Contents

Chapter 1

Attention Allocation Between Information Scope and Depth

1.1 Introduction

When trying to determine which stocks to invest in, an investor can focus on a few stocks or can examine all available stocks. The former case provides the investor with more information on individual stocks, while the latter provides less information on a wider range of stocks. Here, investors face a trade-off between the scope and depth of information and have to make an allocation decision based on a scarce resource, namely their attention.

In this study, I investigate the Arrow-Debreu security model in an environment where investors' information is incomplete in terms of both scope and depth. Here, I focus on the racetrack betting market data where in-

vestors in the model cannot observe the true fundamentals directly and do not consider all feasible investment oppotunities in the market. Instead, investors are assumed to have heterogeneous beliefs and their portfolio choices are confined to a subset of all feasible investment oppotunities, called a consideration set.

Existing studies that use racetrack betting market data mainly investigate how investors form their beliefs based on information depth. Most of these studies examine an anomalous pricing pattern called the favorate long shot bias (FLB), which is occurs when long shots are overpriced. For example, Jullien and Salanié (2000) find empirical evidence that the FLB is the result of bettors' misperceptions about a winning probability, and Snowberg and Wolfers (2010) find evidence supporting a cumulative prospect theory-based explanation. Recently, Gandhi and Serrano-Padial (2014) proposed a heterogeneous belief model. Theoretically, thier model proves that even if investors' have the correct mean belief, the existence of belief heterogeneity is sufficient to generate an FLB. Empirically, their model shows that belief heterogeneity model better explains the data. Furthermore, they find that the degree of belief heterogeneity is responsive to the amount of information.

However, observations based on the Korean racetrack betting market are not well explained by information depth alone. (e.g., anomalous odd gap ratio). The quinella and the exacta are the two most popular types of bets in Korea. In a quinella, a bettor has to pick the two fastest horses, in any order. The exacta is similar, except that the bettor must specify the order as well. By construction, a quinella bet is easily replicated by a linear combination of two exacta bets and, theoretically, the odds of the original quinella bet and

its replicated counterpart should be equal, on average. However, as shown in figure 1.1, 62.2% of original quinella bets are overpriced compared with thier replicated counterparts. An explanation based on information depth, such as the belief heterogeneity model, cannot explain this anomaly because the fundamentals of the original and replicated bets are identical. However, if we relax the assumption that investors have complete information in terms of scope, then the existence of the anomaly is explained. Investors can fail to take into account such strictly preferable opportunities because the investors with a limited information scope overlook some assets in the market.

In this study, I construct a competitive A-D security market model, where investors are bounded in terms of both information depth and scope. Then, I use this model to investigate the equilibrium security price in conjunction with the investors' degree of rationality. In particular, I explore how the two dimensions of information contribute to the investors' returns and to market efficiency. Then, I check the empirical validity of the limited information scope explanation using Korean racetrack betting market data. There are two main findings.

First, at the investor level, the theoretical results indicate that both the depth and the scope of information are valuable to investors. By increasing either the depth or the scope, investors' expected return increase. Thus, if they are not limited in terms of attention, investors will behave in a fully manner because they have enough of an incentive to do so. At the market level, market prices fail to fully reflect the fundamentals of securities if either the depth or the scope is incomplete. Thus, both are necessary to achieve market efficiency.

Second, the results of the empirical analysis identify two types of bettors: generalists and specialists. Generalists pay more attention to the scope of the information, making their prediction naive, while specialists focus on a small number of bets that yield better prediction of the fundamentals. The ratios of generalists to specialists in the quinella and exacta markets turn out to be different; the share of specialists in the quinella market is 87.2% and is 95.9% in the exacta market. This explains the Korean racetrack betting market anomalies.

Then, why do bettors choose different attention allocation strategies? From the viewpoint of standard economics, an investor's attention allocation choice is typical resource allocation problem; that is, bounded rational behavior can be the result of rational choice. Using a classical economic analysis tool, such as an indifference curve, explains the attention strategy choice in terms of the heterogeneity of the information cost structure. This is intuitively appealing. For example, consider a mutual fund company whose fund managers are industry veterans (e.g., the automobile industry). In such a case, it might be a better for them to limit their portfolio to automobile industry-related stocks. On the other hand, for a mutual fund company whose fund managers excel at momentum strategies or picking the market timing, expanding the investment scope can improve the portfolio performance.

The idea that information scope matters to investors' decisions is not new in finance. Merton (1987) models a capital market equilibrium where investors are informed only about a subset of securities in the market. He finds that a bounded information scope can reconcile several empirical anomalies. Odean (1999) finds that investors tend to overly invest in

attention-grabbing stocks, and Barber and Odean (2007) find evidence supporting an attention explanation in the stock market. To discern the attention effect and the effect of fundamentals, Da et al. (2011) use the Google search index as a proxy for attention and confirm the attention hypothesis.

Another strand of the literatures related to this work is attention allocation. Similarly to the model examined here, Sims (2006) proposes a model where the economic agent has a bounded information-processing capacity, which means they cannot respond fully to the given information. In his model, the best that economic agents can do is to allocate their scarce attention in a rational manner. Van Nieuwerburgh and Veldkamp (2010) examine the case where investors decide on an investment and an information acqusition at the same time. They show that seemingly anomalous investment portfolio holdings can be rationalized using the cost structure of the information acqusition. According to this point of view, the coexistence of the generalist and the specialist in the betting market can be rationalized if they are endowed with different attention cost functions. Kacperczyk et al. (2016) support the idea that attention allocation can be viewed as a rational scarce resource allocation problem. Kacperczyk et al. (2016) provide a the formal model of ratioinal attention allocation in conjunction with the business cycle, and show that some investment managers do allocate their attention in a rational way.

This paper is organized as follows. In section 2, the bounded rational model for the A-D security market is introduced. In section 3, I apply the model to analyze the effects of incomplete information depth and scope at both the investor level and the market level. In section 4, I introduce the data set and Korean racetrack betting market anomalies. In section 5, the

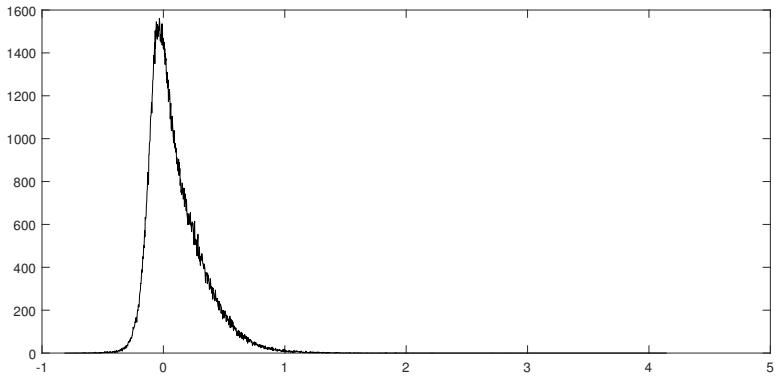

Figure 1.1: Histogram of odds gaps between a quinella bet and a replicated bet

Note: The odds gap ratio is defined by $\frac{\hat{o}_j - o_j}{o_j}$, where o_j and $\hat{o}_j$ are the odds of an original quinella bet j and its replicated counterpart, respectively.

estimation strategy and the results are presented. In section 6, I show how a standard economic analysis can explain the bounded rational investor's attention allocation choice using empirical results. Section 7 concludes the paper.

1.2 Model

1.2.1 The Market

In this section, I develop a bounded rational model for an Arrow-Debreu security market. Instead of assuming fully rational investors, I assume they are bounded rational in two dimensions: i) rather than considerting all feasible investment opportunities, they consider a random subset; and ii)

their beliefs about the fundamentals of securities deviate randomly from the true fundamentals. The market is composed of a set of A-D securities B, bounded rational investors I, and the competitive market demand $\{s_j\}_{j\in B}$. Each $j \in B$ is characterized by a pair (π_j, E_j), where π_j is the A-D security price and E_j is the underlying event of an A-D security. A racetrack bet is a typical example of an A-D securty where the price is equal to the reciprocal of its odds $\frac{1}{o_j}$ and the underlying event is the result of a race. I use $\frac{1}{o_j}$ rather than π_j. Henceforth, I use j and E_j interchangeably if there is no risk of confusion.

Each investor $i \in I$ has subjective belief q_j^i about the winning probability of bet j. Assume risk-neutral investors. Thus, if an investor i considers all A-D securities in the market, then she will invest all of her endowment w^i on security j, such that $q_j^i o_j > q_l^i o_l$ for all $l \in B$. In the model, I assume each investor's investment scope is restricted to a randomly generated consideration set $B^j \subset B$. Define $J^i = \text{argmax}\{q_l^i o_l | l \in B^i\}$. This is a random variable because the subjective probabilies q_j^i and the consideration set B^i are random. Denote investor i's demand for security j by s_j^i. If $\{q_j^i\}_{j\in B}$ have no concentrated measure, the following relation holds:

$$s_j^i = \begin{cases} w^i & \text{if} \quad j = J^i \\ 0 & \text{o.w.} \end{cases}$$

Assume that $\frac{w^i}{\sum_{i\in I} w^i}$ are infinitesimal and $\{q_j^i\}_{j\in B}$ and that B^i are mutually independent. Express the aggregate demand share on bet j as $s_j = \frac{\sum_{i\in I} s_i^j}{\sum_{i\in I} w^i}$. Then, the following equality holds by the law of large numbers, if $|I| = \infty$:

$$s_j = \Pr[i = \text{argmax}\{q_l^i o_l | l \in B^i\}]. \tag{1.1}$$

For the pricing equation that clears A-D security market, I assume a Parimutuel mechanism. This is a winner-takes-all rule that guarantees the fixed portion of sales as the revenue for the market organizer. Under this rule, highly demanded securities are highly priced. Under the Parimutuel mechanism, A-D security prices are set to proportional the sales share as follows:

$$o_j = \rho \frac{1}{s_j} \qquad \forall j \in B. \tag{1.2}$$

Here, $\rho \in (0, 1)$ is the discount factor that reflects the tax and revenue of the racetrack company. In equilibrium, the asset price $\frac{1}{o_i}$ is determined where equations (1.1) and (1.2) are satisfied simultaneously.

1.2.2 Information Depth

Investors seek and process relevant information in a proper way in order to make a good investment decision. In the model, I assume the investor's belief $\{q_j^i\}_{j \in B}$ approaches to the true probability $\{p_j\}_{j \in B}$ as he or she obtains in-depth information. The information depth reflects the amount of attention an investor pays to elicit a belief. For example, consider the following hypothetical case. Suppose N races are conducted independently under the same conditions. Let n_j be the number of the race that bet j wins and q_j be the subjective probability of bet j wins. Then, the conditional distribution $p\left(\{n_j\}_{j \in B} | \{q_j\}_{j \in B}\right)$ is

$$p\left(\{n_j\}_{j \in B} | \{q_j\}_{j \in B}\right) = N! \prod_{j=1}^{k} \frac{q_i^{n_j}}{n_j!}.$$

Here we can interprete $N = \sum_{j \in B} n_j$ as a number of hypothetical races an investor conducts to elicit his or her beliefs. If investors have a flat prior

on $\{q_j\}_{j\in B}$, a direct application of Bayes' theorem results in right-hand side of the above equation being proportional to the conditional distribution $p\left(\{q_j\}_{j\in B}|\{n_j\}_{j\in B}\right)$. Observe that $\lim_{N\to\infty} E[q_j^i] = \frac{n_j}{N} = p_j$ and $\lim_{N\to\infty} var[q_j^i] = 0$, where $|B|$ is finite.

Another classical example is to think of information as a noisy signal. Let $D_n \sim (p_j, \sigma)$ be the noisy signal an investor receives. Assume that investors have to pay attention to obtain an additional signal. If an investor i receives N random signals, then her subjective belief q_j^i is the conditional expectation $E[p_j^i|\{D_n\}_{n=1}^N]$. Observe that as N goes to infinity, her subjective belief converges to the true probability.

In the aforementioned two examples, the number of data observations N can be interpreted as the amount of information consumed by investors. In the extreme case where $N \to \infty$, the subjective belief distribution collapes to a concentrated point measure. Therefore, there must be a single consensus among investors about the fundamentals if they process all relevant information in a reasonable way and are sure of this information. It seems plausible to assume that such a consensus reveals the true fundamentals. Thus, I assume the following, which I refer to as *Bayesian Rationality*. Note that this is a rephrasing of the *Rational Expectation Hyphothesis*.

Assumption 1.1. *(Bayesian Rationality) The subjective belief q_j^i follows the distribution $F(\cdot|p_j, \sigma^i)$, such that*

$$E[q_j^i] = p_j$$

and its standard deviation increases in σ^i.

1.2.3 Information Scope

For the choice probability of consideration set, I impose the following assumption.

Assumption 1.2. *(Independent Random Attention)* $\exists!\{\gamma_j^i\}_{j\in B}$ *s.t.* $\forall E \in 2^B$*, then the following is true:*

$$\Pr[B^i = E] = \prod_{j\in E} \gamma_j^i \prod_{l\in B\setminus E} (1-\gamma_l^i).$$

This assumes that the event that investor i considers as an asset(bet) j is an independent binary event for all $j \in B$. Each asset j is randomly and independently considered by investor i with probability γ_j^i. This is a parsimonious starting point that keeps the model mathematically tractable.

Given the two assumptions on the subjective belief formation and the consideration set generation, the investor's choice probability $\Pr[j = J^i]$ is

$$\Pr[j = J^i] = \int_{-\infty}^{\infty} \prod_{l\in B\setminus\{j\}} \left(1-\gamma_l^i + \gamma_l^i F(s/o_l|p_l,\sigma^i)\right) f(s/o_j|p_j,\sigma^i)ds, \tag{1.3}$$

where $f(\cdot|p_j\sigma^i)$, and $F(\cdot|p_j,\sigma^i)$ are the PDF and CDF of q_j^i, respectively. For the derivation, see the Appendix.

1.3 Bounded Rationality and Asset Market

Under the standard assumption, investors are fully rational, which means they can thoroughly review all feasible investment oppotunities and know the true probability p_j. Thus, the asset price $\frac{1}{o_j}$ is proportional to p_j and all investors have the same expected return in equilibrium, as long as they are risk neutral. This is a desirable consequence of a well-functioning financial

market because prices correctly reflect the fundamentals. However, it is no longer true if investors are bounded rational. The following analysis predicts that investors who are more rational are eligible to earn excess returns from an investment, and that the equilibrium asset prices are distorted in an environment where investors have incomplete information scope and depth.

1.3.1 Value of Information

The value of in-depth information seems clear; as an investor's subjective beliefs approach the true fundamental, the probability of making a correct choice increase. However, the value of information scope is more difficult to discern. As the investment scope increases, the probability of making an incorrect decision increase when the investor's information depth is finite. The following analysis shows that the information scope has value at least in the ex ante sense.

If investor i is considering $E \in 2^B$, the ex ante rate of return can be written as

$$R^i(E) = \max\{q_l^i o_l | l \in E\}.$$

For example, in the standard full attention model, investors are assumed to be capable of considering all feasible assets in the market. Thus, a fully attentive investor i is supposed to earn an ex ante rate of return $R^i(B)$. On the other hand, a bounded rational investor i is supposed to earn $R^i(B^i)$. For i and $i' \in I$, i has a wider information scope than i' if $\gamma_j^i > \gamma_j^{i'}$, for all $j \in B$. Assume they have same information depth. Now, the question is whether $R^i(B^i)$ stochastically dominates $R^{i'}(B^{i'})$. The next result shows that the ex ante returns of more rational (attentive) investors stochastically

dominate those of less rational (attentive) investors. For technical convenience, assume there exists one risk-free asset $d \in B$ that pays $\underline{R}$ with probability one and that $\gamma_d^i = 1$ for all i.

Theorem 1.1. *(Excess Return of Being Rational)*
Assume $\{q_j^i\}_{j \in B}$ *and* $\{q_j^{i'}\}_{j \in B}$ *are generated by the same distribution. If* $\gamma_j^i \geq \gamma_j^{i'}$ *for all* $j \in B$*, then*

$$\Pr[R^i(B^i) < r] \leq \Pr[R^{i'}(B^{i'}) < r].$$

Proof. See the Appendix. □

Theorem 1.1 implies that the scope of information is valuable, at least in the ex ante sense. Thus, investors always have an incentive to extend the scope of their information; that is if attention is not a scarce resource, investors will consider all assets in the market.

1.3.2 Is the Price Right?

Here, I address the question of whether a security's price correctly reveals its fundamental value in an environment where investors have incomplete information. The answer is no. To see why, we require the following definition.

Definition 1.1. *(Distorted Security Price) The set of security prices* $\{\frac{1}{o_j}\}_{j \in B}$ *is "distorted" if* $p_j o_j \neq p_k o_k$ *for some* $i, k \in B$ *and* $s_j = \frac{1}{o_j}$ *for all* $j \in B$*.*

To focus on the impact of incomplete information depth, temporarily assume that all investors are fully attentive. For example, consider the following simplest possible case. Let two horses, say 1 and 2, run in a race. If

all $i \in I$ use the same belief formation process, then demand for security 1 satisfies:

$$s_1 = \Pr[q_1^i o_1 > q_2^i o_2].$$

Under the Parimutuel mechanism, $s_j = \frac{1}{o_j}$ and $\sum_{j \in B} s_j = 1$. Thus, the equilibrium security price s_1 must satisfy the following:

$$s_1 = 1 - F(s_1),$$

where $F(\cdot)$ is the cumulative distribution function of q_1^i. The following proposition shows that the equilibrium market price is a distorted security price, even when the *Bayesian rationality assumption* holds.

Proposition 1.1. *Assume that $Eq_1 = p_1$ and $F(p_1) = 1 - a$. If $p_1 < a$, then $s_i > p_1$.*

Proof. Suppose the proposition does not hold. Then, $p_1 \geq s_i$, and we have the following:

$$s_1 = 1 - F(s_1) \geq 1 - F(p_1) = \alpha > p_1.$$

The first equality is the equilibrium condition and the second inequality is derived from the hypothesis $p_1 \geq s_1$. However, this is a contradiction because it results $s_1 > p_1$. □

The above proposition explains the FLB, which is one of the most well-known stylized facts observed in betting markets. The reason why this distortion arises is because the percentile of belief matters in equilibrium rather than the mean investor's belief. Thus, equilibrium security prices can be distorted as long as the investors have heterogeneous beliefs. Gandhi and Serrano-Padial (2014) provide a rigorous proof when the number of bets

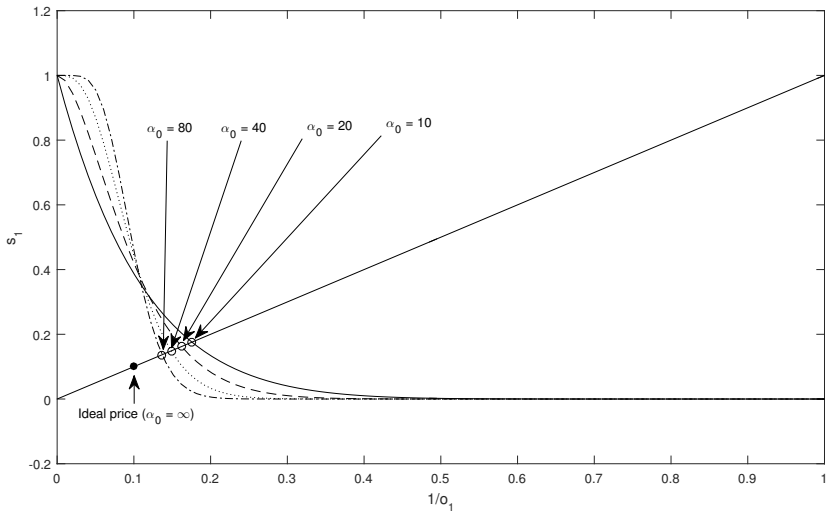

Figure 1.2: Competitive market equilibrium

Note: Assume two A-D securities, say 1 and 2, are available in the market, where the market is complete. Let the probability that security 1 wins be $p_1 = 0.1$. The horizontal axis represents the reciprocal of the odds and the vertical axis represents the demand share. The 45 degree line is the supply curve of the security under the parimutuel mechanism and the downward sloping curves are the demand for security 1 for varying degrees of information depth.

is an arbitrary finite number. Under the *Bayesian rationality assumption*, such distortion tends to be alleviated as the investors' degree of rationality α_0^i increases. For example, let $p_1 = \frac{1}{10}$. Figure 1.3.2 shows the market equilibrium security prices for different values of α_0: 10, 20, 40 and 80[1]. As α_0 increases from 10 to 80, the security price approaches the efficient market price 0.1.

[1]Here, the subjective beliefs are assumed to be distriuted by the Dirichlet distribution and α_0 is the sum of the concentration parametors.

On the other hand, the following theorem implies that even if all investors have complete information depth, security prices can be distorted in equilibrium if they have incomplete information scope. Let $I = \{1, 2, \ldots, |B|\}$ be the index of $j \in B$. To clearly discern the impact of an incomplete information scope on the security price, assume that investors have complete information depth; that is, $q_j^i = p_j$ for all $i \in I$ and $j \in B$. Even in this case, the equilibrium price can be misleading. The following theorem suggests that the equilibrium can be distorted and, furthermore, the set of bets B can be strictly totally ordered by $\succeq$. For notational simplicity, the set of bets B is identified as an index set $I \subset \mathbb{N}$ by an index function. The proofs of all the following theorems all show that when the index function that maps B to I preserves the preference ordering structure of B so that the natural order $>$ of I can represent the preference relation of B, I can be strictly ordered by the natural order $>$ in equilibrium.

Theorem 1.2. *(Price Distortion under Random Attention) Suppose the following two conditions are satisfied.*

- *The index of set I can be rearranged so that $p_i > p_l$ for $i > l$ for all $i, l \in I$.*
- *Let the index be given by $i = 1, 2, \ldots, |I|$ and*

$$\frac{p_{i+1}}{p_i} > \frac{\gamma_{i+1}}{\gamma_i(1-\gamma_i)}.$$

Then, there is $\{s_i\}_{i \in I}$ in equilibrium, such that

$$\frac{p_i}{s_i} > \frac{p_l}{s_l}.$$

whenever $i > l$, which implies the FLB.

Proof. See the Appendix. □

The next question is whether we can correct this distortion using a simple and implementable mechanism. Here, I present an analysis of the role of the default option or the reference security. In the real world, securities such as the U.S. T-bill might be considered by all investors as a feasible investment opportunity and serve as a reference point. If investors have a common reference security with which to compare the profitablity of investment opportunities, the security prices might be adjusted to the efficient level. If this works properly, it is a cheap solution for market efficiency. Unfortunately, the following result shows that the existence of a default option or reference security might alleviate such distortion, but it is not enough to correct the price distortion completely.

Theorem 1.3. *(Role of Reference Securitty) Suppose $k \in B$ is a reference security; i.e.,* $\Pr[k \in B^i] = 1$ *for all $i \in I$. This does not imply $p_j o_j = p_{j'} o_{j'}$ for all $j, j' \in B$. Rather, there could be a subset $B_{>k} \subset B$ such that $p_j o_j \neq p_{j'} o_{j'}$ for all $j, j' \in B_{>k}$.*

Proof. See the Appendix. □

1.4 Data

Two distinctive characteristics of the Korean racetrack betting market are high liquidity and the popularity of exotic bets. On average, the KRW 4.7 billion bets are placed per race and the average share of quinella and exacta bets are 67.9% and 19.9%, respectively. Thus, the odds gap data between the quinella and exacta markets are a valuable source of information

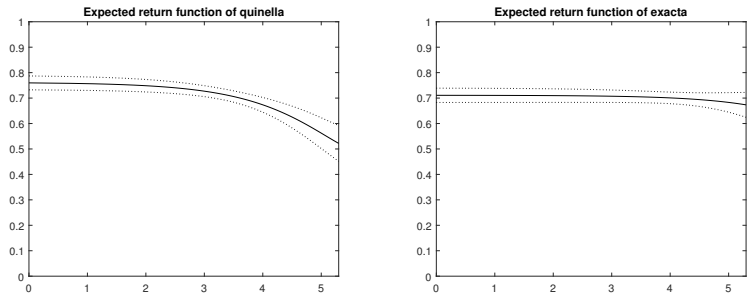

Figure 1.3: Empirical return functions of quinella and exacta betsbet.

Note: The horizontal axis is the log of the odds and the vertical axis is the empirical return. The estimations use the LOESS method. For the detailed procedure, see the Appendix.

for studying the arbitrage gap. A brief analysis shows that quinella bets and exacta bets are close substitues[2], but that their pricing patterns are remarkably different.

The empirical return function $r(o)$, the estimate of the return of a bet conditional on its odds, reveals the relationship between the security price and its fundamentals.[3] From the estimation results shown in of figure 1.3, we can see that empirical return function is downward sloping in the quinella market, which implies FLB. On the other hand, the FLB does not seem to be obvious in the exacta market. One possible explanation is to assume

[2]See the Appendix

[3]For $j \in B$, define a pair of random variables (O_j, W_j) such that W_j is 1 if bet j wins, and zero otherwise, and O_j is the odds of bet j. Then, R_j represents the return of bet j. The empirical return function $r(o)$ is the conditional expection of this return. Specifically,

$$r(o) = E[R_j | O_j = o].$$

that quinella bettors and exacta bettors have different preferences, beliefs, or degree of belief heterogeneity and remain in one market. However, this is not a satisfactory explanation because there is no institutional constraint that prevents bettors from investing in various types of bets. Thus, the fact that the odds of similarly probable exacta bets are higher than those of quinella bets for the long shots, and that this gap is sustained, is not explained. Bettors have enough incentive to alter their betting from one market to another. Statistically, quinella bets and exacta bets seem to be close substitutes. Thus, the bounded rationality hypothesis (bettors are not fully attentive) seems more plausible.

Another observation suggests the presence of heterogeneous degrees of rationality among bettors. The exacta bet was first introduced in May 2000 in Korea and was sold only for randomly selected races untill October 2003; the KRA sold exacta bets for 617 randomly selected races from a total of 2,967 races. KRA offered exacta bets for all races after this trial period. Figure 1.4 shows the two empirical return functions of quinella bets when exacta bets are sold and not sold, repectively. It shows that the pricing pattern of quinella bets is more distorted when exacta bets are available. In addition, statistics suggest that a subset of quinella bettors move to exacta bets when they are available.[4] this implies that bettors who move to exacta bets tend to bet less on long shot. That is, they are seemingly more rational.

The bounded rational investor model seems to better explain the anomalous pricing patterns in the Korean racetrack betting market. In the follow-

[4]For the trial period, I set up a simple linear regression in which the regressant and the regressor are the sales share of exacta and quinella bets, respectively. The result shows their coefficient is -0.97 and R^2-value is 0.94.

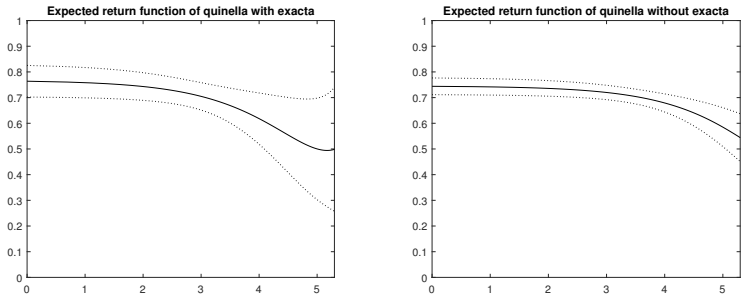

Figure 1.4: Empirical return function of quinella bet, with and without an exacta bet

Note: The horizontal axis is the log of the odds and the vertical axis is the empirical return. The left figure is estimated using data of 617 races from the trial period when exacta bets were sold. The right figure is estimated from 2,350 races with exacta bets.

ing section, I describe the empirical strategy used to estimate the bounded rational investor model using aggregate level data. The data set used in the empirical analysis contains 1,232 races and 203,271 bets.

1.5 Estimation

The parameters of interest in this study are the set of consideration parameters γ_j^i and the belief formation rationality parameters σ^i. To alleviate the computational complexity, I assume i) there are finite types of investors, and ii) investors are indifferent to considering same type of bet; that is, $I = \uplus_{k=1}^{K} I_k$ and if $i, i' \in I^k$, then $\gamma_j^i = \gamma_j^{i'}$, and if two bets j and j' are of the same type, $\sigma^i = \sigma^{i'}$ and $\gamma_j^i = \gamma_{j'}^i$. For the subjective belief, assume $q_j^i = p_j e^{\epsilon_j^i}$, where ϵ_j^i is the Gumbel noise with scale parameter σ^i and p_j denotes the true probability. Thus, the choice probability can be written as

Parameter	One type investor model (std. error)	Two type investor model (std. error)	Three type investor model (std. error)
σ^1	0.1415	1.8428	1.1569
	(0.0001)	(0.0334)	(> 100)
σ^2	-	0.0618	0.1159
	-	(0.0007)	(> 100)
σ^3	-	-	0.0083
	-	-	(> 100)
γ^1_{qu}	0.2958	1	1
	(0.0021)	(0.5218)	(> 100)
γ^1_{ex}	0.0153	0.0246	0.0442
	(0.0001)	(0.0113)	(> 100)
γ^2_{qu}	-	0.3076	0.5114
	-	(0.0114)	(> 100)
γ^2_{ex}	-	0.0622	0.0153
	-	(0.0051)	(> 100)
γ^3_{qu}	-	-	0
	-	-	(0.0022)
γ^3_{ex}	-	-	1.0000
	-	-	(> 100)
π^1	1	0.1160	0.0696
	-	(0.0038)	(> 100)
π^2	-	0.8840	0.7435
	-	-	(> 100)
π^3	-	-	0.1869
	-	-	-
sq error	1.5297	1.1964	0.6921

Table 1.1: Estimation result

follows:

$$\Pr[j = J^i] = \sum_{k=1}^{K} \Pr[j = J^i | i \in I^k] \Pr[i \in I^k].$$

If all of the true probabilities $\{p_j\}_{j \in B}$ are known, the conditional choice probability $\Pr[j = J^i | i \in I^k]$ can be parameterized by $\theta^k = (\gamma^k_{qu}, \gamma^k_{ex}, \sigma^k)$ using equation 1.3. Thus, the unconditional choice probabiltiy $\Pr[j = J^i]$ is parameterized by $\theta = (\theta^1, \ldots, \theta^K, \pi^1 \ldots, \pi^K)$. Denote the unconditional choice probability by $\hat{s}_j(\theta | \{p_j\}_{j \in B})$. Theoretically, if the number of investors

increases, the observed market share s_j should converge to $\hat{s}_j(\theta_0|\{p_j\}_{j\in B})$, where θ_0 is the true parameter. I assume the following data-generating process:

$$\log s_j = \log \hat{s}_j(\theta_0|\{p_j\}_{j\in B}) + \varepsilon_j,$$

where ε_j is an i.i.d. random noise with a finite second moment. The estimator $\hat{\theta}$ is the minimizer of the squared log difference between s_j and $\hat{s}_j(\theta|\{p_j\}_{j\in B})$.

One remaining issue is the unobservability of p_j. To tackle this issue, I use its estimate as a proxy. To obtain a reasonable proxy, I first estimate the empirical winning probability $\hat{p}(o) = \Pr[W_j = 1|O_j = o]$ of the exacta bets using the same method used for the estimation of the empirical return function. Denote the set of bets from the kth race by B^k and $B^k_{qu} = B^k \cap B_{qu}$ and $B^k_{ex} = B^k \cap B_{ex}$, where B_{ex} and B_{qu} are sets of exacta bets and quinella bets, respectively. For each $j \in B^k_{ex}$, I define the proxy $\hat{p}_j$ of the winning probability conditional on k by

$$\hat{p}_j = \frac{\hat{p}(o_j)}{\sum_{l\in B^k_{ex}} \hat{p}(o_l)}.$$

Recall that there are two bets $j1$ and $j2$ in B^k_{ex} for a $j \in B^k_{qu}$, such that $j1 \uplus j2 = j$. By the probability axiom, $\hat{p}_j = \hat{p}_{j1} + \hat{p}_{j2}$ holds for all quinella bets.

The estimation procedure it consists of two steps. In the first step, estimate $\{\hat{p}_j\}_{j\in B}$ using the above mentioned procedure. In the second step, substitute $\{\hat{p}_j\}_{j\in B}$ into $\hat{s}_j(\theta|\cdot)$ and find $\hat{\theta}$ that minimizes the squared error. i.e.:

$$\hat{\theta} = \underset{\theta\in\Theta}{\operatorname{argmin}} \frac{1}{|B|} \sum_{j\in B} \left(\log \hat{s}_j(\theta|\{\hat{p}_j\}_{j\in B} - \log s_j)\right)^2.$$

I use sequential quadratic programming for the numerical minimization.

To complete the model, the last step is to fix the number of investor types. Table 1.1 shows the models fit the data better as K increases as usual. I test the following two hypothesis using Wald statistics:

$$\begin{array}{llll} \text{Test 1} & H_0 : K = 2 & \text{against} & H_A : K = 1 \\ \text{Test 2} & H_0 : K = 3 & \text{against} & H_A : K = 2 \end{array}$$

As shown in table, the test results favor the hypothesis $K = 2$.

Assuming the two-type specification is true, investors attention allocation behavior can be categorized into two types: generalist and specialist. Type 1 bettors, the generalists, fully attend to the quinella bet, but the estimate of σ^1 shows they do not assess the fundamentals in depth. In contrast, the second type of bettors, the specialists, seem to consider less number of bets on average, but the estimate of σ^2 indicates that they have good predictions of the fundamentals. Interestingly, specialists pay more attention to exacta bets than the generalists do. This explains a large number of the empirical anomalies observed in the Korean racetrack betting market. The simulation shows that 95.9% of exacta bettors are specialists. Thus, exacta betting market participants have more rational beliefs on average, than quinella bettors do. Thus, the pricing pattern of the exacta market is less distorted. This also provides a convincing explanation for the anomaly in the trial period. In the absence of exacta betting opportunities, specialists might be betting on quinella bets. As exacta bets became available, specialists altered their choice to the exacta market. As a result, the average level of belief in the quinella market became fuzzier as new betting opportunities attracted more rational bettors.

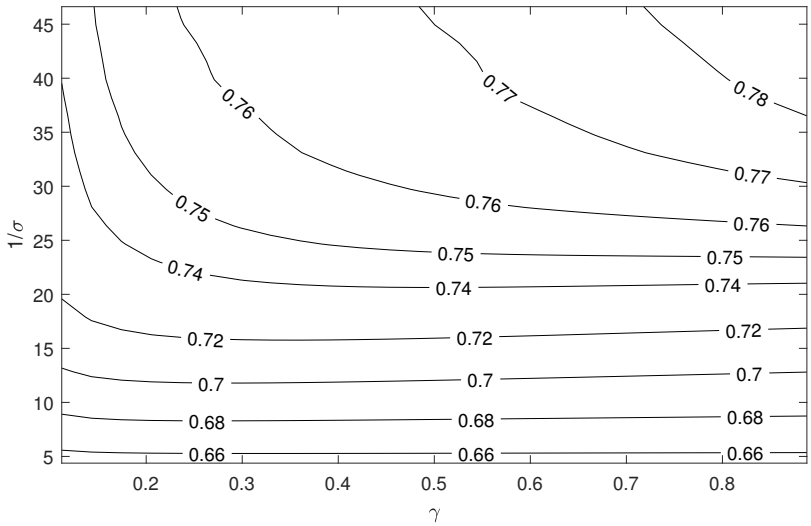

Figure 1.5: Indifference curve of depth and scope

1.6 Discussion: Rational Attention Allocation

The results described in the previous section indicate that there are two types of bettors in the market: generalists and specialists. The generalists seek better investment oppotunities by extending the scope of information, while the specialists delve into fewer investment oppotunities to deepen their understanding of the fundamentals. A natural question to ask is what made them behave in such extremely different ways? In this section, I investigate this questtion from a classical economic perspective to provide a partial answer.

If we accept the view that information consumes attention, then attention is a scarce resource. Thus, the difference in endowed resources could explain why investors choose different attention strategies. Under the stan-

dard economic framework, the fact that an agent chooses an inferior allocation in the presence of a strictly preferable allocation reveals that a strictly preferable allocation does not fealsible. To identify the investors' preference over the attention allocation strategy observed in the betting market, denote a set of generalists by I_G and that of specialists by I_S. The ex post return conditional on attention strategy, that is, $E[R_{J^i}|i \in I_t]$, where $t \in \{G, S\}$, determines the preference over the attention strategy because all investors are assumed to be risk neutral. The estimation indicates $E[R_{J^i}|i \in I_G]$ and $E[R_{J^i}|i \in I_S]$ are 59.6% and 74.4%, respectively. This implies that the specialist strategy is strictly preferable for all investors. Thus, we can conclude specialists are endowed with a greater attention capacity than that of generalists.

A scarce resource allocation problem is explained in terms of an optimization in the presence of a trade-off relationship. Thus, to understand attention allocation in a classical framework, we need to analyze how the scope and the depth of information interact and contribute to the investors' utility. To serve this purpose, I assume that an imaginary investor i has an attention strategy $(\gamma, \frac{1}{\sigma}) \in [0, 1] \times (0, \infty)$; investor i considers both quinella bets and exacta bets with the same probability γ, and the depth parameter is equal to σ. Figure 1.5 plot the isoquant line of ex post returns conditional on i's attention strategy. Each isoquant line corresponds to the indifference curve of risk neutral investor i. If i is sufficiently rational, she can pick a near optimal attention allocation by trial and error because the ex post returns are observable. This shows that both the scope and the depth contribute to the return and that they complement each other.

The generalist strategy corresponds to the lower right area. In this area,

the marginal rate of substitution between the depth and scope shows that depth is more valuable to the generalist. In contrast, the specialist strategy corresponds to the upper left area, where scope provides more valuable information. This suggests two possibilities. First, the relative cost of information is highly heterogeneous by investors. If an investor is good at delving into detailed information, but tends to be distracted by an overwhelming number of tasks, being a specialist is a better option.

The second possiblity is that the information cost function is a concave function. The intuition is simple. If an investor has in-depth information, it might be more difficult to extend the scope of the investment while maintaining the quality of a prediction. By the same token, it is difficult to improve the prediction precision if an investor has a larger investment scope.

The observation that the structure of the information cost is the critical determinant of an investor's attention strategy is meaningful. It implies that the advent of new technologies that influence the structure of the information cost, such as a recommendation algorithm, can result in make a dramatic change in investors' optimal attention strategies. Thus, information technology can be an important determinant of the degree of market efficiency.

1.7 Conclusion

I have proposed a model of bounded rational investors who are bounded in information depth and scope. This model explains anomalies in the Korean racetrack betting market. The empirical analysis identifies that general-

ists and specialists coexist in the market. Interestingly, bounded rational investors' attention allocation choices can be explained within a standard rational framework. A remaining issue is to identify the information cost structure. This is left to future research.

Appendix

1.A Omitted Proofs

1.A.1 Derivation of Equation (1.3)

Proof. From definition,

$$\Pr[j = J^i] = \Pr[j = \operatorname{argmax}\{q_l^i o_l | l \in B^i\}].$$

Using theorem 2.1, substitute $F(u|\alpha_l^i)$ and $DF(u|\alpha_j^i)$ by $F(u/o_l|p_l, \sigma^i)$ and $f(u/o_j|p_j, \sigma^i)du$, respectively. This yields,

$$\begin{aligned}\operatorname{argmax}\{q_l^i o_l | l \in B^i\}] \quad &= \\ &\gamma_j^i \int \prod_{l \in B_{-j}} \left(1 - \gamma_l^i + \gamma_l^i F(u/o_l|p_l, \sigma^i)\right) f(u/o_j|p_j, \sigma^i) du,\end{aligned}$$

where B_{-i} denote the $B \backslash \{i\}$ for a set B and $i \in B$. □

1.A.2 Proof of Theorem 1.1

Proof. It is sufficient to show that $\frac{\partial}{\partial \gamma_k^i} \Pr[R^i(B^i) < r] \leq 0$, for all arbitrary $k \in B\ \{d\}$. Direct calculation shows following:

$$\begin{aligned}\frac{\partial}{\partial \gamma_k^i} \Pr[R^i(B^i) < r] &= \frac{\partial}{\partial \gamma_k^i} \sum_{E \in 2^B} \Pr[R^i(E) < r] \Pr[B^i = E] \\ &= \frac{\partial}{\partial \gamma_k^i} \sum_{E \in 2^{B-k}} A_k^i(E) \big((1 - \gamma_k^i) \Pr[R^i(E) < r] \\ &\qquad + \gamma_k^i \Pr[R^i(E \cup \{k\}) < r]\big) \\ &= \sum_{E \in 2^{B-k}} A_k^i(E) \big(\Pr[R^i(E \cup \{k\}) < r] - \Pr[R^i(E) < r]\big) \\ &\leq 0,\end{aligned}$$

since $\Pr[R^i(E \cup \{k\}) < r] \leq \Pr[R^i(E) < r]$, where $A_k^i(E) = \frac{\Pr[E = B^i]}{1 - \gamma_k^i}$. □

1.A.3 Proof of Theorem 1.2

Proof. Arrange A so that it can satisfies ordering condition, $n_j \geq n_{j'}$ if $p_j o_j > p_{j'} o_{j'}$. In equilibrium, the aggregate demand of j is given as follow:

$$\frac{1}{o_j} = s_j = \gamma_j \frac{\prod_{l>n_j}(1-\gamma_l)}{1-\prod_{l\in A}(1-\gamma_l)}.$$

Thus, as long as A keep satisfies ordering condtion,

$$p_i o_i > p_{i'} o_{i'} \quad (\Leftrightarrow) \quad \frac{p_i}{p_{i'}} > \frac{\gamma_i}{\gamma_{i'}} \frac{\prod_{m>n_i}(1-\gamma_m)}{\prod_{m>n_{i'}}(1-\gamma_m)} = \frac{\gamma_i}{\gamma_{i'} \prod_{n_{i'}<m\leq n_i}(1-\gamma_m)},$$

for $n_i > n_{i'}$. It is equivalent to following inequality.

$$\frac{p_{n_i+1}}{p_{n_i}} > \frac{\gamma_{n_i+1}}{\gamma_{n_i}(1-\gamma_{n_i})} \quad \forall n_i = 1, 2, \ldots, |B|-1.$$

□

1.A.4 Proof of Theorem 1.3

Proof. Define following subsets of B.

$$\begin{aligned} B_{<k} &= \{j \in B | p_j o_j < p_k o_k\} \\ B_{>k} &= \{j \in B | p_k o_k < p_j o_j\} \\ B_{=k} &= \{j \in B | p_j o_j = p_k o_k\}. \end{aligned}$$

Claim 1.1. *$B_{<k}$ is empty if $\{o_j\}_{j\in B}$ is in equilibrium.*

Proof. Suppose not. Then, $\exists j \in B_{<k}$ and o_j which is in equilibrium. i.e.

$$\Pr[B^i \subset B_{<k}] \leq \frac{1}{o_k} < \frac{p_k}{p_j o_j} \leq 1 - \Pr[k \in B^i]$$

since $p_j o_j < p_k o_k$ if $j \in B_{<k}$. But last term is equal to 0 since k is always attended. It is contradiction. □

By claim 1.1, $B = B_{>k} \uplus B_{=k}$. If $B_{>k}$ is not empty, the proof ends. Let $\{n_j\}_{j \in B_{>k}}$ be the natural number index starts from 1 and assume $n_j > n_{j'}$ if $p_j o_j > p_{j'} o_{j'}$. Observe,

$$p_k o_k = \frac{\sum_{l \in B_{=k}} p_l}{\prod_{l \in B_{>k}} (1 - \gamma_l)}$$

since,

$$\sum_{l \in B_{=k}} p_l = p_k o_k \sum_{l \in B_{=k}} s_l = p_k o_k \prod_{l \in B_{>k}} (1 - \gamma_l).$$

In equilibrium,

$$s_j = \frac{1}{o_j} = \gamma_j \prod_{l > n_j} (1 - \gamma_l),$$

and by assumption,

- $p_1 o_1 > p_k o_k$
- $p_{n_j} o_{n_j} > p_{n_{j'}} o_{n_{j'}}$ if $n_j > n_{j'}$ for $j, j' \in B_{>k}$.

It means, the security prices $\frac{1}{o_j}$ are distorted for $j \in B_{>k}$, if following inequalities hold:

$$\begin{aligned} \frac{p_1}{\sum_{l \in B_{=k}} p_l} &> \frac{\gamma_1}{1 - \gamma_1} \\ \frac{p_{n_j+1}}{p_{n_j}} &> \frac{\gamma_{n_j+1}}{\gamma_{n_j}(1 - \gamma_{n_j})} \quad \forall n_j = 1, 2, \ldots, |I_{>k}| - 1. \end{aligned}$$

□

1.B Substitution Between Betting Types

Denote the sales share of betting type t at the rth race by s_t^r, for $r = 1, \ldots, R$ and $t \in \{win, quinella, exacta, others\}$. Here, the betting type *others* is the sum of all other minor bets, such as place or show. To see

	Win	Exacta	Others
β_0	0.72	0.91	0.85
	(0.01)	(0.003)	(0.005)
β_1	-4.72	-1.17	-1.47
	(1.302)	(0.017)	(0.044)
R^2	0.01	0.79	0.48

Table 1.2: Estimation result

the substitution effect, the following simple regression model is analyzed for $t \in \{win, exacta, others\}$:

$$s^r_{quinella} = \beta_0 + \beta_1 s^r_t + \varepsilon^r_t.$$

Table 1.B summarizes the results. The R^2-values for each model indicate that a large part of the variation in the share of quinella bets is explained solely by the variation in exacta bet sales. Furthermore, the esimated rate of substitution between quinella and exacta bets is close to -1 which supports the hypothesis that they are close substitutes.

1.C Estimation of the Empirical Return Function

1.C.1 Data

Let I be a set of bets. For each $i \in I$, the betting data form a pair $(o_i, w_i) \in \mathbb{R}^+ \times \{0, 1\}$, where o_i denotes the odds of bet i and w_i is 1 if bet i wins, and 0 otherwise. Define

$$E_O = \{o \in \mathbb{R} | o = o_i, i \in I\},$$

and denote its elements by o_k for $k = 1, \ldots, K$, where $K = |E_O|$ and $o_k < o_{k'}$ if $k < k'$. Write

$$\begin{aligned} n_k &= |\{i \in I | o_i = o_k\}|, \\ w_k &= |\{i \in I | (o_i, w_i) = (o_k, 1)\}|, \end{aligned}$$

and $r_k = o_i \frac{w_k}{n_k}$. The data used for the estimation are, (o_k, r_k, n_k, w_k) for $k = 1, \ldots, K$.

1.C.2 Estimation Strategy

There are two major issues that need to be addressed to estimate the empirical return function: i) the parametric form of $r(\cdot)$; and ii) the heteroscedasticity. It is not innocuous to assume that an unknown function has a certain parametric form without a sound theoretical or intuitive background. Recall that no specific economic assumptions or interpretations have been made about the empirical return function so far. Instead of resorting to a specific parametric form, I apply a nonparametric estimation method called a locally weighted regression (LOESS), which gives greater weight to data closer to an evaluation point; that is, it implicitly assumes the smoothness of $r(\cdot)$.

The second issue, heteroscedasticity, stems from the difference in the informativeness of an observation. From the data, n_k is the number of bets with odds o_k, and w_k is the number of winning bets with odds o_k. Thus, it is natural to model w_k as the sum of n_k independent Bernoulli random variables with probability $p(o_k)$, where $p(o)$ is the conditional probability $\Pr[W_i = 1 | O_i = o]$. Observe that $E\frac{w_k}{n_k}$ is $p(o_k)$ and $r(o) = op(o)$. Thus, the

data-generating process of r_k is

$$r_k = r(o_k) + \varepsilon_k,$$

where $\varepsilon_k \sim \left(0, o_k^2 \frac{p(o_k)(1-p(o_k))}{n_k}\right)$. Note that $Var(\varepsilon_k)$ is $o_k^2 \frac{p(o_k)(1-p(o_k))}{n_k}$, which implies heteroscedasticity.

If $p(o_k)$ is known, we can calculate the exact value of $Var(\varepsilon_k)$ and adjust the estimation using the generalized least squares estimation method, but this is not yet feasible. My approach is to use a consistent estimate $\hat{p}(o_k)$ as a proxy for $p(o_k)$. The procedure comprises two steps. In the first step, $\hat{p}(o_k)$ is estimated using LOESS and a proper adjustment matrix $\hat{\Sigma}$ is derived from the estimate. In the second step, $\hat{r}(o)$ is estimated using the adjusted LOESS based on the adjustment matrix $\hat{\Sigma}$. The steps of the procedure are as follows.

Step 1: In the first locally weighted regression procedure, the regressor is the triple $(1, o_k, o_k^2)$ and the regressand is the set of observed returns r_k. Write $\mathbf{o} = (1, o, o^2)$, $\mathbf{O}$ as a $k \times 3$ matrix, with kth row equal to $\mathbf{o}_k$, and $\mathbf{R}$ as a $K \times 1$ column vector, with the kth element equal to r_k. The LOESS estimator can be written as follows:

$$\hat{r}_{LOESS}(o) = \mathbf{o}\left(\mathbf{O}^T\mathbf{W}(o)\mathbf{O}\right)^{-1}\mathbf{O}^T\mathbf{W}(o)\mathbf{R}. \tag{1.4}$$

Here, $\mathbf{W}(o)$ is a weighting matrix. Specifically, it is a $K \times K$ diagonal matrix, with each diagonal element $w(x)_{k,k}$ set to

$$w(o)_{k,k} = K\left(\frac{|o - o_k|}{d(o)}\right).$$

Here $K(x) = (1 - x^3)^3$ if $x \in [0, 1]$, and 0 otherwise and $d(o)$ is the distance between the qth nearest o_k and o. Intuitively, $K(\cdot)$ and $d(\cdot)$ can be thought

of as the kernel function and the bandwidth, respectively if the k-nearest kernel is considered. In the analysis, q is fixed to the nearest integer of $0.4K$. Set $\hat{p}(o_k) = \frac{1}{o_k}\hat{r}_{LOESS}(o)$ and $\hat{\sigma}_k^2 = o_k^2 \frac{\hat{p}(o_k)(1-\hat{p}(o_k))}{n_k}$. A proper adjustment matrix $\hat{\Sigma}$ is defined by a $K \times K$ diagonal matrix, with the kth elements equal to $\hat{\sigma}_k^2$.

Step 2: Using $\hat{\Sigma}$, find the following estimate:

$$\hat{r}(o) = \mathbf{o}\left(\mathbf{O}^T\mathbf{W}(o)\hat{\Sigma}^{-1}\mathbf{O}\right)^{-1}\mathbf{O}^T\mathbf{W}(o)\hat{\Sigma}^{-1}\mathbf{R}.$$

Chapter 2

Fast Computation Algorithm for the Random Consideration Model

2.1 Introduction

Random consideration models assume an environment where consumers do not consider all feasible alternatives. However, the computational complexity of these models makes them difficult to apply. The root cause of the complexity is that consumer consideration is not observable, combined with an overwhelmingly large number of possible consideration states. For example, determining the choice probability in a market where 20 products are available involves more than a million states. Thus, the naive calculation method is constrained in terms of applicability. Goeree (2008) proposed a

This chapter is based upon the project co-authored with Prof. Kyoungwon Seo.

Monte Carlo-based method to circumvent this issue but still is complex and inaccurate.

This study proposes a more reliable and faster computation algorithm. Instead of circumventing the complexity issue, I find a simple random utility representation of the choice probability, and then apply quadrature integration methods to find a solution. This approach leaves room for researchers to improve the computational efficiency by choosing an appropriate quadrature rule for their setting.

The new algorithm outperforms the existing Monte Carlo-based method[2] in terms of accuracy. Furthermore, the computational burden increases in proportion to the number of feasible alternatives, while that of the naive algorithm increases exponentially. This improves the efficiency significantly. Theoretically, the new algorithm has a better convergence rate than that of the existing Monte Carlo-based method. Numerical experiments show that to achieve a similar level of error, the existing method requires more than $7,000$ times as many samples as the new algorithm.

The remainder of the paper is structured as follows. In section 2, I derive the random utility representation of a random consideration model that is numerically tractable. In section 3, a quadrature integration-based numerical algorithm is proposed. In section 4, the performance of the algorithms are compared using numerical experiments. Section 5 concludes the paper.

[2] Goeree (2008)

2.2 Analysis

I begin with the standard discrete choice model. Let i and j be the indices of consumers and products, respectively. Let F be the set of all feasible alternatives. Assume that consumer i has utility values U_j^i for each product $j \in F$. I assume the following for the utility value U_j^i:

$$U_j^i \sim F(\cdot|\alpha_j^i).$$

In this setting, if consumer i makes a choice after taking all feasible alternatives into account, s_j^i, the probability of i choosing j is equal to the probability that $U_j^i \geq U_l^i$, for all $l \in F$. That is,

$$\begin{aligned} s_j^i &= \Pr[U_j^i > U_l^i \quad \forall l \in F] \\ &= \int_{-\infty}^{\infty} \prod_{l \in F \setminus \{j\}} F(u|\alpha_l^i) dF(u|\alpha_l^i). \end{aligned} \tag{2.1}$$

In applications, U_j^i is usually assumed to be a Gumbel noise contaminated random variable. For example, the textbook example of Train (2009) uses the following specification:

$$\begin{aligned} U_j^i &= \alpha_j^i + \varepsilon_j^i \\ \alpha_j^i &= X_j \beta^i, \end{aligned}$$

where ε_j^i is an independent Gumbel noise distributed by $e^{-e^{-\varepsilon_j^i}}$, X_j denotes the observable product characteristics of j, and β_i is the consumer's preference. Then, a direct calculation shows that the choice probability has the following simple multinomial logit form:

$$s_j^i = \frac{e^{X_j \beta^i}}{\sum_{l \in F} e^{X_l \beta^i}}.$$

2.2.1 Random Consideration Model

In contrast to the standard discrete choice model, the random consideration model assumes that a comsumer conceives the subset of feasible alternatives stochastically. Let C^i be the consideration set conceived by consumer i. If there exists a probability function $\Pr[C^i = E]$ for each $E \in 2^F$, then the choice probability s^i_j satisfies the following equation:

$$s^i_j = \sum_{j \in E \in 2^F} \Pr[C^i = E] \Pr[U^i_j \geq U^i_l, \quad \forall l \in E]. \tag{2.2}$$

The random consideration model further specifies that the consideration probability $\Pr[C^i = E]$ satisfies a certain condition. It assumes that the event that alternative j is conceived by i is an independent binary event with probability γ^i_j. That is.

$$\Pr[C^i = E] = \prod_{m \in E} \gamma^i_m \prod_{n \notin E} (1 - \gamma^i_n) \quad \forall E \in 2^F.$$

The random consideration model has an advantage in applications, in addition to being axiomatically founded[3]: it allows researchers to explicitly analyze the role of various attention grabbers in a consumption decison. For example, Chiang et al. (1998) analyze the impact of brand recognition on ketchup sales, and Goeree (2008) studies the personal computer industry in conjunction with advertising expenses.

On the other hand, equation (2.2) reveals the major drawback of the model. Equation (2.2) shows that the number of terms in the summation operator is $2^{|F|-1}$, which increases exponentially with the number of feasible alternatives. This severely constrains the applicability of the model

[3]Manzini and Mariotti (2014) provide the axiomatic foundation.

because the postulation that consumers do not consider all feasible alternatives is especially appealing if there is an overwhelmingly large number of alternatives.

2.2.2 Random Utility Representation

The following theorem implies that the exponentially increasing computational complexity can be reduced significantly.

Theorem 2.1. *(Random Utility Representation)*

$$s_j^i = \gamma_j^i \int_{-\infty}^{\infty} \prod_{l \in F \setminus \{j\}} [1 - \gamma_l^i + \gamma_l^i F(u|\alpha_l^i)] dF(u|\alpha_j^i).$$

Proof. See the Appendix. □

Two facts are noteworthy from the above theorem: i) the integrand is composed of the product of $|F|$ terms; and ii) each term of the integrand are cumulative distribtion functions. The first fact implies that the computational complexity of the choice probability calculation using a random utility representation has linear scalability relative to the number of feasible alternatives. Instead of adding $2^{|F|-1}$ terms, we can find the desired solution by multiplying $|F|$ terms, as long as we have a proper numerical integration method.

The name of the theorem is taken from the second fact that each term $[1 - \gamma_l^i + \gamma_l^i F(u|\alpha_l^i)]$ is a cumulative distirubution function derived from $F(u|\alpha_l^i)$.[4] It shows that the random utility representation is numerically tractable as long as $F(u|\alpha_l^i)$, for $l \in F$ is regular because each term is equal to an affine transformation of $F(u|\alpha_l^i)$.

[4]Given $\{\alpha_l^i\}_{l \in F}$ and $\{\gamma_l^i\}_{l \in F}$, each integrand $[1 - \gamma_l^i + \gamma_l^i F(u|\alpha_l^i)]$ is equivalent to the

2.3 Algorithm

The proposed method numerically integrates the random utility representation using a quadrature integration method. Thus, I call it the *Quadrature-based Choice Probability Estimation* (QCPE). In contrast, the existing method proposed by Goeree (2008) is based on a Monte Carlo integration. Thus, I refer to the latter method as the *Monte Carlo-based Choice Probability Estimation* (QCPE). Here, I present their detailed algorithms and convergence properties.

2.3.1 MCPE

To circumvent the computational issues caused by the overwhelmingly large number of terms in the summation operator in equation (2.2), MCPE use randomly generated consideration sets to approximate it. Denote the MCPE estimate of the choice probability by $\hat{s}^i_{j|MCPE}$. The calculation procedure is as follows.

Algorithm 2.1. *(MCPE) The output of this algorithm is the estimated choice probability* $\{\hat{s}^i_{j|MCPE}\}_{j\in F}$*. The input arguments are the number of Monte Carlo samples* N*, attention probability* $\{\gamma^i_j\}_{j\in F}$*, and density function of the random utility* $\{F(\cdot|\alpha^i_j)\}_{j\in F}$*.*

1. *Randomly generate a set of consideration sets* $\{C^i_n\}_{n=1}^N$ *with probabil-*

cumulative distribution function of $\hat{U}^i_j$, such that

$$\hat{U}^i_j = \begin{cases} U^i_j & w.p. \quad \gamma^i_j \\ -\infty & w.p. \quad 1-\gamma^i_j \end{cases}.$$

ity

$$\Pr[E = C_n^i] = \prod_{m \in E} \gamma_m^i \prod_{n \notin E} (1 - \gamma_n^i).$$

2. *Let*

$$\hat{s}_{j|MCPE}^i = \sum_{n=1}^{N} \Pr[U_j^i \geq U_l^i, \quad \forall l \in C_n^i].$$

Here, $\hat{s}_{j|MCPE}^i$ *for* $j \in F$ *yields the desired results.*

Monte Carlo methods usually have an $O(N^{-1/2})$ convergence rate, and the MCPE is no exception. To check the convergence rate, let $\mathbf{1}_n^E$ be a random variable equal to one 1 if $C_j^i = E$, and zero otherwise. Then, the MCPE is defined by the following equation:

$$\hat{s}_{j|MCPE}^i = \sum_{j \in E \in 2^F} \left(\frac{\sum_{n=1}^{N} \mathbf{1}_n^E}{N} \right) \Pr[U_{ij} \geq U^{il}, \quad \forall l \in E].$$

See the following proposition.

Proposition 2.1.

$$\hat{s}_{j|MCPE}^i = s_j^i + O_p(N^{-\frac{1}{2}}).$$

Proof. See the Appendix. □

2.3.2 QCPE

In addition to the reduction in computational complexity, another advantage of working with the QCPE is that the researcher can choose the quadrature rule. Depending on the characteristics of the distribution functions $F(\cdot|\alpha_j^i)$, the precision of the approximation can be improved by choosing an appropriate quadrature rule. For example, if the distribution functions are sufficiently smooth, it would be better to choose a Gaussian quadrature.

Other quadrature rules, such as the trapezoidal rule, can also improve the accuracy.[5]

For ease of exposition, I consider one specific case: the trapezoidal rule is used to approximate the following Stieltjes integration:

$$s_j^i = \gamma_j^i \int_{-\infty}^{\infty} \prod_{l \in F \setminus \{j\}} [1 - \gamma_l^i + \gamma_l^i F(u|\alpha_l^i)] d \ln F(u|\alpha_j^i).$$

The calculation is as follows.

Algorithm 2.2. *(QCPE) The output of this algorithm is the estimated choice probability* $\{\hat{s}_i\}_{i \in A}$*. The input arguments are the tolerance value* ϵ*, number of grid points* N*, attention probability* $\{\gamma_j^i\}_{j \in F}$*, and density function of the random utility* $\{F(\cdot|\alpha_j^i)\}_{j \in F}$*. For notational simplicity, define* $F(\cdot|\alpha_j^i, \gamma_j^i) = 1 - \gamma_j^i + \gamma_j^i F(\cdot|\alpha_j^i)$*.*

1. *Set*

$$\begin{aligned} \bar{u} &= \max\{u | u = F^{-1}(1 - \epsilon/2|\alpha_i), i \in A\} \\ \underline{u} &= \min\{u | u = F^{-1}(\epsilon/2|\alpha_i), i \in A\} \end{aligned}$$

2. *Define the* $(N+1) \times 1$ *column vector* $\mathbf{u}$ *such that* $\mathbf{u}_i = \underline{u} + (\bar{u} - \underline{u}) \frac{(i-1)}{N}$*.*

3. *Define the* $(N+1) \times |A|$ *matrix* $\mathbf{F}$*, where* $\mathbf{F}_{n,i} = F(\mathbf{u}_n | \alpha_i, \pi_i)$*.*

4. *Define an* $N \times |A|$ *matrix* $\mathbf{DF}$ *and an* $N \times 1$ *column vector* $\mathbf{\Pi}$ *such*

[5] In many cases, an evaluation of the convergence rate of the Gaussian quadrature integration method is impractical. Nonetheless, I recommend using the Gaussian quadrature rule because it is known to outperform other quadrature rules as long as the integrand is sufficiently smooth. See Laurie (1985).

that

$$\begin{aligned}\mathbf{DF}_{n,i} &= \ln(\mathbf{F}_{n+1,i}) - \ln(\mathbf{F}_{n,i}) \\ \mathbf{\Pi}_n &= \frac{1}{2}\left(\prod_{i\in A}\mathbf{F}_{n,i} + \prod_{i\in A}\mathbf{F}_{n+1,i}\right).\end{aligned}$$

5. *Let* $\hat{\mathbf{s}}_{QCPE} = \mathbf{DF}^T \cdot \mathbf{\Pi}$. *Then,* $\hat{\mathbf{s}}_i = \hat{s}_i$, *which is the desired result.*

The following proposition shows that the convergence rate of this algorithm is $O(N^{-1})$.

Proposition 2.2. *A QCPE* $\hat{s}^j_{i|QCPE}$ *with* N *grid points satisfies the following equation for an arbitrary constant* $\epsilon > 0$.

$$\hat{s}^j_{i|QCPE} = s^j_i + \epsilon + O(N^{-1}).$$

Proof. See the Appendix. □

As mentioned previously, this convergence rate can be improved further by applying the Gaussian quadrature rule or the trapezoidal rule directly to the random utility representation.

2.4 Numerical Example

The purpose of this numerical experiment is to evaluate the performance of the algorithms for different numbers of samples or nodes so that we can compare them in terms of both accuracy and complexity. Three different algorithms are considered: MCPE, QCPE with trapezoidal rule, and QCPE with the Gaussian quadrature rule.

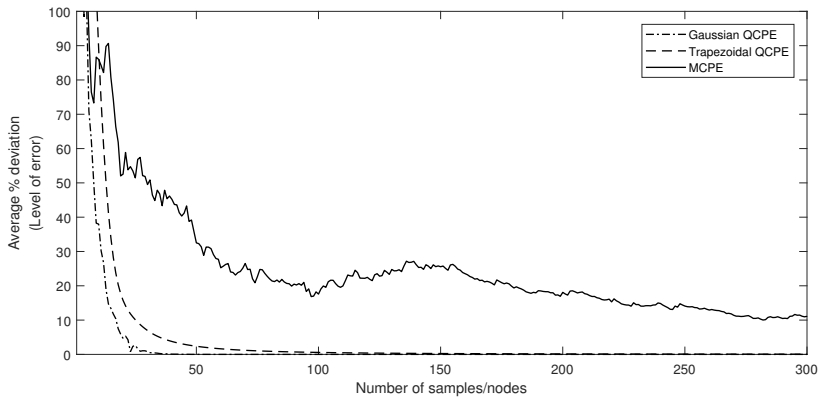

Figure 2.1: Approximation error by number of samples/nodes

2.4.1 Simulation Setting

Let $\hat{s}_j(N|a)$ be the choice probability of the jth product estimated by algorithm a with N samples or nodes. The performance metric $M(N|a)$, given N and a, is defined as follows:

$$M(N|a) = 100 \times \sqrt{\frac{1}{|F|} \sum_{j \in F} \left(\frac{\hat{s}_j - s_j}{s_j} \right)^2},$$

where s_j is the true choice probability of the jth product. This is analogous to the percentage standard deviation of the estimate.

I assume a manageable size of the alternative set $F = \{1, \ldots, 10\}$ for the simulation so that we can determine the true choice probability s_j. For each $j \in F$, the random utility is generated by

$$U_j^i = \alpha_j^i + \varepsilon_j^i,$$

where ε_j^i is independently distributed by $e^{-e^{-\varepsilon_j^i}}$ which is a Gumbel noise. I set α_j^i to j and γ_j^i to $\frac{1}{j}$.

Algorithm	$< 20\%$	$< 10\%$	$< 5\%$	$< 1\%$	$< 0.5\%$
Gaussian QCPE	13	18	20	23	32
Trapezoidal QCPE	19	26	35	77	108
MCPE	89	412	1,704	31,987	245,606

Table 2.1: Number of samples or nodes required to achieve error bound

To focus on the convergence patterns of the algorithms, the random consideration set samples $\{C_n^i\}_{n=1}^N$ are reused to calculate $\hat{s}_j(N+1|\text{MCPE})$ for the MCPE algorithm. Specifically, the consideration set samples $\{C_n^i\}_{n=1}^N \cup \{C_{N+1}^i\}$ are used to calculate $\hat{s}_j(N+1|\text{MCPE})$ for all N, where C_{N+1}^i is an independently generated random consideration set. For the QCPE algorithms, the tolerance value ϵ is set to 10^{-10}.

2.4.2 Simulation Result

The simulation result is consistent with the theoretical prediction. Figure 2.1 shows that both versions of the QCPE algorithms outperform the MCPE algorithm for the same number of nodes and samples. In addition, consistent with the existing literature[6], the Gaussian QCPE yields a more accurate result than that of the trapezoidal QCPE.

Using the manageable number of nodes, the QCPE algorithms achieve an acceptible level of error; a 0.1% level of error is achieved by the Gaussian QCPE using 47 nodes and by the trapezoidal QCPE using 238 nodes. On the other hand, the MCPE fails to achieve this error level with a million samples. Table 2.1 shows the smallest number of samples or nodes required

[6] Laurie (1985)

to achieve a certain level of error for each algorithm. The simulation result shows that 32 nodes are used by the Gaussian QCPE to achieve a 0.5% level of error, while more than $200,000$ samples are required for the MCPE to achieve the same level of precision, which is more than $7,000$ times larger.

2.5 Conclusion

I have proposed a QCPE algorithm for the choice probability estimation of random consideration models. In contrast to the existing MCPE algorithm, which seeks to circumvent the innate complexity of the model, I find a simpler and numerically tractable representation of the model, which I call a random utility representation. By exploiting this representation, the QCPE provides a fast and reliable solution. A remaining issue not covered in this study is that of preference heterogeneity. In many economic applications, especially in empirical industrial organization studies, consumer preferences are assumed to be heterogeneous. It is expected that the QCPE algorithm will be a building block for future studies.

Appendix

2.A Proofs

2.A.1 Proof of theorem 2.1

For the simplicity, I suppress the superscript i and write $I_{-j} = I \backslash \{j\}$ for a $j \in I$.

Lemma 2.1. *If I is the finite index set and there exist a sequence of I indexed number (a_m, b_m) for all $m \in I$ then,*

$$\sum_{E \in 2^I} \prod_{m \in E} a_m \prod_{n \in I \backslash E} b_n = \prod_{l \in I} (a_l + b_l).$$

Proof of Lemma. It is sufficient to show that

$$\sum_{E \in 2^I} \prod_{m \in E} a_m \prod_{n \in I \backslash E} b_n = (a_j + b_j) \sum_{E \in 2^{I_{-j}}} \prod_{m \in E} a_m \prod_{n \in I_{-j} \backslash E} b_n.$$

Observe that for $E \in 2^I$ and $j \in I$, either $j \in E$ or $j \notin E$. And, if $j \in E$, then $j \notin I \backslash E$ and vise verse. Thus,

$$\sum_{E \in 2^I} \prod_{m \in E} a_m \prod_{n \in I \backslash E} b_n = \sum_{E \in 2^{I_{-j}}} \left(a_j \prod_{m \in E} a_m \right) \prod_{n \in I_{-j} \backslash E} b_n + \sum_{E \in 2^{I_{-j}}} \prod_{m \in E} a_m \left(b_j \prod_{n \in I_{-j} \backslash E} b_n \right).$$

□

Proof of Theorem. From equation (2.2),

$$\begin{aligned}
s_j &= \sum_{j \in E \in 2^F} \left(\prod_{m \in E} \gamma_m \prod_{n \notin E} (1 - \gamma_n) \right) \int_{-\infty}^{\infty} \prod_{l \in E_j} F(u|\alpha_l) dF(u|\alpha_j) \\
&= \gamma_j \int_{-\infty}^{\infty} \sum_{E \in 2^{F_{-j}}} \left(\prod_{m \in E} \gamma_m \prod_{n \in F_{-j} \backslash E} (1 - \gamma_n) \right) \prod_{l \in E} F(u|\alpha_l) dF(u|\alpha_j) \\
&= \gamma_j \int_{-\infty}^{\infty} \sum_{E \in 2^{F_{-j}}} \left(\prod_{m \in E} \gamma_m F(u|\alpha_m) \prod_{n \in F_{-j} \backslash E} (1 - \gamma_n) \right) dF(u|\alpha_j)
\end{aligned}$$

Using lemma 2.1, put $a_m = \gamma_m F(u|\alpha_m), b_n = 1 - \gamma_n$ and $I = F_{-1}$ then it proves the theorem. $\square$

2.A.2 Proof of proposition 2.1

Proof. It is clear that $E[\hat{s}^i_{j|MCPE}] = s^i_j$ since $E\left[\frac{\sum_{k=1}^N \mathbf{1}_k^E}{N}\right] = \Pr[C^i = E]$. For the convergence rate, observe followings.

$$\begin{aligned} var(\hat{s}^i_{j|MCPE}) &= E\left[\left(\hat{s}^i_{j|MCPE} - s^i_j\right)^2\right] \\ &= E\left[\left(\sum_{i \in E \in 2^F}\left(\frac{\sum_{k=1}^N \mathbf{1}_k^E}{N} - \Pr[C^i = E]\right)\Pr[U^i_j \geq U^i_l, \quad \forall l \in E]\right)^2\right] \\ &= \frac{1}{N}\left[\sum_{i \in E \in 2^F} \Pr[C^i = E]\Pr[U^i_j \geq U^i_l, \quad \forall l \in E]^2 - (s^i_j)^2\right] \end{aligned}$$

From above, consider $\Pr[U^i_j \geq U^i_l, \quad \forall l \in E]$ as a random variable $S_i(E)$ defined on 2^F and $\Pr[C^i = E]$ as a probability mass function. Then, following is true.

$$var\left(\hat{s}^j_{i|MCPE}\right) = \frac{1}{N} var(S_i)$$

Consequently,

$$std(\hat{s}^j_{i|MCPE}) = \frac{1}{\sqrt{N}} std(S_i).$$

It implies that convergence rate of $\hat{s}^i_{j|MCPE}$ is $O(N^{-\frac{1}{2}})$. $\square$

2.A.3 Proof of proposition 2.2

Proof. Denote the QCPE of s^i_j by $\hat{s}^i_{j|QCPE}$. And recall the set of equally spaced points $\{u_n\}_{n=1}^{N+1}$ such that $u_n = \underline{u} + (\bar{u} - \underline{u})\frac{(n-1)}{N}$. For the nontational

simplicity, write

$$\begin{aligned}
F(u|\alpha_j,\gamma_j) &= 1-\gamma_j^i+\gamma_j^i F(u|\alpha_j^i) \\
\Delta G(u_n) &= G(u_{n+1})-G(u_n) \\
\overline{\prod_{l\in F} F(u|\alpha_l,\gamma_l)} &= \frac{1}{2}\left(\prod_{l\in F} F(u_n|\alpha_l,\gamma_l)+\prod_{l\in F} F(u_{n+1}|\alpha_l,\gamma_l)\right),
\end{aligned}$$

where G is any function of u, and $n=1,\ldots,N$. Using new notations, $\hat{s}^i_{j|QCPE}$ can be written as follow.

$$\hat{s}^i_{j|QCPE}=\sum_{n=1}^{N}\overline{\prod_{l\in F} F(u_n|\alpha_l,\gamma_l)}\Delta\ln F(u_n|\alpha_l,\gamma_l)$$

Then,

$$\begin{aligned}
|s^i_j-\hat{s}^i_{j|QCPE}| &\leq \underbrace{\left|s^i_j-\int_{\underline{u}}^{\bar{u}}\prod_{l\in F\setminus\{j\}} F(u|\alpha_l,\gamma_l)]dF(u|\alpha_j,\gamma_j)\right|}_{(*)} \\
&+ \underbrace{\left|\int_{\underline{u}}^{\bar{u}}\prod_{l\in F\setminus\{j\}} F(u|\alpha_l,\gamma_l)]dF(u|\alpha_j,\gamma_j)-\hat{s}^i_{j|QCPE}\right|}_{(**)}.
\end{aligned}$$

For $(*)$,

$$\begin{aligned}
(*) &= \left|\int_{-\infty}^{\underline{u}}\prod_{l\in F\setminus\{j\}} F(u|\alpha_l,\gamma_l)]dF(u|\alpha_j,\gamma_j)+\int_{\bar{u}}^{\infty}\prod_{l\in F\setminus\{j\}} F(u|\alpha_l,\gamma_l)]dF(u|\alpha_j,\gamma_j)\right| \\
&\leq F(\underline{u}|\alpha_j,\gamma_j)+(1-F(\bar{u}|\alpha_j,\gamma_j))\leq\epsilon.
\end{aligned}$$

For $(**)$,

$$\begin{aligned}
(**) &= \left|\int_{\underline{u}}^{\bar{u}}\prod_{l\in F} F(u|\alpha_l,\gamma_l)]d\ln F(u|\alpha_j,\gamma_j)-\hat{s}^i_{j|QCPE}\right| \\
&= \left|\sum_{n=1}^{N}\int_{u_{n+1}}^{u_n}\prod_{l\in F} F(u|\alpha_l,\gamma_l)]d\ln F(u|\alpha_j,\gamma_j)-\overline{\prod_{l\in F} F(u_n|\alpha_l,\gamma_l)}\Delta\ln F(u_n|\alpha_l,\gamma_l)\right|.
\end{aligned}$$

Observe that there exists $\tilde{u}_n \in (u_n, u_{n+1})$ for all $n = 1, \ldots, N$ such that

$$\int_{u_{n+1}}^{u_n} \prod_{l \in F} F(u|\alpha_l, \gamma_l)] d \ln F(u|\alpha_j, \gamma_j) = \prod_{l \in F} F(\tilde{u}_n|\alpha_l, \gamma_l) \Delta \ln F(u_n|\alpha_j, \gamma_j)$$

since $\prod_{l \in F} F(u|\alpha_l, \gamma_l)$ and $\ln F(u|\alpha_j, \gamma_j)$ are continuous and monotonically increasing. Consequently,

$$\begin{aligned}
(**) &= \left| \sum_{n=1}^{N} \left(\prod_{l \in F} F(\tilde{u}_n|\alpha_l, \gamma_l) - \overline{\prod_{l \in F} F(u_n|\alpha_l, \gamma_l)} \right) \Delta \ln F(u_n|\alpha_j, \gamma_j) \right| \\
&\leq \sum_{n=1}^{N} \kappa |\tilde{u}_n - u_n| |\Delta \ln F(u_n|\alpha_j, \gamma_j)| \\
&\leq \frac{\kappa}{N} \ln \frac{1-\epsilon}{\epsilon}.
\end{aligned}$$

For some constant $\kappa > 0$. First inequality comes from the differentiability of F. $\square$

Chapter 3

Robust Wald Test for an Ill-Posed or Ill-Conditioned Linear Mode

3.1 Introduction

If the economic question behind an econometric model requires high-dimensional hypothesis testing with high-dimensional data, such as Slutsky symmetry, the standard test procedure does not work well. For example, Deaton and Muellbauer (1980), Meisner (1979), and Laitinen (1978) have shown that the tests for Slutsky symmetry and demand homogeneity tend to be overly rejected. In a presence of high dimensionality, a design matrix is usually rank deficient or near singular. I such cases, the estimation equation is ill-posed or ill-conditioned, which is susceptible to overfitting in

This chapter is based upon the project co-authored with Prof. Kyoungwon Seo.

the parameter estimation, including the nuisance parameters. Thus, there is a risk that the critical region is overconfidently constructed, leading to misleading test results.

In this study, I propose a new robust Wald test procedure that is attainable even when the econometric model is ill-posed or ill-conditioned. The proposed method has two robust properties: attainability and conservativeness. The test statistic is attainable even in an environment where the design matrix is singular or near singular and has a smaller type-I error than that of the standard Wald test. Thus, it prevents researchers from overconfidently rejecting the null hypothesis in the event of insufficient information.

Several successful alternative methods have been proposed for regression problems with ill-posed or ill-conditioned econometric models, but the direct application of such methods to the test procedure does not guarantee the impartiality of the test. Lasso[2], ridge[3], and the pseudo inverse estimators have been proved successful in various applications. However, using these estimators for a test is problematic because it is difficult to construct the right critical region. This is discussed in detail in section 2.

The robust Wald test proposed here is built on the idea that the fundamental reason an econometrician has an ill-posed or ill-conditioned econometric model is that she does not have sufficient information to obtain a stable point estimate. In particular, if the model is high dimensional, we cannot be sure that all available covariates are informative because some of them can similar information. In the robust Wald procedure, instead of

[2]Tibshirani (1996)

[3]Hoerl and Kennard (1970)

using all dimensions, seemingly less informative dimensions of information are thresholded out. Using the informative dimensions only, I find the estimator that most favors the null hypothesis from among estimators those are consistent with the given information.

The paper proceeds as follows. In section 2, I show when and why the Wald test fails. In section 3, the robust Wald test procedure is described. In section 4, a numerical example is presented, and section 5 concludes the paper.

3.1.1 Related Literature

Deaton and Muellbauer (1980), Meisner (1979), and Laitinen (1978) have shown that the tests for Slutsky symmetry and demand homogeneity tend to be overly rejected, even when the design matrix has full rank. Anatolyev (2012) provides a theoretical explanation for this phenomenon, showing that the size of the classical test can be distorted if the dimension of the parameter and the null hypothesis increase rapidly with the number of observations. The aforementioned study also proves that classical Wald-type test over-rejects the null hypothesis in such situations, and provides the correct asymptotic distributions of classical test statistics such as the F, likelihood ratio, and Lagrange multiplier tests.

On the other hand, the construction of classical hypothesis testing, which should be preceded by an estimation of the parameters, is challenging if the design matrix is rank deficient. Regularized estimators such as the ridge regression (Hoerl and Kennard, 1970) and lasso (Tibshirani, 1996) have been proved to be successful when estimating high-dimensional models, especially if the true model is suspected to be sparse. In econometrics,

many studies have dealt with regularized estimators in conjunction with their variable selection properties. For example, Belloni et al. (2014) proposed a double-selection method to find the right control variable from among a large number of candidate covariates, and Gillen et al. (2015) devised an choice probability estimation technique that accommodates high-dimensional demographic characteristics in the model.

Despite our increasing knowledge of regularized estimators, it is relatively unclear whether we can use them to construct an impartial test procedure because they are biased. Relatively few studies have examined debiasing issues in conjunction with general hypothesis testing (e.g., $R\beta = c$). Minnier et al. (2011) construct a confidence interval of regularized estimators, but it is not directly applicable to general hypothesis testing. More recently, Bühlmann et al. (2013), Zhang and Zhang (2014), and Javanmard and Montanari (2014) proposed debiased regularized estimators that are aymptotically normal, which means they can be used for general hypothesis testing. Bühlmann et al. (2013) suggest a debiased ridge estimator, Zhang and Zhang (2014) find a bias corrector using nodewise lasso estimates, and Javanmard and Montanari (2014) propose a debiased lasso estimator. In this study, I numerically compare the performance of a robust Wald test to that of the tests of Zhang and Zhang (2014) and Javanmard and Montanari (2014).

3.2 Why Does the Wald Test Not Work?

For regressions in the presence of a rank-deficient design matrix, the regularization methods and pseudo-inverse estimators have proved satisfactory

in many applications. However, the direct application of these techniques to hypothesis testing does not guarantee an impartial test result. The fundamental reason for this is that finding the distribution of the estimator under the null hypothesis is not an easy task. Furthermore, it is usually only possible to find a simple solution in a few cases. In this section, I illustrate the ridge estimator case because it is an exception of it.[4]

Consider the following example. The data-generating process is defined by

$$y_i = x_i^T \beta + \varepsilon_i,$$

where $x_i \sim (0, \Sigma_x), \varepsilon_i \sim N(0,1)$ and $x_i, \beta \in \mathbb{R}^k$. Now, suppose we wish to test the following hypothesis:

$$\begin{aligned} H_0 &: R\beta = c \\ H_A &: R\beta \neq c. \end{aligned}$$

Here, $R \in \mathbb{R}^{l \times k}$ and $c \in \mathbb{R}^l$. Assume a set of data (y_i, x_i) for $i = 1, \ldots, n$. For notational simplicity, write $y = (y_1, \ldots, y_n)^T$, $X = (x_1, \ldots, x_n)^T$, and $\hat{\Sigma} = X^T X$. Given λ, the ridge estimator $\hat{\beta}_{ridge}(\lambda)$ is the solution to the following minimization problem:

$$\min_{\beta \in \mathbb{R}^k} \sum_{i=1}^{n} (y_i - x_i^T \beta)^2 + \lambda \sum_{l=1}^{k} \beta_l^2,$$

where minimizer is

$$\hat{\beta}_{ridge}(\lambda) = \left(\hat{\Sigma} + \lambda I\right)^{-1} X^T y.$$

[4] Another reason is that Moore-Penrose estimators can be viewed as a special case of ridge estimator.

Thus,

$$\hat{\beta}_{ridge}(\lambda) \sim N\left(T(\lambda)\beta, \sigma^2\left(\hat{\Sigma}+\lambda I\right)^{-1} T(\lambda)\right),$$

where $T(\lambda) = I - \lambda\left(\hat{\Sigma}+\lambda I\right)^{-1}$. If $\lambda \to 0+$, $\hat{\beta}_{ridge}(\lambda)$ is the Moore-Penrose estimator. This shows that a ridge estimator is not an unbiased estimator of β. Even when $\lambda \to 0+$, the bias is suppressed as long as $\hat{\Sigma}$ is singular. Unless these estimators are properly debiased, size distortion seems inevitable in the standard Wald test procedure.

To compare the ridge estimator-based Wald test procedure to the standard procedure, suppose $\hat{\beta}$ is an unbiased estimator of β; that is $\hat{\beta} \sim N(\beta, \Sigma_\beta)$. Under the null hypothesis, $R\hat{\beta} \sim N(c, R\Sigma_\beta R^T)$ and thus, its Wald statistic,

$$W(\hat{\beta}) = \left(R\hat{\beta}-c\right)^T \left(R\Sigma_\beta R^T\right)^{-1}\left(R\hat{\beta}-c\right),$$

follows a chi-squared distribution with r_R degree of freedom, where r_R is the rank of $R\Sigma_\beta R^T$. On the other hand, in the case of the ridge estimator, we cannot characterize the distribution of test statistic precisely under the null hypothesis because $ER\hat{\beta}_{ridge}$ is $RT(\lambda)\beta$, not $R\beta$. We could ignore the bias and directly apply the Wald test procedure, treating $\hat{\beta}_{ridge}$ as an unbiased estimator, to construct the test statistic using the following equation:

$$W(\hat{\beta}_{ridge}) = \left(R\hat{\beta}_{ridge}-c\right)^T \left(R\sigma^2\left(\hat{\Sigma}+\lambda I\right)^{-1} T(\lambda)R^T\right)^{-1}\left(R\hat{\beta}_{ridge}-c\right).$$

Table 3.1 shows the empirical type-I error using this statistic. In this simulation, R, β, and σ^2 are set to $[1,-2], [2,1]^T$, and 1, respectively. The covariates x_i are generated by $N(0, \Sigma_x)$, where

$$\Sigma_x = \begin{bmatrix} 1 & 1 \\ 1 & 1+\frac{1}{100} \end{bmatrix},$$

Rejection at	10%	5%	1%
$\lambda = 0.5$	15.1	7.9	2.3
$\lambda = 1$	26.1	17.4	5.8
$\lambda = 1.5$	41.9	31.7	15
$\lambda = 2$	61.7	49.7	27.4
$\lambda = 2.5$	78.9	68.3	44.2
$\lambda = 3$	90.1	83.3	63.9

Table 3.1: Empirical type I-errors with distorted Wald statistics

Note: The entry in each cell is the percentage of rejections of the null hypothesis when the null hypothesis is true.

which is near singular. The result shows this test suffers from severe size distortion and that the degree of distortion is sensitive to the value of λ. The simulation result shows that even in an environment where the true variance of noise σ^2 is known, the bias term can distort the test result significantly. One of the main objectives of the robust Wald test procedure is to overcome the distortion caused by a biased estimator, which it achieves by debiasing the estimator in favor of the null hypothesis.

3.3 Robust Wald Test Procedure

3.3.1 Notation and Settings

In this section, I describe the robust Wald test procedure, which is the main result of this study. Here, I consider a general linear model in conjunction

with the following hypothesis:

$$y_i = x_i^T \beta + \varepsilon_i \qquad \varepsilon_i \sim (0, \sigma^2)$$

and

$$\begin{aligned} H_0 &: R\beta = c \\ H_A &: R\beta \neq c. \end{aligned}$$

Consistent with the previous section, I denote $y = (y_1, \ldots, y_n)^T$, $X = (x_1, \ldots, x_n)^T$, $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^T$, and $\hat{\Sigma} = X^T X$. For ease of presentation, define the following additional notation.

Notation 3.1. *For a symmetric positive semi-definite matrix A and a positive real number γ, define the following:*

i) $A^{-\gamma}$ *is the γ-inverse of A, defined as follows:*

$$A^{-\gamma} = Q\Lambda^{-\gamma}Q^T,$$

where $Q\Lambda Q^T$ is the eigen-decomposition of A and $\Lambda^{-\gamma}$ is defined by

$$\left(\Lambda^{-\gamma}\right)_{ij} = \begin{cases} \frac{1}{\lambda_{ij}} & \text{if} \quad \lambda_{ij} > \gamma \\ 0 & o.w. \end{cases},$$

ii) Denote the column space and the kernel space of A by $\underline{S}(A)$ and $S^{\perp}(A)$, respectively, and their dimensions by $\underline{k}$ and $k_{\perp}$, respectively.

iii) $W(a|\gamma) = (Ra - c)^T \left(R\hat{\Sigma}^{-\gamma}R^T\right)^{-1} (Ra - c)\,.$

iv) $\beta^{-\gamma} = \hat{\Sigma}^{-\gamma}X^T y.$

The γ-inverse is a generalization of the Moore-Penrose inverse; if $\gamma = 0$, the γ-inverse is equal to the Moore-Penrose inverse. It is well defined for all symmetric positive semi-definite matrices and is the essential element of the robust Wald test statistic. For the definition of the new test procedure, I assume the regular hypothesis.

Assumption 3.1 (Regular Hypothesis)**.**

i) R has full row rank.

ii) $R\hat{\Sigma}^{-\gamma}R^T$ is non-singular.

The second assumption imposes an upper bound on the admissible value of γ because as γ increases, the kernel space of $\hat{\Sigma}^{-\gamma}$ expands. In an extreme case, if $\gamma \to \infty$, $R\hat{\Sigma}^{-\gamma}R^T$ collapses to a zero operator. Under the regular hypothesis assumption, the robust Wald test is defined.

3.3.2 Robust Wald Test

The main result of this study, the robust Wald test procedure, is defined as follows:

Definition 3.1 (Robust Wald Test)**.** *Assume γ satisfies the regular hypothesis assumption. Then, the robust Wald statistic $\underline{W}(\gamma)$ is defined as follows:*

$$\underline{W}(\gamma) = \frac{1}{\underline{\sigma}^2} \min_{a \in \beta^{-\gamma} + S_\perp(\hat{\Sigma}^{-\gamma})} W(a|\gamma),$$

where $\underline{\sigma}^2 = \frac{1}{n-\underline{k}}(y - X\hat{a})^T(y - X\hat{a})$ for

$$\hat{a} \in \operatorname*{argmin}_{a \in \beta^{-\gamma} + S_\perp(\hat{\Sigma}^{-\gamma})} W(a).$$

The critical region C_α, where the type-I error is given by α, is

$$C_\alpha = \{w \in \mathbb{R} | \chi^{-1}_{r_{H_0}}(1-\alpha) < w\},$$

where r_{H_0} is the rank of $R\hat{\Sigma}^{-\gamma}R^T$ and $\chi^{-1}_r(\cdot)$ is the inverse cumulative distribution function of a chi-squared random variable with r degrees of freedom.

The robust Wald test procedure has two properties: attainability and conservativeness. The following theorem shows that the robust Wald statistic is attainable as long as the regular hypothesis is satisfied.

Theorem 3.1 (Attainability)**.** *The robust Wald test statistic $\underline{W}(\gamma)$ is well defined under the regular hypothesis assumption.*

The second robust property, conservativeness, guarantees that the type-I error of the robust Wald test is smaller than α whenever the threshold value γ is properly chosen value. If the following assumption is satisfied, then γ is the porperly choosen value.

Assumption 3.2 (Proper Threshold)**.** *Suppose that as the number of observations, n, increases, the number of variables, k_n, increases; that is, $\langle X^{(m)} \rangle$ and $\langle y^{(m)} \rangle$ are the sequence of the $n^{(m)} \times k^{(m)}$ matrix and the $n^{(m)} \times 1$ real vector, respectively. A sequence $\langle \gamma^{(m)} \rangle$ is a proper threshold if the following holds:*

$$\lim_{m \to \infty} \frac{n^{(m)} - \underline{k}^{(m)}}{n^{(m)}} = \kappa > 0,$$

where $\underline{k}^{(m)}$ is the dimension of $\underline{S}(X^{(m)T}X^{(m)})$.

This condition is not strong. For example, for a given $X^{(m)T}X^{(m)}$, let $\lambda^{(m)}$ be its vector of eigenvalues. Then, we can define $\gamma^{(m)}$ by the smallest non-zero element of $\lambda^{(m)}$ or a proper small value that satisfies regular

(%)	Slutsky Symmetry				Demand Homogeneity			
	Classical		Robust		Classical		Robust	
Rejection at	10%	5%	10%	5%	10%	5%	10%	5%
N=5	12.6	8.5	9.5	5.9	14.6	9.6	8.3	5.8
N=8	15.1	8.5	4.4	3.2	14.3	9.2	1.3	0.8
N=11	15.8	9.7	7.5	4.1	16.7	11.2	7.8	4.9
N=14	17.2	11.3	7.3	4.5	15.9	11.1	10.1	7
N=17	15.7	10.5	4.1	2.6	15.8	11.7	7.9	5.6
N=20	18.9	13.4	3.1	1.8	18.2	13.3	9.4	6.5
N=23	18.4	12.3	4.7	3.1	18	11.7	5.6	3.5

Table 3.2: Empirical Type I error

Note: The entries in each cell is the percentage of rejection of the null hypothesis where the null hypothesis is true. For the different number of N, 1,000 data sets have been generated under the null hypothesis where T is uniformly set to 31 .

hypothesis assumption. The following theorem establishes the conservativeness of the test.

Theorem 3.2 (Conservativeness)**.** *The type-I error of the robust Wald test is smaller than α.*

3.4 Numerical Examples

3.4.1 Slutsky Symmetry and Demand Homogeneity

The first numerical exmple considers the tests for Slutsky symmetry and demand homogeneity. I show by example how the robust Wald test can improve the type-I errors of the tests for Slutsky symmetry and demand homogeneity. Classical consumer theory predicts that consumer demand satisfies Slutsky symmetry and homogeneity if consumption decision is gen-

erated by preference maximization (Mas-Colell et al., 1995). In empirical studies, however, these hypotheses tend to be rejected decisively under various demand system specifications, for example, Deaton and Muellbauer (1980). Meisner (1979) and Laitinen (1978) suggest that such test results are statistical artefacts. Using numerical simulations, these studies show that the null hypotheses of Slutsky symmetry and demand heterogeneity are over-rejected as long as the variance-covariance structure of the noise is not sharply estimated.

Assume the following simplified version of an econometric model:

$$dS_j^t = \sum_{i=1}^{N} \beta_{ij} dP_i^t + \varepsilon_j^t.$$

Here, the superscript $t = 1, \ldots, T$ is the time index and the subscript $i = 1, \ldots, N$ denotes the product index. Let dS_i^t, dP_i^t, and ε_j^t be the change in the demand share, price, and random noise of product i at time t, respectively. The parameter β_{ij} is the rate of substitution $\frac{dS_j}{dP_i}$. This is a simplified version of the demand system, but it is sufficient for the purpose of illustration. Under this specification, the null hypotheses of Slutsky symmetry and demand homegeneity are $\beta_{ij} = \beta_{ji}$, for all $i, j = 1, \ldots, N$, and $\sum_{i=1}^{N} \beta_{ij} = 0$, for all $j = 1, \ldots, N$, respectively. I generate 1,000 copies of random data sets under the null hypothesis[5]. Table 3.2 summarizes the results.

Table 3.2 shows several trends. First, the empirical type-I error of the robust Wald test is smaller than the prespecified type-I error, except for one case: demand heterogeneity with $N = 14$ at a 10% type-I error. On the other hand, the classical Wald test overly rejects the null hypothesis. Second, the distortion worsens for the classical test as the number of products N

[5]For the detail, see appendix

(%)	ZZ		JM		RW	
	Type-I	Type-II	Type-I	Type-II	Type-I	Type-II
K = 34	42.4	26.4	31	32	5.8	81
K = 38	46.6	55.6	39.4	65	5	67.8
K = 42	50.2	72.8	67.6	17	8.2	72.2
K = 46	32.4	30.4	39.8	20.2	8.2	38.2
K = 50	88.8	1.4	73.2	10.4	13.8	27.4
K = 54	50.8	14.6	54.6	7.2	5.8	46.6
K = 58	77.4	0.4	97.6	0.2	3.2	43.8

Table 3.3: Empirical error by test

Note: ZZ, JM, and RW represent the Zhang and Zhang (2014), Javanmard and Montanari (2014), and robust Wald test respectively. The number of observations N is fixed to 50. The empirical test error is calculated using 500 simulations. The critical region is set to have a 10 percent type-I error.

increases. However, the empirical type-I errors of the robust Wald test do not seem to suffer from this issue. These observations support the view that the robust Wald test provides the more conservative test criteria.

3.4.2 Comparison with the Debiased Estimators

The second example compares the type-I and type-II errors of the robust Wald test to those of the Wald test based on the two recently suggested debiased regularized estimators: Zhang and Zhang (2014) and Javanmard and Montanari (2014). for the simulation, the following model is assumed:

$$y_i = x_i^T \beta + \varepsilon_i,$$

where x_i is a $K \times 1$ vector, and $i = 1, \ldots, N$. Table 3.3 summarizes the results. Both Zhang and Zhang (2014) and Javanmard and Montanari (2014)

over-reject the null hypothesis, and the type-I error gets worse as the number of covariates increases. On the other hand, the robust Wald test achieves uniformly conservative performance, even when $N < K$. A fundamental reason for this result is that regularized estimators are built on the sparsity assumption, while the model I employ for the simulation is not sparse. This shows that the sparsity assumption can distort the result of hypothesis testing.

3.5 Conclusion

In this study, I proposed a new test framework that is easily applicable when a model is ill-posed or ill-conditioned. The robust Wald-test procedure proposed here is similar to that of shrinkage estimation: cut off the unstable or seemingly redundant information to avoid the potential risk of overfitting at the cost of biasedness. The idea is general, but the topics covered here are limited to the general linear model. There are several possible directions for future research. First, the classical Wald test is applicable to nonlinear hypotheses and nonlinear models. Thus, the robust Wald test framework can be extended to nonlinear cases. Second, the optimal selection of the threshold value γ would enrich the class of models that can be tested using the robust Wald-test.

Appendix

3.A Proofs

3.A.1 Proof of theorem 3.1

Under the regular hypothesis assumption, $R\hat{\Sigma}^{-\gamma}R^T$ is invertible. Denote $\left(R\hat{\Sigma}^{-\gamma}R^T\right)^{-1}$ by M. Define $E \in \mathbb{R}$ as follow:

$$E = \left\{ \left(R(\beta^{-\gamma} + a) - c\right)^T M \left(R(\beta^{-\gamma} + a) - c\right) \in \mathbb{R} \,\middle|\, a \in S_\perp(\hat{\Sigma}^{-\gamma}) \right\}$$

If E has a minimum, then $\underline{W}(\gamma)$ is attainable. Let $Q_\perp$ be the matrix which column vectors are the orthogonal basis of $S_\perp(\hat{\Sigma}^{-\gamma})$, and $k_\perp$ be the dimension of $S_\perp(\hat{\Sigma}^{-\gamma})$. For all $a \in S_\perp(\hat{\Sigma}^{-\gamma})$ there exists $b \in \mathbb{R}^{k_\perp}$ such that $a = Q_\perp b$. Thus, $a \in S_\perp(\hat{\Sigma}^{-\gamma})$ can be substituted by $Q_\perp$, and if following quadratic form is a convex function of $b \in \mathbb{R}^{k_\perp}$, then E has minimum.

$$\left(R(\beta^{-\gamma} + Q_\perp b) - c\right)^T M \left(R(\beta^{-\gamma} + Q_\perp b) - c\right) \tag{3.1}$$

By direct calculation, (3.1) is

$$\left(R\beta^{-\gamma} - c\right)^T M \left(R\beta^{-\gamma} - c\right) + 2\left(R\beta^{-\gamma} - c\right)^T MRQ_\perp b + b^T Q_\perp^T R^T MRQ_\perp b.$$

Observe the $Q_\perp^T R^T MRQ_\perp$ is a positive semi-definite. i.e. (3.1) is a convex function of $b \in \mathbb{R}^{k_\perp}$.

3.A.2 Proof of theorem 3.2

Lemma 3.1. *Write the rank of $R\hat{\Sigma}^{-\gamma}R^T$ by r_{H_0}. Define*

$$W(a) = \left(R(\beta^{-\gamma} + a) - c\right)^T \left(R\hat{\Sigma}^{-\gamma}R^T\sigma^2\right)^{-1} \left(R(\beta^{-\gamma} + a) - c\right).$$

Then $W(a) \sim \chi^2_{r_{H_0}}$ for some $a \in S_\perp(\hat{\Sigma}^{-\gamma})$ under the null hypothesis.

Proof of the lemma. First, see

$$\begin{aligned}
\beta^{-\gamma} &= \hat{\Sigma}^{-\gamma} X^T y && (3.2)\\
&= \hat{\Sigma}^{-\gamma} X^T (X\beta + \varepsilon) && (3.3)\\
&= \underline{\beta} + \hat{\Sigma}^{-\gamma} X^T \varepsilon && (3.4)
\end{aligned}$$

where $\underline{\beta}$ is the projection of β onto the $\underline{S}(\hat{\Sigma}^{-\gamma})$. It implies that there exists a bias corrector a in $S_\perp(\hat{\Sigma}^{-\gamma})$ such that

$$\beta^{-\gamma} + a \sim N(\beta, \hat{\Sigma}^{-\gamma}).$$

Thus,

$$R\left(\beta^{-\gamma} + a\right) - c \sim N(\beta, R\hat{\Sigma}^{-\gamma}R^T)$$

and it proves the lemma. □

Lemma 3.2. *Under the proper threshold assumption, $\sigma^2 \leq \underline{\sigma}^2(\hat{a})$ with probability 1 asymptotically.*

Proof of the lemma. For some $b \in S_\perp(\hat{\Sigma}^{-\gamma})$,

$$\begin{aligned}
\underline{\sigma}^2(\hat{a}) &= \frac{1}{n-\underline{k}} \left(y - X(\beta^{-\gamma} + b)\right)^T \left(y - X(\beta^{-\gamma} + b)\right)\\
&= \frac{1}{n-\underline{k}} \Big(X(\beta - \underline{\beta} - b) + \varepsilon - X\hat{\Sigma}^{-\gamma}X^T\varepsilon)\Big)^T \Big(X(\beta - \underline{\beta} + b) + \varepsilon - X\hat{\Sigma}^{-\gamma}X^T\varepsilon)\Big)\\
&= \frac{1}{n-\underline{k}} \big((\beta - \underline{\beta} - b)^T X^T X(\beta - \underline{\beta} - b)\\
&\qquad + 2(\beta - \underline{\beta} - b)^T X^T (I - X\hat{\Sigma}^{-\gamma}X^T)\varepsilon + \varepsilon^T (I - X\hat{\Sigma}^{-\gamma}X^T)\varepsilon\Big)
\end{aligned}$$

where $\underline{\beta}$ is the projection of β onto the $\underline{S}(\hat{\Sigma}^{-\gamma})$. As n goes to infinity, the first term conveerge to $\frac{1}{\kappa}(\beta - \underline{\beta} - b)^T E[x_i x_i^T](\beta - \underline{\beta} - b)$, second term converge to zero and the third term converge to σ^2. Since $E[x_i x_i^T]$ is a positive semi-definite matrix, $\sigma^2 \leq \underline{\sigma}^2(\hat{a})$. □

Proof of the theorem. What I want to show is for all $w > 0$, following is asymptotically true.

$$\Pr[W < w|H_0] \leq \Pr[\underline{W}(\gamma) < w|H_0]$$

where W is a chi-square random variable with degree of freedom r_{H_o}. Observe that

$$\frac{\underline{\sigma}^2}{\sigma^2}\underline{W}(\gamma) \leq W(a)$$

for all $a \in S_\perp(\hat{\Sigma}^{-\gamma})$ and by the lemma 3.1, $W(b) \sim \chi^2_{r_{H_0}}$ for some $b \in S_\perp(\hat{\Sigma}^{-\gamma})$. The lemma 3.2 shows $1 < \frac{\underline{\sigma}^2}{\sigma^2}$ asymptotically. Thus,

$$\underline{W}(\gamma) \quad \leq \quad \frac{\underline{\sigma}^2}{\sigma^2}\underline{W}(\gamma) \quad \leq \quad W(b) \quad \sim \quad \chi^2_{r_{H_0}}$$

□

3.B Details of Numerical Examples

3.B.1 Slutsky Symmetry and Demand Homogeneity

Generation of β_{ij}

i) Generate $\frac{N(N-1)}{2}$ number of independent Irwin–Hall random number with $n = 2$, that is the sum two uniform random variable with support $[0, 1]$, and assigned each of them to β_{ij} for $i < j$.

ii) For $i > j$, assign $\beta_{ij} = \beta_{ji}$.

iii) For $i = j$, put $\beta_{ij} = -\sum_{l=1}^{N} \beta_{lj}$.

Then, β_{ij} satisfies the Slutsky symmetry and the demand homogeneity.

Generation of dX_i^t

For each $t = 1, \ldots, T$, let dX^t be a $N \times 1$ column vector which i the element is dX_i^t. dX^t are generated by $N(0, \Sigma)$ where

$$(\Sigma)_{mn} = \frac{\mu}{\mu + |m - n|}$$

for some $\mu > 0$. In the simulation I set μ to 20.

Generation of ε_i^t

For $i = 1, \ldots, N$ and $i = 1, \ldots, N$, I_i and T_t are generated by uniform distribution with support $[0, \sqrt{3}]$. Each of them prepresent the random individual effect and the random time effect. For $\hat{\varepsilon}_i^t \sim N(0, 1)$, each ε_i^t is defined as follow:

$$\varepsilon_i^t = I_i T_t \hat{\varepsilon}_i^t.$$

3.B.2 Comparison to the Debiased Estimators

Generation of β_i

Set $\beta_i = 2\frac{i}{K}$ for $i \leq \frac{K}{2}$. I set $\beta_i = \beta_{i-\frac{K}{2}} - 1$ and $\beta_i = \beta_{i-\frac{K}{2}} - 1 - \frac{1}{\sqrt{50}}$ for $i > \frac{K}{2}$ for the null hypothesis and alternative hypothesis respectively.

Generation of x_i

For each $i = 1, \ldots, N$, let x_i is a $K \times 1$ column vector distributed by by $N(0, \Sigma)$ where

$$(\Sigma)_{mn} = \frac{\mu}{\mu + |m - n|}$$

for some $\mu > 0$. In the simulation I set μ to 30 to make the design matrix highly correlated.

Generation of ε_i

Let I_i and T_i be the random variables independently and identically distributed by uniform distribution with support $[0, \sqrt{3}]$ for $i = 1, \ldots, N$. For a $\hat{\varepsilon}_i \sim N(0, 1)$, the noise term ε_i is defined as follow:

$$\varepsilon_i = I_i T_i \hat{\varepsilon}_i.$$

Bibliography

Anatolyev, S. (2012). Inference in regression models with many regressors. *Journal of Econometrics 170*(2), 368–382.

Barber, B. M. and T. Odean (2007). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *The review of financial studies 21*(2), 785–818.

Belloni, A., V. Chernozhukov, and C. Hansen (2014). Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies 81*(2), 608–650.

Bühlmann, P. et al. (2013). Statistical significance in high-dimensional linear models. *Bernoulli 19*(4), 1212–1242.

Chiang, J., S. Chib, and C. Narasimhan (1998). Markov chain monte carlo and models of consideration set and parameter heterogeneity. *Journal of Econometrics 89*(1), 223–248.

Da, Z., J. Engelberg, and P. Gao (2011). In search of attention. *The Journal of Finance 66*(5), 1461–1499.

Deaton, A. and J. Muellbauer (1980). An almost ideal demand system. *The American economic review 70*(3), 312–326.

Gandhi, A. and R. Serrano-Padial (2014). Does belief heterogeneity explain asset prices: the case of the longshot bias. *The Review of Economic Studies*, rdu017.

Gillen, B., S. Montero, H. R. Moon, and M. Shum (2015). Blp-lasso for aggregate discrete-choice models applied to elections with rich demographic covariates. Technical report, Working Paper. California Institute of Technology.

Goeree, M. S. (2008). Limited information and advertising in the us personal computer industry. *Econometrica 76*(5), 1017–1074.

Hoerl, A. E. and R. W. Kennard (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics 12*(1), 55–67.

Javanmard, A. and A. Montanari (2014). Confidence intervals and hypothesis testing for high-dimensional regression. *The Journal of Machine Learning Research 15*(1), 2869–2909.

Jullien, B. and B. Salanié (2000). Estimating preferences under risk: The case of racetrack bettors. *Journal of Political Economy 108*(3), 503–530.

Kacperczyk, M., S. Van Nieuwerburgh, and L. Veldkamp (2016). A rational theory of mutual funds' attention allocation. *Econometrica 84*(2), 571–626.

Laitinen, K. (1978). Why is demand homogeneity so often rejected? *Economics Letters 1*(3), 187–191.

Laurie, D. P. (1985). Practical error estimation in numerical integration. *Journal of Computational and Applied Mathematics 12*, 425–431.

Manzini, P. and M. Mariotti (2014). Stochastic choice and consideration sets. *Econometrica 82*(3), 1153–1176.

Mas-Colell, A., M. D. Whinston, J. R. Green, et al. (1995). *Microeconomic theory*, Volume 1. Oxford university press New York.

Meisner, J. F. (1979). The sad fate of the asymptotic slutsky symmetry test for large systems. *Economics Letters 2*(3), 231–233.

Merton, R. C. (1987). A simple model of capital market equilibrium with incomplete information. *The journal of finance 42*(3), 483–510.

Minnier, J., L. Tian, and T. Cai (2011). A perturbation method for inference on regularized regression estimates. *Journal of the American Statistical Association 106*(496), 1371–1382.

Odean, T. (1999). Do investors trade too much? *American economic review 89*(5), 1279–1298.

Sims, C. A. (2006). Rational inattention: Beyond the linear-quadratic case. *American Economic Review 96*(2), 158–163.

Snowberg, E. and J. Wolfers (2010). Explaining the favorite–long shot bias: Is it risk-love or misperceptions? *Journal of Political Economy 118*(4), 723–746.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 267–288.

Train, K. E. (2009). *Discrete choice methods with simulation.* Cambridge university press.

Van Nieuwerburgh, S. and L. Veldkamp (2010). Information acquisition and under-diversification. *The Review of Economic Studies 77*(2), 779–805.

Zhang, C.-H. and S. S. Zhang (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 76*(1), 217–242.