are considering an oligopoly framework in a preemption game. In the real options game literature, a duopoly game has been studied intensively under complete information (see, for instance, Pawlina and Kort (2006); Kong and Kwok (2007)). Recently, a duopoly game under incomplete information has also been presented (Lambrecht and William (2003); Janssens and Kort (2012)). Anderson, Friedman and Oprea (2010) investigate an N players' preemption game under incomplete information about investors' costs to study a stochastic investment opportunity and provide a laboratory experiment to test some hypotheses implied by the theory. Our structural model extends theirs further by considering a more general payoff structure. In our model, all the players make some positive profits from the investment, while the winner taking all of the returns is assumed in their game. Therefore, it is a one-shot move in Anderson, Friedman and Oprea (2010)'s setup, while our model delivers a sequential move of the players.

The paper also adds to the research on the estimation of timing games. Honoré and de Paula (2010) formulate a game theoretic model with complete information where the durations are endogenously determined. In their model, the timing decision by the pioneer will have a positive effect on the other player's utility and both the sequential timing equilibrium and simultaneous timing equilibria could occur under different parameter realizations of the game. Our model shows the existence of a unique symmetric Bayesian Nash equilibrium under incomplete information and allows for point identification and a full solution approach (nested fixed point approach) in the estimation. In another example of timing game, Takahashi (2013) examines the strategic exit decisions in a declining demand period in the U.S. movie theater industry. In his setting, competition resembles a war of attrition and generates a unique sequential exit under incomplete information. In our model, a sequential entry occurs due to the effect of the first-mover advantage in the preemption game and also players' beliefs about their competitors' costs which is revealed over time. Schmidt-Dengler (2006) investigates the timing of technology adoption with a preemption motive. In his model, it is a discrete time dynamic game in which firms decide every period whether to adopt the new technology or not and subgame perfect equilibria are considered. Our model is a continuous time model with a stochastic payoff structure and we investigate the effect of both the preemption incentive and the incomplete information on the investment timing under uncertainty.

Our model provides a game theoretic MHT model. Abbring and Yu (2015) provide another approach and study a two-player optimal stopping game with complete information. Abbring and Yu (2015) use their setup to investigate the peer effect of quit smoking in a couple. Our model study an N-player optimal stopping game with incomplete information and can be used to study an entry competition between oligopolies in an industry. However the estimation method of the game theoretic MHT model we propose relates to threshold regression models. Lee and Whitmore (2006) has reviewed research on first-hitting-time models with regression structures, referred to as threshold regression models for survival data. The threshold regression models make the threshold and sometimes the parameters of the latent process dependent on the covariates through link functions. The reduced-form model from our game structure resembles the threshold regression models and differs from theirs by incorporating agents' heterogeneity into the regressors. Our investment timing model arises naturally from the real options literature, which has been applied in real estate markets, R&D investment, and many other irreversible investment decisions(Clapp, Bardos and Wong (2012), Bulan, Mayer and Somerville (2009),Weeds (2002) and the others).

There are two general strands of literature that are related to this paper's model setup. The literature on irreversible investments under uncertainty states that there is an option value of waiting and it delays investment compared to the traditional net present value (NPV) rule. The NPV rule is calculated as the deterministic discounted cash flow of an investment project and is considered generally incorrect (Dixit and Pindyck (1994)), because it fails to accommodate the ability to delay the irreversible investment when future return is uncertain. The real options approach values the opportunity to invest as a real option when determining the optimal investment timing. And real options theory has been applied both theoretically and empirically in many areas to characterize this negative relationship between investment and uncertainty. Since the option value of waiting to invest in the real world is often strategically interdependent, recent research has incorporated the game theoretic setting into the real options literature and argued that competition erodes the real options value and the fear to be preempted reduces the investment delay(e.g.,Grenadier (1996); Mason and Weeds (2010) etc.). Azevedo and Paxson (2010) review the recent literature on real options game and its applications. Our model studies the real options game with incomplete information and therefore we can further investigate the effect of asymmetric information along with the effect of preemptive advantage on the investment timing.

Our model is a multiple duration model in continuous time. For a given stochastic process, we have multiple observations of the investment timing durations determined by different thresholds. This is different from the competing risk models in survival analysis. Competing risk models study the multiple causes of failure. For an observed duration, it is induced by multiple latent processes competing with each other. While the duration in our model is determined by players' interactive decisions following the same latent process. However, we develop most of our results similar to the identification ideas in the duration analysis and obtain point identification.

The remainder of the chapter is organized as follows. Section 1 presents the economic model. Section 2 shows the main identification results. Section 3 first illustrates the estimation strategies and then conducts small Monte Carlo simulations and examines the cost of incomplete information on the investment timing. Section 4 concludes and shows possible applications of our structural model. All the mathematical derivations and main proofs details are in the appendix.

1.1 The Economic Model

In this section, we analyze a real options game of oligopoly competition with incomplete information. We consider each firm has private information concerning its own investment cost, and has the opportunity to invest in the same market. The investment opportunity can be a development project for each firm to gain entry into a real estate market. We assume the investment opportunity is perpetual and the investment decision is irreversible. Hence the firms choose their optimal investment timing once with the sunk costs. There are no variable costs after investment. Further, we assume that the firms are value maximizing and risk-neutral, that is, the time discount rate is set to be the risk-free rate r.

1.1.1 Setting and Notation

Suppose that there are N firms. Firm i observes its own investment cost C_i, $i \in \{1, ..., N\}$, at the beginning of the game but is uncertain about its rivals' costs. The game starts at $t = 0$ and the time is continuous. At each instant, each firm decides whether to make the investment or not. Before the investment, the firms have an option to wait for their optimal timing of investment and receive no profits. Once it chooses to invest at time t, firm i pays an amount of C_i, $i \in \{1, ..., N\}$, and obtains an instantaneous profit flow of

$$\pi(y_t, D_j(x)) = y_t D_j(x), \quad j \in \{1, ..., N\},$$

where j stands for the total number of firms that have entered the market and the state variable y_t follows a Geometric Brownian Motion with drift parameter μ and standard deviation parameter σ:

$$dy_t = \mu y_t dt + \sigma y_t dW_t,$$

where y_t can be interpreted as a demand shock and parameter μ can be interpreted as the industry growth rate and parameter σ as the industry volatility. We assume the stochastic process starts from initial state y_0 low enough such that the immediate investment is not optimal for any firm.

The deterministic multiplier $D_j(x)$, $j \in \{1, ..., N\}$, reflects the effect of the competition on profits and the state variables x are covariates to describe market conditions and also used for the identification of function $D_j(\cdot)$. Here we want to impose some assumptions on $D_j(x)$ to characterize the negative market externality, that is, when j increases, the market has to be shared by more firms, which means each active firm in the market has lower profits.

The following assumptions are made:

Assumption 1. *If n firms invest at the same instant t, then only one firm is chosen randomly with probability $\frac{1}{n}$, $n \in \{2, ..., N\}$, and invests successfully.*

Assumption 2. $D_1(x) > D_2(x) > ... > D_N(x)$.

Assumption 3. *The values of C_i, $i \in \{1, ..., N\}$, are drawn independently from the distribution function G, with support on $[C_L, C_U]$.*

Assumption 1 implies if more than one firm choose to enter at the same instant t, a randomization device is introduced to avoid the problem of the non-existence of a pure strategy equilibrium[1]. The game proceeds as follows: conditional on how many firms have already invested, each of the remaining firms decides when to invest. Once one of the firms invests, the other remaining firms revise their planned entry time based on the current information, which reflects how many firms are active in the market. Hence a firm's investment timing strategy depends on both the number of the active firms in the market, and his investment cost rank among the remaining firms. The equilibrium policy function suggests that all the firms will invest sequentially according to their investment costs order.

Assumption 2 imposes negative externalities in the model, which means monopoly profits are highest and oligopoly profits are decreasing in the number of the firms in the market.

Assumption 3 implies that all the firms have the identical priors about each other's investment costs. Note in this incomplete information game, each firm has the same information about the state variable, which represents the public market information. While each firm knows its own investment cost, which represents the private firm specific information, it only knows about its rivals' costs up to a prior G.

1.1.2 Solution

We will look for a symmetric Bayesian Nash equilibrium (BNE) of this game. At each instant, suppose the state variables of each player i consist of

[1]This randomization device rules out the possibility that simultaneous investments happen because of coordination failures. See the discussion about how it works in a N-player preemption game on page 374 in Argenziano and Schmidt-Dengler (2014).

its private investment cost C_i and the common state variables $\omega_t = (l, t, h^t)$, where l is the number of firms entered the market, t is the current time, and h^t is the history of the game up to time t, which involves the history of the stochastic state variable y_t up to time t. Given this information, the firm i decides when to enter the market. The optimal entering time T_i is defined as:

$$T_i = \inf\{t \geq 0 : (\omega_t, C_i) \in \boldsymbol{S} \times [C_L, C_U]\},$$

where $\boldsymbol{S}$ is the state space of all the common state variables ω. T_i is also called the optimal stopping time such that firm i maximizes its expected discounted value to enter the market, due to the stochastic nature of its expected discounted value. As a result, the pure timing strategy T_i we consider here is solved as the best response to a stochastic control problem, which is the first time the stochastic process y_t hits a threshold.

Let $V_i(\tau, T_{-i}, \omega, C_i)$ be the present expected discounted value of firm i when i decides to enter at time τ, and the other firms' strategies are represented by T_{-i}, given the state variables are (ω, C_i). Let $h(C_i|h^t)$ be the firm i's belief about its competitors' investment costs when firm i and all its competitors have survived until time t. The Bayesian Nash equilibrium is defined as

DEFINITION: A set of strategies $(\hat{T}_i(\omega, C_i))_{i=1}^N$ with posterior beliefs $h(C_i|h^t)$ is a perfect Bayesian Nash equilibrium if for all $\omega \in \boldsymbol{S}$ and $C_i \in [C_L, C_U]$,

- if for all i and any strategy T_i,

$$V_i(\hat{T}_i, \hat{T}_{-i}, \omega, C_i) \geq V_i(T_i, T_{-i}, \omega, C_i),$$

  and

- if for any i's competitor j, $h(C_j|h^t)$ is given by Bayes' rule.

Since the optimal decision rule to this game involves a threshold: each player will invest as soon as the Geometric Brownian Motion hits the threshold, we will use the timing strategy and the corresponding threshold interchangeably from now on. The optimal threshold function is a function of its own investment cost, above which a firm immediately invests. Based on the investment timing order of the N firms' decisions, we will classify the entry order of the N firms as the first, ..., the kth, ..., the Nth active player in the market. We will also distinguish the identity of the N firms in the subscript accordingly as firm i or its rival j in the context without confusion, $\forall\ i, j \in \{1, 2, ..., N\}$.

Now suppose that firm i enters as the kth active firm in the market, followed by subsequent $N-k$ firms who seize the investment opportunity at time t_j^*, with expected investment threshold y_j^* accordingly, $j \in \{k+1,...,N\}$, $k \in \{1,...,N-1\}$. **Appendix** A.1 shows the derivation of the expected value functions in details. When firm i chooses to invest at the threshold y_i, the expected value function of the kth active firm at time t is

$$\begin{aligned}&\mathbb{E}[V_i^k(y_t \mid y_i, y_{k+1}^*, ..., y_N^*)] \\ &= \left(\frac{y_i D_k(x)}{r-\mu} - C_i\right)\left(\frac{y_t}{y_i}\right)^{\beta} + \sum_{j=k+1}^{N}\left(\frac{y_j^*(D_j(x)-D_{j-1}(x))}{r-\mu}\right)\left(\frac{y_t}{y_j^*}\right)^{\beta},\end{aligned} \tag{1.1}$$

where β is the positive solution to the fundamental quadratic, i.e. $\frac{1}{2}\sigma^2\beta^2 + (\mu - \frac{1}{2}\sigma^2)\beta - r = 0$. Note that the expected value function of the kth active firm, $\mathbb{E}[V_i^k]$, consists of the net present value from its own investment, $\frac{y_i D_k(x)}{r-\mu} - C_i$, discounting back from the time when y_t hits y_i, plus the net present value of a group of dividends caused by its rivals' eventual investments, $\frac{y_j^*(D_j(x)-D_{j-1}(x))}{r-\mu}$, discounting back from the time when y_t hits y_j^* sequentially, $j \in \{k+1,...,N\}$.

Clearly, the threshold functions depend on firms' entry order in this game and each firm has to consider all the possibilities of its entry order in his strategy. Since the Nth entrant in the market has complete information, his investment trigger is given by $f_i^N = \frac{\beta}{\beta-1}\frac{(r-\mu)C_i}{D_N(x)}$ with investment cost C_i. The derivation of f_i^N is in **Appendix** A.1. Hence the strategy of firm i is denoted as $(\{b^k(C_i)\}_{k=1}^{N-1}, f_i^N)$, where the investment threshold functions $b^k(C_i)$ map his own investment cost to the real line, given his entry order k, $k \in \{1,...,N-1\}$. Note that we make the dependence of the threshold functions $\{b^k(C_i)\}_{k=1}^{N-1}$ on the other state variables implicit. Now let us specify the state variables available for each firm thoroughly. At the time t, a firm i will know the number of active firms in the market is l, the current state value is y_t, the highest possible cost for the rivals to invest is $\hat{C}$, which will be specified later and his own investment cost is C_i. Now after the l investments, firm i has to compete against $N-l-1$ firms, and firm i could invest as the kth active firm in the market with investment threshold $b^k(C_i)$, $k \in \{l+1,...,N-1\}$, or the Nth active firm with investment threshold f_i^N. If firm i wants to invest as the kth firm, then he must invest before all his rivals and this implies all his rivals' investment costs are higher than firm i's. Suppose that the probability for firm i to invest as the kth active firm is

$1-H^k(C_i)$, which will be specified later. Then firm i's total expected value function after l investments is

$$\begin{aligned}
&V(y_i, y^*_{-i}, N-l, y_t, \hat{C}, C_i)\\
&= \sum_{k=l+1}^{N-1} \left(1-H^k(C_i)\right) \mathbb{E}[V_i^k(y_t \mid b^k(C_i), y^*_{k+1}, ..., y^*_N)]\\
&+ \left(1 - \sum_{k=l+1}^{N-1} \left(1-H^k(C_i)\right)\right) V_i^N(y_t \mid f_i^N),
\end{aligned}$$

where $y_i = \{b^k(C_i) | k = l+1, ..., N-1\}$ and $y^*_{-i} = \{y^*_{k+1}, ..., y^*_N | k = l+1, ..., N-1\}$. The expected value function as the kth firm is given in equation (1.1). Now we have to specify $1-H^k(C_i)$, the probability for firm i to invest as the kth active firm, $k \in \{l+1, ..., N-1\}$.

Firstly, let us see what can be inferred from the information set at time t. Since each firm has observed a history of state values of y_t until the time t and obtained the information that none of the remaining $N-l$ firms have invested yet. Then we can infer the probability of the kth entry firm after time t, $\forall\ k \in \{l+1, ..., N-1\}$. Let $\hat{y}$ be the highest value of state variable observed until the time t, $\hat{y} = \max_{0 \le \tau \le t}\{y_\tau\}$ and $\hat{C} = (b^k)^{-1}(\hat{y})$ be the highest cost draw that would lead a rival of the kth entry firm to invest at $\hat{y}$. Suppose that firm i is the kth entry firm, $\forall\ k \in \{l+1, ..., N-1\}$. Then the fact that none of firm i's rivals have invested until $\hat{y}$ is reached means their thresholds are larger than $\hat{y}$, then firm i would learn that $C_j > \hat{C}$, $\forall\ j \in \{k+1, ..., N\}$.

The probability that firm i's rival j, has cost C_j higher than C_i, conditional on $C_j > \hat{C}$ is $\frac{1-G(C_i)}{1-G(\hat{C})}$. When firm i enters as the kth firm, implying firm i has $N-k$ rivals to consider, the probability that none of the $N-k$ rivals will preempt is

$$Pr(t_i < t_j,\ \forall\ j \neq i) = \left(\frac{1-G(C_i)}{1-G(\hat{C})}\right)^{N-k} \equiv 1 - H^k(C_i).$$

Denote $b(m) = \{b^k(m) | k = l+1, ..., N-1\}$, $m \in [\hat{C}, C_U]$. Given investment cost C_i, firm i's problem is reduced to find a threshold function $b(m)$

that solves:

$$\begin{aligned}
&\max_{m\in[\hat{C},C_U]} V(b(m), y_{-i}, N-l, y_t, \hat{C}, C_i) \\
&= \max_{m\in[\hat{C},C_U]} \left(\sum_{k=l+1}^{N-1} \Big(1 - H^k(m)\Big) \mathbb{E}[V_i^k(y_t \mid b^k(m), y_{k+1}^*, ..., y_N^*)] \right. \\
&+ \left. \left(1 - \sum_{k=l+1}^{N-1} \Big(1 - H^k(m)\Big)\right) V_i^N(y_t \mid f_i^N) \right),
\end{aligned} \tag{1.2}$$

where $y_{-i} = \{y_{k+1}^*, ..., y_N^* | k = l+1, ..., N-1\}$.

In equilibrium, the investors choose their true type to obtain the BNE strategy function, such that investor i maximizes equation (1.2) when $m = C_i$. The first order condition (FOC) for $b^k(C_i)$ satisfies the following first order differential equation,

$$\frac{h^k(C_i)}{1 - H^k(C_i)} = \frac{\mathbb{E}[V_i^{k\prime}(y_t \mid b^k(C_i), y_{k+1}^*, ..., y_N^*)]}{\mathbb{E}[V_i^k(y_t \mid b^k(C_i), y_{k+1}^*, ..., y_N^*)] - V_i^N(y_t \mid f_i^N)}, \tag{1.3}$$

where $h^k(C_i) \equiv H^{k\prime}(C_i)$. The intuition behind the above FOC reflects that firm i's expected marginal benefits of winning as the kth entrant is equal to its expected marginal costs of waiting to enter the market later. That is, firm i's expected payoff gain when it wins against its rival to enter the kth market first is $\mathbb{E}[V_i^k(y_t \mid b^k(C_i), y_{k+1}^*, ..., y_N^*)] - V_i^N(y_t \mid f_i^N)$, with probability $h^k(C_i)$ that firm i wins at trigger $b^k(C_i)$. When it loses the competition, firm i's expected payoff loss is $\mathbb{E}[V_i^{k\prime}(y_t \mid b^k(C_i), y_{k+1}^*, ..., y_N^*)]$, with probability $1 - H^k(C_i)$ that one of i's competitors invests first, conditional on not investing until $b^k(C_i)$ is reached.

In the symmetric BNE, all the firms use the same threshold functions $\{b^k(\cdot) | k \in \{1, ..., N-1\}\}$. Therefore, the expected investment trigger y_j^*, $j \in \{k+1, ..., N\}$ in the total expected value function (1.2) is given by

$$\begin{aligned}
y_j^* &= \mathbb{E}[b^j(C) \mid C > C_i, C_i] \\
&= \frac{\int_{C_i}^{C_U} b^j(C) g(C) dC}{1 - G(C_i)} \equiv B_j(C_i), \ j \in \{k+1, ..., N-1\}
\end{aligned} \tag{1.4}$$

$$\begin{aligned}
y_N^* &= \mathbb{E}[f_C^N \mid C > C_i, C_i] \\
&= \frac{\beta}{\beta - 1} \frac{(r-\mu)}{D_N(x)} \frac{\int_{C_i}^{C_U} C g(C) dC}{1 - G(C_i)} \equiv B_N(C_i),
\end{aligned} \tag{1.5}$$

where

$$f_C^N = \frac{\beta}{\beta - 1}\frac{(r-\mu)C}{D_N(x)},$$

and f_C^N is the investment trigger of the Nth active firm. From firm i's perspective, when it invests as the kth active firm, the expected trigger y_j^*, $j \in \{k+1, ..., N\}$ is the same across all of i's rivals, which depends on firm i's information set $\{\forall j \in \{k+1, ..., N\} : C_j > C_i\}$ and firm i's prior for all its competitors. Now, we substitute $y_j^* = B_j(C_i)$ into (1.3), $\forall j \in \{k+1, ..., N\}$, and let $y_t = b^k(C_i)$. Then we will get

$$\frac{(N-k)g(C_i)}{1-G(C_i)} = \frac{I_1}{I_2}, \tag{1.6}$$

where

$$I_1 = (1-\beta)\left[\frac{b^{k\prime}(C_i)D_k(x)}{r-\mu} + \sum_{j=k+1}^{N}\left(B_j'(C_i)\left(\frac{b^k(C_i)}{B_j(C_i)}\right)^{\beta}\left(\frac{(D_j(x)-D_{j-1}(x))}{r-\mu}\right)\right)\right] + \beta\frac{b^{k\prime}(C_i)}{b^k(C_i)}C_i$$

and

$$I_2 = \left(\frac{b^k(C_i)D_k(x)}{r-\mu} - C_i\right) + \sum_{j=k+1}^{N}\left(\frac{B_j(C_i)(D_j(x)-D_{j-1}(x))}{r-\mu}\right)\left(\frac{b^k(C_i)}{B_j(C_i)}\right)^{\beta} - \left(\frac{f_{C_i}^N D_N(x)}{r-\mu} - C_i\right)\left(\frac{b^k(C_i)}{f_{C_i}^N}\right)^{\beta},$$

with boundary condition,

$$\frac{b^k(C_U)D_N(x)}{r-\mu} = C_U + \left(\frac{f_{C_U}^N D_N(x)}{r-\mu} - C_U\right)\left(\frac{b^k(C_U)}{f_{C_U}^N}\right)^{\beta}, \tag{1.7}$$

where

$$\begin{aligned} g(C) &\equiv G'(C), \\ f_y^N &= \frac{\beta}{\beta-1}\frac{(r-\mu)y}{D_N(x)}, \quad y \in \{C_i, C_U\}. \end{aligned}$$

To compute the equilibrium of the model, we note that the game proceeds sequentially as if an $(N-1)$-players game is nested in an N-player game. Therefore, we solve the game sequentially starting from the Nth player. Since the Nth active firm enters the market with complete information, we first solve the $(N-1)$th active firm's investment trigger $b^{N-1}(\cdot)$. The derivation of $b^{N-1}(\cdot)$ is illustrated in **Appendix** A.3. Repeating the procedure, the $b^k(\cdot)$ can be expressed in the following recursive formula,

$$\begin{aligned} &\frac{b^k(C_i)D_k(x)}{r-\mu} \\ &= C_i - R^k(C_i) \\ &+ \int_{C_i}^{C_U} \left(\left(\frac{b^k(C_i)}{b^k(y)} \right)^\beta - \left(\frac{b^k(C_i)}{f_y^N} \right)^\beta \right) \left(\frac{1-G(y)}{1-G(C_i)} \right)^{N-k} dy, \end{aligned} \tag{1.8}$$

where

$$\begin{aligned} R^k(C_i) = \sum_{j=k+1}^{N} &\left(\frac{B_j(C_i)(D_j(x)-D_{j-1}(x))}{r-\mu} \right) \left(\frac{b^k(C_i)}{B_j(C_i)} \right)^\beta \\ &- \left(\frac{b^k(C_i)}{f_{C_i}^N} \right)^\beta \left(\frac{f_{C_i}^N D_N(x)}{r-\mu} - C_i \right), \end{aligned}$$

and

$$f_y^N = \frac{\beta}{\beta-1}\frac{(r-\mu)y}{D_N(x)}, \quad f_{C_i}^N = \frac{\beta}{\beta-1}\frac{(r-\mu)C_i}{D_N(x)},$$

given the expected investment triggers of firm i's rivals $\{B_j(\cdot)\}_{j=k+1}^N$ defined in equation $(1.4-1.5)$.

As proved in **Theorem** 1, the above boundary value problem has a unique solution $b^k(\cdot)$, given the expected triggers $\{B_j(\cdot)\}_{j=k+1}^N$. And the investment threshold functions $\{b^k(\cdot)|k \in \{1, ..., N-1\}\}$ characterize our symmetric BNE.

Theorem 1. *Under **Assumptions** 1–3, we have the following conclusions:*

1. *A function $b^{*k}(\cdot)$ satisfies the recursive formula (1.8) if and only if it solves the boundary problem $(1.6-1.7)$, $k \in \{1, ..., N-1\}$.*

2. *Given $\beta > 0$, boundary value problem $(1.6-1.7)$ has a unique solution: $b^{*k}(\cdot) : [C_L, C_U] \longrightarrow R$, $k \in \{1, ..., N-1\}$.*

3. *The solution to the boundary problem $(1.6-1.7)$, $k \in \{1, ..., N-1\}$ constitutes a symmetric Bayesian Nash Equilibrium in which each*

player invests according to the thresholds $(\{b^{*k}(C_i)\}_{k=1}^{N-1}, f_{C_i}^N)$*, with its own investment cost* C_i*. Furthermore,* $\{b^{*k}(C_i)\}_{k=1}^{N-1}$ *are solved sequentially, starting from* $k = N - 1$ *by the recursive formula (1.8).*

4. *The symmetric equilibrium is unique, if it exists.*

Proof. See **Appendix** A.4 for details. ■

To obtain $b^{*k}(\cdot)$ numerically, we iterate (1.8) from an initial approximation using the upper boundary value (1.7), substitute it for $b^{*k}(\cdot)$ on the right hand side of equation (1.8) to obtain a better approximation, and iterate. The symmetric BNE threshold $b^{*k}(\cdot)$ is a fixed point of (1.8). In fact, the recursive formula (1.8) is a nonhomogeneous nonlinear Volterra integral equation. M. Eshaghi Gordji and Baghani (2011) provides the existence and uniqueness conditions for the solution to this type of integral equation and their conditions are applied in the proof of **Theorem** 1. Note the existence of a symmetric Bayesian Nash equilibrium is not proved in the general case. The argument is similarly to the discussions as in Proposition 5 of Takahashi (2013). However, if a symmetric equilibrium of the game exists, then it is the unique symmetric equilibrium as discussed in the Theorem 1 of our paper. For a duopoly case, the existence and uniqueness of the symmetric BNE is proved in Janssens and Kort (2012).

Anderson, Friedman and Oprea (2010) investigate an N players' preemption game under incomplete information with a payoff structure that only the first firm can make a profit. Our structural model considers a more general payoff structure. In our model, all the followers make some positive profits from the investment, while the winner takes all of the returns in their game. As a result, the symmetric BNE threshold function in Anderson, Friedman and Oprea (2010)'s model is consistent with the boundary problem (1.6 − 1.7) for the special case that the payoff of all the followers' after the first entrant is equal to zero. Therefore, Anderson, Friedman and Oprea (2010) use a different method to prove the existence and uniqueness of the symmetric BNE threshold. The boundary problem in Anderson, Friedman and Oprea (2010) is equivalent to an ODE of $b^{*k}(\cdot)$ and the Picard-Lindelof theorem is applied accordingly to prove the symmetric BNE exists and is unique. In our model, the followers capture a positive profit, and affect the first entrant's expected profit through the expected investment triggers $\{B^j(\cdot)\}_{j=k+1}^N$, which is reflected in the boundary problem (1.6 − 1.7). Naturally, we obtain an equivalent recursive integral equation system of $(b^{*k}(\cdot), \{B^j(\cdot)\}_{j=k+1}^N)$ and adopt an alternative way to prove the existence and uniqueness of the symmetric BNE threshold.

1.2 Identification

For the identification and estimation of the preemption game, we observe a random sample of M markets. In each market m, the game is played among N players and we will classify the entry order of the N firms as the first,...,the kth,..., the Nth active player in the market. We will also distinguish the identity of the N firms in the subscript accordingly as firm i or its rival j, $\forall\ i, j \in \{1, 2, ..., N\}$. The identification proceeds in sequential stages. In the first stage, we determine both the parameters of the payoff structure for the Nth player and the distribution function of the unobserved investment costs. Then we apply the first-stage identification results to the second-stage identification analysis of the remaining parameters, which are from the $(N-1)$th player's payoff structure. As shown in **Theorem** 3, the identification strategy for the $(N-1)$th player's payoff structure can be applied to any players j's payoff structure, $j \in \{1, ..., N-2\}$. Therefore, we will focus on the identification of a duopoly game in the following analysis.

Suppose that we observe a random sample $\{(T_m, x_m)\}_{m=1}^{M}$, where subscript $m \in \{1, ..., M\}$ indicates different markets. Let $T_m = \{T_m^{(N-1)}, T_m^{(N)}\}$ be the investment timing of the players according to the entry order in the market m and $x_m = \{x_m^{(N-1)}, x_m^{(N)}\}$ are the market-specific covariates according to the entry order in the market m.[2] The superscript $(N-1)$ and (N) indicate the entry order of the players in each game. For example, $(N-1)$ indicates the first entrant in the duopoly game and also named as the leader in the game and (N) is called the follower. For the simplicity of the notation, we omit the subscript m in the following analysis without confusion and use subscript i to indicate the player i in the market m. When firm i is the Nth entrant in the market m, its investment timing is $T_i^{(N)}$ with investment threshold f_i^N. Define $T_i^{(N)}$ as

$$T_i^{(N)} = \inf\{t \geq 0 : y_t \geq f_i^N\} = \inf\{t \geq 0 : \ln(y_t) \geq \ln(f_i^N)\}.$$

When firm i is the $(N-1)$th entrant, its investment timing is $T_i^{(N-1)}$ with investment trigger $b^{N-1}(C_i)$,

$$T_i^{(N-1)} = \inf\{t \geq 0 : \ln(y_t) \geq \ln(b^{N-1}(C_i))\}.$$

[2]The function $D_j(x)$ in the profit flow, $j \in \{N-1, N\}$ can still be identified even when $x^{(N-1)} = x^{(N)} = x$. For the identification of $D_j(\cdot)$ we only need some variation in the covariates of x.

Since y_t is a Geometric Brownian Motion with instantaneous drift μ and instantaneous standard deviation σ, $\ln(y_t)$ is a standard Brownian Motion with instantaneous drift $\hat{\mu} = \mu - \frac{1}{2}\sigma^2$ and instantaneous standard deviation $\hat{\sigma} = \sigma$. Therefore, $T_i^{(N)}$ is the first time that the standard Brownian Motion $\ln(y_t)$ crosses a threshold $\ln(f_i^N)$. Abbring (2012) shows that although the distribution of $T_i^{(N)}$ cannot be explicitly expressed, the Laplace transform of the distribution of $T_i^{(N)}$ satisfies

$$\begin{aligned} \mathcal{L}_{T_i^{(N)}}(s) &\equiv \mathbb{E}\left[\exp\left(-sT_i^{(N)}\mathbf{1}\left(T_i^{(N)} < \infty\right)\right)\right] \\ &= \mathbb{E}\left[\exp\left(-\Lambda(s)\ln\left(f_i^N\right)\right)\right], \end{aligned} \tag{1.9}$$

where

$$\Lambda(s) = \frac{\sqrt{\hat{\mu}^2 + 2\hat{\sigma}^2 s} - \hat{\mu}}{\hat{\sigma}^2},$$

and $\Lambda(s)$ is the inverse function of the Laplace exponent of the standard Brownian Motion given by the Lévy-Khintchine formula[3].

Suppose that we observe market covariates $x^{(N)}$ when the (N)th player invests in the market and market covariates $x^{(N-1)}$ when the $(N-1)$th investment happens in the market. Note the market covariates only depends on the order of the investment and not depend on the individual firm i. Therefore, we can write the investment trigger of f_i^N as

$$f_i^N = \frac{\beta}{\beta - 1}\frac{(r-\mu)C_i^{(N)}}{D_N(x^{(N)})}, \tag{1.10}$$

where $C_i^{(N)} = \max_{i\in\{1,\ldots,N\}}\{C_1, \ldots, C_N\}$.

Assumption 4. *$D_j(x) : \chi \longrightarrow (0,\infty)$ is such that $D_j(x_1) \neq D_j(x_2)$ for some x_1, $x_2 \in \chi$, $\forall j \in \{1, \ldots, N\}$.*

Assumption 5. *The time discount factor r is known to the econometrician.*

Assumption 6. *Investment costs C_i and C_j are drawn independently from the same distribution function $G(\cdot)$. $G(\cdot)$ is a continuous function with full support on $\mathbb{R}_+^2$. And the marginal density function $g(\cdot)$ satisfies that $0 < g(\cdot) < \infty$ on $\mathbb{R}_+$[4].*

[3]See Abbring (2012) for an illustration of the Lévy-Khintchine formula in the general case of a Lévy process in Section 4.1

[4]Note the unobservables are the sunk costs of each player, so it is in the economic sense that $C_i > 0$, $i \in \{1, 2, \ldots, N\}$. In Abbring and Yu (2015)'s paper, they consider a form of unobservables of the form $C_i = e^{-v_i}$ and for v_i point mass at zero is allowed.

Theorem 2. ***Identification of*** $D_N(\cdot)$. *Under* ***Assumptions*** *1–6, the function* $D_N(\cdot)$ *and parameters* μ *and* σ *are identified up to scale, and the distribution of investment cost* $G(\cdot)$ *is identified up to scale.*

Proof. Define $\phi_N(x) = \ln\left(\frac{\beta}{\beta-1}\frac{(r-\mu)}{D_N(x)}\right)$. Then from equation (1.10), the logarithm of the trigger value for $T_i^{(N)}$ is $\ln f_i^N = \phi_N(x^{(N)}) + \ln(C_i^{(N)})$. By equation (1.9), we have

$$\begin{aligned}\mathcal{L}_{T_i^{(N)}}(s \mid x^{(N)}) &= \mathbb{E}(\exp[-\Lambda(s)(\phi_N(x^{(N)}) + \ln(C_i^{(N)}))] \mid x^{(N)}) \\ &= \mathbb{E}(\exp[-\Lambda(s)\ln(C_i^{(N)})])\exp[-\Lambda(s)\phi_N(x^{(N)})] \\ &= \mathcal{L}_{\ln C_i^{(N)}}(\Lambda(s))\exp[-\Lambda(s)\phi_N(x^{(N)})],\end{aligned}$$

where $\mathcal{L}_{\ln C_i^{(N)}}$ is the Laplace transform of the distribution of $\ln C_i^{(N)}$, and we transform the Geometric Brownian Motion y_t into a standard Brownian Motion $\ln(y_t)$, so that

$$\Lambda(s) = \frac{\sqrt{(\mu - \frac{1}{2}\sigma^2)^2 + 2\sigma^2 s} - (\mu - \frac{1}{2}\sigma^2)}{\sigma^2}.$$

Given two sets of covariates $x_1^{(N)}$ and $x_2^{(N)}$, the conditional distribution of $T_i^{(N)}$ given $x_1^{(N)}$ and that of $T_i^{(N)}$ given $x_2^{(N)}$ are characterized by

$$\mathcal{L}_{x_1^{(N)}}(s) \equiv \mathcal{L}_{T_i^{(N)}}(s \mid x_1^{(N)}) = \mathcal{L}_{\ln C_i^{(N)}}(\Lambda(s))\exp[-\Lambda(s)\phi_N(x_1^{(N)})],$$

and

$$\mathcal{L}_{x_2^{(N)}}(s) \equiv \mathcal{L}_{T_i^{(N)}}(s \mid x_2^{(N)}) = \mathcal{L}_{\ln C_i^{(N)}}(\Lambda(s))\exp[-\Lambda(s)\phi_N(x_2^{(N)})].$$

Note

$$\begin{aligned}\frac{\mathcal{L}_{x_1^{(N)}}(s)}{\mathcal{L}_{x_2^{(N)}}(s)} &= \exp\left[-\Lambda(s)(\phi_N(x_1^{(N)}) - \phi_N(x_2^{(N)}))\right] \\ &= \exp\left[-\Lambda(s)\ln\left(\frac{D_N(x_2^{(N)})}{D_N(x_1^{(N)})}\right)\right].\end{aligned}$$

When $\mathcal{L}_{x_1^{(N)}}(s)/\mathcal{L}_{x_2^{(N)}}(s) \neq 0$, we can write

$$\frac{\left(\mathcal{L}_{x_1^{(N)}}(s)/\mathcal{L}_{x_2^{(N)}}(s)\right)'}{\left(\mathcal{L}_{x_1^{(N)}}(s)/\mathcal{L}_{x_2^{(N)}}(s)\right)} = -\Lambda'(s)\ln\left(\frac{D_N(x_2^{(N)})}{D_N(x_1^{(N)})}\right). \tag{1.11}$$

Since

$$\lim_{s\to 0}\Lambda'(s) = \left(\mu - \frac{1}{2}\sigma^2\right)^{-1},$$

so

$$\lim_{s\to 0}\frac{\left(\mathcal{L}_{x_1^{(N)}}(s)/\mathcal{L}_{x_2^{(N)}}(s)\right)'}{\left(\mathcal{L}_{x_1^{(N)}}(s)/\mathcal{L}_{x_2^{(N)}}(s)\right)} = -\left(\mu - \frac{1}{2}\sigma^2\right)^{-1}\ln\left(\frac{D_N(x_2^{(N)})}{D_N(x_1^{(N)})}\right). \tag{1.12}$$

Note the LHS of equation (1.12) is not degenerate. To see this, we rewrite it using the definition of the Laplace transform,

$$\begin{aligned}
&\lim_{s\to 0}\frac{\left(\mathcal{L}_{x_1^{(N)}}(s)/\mathcal{L}_{x_2^{(N)}}(s)\right)'}{\left(\mathcal{L}_{x_1^{(N)}}(s)/\mathcal{L}_{x_2^{(N)}}(s)\right)}\\
&= \lim_{s\to 0}\frac{\left(\mathbb{E}(\exp(-sT^{(N)}) \mid x_1^N)/\mathbb{E}(\exp(-sT^{(N)}) \mid x_2^N)\right)'}{\left(\mathbb{E}(\exp(-sT^{(N)}) \mid x_1^N)/\mathbb{E}(\exp(-sT^{(N)}) \mid x_2^N)\right)}\\
&= \mathbb{E}(-T^{(N)} \mid x_1^N) - \mathbb{E}(-T^{(N)} \mid x_2^N).
\end{aligned}$$

Next we substitute equation (1.12) into

$$\frac{\mathcal{L}_{x_1^{(N)}}(s)}{\mathcal{L}_{x_2^{(N)}}(s)} = \exp[-\Lambda(s)\ln\frac{D_N(x_2^{(N)})}{D_N(x_1^{(N)})}],$$

then

$$\frac{\mathcal{L}_{x_1^{(N)}}(s)}{\mathcal{L}_{x_2^{(N)}}(s)} = \exp[\Lambda(s)\left(\mu - \frac{1}{2}\sigma^2\right)\lim_{s\to 0}\frac{\left(\mathcal{L}_{x_1^{(N)}}(s)/\mathcal{L}_{x_2^{(N)}}(s)\right)'}{\left(\mathcal{L}_{x_1^{(N)}}(s)/\mathcal{L}_{x_2^{(N)}}(s)\right)}], \tag{1.13}$$

so μ and σ in $\Lambda(s)$ are identified up to scale. Once $\Lambda(s)$ is known, so is $D_N(\cdot)$ identified up to scale from equation (1.11).

Given μ and σ are identified up to scale and the time discount factor r

is known, β is identified up to scale as well. So $\phi_N(x)$ is identified up to scale. Lastly,

$$\mathcal{L}_{x_1^{(N)}}(s) = \mathcal{L}_{\ln C_i^{(N)}}(\Lambda(s)) \exp[-\Lambda(s)\phi_N(x_1^{(N)})],$$

therefore, $\mathcal{L}_{\ln C_i^{(N)}}(s)$, is identified and so are the distributions of $\ln C_i^{(N)}$ and $C_i^{(N)}$. Suppose $g_{C_i^{(N)}}(x)$ is the density function of $C_i^{(N)}$. Since $C_i^{(N)} = \max_{i\in\{1,...,N\}}\{C_1, ..., C_N\}$ and $C_1, ..., C_N$ are i.i.d., then we can identify the distribution of C_i from

$$g_{C_i^{(N)}}(x) = Ng(x)G(x)^{N-1}.$$

Because the investment costs are drawn from the same distribution $G(\cdot)$, we can use the distribution of $\ln C_i$, $i \in \{1, 2\}$, to identify $G(\cdot)$. ■

Theorem 3. ***Identification of*** $D_{N-1}(\cdot)$. *Under* ***Assumptions*** *1–6, the function* $D_{N-1}(\cdot)$ *is identified up to scale.*

Proof. Suppose that firm i enters the market as the $(N-1)$th player. Then

$$T_i^{(N-1)} = \inf\{t \geq 0 : \ln(y_t) \geq \ln(b^{N-1}(C_i))\}.$$

The Laplace transform of the distribution of $T_i^{(N-1)}$, conditional on C_i satisfies

$$\begin{aligned}\mathcal{L}_{T_i^{(N-1)}|C_i}(s) &\equiv \mathbb{E}[\exp(-sT_i^{(N-1)}\mathbf{1}(T_i^{(N-1)} < \infty)] \\ &= \exp[-\Lambda(s)\ln(b^{N-1}(C_i))].\end{aligned}$$

The unconditional Laplace transform of the distribution of $T_i^{(N-1)}$ is,

$$\begin{aligned}\mathcal{L}_{T_i^{(N-1)}}(s) &\equiv \mathbb{E}[\exp[-\Lambda(s)\ln(b^{N-1}(C_i))]] \\ &= \mathcal{L}_{\ln(b^{N-1}(C_i))}(\Lambda(s)).\end{aligned}$$

By **Theorem** 2, Λ is identified, so the density of $\ln(b^{N-1}(C_i))$ is known. Denote the conditional quantile function of $\ln(b^{N-1}(C_i))$ by $Q_{\ln(b^{N-1}(C_i))|x}(\tau)$, given x at $\tau \in (0, 1)$. Then the policy function $b^{N-1}(\cdot)$ is identified by

$$b^{N-1}(G_{C_i}^{-1}(\tau) \mid x) = Q_{\ln(b^{N-1}(C_i))|x}(\tau),$$

where $G_{C_i}^{-1}(\cdot)$ is the inverse of the conditional distribution function $G_{C_i}(\cdot)$, which is identified by **Theorem** 2.

Once the policy function $b^{N-1}(\cdot \mid x)$ is known, we plug the policy function into the FOCs $(1.6 - 1.7)$ and the only unknown part of the payoff function D_{N-1} is solved.

Finally, the identification for the oligopoly game can be proved just as the duopoly game. We first identify model primitives from the last player and then identify the market externality multipliers D_j, $j = N-1, ..., 1$ from $(N-1)$th player to the first, each following the method from **Theorem** 3. ■

1.3 Parametric Estimation

In this section, we consider the parametric estimation of a duopoly game. Then we can extend the procedure to an $N-$players' game. For the duopoly game, we first estimate the parameters in the threshold function of the last entrant in the market, and then estimate the remaining parameters of the threshold function of the first entrant in the same market. Since the first hitting time in the case of a standard Brownian Motion process follows the inverse Gaussian distribution, we will parameterize the threshold functions to apply the maximum likelihood estimation. The generalized moment condition estimation method may also be applicable but will be difficult to implement due to the complicated moment conditions in our model. The moment conditions of the durations depend on the threshold policy functions, which are solutions to a system of very complicated differential equations. In the two-player complete information case studied by Abbring and Yu (2015), they suggest that a likelihood-based approach is more promising than GMM and worth investigating in the future.

1.3.1 Estimation Strategies

The estimation strategies follow the arguments of the threshold regression in the case of a standard Brownian Motion associated with its inverse Gaussian first hitting time as in Lee and Whitmore (2006). Consider a standard Brownian Motion process y_t with instantaneous drift μ and instantaneous standard deviation σ. The time T when the process hits the threshold a_0 for the first time has an inverse Gaussian distribution. The density function for the first hitting time is given by,

$$f(t \mid \mu, \sigma, a_0) = \frac{a_0}{\sigma\sqrt{2\pi t^3}} \exp\left(-\frac{(a_0 - \mu t)^2}{2\sigma^2 t}\right).$$

The latent process y_t can be given an arbitrary measurement unit. Therefore one parameter need to be normalized. We set the variance parameter σ to unity. The threshold a_0 is linked to regression covariates that are represented by vector $\mathbf{x} = (1, x_1, ..., x_k)$. In the threshold regression analysis, a logarithmic function

$$\ln(a_0) = \beta\mathbf{x} = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k,$$

is used to link a_0 to the covariates. A sample log-likelihood function will be constructed accordingly to define the maximum likelihood estimators for β and μ.

In our model, the latent process y_t is a Geometric Brownian Motion process with drift parameter $\tilde{\mu}$ and standard deviation $\tilde{\sigma}$. To transform the process into a standard Brownian Motion, we obtain $\ln(y_t)$ with instantaneous drift $\mu = \tilde{\mu} - \frac{1}{2}\tilde{\sigma}^2$ and instantaneous standard deviation $\sigma = \tilde{\sigma}$.

Suppose that firm i is the leader and firm j is the follower. For the leader i, we observe the optimal investment timing $T_i^{(1)}$, threshold $b_1(C_i)$, and market covariates $x^{(1)}$. For the follower j, we observe the optimal investment timing $T_j^{(2)}$, threshold f_j, and market covariates $x^{(2)}$. Define $T_i^{(1)}$ and $T_j^{(2)}$ as

$$T_i^{(1)} = \inf\{t \geq 0 : y_t \geq b_1(C_i)\} = \inf\{t \geq 0 : \ln(y_t) \geq \ln(b_1(C_i))\}, \quad (1.14)$$

and

$$T_j^{(2)} = \inf\{t \geq 0 : y_t \geq f_j\} = \inf\{t \geq 0 : \ln(y_t) \geq \ln(f_j)\}, \quad (1.15)$$

where

$$\ln f_j = \ln\left(\frac{\beta}{\beta - 1}\frac{(r - \mu)}{D_2(x^{(2)})}\right) + \ln(C_j),$$

and $b_1(C_i \mid \beta, \mu, \sigma, D_1(x^{(1)}), D_2(x^{(2)}))$ is defined as the solution to the integral equation (1.8). Note that the cost variables C_i and C_j are unobservable in the threshold functions to the econometrician and are assumed to be independently drawn from the distribution function $G(\cdot)$. Following the threshold regression, we take $D_2(x) = \exp(\beta_0^2 + \beta_1^2 x)$, and $D_1(x) =$

$\exp(\beta_0^1 + \beta_1^1 x)$.

Let $\{(T_m, X_m)\}_{m=1}^M$ be a complete random sample induced from our model at the true parameter values, where $T_m = \{T_{i,m}^{(1)}, T_{j,m}^{(2)}\}$ and $X_m = \{X_m^{(1)}, X_m^{(2)}\}$. The parameters $(\mu, \sigma, \beta_0^1, \beta_1^1, \beta_0^2, \beta_1^2)$ can be uniquely estimated from the conditional distribution of T_1 given X and $T^{(2)}$ given X under proper normalizations (Abbring and Salimans (2012)).

We first estimate structural parameters of the follower j's threshold. Rewrite the threshold function $\ln f_j$ for the follower

$$\begin{aligned}
\ln f_j &= \ln\left(\frac{\beta}{\beta-1}\frac{(r-\mu)}{D_2(x^{(2)})}\right) + \ln(C_j) \\
&= \ln\left(\frac{\beta(r-\mu)}{\beta-1}\right) + \beta_0^2 - \beta_1^2 x^{(2)} + \ln(C_j) \\
&= \phi(\mu) - \beta_1^2 x^{(2)} + \ln(C_j),
\end{aligned}$$

where β_0^2 is normalized to 0, and $\phi(\mu)$ is a function of μ, given r and σ. Since the conditional distribution of $T^{(2)}$ given $X^{(2)}$ and C_j follows an inverse Gaussian distribution with density function

$$\begin{aligned}
f(t_2 \mid x^{(2)}, C_j) = &\frac{\phi(\mu) - \beta_1^2 x^{(2)} + \ln(C_j)}{\sqrt{2\pi t_2^3}} \\
&\times \exp\left(-\frac{(\phi(\mu) - \beta_1^2 x^{(2)} + \ln(C_j) - (\mu - \frac{1}{2})t_2)^2}{2t_2}\right).
\end{aligned} \tag{1.16}$$

The likelihood is computed as

$$l_2(\mu, \beta) = \sum_{m=1}^M \ln \int f(t_{2,m} \mid x_m^{(2)}, C_{j,m})\, dG(C_j),$$

where $t_{2,m}$ is a realization of $T^{(2)}$.

After we obtain the estimators for $\hat{\mu}$ and $\hat{\beta}_1^2$, we will substitute them into the conditional density function of $T^{(1)}$, given $X^{(1)}$ and C_i and estimate the remaining parameters β_0^1 and β_1^1 of the leader i's threshold. The conditional density function for $T^{(1)}$, given $X^{(1)}$ and C_i is,

$$f(t_1 \mid x^{(1)}, x^{(2)}, C_i) = \frac{\ln b(C_i, x^{(1)}, x^{(2)}, \hat{\mu}, \hat{\beta}_1^2, \beta_0^1, \beta_1^1)}{\sqrt{2\pi t_1^3}} \times \exp\left(-\frac{(\ln b(C_i, x^{(1)}, x^{(2)}, \hat{\mu}, \hat{\beta}_1^2, \beta_0^1, \beta_1^1) - (\hat{\mu} - \frac{1}{2})t_1)^2}{2t_1}\right). \tag{1.17}$$

Following Abbring and Salimans (2012), we normalize β_0^1 to 1.[5] Here the policy function $b(C_i, x^{(1)}, \hat{\mu}, \hat{\beta}_1^2)$ is a fixed point for the integral equation (1.8) and the resulting MLE estimation will have a nested fixed point problem to solve, which increase the computational time significantly than the MLE estimation for parameters in the threshold of $T^{(2)}$.

1.3.2 Monte Carlo Simulation

The data-generating process is defined as follows: we simulate a standard Brownian Motion y_t with drift parameter $\mu = 1.5$ and standard deviation $\sigma = 1$ and simulate the first hitting time according to (1.15) and (1.14). To simulate the thresholds in (1.15) and (1.14), we set the time discount $r = 5$ and taking $D_2(x) = \exp(\beta_0^2 + \beta_1^2 x)$ and $D_1(x) = \exp(\beta_0^1 + \beta_1^1 x)$. We set $(\beta_0^1, \beta_1^1, \beta_0^2, \beta_1^2) = (1, 0.3, 0, 0.6)$ and assume that the market specific covariables $x = [x^{(1)}, x^{(2)}]$ are drawn independently from uniform distribution on $[-1, 1]^2$. We will assume C_i and C_j are independent and uniformly distributed on $[50, 80]$ in the parametrization following Anderson, Friedman and Oprea (2010).

To make a comparison, we also estimate a linear approximation to the policy function $b(C_1, x_1, \hat{\mu}, \hat{\beta})$ as the misspecification model. Suppose

$$\ln b(C_i, x^{(1)}, \hat{\mu}, \hat{\beta}_1^2) = \phi(\hat{\mu}) - \ln D_1(x^{(1)}) + \ln(C_i), \tag{1.18}$$

[5] The normalization of the threshold is needed because of the existence of both the covariates and the heterogeneity in the threshold. When the covariates are multiplied by a constant, the heterogeneity can also be multiplied by the reciprocal of the constant without changing the magnitude of the threshold. See the argument given in Abbring and Salimans (2012) on page $8-9$.

Table 1.1: Parametric Estimation

	True Model			Linear Approximation		
	True value	Bias	Std.	True value	Bias	Std.
μ	1.5	0.0164	0.0183			
β_1^2	0.6	0.2456	0.1935			
β_1^1	0.3	0.03	0.1014	0.3	0.0707	0.2126

Notes:σ is normalized to 1;β_0^2 is normalized to 0 while β_0^1 is normalized to 1.The true estimation is based on 100 random samples of size 200 observations each.The linear approximation estimation is based on 100 random samples of size 200 observations each.

then

$$f(t_1 \mid x^{(1)}, x^{(2)}, C_i) = \frac{\phi(\hat{\mu}) - \beta_0^1 - \beta_1^1 x^{(1)} + \ln(C_1)}{\sqrt{2\pi t_1^3}} \times \exp\left(-\frac{(\phi(\hat{\mu}) - \beta_0^1 - \beta_1^1 x^{(1)} + \ln(C_1) - (\hat{\mu} - \frac{1}{2})t_1)^2}{2t_1}\right). \tag{1.19}$$

All the estimation results reported in table 1.1 are based on 100 replications of datasets of size 200. In table 1.1, the estimation of true model is based on (1.16) and (1.17), while the estimation of linear approximation is based on (1.18). Because the estimation proceeds in two stages according to the identification, we first estimate μ and β_1^2 from (1.16) in the true model and then estimate β_1^1 from (1.17). The difference between true model and linear approximation is the estimation of the parameters in the second stage. The estimation of linear approximation uses the estimators of $\hat{\mu}$ and $\hat{\beta}_1^2$ from the first stage, and then applies a misspecified policy function to the density (1.19), which is implied by the leader's strategy in a complete information duopoly game. As expected, the estimation of β_1^1 is more biased compared to the result from the true model.

1.4 Conclusion

In this paper, we study the interactive mixed duration models driven by Brownian Motion. This paper relates to the continuous time duration analysis and also contributes to the real options game literature. We use the identification strategies of MHT (Abbring (2012)) for the first-stage identification analysis of the model primitives and we utilize the unique symmetric Bayesian-Nash equilibrium solution to this game for the second-stage identification analysis of the rest of the model primitives.

There are many applications of real options game in the real estate market. The development of a new apartment is an irreversible investment and it takes time for the developers to start the construction. Therefore uncertainty of the market plays an important role in the investment timing decision for the developers. Also the studies of pioneer survival can demonstrate an interesting application. For a market pioneer, both the uncertainty of the market's response to the new product and the future technology development delay the entry timing of firms. But the possible high profit from the monopoly also induces the preemptive competition among firms. Wang and Xie (2014) investigate both the newspaper industry and the high-tech industries to analyze the first-mover's advantage. They estimate and compare the survival distributions by separating the monopoly period and the competition period. We could show the effect of preemptive incentive on the investment timing directly by examining the survival distributions of the first mover and the second mover. And it would be interesting to further quantify the effect of asymmetric information on the investment timing and compare it with the effect of first-mover advantage.

Chapter 2

Identification of Optimal Hitting Time in a Real Options Game with Complete Information

This chapter studies the identification of a real options game with complete information. As in the first chapter, the investment opportunity considered here is irreversible and the payoffs before the investment is normalized to zero while the payoffs after the investment are driven by Geometric Brownian Motion. Each player observes the value of the payoffs at each instant time and all the players' investment costs are known as public information. The optimal timing decision rule to this game involves a threshold: each player will invest as soon as the Geometric Brownian Motion hits the threshold. In other words, the optimal investment timing is a duration determined by the first time a Geometric Brownian Motion hits a threshold, which fits the mixed hitting time (MHT) model framework. The point identification of the model primitives is achieved in a duopoly game by applying the identification strategies for the MHT model and exploring the equilibria's specification. And non-identification results are provided when there are three or more players.

The analysis of investment decisions in the framework of a real options game has been studied intensively for a duopoly game with complete information. Pawlina and Kort (2006) propose a duopoly game with asymmetric costs and Kong and Kwok (2007) study a duopoly game that is asymmetric both in the costs and in the profits for the duopolies. As is well known, the equilibrium for a complete information game is not unique in a more general set-up, and this complicates the identification problem. However, we can achieve point identification of the primitives of an optimal stopping time game in a duopoly case. Although there are two possible different kinds of equilibria in the duopoly game studied in this chapter, the point identification can still be achieved because the equilibria in this duopoly game have a unique characteristic that the entry order is according to the efficiency order, such that the more efficient firm with a lower entry cost always enters the market first, which helps to separate the different equilibria on the parameter space of the unobserved variables.

However, when we consider a three-player game with complete information, the entry order is not necessarily the same as the efficiency order. In fact, even when the three players have only two different types of cost, namely, low-cost and high-cost, Argenziano and Schmidt-Dengler (2012) show that the high-cost firm could preempt the low-cost firm in the equilibrium. In **Appendix** B.4, the equilibria of a three-player game with three different cost types is derived, and there are three possible different types of equilibria. In this three-player game, we cannot separate the parameter space of the investment cost monotonically according to the cost rank and therefore, we cannot achieve point identification of all the model primitives. And the identification of a complete information game that involves multiple equilibria typically relies on the equilibrium selection mechanism (Bresnahan and Reiss (1991), Berry (1992), Einav (2010) etc). Even under the assumption of the same type of players, a multiple-player investment game with complete information may have an equilibrium that two or more invest simutaneously when the preemption race is intensive among the active players, as shown in Argenziano and Schmidt-Dengler (2014) and Bouis, Huisman and Kort (2009). There are some N-player real options games with complete information delivering a unique equilibrium. For example,Wang and Zhou (2006) have studied an N-player game incorporating both stochastic demand and stochastic construction costs in the real estate market and derived a unique Markov subgame perfect equilibrium. The firms behave in the same way as the symmetric firms as in Grenadier (1996) and all the players achieve the same level of utility in the equilibrium.

The contribution of this chapter is point identification results for a duopoly game and non-identification results for a three-player game. For a duopoly game, we observe two outcomes from the two different kind of equilibria results. From the last entrant's investment timing, since the two players share the same payoff structure and the unobserved investment cost are drawn independently from the same distribution, we can achieve point identification of the payoff structure and the distribution of the unobservables as in the the identification in a single agent problem. And then from the first entrant's ivestment timing, we just have to identify the remaining payoff term only related to the first entrant. However, as discussed in a three-player game, the multiple equilibria results also depend on entry orders. Therefore, now we observe three outcomes from the game, but we have to identify three sets of payoff structure terms and the three possible entry orders as well. Even after normalization, we still cannot achieve point identification in this case.

The remainder of this chapter is organized as follows. Section 1 presents the economic model. Section 2 shows the main identification results. Section 3 concludes and shows possible applications of our structural model. All the mathematical derivations and main proofs details are in the appendix.

2.1 The Economic Model

This section describes Kong and Kwok (2007)'s real options game of a duopoly with asymmetric investment costs, which will be used to study the identification. In **Appendix** B.4, we also study a real options game of three players with three different cost types and show that under complete information, there are three different kinds of equilibria based on the entry order.

2.1.1 A Duopoly with Complete Information

Two risk-neutral firms compete for the optimal development timing of real estate projects. The two developers are assumed to have a perpetual investment opportunity on a development site. The investment decision is irreversible and the sunk costs of investment are asymmetric between the two firms. Before the investment, both firms have the option to wait for their optimal timing of investment and receive a zero profit until the investment. At any point in time, a firm i can choose to invest the amount of K_i, $i \in \{1, 2\}$, and obtain an instantaneous profit flow of

$$\pi(y_t, D_N(x)) = y_t D_N(x).$$

The state variable y_t represents demand shock of real estate market and follows a Geometric Brownian Motion:

$$dy_t = \mu y_t dt + \sigma y_t dW_t,$$

where μ and σ are instantaneous drift parameter and instantaneous standard deviation parameter. dt is the time increment and dW_t is the Wiener increment. Parameter μ can be interpreted as the industry growth rate and parameter σ as the industry volatility. We assume the stochastic process starts from y_0 low enough such that immediate investment is not optimal for either firm.

The deterministic multiplier $D_N(x)$, $N \in \{1, 2\}$, stands for the market externality for each firm i, $i \in \{1, 2\}$, where N indicates the number of active firms in the market. $N = 1$ represents a monopoly market and $N = 2$ a duopoly market. Note that x are the covariates to describe the market conditions and are also useful for the identification. We define the profit adjusted cost for firm i as $\widetilde{K_{iN}} = K_i/D_N(\cdot)$, $i \in \{1, 2\}$. We also denote firm i's rival firm as firm j in the context without confusion. Here we consider the asymmetry between the two firms in the sunk costs.

The following assumptions are made:

Assumption 7. $D_1(\cdot) > D_2(\cdot)$.

Assumption 8. *If* $K_i < K_j$, *then* $\widetilde{K_{iN}} < \widetilde{K_{jN}}$, $N \in \{1, 2\}$, $i \in \{1, 2\}$.

Assumption 7 imposes negative externalities in the model, which means monopoly profits are higher than duopoly profits for both firms. Assumption 8 specifies that if firm i is a lower investment cost firm, then firm i also has the lower profit adjusted cost. Under Assumption 7 and Assumption 8, we will show that the low cost firm is always a leader in the equilibrium of a duopoly game and achieve the identification and estimation results using data on firms' timing decisions T_1 and T_2 and their respective covariates x_1 and x_2. Kong and Kwok (2007) provide an equilibrium analysis of a general case of our model. However, our special case shows an equilibrium scenario of keen competition between the two firms upon their entry into a market and the identification and estimation results can be adapted into other sets of strategic equilibriums accordingly.

2.1.2 Value Functions and Investment Thresholds

In this section, we define value functions of the firms given their possible different strategies. Each firm has three possible choices of investment timing. It could choose to invest first, as the leader or invest after its rival, as the follower, or invest simultaneously with the other firm. As a standard approach to analyze complete information games, we use backward induction to solve for the firms' optimal strategies. We first solve the follower's optimal development timing, given the leader has started the investment. We derive the optimal value functions and investment thresholds similarly as in **Appendix** A.1 and the optimal value function for follower i, $i \in \{1, 2\}$, is given as

$$V_i^F(y_t \mid x_i) = \begin{cases} \left(\frac{f_i D_2(x)}{r-\mu} - K_i\right)\left(\frac{y_t}{f_i}\right)^{\beta} & \text{if} \quad y_t \le f_i \quad \text{(wait)}, \\ \frac{y_t D_2(x)}{r-\mu} - K_i & \text{if} \quad y_t > f_i \quad \text{(invest)}. \end{cases} \tag{2.1}$$

When $y_t \le f_i$, the value function of the follower is the value of the option to wait. It reflects the present value of the follower's investment discounted back by a stochastic discount factor $\frac{y_t}{f_i}$ from the random time of reaching the trigger value f_i. When $y_t > f_i$, the value function of the follower is the net present value of the project when immediate investment is optimal.

Follower i's the investment trigger is given by

$$f_i = \frac{\beta}{\beta - 1} \frac{(r-\mu)K_i}{D_2(x)}, \quad i \in \{1, 2\},$$

where $\beta > 1$ is the positive root of the associated fundamental quadratic (see Dixit and Pindyck (1994) and Stokey (2008))

$$Q(\beta) = \frac{1}{2}\sigma^2 \beta(\beta - 1) + \mu\beta - r = 0.$$

When the leader faces a preemptive threat from its rival, it may choose to invest immediately. Then the value function for firm i when the other firm j chooses to invest at f_j is given by

$$V_i^L(y_t \mid f_j) = \begin{cases} \frac{y_t D_1(x)}{r-\mu} - K_i + \frac{f_j (D_2(x) - D_1(x))}{r-\mu}\left(\frac{y_t}{f_j}\right)^{\beta} & \text{if} \quad y_t \le f_j, \\ \frac{y_t D_2(x)}{r-\mu} - K_i & \text{if} \quad y_t > f_j. \end{cases} \tag{2.2}$$

Here the leader's value function is divided into two parts by the follower's investment trigger f_j. When $y_t \le f_j$, the value function of the leader is the net present value from immediate investment plus the decrease in the expected present value caused by its rival's eventual investment. By comparing the value of investing immediately as the leader with the value of waiting to invest as the follower, we could consider the incentive of a firm to preempt its competitor to be the leader in the game.

However, if the leader i chooses its optimal investment timing without a preemptive threat, then the value function becomes

$$V_i^{LL}(y_t \mid f_j) = \begin{cases} \left(\frac{l_i D_1(x)}{r-\mu} - K_i\right)\left(\frac{y_t}{l_i}\right)^\beta + \frac{f_j(D_2(x)-D_1(x))}{r-\mu}\left(\frac{y_t}{f_j}\right)^\beta & \text{if} \quad y_t \le l_i, \\ \frac{y_t D_1(x)}{r-\mu} - K_i + \frac{f_j(D_2(x)-D_1(x))}{r-\mu}\left(\frac{y_t}{f_j}\right)^\beta & \text{if} \quad l_i < y_t \le f_j, \\ \frac{y_t D_2(x)}{r-\mu} - K_i & \text{if} \quad y_t > f_j, \end{cases} \tag{2.3}$$

where the leader's the investment trigger without preemption is given by

$$l_i = \frac{\beta}{\beta-1}\frac{(r-\mu)K_i}{D_1(x)}, \quad i \in \{1,2\}. \tag{2.4}$$

2.1.3 Preemptive and Sequential Equilibrium

There are two types of equilibria played by the two competing firms, namely the preemptive and sequential equilibrium. The different outcomes of strategic equilibria come from the properties of different investment thresholds. Since a firm in the leader position may face the preemption by its rival, we have to consider the existence of firms' preemptive thresholds first.

As in Pawlina and Kort (2006), we construct the preemptive trigger (or break-even points as in Janssens and Kort (2012)) of each firm as b_i and b_j respectively, at which the firms are indifferent between being a leader and being a follower. Like Kong and Kwok (2007), we define the break-even function $\xi_i(y)$, $i \in \{1,2\}$, as

$$\xi_i(y) \equiv \frac{V_i^L(y) - V_i^F(y)}{K_i}.$$

After substituting the value functions of $V_i^L(y)$ and $V_i^F(y)$, we obtain

$$\xi_i(y) = -\frac{1}{\beta-1}\left[\beta\left(\frac{1}{l_i} - \frac{1}{f_i}\right)\left(\frac{1}{f_j}\right)^{\beta-1} + \left(\frac{1}{f_i}\right)^\beta\right] y^\beta + \frac{\beta}{\beta-1}\frac{1}{l_i}y - 1. \tag{2.5}$$

The preemptive threshold b_i is defined as the lowest value of y such that $\xi_i(y) = 0$, $i \in \{1,2\}$. Therefore, the preemptive trigger is the first time a firm's value function in a leader position exceeds in a follower position. In the following analysis of strategic equilibria, we will conclude that the low-cost firm i is the leader and the high-cost firm j is the follower under the additional assumption that both preemptive thresholds b_i and b_j exist.[6]

Theorem 4. *Under **Assumptions** 7–8 and the assumptions that both preemptive thresholds exist, the preemptive trigger of the low-cost firm b_i, $i \in \{1,2\}$, is always lower than the preemptive trigger of the high-cost firm b_j.*

Proof. See **Appendix** B.1. ■

Under the assumption of negative externalities, the low-cost firm always has a lower preemptive threshold than the high-cost firm, which means the equilibrium of simultaneous entry does not exist. Therefore, the high-cost firm is a natural follower in the duopoly game.

Theorem 5. *Under **Assumptions** 7–8 and the assumptions that both preemptive thresholds exist, the optimal strategy for the low-cost firm i, $i \in \{1,2\}$, is to invest at the trigger of $\min(b_j, l_i)$, where b_j is the preemptive trigger of the high-cost firm j and l_i is the stand-alone trigger in the leader position for firm i.*

Proof. See **Appendix** B.2. ■

When $l_i < b_j$, the low-cost firm $i \in \{1,2\}$, invests at the monopolist's threshold l_i and the high-cost firm j invests at the follower's threshold f_j, which we call the sequential equilibrium. On the other hand, the preemptive equilibrium is played when the low-cost firm $i \in \{1,2\}$, chooses to invest at b_j.

Theorem 6. *Suppose that firm i is the low-cost firm and firm j is the high-cost firm. Under **Assumptions** 7–8 and the assumptions that both preemptive thresholds exist.*

1. *When we have*

$$\beta\left(\frac{l_i}{l_j} - \frac{\beta-1}{\beta}\right) < \left[\beta\left(\frac{1}{l_j} - \frac{1}{f_j}\right)\left(\frac{1}{f_i}\right)^{\beta-1} + \left(\frac{1}{f_j}\right)^{\beta}\right] l_i^{\beta} < \frac{l_i}{l_j}, \quad (2.6)$$

   then the sequential equilibrium (l_i, f_j) is played.

[6] Kong and Kwok (2007) discuss different cases of the existence conditions of preemptive thresholds in the Proposition 1 of their paper. To apply their results in our model, b_i exists if $\frac{f_i}{f_j} < 1$; both b_i and b_j exist if $\frac{1}{q(f_i/f_j)} < \frac{f_i}{f_j} < 1$, where $q(x) = (\frac{x^{\beta}-1}{\beta(x-1)})^{\frac{1}{\beta-1}}$.

2. When we have

$$\frac{\beta}{\beta-1}\frac{l_i}{l_j}-1>\frac{1}{\beta-1}\left[\beta\left(\frac{1}{l_j}-\frac{1}{f_j}\right)\left(\frac{1}{f_i}\right)^{\beta-1}+\left(\frac{1}{f_j}\right)^{\beta}\right]l_i^{\beta}, \quad (2.7)$$

then the preemptive equilibrium (b_j, f_j) *is played.*

Proof. See Appendix B.3. ■

Therefore, we can separate the sequential equilibrium from the preemptive equilibrium in terms of the parameters when both the break-even thresholds b_i and b_j exist.

2.2 Identification

In this section, we will investigate how to identify the duopoly model by the duration data. Suppose that we have data on both the optimal timing $\{(T_{i1}, T_{i2})\}_{i=1}^{N}$ and covariates $\{(x_{i1}, x_{i2})\}_{i=1}^{N}$, where subscripts $i1$ and $i2$ refer to firm 1 and firm 2 respectively in the market i. The cost pairs $\{(K_{i1}, K_{i2})\}_{i=1}^{N}$ are unobservable to the econometrician. The optimal stopping time T_{i1} and T_{i2} can be seen as the mixed hitting time defined in Abbring (2012) and therefore we can achieve the identification by applying the identification strategies in Abbring (2012) and utilizing the probability of the events $T_{i1} < T_{i2}$ and $T_{i1} > T_{i2}$ as in Honoré and de Paula (2010).

Assumption 9. *K_{i1} and K_{i2} are jointly distributed with cdf $G(\cdot,\cdot)$. $G(\cdot,\cdot)$ is a continuous function with full support on $\mathbb{R}_+^2$.*

Assumption 10. *For each $i \in \{1, ..., N\}$ and each $j \in \{1, 2\}$, at least one component of x_{ij}, say $x_{ij,k}$, has a support that contains a nonempty open subset of $\mathbb{R}$.*

Assumption 11. *The time discount factor r is known.*

Assumption 12. *Investment costs K_{i1} and K_{i2} are independent and identically distributed, for each $i \in \{1, ..., N\}$.*

Take the event $T_{i1} < T_{i2}$ for example, which means firm $i1$ is the leader and firm $i2$ is the follower, $i \in \{1, ..., N\}$. From the results in **Theorem** 5, the optimal trigger values for firm $i1$ are defined as l_{i1} in the sequential equilibrium and b_{i2} in the preemption equilibrium, while the optimal trigger for firm $i2$ is denoted as f_{i2}, $i \in \{1, ..., N\}$. The optimal timing of the follower, T_{i2}, $i \in \{1, ..., N\}$, is given as

$$T_{i2} = \inf\{t \geq 0 : y_t \geq f_{i2}\} = \inf\{t \geq 0 : \ln(y_t) \geq \ln(f_{i2})\}.$$

And the optimal timing of the leader, T_{i1}, $i \in \{1, ..., N\}$, is defined as,

$$T_{i1} = \inf\{t \geq 0 : y_t \geq \min(l_{i1}, b_{i2})\} = \inf\{t \geq 0 : \ln(y_t) \geq \min(\ln(l_{i1}), \ln(b_{i2}))\}.$$

The identification is divided into two parts. We first apply the identification strategies in Abbring (2012) to the follower's timing duration T_{i2}, $i \in \{1, ..., N\}$. Since y_t is a Geometric Brownian Motion with instantaneous drift μ and instantaneous standard deviation σ, $\ln(y_t)$ is a standard Brownian Motion with instantaneous drift $\hat{\mu} = \mu - \frac{1}{2}\sigma^2$ and instantaneous standard deviation $\hat{\sigma} = \sigma$. Therefore, T_{i2} is the first time that the standard Brownian Motion $\ln(y_t)$ crosses a threshold $\ln(f_{i2})$. Abbring (2012) shows that although the distribution of T_{i2} cannot be explicitly expressed, the Laplace transform of the distribution of T_{i2} satisfies,

$$\begin{aligned}\mathcal{L}_{T_{i2}}(s) &\equiv \mathbb{E}[\exp(-sT_{i2}\mathbf{1}(T_{i2} < \infty)] \\ &= \exp[-\Lambda(s)\ln(f_{i2})],\end{aligned}$$

where

$$\Lambda(s) = \frac{\sqrt{\hat{\mu}^2 + 2\hat{\sigma}^2 s} - \hat{\mu}}{\hat{\sigma}^2},$$

and $\Lambda(s)$ is an inverse function of the Laplace exponent of the standard Brownian Motion given by the Lévy-Khintchine formula.

Theorem 7. ***Identification of*** $D_2(\cdot)$. *Under* ***Assumptions*** *7–11, the function* $D_2(\cdot)$ *and parameters* μ *and* σ *are identified up to scale. Furthermore, with additional Assumption 12, both the marginal distribution functions* G_1 *and* G_2 *are identified up to scale.*

Proof. When $T_{i1} < T_{i2}$, the logarithm of the trigger value for T_{i2} is $\ln f_{i2} = \ln\left(\frac{\beta}{\beta-1}\frac{(r-\mu)}{D_2(x_{i2})}\right) + \ln(K_{i2})$. Let $\phi_2(x) = \ln\left(\frac{\beta}{\beta-1}\frac{(r-\mu)}{D_2(x)}\right)$. Then

$$\begin{aligned}\mathcal{L}_{T_{i2}}(s \mid x_{i2}, T_{i1} < T_{i2}) &= \mathbb{E}(\exp[-\Lambda(s)(\phi_2(x_{i2}) + \ln(K_{i2}))] \mid x_{i2}, T_{i1} < T_{i2}) \\ &= \mathbb{E}(\exp[-\Lambda(s)\ln(K_{i2})] \mid T_{i1} < T_{i2})\exp[-\Lambda(s)\phi_2(x_{i2})] \\ &= \mathcal{L}_{\ln K_{i2}|T_{i1}<T_{i2}}(\Lambda(s))\exp[-\Lambda(s)\phi_2(x_{i2})],\end{aligned}$$

where $\mathcal{L}_{\ln K_{i2}|T_{i1}<T_{i2}}$ is the Laplace transform of the distribution $\ln K_{i2}$ conditional on the event $T_{i1} < T_{i2}$, and we transform the Geometric Brownian

Motion y_t into a standard Brownian Motion $\ln(y_t)$, so that

$$\Lambda(s) = \frac{\sqrt{(\mu - \frac{1}{2}\sigma^2)^2 + 2\sigma^2 s} - (\mu - \frac{1}{2}\sigma^2)}{\sigma^2}.$$

Given two sets of covariates $x_{i2,1}$ and $x_{i2,2}$ for firm $i2$, the conditional distribution of T_{i2} given $x_{i2,1}$ and $x_{i2,2}$ respectively are characterized by

$$\begin{aligned}\mathcal{L}_{x_{i2,1}|T_{i1}<T_{i2}}(s) &\equiv \mathcal{L}_{T_{i2}}(s \mid x_{i2,1}, T_{i1} < T_{i2}) \\ &= \mathcal{L}_{\ln K_{i2}|T_{i1}<T_{i2}}(\Lambda(s)) \exp[\Lambda(s)\phi_2(x_{i2,1})],\end{aligned}$$

and

$$\begin{aligned}\mathcal{L}_{x_{i2,2}|T_{i1}<T_{i2}}(s) &\equiv \mathcal{L}_{T_{i2}}(s \mid x_{i2,2}, T_{i1} < T_{i2}) \\ &= \mathcal{L}_{\ln K_{i2},T_{i1}<T_{i2}}(\Lambda(s)) \exp[-\Lambda(s)\phi_2(x_{i2,2})].\end{aligned}$$

Note

$$\begin{aligned}\frac{\mathcal{L}_{x_{i2,1}|T_{i1}<T_{i2}}(s)}{\mathcal{L}_{x_{i2,2}|T_{i1}<T_{i2}}(s)} &= \exp\left[-\Lambda(s)(\phi_2(x_{i2,1}) - \phi_2(x_{i2,2}))\right] \\ &= \exp\left[-\Lambda(s) \ln \frac{D_2(x_{i2,2})}{D_2(x_{i2,1})}\right].\end{aligned}$$

When $\mathcal{L}_{x_{i2,1}|T_{i1}<T_{i2}}(s)/\mathcal{L}_{x_{i2,2}|T_{i1}<T_{i2}}(s) \neq 0$, we can write

$$\frac{\left(\mathcal{L}_{x_{i2,1}|T_{i1}<T_{i2}}(s)/\mathcal{L}_{x_{i2,2}|T_{i1}<T_{i2}}(s)\right)'}{\left(\mathcal{L}_{x_{i2,1}|T_{i1}<T_{i2}}(s)/\mathcal{L}_{x_{i2,2}|T_{i1}<T_{i2}}(s)\right)} = -\Lambda'(s) \ln\left(\frac{D_2(x_{i2,2})}{D_2(x_{i2,1})}\right). \tag{2.8}$$

Note

$$\lim_{s\to 0} \Lambda'(s) = \left(\mu - \frac{1}{2}\sigma^2\right)^{-1}.$$

So

$$\lim_{s\to 0} \frac{\left(\mathcal{L}_{x_{i2,1}|T_{i1}<T_{i2}}(s)/\mathcal{L}_{x_{i2,2}|T_{i1}<T_{i2}}(s)\right)'}{\left(\mathcal{L}_{x_{i2,1}|T_{i1}<T_{i2}}(s)/\mathcal{L}_{x_{i2,2}|T_{i1}<T_{i2}}(s)\right)} = -\left(\mu - \frac{1}{2}\sigma^2\right)^{-1} \ln\left(\frac{D_2(x_{i2,2})}{D_2(x_{i2,1})}\right). \tag{2.9}$$

Note the LHS of equation (2.9) is not degenerate as discussed in Chapter 1.

Next we substitute equation (2.9) into

$$\frac{\mathcal{L}_{x_{i2,1}|T_{i1}<T_{i2}}(s)}{\mathcal{L}_{x_{i2,2}|T_{i1}<T_{i2}}(s)} = \exp\left(-\Lambda(s)\ln\left(\frac{D_2(x_{i2,2})}{D_2(x_{i2,1})}\right)\right).$$

Hence μ and σ in $\Lambda(s)$ are identified up to scale. Once $\Lambda(s)$ is known, so is $D_2(\cdot)$ identified up to scale from equation (2.8).

Given μ and σ are identified up to scale and time discount factor r is known, β is identified up to scale as well. Hence $\phi_2(x)$ is identified up to scale. Lastly,

$$\mathcal{L}_{x_{i2,1}|T_{i1}<T_{i2}}(s) = \mathcal{L}_{\ln K_{i2}|T_{i1}<T_{i2}}(\Lambda(s))\exp\left(-\Lambda(s)\phi_2(x_{i2,1})\right).$$

therefore $\mathcal{L}_{\ln K_{i2}|T_{i1}<T_{i2}}(s)$ is identified and so is the distribution of $\ln K_{i2}$, $i \in \{1, ..., N\}$. ■

Theorem 8. ***Identification of*** $D_1(\cdot)$. *Under* ***Assumptions*** *7–12, the function* $D_1(\cdot)$ *is identified up to scale.*

Proof. We consider the event $T_{i1} < T_{i2}$ and apply the equilibrium separation conditions we discussed in section 2.1 to achieve the identification of $D_1(\cdot)$. From **Theorem** 5, we know $i1$ would choose a strategy $min\{l_{i1}, b_{i2}\}$. let K_{i2}^S be the investment cost of firm $i2$ such that $\xi_{i2}(l_{i1}; K_{i2}^S, K_{i1}) = 0$. By **Proposition** 5.2 in Janssens and Kort (2012), we know that b_{i2} is an increasing function of K_{i2}. Therefore, when $K_{i1} < K_{i2} < K_{i2}^S$, firm $i1$ would face an preemption threat from firm $i2$, and therefore choose to invest at b_{i2}. When $K_{i2} > K_{i2}^S$, either b_{i2} does not exist as the solution to firm $i2$'s break-even function $\xi_{i2}(\cdot)$ or $l_{i1} < b_{i2}$ holds, therefore, firm $i1$ always chooses l_{i1} to invest. We apply this property to complete the identification.

Firstly,

$$\begin{aligned}
&Pr(T_{i1} < T_{i2} \mid x_{i1}, x_{i2}) \\
&= Pr(K_{i1} < K_{i2} < K_{i2}^S \mid x_{i1}, x_{i2}) + Pr(K_{i1} < K_{i2}^S < K_{i2} \mid x_{i1}, x_{i2}) \\
&= Pr(K_{i1} < \min(K_{i2}, K_{i2}^S) \mid x_{i1}, x_{i2}).
\end{aligned}$$

Then, we consider the following conditional probability

$$
\begin{aligned}
&Pr(T_{i1} < T_{i2} \mid x_{i1}, x_{i2}, K_{i1} < K_{i2}) \\
&= Pr(K_{i1} < K^S_{i2} \mid x_{i1}, x_{i2}, K_{i1} < K_{i2}) \\
&= Pr(K_{i1} < (1 - (\frac{D_2(x_{i2})K_{i1}}{D_1(x_{i1})K_{i2}})^{\beta})^{-1}(\frac{\beta}{\beta - 1})(K_{i1} + \\
&(\frac{D_2(x_{i2}) - D_1(x_{i2})}{D_2(x_{i1})})(\frac{D_2(x_{i1})}{D_1(x_{i1})})^{\beta} K_{i1} - K_{i2}(\frac{D_2(x_{i2})K_{i1}}{D_1(x_{i1})K_{i2}})^{\beta}) \mid x_{i1}, x_{i2}, K_{i1} < K_{i2}) \\
&= Pr(1 - (\frac{D_2(x_{i2})K_{i1}}{D_1(x_{i1})K_{i2}})^{\beta} < (\frac{\beta}{\beta - 1})(1 + (\frac{D_2(x_{i2}) - D_1(x_{i2})}{D_2(x_{i1})})(\frac{D_2(x_{i1})}{D_1(x_{i1})})^{\beta}) \\
&- (\frac{\beta}{\beta - 1})(\frac{D_2(x_{i2})}{D_1(x_{i1})})^{\beta}(\frac{K_{i1}}{K_{i2}})^{\beta - 1} \mid x_{i1}, x_{i2}, K_{i1} < K_{i2}) \\
&= Pr((\frac{\beta}{\beta - 1})(\frac{D_2(x_{i2})}{D_1(x_{i1})})^{\beta}(\frac{K_{i1}}{K_{i2}})^{\beta - 1} - (\frac{D_2(x_{i2})K_{i1}}{D_1(x_{i1})K_{i2}})^{\beta} < (\frac{\beta}{\beta - 1})(1 \\
&+ (\frac{D_2(x_{i2}) - D_1(x_{i2})}{D_2(x_{i1})})(\frac{D_2(x_{i1})}{D_1(x_{i1})})^{\beta}) - 1 \mid x_{i1}, x_{i2}, K_{i1} < K_{i2}) \\
&= Pr((\frac{\beta}{\beta - 1})(\frac{K_{i1}}{K_{i2}})^{\beta - 1} - (\frac{K_{i1}}{K_{i2}})^{\beta} < (\frac{D_2(x_{i2})}{D_1(x_{i1})})^{-\beta}(\frac{\beta}{\beta - 1})(1 \\
&+ (\frac{D_2(x_{i2}) - D_1(x_{i2})}{D_2(x_{i1})})(\frac{D_2(x_{i1})}{D_1(x_{i1})})^{\beta}) - 1) \mid x_{i1}, x_{i2}, K_{i1} < K_{i2}).
\end{aligned}
\tag{2.10}
$$

Since the marginal distribution of K_{i1} and K_{i2}, $i \in \{1, ..., N\}$, are identified in **Theorem** 7, and under the independence assumption between K_{i1} and K_{i2}, distribution of K_{i1}/K_{i2} is identified, $i \in \{1, ..., N\}$. Therefore, the expression of the conditional probability function (2.10) can be used to identify function $D_1(\cdot)$. ■

2.3 Conclusion

This chapter studies a game theoretic MHT model and contributes to the literature of continuous time duration analysis. We use the identification strategies for MHT (Abbring (2012)) for the first stage identification analysis of the model primitives and we achieve point identification for the rest of the model primitives by separating the two different equilibria according to the efficiency order of the two firms. The three-player or N-player real options game with complete information is more difficult to analyze the point identification of all the model primitives because the multiple equilibria of the game cannot be separated in general. See the discussion for details in **Appendix** B.4. Therefore a further investigation about the multiple equilibria selection would be interesting.

Chapter 3

Estimation of Varying Coefficient Models with Matching Data

The purpose of this chapter is to investigate the estimation of the varying coefficient model that involves matching estimators. Suppose that we are interested in estimating the varying coefficient model:

$$Y = X \cdot \beta(Z) + u. \tag{3.1}$$

Unlike a linear parametric regression model with constant coefficients, we study a case of nonlinear parametric regression models with coefficients varying with the models' specification. For example, the marginal propensity to consume would be different between generations and the rate of return to schooling would be different for individuals with different working experience. And when we investigate issues between generations, we sometimes face the problem that (Y, X, Z) can not be obtained from a single data set and we have to combine information from two or more samples drawn from the same population. For example, there is a large literature that studies the intergenerational income mobility. Let Y be son's income, and X_1 be control variables of son's characteristics, such as son's education and/or years of working experience. Let X_2 be father's income or family income at the time of son's childhood and Z be father's education. It is quite likely that X_2 and Y can not be observed in the same sample set. Usually we can only observe (Y, X_1, Z) in one sample and (X_2, Z) in another sample, where Z are some common variables (not necessarily common observations). Combining different data sets is quite common when a complete data is not available. Arellano and Meghir (1992) estimate the female labor supply equation using two survey data sets, the UK Labour Force Survey (LFS) and the Family Expenditure Survey (FES). The labour survey contains the information about the labour supply, which is the dependent variable, and the information on job-search activity, which is one of the explanatory variables, while the budget survey contains information on the wage rate, other income and consumption. Common variables are education, age of husband and regional labor market conditions. In Arellano and Meghir (1992), common variables are excluded from the supply equation and only used for the imputation of wage and other income in the combined data set.

There is a large literature studying how to identify and estimate the joint density of (Y, X, Z) based on the data combination, see the literature review by Ridder and Moffitt (2007). However, as noted in Ridder and Moffitt (2007), what can be recovered from the combined data is largely dependent on the nature of the available samples and the additional assumptions we want to make. For example, when the population moment conditions are additively separable into two samples and the available samples are rich enough such that we can construct required moment conditions from the two samples respectively, then a two-sample generalized method of moments (GMM) estimation is enough and no additional assumptions are needed, see Angrist and Krueger (1995) and Liu, Murtazashvili and Prokhorov (2015). However, when the two-sample GMM is not feasible due to the nature of the available samples, we need more assumptions. Suppose that Y and X are only available in two different data sets. Usually, either the assumption of conditional independence between Y and X given common variables Z, or the exclusion conditions are added for the full inference. Obviously, the conditional independence assumption is not very attractive when we are interested in the estimation of $\mathbb{E}(Y \mid X, Z)$. On the other hand, exclusion conditions are similar to the instrumental variables (IV) approach. We need to find variables that are excluded from the regression we are interested in, but also highly correlated to the missing data we want to impute into the combined data. This approach is also known as two-sample IV. Instead of making more assumptions or requiring rich samples, we can also simply combine the missing data samples by applying some matching method and investigate the properties of the estimators involving matching, which have been rigorously studied in the average treatment effects models, see Abadie and Imbens (2006) and Abadie and Imbens (2011).

We consider a general case where (Y, X_1, Z) and (X_2, Z) are collected from two samples with one sample size potentially greater than the other but of the same order. And we will apply a nearest matching method to match these two samples based on the covariates Z. The first intuition, as in the literature of evaluation research on the average treatment effect, is to approximate the missing X_2 in (Y, X_1, Z) by reasonable X_2 in the larger sample determined through nearest matching over the corresponding Z. Our investigation shows that the simple local linear estimator based on matching is inconsistent, which is due to the "matching discrepancy" termed in Abadie and Imbens (2006). Moreover, it is shown that the rate of convergence of the simple local linear estimator is dominated by the rate of the error term involving matching discrepancy, which in turn depends on the number of matching variables. In particular, the simple local linear estimator reaches the parametric convergence rate only if matching is conducted over one variable, instead of high dimensional Z. In addition to the above results, we discuss the possible bias-corrected estimators.

The rest of this chapter is organized as follows. Section 3.1 shows the inconsistency of the simple local linear estimation of the regression model (1) using matched samples. Section 3.2 proposes the methods for bias-corrected estimators and examines their convergence properties. Section 3.3 conducts Monte Carlo simulations and examines the performance of the bias correction in finite samples. Section 3.4 concludes. All proofs are given in the Appendix. MATLAB codes implementing the estimators are available upon request.

3.1 Two-Sample Matching Estimator

3.1.1 Setting and Notation

Consider the following varying coefficient model

$$\begin{aligned} Y_i &= X_i^T\beta(Z_i) + u_i \\ &= X_{1i}^T\beta_1(Z_i) + X_{2i}^T\beta_2(Z_i) + u_i, \\ &\mathbb{E}(u_i \mid X_i, Z_i) = 0 \\ &\mathbb{E}({u_i}^2 \mid X_i, Z_i) = \sigma^2 \quad a.s., i \in \{1, ...n\} \end{aligned} \tag{3.2}$$

where the dependent variable Y_i is a scalar random variable, and $\beta(\cdot) = (\beta_1(\cdot)^T, \beta_2(\cdot)^T)^T$ is a $d_1 + d_2$ dimensional vector of unknown functions. Let $X_i = (X_{1i}^T, X_{2i}^T)^T \in \mathcal{R}^{d_1+d_2}$, where X_{1i} and X_{2i} denote d_1 and d_2 dimensional vectors of exogenous regressors respectively. $Z_i \in \mathcal{R}^1$ are continuous covariates with compact support.

Suppose that we observe two independent random samples from the same population, namely, $\{Y_i, X_{1i}, Z_{1i}\}_{i=1}^n$ and $\{X_{2j}, Z_{2j}\}_{j=1}^m$. And we construct a matching data set of n observations $\{(Y_i, X_{1i}, X_{2j(i)}, Z_{1i}, Z_{2j(i)})\}_{i=1}^n$, where $Z_{2j(i)}$ is denoted as the nearest match to Z_{1i} and $X_{2j(i)}$ is the observation paired with $Z_{2j(i)}$ in the sample $\{X_{2j}, Z_{2j}\}_{j=1}^{m}$. Define

$$j(i) := \text{argmin}_{j\in\{1,\ldots,m\}}|Z_{2j} - Z_{1i}|. \tag{3.3}$$

In other words, we will match the missing values X_{2i} in $\{X_{1i}, X_{2i}\}_{i=1}^{n}$, with $X_{2j(i)}$ in $\{X_{2j}, Z_{2j}\}_{j=1}^{m}$, such that $j(i)$ is the index of the unit that is the nearest match for unit i in terms of the matching variables Z[7].

Define $C(j)$, the number of times that unit j in the sample $\{X_{2j}, Z_{2j}\}_{j=1}^{m}$ is used as a match to unit i in the sample $\{Y_i, X_{1i}, Z_{1i}\}_{i=1}^{n}$,

$$C(j) = \sum_{i=1}^{n} \mathbb{1}(j = j(i)), \quad j \in \{1, ..., m\}$$

where $\mathbb{1}(\cdot)$ is the indicator function, equal to one if $j = j(i)$ is true and zero otherwise.

We also define $\mathbb{A}(j)$ as the subset of the indices i, $i \in \{1, ..., n\}$, such that j is used as a match to each observation indexed from $\mathbb{A}(j)$; for instance, if $i \in \mathbb{A}(j)$, then $j = j(i)$. Clearly, the number of the elements in the set $\mathbb{A}(j)$ is $C(j)$.

3.1.2 Identification of the Two-sample Estimator

If we can observe a complete sample, the moment condition is $\mathbb{E}(X_i u_i \mid Z_i) = 0_{d_1+d_2}$, or $\begin{pmatrix}\mathbb{E}(X_{1i}u_i \mid Z_i)\\ \mathbb{E}(X_{2i}u_i \mid Z_i)\end{pmatrix} = 0_{d_1+d_2}$. If we now have two samples of missing data $\{Y_i, X_{1i}, Z_{1i}\}_{i=1}^{n}$ and $\{X_{2j}, Z_{2j}\}_{j=1}^{m}$ as well as the constructed sample $\{(Y_i, X_{1i}, X_{2j(i)}, Z_{1i}, Z_{2j(i)})\}_{i=1}^{n}$, the previous moment condition can not be used directly to identify parameters. Instead, the applicable moment condition in this case would be $\begin{pmatrix}\mathbb{E}(X_{1i}u_i \mid Z_i)\\ \mathbb{E}(X_{2j(i)}u_i \mid Z^{1,2})\end{pmatrix} = 0_{d_1+d_2}$. Therefore with the above moment condition, a straight forward calculation gives

$$\begin{aligned}&\begin{pmatrix}\mathbb{E}(X_{1i}y_i \mid Z_i)\\ \mathbb{E}(X_{2j(i)}y_i \mid Z^{1,2})\end{pmatrix}\\ &= \begin{pmatrix}\mathbb{E}(X_{1i}X_{1i}^T \mid Z_i), & \mathbb{E}(X_{1i}X_{2i}^T \mid Z_i)\\ \mathbb{E}(X_{2i} \mid Z_i)\mathbb{E}(X_{1i}^T \mid Z_i), & \mathbb{E}(X_{2i} \mid Z_i)\mathbb{E}(X_{2i}^T \mid Z_i)\end{pmatrix}\begin{pmatrix}\beta_1(Z_i)\\ \beta_2(Z_i)\end{pmatrix}.\end{aligned} \tag{3.4}$$

[7]As discussed in Abadie and Imbens (2006), we can also apply the nearest kth match method such that we firstly find k nearest matches to each i and then average the k matches to impute into the combined sample set. This will improve the performance of our corrected estimator using a single match and further proof will be explored in the future work

To further reduce equation 3.4, let us define some notations. Suppose $\Omega(z) = \mathbb{E}(XX' \mid Z = z)$ is positive definite for each z and uniformly continuous in z. And let $\Omega(z)_{(ij)}$ be the (i, j)th block element of matrix $\Omega(z)$. Further denote the conditional expectations of X_1 and X_2, given Z, as

$$g_1(Z) = \mathbb{E}(X_1 \mid Z), \quad g_2(Z) = \mathbb{E}(X_2 \mid Z).$$

And the conditional expectation errors are

$$v_1 = X_1 - g_1(Z), \quad v_2 = X_2 - g_2(Z).$$

Let $g(Z) = (g_1(Z)^T, g_2(Z)^T)^T$ and $v = (v_1^T, v_2^T)^T$. Denote the conditional variance of v as $\mathbb{E}(vv^T|Z) = \Sigma$, the conditional variance of v_1 as $\mathbb{E}(v_1 v_1^T|Z) = \Sigma_{11}$, the conditional variance of v_2 as $\mathbb{E}(v_2 v_2^T \mid Z) = \Sigma_{22}$, and the conditional cross-covariance as $\mathbb{E}(v_1 v_2^T|Z) = \Sigma_{12}$ and $\mathbb{E}(v_2 v_1^T|Z) = \Sigma_{21}$.

Recall that the model is identifiable if the matrix

$$\begin{pmatrix} \mathbb{E}(X_{1i}X_{1i}^T \mid Z_i) & \mathbb{E}(X_{1i}X_{2i}^T \mid Z_i) \\ \mathbb{E}(X_{2i} \mid Z_i)\mathbb{E}(X_{1i}^T \mid Z_i) & \mathbb{E}(X_{2i} \mid Z_i)\mathbb{E}(X_{2i}^T \mid Z_i) \end{pmatrix} \tag{3.5}$$

is invertible from equation 3.4. Also matrix 3.5 is equivalent to

$$\begin{pmatrix} \Omega(z)_{(11)} & \Omega(z)_{(12)} \\ \Omega(z)_{(21)} & \Omega(z)_{(22)} - \Sigma_{22} \end{pmatrix}$$

is invertible when $\Sigma_{12} = 0$, which means that X_1 and X_2 are uncorrelated conditional on Z.

Firstly, we note matrix 3.5 is symmetric positive definite when $\Sigma_{12} = 0$, and matrix 3.5 is invertible only if diagonal elements are invertible. Therefore we assume both $\mathbb{E}(X_{1i}X_{1i}^T \mid Z_i)$ and $\mathbb{E}(X_{2i} \mid Z_i)\mathbb{E}(X_{2i}^T \mid Z_i)$ are invertible, which means X_{2i}'s dimension is $d_2 = 1$.

Secondly, it is interesting to notice that the conditional distribution of $X \mid Z$ is not needed to obtain the GMM estimator in this two-sample VCM model. Similarly to the discussions in Ridder and Moffitt (2007) about statistical matching methods in the case of a simple OLS regression, this is due to the moment conditions are quadratic in $X = (X_1, X_2)^T$ and only $\mathbb{E}(X \mid Z)$ and $\mathbb{E}(XX^T \mid Z)$ are needed. In the special case parallel to the simple OLS in Ridder and Moffitt (2007)(Page 5499, equation 63), when we have two independent samples (Y, Z) and (X, Z), now the matrix 3.5 is reduced to $\mathbb{E}(X \mid Z)\mathbb{E}(X^T \mid Z)$, which is only invertible when X's dimension is $d = 1$.

3.1.3 Two-Sample Naive Local Linear Estimator

Suppose that $\{Y_i, X_{1i}, X_{2j(i)}, Z_{1i}, Z_{2j(i)}\}_{i=1}^n$ is a matched sample from two samples $\{Y_i, X_{1i}, Z_{1i}\}_{i=1}^n$ and $\{X_{2j}, Z_{2j}\}_{j=1}^m$. Denote $X_{j(i)} = (X_{1i}^T, X_{2j(i)})^T$ as the matching pair for $X_i = (X_{1i}^T, X_{2i})^T$, $i \in \{1, ..., n\}$. We can approximate $\beta(Z_{1i})$ in a small neighbourhood of z by a linear function

$$\beta(Z_{1i}) \approx \theta_0 + \frac{1}{h}(Z_{1i} - z)\theta_1,$$

where $\theta_0 = \beta(z)$, and $\theta_1 = h\beta'(z)$. Then we consider a naive local linear estimator as minimizers with respect to (θ_0, θ_1) of the following weighted local least-squares problem:

$$\sum_{i=1}^{n} \left(Y_i - X_{j(i)}^T \theta_0 - \frac{Z_{2j(i)} - z}{h} X_{j(i)}^T \theta_1 \right)^2 K_h \left(Z_{2j(i)} - z \right), \tag{3.6}$$

where $K(\cdot)$ is a kernel function, h is a bandwidth and $K_h(\cdot) = \frac{K(\cdot/h)}{h}$.

The local linear estimator $\hat{\beta}(z)$ is given by the solution for θ_0 to the problem of minimizing (3.6). And the solution admits the following expression:

$$\left(\hat{\beta}(z)^T, h\hat{\beta}'(z)^T \right)^T = \left(\left(D^{X_m} \right)^T W D^{X_m} \right)^{-1} \left(D^{X_m} \right)^T WY, \tag{3.7}$$

where

$$D^{X_m} = \begin{pmatrix} X_{j(1)}^T & X_{j(1)}^T \frac{Z_{2j(1)} - z}{h} \\ ... & ... \\ X_{j(n)}^T & X_{j(n)}^T \frac{Z_{2j(n)} - z}{h} \end{pmatrix},$$

$$W = diag \left(K_h \left(Z_{2j(1)} - z \right), ..., K_h \left(Z_{2j(n)} - z \right) \right),$$

and

$$Y = (Y_1, ..., Y_n)^T.$$

Let $\Theta(z) = \left(\beta(z)^T, \beta'(z)^T \right)^T$, and define its estimator to be

$$\hat{\Theta}(z) = H^{-1} \left((D^{X_m})^T W D^{X_m} \right)^{-1} \left(D^{X_m} \right)^T WY,$$

where $H = diag(1, ..., 1, h, ..., h)$ is a $2(d_1 + 1) \times 2(d_1 + 1)$ matrix with the first $d_1 + 1$ diagonal elements being 1 and the remaining diagonal elements

h. We can write the estimator of $\beta(z)$ as

$$\hat{\beta}(z)^T = eH^{-1}((D^{X_m})^T W D^{X_m})^{-1}(D^{X_m})^T WY,$$

where e is a $1 \times 2(d_1+1)$ matrix with the first d_1+1 elements being 1 and the rest diagonal elements 0. We call $\hat{\beta}(z)$ the two-sample naive local linear estimator for varying coefficient models. For the following analysis, we modify the expression for the naive local linear estimator in a concise way,

$$H\hat{\Theta}(z) = D_n(z)^{-1}N_n(z),$$

where $D_n(z) = \begin{pmatrix} D_{n,0} & D_{n,1} \\ D_{n,1} & D_{n,2} \end{pmatrix}$ and $N_n(z) = \begin{pmatrix} N_{n,0} \\ N_{n,1} \end{pmatrix}$,

$$\begin{aligned}
D_{n,0}(z) &= \frac{1}{n}\sum_{i=1}^{n} K_h(Z_{2j(i)} - z)X_{j(i)}X_{j(i)}^T, \\
D_{n,1}(z) &= \frac{1}{nh}\sum_{i=1}^{n} K_h(Z_{2j(i)} - z)(Z_{2j(i)} - z)X_{j(i)}X_{j(i)}^T, \\
D_{n,2}(z) &= \frac{1}{nh^2}\sum_{i=1}^{n} K_h(Z_{2j(i)} - z)(Z_{2j(i)} - z)^2 X_{j(i)}X_{j(i)}^T, \\
N_{n,0}(z) &= \frac{1}{n}\sum_{i=1}^{n} K_h(Z_{2j(i)} - z)X_{j(i)}Y_i, \\
N_{n,1}(z) &= \frac{1}{nh}\sum_{i=1}^{n} K_h(Z_{2j(i)} - z)(Z_{2j(i)} - z)X_{j(i)}Y_i.
\end{aligned}$$

3.1.4 Large Sample Properties of the Two-Sample Naive Local Linear Estimator

Assumption 13. *$\{Y_i, X_{1i}, Z_{1i}\}_{i=1}^n$ and $\{X_{2j}, Z_{2j}\}_{j=1}^m$ are two independent samples from the same population $\{Y, X_1, X_2, Z\}$ with missing data.*

Assumption 14. *1. The density function $f(\cdot)$ of Z is bounded, and have continuous second derivatives on a compact set.*

2. The matrix $f(z)\Omega(z)$ is invertible, and so is the matrix

$$f(z)\begin{pmatrix} \Omega(z)_{(11)}, & \Omega(z)_{(12)} \\ \Omega(z)_{(21)}, & \Omega(z)_{(22)} - \Sigma_{22} \end{pmatrix}$$

over the domain of z.

3. $\beta_j(\cdot)$, *with* $j \in \{1, ..., d_1 + d_2\}$ *have continuous second derivatives at each point* z *in the support of* Z

Assumption 15. *The kernel function* $K(\cdot)$ *is symmetric and a bounded second order kernel function with compact support.* $K(\cdot)$ *is Lipschitz continuous. The bandwidth* h *satisfies* $nh \to \infty$ *and* $h \to 0$ *as* $n \to \infty$, $nh^8 \to 0$ *and* $nh^2/(\log n)^2 \to \infty$ *as* $n \to \infty$.

Assumption 16. 1. *Functions* $g_1(\cdot)$ *and* $g_1(\cdot)$ *have continuous second derivatives at each point* z *on the support of* Z,

2. *the fourth moment of the conditional distribution of* Y *given* $Z = z$ *exists and is bounded uniformly in* z,

3. σ^2 *is bounded away from zero,*

4. $\frac{m}{n} \to \kappa \in (0, \infty)$ *as* $n,\ m \to \infty$ *jointly.*

Assumption 13 specifies our two-sample setup. As both sample sizes $n \to \infty$ and $m \to \infty$, we cannot apply the law of large numbers or the central limit theorem to the combined sample constructed by nearest matching, because the i.i.d. property of the combined sample is destroyed by replacing the fixed index i with a random index $j(i)$. Assumptions 14−15 are standard assumptions for the consistency and asymptotic normality of the local linear estimators of the varying coefficient models. Assumption 16 adds additional assumptions needed to reestablish the consistency and asymptotic normality results for our two-sample nearest matching estimators.

We first show that the denominator in our two-sample matching estimator is consistent for its expectation. And then we show that the numerator is also consistent but the resulting two-sample matching estimator is biased. Then we will prove that without the conditional bias term, the matching estimator is $N^{1/2}$ consistent and asymptotically normal. In the following analysis, we denote

$$\mu_j = \int u^j K(u) du,$$

and

$$\nu_j = \int u^j K^2(u) du, \quad j = 1, 2, 3.$$

And we use $\bigotimes$ to denote kronecker product.

Lemma 9. ***Convergence of Denominator***. *Under **Assumptions** 13–16,*

$$D_n(z) \xrightarrow{p} f(z)\Omega(z) \bigotimes \begin{pmatrix} 1 & 0 \\ 0 & \mu_2 \end{pmatrix},$$

Proof. See **Appendix** C.1. ■

Lemma 10. ***Convergence of Numerator***. *Under **Assumptions** 13–16,*

$$N_n(z) - \frac{1}{n}(D^{X_m})^T W D^{X_m} \begin{pmatrix} \beta(z) \\ h\beta'(z) \end{pmatrix} \xrightarrow{p}$$
$$f(z) \begin{pmatrix} 0 & \Sigma_{12} \\ 0 & -\Sigma_{22} \end{pmatrix} \beta(z) \bigotimes \begin{pmatrix} 1 \\ 0 \end{pmatrix} + f(z) \begin{pmatrix} 0 & \Sigma_{12} \\ 0 & -\Sigma_{22} \end{pmatrix} h\beta'(z) \bigotimes \begin{pmatrix} 0 \\ \mu_2 \end{pmatrix}.$$

Proof. See **Appendix** C.2. ■

Theorem 11. ***Inconsistency of the Naive Estimator***. *Suppose that **Assumptions** 13–16 hold. Then*

$$\hat{\beta}(z) - \beta(z) = bias(z)\beta(z) + O_p\left(h^2 + \frac{1}{\sqrt{nh}}\right), \tag{3.8}$$

where

$$bias(z) = \Omega(z)^{-1} \begin{pmatrix} 0 & \Sigma_{12} \\ 0 & -\Sigma_{22} \end{pmatrix}.$$

Proof. **Theorem** 11 holds by combining the results from **Lemma** 9 and 10. ■

Theorem 12. ***Asymptotic Normality for the Naive Estimator***. *Suppose that **Assumptions** 13–16 hold. Then*

$$\sqrt{nh}V(z)^{-1/2}D(z)\left(H(\hat{\Theta}(z) - \Theta(z)) - bias(z) \bigotimes \begin{pmatrix} \beta(z) \\ h\beta'(z) \end{pmatrix}\right) \xrightarrow{d} \mathcal{N}(0, I),$$

where $D(z) = f(z)\Omega(z) \bigotimes \begin{pmatrix} 1 & 0 \\ 0 & \mu_2 \end{pmatrix}$ *and*

$$V(z) = \sum_{j=1}^{m} \sigma^2 (C(j)K_h(Z_{2j} - z))^2 \Omega(Z_{2j}) \bigotimes \begin{pmatrix} 1 & (Z_{2j} - z)/h \\ (Z_{2j} - z)/h & ((Z_{2j} - z)/h)^2 \end{pmatrix}.$$

Proof. From the argument in **Lemma** 10,

$$\sqrt{nh}\left(H(\hat{\Theta}(z)-\Theta(z)) - bias(z)\bigotimes\begin{pmatrix}\beta(z)\\ h\beta'(z)\end{pmatrix}\right) = \sqrt{nh}D_n^{-1}(z)I_4,$$

where $I_4 = \begin{pmatrix}\frac{1}{n}\sum_{i=1}^n X_{j(i)}K_h(Z_{2j(i)}-z)u_i\\ \frac{1}{n}\sum_{i=1}^n X_{j(i)}K_h(Z_{2j(i)}-z)u_i(Z_{2j(i)}-z)/h\end{pmatrix}$. By **Lemma** 9, $D_n(z) \xrightarrow{p} f(z)\Omega(z)\bigotimes\begin{pmatrix}1 & 0\\ 0 & \mu_2\end{pmatrix}$.

Firstly, we will show I_4 is a random vector made up with sums of conditionally independent random variables, given all the Z from the two samples. Denote $I_4 = (I_{4,0}^T, I_{4,1}^T)^T$, where

$$I_{4,0} = \frac{1}{n}\sum_{i=1}^n X_{j(i)}K_h(Z_{2j(i)}-z)u_i,$$

and

$$I_{4,1} = \frac{1}{n}\sum_{i=1}^n X_{j(i)}K_h(Z_{2j(i)}-z)u_i(Z_{2j(i)}-z)/h.$$

Since $\mathbb{A}(j)$ indicates the subset of the index i, $i\in\{1,...,n\}$, such that j is used as a match to each observation indexed by i, $i\in\{1,...,n\}$, we can rewrite $I_{4,0}$ and $I_{4,1}$ as

$$\begin{aligned}I_{4,0} &= \frac{1}{n}\sum_{j=1}^m I_{4,0}^j\\ &= \frac{1}{n}\sum_{j=1}^m\sum_{i\in\mathbb{A}(j)} X_{(i,j)}K_h(Z_{2j}-z)u_i,\end{aligned}$$

and

$$\begin{aligned}I_{4,1} &= \frac{1}{n}\sum_{j=1}^m I_{4,1}^j\\ &= \frac{1}{n}\sum_{j=1}^m\sum_{i\in\mathbb{A}(j)} X_{(i,j)}K_h(Z_{2j}-z)u_i(Z_{2j}-z)/h,\end{aligned}$$

where $X_{(i,j)} = (X_{1i}^T, X_{2j}^T)^T$.

Recall that $Z^{1,2}$ represents all the Z from the two samples. Then conditional on $Z^{1,2}$, the unit-level terms $I_{4,0}^j = \sum_{i\in\mathbb{A}(j)} X_{(i,j)} K_h(Z_{2j}-z)u_i$ are independent with zero means and nonidentical distributions. To see this, first, note the number of elements in the index set $\mathbb{A}(j)$ is $C(j)$, which is the number of times Z_{2j} is used as a match, $j \in \{1,...,m\}$. And conditional on $Z^{1,2}$, $C(j)$ is nonstochastic. Therefore conditional on $Z^{1,2}$, the sum in $I_{4,0}$ are made up with all *i.i.d.* terms for all j, $j \in \{1,...,m\}$. As a result, $I_{4,0}^j$ are independent for all j, $j \in \{1,...,m\}$. The conditional variance of $I_{4,0}^j$ is $(C(j)K_h(Z_{2j}-z))^2\sigma^2\Omega(Z_{2j}) + o(1)$.

Likewise, conditional on all the $Z^{1,2}$, the unit-level terms $I_{4,1}^j$ are also independent with zero means and nonidentical distributions. The conditional variance of $I_{4,1}^j$ is $(C(j)K_h(Z_{2j}-z)(Z_{2j}-z)/h)^2\sigma^2\Omega(Z_{2j}) + o(1)$.

Next, we will use the Cramèr-Wold device and the Lindeberg-Feller central limit theorem to derive the asymptotic distribution of $\sqrt{n}I_4$. Denote the dimension of $X = (X_1, X_2)$ as $p = d_1 + d_2$. For any $2p \times 1$ nonzero vector $\tau = (\tau_1, ..., \tau_{2p})^T$, we have

$$\sqrt{nh}\tau^T I_4 = \sqrt{\frac{m}{n}} \frac{\sqrt{h}}{\sqrt{m}} \sum_{j=1}^{m} \left\{ \sum_{k=1}^{p} \sum_{i\in\mathbb{A}(j)} \tau_k X_{(i,j)k} K_h(Z_{2j}-z)u_i \right.$$
$$\left. + \sum_{k=1}^{p} \sum_{i\in\mathbb{A}(j)} \tau_{k+p} X_{(i,j)k} K_h(Z_{2j}-z)\left((Z_{2j}-z)/h\right) u_i \right\}$$

Similarly to the argument for the conditional independence of $I_{4,0}^j$ and $I_{4,1}^j$ given $Z^{1,2}$,

$$\sqrt{h} I_4^j = \sqrt{h} \left\{ \sum_{k=1}^{p} \sum_{i\in\mathbb{A}(j)} \tau_k X_{(i,j)k} K_h(Z_{2j}-z)u_i \right.$$
$$\left. + \sum_{k=1}^{p} \sum_{i\in\mathbb{A}(j)} \tau_{k+p} X_{(i,j)k} K_h(Z_{2j}-z)\left((Z_{2j}-z)/h\right) u_i \right\}$$

are independent and we can apply the Lindeberg-Feller central limit theorem if the Lindeberg-Feller condition are satisfied. Denote the variance of $\sum_{j=1}^{m} I_4^j$ is V_τ.

For given $Z^{1,2}$, the Lindeberg-Feller condition requires that

$$\frac{1}{mV_\tau}\sum_{j=1}^{m}\mathbb{E}\left[(I_4^j)^2\mathbb{1}\{|I_4^j|\geq\eta\sqrt{mV_\tau}\}\mid Z^{1,2}\right]\to 0$$

for all $\eta > 0$. By applying Hölder's and Markov's inequalities we have

$$\begin{aligned}
&\frac{1}{mV_\tau}\sum_{j=1}^{m}E\left[(I_4^j)^2\mathbb{1}\{|I_4^j|\geq\eta\sqrt{mV_\tau}\}\mid Z^{1,2}\right]\\
&\leq\frac{1}{mV_\tau}\sum_{j=1}^{m}\left(\mathbb{E}\left[(I_4^j)^4\mid Z^{1,2}\right]\right)^{\frac{1}{2}}\frac{\mathbb{E}\left[(I_4^j)^2\mid Z^{1,2}\right]}{\eta^2 mV_\tau}\\
&\leq\frac{1}{mV_\tau}\sum_{j=1}^{m}\left(C(j)^4K_h(Z_{2j}-z)^4\mathbb{E}[u_j^4\mid Z^{1,2}]\Psi(\tau,Z_{2j})\right)^{\frac{1}{2}}\\
&\frac{C(j)^2K_h(Z_{2j}-z)^2\sigma^2\Phi(\tau,Z_{2j})}{\eta^2 mV_\tau}\\
&\leq\frac{\bar{c}^{\frac{1}{2}}}{\eta^2\sigma^2}\frac{1}{m}\left(\frac{1}{m}\sum_{j=1}^{m}C(j)^4\right),
\end{aligned}$$

where $\bar{c} = sup_z\mathbb{E}[u_j^4 \mid Z = z] < \infty$, both $\Psi(\tau, Z_{2j})$ and $\Phi(\tau, Z_{2j})$ are functions composed with deterministic coefficients. Because $\mathbb{E}(C(j)^4)$ is uniformly bounded by **Lemma** 3 in Abadie and Imbens (2006), by Markov's inequality, the last term is bounded in probability. Hence, the Lindeberg-Feller condition is satisfied for almost all Z. As a result,

$$\sqrt{nh}V(z)^{-1/2}I_4\xrightarrow{d}\mathcal{N}(0,\mathrm{I}),$$

where

$$V(z)=\sum_{j=1}^{m}\sigma^2(C(j)K_h(Z_{2j}-z))^2\Omega(Z_{2j})\bigotimes\begin{pmatrix}1 & (Z_{2j}-z)/h\\ (Z_{2j}-z)/h & ((Z_{2j}-z)/h)^2\end{pmatrix}+o(1).$$

The boundedness of $V(z)$ are proved in **Appendix** C.4. Then

$$\sqrt{nh}V(z)^{-1/2}D(z)\left(H(\hat{\Theta}(z)-\Theta(z))-bias(z)\bigotimes\begin{pmatrix}\beta(z)\\ h\beta'(z)\end{pmatrix}\right)\xrightarrow{d}\mathcal{N}(0,I),$$

where $D(z) = f(z)\Omega(z) \bigotimes \begin{pmatrix} 1 & 0 \\ 0 & \mu_2 \end{pmatrix}$. ■

3.2 Bias Correction and the Consistent Estimator

In this section we analyze the asymptotic properties of the bias-corrected matching estimator. From **Theorem** 11, the bias term in the two-sample matching estimator is given by

$$bias(z) \bigotimes \begin{pmatrix} \beta(z) \\ h\beta'(z) \end{pmatrix} = \Omega(z)^{-1} \begin{pmatrix} 0 & \Sigma_{12} \\ 0 & -\Sigma_{22} \end{pmatrix} \bigotimes \begin{pmatrix} \beta(z) \\ \mu_2 h\beta'(z) \end{pmatrix}.$$

This bias comes from the fact that the denominator of the two-sample matching estimator, $D_n(z)$, converges to its expectation $D(z)$, while the numerator $N_n(z)$ converges to $\left(D(z) + f(z) \begin{pmatrix} 0 & \Sigma_{12} \\ 0 & -\Sigma_{22} \end{pmatrix} \bigotimes \begin{pmatrix} 1 & 0 \\ 0 & \mu_2 \end{pmatrix} \right) \begin{pmatrix} \beta(z) \\ h\beta'(z) \end{pmatrix}$. Therefore, we can replace the denominator $D_n(z)$ with a consistent estimator of $D_n(z) + f(z) \begin{pmatrix} 0 & \Sigma_{12} \\ 0 & -\Sigma_{22} \end{pmatrix} \bigotimes \begin{pmatrix} 1 & 0 \\ 0 & \mu_2 \end{pmatrix}$, and leave the numerator $N_n(z)$ unchanged to eliminate the bias in the two-sample matching estimator.

In order to establish the asymptotic properties of the bias-corrected estimator, we need to estimate Σ_{12} and Σ_{22} consistently. For Σ_{22}, we consider the difference-based variance estimator proposed in Rice (1984). In particular,

$$\hat{\Sigma}_{22} = \frac{1}{2(m-1)} \sum_{j=2}^{m} (X_{2(j)} - X_{2(j-1)})(X_{2(j)} - X_{2(j-1)})^T,$$

where $X_{2(j)}$ and $Z_{2(j)}$ are from the ordered sample $\{(X_{2(j)}, Z_{2(j)})\}_{j=1}^{m}$ based on $Z_{2(1)} \leq \cdots \leq Z_{2(m)}$.

However, the estimator of Σ_{12} is more complicate, because Σ_{12} reflects the population correlation between X_1 and X_2, which are not available in a single sample. And there is a large literature about this kind of two-sample combination problems and the solutions to recover the population joint density of X_1 and X_2 are either assuming X_1 and X_2 are conditionally independent or adding more exclusive variables, which are excluded from the regression of Y on X_1 and X_2, but highly correlated to both X_1 and X_2. As a result, here we assume $\Sigma_{12} = 0$, and we can also consistently estimate the density function $f(z)$ by any nonparametric method.

Define bias-corrected two-sample matching estimator as

$$H\hat{\Theta}^{bcll}(z) = \left(D_n(z) + \hat{f}(z)\begin{pmatrix} 0 & 0 \\ 0 & -\hat{\Sigma}_{22} \end{pmatrix} \bigotimes \begin{pmatrix} 1 & 0 \\ 0 & \mu_2 \end{pmatrix}\right)^{-1} N_n(z).$$

Then we can estimate $\beta(z)$ consistently using

$$\hat{\beta}^{bcll}(z) = eH^{-1}\left(D_n(z) + \hat{f}(z)\begin{pmatrix} 0 & 0 \\ 0 & -\hat{\Sigma}_{22} \end{pmatrix} \bigotimes \begin{pmatrix} 1 & 0 \\ 0 & \mu_2 \end{pmatrix}\right)^{-1} N_n(z), \tag{3.9}$$

where e is a $1 \times 2(d_1 + d_2)$ matrix with the first $(d_1 + d_2)$ elements being 1 and the remaining diagonal elements 0.

Theorem 13. ***Asymptotic Normality for the Bias-corrected Matching Estimator***. *Suppose that* ***Assumptions*** *13–16 hold, and assume* $\Sigma_{12} = 0$, *then*

$$\sqrt{nh}V(z)^{-1/2}D(z)\left(H(\hat{\Theta}^{bcll}(z) - \Theta(z))\right) \xrightarrow{d} \mathcal{N}(0, I),$$

where $D(z) = f(z)\Omega(z) \bigotimes \begin{pmatrix} 1 & 0 \\ 0 & \mu_2 \end{pmatrix}$ *and*

$$V(z) = \sum_{j=1}^{m} \sigma^2 (C(j)K_h(Z_{2j} - z))^2 \Omega(Z_{2j}) \bigotimes \begin{pmatrix} 1 & (Z_{2j} - z)/h \\ (Z_{2j} - z)/h & ((Z_{2j} - z)/h)^2 \end{pmatrix}.$$

Proof. Firstly, we will show that

$$\sqrt{n}\left(H\hat{\Theta}^{bcll}(z) - H\hat{\Theta}(z) + bias(z) \bigotimes \begin{pmatrix} \beta(z) \\ h\beta'(z) \end{pmatrix}\right) \xrightarrow{p} 0,$$

where $bias(z) = \Omega(z)^{-1}\begin{pmatrix} 0 & 0 \\ 0 & -\Sigma_{22} \end{pmatrix}$. Note

$$\begin{aligned}
&H\hat{\Theta}^{bcll}(z) - H\hat{\Theta}(z) \\
&= \left(\left(D_n(z) + \hat{f}(z)\begin{pmatrix} 0 & 0 \\ 0 & -\hat{\Sigma}_{22} \end{pmatrix} \bigotimes \begin{pmatrix} 1 & 0 \\ 0 & \mu_2 \end{pmatrix}\right)^{-1} - (D_n(z))^{-1}\right) N_n(z) \\
&= (D_n(z))^{-1}\left(\hat{f}(z)\begin{pmatrix} 0 & 0 \\ 0 & -\hat{\Sigma}_{22} \end{pmatrix} \bigotimes \begin{pmatrix} 1 & 0 \\ 0 & \mu_2 \end{pmatrix}\right) \\
&\qquad \left(D_n(z) + \hat{f}(z)\begin{pmatrix} 0 & 0 \\ 0 & -\hat{\Sigma}_{22} \end{pmatrix} \bigotimes \begin{pmatrix} 1 & 0 \\ 0 & \mu_2 \end{pmatrix}\right)^{-1} N_n(z) \\
&= (D_n(z))^{-1}\left(\hat{f}(z)\begin{pmatrix} 0 & 0 \\ 0 & -\hat{\Sigma}_{22} \end{pmatrix} \bigotimes \begin{pmatrix} 1 & 0 \\ 0 & \mu_2 \end{pmatrix}\right) H\hat{\Theta}^{bcll}(z).
\end{aligned}$$

The consistency of $H\hat{\Theta}^{bcll}(z)$ can be established in line with the proof of **Theorem** 11. Therefore we have $H\hat{\Theta}^{bcll}(z) \xrightarrow{p} H\Theta$. As a result,

$$H\hat{\Theta}^{bcll}(z) - H\hat{\Theta}(z) \xrightarrow{p} \Omega(z)^{-1}\begin{pmatrix} 0 & 0 \\ 0 & -\Sigma_{22} \end{pmatrix} \bigotimes \begin{pmatrix} 1 & 0 \\ 0 & \mu_2 \end{pmatrix} H\Theta.$$

Then the asymptotic normality results will be applied by **Theorem** 12. ■

The result of **Theorem** 13 suggests that the bias-corrected matching estimator has the same asymptotic variance as the naive local linear matching estimator.

3.3 Monte Carlo Simulations

To evaluate the naive local linear and the bias-corrected local linear estimator, we consider the following data generating process (DGP):

$$Y = \beta_0(Z) + \beta_1(Z)X_1 + \beta_2(Z)X_2 + U, \tag{3.10}$$

where $\beta_0(Z) = (1 - e^Z + Z)$ and $\beta_1(Z) = (0.5 + 0.5Z)$. We consider two functional forms of $\beta_2(\cdot)$. We assume $\beta_2^{(1)}(z) = 0.5 + 0.5Z + 0.25Z^2$ for DGP (1) and $\beta_2^{(2)}(z) = 1 + e^z$ for DGP (2). We set $Z \sim N(0,1)$ truncated at ± 2, $X_1 = 2Z + \xi_1$, $X_2 = 3Z + \xi_2$, $(\xi_1, \xi_2)' \sim N(0, \mathbf{I}_2)$, and $U \perp (\xi_1, \xi_2)'$.

The optimal bandwidth h of all the estimators are determined by cross validation (CV) method throughout this section.

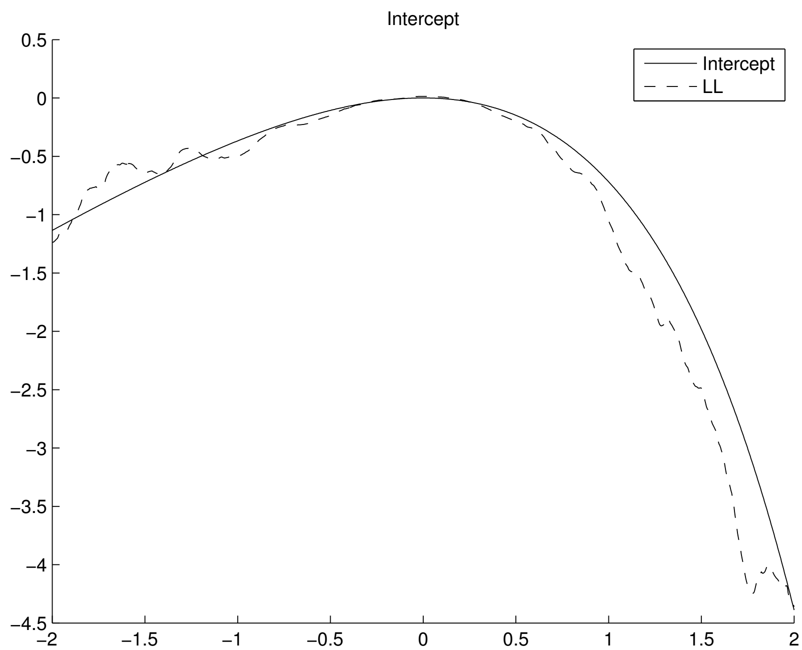

Figure 3.1: Local linear estimator in one complete sample case: intercept, $\beta_0(Z)$

3.3.1 Case with One Complete Sample

In this section we show that the naive local linear estimator performs well when we have one complete sample generated from DPG (1). The sample size $n = 3000$. Figure 3.1 - Figure 3.3 are based on 500 replications. (In fact the performance of the estimator based on one estimation, including the order of bias and variance, is similar with that based on 500 replications.)

3.3.2 Case with Two Missing-data Samples

In this section we first study the behaviour of the naive local linear estimator (LL) and the bias-corrected local linear estimator (BCLL) when we have two i.i.d samples with missing data generated from DPG (1), $\{Y_i, X_{1i}, Z_{1i}\}_{i=1}^n$, $\{X_{2j}, Z_{2j}\}_{j=1}^m$, where $n = 3000$, $m = 4000$. Figure 3.4 - Figure 3.6 are based on 500 replications. (In fact the performance of the LL and BCLL estimators in one estimation, including the order of bias and

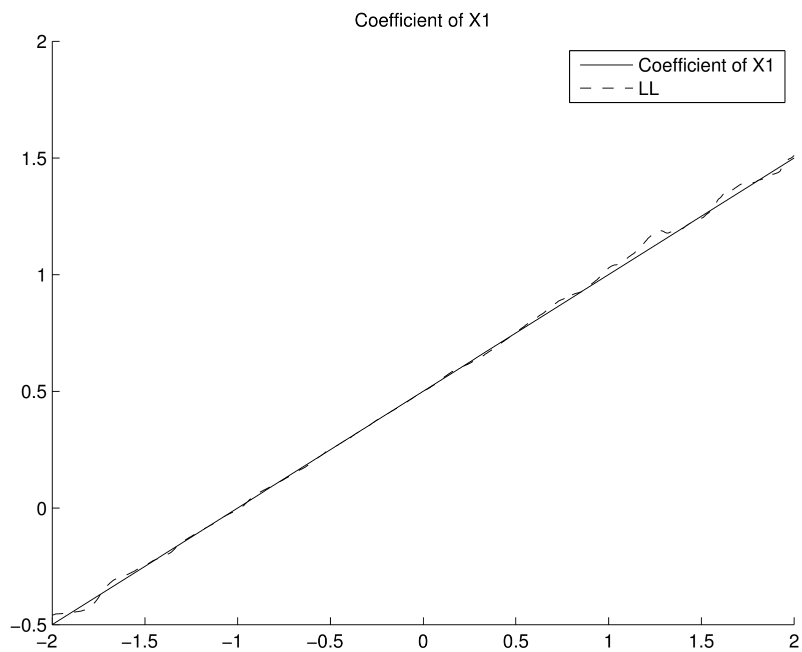

Figure 3.2: Local linear estimator in one complete sample case: coefficient of X_1, $\beta_1(Z)$

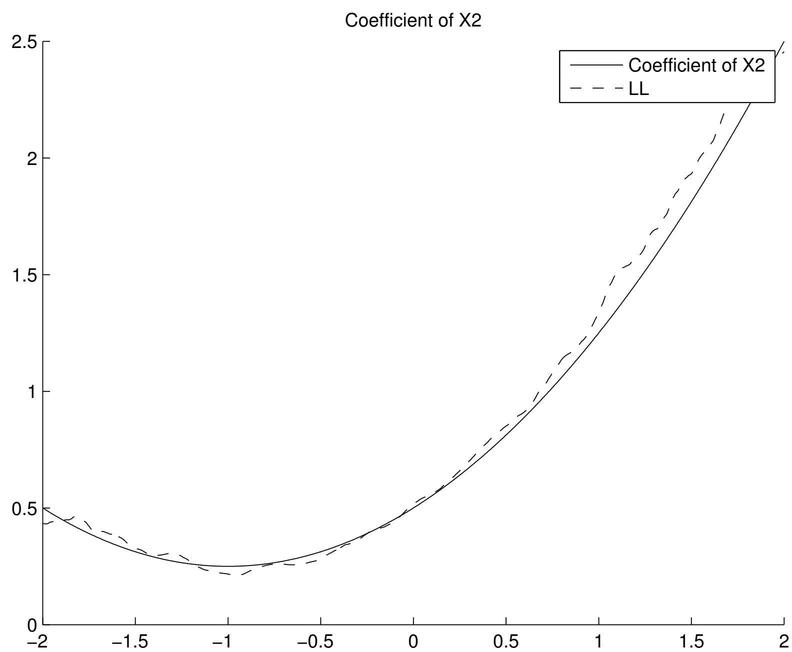

Figure 3.3: Local linear estimator in one complete sample case: coefficient of X_2, $\beta_2(Z)$

variance, are similar with that based on 500 replications, respectively.) Figure 3.4, Figure 3.5 and Figure 3.6 show the performance of the LL and BCLL estimators on the estimation of the intercept, the coefficient function for X_1 and the coefficient function for X_2 respectively. In these figures, the solid lines represent the true function parameters, while the dashed line and the dotted line represent the LL and BCLL respectively. It can be easily seen from Figure 3.4 that, the BCLL estimator has less average bias than the LL estimator in most part of the support of Z despite the boundary effect. In addition, BCLL identifies the right shape of the intercept function while LL does not. Figure 3.6 shows the similar properties of two estimators as in Figure 3.4. For Figure 3.5, the two estimators both identifies the true shape of the coefficient function while BCLL has less average bias.

To evaluate the finite sample performance of the LL and BCLL estimator of the functional coefficient, we calculate both the mean absolute deviation (MAD) and mean squared error (MSE) for each estimate evaluated at 100 evenly spaced points between the support of Z, which is $[-2, 2]$. In this case we consider both DGP (1) and DGP(2) with sample size $n = 300$ and $m = 400$.

Table 3.1 reports the results where the MSEs and MADs are averages over 500 replications for each functional coefficient. As expected, the bias-corrected local linear (BCLL) estimators perform better than the local linear (LL) estimators both in DGP(1) and DGP(2), which are specified differently only for the coefficient of X_2, $\beta_2(Z)$, in equation 3.10. However, all the estimators of the coefficients $\beta_0(Z)$, $\beta_1(Z)$ and $\beta_2(Z)$ are very sensitive to this change in the DGP. This reflects the fact that even with the additional assumption that $\Sigma_{12} = 0$, the BCLL estimator is not just corrected for the coefficient β_2. All the estimators of coefficients are affected because of the inverse of the additive bias-corrected term in the denominator of the BCLL estimator.

Table 3.1: Finite Sample Comparison of the Local Linear (LL) Estimator and the Bias-Corrected Local Linear (BCLL) Estimator

DGP	Estimators	$\beta_1(Z)$		$\beta_2(Z)$		$\beta_3(Z)$	
		MSE	MAD	MSE	MAD	MSE	MAD
$DGP(1)$	LL	1.1625	1.3886	0.7254	0.6272	0.8537	0.6072
	BCLL	0.2233	0.3040	0.2784	0.2297	0.3082	0.2237
$DGP(2)$	LL	1.3908	1.0438	1.4761	1.1572	1.3490	0.9765
	BCLL	0.8634	0.7173	0.1902	0.3199	0.3856	0.5050

Notes: samples are generated from DGP (1) and DGP (2) respectively. The pair of sample sizes are ($n = 300, m = 400$), which is the same for both DGP (1) and DPG (2). MSEs and MADs are averages over 500 replications.

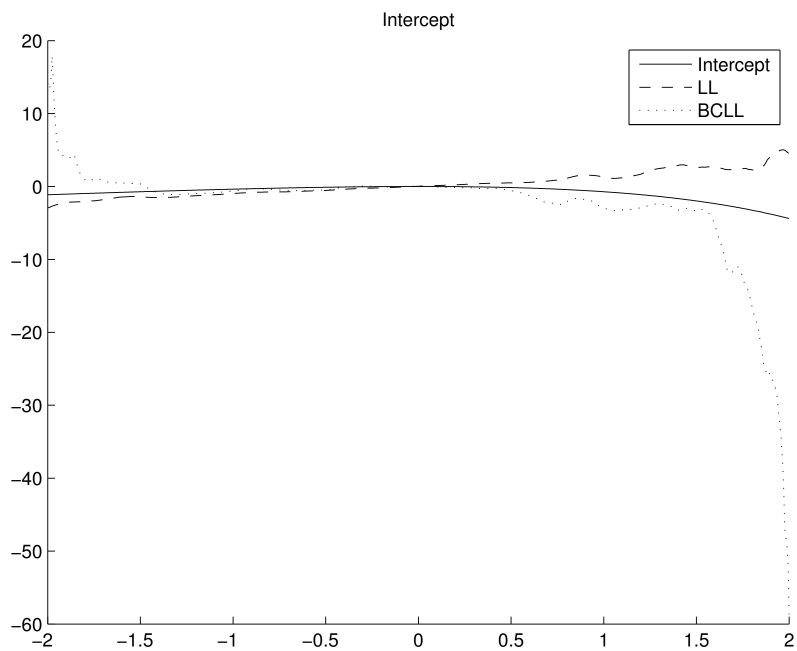

Figure 3.4: Estimators in two-sample case: intercept, $\beta_0(Z)$

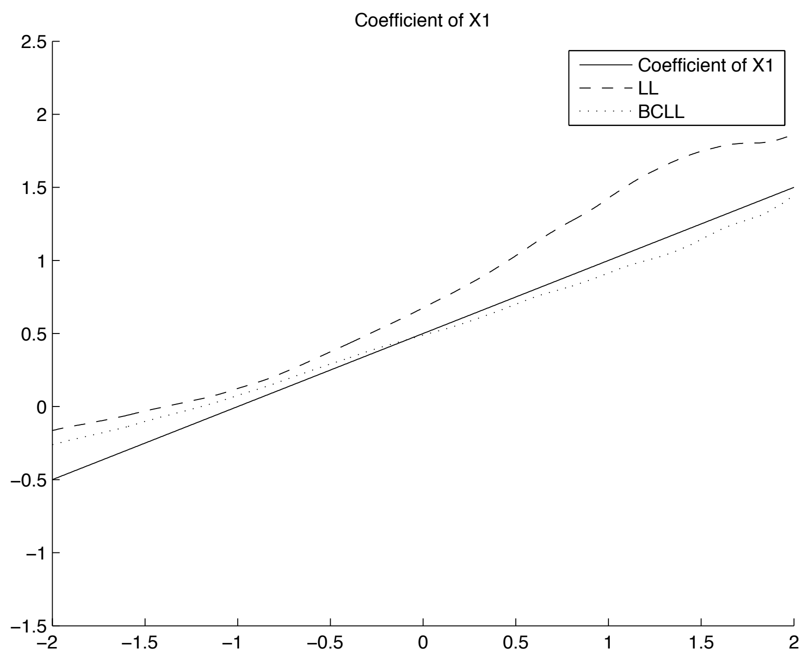

Figure 3.5: Estimators in two-sample case: coefficient of X_1, $\beta_1(Z)$

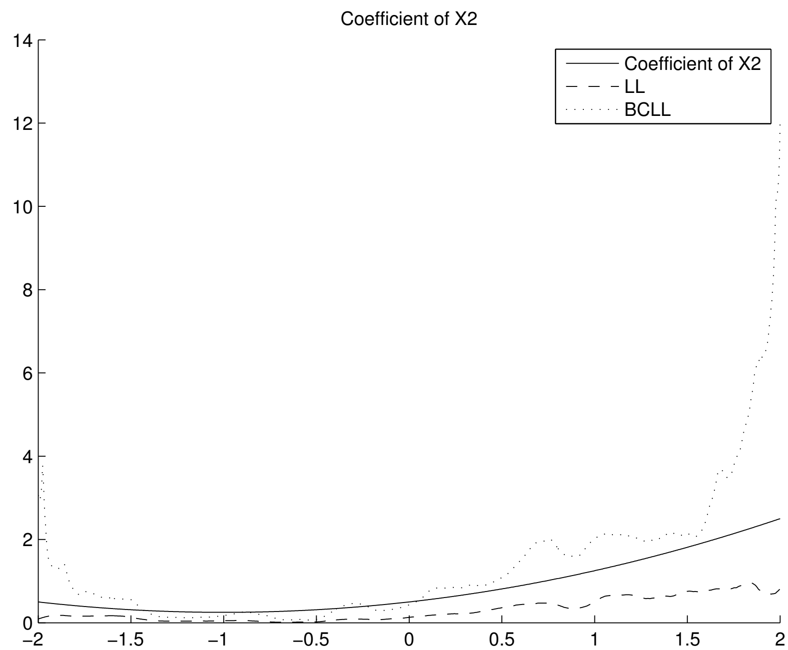

Figure 3.6: Estimators in two-sample case: coefficient of X_2, $\beta_2(Z)$

Bibliography

Abadie, Alberto, and Guido Imbens (2006) 'Large sample properties of matching estimators for average treatment effects.' *Econometrica* 74, 235–267

— (2008a) 'Estimation of the conditional variance in paired experiments.' *Annales d' Économie et de Statistique* (91/92), pp. 175–187

— (2008b) 'On the failure of the bootstrap for matching estimators.' *Econometrica.* 76, 1537–1557

— (2011) 'Bias-corrected matching estimators for average treatment effects.' *Journal of Business and Economic Statistics* 29, 1–11

Abbring, Jaap H. (2010) 'Identification of dynamic discrete choice models.' *Annual Review of Economics* 2(1), 367–394

— (2012) 'Mixed hitting-time models.' *Econometrica* 80(2), 783–819

Abbring, Jaap H., and T. Salimans (2012) 'The likelihood of mixed hitting-time models.' *working paper*

Abbring, Jaap H., and Yifan Yu (2015) 'The empirical content of synchronization games.' *working paper*

Anderson, Steven T., Daniel Friedman, and Ryan Oprea (2010) 'Preemption games: Theory and experiment.' *American Economic Review* 100(4), 1778–1803

Angrist, Joshua D., and Alan B. Krueger (1995) 'Split-sample instrumental variables estimates of the return to schooling.' *Journal of Business & Economic Statistics* 13(2), pp. 225–235

Arellano, Manuel, and Costas Meghir (1992) 'Female labour supply and on-the-job search: An empirical model estimated using complementary data sets.' *The Review of Economic Studies* 59, 537–559

Argenziano, Rossella, and Philipp Schmidt-Dengler (2012) 'Inefficient entry order in preemption games.' *Journal of Mathematical Economics* 48, 445–460

— (2014) 'Clustering in n-player preemption games.' *Journal of the European Economic Association* 12, 368–396

Azevedo, Alcino F., and Dean A. Paxson (2010) 'Real options games models: A review.' *working paper*

Berry, Steven T. (1992) 'Estimation of a model of entry in the airline industry.' *Econometrica* 60(4), pp. 889–917

Bouis, Romain, Kuno J.M. Huisman, and Peter M. Kort (2009) 'Investment in oligopoly under uncertainty: The accordion effect.' *International Journal of Industrial Organization* 27(2), 320 – 331

Bresnahan, Timothy F., and Peter C. Reiss (1991) 'Empirical models of discrete games.' *Journal of Econometrics* 48(12), 57 – 81

Bulan, Laarni, Christopher Mayer, and C. Tsuriel Somerville (2009) 'Irreversible investment, real options, and competition: Evidence from real estate development.' *Journal of Urban Economics* 65(3), 237 – 251

Cai, Z., M. Das, H. Xiong, and X. Wu (2006) 'Functional coefficient instrumental variables models.' *Journal of Econometrics* 133, 207–241

Cai, Zongwu, and Qi Li (2008) 'Nonparametric estimation of varying coefficient dynamic panel data models.' *Econometric Theory* 24(5), 1321–1342

Chaubey, Yogendra P (2002) 'Estimation in inverse gaussian regression: comparison of asymptotic and bootstrap distributions.' *Journal of Statistical Planning and Inference* 100(2), 135 – 143

Clapp, John M., Katsiaryna Salavei Bardos, and S.K. Wong (2012) 'Empirical estimation of the option premium for residential redevelopment.' *Regional Science and Urban Economics* 42(1?), 240 – 256

Dixit, Avinash K, and Robert S Pindyck (1994) *Investment under uncertainty,* first ed. (Princeton University Press)

Einav, Liran (2010) 'Not all rivals look alike: Estimating an equilibrium model of the release date timing game.' *Economic Inquiry* 48(2), 369–390

Fan, Jianqing, and W Y Zhang (2008) 'Statistical methods with varying coefficient models.' *Statistics and Its Interface* 1(1), 179–195

Folks, J. L., and R. S. Chhikara (1978) 'The inverse gaussian distribution and its statistical application–a review.' *Journal of the Royal Statistical Society. Series B (Methodological)* 40(3), pp. 263–289

Folks, J. Leroy, and Anne S. Davis (1981) 'Regression models for the inverse gaussian distribution.' *Statistical Distributions in Scientific Work-NATO Advanced Study Institutes Series* 79(3), 91 – 97

Grenadier, Steven R. (1996) 'The strategic exercise of options: Development cascades and overbuilding in real estate markets.' *The Journal of Finance* 51(5), 1653–1679

Honoré, Bo E., and Áureo de Paula (2010) 'Interdependent durations.' *Review of Economic Studies* 77(3), 1138–1163

_ (2014) 'Interdependent durations in joint retirement decisions.' *working paper*

Hu, Yingyao, and Geert Ridder (2012) 'Estimation of nonlinear models with mismeasured regressors using marginal information.' *Journal of Applied Econometrics* 27(3), 347–385

Janssens, Gert, and Peter M. Kort (2012) 'Real options in an oligopoly under incomplete and asymmetric information.' *working paper*

Kong, Jean J. Kong, and Yue Kuen Kwok (2007) 'Real options in strategic investment games between two asymmetric firms.' *European Journal of Operational Research* 181, 967–985

Lambrecht, Bart, and William (2003) 'Real options and preemption under incomplete information.' *Journal of Economic Dynamics and Control* 27(4), 619 – 643

Lee, Mei-Ling Ting, and G. A. Whitmore (2006) 'Threshold regression for survival analysis: Modeling event times by a stochastic process reaching a boundary.' *Statistical Science* 21(4), pp. 501–513

Liu, Di, Irina Murtazashvili, and Artem Prokhorov (2015) 'Two-sample nonparametric estimation of intergenerational income mobility in the united states and sweden.' *Canadian Journal of Economics*

Lu, Tzon-Tzer, and Sheng-Hua Shiou (2002) 'Inverses of 2 2 block matrices.' *Computers & Mathematics with Applications* 43, 119–129

M. Eshaghi Gordji, H. Baghani, and O. Baghani (2011) 'On existence and uniqueness of solutions of a nonlinear integral equation.' *Journal of Applied Mathematics*

Mason, Robin, and Helen Weeds (2010) 'Investment, uncertainty and preemption.' *International Journal of Industrial Organization* 28(3), 278 – 287

Novy-Marx, Robert (2007) 'An equilibrium model of investment under uncertainty.' *The Review of Financial Studies* 20(5), pp. 1461–1502

Pawlina, Grzegorz, and Peter M. Kort (2006) 'Real options in an asymmetric duopoly: Who benefits from your competitive disadvantage?' *Journal of Economics & Management Strategy* 15(1), 1–35

Rice, John (1984) 'Bandwidth choice for nonparametric regression.' *Annals of Statistics* 12(4), 1215–1230

Ridder, Geert, and Robert Moffitt (2007) 'Chapter 75 the econometrics of data combination.' vol. 6, Part B of *Handbook of Econometrics* (Elsevier) pp. 5469 – 5547

Schmidt-Dengler, Philipp (2006) 'The timing of new technology adoption: The case of mri.' *Society for Economic Dynamics*

Stokey, Nancy L. (2008) *The Economics of Inaction:Stochastic Control Models with Fixed Costs,* first ed. (Princeton University Press)

Su, Liangjun, Irina Murtazashvili, and Aman Ullah (2013) 'Local linear gmm estimation of functional coefficient iv models with an application to estimating the rate of return to schooling.' *Journal of Business & Economic Statistics* 31(2), 184

Takahashi, Yuya (2013) 'Estimating a war of attrition: The case of the u.s. movie theater industry.' *working paper*

Takatsuka, Hajime (2004) 'Existence conditions of the optimal stopping time: The cases of geometric brownian motion and arithmetic brownian motion.' *Journal of the Operations Research Society of Japan* 47(3), 145–162

Wang, Ko, and Yuqing Zhou (2006) 'Equilibrium real options exercise strategies with multiple players: The case of real estate markets.' *Real Estate Economics* 34(1), 1–49

Wang, Qi, and Jinhong Xie (2014) 'Decomposing pioneer survival: Implications for the order-of-entry effect.' *Journal of Product Innovation Management* 31(1), 128–143

Weeds, Helen (2002) 'Strategic delay in a real options model of r and d competition.' *The Review of Economic Studies* 69(3), 729–747

You, Jinhong, Yong Zhou, and Gemai Chen (2006) 'Corrected local polynomial estimation in varying-coefficient models with measurement errors.' *Canadian Journal of Statistics.* 34(3), 391–410

Appendix A

Appendix to Chapter 1

A.1 Expected Value Functions for the Stochastic Control Problem

Based on the timing order of the N firms' investment decisions, we will classify the entry order of the N firms as the first, ..., the kth, ..., the Nth active player in the market. And we will distinguish the identity of the N firms in the subscript accordingly as firm i or its rival j, $\forall i, j \in \{1, 2, ..., N\}$.

To solve the expected value functions of N firms, we start with the Nth player's problem, which is equivalent to a single agent's stochastic control problem. The solution method has been orderly explored in the book of Dixit and Pindyck (1994) and Stokey (2008). We then solve the expected value function of the $(N-1)$th player. Following the same method, we can derive the expected value functions of the kth active firm in the market, $k \in \{1, ..., N\}$.

Suppose that firm i is the Nth player. Before the investment, firm i holds the opportunity to invest, the value of which can also be seen as an option value. Firm i chooses optimal investment time t_{iN} to maximize

$$V_i^N(y_t \mid x) = \sup_{t_{iN}} \mathbb{E}_t \left(\int_{t_{iN}}^{\infty} e^{-r(u-t)} D_N(x) y_u du - e^{-r(t_{iN}-t)} C_i \mid y_t \right). \quad \text{(A.1)}$$

To solve this stochastic control problem, we first define a reward function

$$\begin{aligned} g^N(s, y_s) &= \mathbb{E}_s \left(\int_{s}^{\infty} e^{-r(u-s)} D_N(x) y_u du \mid y_s \right) \\ &= \frac{y_s D_N(x)}{r - \mu}, \end{aligned} \quad \text{(A.2)}$$

and $g^N(t_{iN}, y_{t_{iN}})$ is denoted as the optimal reward function. Then the value function (A.1) is rewritten as

$$V_i^N(y_t \mid x) = \sup_{t_{iN}} \mathbb{E}_t \left(\left(g^N(t_{iN}, y_{t_{iN}}) - C_i \right) e^{-r(t_{iN}-t)} \mid y_t \right).$$

The solution to the optimal investment time t_{iN} is characterized as when the stochastic process y_t first crosses the threshold $y_{t_{iN}}$. Let $y_i^N = y_{t_{iN}}$. The Hamilton-Jacobi-Bellman equation in the continuation region(i.e. before the investment is made) is:

$$rV_i^N(y) = \mu y \frac{\partial V_i^N(y)}{\partial y} + \frac{1}{2}\sigma^2 y^2 \frac{\partial^2 V_i^N(y)}{\partial y^2}. \tag{A.3}$$

The general solution to (A.3) has the following form:

$$V_i^N(y) = Ay^\beta + By^\gamma,$$

where

$$\beta = \frac{1}{2} - \frac{\mu}{\sigma^2} + \sqrt{(\frac{\mu}{\sigma^2} - \frac{1}{2})^2 + \frac{2r}{\sigma^2}} > 1,$$

and

$$\gamma = \frac{1}{2} - \frac{\mu}{\sigma^2} - \sqrt{(\frac{\mu}{\sigma^2} - \frac{1}{2})^2 + \frac{2r}{\sigma^2}} < 0.$$

Applying the following boundary conditions:

$$V_i^N(y_i^N) = \frac{y_i^N D_N(x)}{r-\mu} - C_i, \tag{A.4}$$

$$\frac{\partial V_i^N(y_i^N)}{\partial y} = \frac{D_N(x)}{r-\mu}, \tag{A.5}$$

$$\lim_{y \longrightarrow 0} V_i^N(y) = 0. \tag{A.6}$$

(A.4) is the value-matching condition, (A.5) is the smooth-passing condition and (A.6) is the no bubble condition for value function $V_i^N(y)$ at $y = 0$. So the Nth player's optimal value function and investment trigger is given by

$$V_i^N(y_t \mid x) = \begin{cases} \left(\frac{y_i^N D_N(x)}{r-\mu} - C_i\right)\left(\frac{y_t}{y_i^N}\right)^\beta & \text{if} \quad y_t \le y_i^N \quad (wait) \\ \frac{y_t D_N(x)}{r-\mu} - C_i & \text{if} \quad y_t > y_i^N \quad (invest) \end{cases},$$

$$y_i^N = \frac{\beta}{\beta-1}\frac{(r-\mu)C_i}{D_N(x)}. \tag{A.7}$$

When $y_t \le y_i^N$, the value function is the value of the option to wait. It reflects the present value of the Nth player i's investment discounted back by a stochastic discount factor y_t/y_i^N from the random time of reaching the trigger value y_i^N. When $y_t > y_i^N$, the value function is the net present value of the project when immediate investment is optimal.

Next, we will calculate the $(N-1)$th player's expected option value. Now we assume firm i is the $(N-1)$th player and his optimal control problem is to choose optimal investment time t_{iN-1}, given the Nth player's optimal investment time t_N^*. As I will show in the next section A.2, although the entry order matters for firm i's optimal investment timing decision, in an SBNE, the entry order of firm i's rivals does not affect firm i's decision. In fact, firm i only considers the expected investment time of its rivals. As a result, we use t_N^* as the expected investment time of the Nth player that has not invested yet. Then the optimal stochastic problem of the $(N-1)$th player is:

$$
\begin{aligned}
&\mathbb{E}[V_i^{N-1}(y_t \mid x)] \\
&= \sup_{t_{iN-1}} \mathbb{E}_t \left(\int_{t_{iN-1}}^{t_N^*} e^{-r(u-t)} D_{N-1}(x) y_u du - e^{-r(t_{iN-1}-t)} C_i \right. \\
&\left. + \int_{t_N^*}^{\infty} e^{-r(u-t)} D_N(x) y_u du \mid y_t \right) \\
&= \sup_{t_{iN-1}} \mathbb{E}_t \left(\int_{t_{iN-1}}^{\infty} e^{-r(u-t)} D_{N-1}(x) y_u du - e^{-r(t_{iN-1}-t)} C_i \right. \\
&\left. + \int_{t_N^*}^{\infty} e^{-r(u-t)} \left(D_N(x) - D_{N-1}(x)\right) y_u du \mid y_t \right).
\end{aligned}
$$

Applying the reward function definition as in (A.2), we can divide the expected option value function $\mathbb{E}[V_i^{N-1}]$ into two parts,

$$
\begin{aligned}
&\mathbb{E}[V_i^{N-1}(y_t \mid x)] \\
&= \sup_{t_{iN-1}} \mathbb{E}_t \left(\left(g^{N-1}(t_{iN-1}, y_{t_{iN-1}}) - C_i\right) e^{-r(t_{iN-1}-t)} \mid y_t \right) \\
&+ \mathbb{E}_t \left(\left(g^N(t_N^*, y_{t_N^*}) - g^{N-1}(t_N^*, y_{t_N^*})\right) e^{-r(t_N^*-t)} \mid y_t \right).
\end{aligned}
$$

Therefore, the expected option value function $\mathbb{E}[V_i^{N-1}]$ is the summation of a perpetual payment flow $y_t D_{N-1}$, minus a sunk cost C_i, starting from t_{iN-1} and a perpetual dividend rate $y_t(D_N - D_{N-1})$ starting from t_N^*.

Let $y_N^* = y_{t_N^*}$ be the expected value of the Nth player's optimal threshold, and $y_i^{N-1} = y_{t_{iN-1}}$ be player i's threshold as the $(N-1)$th player. Apply the same solution method as in the Nth player's optimal stochastic control problem, we will have the expected value function before the investment as,

$$\mathbb{E}[V_i^{N-1}(y_t \mid x)]$$
$$= \left(\frac{y_i^{N-1} D_1(x)}{r-\mu} - C_i + \frac{y_N^*(D_N(x) - D_{N-1}(x))}{r-\mu} \left(\frac{y_i^{N-1}}{y_N^*} \right)^{\beta} \right) \left(\frac{y_t}{y_i^{N-1}} \right)^{\beta}.$$

When $y_t \leq y_i^{N-1}$, the expected value function of $(N-1)$th player is the net present value from its own investment plus the decrease in the expected present value caused by its rival's eventual investment.

Now suppose that firm i invests as the kth active firm, followed by subsequent $N-k$ firms who seize the investment opportunity at time t_j^*, with expected investment threshold y_j^* accordingly, $j \in \{k+1, ..., N\}$. When firm i chooses investment threshold y_i, the expected option value function is

$$\mathbb{E}[V_i^k(y \mid y_i, y_{k+1}^*, ...y_N^*)]$$
$$= \left(\frac{y_i D_k(x)}{r-\mu} - C_i + \sum_{j=k+1}^{N} \left(\frac{y_j^*(D_j(x) - D_{j-1}(x))}{r-\mu} \right) \left(\frac{y_i}{y_j^*} \right)^{\beta} \right) \left(\frac{y}{y_i} \right)^{\beta}. \tag{A.8}$$

Therefore, the expected option value function $\mathbb{E}[V_i^k]$ consists of the net present value from its own investment, discounting back from the time when y_t hits y_i, plus the net present dividend caused by its rivals' eventual investment, discounting back from the time when y_t hits y_j^* sequentially, $j \in \{k+1, ..., N\}$.

A.2 Equilibrium Strategies

Recall that each firm makes its entry-time decision after observing the number of active firms in the market and current state value y_t. Suppose that firm i chooses a threshold y_i at time t, when there are l active firms in the market, $l \in \{0, ..., N-1\}$. Then firm i could enter as the kth firm into the market and is followed by subsequent $N-k$ investments of i's rivals, where $k \in \{l+1, ..., N-1\}$. Suppose that firm i's strategy as the kth active firm is $y_i = b^k(C)$, which is a function of investment cost, and firm i's expected value as the kth active firm is given in (A.8).

Now we consider the information set for firm i when firm i chooses to enter as the kth firm into the market. Each firm has observed a history of state values of y_t until the time t and obtained the information that none of the remaining firms have invested yet. And this information can be used to infer the probability of the next first-entry firm after time t. Let $\hat{y}$ be the highest value of state variable observed until the time t, $\hat{y} = \max_{0 \leq \tau \leq t}\{y_\tau\}$ and $\hat{C} = (b^k)^{-1}(\hat{y})$ be the highest cost draw that would lead a rival to invest at $\hat{y}$. And the fact that none of i's rivals has invested until $\hat{y}$ is reached means their thresholds are larger than $\hat{y}$, then i would learn that all of the $N-k$ rivals have the investment costs $C_j > \hat{C}$, $\forall j \in \{k+1, ..., N\}$.

The probability that any of i's rivals j, has cost C_j higher than i's cost C, conditional on $C_j > \hat{C}$ is $\frac{1-G(C)}{1-G(\hat{C})}$. To invest first against all of the $N-k$ rivals, the probability that none of the $N-k$ rivals will preempt is

$$Pr(t_i < t_j,\ \forall\ j \neq i) = \left(\frac{1-G(C_i)}{1-G(\hat{C})}\right)^{N-k} \equiv 1 - H^k(C_i). \tag{A.9}$$

Given investment cost C_i, firm i's problem to find a threshold $y_i = \{b^k(C_i), f_i^N | k = l+1, ..., N-1\}$ is equivalent to find its optimal investment type m, $m \in [\hat{C}, C_U]$, given the threshold function $b^k(m)$. Therefore we solve

$$\begin{aligned} &\max_{m \in [\hat{C}, C_U]} V(y_i, y_{-i}, N-l, y_t, \hat{C}, C_i) = \\ &\sum_{k=l+1}^{N-1} \left(1 - H^k(m)\right) \mathbb{E}[V_i^k(y_t \mid b^k(m), y_{k+1}, ..., y_N)] \\ &+ \left(1 - \sum_{k=l+1}^{N-1} \left(1 - H^k(m)\right)\right) V_i^N(y_t \mid f_i^N), \end{aligned} \tag{A.10}$$

where $y_{-i} = \{y^*_{k+1}, ..., y^*_N | k = l+1, ..., N-1\}$.

In the symmetric BNE, all the firms use the same threshold functions $\{b^k(C) | k \in \{1, ..., N-1\}, C \in [C_L, C_U]\}$. Therefore, the expected investment trigger y_j^*, $j \in \{k+1, ..., N\}$ in the total expected value function A.10 is given by

$$\begin{aligned} y_j^* &= \mathbb{E}[b^j(C) \mid C > C_i, C_i] \\ &= \frac{\int_{C_i}^{C_U} b^j(C) g(C) dC}{1 - G(C_i)} \equiv B_j(C_i),\ j \in \{k+1, ..., N-1\}, \end{aligned}$$

$$
\begin{aligned}
y_N^* &= \mathbb{E}[f_C^N \mid C > C_i, C_i] \\
&= \frac{\beta}{\beta-1}\frac{(r-\mu)}{D_N(x)}\frac{\int_{C_i}^{C_U} Cg(C)dC}{1-G(C_i)} \equiv B_N(C_i),
\end{aligned}
$$

where

$$
f_C^N = \frac{\beta}{\beta-1}\frac{(r-\mu)C}{D_N(x)},
$$

f_C^N is the investment trigger of the Nth active firm. From firm i's perspective, when it invests as the kth active firm, the expected trigger y_j^*, $j \in \{k+1, ..., N\}$ is the same across all of i's rivals, which depends on firm i's information set $\{\forall j \in \{k+1, ..., N\} : C_j > C_i\}$ and firm i's prior for all its competitors. Therefore, when we substitute $y_j = B_j(m)$ into (A.8), $\forall j \in \{k+1, ..., N\}$, the expected utility $\mathbb{E}[V_i^k]$ becomes

$$
\begin{aligned}
&\mathbb{E}[V_i^k(y_t \mid b^k(m), B_{k+1}(m), ..., B_N(m))] \\
&= \left(\frac{b^k(m)D_k(x_k)}{r-\mu} - C_i\right)\left(\frac{y_t}{b^k(m)}\right)^{\beta} \\
&+ \sum_{j=k+1}^{N}\left(\frac{B_j(m)(D_N(x_j) - D_{j-1}(x))}{r-\mu}\right)\left(\frac{y_t}{B_j(m)}\right)^{\beta}.
\end{aligned}
$$

As a result, the optimization problem (A.10) is rewritten as

$$
\begin{aligned}
&\max_{m\in[\hat{C},C_U]} V(b(m), B(m), N-l, y_t, \hat{C}, C_i) \\
&= \max_{m\in[\hat{C},C_U]} \sum_{k=l+1}^{N-1}\left(1 - H^k(m)\right)\mathbb{E}[V_i^k(y_t \mid b^k(m), B_{k+1}(m), ..., B_N(m))] \\
&+ \left(1 - \sum_{k=l+1}^{N-1}\left(1 - H^k(m)\right)\right)V_i^N(y_t \mid f_i^N).
\end{aligned}
\tag{A.11}
$$

The FOC to this optimization problem is

$$\begin{aligned} 0 = &\sum_{k=l+1}^{N-1} \left(1 - H^k(m)\right) \mathbb{E}[V_i^{k\prime}(y_t \mid b^k(m), B_{k+1}(m), ..., B_N(m))] \\ &- \sum_{k=l+1}^{N-1} H^{k\prime}(m)\mathbb{E}[V_i^k(y_t \mid b^k(m), B_{k+1}(m), ..., B_N(m))] \\ &+ \sum_{k=l+1}^{N-1} H^{k\prime}(m) V_i^N(y_t \mid f_i^N). \end{aligned}$$

When $\frac{h^k(m)}{1-H^k(m)}$ is properly defined (i.e. $0 < G(m) < 1$), the above FOC is rearranged as

$$\frac{h^k(m)}{1 - H^k(m)} = \frac{\mathbb{E}[V_i^{k\prime}(y_t \mid b^k(m), B_{k+1}(m), ..., B_N(m))]}{\mathbb{E}[V_i^k(y_t \mid b^k(m), B_{k+1}(m), ..., B_N(m))] - V_i^N(y_t \mid f_i^N)}.$$

where $h^k(m) \equiv H^{k\prime}(m)$.

Substituting the expression for $h^k(m)$ and $H^k(m)$ from (A.9), now the FOC for $b^k(m)$ becomes

$$\frac{(N-k)g(m)}{1 - G(m)} = \frac{\mathbb{E}[V_i^{k\prime}(y_t \mid b^k(m), B_{k+1}(m), ..., B_N(m))]}{\mathbb{E}[V_i^k(y_t \mid b^k(m), B_{k+1}(m), ..., B_N(m))] - V_i^N(y_t \mid f_i^N)}. \tag{A.12}$$

From the FOC (A.12) and equilibrium condition that $m = C_i$, we will derive the optimal strategy functions $b^k(C_i)$, $\forall k \in \{l+1, ..., N-1\}$ under proper technical conditions. Before we prove that the threshold functions $b^k(\cdot)$, $\forall k \in \{l+1, ..., N-1\}$ indeed constitute a SBNE, we will show how to solve the threshold function $b^k(\cdot)$ from the FOC in a two players game.

A.3 A Duopoly Example

When $k = N-1$, we have the duopoly case: firm i competes against its rival to decide when to enter the $(N-1)$th market. Let $b(C) = b^{N-1}(C)$ and $B(C) = B^N(C)$. And the FOC for $b(C)$ is

$$\begin{aligned} &g(C)[\mathbb{E}[V_i^{N-1}(y_t \mid b(C), B(C))] - V_i^N(y_t \mid f_i^N)] \\ &= \mathbb{E}[V_i^{N-1\prime}(y_t \mid b(C), B(C))](1 - G(C)). \end{aligned}$$

The intuition of the above FOC is that firm i's expected payoff gain when it wins against its rival to enter the $(N-1)$th market first is equal to its expected payoff loss when it loses the competition, where $g(C)$ is the probability that firm i wins at trigger $b(C)$ and $1-G(C)$ is the probability that one of i's competitors invest first, conditional on not investing until $b(C)$ is reached.

After substituting the expression for $\mathbb{E}[V_i^{N-1}]$ and V_i^N into the FOC, and evaluating at the equilibrium condition $C_i = C$, we obtain the following differential equation for optimal strategy $b(C)$:

$$\frac{g(C)}{1-G(C)} = \frac{I_1}{I_2},$$

where

$$I_1 = (1-\beta)\left[\frac{b'(C)D_{N-1}(x)}{r-\mu} + B'(C)\left(\frac{b(C)}{B(C)}\right)^{\beta}\left(\frac{(D_N(x)-D_{N-1}(x))}{r-\mu}\right)\right] + \beta\frac{b'(C)}{b(C)}C,$$

$$I_2 = \left(\frac{b(C)D_{N-1}(x)}{r-\mu} - C\right) + \frac{B(C)(D_N(x)-D_{N-1}(x))}{r-\mu}\left(\frac{b(C)}{B(C)}\right)^{\beta} - \left(\frac{f_C^N D_N(x)}{r-\mu} - C\right)\left(\frac{b(C)}{f_C^N}\right)^{\beta},$$

and

$$f_C^N = \frac{\beta}{\beta-1}\frac{(r-\mu)C}{D_N(x)}.$$

To obtain an explicit recursion formula for $b(C)$, we first rewrite the objective function (A.11) in the duopoly case:

$$\begin{aligned} &V(b(m), B(m), 2, y_t, \hat{C}, C_i) \\ &= \left(\frac{1-G(m)}{1-G(\hat{C})}\right)\left(\frac{y_t}{b(m)}\right)^{\beta}\left(\frac{b(m)D_{N-1}(x)}{r-\mu} - C_i \right. \\ &+ \left.\left(\frac{B(m)(D_N(x)-D_{N-1}(x))}{r-\mu}\right)\left(\frac{b(m)}{B(m)}\right)^{\beta}\right) \\ &+ \left(1-\frac{1-G(m)}{1-G(\hat{C})}\right)\left(\frac{y_t}{b(m)}\right)^{\beta}\left(\frac{b(m)}{f_i^N}\right)^{\beta}\left(\frac{f_i^N D_N(x)}{r-\mu} - C_i\right), \end{aligned} \tag{A.13}$$

where

$$f_i^N = \frac{\beta}{\beta - 1}\frac{(r-\mu)C_i}{D_N(x)}.$$

It is convenient to define the following functions:

$$1 - W(m) = \left(\frac{1 - G(m)}{1 - G(\hat{C})}\right)\left(\frac{y_t}{b(m)}\right)^{\beta},$$

$$R(m; C_i) = \left(\frac{B(m)(D_N(x) - D_{N-1}(x))}{r - \mu}\right)\left(\frac{b(m)}{B(m)}\right)^{\beta}$$
$$- \left(\frac{b(m)}{f_i^N}\right)^{\beta}\left(\frac{f_i^N D_N(x)}{r-\mu} - C_i\right).$$

As a result, the optimization problem (A.13) is reduced as

$$\begin{aligned}&\max_{m\in[\hat{C}, C_U]} V(b(m), B(m), 2, y_t, \hat{C}, C_i)\\ &= \max_{m\in[\hat{C}, C_U]} (1 - W(m))\left(\frac{b(m)D_{N-1}(x)}{r-\mu} - C_i + R(m; C_i)\right).\end{aligned} \tag{A.14}$$

The FOC is

$$\begin{aligned}0 = &-w(m)\left(\frac{b(m)D_{N-1}(x)}{r-\mu} - C_i + R(m; C_i)\right)\\ &+ (1 - W(m))\left(\frac{b'(m)D_{N-1}(x)}{r-\mu} + R'(m; C_i)\right),\end{aligned} \tag{A.15}$$

where w is W's positive and continuous derivative.

Here we insert the equilibrium condition $m = C_i \equiv C$ into (A.15) to obtain

$$\begin{aligned}-Cw(C) = &-w(C)\left(\frac{b(C)D_{N-1}(x)}{r-\mu} + R(C; C)\right)\\ &+ (1 - W(C))\left(\frac{b'(C)D_{N-1}(x)}{r-\mu} + R'(C; C)\right).\end{aligned} \tag{A.16}$$

To simplify the notation, we define $R(C) = R(C; C)$. However, $R'(C) \neq R'(C; C)$ and

$$R'(C) = R'(C; C) + \left(\frac{b(C)}{f_C^N}\right)^{\beta},$$

where

$$f_C^N = \frac{\beta}{\beta-1}\frac{(r-\mu)C}{D_N(x)}.$$

Integrate both sides of (A.16) from C to C_U

$$-\int_C^{C_U} yw(y)\,dy =$$
$$\int_C^{C_U}\left(\frac{d}{dy}\left[(1-W(y))\,(\frac{b(y)D_{N-1}(x)}{r-\mu}+R(y))\right]-(1-W(y))\left(\frac{b(y)}{f_y^N}\right)^\beta\right)dy,$$

where

$$f_y^N = \frac{\beta}{\beta-1}\frac{(r-\mu)y}{D_N(x)}.$$

Then use the fact that $1-W(C_U)=0$ since $1-G(C_U)=0$,

$$\int_C^{C_U} yw(y)\,dy$$
$$= (1-W(C))\left(\frac{b(C)D_{N-1}(x)}{r-\mu}+R(C)\right)+\int_C^{C_U}(1-W(y))\left(\frac{b(y)}{f_y^N}\right)^\beta dy. \tag{A.17}$$

Rearrange (A.17)

$$\frac{b(C)D_{N-1}(x)}{r-\mu}+R(C)$$
$$=\frac{1}{1-W(C)}\int_C^{C_U} yw(y)\,dy-\int_C^{C_U}\left(\frac{1-W(y)}{1-W(C)}\right)\left(\frac{b(y)}{f_y^N}\right)^\beta dy. \tag{A.18}$$

To obtain an explicit recursion formula, we first integrate $\int_C^{C_U} yw(y)\,dy$ by parts and then substitute the expression for $W(y)$ and $1-W(C)$ to get,

$$\frac{1}{1-W(C)}\int_C^{C_U} yw(y)\,dy = \frac{1}{1-W(C)}\left(C_U - CW(C) - \int_C^{C_U} W(y)\,dy\right)$$
$$=\frac{1}{1-W(C)}\left(C_U - CW(C)-\int_C^{C_U}\left(1-\left(\frac{1-G(y)}{1-G(\hat{C})}\right)\left(\frac{y_t}{b(y)}\right)^\beta\right)dy\right)$$
$$= C+\left(\int_C^{C_U}\left(\frac{1-G(y)}{1-G(C)}\right)\left(\frac{y_t}{b(y)}\right)^\beta dy\right).$$

Also,

$$\int_C^{C_U} \left(\frac{1-W(y)}{1-W(C)}\right)\left(\frac{b(y)}{f_y^N}\right)^{\beta} dy = \int_C^{C_U} \left(\frac{1-G(y)}{1-G(C)}\right)\left(\frac{b(C)}{f_y^N}\right)^{\beta} dy.$$

Therefore, (A.18) is rewritten as

$$\frac{b(C)D_{N-1}(x)}{r-\mu} = C - R(C) + \int_C^{C_U} \left(\left(\frac{b(C)}{b(y)}\right)^{\beta} - \left(\frac{b(C)}{f_y^N}\right)^{\beta}\right)\left(\frac{1-G(y)}{1-G(C)}\right) dy, \tag{A.19}$$

where

$$R(C) = \left(\frac{B(C)(D_N(x)-D_{N-1}(x))}{r-\mu}\right)\left(\frac{b(C)}{B(C)}\right)^{\beta} - \left(\frac{b(C)}{f_C^N}\right)^{\beta}\left(\frac{f_C^N D_N(x)}{r-\mu} - C\right),$$

$$B(C) = \frac{\beta}{\beta-1}\frac{(r-\mu)}{D_N(x)}\frac{\int_{C_i}^{C_U} Cg(C)dC}{1-G(C_i)}.$$

(A.19) shows that when C approaches its upper endpoint $C = C_U$, we have

$$\frac{b(C_U)D_N(x)}{r-\mu} \to C_U + V_{C_U}^F,$$

where

$$V_{C_U}^F = (\frac{f_{C_U}^N D_N(x)}{r-\mu} - C_U)(\frac{b(C_U)}{f_{C_U}^N})^{\beta}.$$

A.4 Proof of Theorem 1

Proof of Theorem 1: For the first part of the theorem, suppose that $b^{*k}(C)$ satisfies the recursive equation (1.8). Define the last term of equation (1.8) as $f_k(C;y)$, where

$$f_k(C;y) = \int_{C_i}^{C_U} \left(\left(\frac{b^k(C_i)}{b^k(y)}\right)^{\beta} - \left(\frac{b^k(C_i)}{f_y^N}\right)^{\beta}\right)\left(\frac{1-G(y)}{1-G(C_i)}\right)^{N-k} dy.$$

Note first that the integrand $f_k(C;y)$ is the product of two increasing functions of C. Differentiating the integrand $f_k(C;y)$ with C and using the product rule,

$$
\begin{aligned}
&\frac{\partial f_k(C;y)}{\partial C} = \\
&\int_C^{C_U} \beta \frac{b^{*k\prime}(C)}{b^{*k}(C)} \left(\left(\frac{b^{*k}(C)}{b^{*k}(y)}\right)^{\beta} - \left(\frac{b^{*k}(C)}{f_y^N}\right)^{\beta} \right) \left(\frac{1-G(y)}{1-G(C)}\right)^{N-k} dy \\
&- (N-k)\frac{g(C)}{1-G(C)} \int_C^{C_U} \left(\left(\frac{b^{*k}(C)}{b^{*k}(y)}\right)^{\beta} - \left(\frac{b^{*k}(C)}{f_y^N}\right)^{\beta} \right) \left(\frac{1-G(y)}{1-G(C)}\right)^{N-k} dy \\
&- \left(1 - (\frac{b^{*k}(C)}{f_C^N})^{\beta}\right).
\end{aligned}
$$

Differentiating $R^k(C)$ with C,

$$
\begin{aligned}
&R^{k\prime}(C) \\
&= (1-\beta) \sum_{j=k+1}^{N} B_j'(C) \left(\frac{(D_j(x) - D_{j-1}(x))}{r-\mu}\right) \left(\frac{b^{*k}(C)}{B_j(C)}\right)^{\beta} \\
&+ \sum_{j=k+1}^{N} \beta b^{*k\prime}(C) \left(\frac{(D_j(x) - D_{j-1}(x))}{r-\mu}\right) \left(\frac{b^{*}(C)}{B_j(C)}\right)^{\beta-1} \\
&- \beta \frac{b^{*k\prime}(C)}{b^{*k}(C)} \left(\frac{b^{*k}(C)}{f_C^N}\right)^{\beta} \left(\frac{f_C^N D_N(x)}{r-\mu} - C\right) + \left(\frac{b^{*k}(C)}{f_C^N}\right)^{\beta}.
\end{aligned}
$$

Now differentiating both sides of (1.8), we obtain

$$
\frac{b^{*k\prime}(C) D_k(x)}{r-\mu} = 1 - R^{k\prime}(C) + \frac{\partial f_k(C;y)}{\partial C}.
$$

After substituting the expression for $R^{k\prime}(C)$ and $\frac{\partial f_k(C;y)}{\partial C}$ into the above equation, it is equivalent to the boundary problem $(1.6 - 1.7)$.

Conversely, suppose that $b^{*k}(C)$ satisfies the boundary problem $(1.6 - 1.7)$. To prove it is also the solution to the recursive equation (1.8), let us use the duopoly case in **Appendix** A.3 for an example. In the duopoly case, the recursive equation (1.8) is the FOC, derived from (A.13). The boundary problem $(1.6 - 1.7)$ is the FOC for the same problem with a different but equivalent expression for the value function (A.11). For a general case, we follow the argument as in the duopoly case. Now the value function can be rewritten as

$$\max_{m\in[\hat{C},C_U]} \sum_{k=l+1}^{N-1} V^k(b^k(m), B^{k+1}(m), ..., B^N(m), N-l, y_t, \hat{C}, C_i)$$
$$= \max_{m\in[\hat{C},C_U]} \sum_{k=l+1}^{N-1} \Big(1 - W^k(m)\Big) \left(\frac{b^k(m)D_k(x)}{r-\mu} - C_i + R^k(m; C_i)\right),$$

where

$$1 - W^k(m) = \left(\frac{1-G(m)}{1-G(\hat{C})}\right)^{N-k} \left(\frac{y_t}{b^k(m)}\right)^{\beta},$$

$$R^k(m; C_i) = \sum_{j=k+1}^{N} \left(\frac{B_j(m)(D_j(x) - D_{j-1}(x))}{r-\mu}\right) \left(\frac{b^k(m)}{B_j(m)}\right)^{\beta} -$$
$$\left(\frac{b^k(m)}{f_i^N}\right)^{\beta} \left(\frac{f_i^N D_N(x)}{r-\mu} - C_i\right).$$

And the recursive equation (1.8) is the FOC to the above maximized value function.

For the second part of the theorem, we use the recursive equation (1.8) to prove the existence and uniqueness of $b^{*k}(\cdot)$, since the boundary problem (1.6 − 1.7) is equivalent to the recursive equation (1.8). To prove this, we first use the theorem from M. Eshaghi Gordji and Baghani (2011), which prove the existence and uniqueness of the solution to the following nonhomogeneous nonlinear Volterra integral equation:

$$b(c) = f(c) + \varphi\left(\int_{C_L}^{c} F(c, y, b(y))dy\right), \tag{A.20}$$

where $c \in (C_L, C_U)$, $b \in X$, $f(\cdot) : (C_L, C_U) \to R$ is a mapping, and F is a continuous function. X is a functional space. We take (X, d) to be a complete metric space with

$$d(f, g) = max_{x\in(C_L,C_U)} \mid f(x) - g(x) \mid,$$

for all $f, g \in X$ and assume that φ is a linear bounded transformation on X.

Theorem 14. *Suppose the integral equation (A.20) satisfies the following conditions*

1. *$\varphi : X \to X$ is a bounded linear transformation,*

2. *$f : (C_L, C_U) \to \mathrm{R}$ and $F : (C_L, C_U) \times (C_L, c) \times X$ are continuous,*

3. *there exists a integrable function $p : (C_L, C_U) \times (C_L, C_U) \to \mathrm{R}$ such that,*
$$|F(c, y, u) - F(c, y, v)| \leq p(c, y)\phi(| u - v |),$$
for each $c, y \in (C_L, C_U)$ and $u, v \in X$, where $\phi : [0, \infty) \to [0, \infty)$ satisfies the following conditions:
   - *ϕ is increasing,*
   - *for each x, $\phi(x) < x$,*
   - $\limsup_{x \longrightarrow t^+} \frac{\phi(x)}{x} < 1, \forall t \in [0, \infty)$.

4. $$\sup_{c \in (C_L, C_U)} \int_{C_L}^{C_U} p^2(c, y)dy \leq \frac{1}{\|\varphi\|^2}(C_U - C_L),$$

   then integral equation (A.20) has a unique fixed point in X.

We will check each conditions of the above theorem for integral equation (1.8):

$$\frac{b^k(C)D_k(x)}{r - \mu} = C - R^k(C) + \int_C^{C_U} \left(\left(\frac{b(C)}{b(y)}\right)^\beta - \left(\frac{b(C)}{f_y^N}\right)^\beta \right) \left(\frac{1 - G(y)}{1 - G(C)}\right)^{N-k} dy,$$

where

$$R^k(C) = \sum_{j=k+1}^{N} \left(\frac{B^j(C)(D_j(x) - D_{j-1}(x))}{r - \mu}\right) \left(\frac{b^k(C)}{B^j(C)}\right)^\beta - \left(\frac{b^k(C)}{f_C^N}\right)^\beta \left(\frac{f_C^N D_N(x)}{r - \mu} - C\right),$$

given $B^{K+1}(\cdot), ..., B^N(\cdot)$ defined in (1.4 $-$ 1.5).

Conditions $1 - 2$ are straightforward. We will focus on conditions 3 and 4. First, we define the following functions:

$$\left(\frac{b_1(C)}{b_1(y)}\right)^\beta - \left(\frac{b_1(C)}{f_y^N}\right)^\beta \equiv G(b_1(c), b_1(y), f_y^N) \equiv G(p_1, q_1, f_y^N),$$

where $G(p, q, f_y^N) = p^\beta(1 - (q/f_y^N)^\beta)/q^\beta$.

Similarly, we have

$$\left(\frac{b_2(C)}{b_2(y)}\right)^\beta - \left(\frac{b_2(C)}{f_y^N}\right)^\beta \equiv G(b_2(c), b_2(y), f_y^N) \equiv G(p_2, q_2, f_y^N).$$

Since $G(p, q, f_y)$ is continuously differentiable in p and q, we have

$$\begin{aligned}
G(p_1, q_1, f_y^N) - G(p_2, q_2, f_y^N) \\
&= G_1(p^\star, q^\star, f_y^N)|p_1 - p_2| + G_2(p^\star, q^\star, f_y^N)|q_1 - q_2| \\
&\leq (G_1(p^\star, q^\star, f_y^N) + G_2(p^\star, q^\star, f_y^N)) \max_{y\in(C_L,C_U)} |b_1(y) - b_2(y)| \\
&= \frac{G_1(p^\star, q^\star, f_y^N) + G_2(p^\star, q^\star, f_y^N)}{\mu} \mu \mid b_1(y^*) - b_2(y^*) \mid,
\end{aligned}$$

where G_1 and G_2 are the partial differential functions of G with respect to the first and the second variable; $p^\star$ and $q^\star$ lie in $[p_1, p_2]$ and $[q_1, q_2]$ respectively; constant μ is chosen such that $0 < \mu < 1$; $y^* = arg \max_{y\in(C_L,C_U)} |b_1(y) - b_2(y)|$. Now, we set $\phi(t) = \mu t$ and

$$p(c, y) = \frac{G_1(p^\star, q^\star, f_y^N) + G_2(p^\star, q^\star, f_y^N)}{\mu} \left(\frac{1 - G(y)}{1 - G(c)}\right)^{N-k}.$$

Hence

$$\begin{aligned}
&\mid F(c, y, b_1) - F(c, y, b_2) \mid \\
&= \mid \left(G(p_1, q_1, f_y^N) - G(p_2, q_2, f_y^N)\right) \left(\frac{1 - G(y)}{1 - G(c)}\right)^{N-k} \mid \\
&\leq p(c, y)\mu \mid b_1(y^*) - b_2(y^*) \mid,
\end{aligned}$$

then condition 3 is satisfied.

To check condition 4, note

$$\sup_{c\in(C_L,C_U)} \int_{C_L}^{C_U} p^2(c, y)dy \leq \left(\frac{G_1(p^\star, q^\star, f_{\bar{y}}^N) + G_2(p^\star, q^\star, f_{\bar{y}}^N)}{\mu}\right)^2 (C_U - C_L),$$

where $\bar{y}$ lie in (C_L, C_U). We will choose an interval (C_L, C_U) large enough and a value of μ to satisfy condition 4. Therefore, integral equation (1.8) has a unique fixed point.

For part 3 of the theorem, we need to establish that setting $m = C_i$, i.e., truth-telling, always maximizes the expected payoff. Let $V^k(b^k(m), C_i)$ be the expected payoff to player i, when his investment cost is C_i and his threshold is $b^k(m)$. Assume that all of his rivals use thresholds $b^k(C_j)$, $\forall j \neq i$. By **Theorem** 1, we solve the threshold functions backward. Therefore we will complete the proof by showing $V^k(b^k(C_i), C_i) > V^k(b^k(m), C_i)$, $\forall k \in \{1, ..., N-1\}$.

Firstly, we recall that

$$V^k(b^k(m), C_i) = \Big(1 - W^k(m)\Big)\left(\frac{b^k(m)D_k(x)}{r-\mu} - C_i + R^k(m; C_i)\right),$$

where

$$1 - W^k(m) = \left(\frac{1-G(m)}{1-G(\hat{C})}\right)^{N-k}\left(\frac{y_t}{b^k(m)}\right)^{\beta},$$

$$R^k(m; C_i) = \sum_{j=k+1}^{N}\left(\frac{B_j(m)(D_j(x) - D_{j-1}(x))}{r-\mu}\right)\left(\frac{b^k(m)}{B_j(m)}\right)^{\beta} - \left(\frac{b^k(m)}{f_i^N}\right)^{\beta}\left(\frac{f_i^N D_N(x)}{r-\mu} - C_i\right).$$

Substituting the optimal threshold function $b^k(\cdot)$ into $V^k(b^k(m), C_i)$, we have

$$V^k(b^k(m), C_i) = \int_m^{C_U} y w^k(y)dy - \Big(1 - W^k(m)\Big)C_i,$$

where $w^k(\cdot)$ is $W^k(\cdot)$'s derivative. Hence,

$$V^k(b^k(C_i), C_i) - V^k(b^k(m), C_i) = (m - C_i)W^k(m) + \int_m^{C_i} W(y)dy.$$

Applying the mean value theorem, there exists a point x between C_i and m, such that

$$V^k(b^k(C_i), C_i) - V^k(b^k(m), C_i) = (m - C_i)(W^k(m) - W^k(x)) \geq 0.$$

since $W^k(\cdot)$ is an increasing function and x is between C_i and m guaranteed by mean value theorem. Hence when $m = C_i$, $b^k(\cdot)$ is a best response function and $b^k(\cdot)$ is a symmetric BNE.

Appendix B

Appendix to Chapter 2

B.1 Proof of Theorem 4

Suppose that $\xi_i(y)$ and $\xi_j(y)$ are the break-even functions for each firm. We study the behaviour of both functions when the domain of the functions are $(0, f_i)$. First, we rewrite these functions in the following form

$$\xi_i(y) = -a_i y^\beta + c_i y - 1,$$

where the coefficients are given by

$$\begin{aligned} a_i &= \frac{1}{\beta-1}\left[\beta\left(\frac{1}{l_i}-\frac{1}{f_i}\right)\left(\frac{1}{f_j}\right)^{\beta-1}+\left(\frac{1}{f_i}\right)^{\beta}\right], \\ c_i &= \frac{\beta}{\beta-1}\frac{1}{l_i}. \end{aligned}$$

Similarly, we have the expression of $\xi_j(y)$ for firm j by switching the indexes i and j in the expression of $\xi_i(y)$.

Clearly, $\xi_i(y)$ and $\xi_j(y)$ intersect at a unique non-zero point

$$\hat{y} = \left(\frac{c_j - c_i}{a_j - a_i}\right)^{\frac{1}{\beta-1}},$$

where $c_i > c_j$ because of $l_i < l_j$ and $sign(a_i - a_j) = sign(\hat{y})$.

We will show $\xi_i(y) > \xi_j(y)$ holds when $y \in (0, f_i)$. Firstly, we will prove $\xi_i(f_i) > 0$ and $\xi_j(f_i) \leq 0$ as follows:

$$\begin{aligned} \xi_i(f_i) &= \frac{1}{K_i}\left[V_i^L(f_i) - V_i^F(f_i)\right] \\ &= \frac{1}{K_i}\frac{1}{f_j}\left[\left(\frac{D_1-D_2}{r-\mu}\right)\left(\frac{f_i}{f_j}\right) - \left(\frac{D_1-D_2}{r-\mu}\right)\left(\frac{f_i}{f_j}\right)^{\beta}\right] > 0, \end{aligned} \tag{B.1}$$

$$\begin{aligned}\xi_j(f_i) &= \frac{1}{K_j}\left[V_j^L(f_i) - V_j^F(f_i)\right] \\ &= \frac{1}{K_j}\left[\frac{D_2 f_i}{r-\mu} - K_j - \left(\frac{D_2 f_j}{r-\mu} - K_j\right)\left(\frac{f_i}{f_j}\right)^{\beta}\right] \le 0.\end{aligned} \tag{B.2}$$

Note that (B.1) holds due to the following properties: $D_1 > D_2$, $f_i < f_j$, and $\beta > 1$. And (B.2) is from the expression of the value function of firm j being a follower: when evaluated at the time that $y = f_i < f_j$, the opportunity value before investment is larger than the value after the investment.

As a result, $\xi_i(f_i) > \xi_j(f_i)$ while is equivalent to the following expression

$$(a_j - a_i)f_i^{\beta} + (c_i - c_j)f_i > 0.$$

Therefore, we must have $f_i < \left(\frac{c_j - c_i}{a_j - a_i}\right)^{\frac{1}{\beta-1}} = \hat{y}$ when $a_j - a_i < 0$. This means $\xi_i(y)$ and $\xi_j(y)$ do not intersect on the interval $(0, f_i)$ and we must have $\xi_i(y) > \xi_j(y)$ when $y \in (0, f_i)$. However, when $a_j - a_i > 0$, since $\hat{y} < 0$ and $\xi_i(f_i) > \xi_j(f_i)$, we also have $\xi_i(y) > \xi_j(y)$ when $y \in (0, f_i)$.

Since b_i and b_j are the lowest value that $\xi_i(y) = 0$ and $\xi_j(y) = 0$ respectively when $y \in (0, f_i)$, b_i is always lower than b_j.

B.2 Proof of Theorem 5

Firstly, we show that $b_i < l_i < f_i < f_j$. We evaluate firm i's break-even function at l_i,

$$\begin{aligned}&\xi_i(l_i) \\ &= \frac{1}{K_i}\left(\frac{l_i D_1(x)}{r-\mu} - K_i + \frac{f_j(D_2(x) - D_1(x))}{r-\mu}\left(\frac{l_i}{f_j}\right)^{\beta} - \right. \\ &\left.\left(\frac{f_i D_2(x)}{r-\mu} - K_i\right)\left(\frac{l_i}{f_i}\right)^{\beta}\right) \\ &= \frac{1}{K_i}\left(\frac{1}{\beta-1}K_i + \frac{f_j(D_2(x) - D_1(x))}{r-\mu}\left(\frac{l_i}{f_j}\right)^{\beta} - \frac{f_i D_2(x)}{\beta(r-\mu)}\left(\frac{l_i}{f_i}\right)^{\beta}\right) \\ &= \frac{1}{\beta-1}\left(1 - \left(\frac{D_2(x)}{D_1(x)}\right)^{\beta-1}\left(\beta\left(\frac{f_i}{f_j}\right)^{\beta-1}\left(1 - \frac{D_2(x)}{D_1(x)}\right) + \frac{D_2(x)}{D_1(x)}\right)\right) \\ &> \frac{1}{\beta-1}\left(1 - \left(\frac{D_2(x)}{D_1(x)}\right)^{\beta-1}\left(\beta\left(1 - \frac{D_2(x)}{D_1(x)}\right) + \frac{D_2(x)}{D_1(x)}\right)\right).\end{aligned}$$

where the last inequality comes from the assumption that $D_2(x)/D_1(x) < 1$ (negative externalities) and $f_i < f_j$ (cost assumption of $\widetilde{K_{i2}} < \widetilde{K_{j2}}$ when $K_i < K_j$), which implies $(f_i/f_j)^{\beta-1} < 1$ when $\beta - 1 > 0$.

Denote $x = \frac{D_2(x)}{D_1(x)}$ and define a function $g(x)$ as the following:

$$g(x) = 1 - (x)^{\beta-1}(\beta(1-x) + x).$$

We have $g(1) = 0$ and $g(x)' = \beta(\beta-1)x^{\beta-2}(x-1) < 0$ when $x < 1$. Consequently, $g(D_2(x)/D_1(x)) > 0$ and $\xi(l_i) > 0$ implies $b_i < l_i$.

$l_i < f_i$ holds from the negative externalities assumption that $D_2(x) < D_1(x)$. Also, $b_j < f_i$ because firm j's break-even function $\xi_j(y)$ can only be positive when $y < f_i$. Therefore, inequalities $b_i < \min(b_j, l_i) < f_i < f_j$ hold in this game.

By considering the thresholds inequalities and the properties that both the leader's and the follower's value functions are strictly increasing (Pawlina and Kort (2006), Janssens and Kort (2012)), it is better off for the low cost firm i to be the leader than be the follower. Therefore firm i's optimal strategy is to invest at $min(b_j, l_i)$.

B.3 Proof of Theorem 6

Firstly, we study the conditions when the sequential equilibrium (l_i, f_j) is played, which is implied by $l_i < b_j$. Since b_j is the smallest value to make $\xi_j(y) = 0$, therefore we have

$$l_i < b_j \Leftrightarrow \begin{cases} \xi_j(l_i) < 0 & \text{, and} \\ \xi_j'(l_i) > 0. \end{cases}$$

Substituting the expressions of $\xi_j(l_i)$ and $\xi_j'(l_i)$ into the above two inequalities, we obtain

$$\left(\frac{\beta}{\beta-1}\right)\frac{l_i}{l_j} - 1 < \frac{1}{\beta-1}\left[\beta\left(\frac{1}{l_j} - \frac{1}{f_j}\right)\left(\frac{1}{f_i}\right)^{\beta-1} + \left(\frac{1}{f_j}\right)^{\beta}\right] l_i^{\beta},$$

$$\left(\frac{\beta}{\beta-1}\right)\frac{l_i}{l_j} > \frac{\beta}{\beta-1}\left[\beta\left(\frac{1}{l_j} - \frac{1}{f_j}\right)\left(\frac{1}{f_i}\right)^{\beta-1} + \left(\frac{1}{f_j}\right)^{\beta}\right] l_i^{\beta}.$$

Combining the above two inequalities, we have

$$\begin{aligned}\left(\frac{\beta}{\beta-1}\right)\frac{l_i}{l_j}-1 &< \frac{1}{\beta-1}\left[\beta\left(\frac{1}{l_j}-\frac{1}{f_j}\right)\left(\frac{1}{f_i}\right)^{\beta-1}+\left(\frac{1}{f_j}\right)^{\beta}\right]l_i^{\beta} \\ &< \left(\frac{1}{\beta-1}\right)\frac{l_i}{l_j}.\end{aligned} \tag{B.3}$$

Next, we study the conditions that the preemptive equilibrium (b_j, f_j) is played, which means $b_j < l_i$. Similarly, condition $\xi_j(l_i) > 0$ holds

$$\left(\frac{\beta}{\beta-1}\right)\frac{l_i}{l_j}-1 > \frac{1}{\beta-1}\left[\beta\left(\frac{1}{l_j}-\frac{1}{f_j}\right)\left(\frac{1}{f_i}\right)^{\beta-1}+\left(\frac{1}{f_j}\right)^{\beta}\right]l_i^{\beta}. \tag{B.4}$$

Also note because of $l_i < l_j$ the following inequality always holds

$$\frac{1}{\beta-1}\frac{l_i}{l_j} > \left(\frac{\beta}{\beta-1}\frac{l_i}{l_j}-1\right).$$

Then, we must also have $\xi_j'(l_i) > 0$ when $\xi_j(l_i) > 0$. As a result, $b_j < l_i$ when condition (B.4) holds.

B.4 A Three-player Real Options Game with Complete Information

This section develops a three-player real options game which is an extension of the asymmetric duopoly game studied intensively by Pawlina and Kort (2006) and Kong and Kwok (2007), etc. In the two-player game, the entry order is according to the efficiency order, such that the more efficient firm with a lower entry cost always enters the market first. In the three-player game, however, the entry order is not necessarily the same as the efficiency order. Argenziano and Schmidt-Dengler (2012) show that the inefficient firm can preempt the efficient firm in the three-player game even when there are two types of players.

B.4.1 Setup

Assume there are three different risk-neutral developers who compete for the optimal timing of real estate development projects. Each developer is assumed to have a perpetual investment opportunity on a development site. The investment decision is irreversible and the sunk cost of investment is different for each firm. Before the investment, the developers have an option to wait for their optimal timing of investment and receive no profits. At any point in time, a firm i can choose to invest an amount of K_i, $i = \{A, B, C\}$, and obtain an instantaneous profit flow of

$$\pi(y_t, D_j(x)) = y_t D_j(x), j = \{1, 2, 3\},$$

where j stands for the total number of firms that have entered. For example, when $j = 1$, $\pi(y_t, D_1(x))$ stands for the monopoly profit for firm i.

The state variable y_t represents demand shock of real estate market and follows a Geometric Brownian Motion:

$$dy_t = \mu y_t dt + \sigma y_t dW_t,$$

where μ and σ are instantaneous drift parameter and instantaneous standard deviation parameter. dt is the time increment and dW_t is the Wiener increment. Parameter μ can be interpreted as the industry growth rate and parameter σ as the industry volatility. We assume the stochastic process starts from y_0 low enough such that immediate investment is not optimal for either firm.

The deterministic multiplier $D_j(x_i)$, $i = \{A, B, C\}$, stands for the market externality. Note the firms are asymmetry in the sunk costs. x_i are covariates used for identification. We define the profit adjusted cost for firm i, $i = \{A, B, C\}$, as $\widetilde{K_{ij}} = K_i / D_j(x)$, $j = \{1, 2, 3\}$.

The following assumptions are made:

Assumption 17. $D_1(x) > D_2(x) > D_3(x)$.

Assumption 18. $K_A < K_B < K_C$.

Assumption 19. $\widetilde{K_{Aj}} < \widetilde{K_{Bj}} < \widetilde{K_{Cj}}$, $j = \{1, 2, 3\}$.

B.4.2 The Two-Firm Subgame

In this section, we use backward induction to analyze this complete information game. Suppose that the first firm has already entered the market, and we consider the outcome of the ensuing two-firm subgame. This subgame is analogous to the asymmetric duopoly game analyzed by Pawlina and Kort (2006). The outcome is that the lower cost firm always enters the market first. The optimal development time for the last player is derived in Appendix A.1. The optimal investment thresholds for the follower

i, $i = \{A, B, C\}$, is given as

$$f_i = \frac{\beta}{\beta - 1} \frac{(r - \mu)K_i}{D_2(x)}.$$

In this duopoly game, the leader i's the investment trigger without preemption threat is given by

$$l_i = \frac{\beta}{\beta - 1} \frac{(r - \mu)K_i}{D_1(x)}. \tag{B.5}$$

This is analogous to the stand-alone investment trigger in Argenziano and Schmidt-Dengler (2012). Without preemption threat, the leader will act as a single decision maker and choose the optimal investment time to receive the perpetual payment of $y_t D_1(x)$ minus the investment cost K_i, and the perpetual dividend of $y_t(D_2(x) - D_1(x))$, starting at T_j. And the follower will invest at f_j

With preemption threat, we have to consider each firm's incentive to be the leader in the two-firm subgame. As in Kong and Kwok (2007), we define the function

$$\xi_i(y) \equiv \frac{V_i^L(y) - V_i^F(y)}{K_i}, i = \{A, B, C\},$$

as a measurement of the preemptive incentive of firm i. If $\xi_i(y)$ is positive, then firm i prefers to be the leader rather than the follower. Define b_i as the lowest value of y such that $\xi_i(y) = 0$, and b_i is the preemptive trigger of firm i, $i = \{A, B, C\}$. Therefore, the preemptive trigger is the first time a firm's value function in a leader position exceeds in a follower position. Consequently, one firm's preemptive trigger must occur before both its rival followers' triggers (when the firm is in a leader position) and the follower's trigger for itself (when the firm is in a follower position).

B.4.3 Multiple Equilibria in the Three-Player Game

Before we analyze the equilibrium strategy of each firm in the three-firm game, we need to identify the entry order first. However, once the the first identity is known, the following two firm enter the market according to their efficiency order (lower cost firm enters first). Therefore, if firm A is the first to invest, then the order of entry must be $A - B - C$; if firm B is the first investor, then the entry order becomes $B - A - C$; and if firm C is the leader, then the order is $C - A - B$. Like in the two-player game, we need to analyze each firm's incentive to be the leader and determine their optimal entry time accordingly.

Firstly, we analyze the equilibrium conditions for the equilibrium result $A-B-C$. We start with analyzing firm A's incentive to be the first investor, given the strategies of B and C, which is the relative positions of B and C in the entry order. We need to compare firm A's leader value function and its follower valuer function, which is equivalent to compare the value function of A in the entry order $A-B-C$ and in the entry order $B-A-C$. Next, we need to make sure firm B has no incentive to be the first investor, so we need to compare firm B's leader value function and its follower valuer function, which is equivalent to compare the value function of B in the entry order $B-A-C$ and in the entry order $A-B-C$. Finally, we need to verify firm C has no incentive to deviate its follower's position in the entry order. So we need to compare firm C's leader value function and its follower valuer function, which is equivalent to compare the value function of C in the entry order $C-A-B$ and in the entry order $A-B-C$.

Similarly we can analyze the equilibrium conditions for the other types of equilibrium, $B-A-C$ and $C-A-B$. But we can not separate these three different equilibria like in the duopoly case and as a result, we can not achieve point identification for all the model primitives.

Suppose that we are considering the entry order $i-j-k$, then value function of the first entry in the three-player game is

$$
\begin{aligned}
&V_i^{L1}(y_t \mid y_{2j}, y_{3k}) \\
&= \begin{cases}
\dfrac{y_t D_1(x)}{r-\mu} - K_i + \dfrac{y_{2j}(D_2(x)-D_1(x))}{r-\mu}\left(\dfrac{y_t}{y_{2j}}\right)^{\beta} + \dfrac{y_{3k}(D_3(x)-D_2(x))}{r-\mu}\left(\dfrac{y_t}{y_{3k}}\right)^{\beta} & \text{if } y_t \le y_{2j}, \\[2ex]
\dfrac{y_t D_2(x)}{r-\mu} - K_i + \dfrac{y_{3k}(D_3(x)-D_2(x))}{r-\mu}\left(\dfrac{y_t}{y_{3k}}\right)^{\beta} & \text{if } y_{2j} < y_t \le y_{3k}, \\[2ex]
\dfrac{y_t D_2(x)}{r-\mu} - K_i & \text{if } y_t > y_{3k},
\end{cases}
\end{aligned}
\tag{B.6}
$$

where y_{2j} is the threshold of the second entry, and y_{3k} is the threshold of the third entry. y_{2j} and y_{3k} are derived as in the two-players subgame.

After the first entry, if firm i enters the market as the second entry, then firm i's value function is

$$V_i^{L2}(y_t \mid y_{3k}) = \begin{cases} \frac{y_t D_1(x)}{r-\mu} - K_i + \frac{y_{3k}(D_2(x)-D_1(x))}{r-\mu}\left(\frac{y_t}{y_{3k}}\right)^{\beta} & \text{if} \quad y_t \leq y_{3k}, \\ \frac{y_t D_2(x)}{r-\mu} - K_i & \text{if} \quad y_t > y_{3k}. \end{cases} \tag{B.7}$$

Finally, if firm i is the third entry, then its value function is

$$V_i^{3F}(y_t \mid x) = \begin{cases} \left(\frac{y_{3i} D_3(x)}{r-\mu} - K_i\right)\left(\frac{y_t}{y_{3i}}\right)^{\beta} & \text{if} \quad y_t \leq y_{3i}, \\ \frac{y_t D_3(x)}{r-\mu} - K_i & \text{if} \quad y_t > y_{3i}, \end{cases} \tag{B.8}$$

where

$$y_{3i} = \frac{\beta}{\beta - 1}\frac{(r-\mu)K_i}{D_3(x)}.$$

Next we evaluate the incentive for each firm, such that $i - j - k$ is the equilibrium entry order. For firm i, define i's incentive function as

$$\xi_i(y_t) = V_i^{L1}(y_t) - V_i^{L2}(y_t), \quad y_t < y_{2j}.$$

Suppose that b_{1i} is the smallest value such that $\xi_i(y_t) = 0$, which is the break-even value of firm i.

Similarly, for firm j, define j's incentive function as

$$\xi_j(y_t) = V_j^{L1}(y_t) - V_j^{L2}(y_t), \quad y_t < y_{2i}.$$

Suppose that b_{1j} is the smallest value such that $\xi_j(y_t) = 0$, which is the break-even value of firm j.

For firm k, define k's incentive function as

$$\xi_k(y_t) = V_k^{L1}(y_t) - V_k^{3F}(y_t), \quad y_t < y_{3k}.$$

Suppose that b_{1k} is the smallest value such that $\xi_k(y_t) = 0$, which is the break-even value of firm k.

When $b_{1i} < b_{1j} < b_{1k} < y_{2j}$, then firm i's optimal threshold is $min\{l_{i1}, b_{1j}\}$,

$$l_{i1} = \frac{\beta}{\beta - 1}\frac{(r-\mu)K_i}{D_1(x)}.$$

And firm j and firm k's optimal threshold are y_{2j} and y_{3k} respectively.

When $b_{1i} < b_{1k} < b_{1j} < y_{2j}$, then firm i's optimal threshold is $min\{l_{i1}, b_{1k}\}$, and firm j and firm k's optimal threshold are y_{2j} and y_{3k} respectively.

However, both case $b_{1i} < b_{1j} < b_{1k} < y_{2j}$ and case $b_{1i} < b_{1k} < b_{1j} < y_{2j}$ are overlapped in the parameter space of investment cost, and can be shown to be nonempty numerically (see Argenziano and Schmidt-Dengler (2012) and Bouis, Huisman and Kort (2009)). In fact, because of the entry order, now the parameter space of the model to be identified is larger than the duopoly case. In the duopoly case, we observe the actions of the players (T_N, T_{N-1}) and need to identify the parameters of the payoff structure, which includes two groups of parameters, $(D_N(x), \mu, \sigma, G(\cdot))$ and $D_{N-1}(x)$. Now in the three-player case, we have to identify three groups of parameters in the payoff structure and two of the possible entry orders as well, while we only observe three possible actions (T_N, T_{N-1}, T_{N-2}). Therefore the parameters of the model can not be identified. As a result, the identification of this model has to be achieved in other ways.

Appendix C

Appendix to Chapter 3

C.1 Convergence of the Denominator

Proof of Lemma 9: Recall that

$$D_{n,0}(z) = \frac{1}{n}\sum_{i=1}^{n} K_h(Z_{2j(i)} - z)\begin{pmatrix} X_{1i}X_{1i}^T & X_{1i}X_{2j(i)} \\ X_{2j(i)}X_{1i}^T & X_{2j(i)}^2 \end{pmatrix}.$$

Let $D_{n,0_{(j_1,j_2)}}(z)$ denote the (j_1, j_2)th block matrix element of $D_{n,0}(z)$. Then we will prove that $D_{n,0_{(22)}}(z) \xrightarrow{p} f(z)\Omega_{(22)}(z)$ by showing that $D_{n,0_{(22)}}(z)$ converges to $f(z)\Omega_{(22)}(z)$ in mean-square. The convergence of the rest of $D_{n,0_{(j_1,j_2)}}(z)$ can be shown in a similar way.

Before we proceed, we first outline some properties of $Z_{2j(i)}, i \in \{1, ..., n\}$. Let $f_{j(i)}(\cdot), i \in \{1, ..., n\}$ denote the density of $Z_{2j(i)}, i \in \{1, ..., n\}$, which is also the distribution of Z_{2j} conditional on it being a nearest match to Z_{1i}, $j \in \{1, ...m\}$. Because the density of Z_{2j} is $f(z)$, then

$$\begin{aligned} f_{j(i)}(z) &= \sum_{j=1}^{m} Pr(j(i) = j \mid Z_{2j} = z)f(z) \\ &= f(z) \cdot \sum_{j=1}^{m}(\frac{1}{m} + o(\frac{1}{m})) \\ &= f(z)(1 + o(1)), \end{aligned}$$

where the second equality comes from the conditional probability result derived in Abadie and Imbens (2006)'s Additional Proofs on page 5. The above implies that the density of any $Z_{2j(i)}$ differs from another density of $Z_{2j(k)}$ and the population density by $o(1)$, $i, k \in \{1, ..., n\}, k \neq i$. Also, $C(j)$ is defined as the number of times that unit j in the sample $\{X_{2j}, Z_{2j}\}_{j=1}^{m}$ is used as a match to unit i in the sample $\{Y_i, X_{1i}, Z_{1i}\}_{i=1}^n$,

$$C(j) = \sum_{i=1}^{n} \mathbb{1}(j = j(i)), \quad j \in \{1, ...m\}$$

where $\mathbb{1}(\cdot)$ is the indicator function, equal to one if $j = j(i)$ is true and zero otherwise. Then $\mathbb{E}(C(j) \mid Z_{2j}) = \frac{n}{m}(1+o(1))$ is given in Abadie and Imbens (2006)'s Additional Proofs on page 11.

Let $Z^{1,2}$ denote all the Z from the two samples. Note that, the conditional expectation of $D_{n,0_{(22)}}(z)$ is

$$\begin{aligned}
\mathbb{E}\left(D_{n,0_{(22)}}(z) \mid Z^{1,2}\right) &= \frac{1}{n}\sum_{i=1}^{n} K_h(Z_{2j(i)} - z)\Omega_{(22)}(Z_{2j(i)}) \\
&= \frac{1}{n}\sum_{j=1}^{m} C(j) K_h(Z_{2j} - z)\Omega_{(22)}(Z_{2j}).
\end{aligned}$$

And the unconditional expectation is

$$\begin{aligned}
\mathbb{E}\left(D_{n,0_{(22)}}(z)\right) &= \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} K_h(Z_{2j(i)} - z)\Omega_{(22)}(Z_{2j(i)})\right) \\
&= \mathbb{E}\left(\frac{1}{n}\sum_{j=1}^{m} C(j) K_h(Z_{2j} - z)\Omega_{(22)}(Z_{2j})\right) \\
&= \frac{1}{n}\sum_{j=1}^{m} \mathbb{E}\left(\mathbb{E}(C(j) \mid Z_{2j}) K_h(Z_{2j} - z)\Omega_{(22)}(Z_{2j})\right) \\
&= \frac{m}{n}\,\mathbb{E}\left(\frac{n}{m}(1+o(1)) K_h(Z_{2j} - z)\Omega_{(22)}(Z_{2j})\right) \\
&= f(z)\Omega_{(22)}(z)(1 + o(1) + O(h^2)).
\end{aligned}$$

Next we are going to show the convergence of the variance of $D_{n,0_{(22)}}(z)$. Apply the law of total variance,

$$\begin{aligned}
&Var\left(D_{n,0_{(22)}}(z)\right) \\
&= \mathbb{E}\left(Var\left(D_{n,0_{(22)}}(z) \mid Z^{1,2}\right)\right) + Var\left(\mathbb{E}\left(D_{n,0_{(22)}}(z) \mid Z^{1,2}\right)\right).
\end{aligned}$$

First we consider $Var\left(\mathbb{E}\left(D_{n,0_{(22)}}(z) \mid Z^{1,2}\right)\right)$. Since $\Omega(z) = E(XX' \mid z)$ is a matrix, we transform it into a vector for its convergence. Let $\Omega_{(22)}(z)$ be the $(2,2)$th block matrix element of $\Omega(z)$. Denote

$$\begin{aligned} A_j &= vec\left(K_h(Z_{2j}-z)\Omega_{(22)}(Z_{2j}) - f(z)\Omega_{(22)}(z)\right), \\ A_{j(i)} &= vec\left(K_h(Z_{2j(i)}-z)\Omega_{(22)}(Z_{2j(i)}) - f(z)\Omega_{(22)}(z)\right). \end{aligned}$$

Then $Var\left(\mathbb{E}\left(D_{n,0_{(22)}}(z) \mid Z^{1,2}\right)\right)$ can be rewritten as

$$\begin{aligned} & Var\left(\mathbb{E}\left(D_{n,0_{(22)}}(z) \mid Z^{1,2}\right)\right) \\ &= \mathbb{E}\left((\frac{1}{n}\sum_{i=1}^{n} A_{j(i)})(\frac{1}{n}\sum_{i=1}^{n} A_{j(i)})^T \right) \\ &= \frac{1}{n^2}\mathbb{E}\left(\sum_{i=1}^{n}\sum_{t=1}^{n} A_{j(i)}A_{j(t)}^T\right) \\ &= \frac{1}{n^2}\mathbb{E}\left(\sum_{i=1}^{n}\sum_{\substack{j(t)=j(i)\\ t\in\{1,\dots,n\}}} A_{j(i)}A_{j(t)}^T\right) + \frac{1}{n^2}\mathbb{E}\left(\sum_{i=1}^{n}\sum_{\substack{j(t)\neq j(i)\\ t\in\{1,\dots,n\}}} A_{j(i)}A_{j(t)}^T\right) \\ &= \frac{1}{n^2}\mathbb{E}\left(\sum_{i=1}^{n} A_{j(i)}C(j(i))A_{j(i)}^T\right) + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{\substack{j(t)\neq j(i)\\ t\in\{1,\dots,n\}}} \mathbb{E}\left(A_{j(i)}\right)\mathbb{E}\left(A_{j(t)}^T\right) \\ &\leq \frac{1}{n^2}\mathbb{E}\left(\sum_{i=1}^{n}(\max_{j=1,\dots,m} C(j))A_{j(i)}A_{j(i)}^T\right) + \frac{n^2-n}{n^2}(o(1)^2 + O(h^2)^2) \\ &= \frac{1}{n^2}\mathbb{E}\left(\sum_{j=1}^{m}(\max_{j=1,\dots,m} C(j))C(j)A_jA_j^T\right) + O(h^4) + o(1)^2 \\ &= \frac{1}{\sqrt{n}}\left(\frac{m}{n}\right)^{\frac{3}{2}}\mathbb{E}\left((\frac{1}{\sqrt{m}}\max_{j=1,\dots,m} C(j))C(j)A_jA_j^T\right) + O(h^4) + o(1)^2 \\ &= \frac{1}{\sqrt{n}}\left(\frac{m}{n}\right)^{\frac{3}{2}}\mathbb{E}\left(\frac{1}{\sqrt{m}}(\max_{j=1,\dots,m} C(j))^2A_jA_j^T\right) + O(h^4) + o(1)^2 \\ &\leq \frac{1}{\sqrt{n}}\left(\frac{m}{n}\right)^{\frac{3}{2}}\mathbb{E}\left(\frac{1}{\sqrt{m}}(\max_{j=1,\dots,m} C(j))^2\right)\bar{\mu}_2 + O(h^4) + o(1)^2, \end{aligned}$$

where $\bar{\mu}_2 = \sup_{Z_j}\|A_jA_j^T\|$. $\bar{\mu}_2$ is finite due to **Assumption** 14 and **Assumption** 15, which implies that Ω and K_h are continuous on the finite support of Z, and in turn implies that K_h and Ω satisfy the Lipschitz condition. Next we are going to prove that $\mathbb{E}\left(\frac{1}{\sqrt{m}}(\max_{j=1,\dots,m} C(j))^2\right)$ is bounded uniformly in m, which implies that $Var\left(\mathbb{E}\left(D_{n,0_{(22)}}(z) \mid Z^{1,2}\right)\right)$ converges to zero matrix as n, m go to infinity of the same order.

Following the proofs on Page 23 in Abadie and Imbens (2006)'s Additional Proofs, by Bonferroni's inequality:

$$
\begin{aligned}
\mathbb{E}\left((\frac{1}{\sqrt{m}}(\max_{j=1,\dots,m} C(j))^2)^2\right) &= \frac{1}{m}\mathbb{E}\left(\max_{j=1,\dots,m} C(j)^4\right) \\
&= \frac{1}{m}\sum_{N=0}^{\infty}\Pr(\max_{j=1,\dots,m} C(j)^4 > N) \\
&\leq \frac{1}{m}\sum_{N=0}^{\infty} m\Pr(C(j)^4 > N) \\
&\leq \mathbb{E}\left(C(j)^4\right)
\end{aligned}
\tag{C.1}
$$

The first equation holds because $C(j)$ is a positive integer for all $j = 1, \dots, m$. According to **Lemma** 3 in Abadie and Imbens (2006) , given **Assumption** 13 and part 1 of **Assumption** 14, $\mathbb{E}\left(C(j)^q\right)$ is bounded uniformly in m for all $q > 0$. Since

$$
\mathbb{E}\left(\frac{1}{\sqrt{m}}(\max_{j=1,\dots,m} C(j))^2\right)^2 \leq \mathbb{E}\left((\frac{1}{\sqrt{m}}(\max_{j=1,\dots,m} C(j))^2)^2\right), \tag{C.2}
$$

which implies the boundedness of $\mathbb{E}\left(\frac{1}{\sqrt{m}}(\max_{j=1,\dots,m} C(j))^2\right)$, and in turn implies that $Var\left(\mathbb{E}\left(D_{n,0_{(22)}}(z) \mid Z^{1,2}\right)\right)$ converges to zero matrix as n, m go to infinity of the same order. By the same method, we can easily show that under certain finite moment assumptions, the expectation of $Var\left(D_{n,0_{(22)}}(z) \mid Z^{1,2}\right)$ goes to zero as sample sizes go to infinity.

To conclude, we have showed that $Var\left(D_{n,0_{(22)}}(z)\right)$ goes to zero as sample sizes go to infinity and therefore $D_{n,0_{(22)}}(z)$ converges to $f(z)\Omega_{(22)}(z)$ in mean-square. Similarly, we can show that all the block matrix elements of $D_{n,0}(z)$ converges to the respective block matrix elements of $f(z)\Omega(z)$ in mean-square. The convergence of $D_n(z)$ to $f(z)\Omega(z) \bigotimes \begin{pmatrix} 1 & 0 \\ 0 & \mu_2 \end{pmatrix}$ is proved in the same line of arguments.

C.2 Convergence of the Numerator

Proof of Lemma 10: We approximate the expression of $Y = X_i^T\beta(Z_{1i}) + u_i \simeq X_i^T\beta(Z_{2j(i)}) + u_i$ by a Taylor expansion in the neighbourhood of $|Z_{1i} - z| < h$ and $|Z_{2j(i)} - z| < h$.

$$X_i^T\beta(Z_{1i}) = X_i^T\beta(z) + (Z_{2j(i)} - z)X_i^T\beta'(z) + \frac{h^2}{2}\left(\frac{Z_{2j(i)} - z}{h}\right)^2 \beta(z)'' + o(h^2) \;\; a.s.,$$

where $\beta'(z)$ and $\beta(z)''$ are the vectors consisting of the first and the second derivatives of the function $\beta(z)$.

Then we substitute the approximate expression of Y into the expression of $N_n(z)$.

$$\begin{aligned} N_n(z) &= \frac{1}{n}(D^{X_m})^T WY \\ &= \frac{1}{n}(D^{X_m})^T W \left(D^X \begin{pmatrix}\beta(z)\\ h\beta'(z)\end{pmatrix} + \frac{h^2}{2}A_z X^T\beta''(z) + u + o(h^2)\right), \end{aligned}$$

where

$$D^{X_m} = \begin{pmatrix} X_{j(1)}^T & X_{j(1)}^T\frac{Z_{2j(1)}-z}{h} \\ \cdots & \cdots \\ X_{j(n)}^T & X_{j(n)}^T\frac{Z_{2j(n)}-z}{h} \end{pmatrix},$$

$$W = diag\left(K_h(Z_{2j(1)} - z), ..., K_h(Z_{2j(n)} - z)\right),$$

$$D^X = \begin{pmatrix} X_1^T & X_1^T\frac{Z_{2j(1)}-z}{h} \\ \cdots & \cdots \\ X_n^T & X_n^T\frac{Z_{2j(n)}-z}{h} \end{pmatrix},$$

and

$$A_z = diag((\frac{Z_{2j(1)} - z}{h})^2, ..., (\frac{Z_{2j(n)} - z}{h})^2).$$

Recall that $X_{j(i)} = (X_{1i}^T, X_{2j(i)})^T$ is the matching pair for $X_i = (X_{1i}^T, X_{2i})^T$, $i \in \{1, ...n\}$.

Next we decompose $N_n(z) - \frac{1}{n}(D^{X_m})^T W D^{X_m}\begin{pmatrix}\beta(z)\\ h\beta'(z)\end{pmatrix} = I_1 + I_2 + I_3 + I_4$, where

$$I_1 = \frac{1}{n}(D^{X_m})^T W \left(D^X\begin{pmatrix}\beta(z)\\ h\beta'(z)\end{pmatrix} - \mathbb{E}(D^X \mid Z^{1,2})\begin{pmatrix}\beta(z)\\ h\beta'(z)\end{pmatrix}\right),$$

$$I_2 = \frac{1}{n}(D^{X_m})^T W \left(\mathbb{E}(D^X \mid Z^{1,2}) \begin{pmatrix} \beta(z) \\ h\beta'(z) \end{pmatrix} - D^{X_m} \begin{pmatrix} \beta(z) \\ h\beta'(z) \end{pmatrix} \right),$$

$$I_3 = \frac{1}{n}(D^{X_m})^T W \frac{h^2}{2} A_z X^T \beta''(z),$$

$$I_4 = \frac{1}{n}(D^{X_m})^T W u.$$

Firstly, we consider I_1. Since

$$D^X - \mathbb{E}(D^X \mid Z^{1,2}) = \begin{pmatrix} (X_1 - g(Z_{11}))^T & (X_1 - g(Z_{11}))^T \frac{(Z_{2j(1)} - z)}{h} \\ \cdots & \cdots \\ (X_n - g(Z_{1n}))^T & (X_n - g(Z_{1n}))^T \frac{(Z_{2j(n)} - z)}{h} \end{pmatrix},$$

then

$$I_1 = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n X_{j(i)} \left(X_i - g\left(Z_{1i}\right)\right)^T K_h\left(Z_{2j(i)} - z\right) \\ \frac{1}{nh}\sum_{i=1}^n X_{j(i)} \left(X_i - g\left(Z_{1i}\right)\right)^T \left(Z_{2j(i)} - z\right) K_h\left(Z_{2j(i)} - z\right) \end{pmatrix} \beta(z)$$
$$+ \begin{pmatrix} \frac{1}{nh}\sum_{i=1}^n X_{j(i)} \left(X_i - g\left(Z_{1i}\right)\right)^T \left(Z_{2j(i)} - z\right) K_h\left(Z_{2j(i)} - z\right) \\ \frac{1}{nh^2}\sum_{i=1}^n X_{j(i)} \left(X_i - g(Z_{1i})\right)^T \left(Z_{2j(i)} - z\right)^2 K_h\left(Z_{2j(i)} - z\right) \end{pmatrix} h\beta'(z).$$

We will show that

$$I_1 \xrightarrow{p} f(z) \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ 0 & 0 \end{pmatrix} \beta(z) \bigotimes \begin{pmatrix} 1 \\ 0 \end{pmatrix} + f(z) \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ 0 & 0 \end{pmatrix} h\beta'(z) \bigotimes \begin{pmatrix} 0 \\ \mu_2 \end{pmatrix}.$$

Denote

$$I_{10} = \frac{1}{n}\sum_{i=1}^n X_{j(i)}(X_i - g(Z_{1i}))^T K_h(Z_{2j(i)} - z),$$

$$I_{11} = \frac{1}{nh}\sum_{i=1}^n X_{j(i)}(X_i - g(Z_{1i}))^T (Z_{2j(i)} - z) K_h(Z_{2j(i)} - z),$$

and

$$I_{12} = \frac{1}{nh}\sum_{i=1}^n X_{j(i)}(X_i - g(Z_{1i}))^T (Z_{2j(i)} - z)^2 K_h(Z_{2j(i)} - z).$$

Next, we will show that $I_{10} \xrightarrow{p} f(z) \begin{pmatrix} \Sigma_{11}(z) & \Sigma_{12}(z) \\ 0 & 0 \end{pmatrix}$. Suppose that $j(i) = j$, then X_{2j} is the match to X_{2i}. Denote $X_{(i,j)} = (X_{1i}^T, X_{2j}^T)^T$ as the match to $X_i = (X_{1i}^T, X_{2i})$, when $j(i) = j$.

$$
\begin{aligned}
\mathbb{E}(I_{10} \mid Z^{1,2}) &= \frac{1}{n}\sum_{j=1}^{m}\mathbb{E}\left(\sum_{i\in\mathbb{A}(j)} X_{(i,j)}(X_i - g(Z_{1i}))^T K_h(Z_{2j}-z) \mid Z^{1,2}\right) \\
&= \frac{1}{n}\sum_{j=1}^{m} K_h(Z_{2j}-z)\mathbb{E}\left(\sum_{i\in\mathbb{A}(j)} X_{(i,j)}(X_i - g(Z_{1i}))^T \mid Z^{1,2}\right) \\
&= \frac{1}{n}\sum_{j=1}^{m} K_h(Z_{2j}-z)\sum_{i\in\mathbb{A}(j)}\mathbb{E}\left(\begin{pmatrix} X_{1i} \\ X_{2j}\end{pmatrix}(v_{1i}, v_{2i})^T \mid Z^{1,2}\right) \\
&= \frac{1}{n}\sum_{j=1}^{m} K_h(Z_{2j}-z)\sum_{i\in\mathbb{A}(j)}\begin{pmatrix} \Sigma_{11}(Z_{1i}) & \Sigma_{12}(Z_{1i}) \\ 0 & 0\end{pmatrix} \\
&= \frac{1}{n}\sum_{j=1}^{m} K_h(Z_{2j}-z)C(j)\begin{pmatrix} \Sigma_{11}(Z_{2j}) & \Sigma_{12}(Z_{2j}) \\ 0 & 0\end{pmatrix}(1+o(1)) .
\end{aligned}
$$

The next to the last equality is from the fact that $\{X_{1i}\}_{i=1}^{n}$ and $\{X_{2j}\}_{j=1}^{m}$ are two independent samples from the same population. Since $\mathbb{A}(j)$ is defined as the subset of the index i, $i \in \{1, ..., n\}$, such that j is used as a match to each indexed observation, then the number of elements in the set $\mathbb{A}(j)$ is $C(j)$, $j \in \{1, ...m\}$. Also $C(j)$ is nonstochastic conditional on $Z^{1,2}$. As we discussed before, for a pair of match, $Z_{2j(i)}$ and Z_{1i}, their density only differ by $o(1)$. So for all $i \in \mathbb{A}(j)$, the density of Z_{1i} and Z_{2j} differ by $o(1)$ as well. Therefore, the last equality holds.

To compute the unconditional expectation of I_{10}, we apply the result of $\mathbb{E}(C(j)|Z^{1,2}) = \frac{n}{m}(1 + o(1))$ from Abadie and Imbens (2006). Then

$$
\mathbb{E}(I_{10}) = f(z)\begin{pmatrix} \Sigma_{11}(z) & \Sigma_{12}(z) \\ 0 & 0\end{pmatrix} .
$$

The convergence of the variance $Var(I_{10})$ is in the **Appendix** C.3. As a result, we have the convergence result of I_{10}. Similarly, we can show the convergence of I_{11} and I_{12}. And then the convergence of I_1 is straightforward.

Next we consider the convergence of I_2. Since

$$
\mathbb{E}(D^X \mid Z^{1,2}) - D^{X_m} = \begin{pmatrix} (g(Z_{11}) - X_{j(1)})^T & (g(Z_{11}) - X_{j(1)})^T \frac{(Z_{1j(1)}-z)}{h} \\ ... & ... \\ (g(Z_{1n}) - X_{j(n)})^T & (g(Z_{1n}) - X_{j(n)})^T \frac{(Z_{nj(n)}-z)}{h}\end{pmatrix},
$$

then

$$I_2 = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n X_{j(i)}(g(Z_{1i}) - X_{j(i)})^T K_h(Z_{2j(i)} - z) \\ \frac{1}{nh}\sum_{i=1}^n X_{j(i)}(g(Z_{1i}) - X_{j(i)})^T (Z_{1i} - z) K_h(Z_{2j(i)} - z) \end{pmatrix} \beta(z)$$
$$+ \begin{pmatrix} \frac{1}{nh}\sum_{i=1}^n X_{j(i)}(g(Z_{1i}) - X_{j(i)})^T (Z_{1i} - z) K_h(Z_{2j(i)} - z) \\ \frac{1}{nh^2}\sum_{i=1}^n X_{j(i)}(g(Z_{1i}) - X_{j(i)})^T (Z_{1i} - z)^2 K_h(Z_{2j(i)} - z) \end{pmatrix} h\beta'(z).$$

Following a similar argument, we can show that

$$I_2 \xrightarrow{p} f(z) \begin{pmatrix} -\Sigma_{11} & 0 \\ 0 & -\Sigma_{22} \end{pmatrix} \beta(z) \bigotimes \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$
$$+ f(z) \begin{pmatrix} -\Sigma_{11} & 0 \\ 0 & -\Sigma_{22} \end{pmatrix} h\beta'(z) \bigotimes \begin{pmatrix} 0 \\ \mu_2 \end{pmatrix}$$

Finally, we can compute in a similar way to prove that $I_3 = o(h^2)$ and $I_4 = o_p(1)$.

C.3 Convergence of $Var(I_{10})$

Let $vec(\cdot)$ denote the vectorization transformation of matrix and $VI_{10} = vec(I_{10})$.

Consider the variance decomposition,

$$Var\left(VI_{10}\right) = \mathbb{E}\left(Var(VI_{10} \mid Z^{1,2})\right) + Var\left(\mathbb{E}(VI_{10} \mid Z^{1,2})\right).$$

In the following, we will prove that the convergence of both $\mathbb{E}\left(Var\left(VI_{10} \mid Z^{1,2}\right)\right)$ and $Var\left(\mathbb{E}\left(VI_{10} \mid Z^{1,2}\right)\right)$ as sample sizes go to infinity.

First consider $\mathbb{E}\left(Var\left(VI_{10} \mid Z^{1,2}\right)\right)$. Let

$$B_{(i,j)} = K_h(Z_{j(i)} - z) \cdot vec\left(X_{j(i)}(X_i - g(Z_i))^T - \begin{pmatrix} \Sigma_{11}(Z_i) & \Sigma_{12}(Z_i) \\ 0 & 0 \end{pmatrix}\right),$$
$$B_{(i,i)} = K_h(Z_i - z) \cdot vec\left(X_i(X_i - g(Z_i))^T - \Sigma\right).$$

Then

$$\begin{aligned}
&\mathbb{E}\left(Var\left(VI_{10} \mid Z^{1,2}\right)\right)\\
&= \mathbb{E}\left(\left(VI_{10} - \mathbb{E}\left(VI_{10} \mid Z^{1,2}\right)\right)\left(VI_{10} - \mathbb{E}\left(VI_{10} \mid Z^{1,2}\right)\right)^T\right)\\
&= \mathbb{E}\left((\frac{1}{n}\sum_{i=1}^{n} B_{(i,j)})(\frac{1}{n}\sum_{i=1}^{n} B_{(i,j)})^T\right)\\
&\leq \frac{1}{\sqrt{n}}\left(\frac{m}{n}\right)^{\frac{3}{2}} \mathbb{E}\left(\frac{1}{\sqrt{m}}(\max_{j=1,\ldots,m} C(j))^2 B_{(i,i)} B_{(i,i)}^T\right) + O(h^4) + o(1)^2
\end{aligned}$$

Following the same assumptions and argument in proof of **Lemma** 9, we have that $\mathbb{E}\left(\frac{1}{\sqrt{m}}(\max_{j=1,\ldots,m} C(j))^2 B_{(i,i)} B_{(i,i)}^T\right)$ is bounded uniformly in all m and therefore $\mathbb{E}\left(Var\left(VI_{10} \mid Z^{1,2}\right)\right)$ converges to zero matrix as sample size n, m go to infinity of the same order.

Before we proceed, we introduce a useful lemma from Abadie and Imbens (2006) about the distribution of the *matching discrepancy*. Suppose that we have a random sample $Z_1, \ldots, Z_N$, with density f over bounded support Z. Now consider the closest match to a $z \in$ Z in the sample. Let $j_1 = \text{argmin}_{j=1,\ldots,N}\|Z_j - z\|$ and let $U_{j_1} = Z_{j_1} - z$ be the matching discrepancy.

Lemma 15. ***Matching Discrepancy - Asymptotic Properties:*** *Suppose that f is differentiable in a neighbourhood of z. Then $U_{j_1} = O_p(N^{-1})$. Moreover, the first two moments of U_{j_1} are $O(N^{-\frac{1}{2}})$.*

Now consider $Var\left(\mathbb{E}\left(VI_{10} \mid Z^{1,2}\right)\right)$. Define

$$A_{i,j(i)} = K_h(Z_{j(i)} - z) \cdot vec\begin{pmatrix} \Sigma_{11}(Z_i) & \Sigma_{12}(Z_i) \\ 0 & 0 \end{pmatrix} - f_z(z) \cdot vec\begin{pmatrix} \Sigma_{11}(z) & \Sigma_{12}(z) \\ 0 & 0 \end{pmatrix},$$

$$A_{i,i} = K_h(Z_i - z) \cdot vec\begin{pmatrix} \Sigma_{11}(Z_i) & \Sigma_{12}(Z_i) \\ 0 & 0 \end{pmatrix} - f_z(z) \cdot vec\begin{pmatrix} \Sigma_{11}(z) & \Sigma_{12}(z) \\ 0 & 0 \end{pmatrix}.$$

By Lemma 15 and Lipschitz assumption on K_h, $\|A_{i,j(i)} - A_{i,i}\| = O_p(N^{-1})$. Also note that by simple calculation, we have

$$\mathbb{E}\left(K_h(Z_i - z)\begin{pmatrix} \Sigma_{11}(Z_i) & \Sigma_{12}(Z_i) \\ 0 & 0 \end{pmatrix}\right) = f_z(z)\begin{pmatrix} \Sigma_{11}(z) & \Sigma_{12}(z) \\ 0 & 0 \end{pmatrix} + O(h^2),$$

which implies that $\mathbb{E}(A_{i,i}) = O(h^2)$. Therefore,

$$\begin{aligned}
&Var\left(\mathbb{E}(VI_{10} \mid Z^{1,2})\right)\\
&= \mathbb{E}\left((\frac{1}{n}\sum_{i=1}^{n} A_{i,j(i)})(\frac{1}{n}\sum_{i=1}^{n} A_{i,j(i)})^T \right)\\
&= \frac{1}{\sqrt{n}}\left(\frac{m}{n}\right)^{\frac{3}{2}} \mathbb{E}\left(\frac{1}{\sqrt{m}}(\max_{j=1,\ldots,m} C(j))^2 A_{(i,i)} A_{(i,i)}^T\right) + O(h^4) + o(1)^2
\end{aligned}$$

Note that under the Lipschitz assumption on K_h and Σ, $Var\left(\mathbb{E}(VI_{10} \mid Z^{1,2})\right)$ goes to zero matrix as sample sizes go to infinity. Therefore we have demonstrated that I_{10} converges in probability to $f(z)\begin{pmatrix} \Sigma_{11}(z) & \Sigma_{12}(z) \\ 0 & 0 \end{pmatrix}$.

C.4 Boundedness of $V(z)$

The following shows that the expectation of $V(z)$ is finite. The result follows **Lemma** 3 in Abadie and Imbens (2006).

$$V(z) = \sum_{j=1}^{m} \sigma^2 (C(j) K_h(Z_j - z))^2 \Omega(Z_j) \bigotimes \begin{pmatrix} 1 & (Z_j - z)/h \\ (Z_j - z)/h & ((Z_j - z)/h)^2 \end{pmatrix}.$$

From **Lemma** 3 in Abadie and Imbens (2006), $C(j)$ is $O(1)$, $j \in \{1, ..., m\}$. By **Assumptions** 15−16, the kernel functions we consider are Lipschitz continuous on a compact set, and $\nu_j = \int u^j K^2(u) du, j = 1, 2, 3$ are also bounded, therefore $V(z)$ is also bounded.