Title
Essays on Identification

Author
Burkovskaya, Anastasia

Publication Date
2016

University of California
Los Angeles

Essays on Identification

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Economics

by

Anastasia Burkovskaya

2016

ABSTRACT OF THE DISSERTATION

Essays on Identification

by

Anastasia Burkovskaya
Doctor of Philosophy in Economics
University of California, Los Angeles, 2016
Professor Rosa Liliana Matzkin, Co-chair
Professor William R. Zame, Co-chair

Chapter 1 considers agents that aggregate several states of the world (i.e., the possible different realizations of an uncertain situation) into an event. They then compare bundles of consumption on these aggregated events, instead of having to consider bundles on each of the single possible realization. This simplifies the decision-making, since it decreases the number of distinguished realizations to consider. The collection of aggregated events of this type is called a subjective partition. In the case of Subjective Expected Utility, this kind of decision process does not affect agents' choices. However, the presence of ambiguity (i.e., when agents do not know the probability associated with each possible realization of uncertain situations) changes their behavior dramatically: different subjective partitions lead to different preferences over bundles. Chapter 1 deals with preferences: we provide axioms for a state aggregation model with ambiguity and show under which conditions preferences imply uniqueness of the subjective partition.

Chapter 2 discusses choices obtained from the state aggregation model described in Chapter 1. We show identification of the subjective partition from choices in a complete market setting. In addition, we demonstrate that the results can be applied to deductible insurance choices. Finally, we offer a set of

revealed preference inequalities that allow for testing the model on a dataset.

Chapter 3 introduces a model of electoral choice that allows for derivation of joint distribution of turnout and voter share from unobservable joint distribution of costs of voting and preferences. Under a set of mild assumptions, we show identification and provide non-parametric estimators of joint distribution of costs of voting and preferences over candidates from observable electoral data. All estimators are consistent and asymptotically normal.

The dissertation of Anastasia Burkovskaya is approved.

Jay Yin Jie Lu

Judea Pearl

William R. Zame, Committee Co-chair

Rosa Liliana Matzkin, Committee Co-chair

University of California, Los Angeles

2016

To my father.

Table of Contents

List of Figures

Acknowledgments

To my committee, thank you for the invaluable advice and discussions. I am deeply indebted to the co-chairs Rosa Matzkin and Bill Zame for all the support. Without their guidance, efforts and incredible patience, my academic career would not be possible.

I would also like to thank audiences at UCLA, ITAM, 2015 Latin American Workshop on Economic Theory, USydney and Midwest Economic Theory conference for helpful comments.

There is no way to overestimate amount of support and blind faith that I had from Josephine Wallace. My pets and I would literally live on the street without her kindness.

And, finally, I am grateful to Stefano Fiorin for all the struggle I put him through during the job market, and, of course, for reading my paper 100 times. I was a terrible human being and I cannot say how much I appreciate his patience and care.

Vita

2005–2010	Specialist in Computer Science, Moscow State University.
2009–2011	M.A. in Economics, New Economic School.
2011–2012	M.A. in Economics, UCLA.
2012–2016	Teaching Fellow, Department of Economics, UCLA.

Publications and Presentations

Monetary Political Business Cycles: New Democracy Setting, *Quantile*, No.11, 2013, pp. 75-90.

Midwest Economic Theory Conference, University of Rochester, 2016: *Identification of Subjective Partition of State Space under Ambiguity.*

Job market seminar, University of Sydney, 2016: *Identification of Subjective Partition of State Space under Ambiguity.*

2015 Latin American Workshop in Economic Theory, San Luis Potosi, Mexico: *Identification of Subjective Partition of State Space under Ambiguity.*

The 10th Annual Economic Graduate Student Conference, WUSTL, 2015: *Identification of Subjective Partition of State Space under Ambiguity.*

The ITAM CIE brown bag seminar, ITAM, 2015: *Identification of Information Structure and Priors under Ambiguity with Intermediate Information*.

The 29th NES Research Conference, Moscow, Russia, 2011: *Opportunistic Political Business Cycles in Monetary Policy and Central Bank Independence*.

CHAPTER 1

Subjective Partition and Its Uniqueness under Ambiguity

Economists usually assume that people have perfect computational abilities and a good understanding of how the world works. However, in reality, there are many states of the world (i.e., many possible realizations of uncertain situations), and optimizing over all of them together might be difficult even for a machine. The behavioral literature provides good insights into different ways that people might simplify decision-making. In our model, agents split one problem into several problems with fewer variables to optimize. An agent combines several states into an event, evaluates an act at the event, and then assesses the final value by treating events as aggregated states. We call a collection of these events a subjective partition. Note that the agent perfectly realizes the existence of each state of the world, but prefers to bundle states together. State aggregation does not affect choices when agents know the probabilities of different realizations of uncertainty that they face – i.e., when agents' behavior can be described by Subjective Expected Utility (SEU). However, a vast literature (Ellsberg (1961), Kahneman and Tversky (1979), Halevy (2007), etc.) demonstrates that in the presence of ambiguity - i.e., when agents do not know the probabilities of each possible realization of uncertainty - people make choices that are not consistent with SEU. Moreover, under ambiguity, various methods of state aggregation lead to different choices.

To illustrate the idea of subjective partition, consider the following example.

Suppose that there is an agent who needs to choose car insurance. Possible states of the world include an accident with the agent at fault, an accident with the other party at fault, and no accident. An insurance policy consists of two deductibles: one applies if the agent caused the accident, and the other applies if the agent was not responsible. Probabilities of the states are unknown and described by a set of multiple priors – i.e., all agents face ambiguity. Next, suppose that the agent is a person that does not aggregate states. If she is ambiguity-averse then, generally, her choice is a bundle of deductibles such that her consumption is smoothed between all states. Note that, in this case, there is no state aggregation, and her subjective partition is the whole state space of the world. Next, imagine a person who has difficulty analyzing all three states of the world together. In order to simplify the decision process, she combines both states that involve a collision into the event "accident," while keeping the "no accident" state separate from them. In this case, the state space of the economy is partitioned by events "accident" (with the agent or the other party at fault) and "no accident." The presence of state aggregation changes the way that the agent thinks about the world. Instead of an accident with the agent at fault, an accident with the other party at fault, and no accident, her subjective states now are "accident" and "no accident." Moreover, the individual generally wants to smooth her consumption between her aggregated states. As a result, she will buy less insurance (greater deductibles) for both types of accidents, compared with what would be optimal for her.

The idea of a subjective partition is closely related to a subjective state space, which is a private understanding of outcome space due to unforeseen contingencies. Both notions are supposed to simplify decision-making in one or another way; however, the relationship between them is even closer than that. If we restrict the whole act space to partition-measurable acts only, then our subjective partition can be interpreted as a subjective state space similar to that in Ahn and Ergin

(2010), where agents aggregate into an event unforeseen contingencies that give the same payoff. In the above example, an insurance policy with the same deductible for any accident type is an example of a partition-measurable act. In contrast to our work, Ahn and Ergin (2010) assume that even though the agent is aware of the existence of unforeseen contingencies, she does not recognize the existence of different kinds of accidents. However, the agent still, generally, smooths her consumption between the events "accident" and "no accident," as in our model. Thus, an implication about lower demand for insurance will hold for the subjective state space with the restriction to partition-measurable acts, too.

In most economic situations, researchers treat the decision-making process like a black box, and all differences in behavior are usually explained by differences in preferences or information. This paper allows for situations in which agents with the same information and preferences might make different choices because of heterogeneity in the subjective partition. Moreover, such behavior cannot be obtained by providing the agent with some kind of incomplete information, which is present in the model in a form of the sets of priors over states. As a result, understanding the subjective partition is critical for researchers and policy-makers. For example, in the discussed model, agents with various subjective partitions will react differently to information provided by policy-makers. Thus, in order to achieve the desired improvements in welfare, we need to understand what kind of information to release to agents depending on their subjective partition.

In this Chapter, we (1) offer a model of state aggregation and axioms that provide the representation of the preferences; and, (2) show that under full-dimensionality of priors and strict α-MEU ambiguity model, preferences imply a unique subjective partition.

First, we provide axioms and the representation of the preferences with state aggregation. We assume that only preferences over acts are observed. Thus, our primitive is preferences over acts under some subjective partition from which we derive conditional preferences. We define conditional preferences only for events that we call "aggregating." Such events satisfy a property similar to Savage's Sure Thing Principle. After that, if there is a partition of the whole state space that consists of aggregating events only, then we assume the α-MEU model of Ghirardato, Maccheroni, and Marinacci (2004) for conditional preferences and preferences over observed acts, which we call ex-ante preferences.

Second, we find that if all sets of priors are full-dimensional, then preferences with a strict (i.e., $0 < \alpha < 1$) α-MEU stage model imply a unique way of state aggregation. This means that different state spaces generate distinct preference relations. We also show that in the case of the MEU stage model, full identification is possible only when the agent does not aggregate states and has non-rectangular priors. When priors satisfy rectangularity, a property that provides dynamic consistency of MEU (Epstein and Schneider (2003)), we are able to achieve only partial identification: it is not possible to distinguish between the whole state space with rectangular priors and the aggregation into a partition, for which rectangularity of priors holds.

This paper contributes to the literature on subjective state spaces. The idea of an agent with a coarse understanding of the state space was introduced by Kreps (1992) and followed by Dekel, Lipman, and Rustichini (2001), Epstein, Marinacci, and Seo (2007), Ghirardato (2001), and Mukerji (1997). The closest work in this field to our paper is Ahn and Ergin (2010). The authors characterize partition-dependent expected utility on partition-measurable acts. They interpret the framing of contingencies into a partition that represents a subjective state space, where unforeseen contingencies have the same payoff and are grouped together into an event. We also group contingencies together; however, we do not require partition-measurability of acts. In addition, in order to obtain identification results, the authors use non-additive probability. In contrast, we obtain similar features of behavior by assuming ambiguity. Thus, if we restrict the act space to partition-measurable acts, our subjective partition can be interpreted as a subjective state space under ambiguity. Our paper also contributes to the literature on decision heuristics or rules of thumb. Pioneering work in the field was introduced by Simon (1955) and followed by a vast literature afterwards (e.g., Akerlof and Yellen (1985), Haltiwanger and Waldman (1985), Ellison and Fudenberg (1993)), as well as one of the most widely known papers on prospect theory by Kahneman and Tversky (1979). Recently, related research has evolved in the direction of discovery of the decision rule (e.g., El-Gamal and Grether (1995) and Houser, Keane, and McCabe (2004)).

In addition, our work is related to the literature on dynamic consistency. The relationship between non-expected utility and problems with dynamic consistency is well known (Machina (1989), Karni and Schmeidler (1991), Epstein and LeBreton (1993), Ghirardato (2002), etc.). The literature has developed mainly in two directions: dynamically consistent models (e.g., Klibanoff, Marinacci, and Mukerji (2009)) and the extension of existing ambiguity models to a dynamically consistent framework (Epstein and Schneider (2003), Wang (2003), Hayashi (2005), etc.). Our results are closely related to those of Epstein and Schneider (2003), who find that the rectangularity of priors is an important condition for dynamic consistency of the MEU model. However, we find that a lack of dynamic consistency often allows for identification of the unique subjective partition. That is why we can uniquely identify it under the MEU stage model only when the set of priors is not rectangular. In addition, a strict α-MEU model with full dimensional priors is generally dynamically inconsistent and, as a result, provides full identification. We expect similar results to hold for other ambiguity stage models.

Finally, the notion of a subjective partition is very close to intermediate information. Li (2011) introduced a concept of intermediate information – information that arrives after a choice has been made and before an outcome is realized. The author axiomatizes preference relations over pairs of acts and intermediate information together. In order to obtain such preference relations over pairs, Li derives preference relations between acts under some specific information by using conditional preferences and basic, "no information" preferences as primitives. A subjective partition can be interpreted as fixed intermediate information. However, it requires different axiomatization because observable primitives are different. Together with the Li's work, our paper is related to the literature on temporal resolution of uncertainty. In the pioneering paper in this field, Kreps and Porteus (1978) axiomatize preferences over timing when risk is resolved. Later, their work was extended and applied to asset pricing and macroeconomics (Epstein and Zin (1989, 1991), Tallarini (2000), etc.). Grant, Kajii and Polak (1998, 2000) reject the reduction of compound lotteries in the Kreps-Porteus framework and connect intrinsic preferences for information with risk aversion. However, Halevy (2007) demonstrates empirical evidence of the relationship between non-reduction of compound lotteries and ambiguity aversion. In the model we characterize, the reduction also does not hold, and the presence of ambiguity creates preferences over intermediate information similarly to previous works.

1.1 The Model and Notation

Suppose that X is a convex subset of consequences in $\mathbb{R}$, and Ω is a finite set of states of the world with an algebra Σ of subsets of Ω. We denote $\mathcal{F}$ a set of all acts, Σ-measurable finite step functions: $\Omega \to \Delta X$. Let $\mathcal{M}$ be a set of elements π, such that $\pi \subset \Sigma$ is a partition of Ω. Partitions of the state space represent different ways of state aggregation.

1.1.1 α-MEU preferences

As stage preferences, we assume the α-MEU model of Ghirardato, Maccheroni and Marinacci (2004). If $\alpha \in [0, 1]$ is a parameter of ambiguity aversion, $u(\cdot)$ is an increasing continuous utility function, $\mathcal{P}$ is a set of priors, and $x \in \mathcal{F}$ is an act, then the value of x is assessed by

$$V(x) = \alpha \min_{P \in \mathcal{P}} \left[\sum_{s \in \Omega} u(x(s)) P(s) \right] + (1-\alpha) \max_{P \in \mathcal{P}} \left[\sum_{s \in \Omega} u(x(s)) P(s) \right]$$

If we denote $P_{\min x}(s)$ and $P_{\max x}(s)$ to be probabilities that minimize and maximize "expected" utility, then

$$V(x) = \sum_{s \in \Omega} u(x(s)) \left(\alpha P_{\min x}(s) + (1-\alpha) P_{\max x}(s) \right) = \sum_{s \in \Omega} u(x(s)) P_{\alpha;x}(s).$$

1.1.2 Costly computation

Every time the agent needs to make a choice, she has to compare acts. For example, when the agent buys a car, she needs to choose an insurance plan, and in order to discover her preferences among them, she will spend some time comparing all available plans. Discovering preferences is costly for the agent and depends on how she aggregates states into some partition π of the state space Ω. For instance, it might be easier for the agent to compare plans based on the event "accident" instead of separating it into states dependent on who is at fault. The agent is aware that she will have to solve a similar problem again in the future for n periods (e.g., renew her insurance plan every year). However, if the agent has already established her preferences under some subjective partition before, she does not need to do it again for the same subjective partition in the future. The agent will carry costs in future periods only if she decides to update her preferences on another subjective partition.

Thus, suppose that there is a discovery cost function $\hat{c}(\cdot) : \mathcal{M} \to \mathbb{R}$. We define the computation cost function at every time period t and partition $\pi \in \mathcal{M}$:

$$c_t(\pi) = \begin{cases} \hat{c}(\pi), & \text{if } \pi \text{ has not been chosen before time } t \\ 0, & \text{if } \pi \text{ has been chosen before time } t \end{cases}$$

We denote one period value of act f given subjective partition π as $V(f|\pi)$ and a time t value functional of the agent as $I_t(\pi) = \max_f V(f|\pi) + \beta \max_\pi (I_{t+1}(\pi) - c_{t+1}(\pi))$, where β is a discount rate. Then, at moment t, the agent solves the following problem:

$$I_t(\pi) - c_t(\pi) \to \max_\pi .$$

For the purposes of this paper, we assume that discovery costs $\hat{c}(\cdot)$ are so high that the agent prefers to choose a subjective partition once and uses it for all future periods. For example, when the agent needs to renew her insurance policy, she spends less time choosing the plan because she uses her "old," already discovered preferences. We also assume that in everything that follows, choices in all periods are mutually independent given the subjective partition; thus, each period can be analyzed as a static case.

1.1.3 Subjective partition

A stage in which several states are combined into one event will be called the conditional stage, and the set of priors $\mathcal{P}(A)$ used for aggregation into event A is the set of conditional priors. We will also call the ex-ante stage (or problem) a stage in which evaluation of an act across events happens. Given the subjective partition π, the set of priors over events $\mathcal{P}(\pi)$ will be called the set of ex-ante priors.

Definition 1.1. *The agent's behavior is said to exhibit α-MEU State Aggregation Representation (α-MEU SAR) if there exist a partition of the state space π,* *nonempty weak compact and convex sets $\mathcal{P}(A)$ and $\mathcal{P}(\pi)$ for any event $A \in \pi$ of probabilities on Σ, and a continuous monotone function $u : X \to \mathbb{R}$, $\alpha \in [0,1]$ such that the agent optimizes the functional $V(\cdot|\pi)$:*

$$V(x|A) = \alpha \min_{P(\cdot|A)\in\mathcal{P}(A)} \sum_{s\in A} u(x(s))P(s|A) + (1-\alpha) \max_{P(\cdot|A)\in\mathcal{P}(A)} \sum_{s\in A} u(x(s))P(s|A)$$

$$V(x|\pi) = \alpha \min_{P\in\mathcal{P}(\pi)} \sum_{A\in\pi} V(x|A)P(A|\pi) + (1-\alpha) \max_{P\in\mathcal{P}(\pi)} \sum_{A\in\pi} V(x|A)P(A|\pi).$$

A set of axioms for the above representation is provided in Section 6. Definition 1 implies that the agent evaluates each act given a partition by a folding-back procedure: First, she aggregates states into event A and evaluates the act x given every such event A in the partition π with the set of conditional priors $\mathcal{P}(A)$:

$$V(x|A) = \alpha \min_{P(\cdot|A)\in\mathcal{P}(A)} \left[\sum_{s\in A} u(x(s))P(s|A)\right] + (1-\alpha) \max_{P(\cdot|A)\in\mathcal{P}(A)} \left[\sum_{s\in A} u(x(s))P(s|A)\right],$$

where α is an agent's coefficient of ambiguity aversion. Second, the agent evaluates the act across events that form her subjective partition. Thus, at the ex-ante stage with the set of ex-ante priors $\mathcal{P}(\pi)$, the agent's value of the act is obtained similarly to the way it is obtained in the conditional stage:

$$V(x|\pi) = \alpha \min_{P(\cdot|\pi)\in\mathcal{P}(\pi)} \left[\sum_{A\in\pi} V(x|A)P(A|\pi)\right] + (1-\alpha) \max_{P(\cdot|\pi)\in\mathcal{P}(\pi)} \left[\sum_{A\in\pi} V(x|A)P(A|\pi)\right].$$

Example 1.1. Consider the example from the introduction: the agent chooses an insurance policy under three states of the world: an accident with the agent at fault, an accident with the other party at fault, and no accident.

Suppose that the probability of no accident (s_1) is not smaller than 50%. However, the probabilities of the accident with the agent at fault (s_2) and the accident with the other party at fault (s_3) are still unknown. Imagine two different situations: (1) the agent does not aggregate states, and her subjective partition of the world $\pi_0 = \{s_1, s_2, s_3\}$ is the whole outcome space; and (2) the agent has difficulty optimizing over three states, and she splits the whole outcome space into events "no accident" $A_1 = \{s_1\}$ and "accident" $A_2 = \{s_2, s_3\}$. We denote the subjective partition in this case as $\pi = \{A_1, A_2\}$.

In situation (1), the ex-ante stage implies regular evaluation over the whole state space. Then, the set of ex-ante priors is $\mathcal{P}(\pi_0) = \{(p_1, p_2, p_3) \in \mathbb{R}^3 : p_1 \geq 0.5; \sum p_i = 1\}$. All conditional stages are degenerate in this case because each event consists of exactly one state.

In situation (2), the agent's subjective partition π consists of two events, A_1 and A_2. In addition, consider that the agent updates prior-by-prior using Bayes' rule. Thus, the set of ex-ante priors is $\mathcal{P}(\pi) = \{(p_{A_1}, p_{A_2}) \in \mathbb{R}^2 : p_{A_1} \geq 0.5; p_{A_1} + p_{A_2} = 1\}$. In the case of no accident, the set of conditional on event A_1 priors is degenerate: $\mathcal{P}(A_1) = \{p_1 \in \mathbb{R} : p_1 = 1\}$. If the event is "accident," then the set of conditional on event A_2 priors is $\mathcal{P}(A_2) = \{(p_2, p_3) \in \mathbb{R}^2 : p_2 + p_3 = 1\}$.

Now, for simplicity, consider utility function $u(x) = x$ and a parameter of ambiguity aversion $0 < \alpha < 1$. First, imagine the agent without state aggregation. Then, her value of act $x = (x_1, x_2, x_3)$ that represents the amount of money/consumption in every state is

$$V(x|\pi_0) = \frac{1}{2}x_1 + \frac{\alpha}{2}\min_i(x_i) + \frac{1-\alpha}{2}\max_i(x_i).$$

In case (2), the value of act x at event A_2 is

$$V(x|A_2) = \alpha \min(x_2, x_3) + (1-\alpha)\max(x_2, x_3),$$

while the value of act x under subjective partition π is

$$V(x|\pi) = \frac{1}{2}x_1 + \frac{\alpha}{2}\min(x_1, V(x|A_2)) + \frac{1-\alpha}{2}\max(x_1, V(x|A_2)).$$

As long as $0 < \alpha < 1$, $V(x|\pi_0) \neq V(x|\pi)$. To easily see this, consider, for example, x such that $x_1 > x_2 > x_3$. Then, $V(x|\pi_0) = \frac{2-\alpha}{2}x_1 + \frac{\alpha}{2}x_3$, while $V(x|\pi) = \frac{2-\alpha}{2}x_1 + \frac{\alpha(1-\alpha)}{2}x_2 + \frac{\alpha^2}{2}x_3$. Thus, with the subjective partition π, preferences differ from the optimal. Notice that if $\alpha = 0, 1$, then $V(x|\pi_0) = V(x|\pi)$.

1.2 Axiomatization

1.2.1 Preliminaries

We denote for all $f, g \in \mathcal{F}, A \in \Sigma$, fAg an act: $fAg(s) = f(s)$ if $s \in A$, and $fAg(s) = g(s)$ if $s \notin A$. The mixture of acts is defined statewise. We will abuse notation and define X as a set of constant acts in what follows below.

1.2.2 Axioms and representation

The purpose of this section is to provide axioms of preference relation $\succeq$ between acts f over $\mathcal{F}$ that can be represented by the model of state aggregation, as described in Section 1.1. We take as a primitive a preference relation between acts. After that, we induce conditional preferences whenever we can guarantee their completeness.

First, we assume that $\succeq$ satisfies the classical axioms of Gilboa and Schmeidler (1989):

Axiom 1.1 (Invariant Biseparable Preferences - IBP)**.** *For all $f, g, h \in \mathcal{F}$ and $x \in X$: (i) $\succeq$ is complete and transitive; (ii) if $\lambda \in (0, 1]$: $f \succeq g \Leftrightarrow \lambda f + (1-\lambda)x \succeq \lambda g + (1-\lambda)x$; (iii) if $f \succ g$, and $g \succ h$, then there exist $\lambda, \mu \in (0, 1)$ such that $\lambda f + (1-\lambda)h \succ g$ and $g \succ \mu f + (1-\mu)h$; (iv) if $f(s) \succeq g(s)$ for all $s \in \Omega$, then $f \succeq g$; and (v) $\succeq$ is not degenerate.*

In order to introduce state aggregation and obtain complete conditional preferences, we define a concept of an aggregating event:

Definition 1.2. *Event A is called aggregating if for any $f, g, h, h' \in \mathcal{F}$: $fAh \succeq gAh \Leftrightarrow fAh' \succeq gAh'$.*

An aggregating event satisfies a property similar to Savage's Sure Thing Principle. The independence axiom does not hold in this model, so it is not equivalent to the existence of a unique probability. The aggregating event implies that the value of the act at the event is the only aspect that matters in act evaluation, and not each state value separately.

Denote a set of all aggregating events $\mathcal{A}$. Note that $\{\Omega\}$ trivially belongs to it, and each separate state is in there due to monotonicity. Now, for all aggregating events A, we define conditional preferences $\succeq_A$:

Axiom 1.2 (Conditional Preferences - CP)**.** *For all $f, g, h \in \mathcal{F}$: $f \succeq_A g \Leftrightarrow fAh \succeq gAh$.*

Note that conditional preferences $\succeq_A$ are complete and satisfy all analogous axioms for IBP.

In order to obtain the $\alpha-$MEU stage preferences, we follow Ghirardato, Maccheroni, and Marinacci (2004) to define unambiguous preferences $\succeq_A^*$:

Definition 1.3. *For any $A \in \mathcal{A}$ and acts $f, g, h \in \mathcal{F}$, f is unambiguously preferred to g given event A, $f \succeq_A^* g$, if $\lambda f + (1-\lambda)h \succeq_A \lambda g + (1-\lambda)h$ for all $\lambda \in (0, 1]$.*

For each possible partition π, we define a set of π-measurable acts $\mathcal{F}_\pi$. Then, the unambiguous π-measurable preference relation is $\succeq_\pi^*$:

Definition 1.4. *For any acts $f, g, h \in \mathcal{F}_\pi$, f is unambiguously preferred to g given π, $f \succeq_\pi^* g$, if $\lambda f + (1-\lambda)h \succeq \lambda g + (1-\lambda)h$ for all $\lambda \in (0, 1]$.*

We now define "possible" certainty equivalence sets for act f on event A:

$$C_A(f) = \{x \in X : \forall y \in X : \text{ if } y \succeq_A^* f, \text{ then } y \succeq_A^* x, \text{ and if } y \preceq_A^* f, \text{ then } y \preceq_A^* x\}$$

Similarly, the "possible" certainty equivalence set for π-measurable act $f \in \mathcal{F}_\pi$ is:

$$C_\pi(f) = \{x \in X : \forall y \in X : \text{ if } y \succeq_\pi^* f \text{ then } y \succeq_\pi^* x, \text{ and if } y \preceq_\pi^* f \text{ then } y \preceq_\pi^* x\}.$$

The above sets contain all constant acts with which unambiguous preferences for f are not defined. Now we adapt Axiom 7 from Ghirardato, Maccheroni, and Marinacci (2004) to fix parameters of ambiguity aversion at each stage for all acts.

Axiom 1.3 (α-MEU)**.** *For all acts $f, g, h, h' \in \mathcal{F}$, there exist a partition π of Ω such that $\pi \subset \mathcal{A}$, and aggregating events $A, B \in \pi$: (i) If $C_A(f) = C_A(g)$, then $f \sim_A g$; (ii) if $C_A(f) = C_B(g)$, then there exists $x \in X$: $f \sim_A x$ and $g \sim_B x$; and (iii) if $C_\pi(f) = C_A(g)$, then there exists $x \in X$: $f \sim x$ and $g \sim_A x$.*

Theorem 1.1. *(Representation Theorem) A binary relation $\succeq$ satisfies axioms 1-3 if and only if there exist nonempty weak compact and convex sets $\mathcal{P}(A)$ and $\mathcal{P}(\pi)$ for any event $A \in \pi$ of probabilities on Σ, a unique up to affine transformation nonconstant continuous monotone function $u : X \to \mathbb{R}$, and $\alpha \in [0, 1]$ such that $\succeq_A$ is represented by the unique preference functional $V(\cdot|A) : \mathcal{F} \to \mathbb{R}$, and $\succeq$ is represented by unique $V(\cdot|\pi) : \mathcal{F} \to \mathbb{R}$ such that agent's behavior exhibits α-MEU SAR, and $u(\cdot), V(\cdot|\cdot), \mathcal{P}(A)$ and $\mathcal{P}(\pi)$ represent $\succeq_A^*$ and $\succeq_\pi^*$, as defined above.*

Moreover, for each A and π, $\mathcal{P}(A)$ and $\mathcal{P}(\pi)$ are unique, and α is unique if $\mathcal{P}(\pi)$ is not a singleton, and there exist $A \in \pi$ such that $\mathcal{P}(A)$ is not a singleton, too.

The uniqueness of representation is related to specific partition π, which is studied below. Also parameters and priors are unique in their representation of unambiguous preferences. However, Ghirardato, Maccheroni, and Marinacci (2004) find that some other set of priors and parameters representing the same preferences might exist, but they will not be related to the unambiguous preferences $\succeq^*$. Moreover, the proposed set of priors will be bigger, and the parameter of ambiguity aversion will be closer to $\frac{1}{2}$.

1.3 Uniqueness

This section discusses conditions when preferences $\succeq$ imply a unique subjective partition, i.e., when the subjective partition is identified from preference relations. It is important to note that the whole state space partition Ω and a trivial partition $\{\Omega\}$, which consists of one event that includes all states, always result in the same preferences, and, as a result, it is not possible to distinguish between them. From now on, we will treat them as one partition π_0.

As the above example demonstrates, strict α-MEU and MEU models might provide different results when it comes to identification. The two following subsections discuss uniqueness under these models separately.

1.3.1 Strict α-MEU

Assumption 1.1. *For any event $A \in \pi$, the sets of priors $\mathcal{P}(A)$ and $\mathcal{P}(\pi)$ are full-dimensional.*

The differences in preferences that arise due to distinct state aggregations come from various ways of bundling ambiguity. If assumption 1.1 does not hold, we know that some information is not ambiguous. Thus, we might get different subjective partitions depending on which states we bundle together. In order to see this, consider the following example:

Example 1.2. Suppose that the objective probability of no accident is 0.15. The agent's subjective partition is $\pi = \{A_1, A_2\}$, where $A_1 = \{s_1\}$ (no accident) and $A_2 = \{s_3, s_2\}$ (accident). The set of priors $\mathcal{P}(\Omega)$ is such that $P(s_3|\Omega) + P(s_2|\Omega) = 0.85$ and $P(s_1|\Omega) = 0.15$. Notice that $\mathcal{P}(\Omega)$ is not a full dimensional polytope. Let's compare behavior under subjective partition π with no state aggregation π_0. The probability of no accident is objective and does not depend on act x itself $P_x(s_1|\pi) = P_x(s_1|\pi_0) = 0.15$. We denote payoff at state i as $x_i = x(s_i)$; then, in order to evaluate an act, the agent without state aggregation has to solve the following problem:

$$\pi_0: \; P(s_1)u(x_1) + P(s_2)u(x_2) + P(s_3)u(x_3) \to \max / \min$$

$$\text{s.t. } P(s_1) = 0.15; \; P(s_2) + P(s_3) = 0.85.$$

The agent with subjective partition π will have to split the problem into two tasks. First, the conditional stage:

$$P(s_2|A_2)u(x_2) + P(s_3|A_2)u(x_3) \to \max / \min$$

$$\text{s.t. } P(s_2|A_2) + P(s_3|A_2) = 1.$$

Second, the ex-ante stage, which is degenerate in this case:

$$P(s_1)u(x_1) + P(A_2)V(x|A_2) \to \max / \min$$

$$\text{s.t. } P(s_1) = 0.15; \; P(A_2) = 0.85.$$

Note that the problem of the agent without state aggregation is equivalent to the set of problems with the subjective partition π. Thus, $P_x(s_i|\pi) = P_x(s_i|\pi_0)$. This implies that $V(x|\pi) = V(x|\pi_0)$ for any act $x \in \mathcal{F}$. It happens because unambiguous state s_1 might be put together with or separated from all other states as long as it does not affect the bundling of ambiguity.

Proposition 1.1. *Suppose that $0 < \alpha < 1$, assumption 1.1 holds, and $\pi, \pi' \in \mathcal{A}$ both α-MEU SAR of preferences $\succeq$; then, $\pi = \pi'$.*

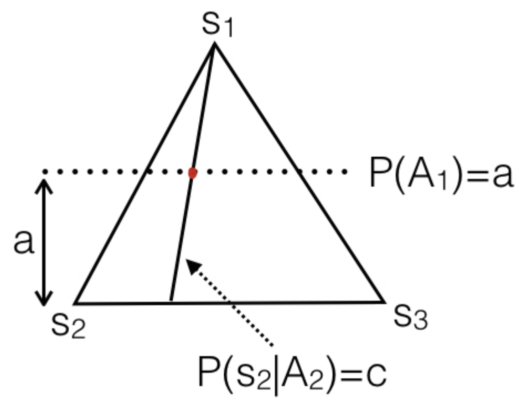

Figure 1.1: Probability simplex

1.3.2 MEU

Under the MEU model, full identification is not always possible. Rectangularity of priors is an important property for identification in this case.

Definition 1.5. *The set of priors $\mathcal{P}(\pi_0)$ is called π-rectangular if and only if there is a set $\mathcal{P}(\pi)$ and for any event $A \in \pi$, there exist sets $\mathcal{P}(A)$ such that for any state $s \in A$:*

$$\mathcal{P}(\pi_0)(s) = \{p_s p_A : p_s \in \mathcal{P}(A)(s), p_A \in \mathcal{P}(\pi)(A)\} = \mathcal{P}(A)(s)\mathcal{P}(\pi)(A).$$

It is easy to understand rectangularity graphically in a probability simplex. Let $\Omega = \{s_1, s_2, s_3\}$ and partition $\pi = \{A_1, A_2\}$, where $A_1 = s_1$ and $A_2 = \{s_2, s_3\}$. A point in the triangle represents one prior: the probability of state is the distance from the point to the side opposite to the vertex corresponding to the state. Now, suppose that we are interested in showing, conditional on event A_2, probabilities in the simplex. First, if $P(A_1) = a$, then all priors that satisfy it can be represented in the simplex as a line parallel to the side $s_2 s_3$ at distance a from it. In this case, $P(s_2|A_2) = c$ is the point such that the ratio of the right part of the line $P(A_1) = a$ to the whole line is c (see Figure 1.1). Thus, the conditional probability splits the line of $P(A_1) = a$ in a given ratio. Next, we allow the probability of event A_1 to

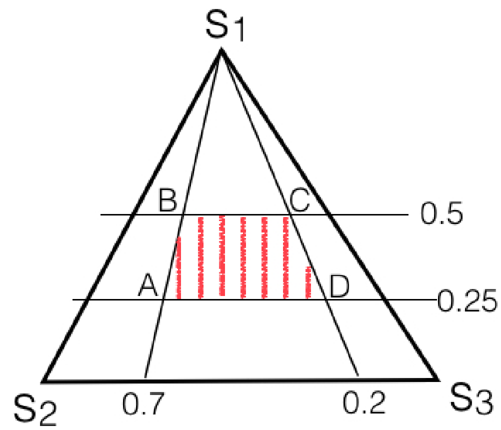

Figure 1.2: Rectangular priors

vary from 0 to 1. In this case, $P(s_2|A_2) = c$ is a line that starts in s_1 and ends at the side s_2s_3.

Now we consider what rectangular priors look like in the simplex. The definition specifies that a rectangular set can be obtained from the multiplication of conditional and ex-ante sets of priors. For example, suppose that $\mathcal{P}(\pi) = \{(p_{A_1}, p_{A_2}) \in \mathbb{R}^2 : 0.25 \le p_{A_1} \le 0.5;\ p_{A_1} + p_{A_2} = 1\}$ and $\mathcal{P}(A_2) = \{(p_2, p_3) \in \mathbb{R}^2 : 0.2 \le p_2 \le 0.7;\ p_2 + p_3 = 1\}$. Then, ABCD in Figure 1.2 is the π-rectangular set of priors that is obtained from the intersection of the sets $\mathcal{P}(\pi)$ and $\mathcal{P}(A_2)$.

Rectangularity is the reason for dynamic consistency for MEU in Epstein and Schneider (2003). In contrast, we find that a lack of dynamic consistency allows for the identification of the subjective partition. Unfortunately, with MEU, behavior under any subjective partition π could be also modeled by the whole state space as the subjective partition with π-rectangular priors. Thus, in this situation, we can obtain only partial identification.

Proposition 1.2. *Suppose that $\alpha = 1$, and a subjective partition π provides α-MEU SAR of preferences $\succeq$; then, the preferences can also be represented by partition π_0 with π-rectangular priors.*

However, when the subjective partition is the whole state space and priors are non-rectangular, we can uniquely identify the partition.

Proposition 1.3. *Suppose that $\alpha = 1$, assumption 1.1 holds, π_0, π both provide α-MEU SAR of preferences $\succeq$, and priors $\mathcal{P}(\pi_0)$ are not π-rectangular; then, $\pi = \pi_0$.*

1.4 Conclusions

We have provided axiomatization of the preferences over acts for the state aggregation model. The model assumes the presence of ambiguity and the α-MEU stage behavior. We have also shown that the lack of dynamic consistency in the stage model might implicitly suggest a unique subjective partition for preferences.

1.5 Appendix to Chapter 1

Proof of Theorem 1.1. First, due to Proposition 19 from Ghirardato, Maccheroni, Marinacci (2004), axioms 1.1, 1.2, and 1.3(i) guarantee unique $V(\cdot|A), \mathcal{P}(A), \beta(A)$ and a unique up to affine transformation $u(\cdot)$ that represent $\succeq_A^*$. Moreover,

$$V(f|A) = \beta(A) \min_{P(\cdot|A)\in\mathcal{P}(A)} \sum_{s\in A} u(f(s))P(s|A) + (1-\beta(A)) \max_{P(\cdot|A)\in\mathcal{P}(A)} \sum_{s\in A} u(f(s))P(s|A).$$

We now show that $\beta(A) = \beta$ for all $A \in \mathcal{A}$. Suppose that $C_A(f) = C_B(g) = [x_1, x_2]$. Then, by axiom 1.3(ii), $f \sim_A \beta(A)x_1 + (1 - \beta(A))x_2 \sim_A x$; $g \sim_B \beta(B)x_1 + (1 - \beta(B))x_2 \sim_B x$. The last implies that $\beta(A) = \beta(B) = \beta$.

Next, we demonstrate that monotonicity with respect to events in π holds for the preferences $\succeq$. We need to show that if $f \succeq_A g$ for all $A \in \pi$, then $f \succeq g$. Suppose that $\pi = \{A_1, A_2, ..., A_n\}$. Consider a constant act $x_f^{A_i}$ such that

$f \sim_{A_i} x_f^{A_i}$ and $fA_ih \sim x_f^{A_i}A_ih$. Also, notice that

$$f \sim x_f^{A_1}A_1f \sim x_f^{A_1}A_1x_f^{A_2}A_2f \sim ... \sim x_f^{A_1}A_1x_f^{A_2}A_2...x_f^{A_n}A_n.$$

The same will hold for act g, by analogy. However, if $f \succeq_{A_i} g$, then $x_f^{A_i} \succeq_{A_i} x_g^{A_i}$, meaning that x_f has a value higher than or equal x_g at each state. Thus, by monotonicity

$$x_f^{A_1}A_1x_f^{A_2}A_2...x_f^{A_n}A_n \succeq x_g^{A_1}A_1x_g^{A_2}A_2...x_g^{A_n}A_n. \Rightarrow f \succeq g$$

Now we demonstrate that axiom 1.3(iii) implies that if $C_\pi(f) = C_\pi(g)$, then $f \sim g$. Suppose that $C_\pi(f) = C_\pi(g)$; then, we take an act h such that there exists an event $A \in \pi$: $C_\pi(f) = C_\pi(g) = C_A(h)$. Then, there exists $x \in X$: $f \sim x$ and $g \sim x$, so $f \sim g$.

Next, we apply Proposition 19 from Ghirardato, Maccheroni, Marinacci (2004) again. However, we will treat events from π as states. We obtain unique $V(\cdot|\pi)$, $\mathcal{P}(\pi)$, α and unique up to affine transformation $I(\cdot)$ that represent $\succeq^*$, and the representation

$$V(f|\pi) = \alpha \min_{P\in\mathcal{P}(\pi)} \sum_{A\in\pi} I(f|A)P(A|\pi) + (1-\alpha) \max_{P\in\mathcal{P}(\pi)} \sum_{A\in\pi} I(f|A)P(A|\pi).$$

Note that $I(x|A)$ and $V(x|A)$ represent the same preferences, and this means that they are monotone transformations of each other. However, they have to be equal on constant acts, so they coincide. Also, notice that $\alpha = \beta$ follows straight from axiom 1.3(iii). $\square$

Suppose that there are two different partitions, π and π'. Denote a set $\tilde{\pi} = \{C \in \Sigma : \exists A \in \pi, B \in \pi' : C = A \cap B\}$.

Lemma 1.1. *If $\pi, \pi' \in \mathcal{A}$ and $C \in \tilde{\pi}$, then $C \in \mathcal{A}$.*

Proof. Take events $A \in \pi$ and $B \in \pi'$ such that $C = A \cap B$. Then, $fCx \succeq_A gCx \Leftrightarrow fCx(A \setminus C)h \succeq gCx(A \setminus C)h$. The last relation is equivalent to $fCh(B \setminus C)x(A \setminus C)h \succeq gCh(B \setminus C)x(A \setminus C)h \Leftrightarrow fCh \succeq_B gCh$.

Now, by providing the same argument from event B back to A, one can easily obtain that $fCh \succeq_A gCh$. □

Lemma 1.2. *Suppose that $f(C_1 \cup C_2)x_f \sim_B x_f$; $fC_1x_1 \sim_{A_1} x_1$ and $fC_2x_2 \sim_{A_2} x_2$, $x_1 \neq x_2$ and $A_i \cap B = C_i$. Then, there exists i (1 or 2) such that $x_i = x_f$ if and only if $\alpha = 0, 1$ and for any $s \in B$: $P(s|B) = 0$ is in the set of priors $\mathcal{P}(B)$.*

Proof. $x_1C_1x_2C_2x_f \sim_B x_f$. Then,

$$V(x_1C_1x_2C_2x_f) - u(x_f) = \alpha \min_p \left(P(C_1|B)(u(x_1) - u(x_f)) + P(C_2|B)(u(x_2) - u(x_f))\right) +$$
$$+(1-\alpha) \max_p \left(P(C_1|B)(u(x_1) - u(x_f)) + P(C_2|B)(u(x_2) - u(x_f))\right) = 0.$$

Suppose that $x_j = x_f$, and, for simplicity, $x_i > x_f$. Then,

$$V(x_jC_jx_iC_ix_f) - u(x_f) = (u(x_i) - u(x_f))(\alpha P_{\min} + (1-\alpha)P_{\max}) = 0,$$

which implies that either $\alpha = 0$ and $P_{\max}(C_i|B) = 0$ or $\alpha = 1$ and $P_{\min}(C_i|B) = 0$. Thus, $P(C_i|B) = 0$ is available in the set of priors.

Now, suppose that for each state s: $P(s|B) = 0$ is available. Suppose that $x_j > x_f > x_i$:

$$(u(x_j) - u(x_f))(\alpha P^j_{\min} + (1-\alpha)P^j_{\max}) + (u(x_i) - u(x_f))(\alpha P^i_{\min} + (1-\alpha)P^i_{\max}) = 0.$$

Then, $P^j_{\min} = P^i_{\max} = 0$ and $P^i_{\min} = P^j_{\max} = 1$. This implies that $u(x_f) = \alpha u(x_i) + (1-\alpha)u(x_j)$. Note that if $\alpha \neq 0, 1$, then there is no such i that $x_f = x_i$. □

Lemma 1.3. *If there exists $C \notin \pi$, but $C \in \mathcal{A}$, and there exists $A \in \pi$: $C \subset A$ and C consists of two or more states, then one of the following must be true:*

1. $P(C|A) = \sum_{s\in C} P(s|A)$ *is a constant;*

2. *For all* $s \in C$*:* $P(s|C) = \frac{P(s|A)}{\sum_{s\in C} P(s|A)}$ *is a constant;*

3. $\alpha = 0, 1$ *and either* $\exists s \in C$*:* $P(s|A) = 0$ *is in the set of priors* $\mathcal{P}(A)$ *or* $\mathcal{P}(A)$ *is* C*-rectangular.*

Proof. We define constant act x_f: $fCx_f \sim_A x_f$. This implies that $fCh \sim_A x_fCh$ for any act h. To simplify calculations, instead of h, we take some constant act y. Thus, $V(fCy|A) = V(x_fCy|A)$, where values are

$$V(fCy|A) = \alpha \min_{p(\cdot|A)} \left(\sum_{s\in C} p(s|A)u(f(s)) + \sum_{s\in A\setminus C} p(s|A)u(y) \right)$$
$$+(1-\alpha) \max_{p(\cdot|A)} \left(\sum_{s\in C} p(s|A)u(f(s)) + \sum_{s\in A\setminus C} p(s|A)u(y) \right)$$

$$V(x_fCy|A) = \alpha \min_{p(\cdot|A)} \left(P(C|A)u(x_f) + \sum_{s\in A\setminus C} p(s|A)u(y) \right)$$
$$+(1-\alpha) \max_{p(\cdot|A)} \left(P(C|A)u(x_f) + \sum_{s\in A\setminus C} p(s|A)u(y) \right).$$

Now, by subtracting $u(x_f)$ from both expressions and setting them equal to each other, one can obtain the following:

$$\alpha \min_{p(\cdot|A)} \left(\sum_{s\in C} p(s|A)(u(f(s)) - u(x_f)) + (1 - P(C|A))(u(y) - u(x_f)) \right) +$$
$$+(1-\alpha) \max_{p(\cdot|A)} \left(\sum_{s\in C} p(s|A)(u(f(s)) - u(x_f)) + (1 - P(C|A))(u(y) - u(x_f)) \right) =$$
$$= \alpha \min_{p(\cdot|A)} (1 - P(C|A))(u(y) - u(x_f)) + (1-\alpha) \max_{p(\cdot|A)} (1 - P(C|A))(u(y) - u(x_f)).$$

The above expression must hold for any acts y and f such that $fCx_f \sim_A x_f$.

If the C-related part inside min and max is always 0, then one of the following is true: (1) for any $s \in C$: $P(s|C) = const$; (2) $\alpha = 0, 1$ and there always exists s such that $P(s|A) = 0$ is available in the set of priors that, by Lemma 1.2, would imply that there is always $u(f(s)) = u(x_f)$.

If the C-related part is not always 0, then the set of priors has to satisfy some kind of separability between events C and $A \setminus C$. It means one of the following must hold: (1) $P(C|A) = const$; (2) $\alpha = 0, 1$ and $\mathcal{P}(A)$ is C-rectangular. □

Lemma 1.4. *Suppose that there are events $A_1, A_2 \in \pi$ and $B \in \pi'$, $A_i \cap B = C_i$ and there is i such that $A_i \neq C_i$; then, one of the following holds:*

1. *$P(A_i)$ is a constant;*
2. *$\frac{P(A_i)}{P(A_1)+P(A_2)}$ is a constant;*
3. *$\alpha = 0, 1$ and there is state $s \in C_i$ such that $P(s|A_i) = 0$ is available in the set of priors $\mathcal{P}(A_i)$ or $P(A_i) = 0$ is available in the set of priors $\mathcal{P}(\pi)$.*

Proof. Define $f(C_1 \cup C_2)x_f \sim_\pi x_f$ and $fC_ix_i \sim_{A_i} x_i$. Note that $f(C_1 \cup C_2)h \sim_\pi x_f(C_1 \cup C_2)h$. Now, instead of act h, we take an act such that it gives x_i on $A_i \setminus C_i$ and some constant act y everywhere else. Then,

$$x_fC_1x_1(A_1 \backslash C_1)x_fC_2x_2(A_2 \backslash C_2)y \sim_\pi f(C_1 \cup C_2)x_1(A_1 \backslash C_1)x_2(A_2 \backslash C_2)y \sim_\pi x_1A_1x_2A_2y.$$

We denote act $g = x_fC_1x_1(A_1 \setminus C_1)x_fC_2x_2(A_2 \setminus C_2)y$. Note that

$$V(g) - u(x_f) = \alpha \min_{P(\cdot|\pi)} \left(\sum_{A_i} P(A_i|\pi)(V(x_fC_ix_i|A_i) - u(x_f)) + (1 - \sum_{A_i} P(A_i))(u(y) - u(x_f)) \right) +$$
$$+(1-\alpha) \max_{P(\cdot|\pi)} \left(\sum_{A_i} P(A_i|\pi)(V(x_fC_ix_i|A_i) - u(x_f)) + (1 - \sum_{A_i} P(A_i))(u(y) - u(x_f)) \right),$$

where

$$V(x_f C_i x_i|A_i) - u(x_f) = \alpha \min_{P(\cdot|A_i)} \left(\sum_{s \in A_i \setminus C_i} P(s|A_i)(u(x_i) - u(x_f)) \right) +$$

$$+(1-\alpha) \max_{P(\cdot|A_i)} \left(\sum_{s \in A_i \setminus C_i} P(s|A_i)(u(x_i) - u(x_f)) \right) = (u(x_i) - u(x_f))\,(1 - (\alpha P_{\min}(C_i|A_i)+$$

$$+(1-\alpha) P_{\max}(C_i|A_i)) = (u(x_i) - u(x_f))\,(1 - P_\alpha(C_i|A_i)).$$

$V(x_f C_i x_i|A_i) - u(x_f)$ can be 0 either if $C_i = A_i$ or if $x_i = x_f$. By Lemma 1.2, the last implies that $\alpha = 0, 1$, and there is $s \in C_i$ such that $P(s|A_i) = 0$ is in the set of priors $\mathcal{P}(A_i)$. Thus,

$$V(g) - u(x_f) = \alpha \min_{P(\cdot|\pi)} \left(\sum_{A_i} (u(x_i) - u(x_f))\,(1 - P_\alpha(C_i|A_i)) P(A_i) + (1 - \sum_{A_i} P(A_i))(u(y) - u(x_f)) \right)$$

$$+(1-\alpha) \max_{P(\cdot|\pi)} \left(\sum_{A_i} (u(x_i) - u(x_f))\,(1 - P_\alpha(C_i|A_i)) P(A_i) + (1 - \sum_{A_i} P(A_i))(u(y) - u(x_f)) \right)$$

$$= \sum_{A_i} (u(x_i) - u(x_f))\,(1 - P_\alpha(C_i|A_i)) \tilde{P}_\alpha(A_i) + (1 - \sum_{A_i} \tilde{P}_\alpha(A_i))(u(y) - u(x_f)).$$

We now compare it with the value of $x_1 A_1 x_2 A_2 y$:

$$V(x_1 A_1 x_2 A_2 y) - u(x_f) = \alpha \min_{P(\cdot|\pi)} \left(\sum_{A_i} P(A_i)(u(x_i) - u(x_f)) + (1 - \sum_{A_i} P(A_i))(u(y) - u(x_f)) \right) +$$

$$+(1-\alpha) \max_{P(\cdot|\pi)} \left(\sum_{A_i} P(A_i)(u(x_i) - u(x_f)) + (1 - \sum_{A_i} P(A_i))(u(y) - u(x_f)) \right) =$$

$$= \sum_{A_i} P_\alpha(A_i)(u(x_i) - u(x_f)) + (1 - \sum_{A_i} P_\alpha(A_i))(u(y) - u(x_f)).$$

The above expressions will trivially coincide if A_i-related parts are always 0. This is possible in one of the following situations: (1) $\frac{P(A_i)}{P(A_1)+P(A_2)}$ is a constant and $C_i = A_i$; (2) $P(A_i)$ is constant; (3) there exists some i: $x_i = x_f$, and $P(A_j) = 0$ is in the set of priors $\mathcal{P}(\pi)$, which imply that $\alpha = 0, 1$, and $s_i \in A_i$ such that $P(s_i|A_i) = 0$ is in the set of priors.

Now, suppose that A_i-related parts are not always 0. The expressions must be equal to each other for any value of y, which implies that $\tilde{P}_\alpha(A_1) + \tilde{P}_\alpha(A_2) = P_\alpha(A_1) + P_\alpha(A_2)$. Note that $u(x_f) = wu(x_1) + (1-w)u(x_2)$, where $w \in (0,1)$. By substituting it into both expressions and setting them equal to each other, one can obtain:

$$w = \frac{P_\alpha(A_1) - (1 - P_\alpha(C_1|A_1))\tilde{P}_\alpha(A_1)}{P_\alpha(A_1) - (1 - P_\alpha(C_1|A_1))\tilde{P}_\alpha(A_1) + P_\alpha(A_2) - (1 - P_\alpha(C_2|A_2))\tilde{P}_\alpha(A_2)}.$$

The above ratio must be constant for any value of y, while $P_\alpha(C_i|A_i)$ does not change, depending on y. This implies that $\frac{P_\alpha(A_1)-(1-P_\alpha(C_1|A_1))\tilde{P}_\alpha(A_1)}{P_\alpha(A_2)-(1-P_\alpha(C_2|A_2))\tilde{P}_\alpha(A_2)} = c$, where c is a constant. If $P_\alpha(A_2)$ and $\tilde{P}_\alpha(A_2)$ are not equal to 0 (this would imply that 0 is in the set of priors for A_i), then the last statement can be rewritten as follows

$$\left(\frac{P_\alpha(A_1)}{P_\alpha(A_2)} - c\right)\frac{P_\alpha(A_2)}{\tilde{P}_\alpha(A_2)} = (1 - P_\alpha(C_1|A_1))\frac{\tilde{P}_\alpha(A_1)}{\tilde{P}_\alpha(A_2)} - c(1 - P_\alpha(C_2|A_2)).$$

On the other hand, $\tilde{P}_\alpha(A_1) + \tilde{P}_\alpha(A_2) = P_\alpha(A_1) + P_\alpha(A_2)$, which can be rewritten as

$$\left(\frac{P_\alpha(A_1)}{P_\alpha(A_2)} + 1\right)\frac{P_\alpha(A_2)}{\tilde{P}_\alpha(A_2)} = \frac{\tilde{P}_\alpha(A_1)}{\tilde{P}_\alpha(A_2)} + 1.$$

By dividing the first expression by the second one, we get:

$$\frac{\frac{P_\alpha(A_1)}{P_\alpha(A_2)} - c}{\frac{P_\alpha(A_1)}{P_\alpha(A_2)} + 1} = \frac{(1 - P_\alpha(C_1|A_1))\frac{\tilde{P}_\alpha(A_1)}{\tilde{P}_\alpha(A_2)} - c(1 - P_\alpha(C_2|A_2))}{\frac{\tilde{P}_\alpha(A_1)}{\tilde{P}_\alpha(A_2)} + 1}.$$

The above expression depends only on the ratios $\frac{P_\alpha(A_1)}{P_\alpha(A_2)}$ and $\frac{\tilde{P}_\alpha(A_1)}{\tilde{P}_\alpha(A_2)}$ and not on the actual values. It is possible in one of the following situations:

1. $c = 0$ implies $P_\alpha(C_1|A_1) = 0$ and $P_\alpha(A_i) = \tilde{P}_\alpha(A_i)$. It means that either $\alpha = 0, 1$ and $P(C_1|A_1) = 0$ is available in $\mathcal{P}(A_1)$ or $P(C_1|A_1) = 0$.

2. $c \neq 0$ implies that $P_\alpha(C_i|A_i) = 0$.

$\square$

Proof of Proposition 1.1. Suppose that $\pi \neq \pi'$. Then, in one of the partitions (suppose π), there exists event A such that it intersects with at least two events from another partition (π'). Then, only two options are possible: either $A = \cup_i B_i$ or $A \neq \cup_i B_i$ for some $B_i \in \pi'$.

First, suppose that $A = \cup_i B_i$. Now, in order to use Lemma 1.3, replace s_i with B_i, C with A, and A with π. Then, either $P(A|\pi')$ or $P(B_i|A)$ is a constant. Unless $A = \Omega$, the last two implications suggest that priors are not full-dimensional. If at least one of B_i consists of two or more states, then, again, by Lemma 1.3, either $P(B_i|A)$ is a constant or for $s \in B_i$ $P(s|B_i)$ is a constant. Thus, the only possibility left is when $A = \Omega$ and all $B_i = s_i$, which means that we are comparing $\pi_0 = \{\Omega\}$ and $\pi^* = \Omega$, that we consider to be the same in this paper.

Finally, suppose that $A \neq \cup_i B_i$. Then, by Lemma 1.4, we know that either $P(B_i)$ or $\frac{P(B_i)}{\sum P(B_i)}$ is a constant, which contradicts full-dimensionality of priors. $\square$

Proof of Proposition 1.2. Suppose that preferences are represented by some π and $\alpha = 1$ with priors $\mathcal{P}(\pi)$ and $\mathcal{P}(A)$ for each $A \in \pi$. Then, the value functional is

$$V(f|\pi) = \min_{q \in \mathcal{P}(\pi)} \sum_{A \in \pi} q(A) \left(\min_{p \in \mathcal{P}(A)} pu(f) \right) = \min_{p \in \mathcal{P}(\pi_0)} pu(f),$$

where $\mathcal{P}(\pi_0)(s) = \mathcal{P}(\pi)(A) \cdot \mathcal{P}(A)(s)$ if $s \in A$ – i.e., $\mathcal{P}(\pi_0)$ is π-rectangular. $\square$

Proof of Proposition 1.3. Preferences are represented by some partition π, so by Proposition 1.2, the value functional can be rewritten through the trivial partition π_0 with π-rectangular priors $\tilde{\mathcal{P}}(\pi_0)$. On the other hand, preferences can be represented by the trivial partition with non π-rectangular priors $\mathcal{P}(\pi_0)$. Due to the uniqueness of representation of MEU preferences, priors must coincide. Thus, $\pi = \pi_0$.

$\square$

CHAPTER 2

Identification of Subjective Partition from Choices

Chapter 1 introduces a model of subjective partition under ambiguity and demonstrates that if the agent is allowed to have a set of multiple priors instead of single probabilities of states, then various methods of state aggregation lead to different preferences. Chapter 2 demonstrates the same phenomena for choices that the agent makes: different partitions imply different choices. For this reason, the subjective partition of each individual might be identified from observed data.

To explain the idea in more detail consider the example from Chapter 1: Suppose that there is an agent who needs to choose car insurance. Possible states of the world include an accident with the agent at fault, an accident with the other party at fault, and no accident. An insurance policy consists of two deductibles: one applies if the agent caused the accident, and the other applies if the agent was not responsible. Probabilities of the states are unknown and described by a set of multiple priors – i.e., all agents face ambiguity. Next, suppose that the agent is a person that does not aggregate states. If she is ambiguity-averse then, generally, her choice is a bundle of deductibles such that her consumption is smoothed between all states. Note that, in this case, there is no state aggregation, and her subjective partition is the whole state space of the world. Next, imagine a person who has difficulty analyzing all three states of the world together. In order to simplify the decision process, she combines both states that involve a collision into the event "accident," while keeping the "no accident" state separate from them. In this case, the state space of the economy is partitioned by events "accident" (with the agent or the other party at fault) and "no accident." The presence of state aggregation changes the way that the agent thinks about the world. Instead of an accident with the agent at fault, an accident with the other party at fault, and no accident, her subjective states now are "accident" and "no accident." Moreover, the individual generally wants to smooth her consumption between her aggregated states. As a result, she will buy less insurance (greater deductibles) for both types of accidents, compared with what would be optimal for her. Thus, choices differ with the subjective partition, implying that under certain conditions the subjective partition might be identified from choices.

Also, it is important to pay attention to the fact that if the individual's behavior were described by the SEU model, then the choice of the insurance would not depend on state aggregation. The agent would always choose the same amount of deductibles and, as a result, keep researchers from identifying the subjective partition. This occurs because the probability of each state used for evaluating an act is an actual subjective probability of this state and it does not depend on the act itself. Thus, when the agent aggregates several states with subjective probabilities, the final weight of the state in the act's value is a product of the state's conditional probability given the aggregated event, and the probability of the event. Thus, the weight is just the subjective probability of the state. As a result, the ranking of acts is not affected by the aggregation of states. On the other hand, in the presence of ambiguity, "probabilities" (weights) of states used for act evaluation do depend on the actual act, and the ranking of acts changes with the subjective partition.

In this Chapter, we (1) identify the subjective partition from choices; and, (2) suggest testable restrictions of the model.

First, we show identification of the subjective partition from choices and prices in a complete market. We make an additional assumption of polytope priors that allows for SEU behavior in different regions of the act space. This type of behavior implies that the agent uses the same subjective probabilities in order to evaluate an act in some region; however, these probabilities differ across regions. The described property permits construction of a method of identification that might be adapted to estimate the model in the future. The proposed method consists of two steps: (1) to identify region-specific probabilities together with utility function; and (2) to obtain the subjective partition from the way that these probabilities change across the regions.

Second, we provide a finite system of revealed preferences inequalities that allows for testing our model by a dataset. An example of the potential dataset is data on choices of car insurance plans that consist of deductibles and income. This paper is related to the literature, started by Afriat (1967), that suggested that a dataset can be rationalized by an increasing concave continuous utility function if and only if there exists a solution to a system of linear inequalities obtained from the dataset. Similar tests for expected utility and its generalizations were developed afterwards (Varian (1983), Diewert (2012), Kubler, Selden, and Wei (2014), Echenique and Saito (2015), etc.). Polisson, Quah, and Renou (2015) develop a testing procedure that can be used for many different models of choice under uncertainty. They allow for non-concave utility, and, as a result, test for an actual model without extra assumptions. However, the price of no additional assumptions is that the system of inequalities becomes non-linear. The procedure from Polisson et al. is applicable to the model developed in this paper after additional inequalities on priors are derived. The final system of inequalities for testing the model is finite and bilinear. The previous literature (e.g. Brown and Matzkin (1996), etc.) suggests applying the Tarski-Seidenberg algorithm in this case.

2.1 Identification in the Market

In this section, we show how to identify the subjective partition from choices of Arrow securities and their prices. Even if Arrow assets are not directly available in the market, as long as the market is complete, Arrow prices can always be uniquely recovered.

2.1.1 Deductibles as Arrow assets

Note that the problem of choosing an insurance plan that contains deductibles can be represented as a choice among Arrow securities. In order to see this, consider the following example.

Example 2.1. Suppose that an agent needs to choose an insurance plan while considering three states of the world, as before: an accident with the agent at fault with losses L_1; an accident with the other party at fault with losses L_2; and no accident, which implies no loss. The price of insurance consists of the sum of prices of chosen deductibles in both states of the accident. We assume that price $q(d)$ of each deductible d is linear in its amount d: $q(d) = a - p_d d$, where a is a constant state premium. Thus, we can define the price of a \$1 deductible decrease as $p_d = q(d) - q(d+1)$, which is constant for any amount d due to assumed linearity. Notice that p_d is the price of a corresponding Arrow security: decreasing the deductible by \$1 in some state is equivalent to increasing consumption by \$1 in the same state.

We denote the deductible in the case of an accident with the agent at fault as d_1, the price of \$1 of deductible in this state as p_{d_1}, and a_1 as the state premium. Similarly, d_2, p_{d_2}, and a_2 are the deductible, the \$1 price in the case of an accident with the other party at fault, and the state premium, respectively. Thus, the total price of the insurance plan (d_1, d_2) is $p = (L_1 - d_1)p_{d_1} + a_1 + (L_2 - d_2)p_{d_2} + a_2$. The agent has income I and optimizes consumption $c = (c_1, c_2, c_3)$ in all states of the world given some subjective partition π. All the money left (s) after paying for insurance is used for consumption during the year. Thus, the agent's problem is

$$V(c|\pi) \to \max$$

$$\text{s.t. } (L_1 - d_1)p_{d_1} + (L_2 - d_2)p_{d_2} + a_1 + a_2 + s = I$$

$$c_1 = s - d_1;\ c_2 = s - d_2;\ c_3 = s.$$

We now rewrite the problem in terms of Arrow securities. Notice that consumption bundle $c = (c_1, c_2, c_3)$ is exactly the portfolio of Arrow assets that the agent chooses. The only thing left to do is to find Arrow prices (p_1, p_2, p_3) and rewrite the budget constraint in an appropriate form. Denote $\tilde{I} = I - L_1p_1 - L_2p_2 - a_1 - a_2$ the agent's total endowment when taking potential losses and all state premiums into account. Also, as mentioned above, $p_1 = p_{d_1}$ and $p_2 = p_{d_2}$. Then, we can rewrite the budget constraint as

$$-d_1p_{d_1} - d_2p_{d_2} + s = (c_1 - s)p_1 + (c_2 - s)p_2 + s = c_1p_1 + c_2p_2 + c_3(1 - p_1 - p_2) = \tilde{I}.$$

Thus, $p_3 = 1 - p_1 - p_2$ and the agent's problem is

$$V(c|\pi) \to \max$$

$$\text{s.t. } c_1p_1 + c_2p_2 + c_3p_3 = \tilde{I}.$$

As a consequence, if we observe the choices of deductibles with their prices and the agent's income, we can recover the amount of Arrow securities/consumption in each state.

2.1.2 Notation

Let us define some notation that is used below. Given event A and act $x \in \mathcal{F}$, we denote the conditional priors that minimize and maximize "expected" utility $\sum_{s\in A} u(x(s))P(s|A)$ at the conditional stage, where states are aggregated into event A, as follows:

$$P_{\min x}(\cdot|A) = \underset{P(\cdot|A)\in\mathcal{P}(A)}{\operatorname{argmin}} \sum_{s\in A} u(x(s))P(s|A)$$
$$P_{\max x}(\cdot|A) = \underset{P(\cdot|A)\in\mathcal{P}(A)}{\operatorname{argmax}} \sum_{s\in A} u(x(s))P(s|A).$$

Similarly, given partition π, the ex-ante priors that minimize and maximize "expected" value at the ex-ante stage are:

$$P_{\min x}(\cdot|\pi) = \underset{P(\cdot|\pi)\in\mathcal{P}(\pi)}{\operatorname{argmin}} \sum_{A\in\pi} V(x|A)P(A|\pi)$$
$$P_{\max x}(\cdot|\pi) = \underset{P(\cdot|\pi)\in\mathcal{P}(\pi)}{\operatorname{argmax}} \sum_{A\in\pi} V(x|A)P(A|\pi).$$

To simplify the notation, we denote

$$P_{\alpha;x}(A|\pi) = \alpha P_{\min x}(A|\pi) + (1-\alpha)P_{\max x}(A|\pi)$$
$$P_{\alpha;x}(s|A) = \alpha P_{\min x}(s|A) + (1-\alpha)P_{\max x}(s|A)$$

$$P_x(s|\pi) \equiv P_{\alpha;x}(s|A)P_{\alpha;x}(A|\pi).$$

where $s \in A \in \pi$. We abuse the notation by calling $P_{\alpha;x}(A|\pi)$ and $P_{\alpha;x}(s|A)$ the probabilities of ex-ante and conditional stages, where no confusion should arise. We also call $P_x(s|\pi)$ a probability.[1]

[1] Note that $P_x(s|\pi)$ behaves like a probability only for act x, but it is not a probability in a regular sense and changes across acts.

2.1.3 Non-parametric example

Suppose that $\Omega = \{s_1, s_2, s_3\}$, partition $\pi = \{A_1, A_2\}$, where $A = \{s_1\}$ and $A_2 = \{s_2, s_3\}$, and sets of priors are $\mathcal{P}(A_2) = \{(p_2, p_3) \in \mathbb{R}^2 : p_3 \geq \phi,\ p_2 + p_3 = 1\}$ and $\mathcal{P}(\pi) = \{(p_{A_1}, p_{A_2}) \in \mathbb{R}^2 : p_{A_2} \geq \phi,\ p_{A_1} + p_{A_2} = 1\}$. Consider that a complete set of Arrow securities is available on the market.[2] The agent purchases a bundle of Arrow securities that maximizes her value given a certain amount of income I and the price p_i of an Arrow security that pays 1 in state i:

$$V((x_1, x_2, x_3)|\pi) \to \max_x$$
$$\text{s.t. } p_1x_1 + p_2x_2 + p_3x_3 = I.$$

Choices $x = (x_1, x_2, x_3)$, prices $p = (p_1, p_2, p_3)$ and income I are observed. We assume that all possible combinations of (p, I) are available. The purpose is to identify the subjective partition π, utility function $u(\cdot)$, sets of priors $\mathcal{P}(\pi)$ and $\mathcal{P}(A_2)$, and parameter α.

Note that the proposed set of priors is a full-dimensional polytope, so it implies that the "probability" of each state at a given act is a linear combination of solutions to some linear programming problems. Moreover, these probabilities are constant in some region around the act due to the fact that linear programming provides corner solutions. Thus, we can obtain these regions together with the related probabilities that the agent uses to solve the problem.

Consider the conditional on the event A_2 stage. This stage and its priors will produce one boundary $x_2 = x_3$, such that the regional conditional probabilities will be as follows:

1. If $x_2 > x_3$, then $P_{\min x}(s_3|A_2) = \phi$; $P_{\max x}(s_3|A_2) = 1$; $P_{x;\alpha}(s_3|A_2) = 1 - \alpha(1 - \phi)$.

[2] An Arrow security pays one unit in a specified state and zero otherwise.

2. If $x_2 < x_3$, then $P_{\min x}(s_3|A_2) = 1$; $P_{\max x}(s_3|A_2) = \phi$; $P_{x;\alpha}(s_3|A_2) = \alpha + (1-\alpha)\phi$.

Now, consider the ex-ante stage. This stage boundary is

$$x_1 = u^{-1}\left(P_x(s_2|A_2)u(x_2) + P_x(s_3|A_2)u(x_3)\right).$$

We denote the above boundary as $b(x_2, x_3)$. This stage has the following ex-ante probabilities:

1. If $x_1 > b(x_2, x_3)$, then $P_{\min x}(A_2|\pi) = 1$; $P_{\max x}(A_2|\pi) = \phi$; $P_{f;\alpha}(A_2|\pi) = \alpha + (1-\alpha)\phi$.

2. If $x_1 < b(x_2, x_3)$, then $P_{\min x}(A_2|\pi) = \phi$; $P_{\max x}(A_2|\pi) = 1$; $P_{f;\alpha}(A_2|\pi) = \alpha\phi + 1 - \alpha$.

Thus, in total, we have four different regions with the following probabilities:

1. If $x_3 > x_2$ and $x_1 > b(x_2, x_3)$, then $P_x(s_1|\pi) = (1-\alpha)(1-\phi)$; $P_x(s_2|\pi) = (\alpha + (1-\alpha)\gamma)\alpha(1-\phi)$; $P_x(s_3|\pi) = (\alpha + (1-\alpha)\phi)(1-\alpha(1-\phi))$.

2. If $x_3 < x_2$ and $x_1 > b(x_2, x_3)$, then $P_x(s_1|\pi) = (1-\alpha)(1-\phi)$; $P_x(s_2|\pi) = (\alpha + (1-\alpha)\phi)(1-\phi)(1-\alpha)$; $P_x(s_3|\pi) = (\alpha + (1-\alpha)\phi)(\phi + (1-\phi)\alpha)$.

3. If $x_3 > x_2$ and $x_1 < b(x_2, x_3)$, then $P_x(s_1|\pi) = \alpha(1-\phi)$; $P_x(s_2|\pi) = (1-\alpha(1-\phi))\alpha(1-\phi)$; $P_x(s_3|\pi) = (1-\alpha(1-\phi))(1-\alpha(1-\phi))$.

4. If $x_3 < x_2$ and $x_1 < b(x_2, x_3)$, then $P_x(s_1|\pi) = \alpha(1-\phi)$; $P_x(s_2|\pi) = (1-\alpha(1-\phi))(1-\phi)(1-\alpha)$; $P_x(s_3|\pi) = (1-\alpha(1-\phi))(\phi + (1-\phi)\alpha)$.

The agents solves her maximization problem and chooses some bundle x:

$$\begin{aligned} P_x(s_1|\pi)u(x_1) + P_x(s_2|\pi)u(x_2) + P_x(s_3|\pi)u(x_3) &\to \max_x \\ \text{s.t. } p_1x_1 + p_2x_2 + p_3x_3 &= I. \end{aligned}$$

For simplicity, assume that $\alpha > 0.5$ – i.e., the agent is ambiguity-averse. In this case, the indifference curves will have kinks at the bundles that connect different regions. Moreover, they will be observed as choices made under the same income, but under a range of prices. Thus, we will be able to separate the regions from each other in our observations.

Inside each region and for each chosen bundle and any states i and j, the optimality condition holds:

$$\frac{p_i}{p_j} = \frac{P_x(s_i|\pi)}{P_x(s_j|\pi)}\frac{u'(x_i)}{u'(x_j)}. \tag{2.1}$$

Now we analyze identification, for example, in region 1. For some fixed value of $\bar{x}_2$, we observe bundles with $x_3 > \bar{x}_2$ (and $x_1 > b_1(\bar{x}_2, x_3)$). By taking two bundles with different x_3 but the same $\bar{x}_2$ from this region:

$$\frac{p_3^1}{p_2^1} = \frac{P_x(s_3|\pi)}{P_x(s_2|\pi)}\frac{u'(x_3^1)}{u'(\bar{x}_2)} \text{ and } \frac{p_3^2}{p_2^2} = \frac{P_x(s_3|\pi)}{P_x(s_2|\pi)}\frac{u'(x_3^2)}{u'(\bar{x}_2)}$$

$$\frac{u'(x_3^2)}{u'(x_3^1)} = \frac{p_3^2 p_2^1}{p_2^2 p_3^1}.$$

Thus, the ratio of derivatives of the utility function can be identified in the interval $[\,\bar{x}_2; +\infty\,)$. Now, we choose another bundle with $\tilde{x}_2$, such that $\tilde{x}_2 > \bar{x}_2$, and $x_3 > \tilde{x}_2$, and some x_1 that satisfies conditions of region 1. The ratio of derivatives $\frac{u'(x_3)}{u'(\tilde{x}_2)}$ has already been identified, so the probability ratio is identified, as well:

$$\frac{P_x(s_3|\pi)}{P_x(s_2|\pi)} = \frac{p_3}{p_2}\frac{u'(\tilde{x}_2)}{u'(x_3)}.$$

By analogy, we can identify all probability ratios $\frac{P_x(s_i|\pi)}{P_x(s_j|\pi)}$ in region 1. Then, the probabilities can be recovered:

$$\frac{P_x(s_1|\pi)}{P_x(s_2|\pi)} = c_1; \text{ and } \frac{P_x(s_3|\pi)}{P_x(s_2|\pi)} = c_2 \Rightarrow P_x(s_2|\pi) = \frac{1}{1 + c_1 + c_2}.$$

We can identify probabilities in other regions in a similar manner. Region 1 will have regions 2 and 3 as its neighbors. As long as $\alpha \neq 0.5$, we know that by crossing a border between neighboring regions, either $P_\alpha(\cdot|A_2)$ or $P_\alpha(\cdot|\pi)$ changes. If it is $P_\alpha(\cdot|\pi)$, then all probabilities change; if it is $P_\alpha(\cdot|A_2)$, then only probabilities of states related to one of the events will change. By looking at regions 1 and 2, we can see that $P(s_1|\pi)$ stays the same, while other probabilities change. This implies that the subjective partition is $\pi = \{s_1, A_2\}$, where $A_2 = \{s_2, s_3\}$.

From the fact that $P_x(s_1|\pi) = P_\alpha(s_1|\pi)$ in region 1, we can recover $P_\alpha(A_2|\pi) = 1 - P_x(s_1|\pi)$; then, $P_\alpha(s_i|A_2) = \frac{P_x(s_i|\pi)}{P_\alpha(A_2|\pi)}$. Thus, all $P_\alpha(\cdot|\pi)$ and $P_\alpha(\cdot|A_2)$ in all regions can be identified. Regions 1 and 2 are from different sides of the same conditional stage boundary, meaning that $P_{\max}(\cdot)$ and $P_{\min}(\cdot)$ are the same but exchange places in order to obtain $P_\alpha(\cdot|A_2)$. Thus, we know that

$$\alpha x + (1-\alpha)y = \alpha\phi + 1 - \alpha = c_1$$
$$\alpha y + (1-\alpha)x = \alpha + (1-\alpha)\phi = c_2.$$

where x and y are vertices of the set $\mathcal{P}(A_2)$ that denote boundaries for $P(s_2|A_2)$. Note that it is a system of two equations and three unknowns. Even though a similar system of equations can be obtained from $P_\alpha(A_2|\pi)$, in total, there will be four equations and five unknowns. Thus, unique identification of the parameter α and sets of priors is not possible. However, we can achieve partial identification of parameters. The above system of equations suggests that $x+y = c_1 + c_2 = 1 + \phi > 1$. Thus, by taking into account that $0 \leq x < y \leq 1$, the biggest possible set of priors that satisfies the above system of equations is $P(s_2|A_2) \in [c_1 + c_2 - 1; 1] = [\phi; 1]$. Moreover, other possible sets of priors belong to $[\phi; 1]$.

We can use the system of equations in order to obtain partial identification of the parameter α. Note that $\alpha = \frac{c_2 - x}{y - x}$ is a monotone function of x. Because $c_2 > c_1$ (due to $\alpha > 0.5$), the function is strictly increasing in x. Thus, the smallest possible value of α is obtained when $x = \phi$ – i.e., $\alpha_{\min} = \frac{c_2-\phi}{1-\phi} = \alpha$, which is the true value of the parameter. Thus, the interval is $\mathcal{A} = [\alpha; 1)$.

The same idea can be applied for partial identification of the ex-ante set of priors $\mathcal{P}(\pi)$. In this example, intervals obtained from both problems for α coincide. However, generally, one needs to take into account all available equations.

2.1.4 Identification

Thus, the agent wants to buy a portfolio of securities and aggregates states of the world into some partition. The purpose of this section is to identify the agent's subjective partition and, when possible, priors from choices of Arrow assets and their prices. In order to do so, we generalize the above non-parametric example; however, the idea behind the method stays the same.

A central assumption for identification from choices in this paper is polytope priors that allow for the agent's choice behavior to be described by the SEU in different regions of the act space with different probabilities across regions. Moreover, this act space separation is subjective-partition-unique for the strict α-MEU stage model.

Assumption 2.1. *Sets of priors $\mathcal{P}(A)$ and $\mathcal{P}(\pi)$ for any $A \in \pi$ are closed convex non-empty polytopes.*

Variation of probabilities across regions creates kinks in indifference curves, as can be seen in Figure 2.1. However, we do not observe choices from indifference curves completely. Our data are choices of bundles of Arrow securities together with their prices. Thus, if indifference curves are not convex (as in Figure 2.1), then we can recover only the intersection of an indifference curve with the convex hull of its upper counter set, as shown in Figure 2.2.

In order to achieve convexity of indifference curves and have the optimality condition hold in each region, we assume the following:

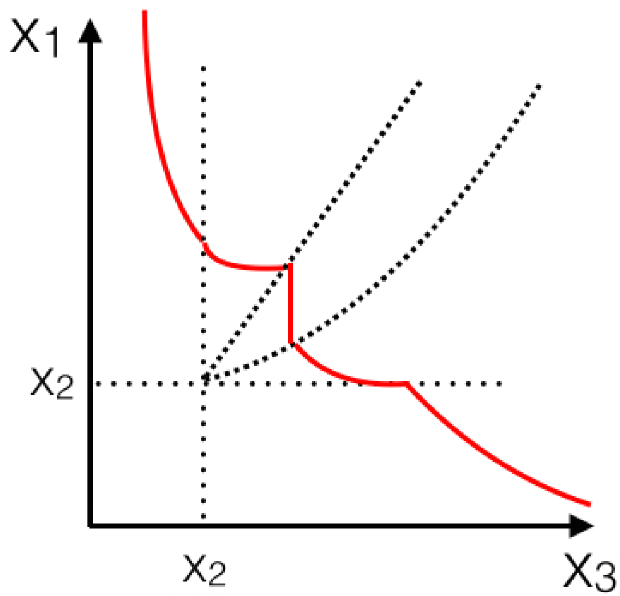

Figure 2.1: Typical indifference curve

Assumption 2.2. *Utility function $u(\cdot)$ is concave and differentiable.*

Only one connected set of indifferent bundles will be recovered in each observable region because indifference curves are convex there. Unfortunately, there is no guarantee that each region is represented by a set of choices. As Figure 2.2 shows, two middle regions are missing in the observations. When behavior in some region is not observed, the identification becomes impossible, which prompts the following assumption:

Assumption 2.3. *There exist prices and incomes such that more than one bundle is chosen in each region.*

Assumptions 1.1, 2.1, 2.2 and 2.3 help us identify the utility function up to affine transformation in each separate region. However, for the identification of the probabilities and the subjective partition, we need to connect these recovered functions with each other. In order to avoid this problem, we make another assumption:

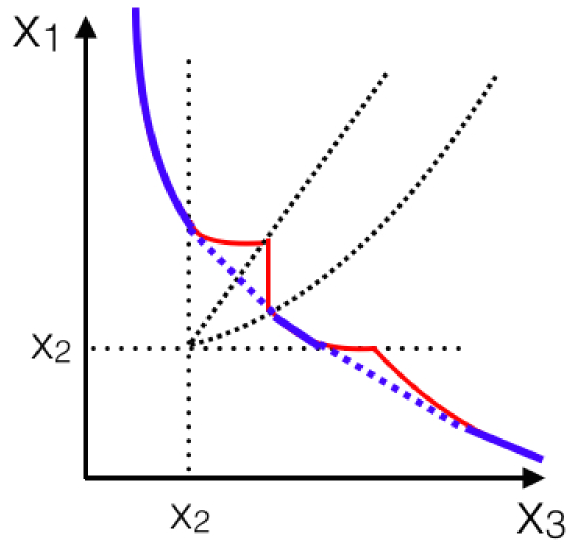

Figure 2.2: Intersection of IC with its convex hull

Assumption 2.4. *Observed choices include payoffs from the whole compact subset X_S of the support X of $u(\cdot)$.*

Proposition 2.1. *If assumptions 1.1 and 2.1–2.4 hold, and $0 < \alpha < 1$, then the following can be uniquely identified from choice behavior:*

1. *utility $u(\cdot)$ up to affine transformation on the set X_S;*

2. *state aggregation π;*

3. *$P_x(s|\pi)$, $P_{\alpha,x}(s|A)$ and $P_{\alpha,x}(A|\pi)$ in each region.*

   In addition, α and sets of priors, $\mathcal{P}(A)$ and $\mathcal{P}(\pi)$, can be partially identified:

4. *there exists an interval $\mathcal{A} \subset (0,1)$ such that $\alpha \in \mathcal{A}$:*

   (a) *if $\alpha > 0.5$, then $\mathcal{A} \subset (0.5, 1)$;*

   (b) *if $\alpha < 0.5$, then $\mathcal{A} \subset (0, 0.5)$;*

5. *there exist sets of priors $\widetilde{\mathcal{P}}(A)$ and $\widetilde{\mathcal{P}}(\pi)$ such that $\mathcal{P}(A) \subseteq \widetilde{\mathcal{P}}(A)$ and $\mathcal{P}(\pi) \subseteq \widetilde{\mathcal{P}}(\pi)$.*

The proof of the above proposition suggests an identification method to proceed. First, the researcher needs to determine regions of the act space in which the agent evaluates acts according to SEU with probabilities that are different across the regions. Then, utility, together with the region-specific probabilities, can be identified. And, finally, the way that probabilities change between neighboring regions suggests the subjective partition and ambiguity attitude. However, the actual parameter of ambiguity aversion α and the sets of ex-ante and conditional priors are not identified due to a non-unique representation of preferences (this problem is generally present in the α-MEU model). However, as non-parametric example shows, partial identification can be achieved. In order to obtain unique identification, some normalization conditions are required.

Note that if we apply the identification method to the MEU model, then the identified subjective partition might not be unique due to the reasons described in Chapter 1. However, the set of priors is identified in this case because the value of ambiguity aversion parameter is known – i.e., $\alpha = 1$. If we obtain that the subjective partition π is different from $\pi_0 = \Omega$, then it means that the subjective partition is not unique. In this case, the agent with the subjective partition π_0 and π-rectangular priors will demonstrate identical behavior. However, if the identification method provides the subjective partition π_0, then we know that it is unique, and the agent does not aggregate states.

2.2 Testable Restrictions

This section provides a set of bilinear Afriat inequalities that allow us to test the subjective partition model with a finite dataset. Since the system of inequalities is bilinear, the Tarski-Seidenberg algorithm can be applied to determine whether the solution exists.

An example of a dataset that could be used for testing the model is yearly individual data on car insurance deductible choices and income. The insurance plan should have at least two deductibles that represent states of the world of different occurrences. We also need to observe the whole choice set of insurance plans with all prices, and the choice set should be constant over time. Another requirement on the structure of the insurance plan must hold: the insurance price must be linear in its deductible prices – a property that is easy to verify in the data. The insurance plan with deductibles can be treated as a portfolio of Arrow assets, as discussed in Section 2.1.1. However, in order to recover all Arrow prices and the amount of Arrow securities, we also need to have data on individual income and losses in each state.

A dataset is a finite collection of pairs $(x^i, p^i)_{i=1}^N \in \mathbb{R}_+^\Omega \times \mathbb{R}_{++}^\Omega$, where N is a number of observations. We define a budget set as $B^i = \{x \in \mathbb{R}_+^\Omega : p^i x \leq p^i x^i\}$ and a boundary of the budget set as $\partial B^i = \{x \in \mathbb{R}_+^\Omega : p^i x = p^i x^i\}$. Also, $\mathcal{X} = \{x \in \mathbb{R}_+ : x = x_s^i \text{ for some } t, s\} \cup \{0\}$ is observed at some state consumption set together with 0; and lattice $\mathcal{L} = \mathcal{X}^\Omega$ is its product over states.

Definition 2.1. *A dataset $(x^i, p^i)_{i=1}^N$ is α-MEU π state aggregation rational (π-rational) if there are sets of priors $\mathcal{P}(\pi)$ and $\mathcal{P}(A)$, $A \in \pi$, a parameter α, and a continuous and increasing function $u : \mathbb{R}_+ \to \mathbb{R}$ such that for all i*

$$y \in B^i \Rightarrow V(y|\pi) \leq V(x^i|\pi).$$

Proposition 2.2. *A dataset is π-rational if and only if for each element of lattice $x \in \mathcal{L}$, there exist non-negative numbers $\mu_s(x)$, $\mu_{\alpha;A}(x)$, $\mu_{\alpha;s}(x)$, $q_{1;s}^A(x)$, $q_{2;s}^A(x)$, $w_{1,A}(x)$, $w_{2;A}(x)$, $0 \leq a \leq 1$, and an increasing function $\bar{u} : \mathcal{X} \to \mathbb{R}$ such that*

$$\sum_{s \in \Omega} \mu_s(x) = \sum_{A \in \pi} \mu_{\alpha;A}(x) = \sum_{s \in \Omega} \mu_{\alpha;s}(x) = 1$$

$$\sum_{s \in A} q_{1;s}^A(x) = \sum_{s \in A} q_{2;s}^A(x) = \sum_{A \in \pi} w_{1,A}(x) = \sum_{A \in \pi} w_{2,A}(x) = 1$$

$$\mu(x^i)\bar{u}(x^i) \geq \mu(x)\bar{u}(x) \ \forall x \in \mathcal{L} \cap B^i$$

$$\mu(x^i)\bar{u}(x^i) > \mu(x)\bar{u}(x) \ \forall x \in \mathcal{L} \cap (B^i \setminus \partial B^i)$$

$$\mu_s(x) = \mu_{\alpha;A}(x)\mu_{\alpha;s}(x), s \in A, x \in \mathcal{L}$$

$$\mu_{\alpha;s}(x) = aq_{1;s}(x) + (1-a)q_{2;s}(x), x \in \mathcal{L}$$

$$\mu_{\alpha;A}(x) = aw_{1;A}(x) + (1-a)w_{2;A}(x), x \in \mathcal{L}$$

$$\sum_{s \in A} q^A_{1;s}(x)\bar{u}(x_s) \leq \sum_{s \in A} q^A_{t;s}(x')\bar{u}(x_s), \ \forall x, x' \in \mathcal{L}, t = 1, 2$$

$$\sum_{s \in A} q_{2;s}(x)\bar{u}(x_s) \geq \sum_{s \in A} q^A_{t;s}(x')\bar{u}(x_s), \ \forall x, x' \in \mathcal{L}, t = 1, 2$$

$$\sum_{A \in \pi} w_{1;A}(x) \sum_{s \in A} \mu_{\alpha;s}(x)\bar{u}(x_s) \leq \sum_{A \in \pi} w_{t;A}(x') \sum_{s \in A} \mu_{\alpha;s}(x)\bar{u}(x_s), \ \forall x, x' \in \mathcal{L}, t = 1, 2$$

$$\sum_{A \in \pi} w_{2;A}(x) \sum_{s \in A} \mu_{\alpha;s}(x)\bar{u}(x_s) \geq \sum_{A \in \pi} w_{t;A}(x') \sum_{s \in A} \mu_{\alpha;s}(x)\bar{u}(x_s), \ \forall x, x' \in \mathcal{L}, t = 1, 2.$$

For the purposes of the proof of Proposition 2.2, we derive conditions for the sets of priors on the lattice and apply a result from Polisson, Quah and Renou (2015), which states that if there exists an increasing utility function on the lattice, then it can be extended to a continuous increasing utility function on the whole support.

Example 2.2. To simplify the example as much as possible, we assume that the set of priors is the whole simplex and $\Omega = \{s_1, s_2, s_3\}$. Suppose that we are interested in testing whether the agent does not aggregate states – i.e., $\pi = \Omega$. With the whole simplex priors, the value of each bundle x such that $x_i > x_j > x_k$ will be attained by using probabilities $\mu_i = 1 - \alpha$, $\mu_j = 0$, and $\mu_k = \alpha$. Consider the agent who chooses an allocation $(3, 1, 3)$ with prices $(2, 1, 2)$. Note that allocation $(2, 2, 3)$ is also available and is strictly inside the budget set. Thus, $(3, 1, 3) \succ (2, 2, 3)$. However, $V((2, 2, 3)|\Omega) \geq V((3, 1, 3)|\Omega)$ if the utility function is increasing:

$$V((3,1,3)|\Omega) = (1-\alpha)u(3) + \alpha u(1)$$

$$V((2,2,3)|\Omega) = (1-\alpha)u(3) + \alpha u(2).$$

Thus, $(3,1,3)$ cannot be chosen if the agent has the whole simplex priors and does not aggregate states.

Next, consider a subjective partition $\pi = \{\{s_1, s_2\}, s_3\}$. In this case, bundles' values are

$$V((3,1,3)|\pi) = (1-\alpha)u(3) + \alpha\left((1-\alpha)u(3) + \alpha u(1)\right)$$

$$V((2,2,3)|\pi) = (1-\alpha)u(3) + \alpha u(2).$$

Thus, $(3,1,3) \succ (2,2,3)$ if and only if $(1-\alpha)u(3)+\alpha u(1) > u(2)$, which is generally possible with some increasing function $u(\cdot)$. However, note that $u(3) - u(2) > \frac{\alpha}{1-\alpha}(u(2)-u(1))$. The last expression implies that if the agent is ambiguity-averse, then her utility function is not concave. Thus, if we require both ambiguity and risk aversion, then this data point rejects the subjective partition π with the whole simplex priors.

2.3 Conclusions

We have explored how the subjective partition might be identified from the complete market choices and prices. We have also shown that under some assumptions on observable data, the results can be applied to deductible insurance plans. In addition, we have offered testable restrictions of our model that might be used on a dataset containing insurance deductible choices and income.

2.4 Appendix to Chapter 2

Lemma 2.1. *For any act $x \in \mathcal{F}$ and partition π, the value $V(x|\pi)$ can be written as follows:*

$$V(x|\pi) = \sum_s P_x(s|\pi)u(x(s)).$$

Proof. From the representation theorem:

$$V(x|A) = \alpha \min_{P(\cdot|A)\in\mathcal{P}(A)} \sum_{s\in A} u(x(s))P(s|A) + (1-\alpha) \max_{P(\cdot|A)\in\mathcal{P}(A)} \sum_{s\in A} u(x(s))P(s|A) =$$
$$= \sum_{s\in A} \left(\alpha P_{\min x}(s|A) + (1-\alpha)P_{\max x}(s|A)\right) u(x(s)) = \sum_{s\in A} P_{\alpha;x}(s|A)u(x(s))$$

$$V(x|\pi) = \alpha \min_{P\in\mathcal{P}(\pi)} \sum_{A\in\pi} V(x|A)P(A|\pi) + (1-\alpha) \max_{P\in\mathcal{P}(\pi)} \sum_{A\in\pi} V(x|A)P(A|\pi) =$$
$$= \sum_{A\in\pi} \left(\alpha P_{\min x}(A|\pi) + (1-\alpha)P_{\max x}(A|\pi)\right) V(x|A) = \sum_{A\in\pi} P_{\alpha;x}(A|\pi)V(x|A) =$$
$$= \sum_{A\in\pi} P_{\alpha;x}(A|\pi) \sum_{s\in A} P_{\alpha;x}(s|A)u(x(s)) = \sum_s P_{\alpha;x}(A|\pi)P_{\alpha;x}(s|A)u(x(s)),$$

where $s \in A \in \pi$. $\square$

Lemma 2.2. *If assumptions 1.1 and 2.1 hold, then the act space $\mathcal{F}$ can be split into a finite number of regions with $P_x(s|\mathcal{A})$ constant inside each of them.*

Proof. Consider the conditional stage when event A is observed. If event A has only one state s, then $\mathcal{P}(A)$ is a singleton, and $P(s|A) = 1$ for any act. Thus, the act space is not being split by this event.

Now, suppose that event A consists of two or more states. Remember that

$P_{\alpha;x}(s|A) = \alpha P_{\min x}(s|A) + (1-\alpha)P_{\max x}(s|A)$, where

$$P_{\min x}(\cdot|A) = \underset{P(\cdot|A)\in\mathcal{P}(A)}{\operatorname{argmin}} \sum_{s\in A} P(s|A)u(x(s))$$
$$P_{\max x}(\cdot|A) = \underset{P(\cdot|A)\in\mathcal{P}(A)}{\operatorname{argmax}} \sum_{s\in A} P(s|A)u(x(s)).$$

Since $\mathcal{P}(A)$ is a closed convex bounded polytope, $P_{\min x}(s|A)$ and $P_{\max x}(s|A)$ are solutions to the linear-programming problem. As a result, they lie on the boundary of the set $\mathcal{P}(A)$ and depend on the direction of the vector $(u(x(s_1)), ..., u(x(s_k)))$, where $A = \{s_1, ..., s_k\}$. Moreover, when the solution is at the corner of $\mathcal{P}(A)$, it will be the same for all vectors $(u(x(s_1)), ..., u(x(s_k)))$ in between two hyperplanes that created the corner in the conditional set of priors. Because the number of corners in a closed convex polytope is finite, there is a finite number of regions in the act space where $P_{\min x}(\cdot|A)$ is constant. The same is also true for $P_{\max x}(\cdot|A)$, implying it for $P_{\alpha;x}(\cdot|A)$.

The same argument can be applied to the ex-ante stage and $P_{\alpha;x}(\cdot|\mathcal{A})$. □

Lemma 2.3. *If assumptions 1.1 and 2.1 hold,* $0 < \alpha < 1$*, and two choice sets are generated by preferences represented by partitions* $\pi_1 \neq \pi_2$*, then the obtained sets of regions in the act space differ.*

Proof. At the conditional stage given event $A = \{s_1, ..., s_n\}$, the agent's problem of finding probabilities can be reformulated as follows:

$$\sum_{i=1}^{n-1}(u(x(s_i)) - u(x(s_n)))p_i \to \max / \min$$
$$\text{s.t. } p \in \mathcal{P}(A).$$

The direction of utility growth is defined by the vector

$$\mathbf{u} = (u(x(s_1)) - u(x(s_n)), ..., u(x(s_{n-1})) - u(x(s_n)))'.$$

$\mathcal{P}(A)$ is a polytope, so it is constructed by the intersection of several hyperplanes. Each hyperplane has two (opposite) normal unit vectors. Thus, $\mathcal{P}(A)$ creates a set of unit vectors $W(\mathcal{P}(A))$ perpendicular to the hyperplanes of $\mathcal{P}(A)$. $W(\mathcal{P}(A))$ splits the vector space it belongs to into different regions. When $\mathbf{u}$ lies in between $n-1$ neighbor vectors from $W(\mathcal{P}(A))$, the solution from $\mathcal{P}(A)$ is the same for all acts with vector $\mathbf{u}$ in this region. Thus, the act space gets separated by boundaries where $\mathbf{u}$ has the same direction as some $v = (v_1, ...v_{n-1})' \in W(\mathcal{P}(A))$. Then, the boundary can be described by the following system of equations:

$$
\begin{aligned}
u(x(s_1)) - u(x(s_n)) &= \lambda v_1 \\
&\ldots \\
u(x(s_{n-1})) - u(x(s_n)) &= \lambda v_{n-1} \\
\lambda &= \sqrt{\sum_{i=1}^{n-1} (u(x(s_i)) - u(x(s_n)))^2}.
\end{aligned}
$$

Due to the symmetry of $W(\mathcal{P}(A))$, for each $v \in W(\mathcal{P}(A))$, all conditional stage boundaries will be defined by the following condition:

$$
u(x(s_i)) = \frac{v_i}{v_j} u(x(s_j)) + \left(1 - \frac{v_i}{v_j}\right) u(x(s_n)) \text{ for any } i = \overline{1, n-2},
$$

where s_j is a state for which $v_j \neq 0$ for the given vector (such state always exists). Thus, notice that the conditional stage boundaries depend on payoffs in the states in event A only.

One potential problem might arise here. Suppose that there is another partition that includes two events B and C such that $A = B \cup C$. Is it possible that the conditional boundaries produced by A are the same as those produced by B and C? In order for this to happen, each boundary must split states into two groups. It is possible only if $u(x(s_i)) - u(x(s_n)) = \lambda v$ for all i such that $s_i \in B$, and $u(x(s_j)) - u(x(s_n)) = 0$ for all j such that $s_j \in C$. This must hold for all conditional boundaries of event A. However, it implies that either $\mathcal{P}(A)$ is not a polytope or it is not full-dimensional. Thus, under assumptions 1.1 and 2.1, the problem does not arise.

By applying the same procedure to the ex-ante stage, the ex-ante boundary can be obtained, too:

$$V(x|A_i) = \frac{w_i}{w_j} V(x|A_j) + \left(1 - \frac{w_i}{w_j}\right) V(x|A_k) \text{ for any } i = \overline{1, k-1},$$

where $w \in W(\mathcal{P}(\pi))$, $\pi = \{A_1, ..., A_k\}$ and A_j is an event for which $w_j \neq 0$. Note that

$$V(x|A) = \sum_{s \in A} P_{\alpha;x}(s|A) u(x(s)),$$

and it implies that the ex-ante boundary depends on all states in Ω.

Thus, if all boundaries are the same for two different people, it means that the conditional stage boundaries must coincide. This occurs because the conditional stage boundary depends only on payoffs in the states of the specific event, while the ex-ante stage boundary includes all states. They can coincide only between π_0 and π^*. Now, note that the conditional stage boundaries imply the partition. Thus, both agents have the same subjective partition. $\square$

Proof of Proposition 2.1. Assumption 2.3 guarantees that we observe each region. Regions are separated from each other either by corner points or by holes. Corner points are those bundles that are chosen under the range of price ratios. Holes are bundles that are never chosen under any combination of price ratios and income. Inside some region R, the value functional is SEU with probabilities $P_x(s|\pi)$ of each state s. Then, inside the region the optimality condition holds:

$$\frac{p_j}{p_i} = \frac{P_x(s_j|\pi)}{P_x(s_i|\pi)} \frac{u'(x(s_j))}{u'(x(s_i))}.$$

We fix some value of $x(s_i) = \bar{x}$ and take a look at the values of $x(s_j)$ such that the bundle is still in region R. Then, observed $x(s_j)$ will belong to some closed interval $[a(\bar{x}), b(\bar{x})]$ ($b(\bar{x})$ might be $+\infty$). Now, note that for any $x_1, x_2 \in [a(\bar{x}), b(\bar{x})]$, there exist prices p_i^1, p_i^2, p_j^1 and p_j^2 such that:

$$\frac{p_j^1}{p_i^1} = \frac{P_f(s_j|\pi)}{P_f(s_i|\pi)} \frac{u'(x_1)}{u'(\bar{x})}$$

$$\frac{p_j^2}{p_i^2} = \frac{P_f(s_j|\pi)}{P_f(s_i|\pi)} \frac{u'(x_2)}{u'(\bar{x})}$$

$$\Rightarrow \frac{u'(x_1)}{u'(x_2)} = \frac{p_j^1}{p_i^1} \frac{p_i^2}{p_j^2}.$$

Thus, the ratio $\frac{u'(x_1)}{u'(x_2)}$ is identified in the interval $[a(\bar{x}), b(\bar{x})]$. By analogy, such a ratio is identified in all intervals that are observed in different states.

If $\frac{u'(x_1)}{u'(x_2)}$ is identified in two different intervals $[a_1, b_1]$ and $[a_2, b_2]$ and $a_1 \leq a_2 \leq b_1 \leq b_2$, then it is also identified in $[a_1, b_2]$: for any $x_1 \in [a_1, b_1]$, $x_2 \in [a_2, b_2]$ and $z \in [a_2, b_1]$, the following holds

$$\frac{u'(x_1)}{u'(x_2)} = \frac{u'(x_1)/u'(z)}{u'(x_2)/u'(z)}.$$

The only potential problem with identification of $\frac{u'(x_1)}{u'(x_2)}$ might arise if intervals do not intersect. However, Assumption 4.4 rules this problem out. Thus, for any observable payoffs $x_1, x_2 \in X_S$, the ratio $\frac{u'(x_1)}{u'(x_2)}$ is identified. This implies that the utility function $u(\cdot)$ is identified up to affine transformation on X_S.

From the results of Lemma 2.3, all boundaries are classified by stages. If two regions are separated by the conditional stage boundary related to event A, then $P_{\alpha,x}(A|\pi)$ for all $A \in \pi$ and $P_{\alpha,x}(s|B)$ for all $B \in \pi \backslash A$ stay the same, while only $P_{\alpha,x}(s|A)$ change. If regions are separated by the ex-ante stage boundary, then all probabilities change. By looking at which probabilities change from one region to another, we can identify the partition.

By denoting different neighboring regions through x_i and x_j, one can obtain:

$$\frac{P_{\alpha;x_i}(s|A)}{P_{\alpha;x_j}(s|A)} = \frac{P_{x_i}(s|\pi)}{P_{x_j}(s|\pi)}.$$

If k is a number of different conditional stage regions for event A, and n is a number of states in event A, then due to full-dimensionality, $k \geq n$. Because all ratios $\frac{P_{\alpha;x_i}(s|A)}{P_{\alpha;x_j}(s|A)}$ for conditional stage regions are known and $\sum_{s \in A} P_{\alpha;x_i}(s|A) = 1$, we have kn unknowns and $n(k-1)+k$ equations. Because $n(k-1)+k \geq kn$, all $P_{\alpha;x_i}(s|A)$ are identified for all i.

Identification of $P_{\alpha;x}(A|\pi)$ is trivial now:

$$P_{\alpha;x}(A|\pi) = \frac{P_x(s|\pi)}{P_{\alpha;x}(s|A)}.$$

$P_{\min x}(\cdot)$ and $P_{\max x}(\cdot)$ are solutions to the linear programming problem at each stage. However, in symmetric regions, $P_{\min x}(\cdot)$ and $P_{\max x}(\cdot)$ will exchange places due to the direction of utility growth, which will be the opposite. Thus, if we denote each pair of these probabilities as (x_s, y_s) for conditional priors, we will obtain two equations:

$$\alpha x_s + (1-\alpha) y_s = P_{\alpha;x_1}(s|A)$$
$$\alpha y_s + (1-\alpha) x_s = P_{\alpha;x_2}(s|A),$$

where x_1 and x_2 are acts from symmetric regions. We can write down such systems for all states and all symmetric regions for conditional probabilities. In addition, we can do the same for the ex-ante probabilities:

$$\alpha x_A + (1-\alpha) y_A = P_{\alpha;x_1}(A|\pi)$$
$$\alpha y_A + (1-\alpha) x_A = P_{\alpha;x_2}(A|\pi).$$

In the final system of equations, there is always one unknown variable more than the number of equations; thus, the system does not have a unique solution.

First, note that $\alpha = 0.5$ if and only if $P_{\alpha;x_1}(s|A) = P_{\alpha;x_2}(s|A)$ and $P_{\alpha;x_1}(A|\pi) = P_{\alpha;x_2}(A|\pi)$ for all symmetric regions, states, and events. Thus, we can always identify whether $\alpha = 0.5$ or $\alpha \neq 0.5$. In addition, it is possible to identify whether the agent is ambiguity-averse/-loving. If we fix bundle payoffs at all states except $s_i, s_j \in A \in \pi$ and move along one IC from high x_i and low x_j to low x_i and high x_j inside one ex-ante region, then, if $\alpha > 0.5$, $\frac{P_{\alpha;x}(s_j|A)}{P_{\alpha;x}(s_i|A)}$ will decrease, while for $\alpha < 0.5$ the probability ratio will increase.

Consider the case of ambiguity aversion. Note that $\alpha = 1$ is always possible. In addition, there always exists the lowest possible value, $\underline{\alpha} > 0.5$, for α that satisfies the system of equations. Then, due to the bilinearity of the system, $\mathcal{A} = [\underline{\alpha}, 1)$. In addition, the greater the value of α, the smaller are the sets of priors $\mathcal{P}(A)$ and $\mathcal{P}(\pi)$. Thus, the sets of priors $\widetilde{\mathcal{P}}(A)$ and $\widetilde{\mathcal{P}}(\pi)$ will correspond to $\underline{\alpha}$ in the above system of equations.

By analogy, if $\alpha < 0.5$, then there exists an upper bound $\bar{\alpha}$ such that $\mathcal{A} = (0, \bar{\alpha}]$, and $\widetilde{\mathcal{P}}(A)$ and $\widetilde{\mathcal{P}}(\pi)$ are sets of priors that correspond to $\bar{\alpha}$.

□

Proof of Proposition 2.2. The proof is similar to the proof of Proposition 1 in Polisson, Quah, and Renou (2015).

For sufficiency, we first denote the following sets: for each $A \in \pi$, Q_A is a convex hull of $\{q_t^A(x)\}_{t=1,2}$; W is a convex hull of $\{w_t(x)\}_{t=1,2}$ for all $x \in \mathcal{L}$. Now we show that $q_1^A(x)\bar{u}(x) = \min_{q \in Q_A} q\bar{u}(x)$. Suppose that it's not true – i.e., there exists $\tilde{x} \in \mathcal{L}$ and $q \in Q_A$ such that $q\bar{u}(\tilde{x}) < q_1^A(\tilde{x})\bar{u}(\tilde{x})$. Q_A is a convex hull of $\{q_t^A(x)\}_{t=1,2}$; thus, q is a convex combination of $\{q_t^A(x)\}_{t=1,2}$. This means that there exists $x' \in \mathcal{L}$ such that $q_t^A(x')\bar{u}(\tilde{x}) < q_1^A(\tilde{x})\bar{u}(\tilde{x})$ – i.e., a contradiction. By analogy, one can show that $q_2^A(x)\bar{u}(x) = \max_{q \in Q_A} q\bar{u}(x)$. Thus, $\mu_\alpha(x)\bar{u}(x) = a \min_{q \in Q_A} q\bar{u}(x) + (1-a) \max_{q \in Q_A} q\bar{u}(x)$.

Note, also, that

$$\sum_{A\in\pi} w_{1;A}(x) \sum_{s\in A} \mu_{\alpha;s}(x)\bar{u}(x_s) = \min_{w\in W} \sum_{A\in\pi} w \sum_{s\in A} \mu_{\alpha;s}(x)\bar{u}(x_s)$$

$$\sum_{A\in\pi} w_{2;A}(x) \sum_{s\in A} \mu_{\alpha;s}(x)\bar{u}(x_s) = \max_{w\in W} \sum_{A\in\pi} w \sum_{s\in A} \mu_{\alpha;s}(x)\bar{u}(x_s).$$

Thus,

$$\mu(x)\bar{u}(x) = a \min_{w\in W} \sum_{A\in\pi} w \sum_{s\in A} \mu_{\alpha;s}(x)\bar{u}(x_s) + (1-a) \max_{w\in W} \sum_{A\in\pi} w \sum_{s\in A} \mu_{\alpha;s}(x)\bar{u}(x_s).$$

Now, we define $\phi : \mathbb{R}^{\Omega}_{+} \to \mathbb{R}$ as

$$\begin{aligned}\phi(\bar{u}) &= a \min_{w\in W} w \circ \left(a \min_{q\in Q_A} q\bar{u} + (1-a) \max_{q\in Q_A} q\bar{u} \right) \\ &+ (1-a) \max_{w\in W} w \circ \left(a \min_{q\in Q_A} q\bar{u} + (1-a) \max_{q\in Q_A} q\bar{u} \right).\end{aligned}$$

We can apply Theorem 1 from Polisson, Quah and Renou (2015) to guarantee the existence of u that rationalizes the dataset.

Note that necessity follows straightforward from the model.

□

CHAPTER 3

Electoral Model and Its Non-Parametric Estimation

Some people vote and others do not. Absent voters make it difficult to evaluate political preferences in the entire population. Thus, any small change in policy or characteristics related to voters' choice might bring unexpected consequences. An example is an unusual spike in African-American turnout in the 2008 U.S. presidential elections due to the Obama candidacy.[1] In order to be able to study and predict changes in voting behavior of the entire population, we offer a structural model that allows for derivation of joint distribution of turnout and voter share. We show how to identify an unobservable joint distribution of costs of voting and preferences from observable joint distribution of turnout and voter share. In addition, we provide non-parametric estimators of identified functions. All estimators are consistent and asymptotically normal.

This paper supposes that a voter needs to decide between two candidates. Similarly to the probabilistic voting model (Lindbeck and Weibull (1987), Persson and Tabellini (2000)), we assume that political preferences over candidates of an individual can be separated into several components: personal, local and regional. In addition, the person might find it too costly to participate in elections (for example, to wait in line in a polling station) if it makes little difference who wins. Thus, people with preferences close to indifference between candidates will abstain from elections. In order to identify the preferences in the whole population,

[1] *2008 Surge in Black Voters Nearly Erased Racial Gap*, NY Times, July 20, 2009.

we impose a structure that includes (1) additivity of preference components, (2) independence of personal and regional components on other characteristics, and, (3) linearity of the regional component in regional characteristics.

Political preferences have been studied by economists before. Several papers estimate preferences of the voters relying on the spatial model (Downs (1957), Riker and Ordeshook (1968), Hinich and Munger (1994)) when building the structure. This model assumes that each voter has a preferred policy and evaluates candidates based on the distances between proposed and preferred policies. Degan (2007) develops and estimates a dynamic spatial model of voting that allows for separation between preferences for ideology and candidates' attractiveness. Kernell (2009) obtains ideological distribution within districts by using data from multiple elections and applying a least squared error model and a Bayesian model. Merlo and de Paula (2015) provide non-parametric estimation of ideological preferences. However, these papers do not analyze turnout in the estimation and, thus, the results are valid only for the portion of voters that participated in the elections.

Another direction of the research focuses on estimation of voter turnout. Coate and Conlin (2004) structurally estimate a group rule-utilitarian model of voter turnout in Texas liquor referenda, and show that ethical-voter models[2] fit the data. McMurray (2012) proposes a model of equilibrium turnout where citizens vote if they have a reasonable level of expertise and finds support of the model in the data.

The above papers study the questions of turnout and voter preferences separately. Degan and Merlo (2011) combine both the spatial model of voting and uncertain-voter model[3] for turnout to study multiple elections. The main problem

[2]Models where a voter understands that she is not pivotal, but morality makes her participate in elections.

[3]A uninformed voter leaves the decision to the informed part of population.

with the estimation of any version of the spatial model is that it assumes that voters only care about ideology and it requires construction of the ideology metric. The idea of the ideology index is good in theory, however, in practice, is very subjective. Moreover, differently constructed indices might drive any estimation results. Additionally, voter preferences are often driven by personal characteristics of the candidate, not just political ideology. For example, it is difficult to explain the popularity of Donald Trump by his political agenda.

This paper offers a different perspective on the preference estimation. We do not discuss the way the preferences were formed and variables they depend on. The preferences in our model are exogenous. We are interested in understanding various components (personal, local and regional) of the preferences and voting costs. In addition, we study a single election.

Notice that exogeneity of preferences accounts for different ways of preference formation. For example, if voters are ideological as in the spatial model, then the preference distribution is the distribution of relative distance from the candidates. If one adds additional assumptions about the relative distance metric and uses the data from multiple elections, then it will be possible to identify the ideological distribution of the population. Our model also includes the uncertain-voter model – a more confident voter derives higher utility from making the right choice, i.e., voting for a specific candidate. The personal preferences account for an asymmetry of information in this case. Thus, the preferences we identify can be treated as a "reduced form" obtained from different models.

First, we offer a model of voting that assumes that (1) preferences over candidates can be separated into personal, local and regional components, and, (2) a voter participates in election if the difference in utility from candidates is higher than the costs of voting. We suppose that the local component and costs of voting are the same for all people in one polling station and regional characteristics affect all population in the region in the same way. From these assumptions, we derive values of turnout and voter share in the population.

Second, we assume that (1) personal component of preferences is independent on other components and costs of voting, (2) the joint distribution of costs of voting and the local component is independent on the regional characteristics, and (3) the regional component is a linear function of observable regional characteristics. Under this set of assumptions, we show identification of the distribution of all components and costs of voting in the whole population.

Finally, we construct non-parametric kernel estimators based on the proposed identification strategy and show their consistency and asymptotic normality. The electoral data is often large enough and is well-suited for the non-parametric approach that might help to avoid unnecessary overparametrization and mistakes in the model specification.

3.1 Electoral model

There are two candidates, A and B, running for office. Each voter has preferences over candidates, defined in a way similar to the probabilistic voting model. Voter i in a polling station j in region K chooses candidate A over B if

$$\sigma_A^{ijK} + \delta_A^{jK} + \mu_A^K > \sigma_B^{ijK} + \delta_B^{jK} + \mu_B^K$$

where σ^{ijK} is a parameter of individual "pure" preferences towards the candidates, δ^{jK} is popularity of a candidate in the area of the polling station and is the same for one polling station j, μ^K is a regional effect in popularity of each candidate and it is a function of some observable characteristics of the region X_K, such as average income, level of education, share of old population, etc. Thus, $\mu^K = \tilde{h}(X_K)$.

In order to obtain the reduced form of the model, we define parameters of difference in preferences between candidates $\sigma^{ijK} = \sigma_B^{ijK} - \sigma_A^{ijK}$, $\delta^{jK} = \delta_B^{jK} - \delta_A^{jK}$

and $\mu^K = \mu_B^K - \mu_A^K = \tilde{h}_B(X_K) - \tilde{h}_A(X_K) \equiv h(X_K)$. Therefore, voter i in a polling station j in region K chooses candidate A over B if

$$\sigma^{ijK} + \delta^{jK} + h(X_K) < 0$$

where σ^{ijK} is personal "pure" preference for the candidate B, δ^{jK} and μ_K are polling station j and regional effects on preferences, correspondingly. Moreover, $\mu^K = h(X_K)$, where $h(\cdot)$ is continuous and monotone in all arguments function.

In order to include turnout in the model, we make voting costly. A voter chooses to participate in elections if the difference in her preferences from different candidates is higher than costs of participation:

$$|\sigma^{ijK} + \delta^{jK} + h(X_K)| \geq c^{jK}$$

Participation costs c^{jK} are random and they are the same for all voters in the same polling station, but different across different polling stations. Costs might represent the length of line to vote, the weather, etc. Such representation of participation implies that if the preferences of a voter are close to indifference between the candidates, then she does not attend elections. And, in contrast, if a person has very strong preferences towards one or the other candidate, then she comes to the polling station even when costs are high.

Individual "pure" preferences σ^{ijK} are independent identically distributed variables with density $g(\cdot)$ and cumulative distribution $G(\cdot)$. Costs of voting c^{jK} and local preferences δ^{jK} are independent identically distributed variables with joint density $f_{\delta,c}(\cdot,\cdot)$.

Next we introduce "swing voters", σ_A^{jK} and σ_B^{jK}, in every polling station j of region K, who are indifferent between participating and not participating in elections:

$$\sigma_A^{jK} = -\delta^{jK} - \mu^K - c^{jK}$$
$$\sigma_B^{jK} = \sigma_A^{jK} + 2c^{jK}$$

Notice that people with "pure" preferences $\sigma^{ijK} < \sigma_A^{jK}$ will vote for the candidate A, people with $\sigma^{ijK} > \sigma_B^{jK}$ will choose the candidate B, and everybody in between the swing voters will abstain from elections. As a result, the number of people who vote for A in a polling station j in region K is $n_A^{jK} = \int_{-\infty}^{\sigma_A^{jK}} dG(x) = G(\sigma_A^{jK})$. The same number for candidate B is $n_B^{jK} = \int_{\sigma_B^{jK}}^{+\infty} dG(x) = 1 - G(\sigma_B^{jK})$. Thus, turnout in the polling station is $\tau^{jK} = 1 - G(\sigma_B^{jK}) + G(\sigma_A^{jK})$. A's share of votes is $\pi_A^{jK} = \frac{n_A^{jK}}{n_A^{jK}+n_B^{jK}} = \frac{G(\sigma_A^{jK})}{1-G(\sigma_B^{jK})+G(\sigma_A^{jK})}$.

The only data available in any elections is voter share, π_A^{jK}, and turnout, τ^{jK}, across all polling stations j and all regions K. However, the following electoral variables can be easily recovered from the data:

$$G(\sigma_A^{jK}) = \pi_A^{jK}\tau^{jK}$$
$$G(\sigma_B^{jK}) = 1 - \tau^{jK} + G(\sigma_A^{jK}) = 1 - \tau^{jK} + \pi_A^{jK}\tau^{jK}$$

In everything that follows, we denote the observable electoral variables $Y = G(\sigma_A^{jK})$ and $Z = G(\sigma_B^{jK})$. Vector $X \in \mathbb{R}^L$ is a vector of L regional characteristics.

3.2 Identification of the model

This section discusses how to identify unobservable $g(\cdot)$, $h(\cdot)$ and $f_{\delta,c}(\cdot,\cdot)$ from observable joint distribution of Y and Z conditional on X.

Assumption 3.1. *Personal preferences σ is independent on X, δ and c, its sup-*

port in $\mathbb{R}$ is compact, and it has continuously differentiable density $g(\cdot)$, strictly increasing on the support cumulative distribution function $G(\cdot)$, and $E\sigma = 0$.

Assumption 3.2. *Local preferences δ and costs of voting c have continuously differentiable joint density $f_{\delta,c}(\cdot,\cdot)$, cumulative distribution function $F_{\delta,c}(\cdot,\cdot)$, and they are independent on regional characteristics X.*

The joint distribution of swing voters across polling stations in region K with characteristics X is:

$$F_{\sigma_A,\sigma_B|X}(y,z) = P(\sigma_A < y, \sigma_B < z|X) = P(-\delta - c - h(X) < y, -\delta + c - h(X) < z) =$$
$$= F_{-\delta-c,-\delta+c}\left(h(X) + y, h(X) + z\right)$$

Now we obtain the joint distribution of Y and Z conditional on X, $F_{Y,Z|X}(\cdot,\cdot)$, as follows:

$$F_{Y,Z|X}(y,z) = P(G(\sigma_A) < y, G(\sigma_B) < z|X) = P(\sigma_A < G^{-1}(y), \sigma_B < G^{-1}(z)) =$$
$$= F_{-\delta-c,-\delta+c}\left(h(X) + G^{-1}(y), h(X) + G^{-1}(z)\right) \tag{3.1}$$

Cumulative distribution function $F_{Y,Z|X}(\cdot,\cdot)$ of electoral variables conditional on the regional characteristics is observed. The right-hand side of the above equation is completely unknown and is to be identified. It is important to notice that the proposed model is not identified without additional assumptions. In order to see this, consider that there are true functions $F_{\delta,c}(\cdot)$, $G^{-1}(\cdot)$ and $h(\cdot)$. Suppose that $F_{-\delta-c,-\delta+c}(t_1,t_2) = \frac{t_1}{t_2}$. Then some other functions $\tilde{h}(X) = 2h(X)$ and $\tilde{G}^{-1}(y) = 2G^{-1}(y)$ will generate the same data.

In order to avoid the problem with identification, we assume linearity of $h(X) = \beta' X$ and normalize coefficients $\|\beta\| = 1$. Potentially, a weaker assumption could deliver the identification in this model. However, another reason for linearity of $h(X)$ in this paper is data availability. Notice that there are many polling stations in one region. As a result, there are many data points for estimation of joint distribution of Y and Z given X and kernel approach can be used for this. Unfortunately, the number of regions in a country is not big enough for completely non-parametric estimation. For example, in Russia there are 83 federal subjects[4], the United States include 50 voting states, Mexico only consists of 31 states. Thus, it is not possible to non-parametrically estimate function $h(X)$ from insufficient number of observations.

Theorem 3.1. *If Assumptions 3.1 and 3.2 hold,* $h(X) = \beta' X$ *and* $\|\beta\| = 1$*, then* $f_{\delta,c}(\cdot,\cdot)$*,* $g(\cdot)$ *and coefficients* β *are identified.*

Proof. Notice that $g(\cdot)$ is a derivative of $G(\cdot)$, thus, $\frac{\partial G^{-1}(x)}{\partial x} = \frac{1}{g(G^{-1}(x))}$ and by taking partial derivatives of the equation (3.1), we obtain the following:

$$\frac{\partial F_{Y,Z|X}(y,z)}{\partial y} = \frac{F^1_{-\delta-c,-\delta+c}\left(h(X) + G^{-1}(y), h(X) + G^{-1}(z)\right)}{g(G^{-1}(y))}$$

$$\frac{\partial F_{Y,Z|X}(y,z)}{\partial z} = \frac{F^2_{-\delta-c,-\delta+c}\left(h(X) + G^{-1}(y), h(X) + G^{-1}(z)\right)}{g(G^{-1}(z))}$$

$$\begin{aligned}\frac{\partial F_{Y,Z|X}(y,z)}{\partial X^i} &= F^1_{-\delta-c,-\delta+c}\left(h(X) + G^{-1}(y), h(X) + G^{-1}(z)\right) h^i(X) + \\ &+ F^2_{-\delta-c,-\delta+c}\left(h(X) + G^{-1}(y), h(X) + G^{-1}(z)) h^i(X)\right) = \\ &= \left[g(G^{-1}(y))\frac{\partial F_{Y,Z|X}(y,z)}{\partial y} + g(G^{-1}(z))\frac{\partial F_{Y,Z|X}(y,z)}{\partial z}\right] h^i(X)\end{aligned}$$

where $F^i(\cdot,\cdot)$ and $h^i(\cdot)$ denote partial derivatives with respect to i-th element. The last equality is available for different values of y and z, so we integrate all the

[4] Since March 18, 2014, Crimea and Sevastopol became federal subjects of Russia, making the total number to 85 federal subjects, although they are internationally recognized as part of Ukraine.

points over some measure $\mu(y,z)$:

$$\int_0^1\int_0^1 \frac{\partial F_{Y,Z|X}(y,z)}{\partial X^i}\mu(y,z)dydz =$$
$$= h^i(X)\int_0^1\int_0^1 \left[g(G^{-1}(y))\frac{\partial F_{Y,Z|X}(y,z)}{\partial y} + g(G^{-1}(z))\frac{\partial F_{Y,Z|X}(y,z)}{\partial z}\right]\mu(y,z)dydz$$

Substituting $h^i(X) = \beta_i$ for different i in the last equation and considering $\|\beta\| = 1$, provides us with identification of $h(\cdot)$. After that by taking equality at points $y = z$, we obtain that

$$\left.\frac{\partial F_{Y,Z|X}(y,z)}{\partial X^i}\right|_{y=z} = \beta_i g(G^{-1}(z))\left[\left.\frac{\partial F_{Y,Z|X}(y,z)}{\partial y}\right|_{y=z} + \left.\frac{\partial F_{Y,Z|X}(y,z)}{\partial z}\right|_{y=z}\right]. \quad (3.2)$$

However, notice that $Y \leq Z$ implies that

$$\left.\frac{\partial F_{Y,Z|X}(y,z)}{\partial y}\right|_{y=z} = 0, \quad \left.\frac{\partial F_{Y,Z|X}(y,z)}{\partial X^i}\right|_{y=z} = \frac{\partial F_{Z|X}(z)}{\partial X^i}$$
$$\left.\frac{\partial F_{Y,Z|X}(y,z)}{\partial z}\right|_{y=z} = \frac{\partial F_{Z|X}(z)}{\partial z}.$$

Thus, we identify the function $\phi(x) = g(G^{-1}(x))$, and, as a result, $G(\cdot)$: Notice that $g(G^{-1}(x))$ is density at the point where cumulative distribution function $G(\cdot)$ takes value x. It implies that we know the structure of the density but we are missing the axis. However, expectation of personal preferences is assumed to be 0 (Assumption 3.1), which is sufficient for identification of $G(\cdot)$:

$$\int_0^z \frac{1}{\phi(x)}dx = \int_0^z \frac{1}{g(G^{-1}(x))}dx = G^{-1}(z) + const.$$

Using the condition for expectation:

$$\int_0^1 G^{-1}(z)g(G^{-1}(z))dG^{-1}(z) = \int_0^1 G^{-1}(z)dz = 0$$

$$\int_0^1 \left[\int_0^z \frac{1}{\phi(x)} dx - const \right] dz = 0.$$

Thus, unknown constant is

$$const = \int_0^1 \int_0^z \frac{1}{\phi(x)} dx dz = \int_0^1 \int_x^1 \frac{1}{\phi(x)} dz dx = \int_0^1 \frac{1-x}{\phi(x)} dx.$$

$$G^{-1}(z) = \int_0^z \frac{1}{\phi(x)} dx - \int_0^1 \frac{1-x}{\phi(x)} dx.$$

It follows that $G(\cdot)$ is identified from its inverse function. After obtaining functions $G^{-1}(\cdot)$ and $h(\cdot)$, it is possible to identify the joint density of swing voters:

$$f_{-\delta-c,-\delta+c}\left(h(X) + G^{-1}(y), h(X) + G^{-1}(z)\right) = f_{Y,Z|X}(y,z) g(G^{-1}(y)) g(G^{-1}(z)).$$

Finally, by applying the transformation theorem, we derive the joint density of local preferences and costs of voting, δ and c:

$$f_{\delta,c}(x,y) = 2 f_{-\delta-c,-\delta+c}(-x-y, -x+y).$$

□

3.3 Estimation

In this section, first, we review the standard kernel estimators used in the paper. Then, we construct the estimators of partial derivatives of $f_{Y,Z|X}(\cdot,\cdot)$. Finally, we provide the estimators of the model and their asymptotic distributions.

3.3.1 Standard kernel estimators

In order to simplify calculations, suppose that all variables are normalized, and the same bandwidth h is chosen for all of them. Standard kernel estimators of densities and derivatives can be obtained as follows (see, for example, Li and Racine (2007)).

$$\hat{f}(X, y, z) = \frac{1}{Nh^{L+2}} \sum_{i=1}^{N} \tilde{k}\left(\frac{X_i - X}{h}\right) k\left(\frac{Y_i - y}{h}\right) k\left(\frac{Z_i - z}{h}\right)$$

$$\hat{f}(X, z) = \frac{1}{Nh^{L+1}} \sum_{i=1}^{N} \tilde{k}\left(\frac{X_i - X}{h}\right) k\left(\frac{Z_i - z}{h}\right)$$

$$\hat{f}(X) = \frac{1}{Nh^{L}} \sum_{i=1}^{N} \tilde{k}\left(\frac{X_i - X}{h}\right)$$

$$\frac{\widehat{\partial f(X, y, z)}}{\partial X^j} = \frac{-1}{Nh^{L+3}} \sum_{i=1}^{N} \tilde{k_j}\left(\frac{X_i - X}{h}\right) k\left(\frac{Y_i - y}{h}\right) k\left(\frac{Z_i - z}{h}\right)$$

$$\frac{\widehat{\partial f(X, z)}}{\partial X^j} = \frac{-1}{Nh^{L+2}} \sum_{i=1}^{N} \tilde{k_j}\left(\frac{X_i - X}{h}\right) k\left(\frac{Z_i - z}{h}\right)$$

$$\frac{\widehat{\partial f(X)}}{\partial X^j} = \frac{-1}{Nh^{L+1}} \sum_{i=1}^{N} \tilde{k_j}\left(\frac{X_i - X}{h}\right)$$

where $\tilde{k}(\frac{X_i - X}{h}) = \prod_{j=1}^{L} k(\frac{X_i^j - X^j}{h})$ is a multiplicative kernel of L regional characteristics and $\tilde{k_j}(\frac{X_i - X}{h}) = k'(\frac{X_i^j - X^j}{h}) \prod_{t=1, t\neq j}^{L} k(\frac{X_i^t - X^t}{h})$.

3.3.2 Estimation of partial derivatives of $F_{Y,Z|X}(y, z)$

The identification strategy suggests using partial derivatives of $F_{Y,Z|X}(y, z)$ for model estimation. Denote joint density of X, Y and Z as $f(X, y, z)$, and density of regional characteristics X as $f(X)$. By definition of the conditional cumulative distribution function:

$$F_{Y,Z|X}(y, z) = \int_{-\infty}^{y} \int_{-\infty}^{z} \frac{f(X, w, t)}{f(X)} dt dw.$$

Then, the partial derivatives with respect to y and z are

$$f_{Z|X}(z) = \frac{\partial F_{Z|X}(z)}{\partial z} = \frac{f(X,z)}{f(X)}$$

$$f_{Y,Z|X}(y,z) = \frac{\partial^2 F_{Y,Z|X}(y,z)}{\partial y \partial z} = \frac{f(X,y,z)}{f(X)}.$$

The analog non-parametric estimators of $f_{Y,Z|X}(y,z)$ and $f_{Z|X}(z)$ can be obtained by substituting the standard kernel estimators in the above equations for derivatives.

In order to obtain estimators of partial derivatives of $F_{Y,Z|X}(y,z)$ and $F_{Z|X}(z)$ with respect to j-th regional characteristic, notice that:

$$\begin{aligned}\frac{\partial F_{Y,Z|X}(y,z)}{\partial X^j} = \frac{1}{f(X)} \int_{-\infty}^{y} \int_{-\infty}^{z} \frac{\partial f(X,w,t)}{\partial X^j} dtdw - \\ - \frac{\frac{\partial f(X)}{\partial X^j}}{(f(X))^2} \int_{-\infty}^{y} \int_{-\infty}^{z} f(X,w,t)dtdw\end{aligned} \tag{3.3}$$

$$\frac{\partial F_{Z|X}(z)}{\partial X^j} = \frac{1}{f(X)} \int_{-\infty}^{z} \frac{\partial f(X,t)}{\partial X^j} dt - \frac{\frac{\partial f(X)}{\partial X^j}}{(f(X))^2} \int_{-\infty}^{z} f(X,t)dt. \tag{3.4}$$

The estimators of the integral of the derivatives can be obtained as follows:

$$\begin{aligned}\int_{-\infty}^{y} \int_{-\infty}^{z} \frac{\partial \widehat{f(X,w,t)}}{\partial X^j} dtdw &= \frac{-1}{Nh^{L+3}} \sum_{i=1}^{N} \tilde{k}_j \left(\frac{X_i - X}{h} \right) \int_{-\infty}^{y} k \left(\frac{Y_i - w}{h} \right) dw \int_{-\infty}^{z} k \left(\frac{Z_i - t}{h} \right) dt \\ &= \frac{-1}{Nh^{L+1}} \sum_{i=1}^{N} \tilde{k}_j \left(\frac{X_i - X}{h} \right) K \left(\frac{y - Y_i}{h} \right) K \left(\frac{z - Z_i}{h} \right)\end{aligned}$$

and

$$\int_{-\infty}^{z} \frac{\partial \widehat{f(X,t)}}{\partial X^j} dt = \frac{-1}{Nh^{L+1}} \sum_{i=1}^{N} \tilde{k}_j \left(\frac{X_i - X}{h} \right) K \left(\frac{z - Z_i}{h} \right),$$

where $K(\cdot)$ is a kernel cdf. In addition, the estimators for the integral of densities are

$$\int_{-\infty}^{y}\int_{-\infty}^{z}\widehat{f(X,w,t)}dtdw = \frac{1}{Nh^{L+2}}\sum_{i=1}^{N}\tilde{k}_j\left(\frac{X_i - X}{h}\right)\int_{-\infty}^{y}k\left(\frac{Y_i - w}{h}\right)dw\int_{-\infty}^{z}k\left(\frac{Z_i - t}{h}\right)dt =$$

$$= \frac{1}{Nh^{L}}\sum_{i=1}^{N}\tilde{k}_j\left(\frac{X_i - X}{h}\right)K\left(\frac{y - Y_i}{h}\right)K\left(\frac{z - Z_i}{h}\right)$$

$$\int_{-\infty}^{z}\widehat{f(X,t)}dt = \frac{1}{Nh^{L}}\sum_{i=1}^{N}\tilde{k}_j\left(\frac{X_i - X}{h}\right)K\left(\frac{z - Z_i}{h}\right).$$

Finally, the analog estimators for $\frac{\partial \widehat{F_{Y,Z|X}(y,z)}}{\partial X^j}$ and $\frac{\partial \widehat{F_{Z|X}(z)}}{\partial X^j}$ can be obtained by plugging estimators of the parts into the formulas (3.3) and (3.4).

3.3.3 Estimation of the model

The following assumptions guarantee asymptotics and the uniform convergence results in the paper.

Assumption 3.3. *The sequence* $\{X_i, Y_i, Z_i\}$ *is i.i.d.*

Assumption 3.4. $f(X,y,z)$ *has compact support in* $\mathbb{R}^{L+2}$ *and is continuously differentiable up to the order* s'*, for some* $s' > 2$.

Assumption 3.5. *The kernel function* $k(\cdot)$ *is differentiable of order* $\tilde{s}$*, the derivatives of* k *of order* $\tilde{s}$ *are Lipscitz,* $k(\cdot)$ *vanishes outside a compact set, integrates to 1, and is of order* s''*, where* $\tilde{s} + s'' \leq s'$.

Assumption 3.6. *As* $N \to \infty$, $h \to 0$, $\frac{\ln N}{Nh^{L+4}} \to 0$, $Nh^{L+2} \to \infty$, $Nh^{L+2+2s''} \to 0$, *and* $\sqrt{N}\left(\sqrt{\frac{\ln N}{Nh^{L+4}}} + h^{s''}\right)^2 \to 0$.

Assumption 3.7. $0 < f(X,y,z) < \infty$.

Define V_{ayz} by

$$V_{ayz} = \frac{R^{L+2}(k) f(X,y,z)}{f^2(X)}.$$

Theorem 3.2. *Let the estimator of $f_{Y,Z|X}(y,z)$ be $\hat{f}_{Y,Z|X}(y,z) = \frac{\hat{f}(X,y,z)}{\hat{f}(X)}$. If Assumptions 3.3–3.7 are satisfied, then,*

$$\sup_{(y,z)\in\mathbb{R}^2} \left| \frac{\hat{f}(X,y,z)}{\hat{f}(X)} - \frac{f(X,y,z)}{f(X)} \right| \xrightarrow{p} 0$$
$$\sqrt{Nh^{L+2}} \left(\hat{f}_{Y,Z|X}(y,z) - f_{Y,Z|X}(y,z) \right) \xrightarrow{d} N\left(0, V_{ayz}\right).$$

Let $\mu(X,y,z)$ be a weighting function such that $\iiint \mu(X,y,z)dydzdX = 1$. Then the estimator of β_i can be defined by

$$\hat{\beta}_i = \frac{\iiint \widehat{\frac{\partial F_{Y,Z|X}(y,z)}{\partial X^i}} \mu(X,y,z)dydzdX}{\sqrt{\sum_{j=1}^{L} \left(\iiint \widehat{\frac{\partial F_{Y,Z|X}(y,z)}{\partial X^j}} \mu(X,y,z)dydzdX \right)^2}}.$$

In order to simplify the notation, denote

$$\bar{F}_{Y,Z|X}(X) = \iint F_{Y,Z|X}(y,z)\mu(X,y,z)dydz$$
$$\bar{F}_{Y,Z|XX}(X) = \frac{\bar{F}_{Y,Z|X}(X)}{f(X)}$$
$$F_i = \iiint \frac{\partial F_{Y,Z|X}(y,z)}{\partial X^i} \mu(X,y,z)dydzdX$$
$$\ddot{F}_{Z|X}(z) = \int \frac{\partial F_{Z|X}(z)}{\partial X^i} \mu(X)dX$$
$$\tilde{F} = \sum_{j=1}^{L} \left(\iiint \frac{\partial F_{Y,Z|X}(y,z)}{\partial X^j} \mu(X,y,z)dydzdX \right)^2 = \sum_{j=1}^{L} F_j^2$$

$$M_j(X,y,z) = \frac{1}{f(X)}\left(\int_y^{+\infty}\int_z^{+\infty}\frac{\partial\mu(X,w,t)}{\partial X^j}dtdw - \right.$$
$$\left. -\frac{\partial f(X)}{\partial X^j}\frac{1}{f(X)}\int_y^{+\infty}\int_z^{+\infty}\mu(X,w,t)dtdw\right)$$

and $F = (F_1, ..., F_L)'$.

In addition, define V_{β_i} by

$$V_{\beta_i} = \tilde{F}^{-3}\left(\left(\sum_{j\neq i}F_j^2\right)^2 V_{F_i} + F_i^2\sum_{j\neq i}F_j^2 V_{F_j} - 2(\tilde{F}-F_i^2)F_i\sum_{j\neq i}F_jC_{ij} + F_i^2\sum_{j\neq i,k\neq j,i}F_jF_kC_{jk}\right),$$

where

$$C_{ij} = \mathrm{cov}(M_i(X,y,z), M_j(X,y,z)) - \mathrm{cov}\left(M_i(X,y,z), \frac{\partial\bar{F}_{Y,Z|XX}(X)}{\partial X^j}\right)$$
$$-\mathrm{cov}\left(M_j(X,y,z), \frac{\partial\bar{F}_{Y,Z|XX}(X)}{\partial X^i}\right) + \mathrm{cov}\left(\frac{\bar{F}_{Y,Z|XX}(X)}{\partial X^i}, \frac{\bar{F}_{Y,Z|XX}(X)}{\partial X^j}\right)$$

and

$$V_{F_j} = V\left(M_j(X,y,z)\right) + V\left(\frac{\partial\bar{F}_{Y,Z|XX}(X)}{\partial X^j}\right) - 2\mathrm{cov}\left(M_j(X,y,z), \frac{\partial\bar{F}_{Y,Z|XX}(X)}{\partial X^j}\right).$$

Theorem 3.3. *Let the estimator of β_i be as defined above. If Assumptions 3.3–3.7 are satisfied, then,* $\sqrt{N}\left(\hat{\beta}_i - \beta_i\right) \xrightarrow{d} N(0, V_{\beta_i})$.

Now using equation (3.2), we derive estimators for $\phi(z) = g(G^{-1}(z))$ and $G^{-1}(z)$:

$$\hat{\phi}(z) = \hat{g}(G^{-1}(z)) = \frac{\int \widehat{\frac{\partial F_{Z|X}(z)}{\partial X^i}}\mu(X)dX}{\hat{\beta}_i\int\widehat{\frac{\partial F_{Z|X}(z)}{\partial z}}\mu(X)dX}$$
$$\widehat{G^{-1}(u)} = \int_0^u \frac{1}{\hat{\phi}(x)}dx - \int_0^1\frac{1-x}{\hat{\phi}(x)}dx.$$

In order to simplify the expression for the variance of $\widehat{G^{-1}(\cdot)}$, we introduce the following notation:

$$V_u = V\left(f_{Z|X}(z)\frac{\ddot{F}_{Z|X}(z)\mu(X)}{\beta_i\phi^2(z)}\mathbb{1}(z<u)\right)$$
$$V_c = V\left(f_{Z|X}(z)\frac{(1-z)\ddot{F}_{Z|X}(z)\mu(X)}{\beta_i\phi^2(z)}\right)$$
$$\Sigma_{uc} = \mathrm{cov}\left(f_{Z|X}(z)\frac{\ddot{F}_{Z|X}(z)\mu(X)}{\beta_i\phi^2(z)}\mathbb{1}(z<u), f_{Z|X}(z)\frac{(1-z)\ddot{F}_{Z|X}(z)\mu(X)}{\beta_i\phi^2(z)}\right).$$

Finally, define $V_{G^{-1}(u)}$ by

$$V_{G^{-1}(u)} = V_u + V_c - 2\Sigma_{uc}.$$

Theorem 3.4. *Let the estimator of $G^{-1}(z)$ be as defined above. If Assumptions 3.3–3.7 are satisfied, then,*

$$\sup_{z\in\mathbb{R}}\left|\widehat{G^{-1}(u)} - G^{-1}(u)\right| \xrightarrow{p} 0$$
$$\sqrt{N}\left(\widehat{G^{-1}(u)} - G^{-1}(u)\right) \xrightarrow{d} N(0, V_{G^{-1}(u))}).$$

Additionally, the asymptotic distributions of $\hat{\phi}(z)$ and $\hat{g}(\widehat{G^{-1}(z)})$ are provided in Lemma 3.8 and 3.12 in Appendix.

Joint density of local preference and costs of voting conditional on the regional characteristics at the point

$$\theta = \left(-\frac{1}{2}(2\sum_{i=1}^{L}\beta_i X^i + G^{-1}(y) + G^{-1}(z)), \frac{1}{2}(G^{-1}(z) - G^{-1}(y))\right)$$

can be estimated as follows:

$$\hat{f}_{\delta,c}(\theta) = 2\hat{f}_{Y,Z|X}(y,z)\hat{\phi}(y)\hat{\phi}(z).$$

The estimators for the components of the vector θ are defined by

$$\hat{\theta}_1 = -0.5\left(2\sum_{i=1}^{L}\hat{\beta}_i X^i + \widehat{G^{-1}(y)} + \widehat{G^{-1}(z)}\right)$$
$$\hat{\theta}_2 = 0.5\left(\widehat{G^{-1}(z)} - \widehat{G^{-1}(y)}\right).$$

Theorem 3.5. *Let the estimators of θ and $f_{\delta,c}(\theta)$ be as defined above. If Assumptions 3.3–3.7 are satisfied, then,*

$$\sup_{\theta\in\mathbb{R}^2}\left|\hat{f}_{\delta,c}(\hat{\theta}) - f_{\delta,c}(\theta)\right| \xrightarrow{p} 0$$
$$\sqrt{Nh^{L+2}}\left(\hat{f}_{\delta,c}(\hat{\theta}) - f_{\delta,c}(\theta)\right) \xrightarrow{d} N\left(0, 4\phi^2(y)\phi^2(z)V_{ayz}\right).$$

3.4 Conclusions

This Chapter has introduced a new electoral model that allows derivation of the joint distribution of turnout and voter share from preferences and costs of voting. Because joint distribution of turnout and voter share can be easily obtained from the data, the unobserved distributions of preferences and costs of voting can be identified and estimated from it. The estimators are naturally derived from the identification strategy that is introduced in the paper. All estimators are consistent and asymptotically normal.

The results of this paper can be useful in research on development of electoral preferences over time in democratic fraud-free countries or in predictions of the effect of the electoral policy changes. Such investigations could help provide implications about electoral patterns in developing democracies. Another potential application of this paper is non-parametric estimation of electoral fraud.

3.5 Appendix to Chapter 3

Let $||g||$ denote the maximum of the supremum of the values and derivatives up to the second order of g and $g(X) = \int\int g(X, y, z)dydz$.

3.5.1 Standard Asymptotics

Asymptotic distributions of kernel estimators of density and derivatives are standard:

$$\sqrt{Nh^{L+2}}\left(\hat{f}(X, y, z) - f(X, y, z)\right) \xrightarrow{d} N\left(0, f(X, y, z)R^{L+2}(k)\right)$$

$$\sqrt{Nh^{L}}\left(\hat{f}(X) - f(X)\right) \xrightarrow{d} N\left(0, f(X)R^{L}(k)\right)$$

$$\sqrt{Nh^{L+4}}\left(\frac{\partial \widehat{f(X, y, z)}}{\partial X^j} - \frac{\partial f(X, y, z)}{\partial X^j}\right) \xrightarrow{d} N\left(0, f(X, y, z)R^{L+1}(k)R(k')\right)$$

$$\sqrt{Nh^{L+2}}\left(\widehat{\frac{\partial f(X)}{\partial X^j}} - \frac{\partial f(X)}{\partial X^j}\right) \xrightarrow{d} N\left(0, f(X)R^{L-1}(k)R(k')\right),$$

where $R(f) = \int [f(u)]^2 du$, $k(\cdot)$ is kernel density and $K(\cdot)$ is kernel cdf.

3.5.2 Proof of Theorem 3.2

Denote

$$V_{az} = R(k)\int \frac{f(X, z)\mu^2(X)}{f^2(X)}dX.$$

Lemma 3.1. *Let* $\int \hat{f}_{Z|X}(z)\mu(X)dX$ *be the estimator of* $\int f_{Z|X}(z)\mu(X)dX$*. Sup-*

pose that Assumptions 3.3–3.7 hold, then,

$$\sup_{z\in\mathbb{R}}\left|\int\left(\hat{f}_{Z|X}(z)-f_{Z|X}(z)\right)\mu(X)dX\right|\xrightarrow{p}0$$
$$\sqrt{Nh}\left(\int\left(\hat{f}_{Z|X}(z)-f_{Z|X}(z)\right)\mu(X)dX\right)\xrightarrow{d}N(0,V_{az}).$$

Proof. Define the functional $\Phi(g)=\int g_{Z|X}(z)\mu(X)dX$. Then,

$$\Phi(\hat{f})=\int\hat{f}_{Z|X}(z)\mu(X)dX \text{ and } \Phi(f)=\int f_{Z|X}(z)\mu(X)dX.$$

For any h such that $||h||$ is sufficiently small: $|h(X)|\leq a||h||$, $\int\frac{h(X,z)\mu(X)}{f(X)}dX\leq a||h||$, and $|f(X)+h(X)|\geq b|f(X)|$ for some $0<a,b<\infty$. Also,

$$\Phi(f+h)-\Phi(f)=D\Phi(f,h)+R\Phi(f,h),$$

where

$$D\Phi(f,h)=\int\frac{h(X,z)\mu(X)}{f(X)}dX-\int\frac{h(X)f_{Z|X}(z)\mu(X)}{f(X)}dX$$
$$R\Phi(f,h)=\left[\int\frac{h(X,t)\mu(X)}{f(X)}\frac{-h(X)}{f(X)+h(X)}dX-\int\frac{h(X)f_{Z|X}(z)\mu(X)}{f(X)}\frac{-h(X)}{f(X)+h(X)}dX\right].$$

Thus, for some $c<\infty$:

$$|D\Phi(f,h)|\leq c||h|| \text{ and } |R\Phi(f,h)|\leq c||h||^2.$$

Lemma B3 in Newey (1994) guarantees that $||h||\xrightarrow{p}0$. As a result, $\sup|\Phi(\hat{f})-\Phi(f)|\leq c||h||+c||h||^2\xrightarrow{p}0$.

Now for $h=\hat{f}-f$:

$$D\Phi(f,\hat{f}-f)=\int\frac{\left(\hat{f}(X,z)-f(X,z)\right)\mu(X)}{f(X)}dX-\int\frac{\left(\hat{f}(X)-f(X)\right)f_{Z|X}(z)\mu(X)}{f(X)}dX.$$

Notice that the second term converges faster than the first term. Also,

$$
\begin{aligned}
&E\left(\int \tilde{k}\left(\frac{X_i - X}{h}\right) k\left(\frac{Z_i - z}{h}\right) \frac{\mu(X)}{f(X)} dX\right)^2 \\
&= \iint \left(\int \tilde{k}\left(\frac{X_i - X}{h}\right) \frac{\mu(X)}{f(X)} dX\right)^2 k^2\left(\frac{Z_i - z}{h}\right) f(X_i, Z_i) dX_i dZ_i \\
&= h^{2L+1} R(k) \int \frac{\mu^2(X) f(X, z)}{f^2(X)} dX
\end{aligned}
$$

and

$$
E\left(\int \tilde{k}\left(\frac{X_i - X}{h}\right) k\left(\frac{Z_i - z}{h}\right) \frac{\mu(X)}{f(X)} dX\right) = h^{L+1} \int \frac{\mu(X) f(X, z)}{f(X)} dX.
$$

Thus,

$$
\sqrt{Nh} \int \frac{h(X, z)\mu(X)}{f(X)} dX \xrightarrow{d} N\left(0, V_{az}\right).
$$

Hence, $\sqrt{Nh} D\Phi(f, \hat{f} - f) \xrightarrow{d} N(0, V_{az})$. Now by Lemma B.3 in Newey (1994) and Assumption 3.6, $\sqrt{Nh} R\Phi(f, h) \xrightarrow{p} 0$ and it follows that

$$
\sqrt{Nh} \int h_{Z|X}(z) \mu(X) dX = \sqrt{Nh} D\Phi(f, \hat{f} - f) \xrightarrow{d} N(0, V_{az}).
$$

□

Recall that

$$
V_{ayz} = \frac{R^{L+2}(k) f(X, y, z)}{f^2(X)}.
$$

Lemma 3.2. *Let $\hat{f}_{Y,Z|X}(y, z)$ be the estimator of $f_{Y,Z|X}(y, z)$. Suppose that Assumptions 3.3–3.7 hold, then,*

$$
\begin{gathered}
\sup_{(y,z)\in\mathbb{R}^2} \left|\hat{f}_{Y,Z|X}(y, z) - f_{Y,Z|X}(y, z)\right| \xrightarrow{p} 0 \\
\sqrt{Nh^{L+2}} \left(\hat{f}_{Y,Z|X}(y, z) - f_{Y,Z|X}(y, z)\right) \xrightarrow{d} N\left(0, V_{ayz}\right).
\end{gathered}
$$

Proof. By definition $\hat{f}_{Y,Z|X}(y,z) = \frac{\hat{f}(X,y,z)}{\hat{f}(X)}$, thus, define the functional $\Phi(g) = g_{Y,Z|X}(y,z)$. Then, $\Phi(\hat{f}) = \hat{f}_{Y,Z|X}(y,z)$ and $\Phi(f) = f_{Y,Z|X}(y,z)$. For any H such that $||H||$ is sufficiently small: $|h(X)| \leq a||H||$and $|f(X)+h(X)| \geq b|f(X)|$ for some $0 < a, b < \infty$. Also,

$$\Phi(f+h) - \Phi(f) = D\Phi(f,h) + R\Phi(f,h),$$

where

$$D\Phi(f,h) = \frac{h(X,y,z) - h(X)f_{Y,Z|X}(y,z)}{f(X)}$$
$$R\Phi(f,h) = \left[\frac{h(X,y,z) - h(X)f_{Y,Z|X}(y,z)}{f(X)}\right]\left[\frac{-h(X)}{f(X)+h(X)}\right].$$

Thus, for some $c < \infty$:

$$|D\Phi(f,h)| \leq \frac{c||H||}{f(X)} \text{ and } |R\Phi(f,h)| \leq \frac{c||H||^2}{f^2(X)}.$$

The last implies uniform convergence as before. Now for $h = \hat{f} - f$:

$$D\Phi(f, \hat{f} - f) = \frac{1}{f(X)}\left[\hat{f}(X,y,z) - f(X,y,z) - f_{Y,Z|X}(y,z)(\hat{f}(X) - f(X))\right].$$

However, notice that $\hat{f}(X,y,z)$ has $\sqrt{Nh^{L+2}}$ rate of convergence, while $\hat{f}(X)$ converges faster, as $\sqrt{Nh^L}$. Thus,

$$\sqrt{Nh^{L+2}} D\Phi(f, \hat{f} - f) =$$
$$= \sqrt{Nh^{L+2}}\left[\frac{1}{f(X)}\left(\hat{f}(X,y,z) - f(X,y,z)\right)\right] + o_p(1) \xrightarrow{d} N\left(0, V_{ayz}\right).$$

By Lemma B.3 in Newey (1994) and under Assumption 3.6, $\sqrt{Nh^{L+2}} R\Phi(f,h) \xrightarrow{p} 0$

and it follows that

$$\sqrt{Nh^{L+2}}\left(\hat{f}_{Y,Z|X}(y,z)-f_{Y,Z|X}(y,z)\right)=\sqrt{Nh^{L+2}}D\Phi(f,\hat{f}-f)\xrightarrow{d}N(0,V_{ayz}).$$

$\square$

Denote V_{yzxj} by

$$V_{yzxj}=\frac{R^{L-1}(k)R(k')}{f(X)}F_{Y,Z|X}(y,z)\left[1-F_{Y,Z|X}(y,z)\right].$$

Lemma 3.3. *Let* $\widehat{\frac{\partial F_{Y,Z|X}(y,z)}{\partial X^j}}$ *be the estimator of* $\frac{\partial F_{Y,Z|X}(y,z)}{\partial X^j}$*. Suppose that Assumptions 3.3–3.7 hold. Then,*

$$\sup_{(y,z)\in\mathbb{R}^2}\left|\widehat{\frac{\partial F_{Y,Z|X}(y,z)}{\partial X^j}}-\frac{\partial F_{Y,Z|X}(y,z)}{\partial X^j}\right|\xrightarrow{p}0$$
$$\sqrt{Nh^{L+2}}\left(\widehat{\frac{\partial F_{Y,Z|X}(y,z)}{\partial X^j}}-\frac{\partial F_{Y,Z|X}(y,z)}{\partial X^j}\right)\xrightarrow{d}N(0,V_{yzxj}).$$

Proof. Define the functional $\Phi(g)=\frac{\partial G_{Y,Z|X}(y,z)}{\partial X^j}$. Then, $\Phi(\hat{f})=\widehat{\frac{\partial F_{Y,Z|X}(y,z)}{\partial X^j}}$ and $\Phi(f)=\frac{\partial F_{Y,Z|X}(y,z)}{\partial X^j}$. For any h such that $||h||$ is sufficiently small: $\frac{\partial h(X)}{\partial X^j}\le a||h||$, $|\int_{-\infty}^{z}\int_{-\infty}^{y}\frac{\partial h(X,w,t)}{\partial X^j}dwdt|\le a||h||$, $|\int_{-\infty}^{z}\int_{-\infty}^{y}h(X,w,t)dwdt|\le a||h||$, $h(X)\le a||h||$ and $|f(X)+h(X)|\ge b|f(X)|$ for some $0<a,b<\infty$. Also,

$$\Phi(f+h)-\Phi(f)=D\Phi(f,h)+R\Phi(f,h),$$

where

$$D\Phi(f,h)=\frac{1}{f^2(X)}\left[f(X)\int_{-\infty}^{z}\int_{-\infty}^{y}\frac{\partial h(X,w,t)}{\partial X^j}dwdt-\frac{\partial f(X)}{\partial X^j}\int_{-\infty}^{z}\int_{-\infty}^{y}h(X,w,t)dwdt-\right.$$
$$\left.-\frac{\partial h(X)}{\partial X^j}\int_{-\infty}^{z}\int_{-\infty}^{y}f(X,w,t)dwdt-h(X)\int_{-\infty}^{z}\int_{-\infty}^{y}\left(\frac{\partial f(X,w,t)}{\partial X^j}-2\frac{\partial f(X)}{\partial X^j}f_{Y,Z|X}(w,t)dwdt\right)\right].$$

Thus, for some $c < \infty$:

$$|D\Phi(f,h)| \leq \frac{c||h||}{f^2(X)} \text{ and } |R\Phi(f,h)| \leq \frac{c||h||^2}{f^3(X)},$$

and the uniform convergence result follows. Now for $h = \hat{f} - f$:

$$\begin{aligned} D\Phi(f,\hat{f}-f) = \frac{1}{f(X)} & \left[\int_{-\infty}^{z}\int_{-\infty}^{y} \frac{\partial \hat{f}(X,w,t)}{\partial X^j} dwdt - \int_{-\infty}^{z}\int_{-\infty}^{y} \frac{\partial f(X,w,t)}{\partial X^j} dwdt\right] \\ & + \frac{\frac{\partial f(X)}{\partial X^j}}{f^2(X)} \left[\int_{-\infty}^{z}\int_{-\infty}^{y} \hat{f}(X,w,t)dwdt - \int_{-\infty}^{z}\int_{-\infty}^{y} f(X,w,t)dwdt\right] \\ & - \frac{1}{f^2(X)} \int_{-\infty}^{z}\int_{-\infty}^{y} f(X,w,t)dwdt \left[\frac{\partial \hat{f}(X)}{\partial X^j} - \frac{\partial f(X)}{\partial X^j}\right] \\ & - \int_{-\infty}^{z}\int_{-\infty}^{y} \left(\frac{\partial f(X,w,t)}{\partial X^j} - 2\frac{\partial f(X)}{\partial X^j} f_{Y,Z|X}(w,t)dwdt\right) \left[\hat{f}(X) - f(X)\right]. \end{aligned}$$

However, notice that $\hat{f}(X)$ and $\int_{-\infty}^{y}\int_{-\infty}^{z} \hat{f}(X,w,t)dtdw$ converge with rate $\sqrt{Nh^L}$ and $\frac{\partial \hat{f}(X)}{\partial X^j}$ with $\sqrt{Nh^{L+2}}$. Moreover,

$$\begin{aligned} & \sqrt{Nh^{L+2}} \left(\int_{-\infty}^{y}\int_{-\infty}^{z} \frac{\partial \widehat{f(X,w},t)}{\partial X^j} dtdw - \int_{-\infty}^{y}\int_{-\infty}^{z} \frac{\partial f(X,w,t)}{\partial X^j} dtdw\right) \xrightarrow{d} \\ & \xrightarrow{d} N\left(0, R^{L-1}(k)R(k') \int_{-\infty}^{z}\int_{-\infty}^{y} f(X,w,t)dwdt\right). \end{aligned}$$

Additionally,

$$\begin{aligned} & \operatorname{cov}\left(\int_{-\infty}^{y}\int_{-\infty}^{z} \frac{\partial \widehat{f(X,w},t)}{\partial X^j} dtdw, \frac{\partial \hat{f}(X)}{\partial X^j}\right) = \\ & = \frac{1}{Nh^{2L+2}} \operatorname{cov}\left(\tilde{k}_j\left(\frac{X_i - X}{h}\right) K\left(\frac{y - Y_i}{h}\right) K\left(\frac{z - Z_i}{h}\right), \tilde{k}_j\left(\frac{X_i - X}{h}\right)\right) \\ & \approx \frac{1}{Nh^{L+2}} R^{L-1}(k)R(k') \int_{-\infty}^{z}\int_{-\infty}^{y} f(X,w,t)dwdt. \end{aligned}$$

Thus, $D\Phi(f,h)$ will converge as follows:

$$\sqrt{Nh^{L+2}}D\Phi(f,\hat{f}-f)=\sqrt{Nh^{L+2}}\frac{1}{f(X)}\int_{-\infty}^{z}\int_{-\infty}^{y}f_{Y,Z|X}(w,t)dwdt\left[\frac{\partial\hat{f}(X)}{\partial X^j}-\frac{\partial f(X)}{\partial X^j}\right]$$
$$+\sqrt{Nh^{L+2}}\frac{1}{f(X)}\left[\int_{-\infty}^{z}\int_{-\infty}^{y}\frac{\partial\hat{f}(X,w,t)}{\partial X^j}dwdt-\int_{-\infty}^{z}\int_{-\infty}^{y}\frac{\partial f(X,w,t)}{\partial X^j}dwdt\right]+o_p(1)\xrightarrow{d}$$
$$\xrightarrow{d}N\left(0,V_{yzxj}\right).$$

Now by Lemma B.3 in Newey (1994), $\sqrt{Nh^{L+2}}R\Phi(f,h)\xrightarrow{p}0$ and it follows that

$$\sqrt{Nh^{L+2}}\left(\frac{\partial\widehat{F_{Y,Z|X}(y,z)}}{\partial X^j}-\frac{\partial F_{Y,Z|X}(y,z)}{\partial X^j}\right)=\sqrt{Nh^{L+2}}D\Phi(f,\hat{f}-f)\xrightarrow{d}N(0,V_{yzxj}).$$

$\square$

Denote
$$V_{zxj}=\frac{R^{L-1}(k)R(k')}{f(X)}F_{Z|X}(z)\left[1-F_{Z|X}(z)\right].$$

Corollary 3.6. *Let* $\frac{\partial\widehat{F_{Z|X}(z)}}{\partial X^j}$ *be the estimator of* $\frac{\partial F_{Z|X}(z)}{\partial X^j}$*. Suppose that Assumptions 3.3–3.7 hold. Then,*

$$\sup_{z\in\mathbb{R}}\left|\frac{\partial\widehat{F_{Z|X}(z)}}{\partial X^j}-\frac{\partial F_{Z|X}(z)}{\partial X^j}\right|\xrightarrow{p}0$$
$$\sqrt{Nh^{L+2}}\left(\frac{\partial\widehat{F_{Z|X}(z)}}{\partial X^j}-\frac{\partial F_{Z|X}(z)}{\partial X^j}\right)\xrightarrow{d}N(0,V_{zxj}).$$

3.5.3 Proof of Theorem 3.3

Lemma 3.4. *If Assumptions 3.3–3.7 hold, then,*

$$N\,cov\left(\int r(X)\frac{\partial\hat{f}(X)}{\partial X_i}dX,\int\tilde{r}(X)\frac{\partial\hat{f}(X)}{\partial X_j}dX\right)\to cov\left(\frac{\partial r(X)}{\partial X_i},\frac{\partial\tilde{r}(X)}{\partial X_j}\right)$$

and

$$\sqrt{N}\left(\int r(X)\frac{\partial h(X)}{\partial X_i}dX\right)\xrightarrow{d} N\left(0, V\left(\frac{\partial r(X)}{\partial X_i}\right)\right).$$

Proof. Note that

$$\begin{aligned}
\operatorname{cov}\left(\int r(X)\frac{\partial \hat f(X)}{\partial X_i}dX, \int \tilde r(X)\frac{\partial \hat f(X)}{\partial X_j}dX\right) = \\
= \frac{1}{Nh^{2L+2}}\operatorname{cov}\left(\int r(X)\tilde k_i\left(\frac{X_i - X}{h}\right)dX, \int \tilde r(X)\tilde k_j\left(\frac{X_i - X}{h}\right)dX\right)\\
= \frac{1}{N}\operatorname{cov}\left(\int \frac{\partial r(X_i - h\sigma_x)}{\partial(X_i - h\sigma_x)}\tilde k(\sigma_x)d\sigma_x, \int \frac{\partial \tilde r(X_i - h\sigma_x)}{\partial(X_i - h\sigma_x)}\tilde k(\sigma_x)d\sigma_x\right)
\end{aligned}$$

and

$$\operatorname{cov}\left(\int \frac{\partial r(X_i - h\sigma_x)}{\partial(X_i - h\sigma_x)}\tilde k(\sigma_x)d\sigma_x, \int \frac{\partial \tilde r(X_i - h\sigma_x)}{\partial(X_i - h\sigma_x)}\tilde k(\sigma_x)d\sigma_x\right) \to \operatorname{cov}\left(\frac{\partial r(X)}{\partial X_i}, \frac{\partial \tilde r(X)}{\partial X_j}\right).$$

Thus, $NV\left(\int r(X)\frac{\partial \hat f(X)}{\partial X_i}dX\right) \to V\left(\frac{\partial r(X)}{\partial X_i}\right)$. □

Recall that

$$M_j(X,y,z) = \frac{1}{f(X)}\left(\int_y^{+\infty}\int_z^{+\infty}\frac{\partial \mu(X,w,t)}{\partial X^j}dtdw - \frac{\partial f(X)}{\partial X^j}\frac{1}{f(X)}\int_y^{+\infty}\int_z^{+\infty}\mu(X,w,t)dtdw\right).$$

Lemma 3.5. *If Assumptions 3.3–3.7 hold, then,*

$$\sqrt{N}\iiint \frac{\mu(X,y,z)}{f(X)}\int_{-\infty}^{y}\int_{-\infty}^{z}\frac{\partial h(X,w,t)}{\partial X^j}dtdwdXdydz \xrightarrow{d} N\left(0, V\left(M_j(X,y,z)\right)\right).$$

In addition,

$$\begin{aligned}
N\,cov\left(\int r(X)\frac{\partial h(X)}{\partial X^i}dX, \iiint \frac{\mu(X,y,z)}{f(X)}\int_{-\infty}^{y}\int_{-\infty}^{z}\frac{\partial h(X,w,t)}{\partial X^j}dtdwdXdydz\right) \to \\
\to cov\left(\frac{\partial r(X)}{\partial X^i}, M_j(X,y,z)\right)
\end{aligned}$$

and

$$N\,cov\left(\iiint \frac{\mu(X,y,z)}{f(X)}\int_{-\infty}^{y}\int_{-\infty}^{z}\frac{\partial h(X,w,t)}{\partial X^i}dtdwdXdydz,\right.$$
$$\left.\iiint \frac{\mu(X,y,z)}{f(X)}\int_{-\infty}^{y}\int_{-\infty}^{z}\frac{\partial h(X,w,t)}{\partial X^j}dtdwdXdydz\right) \to cov\left(M_i(X,y,z), M_j(X,y,z)\right).$$

Proof. First, note that

$$\int \mu(X,y,z)K\left(\frac{y-Y_i}{h}\right)dy = \mu(X,z) - \int\int_{-\infty}^{y}\mu(X,w,z)dwdK\left(\frac{y-Y_i}{h}\right) =$$
$$= \mu(X,z) - \int_{-\infty}^{Y_i}\mu(X,w,z)dw.$$

Thus,

$$\int\int \mu(X,y,z)K\left(\frac{y-Y_i}{h}\right)K\left(\frac{z-Z_i}{h}\right)dydz = \mu(X) - \int_{-\infty}^{Z_i}\int_{-\infty}^{Y_i}\mu(X,w,t)dwdt.$$

Then,

$$V\left(\iiint \frac{\mu(X,y,z)}{f(X)}\int_{-\infty}^{y}\int_{-\infty}^{z}\frac{\partial h(X,w,t)}{\partial X^j}dtdwdXdydz\right) =$$
$$= \frac{1}{Nh^{2L+2}}V\left(\iiint \frac{\mu(X,y,z)}{f(X)}\tilde{k}_j\left(\frac{X_i-X}{h}\right)K\left(\frac{y-Y_i}{h}\right)K\left(\frac{z-Z_i}{h}\right)dXdydz\right),$$

where

$$E\left(\iiint \frac{\mu(X,y,z)}{f(X)}\tilde{k}_j\left(\frac{X_i-X}{h}\right)K\left(\frac{y-Y_i}{h}\right)K\left(\frac{z-Z_i}{h}\right)dXdydz\right)^2 =$$
$$= h^{2L+2}\iiint \left[M_j(X,y,z)\right]^2 f(X,y,z)dXdydz$$

and

$$E\left(\iiint \frac{\mu(X,y,z)}{f(X)}\tilde{k}_j(X)K\left(\frac{y-Y_i}{h}\right)K\left(\frac{z-Z_i}{h}\right)dXdydz\right) =$$
$$= h^{L+1}\iiint M_j(X,y,z)f(X,y,z)dXdydz.$$

Thus,

$$V\left(\iiint \frac{\mu(X,y,z)}{f(X)}\int_{-\infty}^{y}\int_{-\infty}^{z}\frac{\partial h(X,w,t)}{\partial X^j}dtdwdXdydz\right) =$$
$$= \frac{1}{N}\left[\iiint f(X,y,z)M_j^2(X,y,z)dXdydz - \left(\iiint f(X,y,z)M_j(X,y,z)dXdydz\right)^2\right].$$

The result for the first covariance follows from

$$E\left(\int r(X)\frac{\partial h(X)}{\partial X^i}dX\iiint \frac{\mu(X,y,z)}{f(X)}\int_{-\infty}^{y}\int_{-\infty}^{z}\frac{\partial h(X,w,t)}{\partial X^j}dtdwdXdydz\right) =$$
$$= h^{2L+2}\iiint \frac{\partial r(X)}{\partial X^i}M_j(X,y,z)f(X,y,z)dXdydz.$$

The second covariance is obtained by analogy with variance. $\square$

Recall the notation:

$$G_i = \iiint \frac{\partial G_{Y,Z|X}(y,z)}{\partial X^i}\mu(X,y,z)dydzdX$$
$$\tilde{G} = \sum_{j=1}^{L}\left(\iiint \frac{\partial G_{Y,Z|X}(y,z)}{\partial X^j}\mu(X,y,z)dydzdX\right)^2$$
$$\bar{F}_{Y,Z|X}(X) = \iint F_{Y,Z|X}(y,z)\mu(X,y,z)dydz$$
$$\int \bar{F}_{Y,Z|X}(X)dX = F_{Y,Z|X}$$
$$\bar{F}_{Y,Z|XX}(X) = \frac{\bar{F}_{Y,Z|X}(X)}{f(X)}$$

and $F = (F_1, ..., F_L)'$. Also denote

$$V_{F_j} = V\left(M_j(X,y,z)\right) + V\left(\frac{\partial \bar{F}_{Y,Z|XX}(X)}{\partial X^j}\right) - 2\text{cov}\left(M_j(X,y,z), \frac{\partial \bar{F}_{Y,Z|XX}(X)}{\partial X^j}\right)$$

and

$$C_{ij} = \text{cov}(M_i(X,y,z), M_j(X,y,z)) - \text{cov}\left(M_i(X,y,z), \frac{\partial \bar{F}_{Y,Z|XX}(X)}{\partial X^j}\right)$$
$$-\text{cov}\left(M_j(X,y,z), \frac{\partial \bar{F}_{Y,Z|XX}(X)}{\partial X^i}\right) + \text{cov}\left(\frac{\bar{F}_{Y,Z|XX}(X)}{\partial X^i}, \frac{\bar{F}_{Y,Z|XX}(X)}{\partial X^j}\right).$$

Lemma 3.6. *Let $\hat{F}_j$ be the estimator of F_j. If Assumptions 3.3–3.7 hold, then,*

$$||\hat{F}_j - F_j|| \xrightarrow{p} 0 \text{ and } \sqrt{N}\left(\hat{F}_j - F_j\right) \xrightarrow{d} N\left(0, V_{F_j}\right).$$

In addition, $N\,cov(\hat{F}_i, \hat{F}_j) \to C_{ij}$.

Proof. From above,

$$\frac{\partial \widehat{F_{Y,Z|X}(y,z)}}{\partial X^j} - \frac{\partial F_{Y,Z|X}(y,z)}{\partial X^j} = \frac{1}{f(X)}\int_{-\infty}^{y}\int_{-\infty}^{z}\frac{\partial h(X,w,t)}{\partial X^j}dtdw - \frac{F_{Y,Z|X}(y,z)}{f(X)}\frac{\partial h(X)}{\partial X^j} + Res.$$

Thus, if we denote $r(X) = \bar{F}_{Y,Z|XX}(X)$ and use Lemmas 3.4 and 3.5, we obtain the desired result. The covariance follows by analogy.

□

Now define $\Phi(G) = \frac{\iiint \frac{\partial G_{Y,Z|X}(y,z)}{\partial X^i}\mu(X,y,z)dydzdX}{\sqrt{\sum_{j=1}^{L}\left(\iiint \frac{\partial G_{Y,Z|X}(y,z)}{\partial X^j}\mu(X,y,z)dydzdX\right)^2}}$. Then, $\Phi(\hat{F}) = \hat{\beta}_i$ and $\Phi(F) = \beta_i$.

For any h such that $||h||$ is sufficiently small: $|H_i| \leq a||h||$ for all i, and $|\tilde{F} + \tilde{H}| \geq b|\tilde{F}|$ for some $0 < a, b < \infty$. Also,

$$\Phi(F+H) - \Phi(F) = D\Phi(F,H) + R\Phi(F,H),$$

where

$$D\Phi(F,H) = H_i \tilde{F}^{-1.5}(\tilde{F} - F_i^2) - \sum_{j \neq i} H_j \tilde{F}^{-1.5} F_i F_j.$$

And for some $c < \infty$:

$$|D\Phi(F,H)| \leq \frac{c||h||}{\tilde{F}^{1.5}} \text{ and } |R\Phi(F,H)| \leq \frac{c||h||^2}{\tilde{F}^{2.5}}.$$

Thus, by Lemma B3 in Newey (1994), $\sup |\Phi(F+H) - \Phi(F)| \xrightarrow{p} 0$. Now for $H = \hat{F} - F$:

$$D\Phi(F, \hat{F} - F) = (\hat{F}_i - F_i)\tilde{F}^{-1.5}(\tilde{F} - F_i^2) - \sum_{j \neq i}(\hat{F}_j - F_j)\tilde{F}^{-1.5} F_i F_j.$$

By taking into account the results of Lemma 3.6,

$$\sqrt{N} D\Phi(F, \hat{F} - F) \xrightarrow{d} N(0, V_{\beta_i}),$$

where

$$V_{\beta_i} = \tilde{F}^{-3}\left(\left(\sum_{j \neq i} F_j^2\right)^2 V_{F_i} + F_i^2 \sum_{j \neq i} F_j^2 V_{F_j} - 2(\tilde{F} - F_i^2)F_i \sum_{j \neq i} F_j C_{ij} + F_i^2 \sum_{j \neq i, k \neq j, i} F_j F_k C_{jk}\right).$$

Now by Lemma B.3 in Newey (1994), $\sqrt{N} R\Phi(F,H) \xrightarrow{p} 0$ and it follows that

$$\sqrt{N}\left(\hat{\beta}_i - \beta_i\right) = \sqrt{N} D\Phi(F, \hat{F} - F) \xrightarrow{d} N(0, V_{\beta_i}).$$

3.5.4 Proof of Theorem 3.4

Denote

$$V_{dX^j} = V_X\left(\frac{\partial F_{Y,Z|X}(y,z)}{\partial X^j}\frac{\mu(X)}{f(X)}\right) - V_X(F_{Y,Z|X}(y,z)M_j(X)) +$$
$$+ E_X(M_j(X)F_{Y,Z|X}(y,z)(1 - F_{Y,Z|X}(y,z))).$$

Lemma 3.7. *If Assumptions 3.3–3.7 hold, then,*

$$\sqrt{N}\int\left(\frac{\partial \widehat{F_{Y,Z|X}(y,z)}}{\partial X^i} - \frac{\partial F_{Y,Z|X}(y,z)}{\partial X^i}\right)\mu(X)dX \xrightarrow{d} N\left(0, V_{dX^i}\right).$$

Proof. From before,

$$\frac{\partial \widehat{F_{Y,Z|X}(y,z)}}{\partial X^j} - \frac{\partial F_{Y,Z|X}(y,z)}{\partial X^j} = \frac{1}{f(X)}\int_{-\infty}^{y}\int_{-\infty}^{z}\frac{\partial h(X,w,t)}{\partial X^j}dtdw - \frac{F_{Y,Z|X}(y,z)}{f(X)}\frac{\partial h(X)}{\partial X^j} + Res.$$

Thus, if we denote $r(X) = \frac{F_{Y,Z|X}(y,z)\mu(X)}{f(X)}$ and use Lemma 3.4:

$$\sqrt{N}\int\frac{F_{Y,Z|X}(y,z)\mu(X)}{f(X)}\frac{\partial h(X)}{\partial X^j}dX \xrightarrow{d}$$
$$\xrightarrow{d} N\left(0, V_X\left(\frac{\partial F_{Y,Z|X}(y,z)}{\partial X^j}\frac{\mu(X)}{f(X)} + F_{Y,Z|X}(y,z)M_j(X)\right)\right).$$

By analogy with Lemma 3.5, one can obtain

$$\sqrt{N}\int\frac{\mu(X)}{f(X)}\int_{-\infty}^{y}\int_{-\infty}^{z}\frac{\partial h(X,w,t)}{\partial X^j}dtdwdX \xrightarrow{d}$$
$$\xrightarrow{d} N\left(0, \left[E_X(M_j^2(X)F_{Y,Z|X}(y,z)) - \left(E_X(M_j(X)F_{Y,Z|X}(y,z))\right)^2\right]\right)$$

and

$$
\begin{aligned}
&N\text{cov}\left(\int \frac{\mu(X)}{f(X)} \int_{-\infty}^{y} \int_{-\infty}^{z} \frac{\partial h(X,w,t)}{\partial X^j} dt dw dX, \int \frac{F_{Y,Z|X}(y,z)\mu(X)}{f(X)} \frac{\partial h(X)}{\partial X^j} dX\right) = \\
&N\text{cov}\left(M_j(X_i)K\left(\frac{y-Y_i}{h}\right)K\left(\frac{z-Z_i}{h}\right), \frac{\partial F_{Y,Z|X}(y,z)}{\partial X^j}\frac{\mu(X_i)}{f(X_i)} + F_{Y,Z|X_i}(y,z)M_j(X_i)\right) \to \\
&\quad \to \text{cov}_X\left(M_j(X)F_{Y,Z|X}(y,z), \frac{\partial F_{Y,Z|X}(y,z)}{\partial X^j}\frac{\mu(X_i)}{f(X_i)} + F_{Y,Z|X_i}(y,z)M_j(X_i)\right) = \\
&\quad = \text{cov}_X\left(M_j(X)F_{Y,Z|X}(y,z), \frac{\partial F_{Y,Z|X}(y,z)}{\partial X^j}\frac{\mu(X_i)}{f(X_i)}\right) + V_X(M_j(X)F_{Y,Z|X}(y,z)).
\end{aligned}
$$

And the result will follow.

□

Recall the notation $\phi(z) = g(G^{-1}(z))$ and define

$$
V_{\phi(z)} = \frac{\ddot{F}^2_{Z|X}(z)}{\beta_i^2} V_{az}.
$$

Lemma 3.8. *Let $\hat{\phi(z)}$ be the estimator of $\phi(z)$. Then, under Assumptions 3.3–3.7,*

$$
\sup_z \left|\hat{\phi}(z) - \phi(z)\right| \xrightarrow{p} 0
$$

$$
\sqrt{Nh}\left(\hat{\phi(z)} - \phi(z)\right) \xrightarrow{d} N(0, V_{\phi(z)}).
$$

Proof. Define the functional

$$
\Phi(G) = \frac{\int \frac{\partial G_{Z|X}(z)}{\partial X^i}\mu(X)dX}{\hat{\beta}_i(G)\int \frac{\partial G_{Z|X}(z)}{\partial z}\mu(X)dX}.
$$

Then, $\Phi(\hat{F}) = \hat{\phi}(z)$ and $\Phi(F) = \phi(z)$. In order to simplify the notation, denote

$$
\ddot{F}_{Z|X}(z) = \int \frac{\partial F_{Z|X}(z)}{\partial X^i}\mu(X)dX
$$

$$H_X = \int \frac{\partial H_{Z|X}(z)}{\partial X^i} \mu(X) dX$$
$$H_\beta = \hat{\beta}_i - \beta_i$$
$$H_z = \int h_{Z|X}(z) \mu(X) dX.$$

For any h such that $||h||$ is sufficiently small: $|H_X| \leq a||h||$, $|H_\beta| \leq a||h||$, $|H_z| \leq a||h||$ for some $0 < a < \infty$. Also,

$$\Phi(F+H) - \Phi(F) = D\Phi(F,H) + R\Phi(F,H),$$

where

$$D\Phi(F,H) = \frac{\phi(z) H_X}{\ddot{F}_{Z|X}(z)} - \frac{\phi(z)}{\beta_i} H_\beta - \frac{\ddot{F}_{Z|X}(z)}{\beta_i} H_z.$$

Thus, for some $c < \infty$:

$$|D\Phi(F,H)| \leq c||h|| \text{ and } |R\Phi(f,h)| \leq c||h||^2.$$

It implies that $\sup |\Phi(\hat{F}) - \Phi(F)| \xrightarrow{p} 0$. By Lemma 3.1, H_z converges with the speed $\sqrt{Nh}$; by Lemma 3.7, H_X converges as $\sqrt{N}$; and by Theorem 3.3, H_β has rate $\sqrt{N}$. Thus, only H_z defines the limiting distribution of $D\Phi(F,H)$, hence,

$$\sqrt{Nh} D\Phi(F, \hat{F} - F) \xrightarrow{d} N(0, V_{\phi(z)}),$$

where $V_{\phi(z)} = \frac{\ddot{F}^2_{Z|X}(z)}{\beta_i^2} V_{az}$.

Now by Lemma B.3 in Newey (1994), $\sqrt{Nh} R\Phi(F,H) \xrightarrow{p} 0$ and it follows that

$$\sqrt{Nh}\left(\hat{\phi(z)} - \phi(z)\right) \xrightarrow{d} N(0, V_{\phi(z)}).$$

$\square$

Lemma 3.9. *If Assumptions 3.3–3.7 hold, then,*

$$N\,cov\left(\int_0^{u_1}\int h_{Z|X}(z)r(X,z)dXdz, \int_0^{u_2}\int h_{Z|X}(z)\tilde{r}(X,z)dXdz\right) \to$$
$$\to cov\left(r(X,z)f_{Z|X}(z)\mathbb{1}(z<u_1), \tilde{r}(X,z)f_{Z|X}(z)\mathbb{1}(z<u_2)\right).$$

Proof. Note the following:

$$N\int_0^{u_1}\int h_{Z|X}(z)r(X,z)dXdz = \frac{1}{h^{L+1}}\sum\int_0^{u_1}\int r(X,z)\tilde{k}\left(\frac{X_i-X}{h}\right)k\left(\frac{Z_i-z}{h}\right)dXdz$$
$$= \frac{h^{L+1}}{h^{L+1}}\sum\int_0^{u_1}\int r(X_i-\sigma_x h, Z_i-\sigma_z h)\tilde{k}(\sigma_x)k(\sigma_z)d\sigma_x d\sigma_z,$$

where

$$\int_0^{u_1}\int r(X_i-\sigma_x h, Z_i-\sigma_z h)\tilde{k}(\sigma_x)k(\sigma_z)d\sigma_x d\sigma_z \to r(X_i,Z_i)\mathbb{1}(Z_i<u_1).$$

And the result will follow. □

Recall the following notation:

$$V_u = V\left(f_{Z|X}(z)\frac{\ddot{F}_{Z|X}(z)\mu(X)}{\beta_i\phi^2(z)}\mathbb{1}(z<u)\right)$$
$$V_c = V\left(f_{Z|X}(z)\frac{(1-z)\ddot{F}_{Z|X}(z)\mu(X)}{\beta_i\phi^2(z)}\right)$$

and

$$\Sigma_{uc} = \text{cov}\left(f_{Z|X}(z)\frac{\ddot{F}_{Z|X}(z)\mu(X)}{\beta_i\phi^2(z)}\mathbb{1}(z<u), f_{Z|X}(z)\frac{(1-z)\ddot{F}_{Z|X}(z)\mu(X)}{\beta_i\phi^2(z)}\right).$$

Lemma 3.10. *If Assumptions 3.3–3.7 hold, then,*

$$\sqrt{N}\int_0^u\frac{\hat{\phi}(z)-\phi(z)}{\phi^2(z)}dz \xrightarrow{d} N\left(0,V_u\right)$$

$$\sqrt{N}\int_0^1 \frac{(1-z)(\hat{\phi}(z)-\phi(z))}{\phi^2(z)}dz \xrightarrow{d} N\left(0, V_c\right)$$
$$N cov\left(\int_0^u \frac{\hat{\phi}(z)}{\phi^2(z)}dz, \int_0^1 \frac{(1-z)\hat{\phi}(z)}{\phi^2(z)}dz\right) \to \Sigma_{uc}$$
$$N cov\left(\int_0^u \frac{\hat{\phi}(z)}{\phi^2(z)}dz, \hat{\phi}(z)\right) \to 0 \ and \ N cov\left(\int_0^1 \frac{(1-z)\hat{\phi}(z)}{\phi^2(z)}dz, \hat{\phi}(z)\right) \to 0.$$

Proof. Recall from Lemma 3.8 that

$$\sqrt{Nh}(\hat{\phi}(z)-\phi(z)) = \sqrt{Nh}\left(-\frac{\ddot{F}_{Z|X}(z)}{\beta_i}H_z\right) + o_p(1).$$

To obtain V_u use the previous lemma with $r(X,z) = \frac{\ddot{F}_{Z|X}(z)\mu(X)}{\beta_i\phi^2(z)}$, for V_c take $r(X,z) = (1-z)\frac{\ddot{F}_{Z|X}(z)\mu(X)}{\beta_i\phi^2(z)}$ and for Σ_{uc} substitute both functions accordingly.

To obtain zero covariances, notice that $\hat{\phi}(z)$ converges slower than the integrals.

□

Proof of Theorem 3.4. Define the functional $\Phi(g) = \int_0^u \frac{1}{g(z)}dz - \int_0^1 \frac{1-z}{g(z)}dz$. Then, $\Phi(\hat{\phi}) = \hat{\phi}(z)$ and $\Phi(\phi) = \phi(z)$. For any h such that $||h||$ is sufficiently small: $|\int_0^u h_\phi(z)dz| \leq a||h||$, $|\phi(z)| > \phi$ and $|\phi(z)+h_\phi(z)| \geq b|\phi(z)|$ for some $0 < a, b < \infty$. Also,

$$\Phi(\phi+h_\phi) - \Phi(\phi) = D\Phi(\phi, h_\phi) + R\Phi(\phi, h_\phi),$$

where

$$D\Phi(\phi, h_\phi) = -\int_0^u \frac{h_\phi(z)}{\phi^2(z)}dz + \int_0^1 \frac{(1-z)h_\phi(z)}{\phi^2(z)}dz$$
$$R\Phi(\phi, h_\phi) = \int_0^u \frac{h_\phi^2(z)}{\phi^2(z)(\phi(z)+h_\phi(z))}dz - \int_0^1 \frac{(1-z)h_\phi^2(z)}{\phi^2(z)(\phi(z)+h_\phi(z))}dz.$$

Thus, for some $c < \infty$:

$$|D\Phi(\phi, h_\phi)| \leq \frac{c||h||}{\phi^2} \text{ and } |R\Phi(\phi, h_\phi)| \leq \frac{c||h||^2}{\phi^3}.$$

And the uniform convergence follows. Now by Lemma 3.10:

$$\begin{aligned}
\sqrt{N}\int_0^u \frac{h_\phi(z)}{\phi^2(z)}dz &\xrightarrow{d} N\left(0, V_u\right) \\
\sqrt{N}\int_0^1 \frac{(1-z)h_\phi(z)}{\phi^2(z)}dz &\xrightarrow{d} N\left(0, V_c\right) \\
N\text{cov}\left(\int_0^u \frac{h_\phi(z)}{\phi^2(z)}dz, \int_0^1 \frac{(1-z)h_\phi(z)}{\phi^2(z)}dz\right) &\rightarrow \Sigma_{uc}.
\end{aligned}$$

Thus, $\sqrt{N}D\Phi(\phi, h_\phi) \xrightarrow{d} N(0, V_{G^{-1}(u)})$, where $V_{G^{-1}(u)} = V_u + V_c - 2\Sigma_{uc}$. By Lemma B.3 in Newey (1994), $\sqrt{N}R(\phi, h_\phi) \xrightarrow{p} 0$ and it follows that

$$\sqrt{N}\left(\widehat{G^{-1}(u)} - G^{-1}(u)\right) \xrightarrow{d} N(0, V_{G^{-1}(u)}).$$

□

3.5.5 Proof of Theorem 3.5.

Lemma 3.11. *If Assumptions 3.3–3.7 hold, then*

$$Nh\,cov\left(\int_0^u \frac{1}{\hat{\phi}(z)}dz, \hat{\phi}(z)\right) \rightarrow 0 \text{ and } Nh\,cov\left(\int_0^1 \frac{1-z}{\hat{\phi}(z)}dz, \hat{\phi}(z)\right) \rightarrow 0.$$

Proof. Notice that

$$\sqrt{N}\int_0^u \frac{1}{\hat{\phi}(z) - \phi(z)}dz = \sqrt{N}\int_0^u \frac{-\hat{\phi}(z) + \phi(z)}{\phi^2(z)}dz + o_p(1).$$

Then, by Lemma 3.10, the results follow. □

Lemma 3.12. *Let $\hat{g}(G^{-1}(z))$ and $\widehat{G^{-1}(u)}$ be the estimators of $g(G^{-1}(z))$ and*

$G^{-1}(u)$. *Then, under Assumptions 3.3–3.7,*

$$\sup_z \left| \hat{g}(\widehat{G^{-1}(z)}) - g(G^{-1}(z)) \right| \xrightarrow{p} 0$$
$$\sqrt{Nh}\left(\hat{g}(\widehat{G^{-1}(z)}) - g(G^{-1}(z))\right) \xrightarrow{d} N(0, V_{g(G^{-1}(z))}).$$

Proof. First, notice the following:

$$\begin{aligned}\hat{g}(\widehat{G^{-1}(z)}) - g(G^{-1}(z)) &= \hat{g}(\widehat{G^{-1}(z)}) - \hat{g}(G^{-1}(z)) + \hat{g}(G^{-1}(z)) - g(G^{-1}(z)) \\ &= \left.\frac{\partial \hat{g}(t)}{\partial t}\right|_{t=G^{-1}(z)} \left(\widehat{G^{-1}(z)} - G^{-1}(z)\right) + \hat{\phi}(z) - \phi(z) + Res,\end{aligned}$$

where $|Res| \leq a||\widehat{G^{-1}(z)} - G^{-1}(z)||^2$ for some $0 < a < \infty$. In addition,

$$\left.\frac{\partial \hat{g}(t)}{\partial t}\right|_{t=G^{-1}(z)} = \hat{\phi}'(z)\hat{\phi}(z),$$

and, as a result,

$$\begin{aligned}\hat{\phi}'(z)\hat{\phi}(z)\left(\widehat{G^{-1}(z)} - G^{-1}(z)\right) &= \phi'(z)\phi(z)\left(\widehat{G^{-1}(z)} - G^{-1}(z)\right) + \phi(z)h_{\phi'}(z)\left(\widehat{G^{-1}(z)} - G^{-1}(z)\right) \\ &+\phi'(z)h_{\phi}\left(\widehat{G^{-1}(z)} - G^{-1}(z)\right) + h_{\phi}h_{\phi'}\left(\widehat{G^{-1}(z)} - G^{-1}(z)\right).\end{aligned}$$

By Lemma 3.8, $\sup |h_{\phi}| \xrightarrow{p} 0$, and by Theorem 3.4, $\sup \left(\widehat{G^{-1}(z)} - G^{-1}(z)\right) \xrightarrow{p} 0$, thus, we need to show only that $|h_{\phi'}|$ is bounded for uniform convergence of $\hat{g}(\widehat{G^{-1}(z)}) - g(G^{-1}(z))$.

Notice that

$$\phi'(z) = \phi(z)\left(\frac{\int \frac{\partial f_{Z|X}(z)}{\partial X^i}\mu(X)dX}{\int \frac{\partial F_{Z|X}(z)}{\partial X^i}\mu(X)dX} - \frac{\int \frac{\partial f_{Z|X}(z)}{\partial z}\mu(X)dX}{\int f_{Z|X}(z)\mu(X)dX}\right).$$

By Taylor expansion, we obtain

$$h_{\phi'} = \hat{\phi}'(z) - \phi'(z) = \frac{\phi'(z)}{\phi(z)} h_\phi + \frac{\phi(z)}{\int \frac{\partial F_{Z|X}(z)}{\partial X^i} \mu(X) dX} h_1 - \phi(z) \frac{\int \frac{\partial f_{Z|X}(z)}{\partial X^i} \mu(X) dx}{\left[\int \frac{\partial F_{Z|X}(z)}{\partial X^i} \mu(X) dX\right]^2} H_X$$
$$- \frac{\phi(z)}{\int f_{Z|X}(z) \mu(X) dX} h_2 + \phi(z) \frac{\int \frac{\partial f_{Z|X}(z)}{\partial z} \mu(X) dX}{\left[\int f_{Z|X}(z) \mu(X) dX\right]^2} h_3 + Res_2.$$

Now we show that $|h_1|$ is bounded. Note that

$$\frac{\partial f_{Z|X}(z)}{\partial X^i} = \frac{1}{f(X)} \frac{\partial f(X,z)}{\partial X^i} - \frac{f(X,z)}{f^2(X)} \frac{\partial f(X)}{\partial X^i}.$$

Then,

$$h_1 = \int \frac{\partial h_{Z|X}(z)}{\partial X^i} \mu(X) dX = \int \frac{\mu(X)}{f(X)} \frac{\partial h(X,z)}{\partial X^i} dX - \int \frac{h(X)\mu(X)}{f^2(X)} \frac{\partial f(X,z)}{\partial X^i} dX$$
$$- \int \frac{\mu(X)}{f^2(X)} \frac{\partial f(X)}{\partial X^i} h(X,z) dX - \int \frac{\mu(X)}{f^2(X)} \frac{\partial h(X)}{\partial X^i} f(X,z) dX +$$
$$+2 \int \frac{h(X)\mu(X)}{f^3(X)} \frac{\partial f(X)}{\partial X^i} f(X,z) dX + Res_3.$$

For any sufficiently small $||h||$: $|\int r(X) \frac{\partial h(X,z)}{\partial X^i} dX| \leq a||h||$, $|\int h(X)\tilde{r}(X) dX| \leq a||h||$, as well as $|\int \bar{r}(X) h(X,z) dX| \leq a||h||$ and $|\int \ddot{r}(X) \frac{\partial h(X)}{\partial X^i} dX| \leq a||h||$ for some $0 < a < \infty$. Moreover, by Taylor Theorem, $|Res_3| \leq b||h||^2$ for some $0 < b < \infty$. Thus, for some $0 < c < \infty$, $|h_1| \leq c||h|| + b||h||^2$.

Note that by Lemma 3.1 and 3.7, $|H_X| \leq d||h|| + d||h||^2$ and $|h_3| \leq d||h|| + d||h||^2$ for some $0 < d < \infty$. For the uniform convergence of $\hat{g}(\widehat{G^{-1}(z)}) - g(G^{-1}(z))$, we are left to show boundedness of h_2. Notice that

$$\frac{\partial f_{Z|X}(z)}{\partial z} = \frac{1}{f(X)} \frac{\partial f(X,z)}{\partial z}.$$

Hence,

$$h_3 = \int \frac{\partial h_{Z|X}(z)}{\partial z}\mu(X)dX + Res_4 =$$
$$\int \frac{\mu(X)}{f(X)}\frac{\partial h(X,z)}{\partial z}dX - \int \frac{h(X)\mu(X)}{f^2(X)}\frac{\partial f(X,z)}{\partial z}dX + Res_4 + Res_5.$$

For some $0 < f < \infty$, by Taylor Theorem and previous results, $|h_3| \leq e||h|| + e||h||^2$, implying that for some $0 < f < \infty$, by Taylor theorem, $|Res_2| \leq f||h||^2$, $\left| \frac{\partial \hat{g}(t)}{\partial t}\Big|_{t=G^{-1}(z)} \right| \leq f||h|| + f||h||^2$ and $\left| \hat{g}(\widehat{G^{-1}(z)}) - g(G^{-1}(z)) \right| \leq f||h|| + f||h||^2$.

We know from Theorem 3.4 and Lemma 3.8 that $\hat{\phi(z)}$ converges slower than $\widehat{G^{-1}(z)}$, and by Lemma 3.11, $Nh\text{cov}\left(\widehat{G^{-1}(z)}, \hat{\phi}(z)\right) \to 0$. Thus,

$$\sqrt{Nh}\left(\hat{g}(\widehat{G^{-1}(z)}) - g(G^{-1}(z))\right) \xrightarrow{d} N\left(0, V_{g(G^{-1}(z))}\right),$$

where $V_{g(G^{-1}(z))} = V_{\phi(z)}$. □

Denote $\theta = \left(-\frac{1}{2}(2\sum_{i=1}^{L}\beta_i X^i + G^{-1}(y) + G^{-1}(z)), \frac{1}{2}(G^{-1}(z) - G^{-1}(y))\right)$.

Lemma 3.13. *Let $\hat{f}_{\delta,c}(\theta)$ be the estimator of $f_{\delta,c}(\theta)$. Then, under Assumptions 3.3–3.7,*

$$\sup_{\theta}\left|\hat{f}_{\delta,c}(\theta) - f_{\delta,c}(\theta)\right| \xrightarrow{p} 0$$
$$\sqrt{Nh^{L+2}}(\hat{f}_{\delta,c}(\theta) - f_{\delta,c}(\theta)) \xrightarrow{d} N(0, 4\phi^2(y)\phi^2(z)V_{ayz}).$$

Proof. Define the functional $\Phi(s,G) = 2s(y)s(z)g_{Y,Z|X}(y,z)$. Then, $\Phi(\hat{\phi},\hat{F}) = \hat{f}_{\delta,c}(\theta)$ and $\Phi(\phi,F) = f_{\delta,c}(\theta)$. For any h such that $||h||$ is sufficiently small: $|h(y)| \leq a||h||$ and $|h_{Y,Z|X}(y,z)| \leq a||h||$ for some $0 < a < \infty$. Also,

$$\Phi(\phi + h, F + H) - \Phi(\phi, F) = D\Phi(\phi, F, h, H) + R\Phi(\phi, F, h, H),$$

where

$$D\Phi(\phi, F, h, H) = 2\phi(y)\phi(z)h_{Y,Z|X}(y, z) + 2\phi(y)f_{Y,Z|X}(y, z)h_g(z) + 2\phi(z)f_{Y,Z|X}(y, z)h_g(y).$$

Thus, for some $c < \infty$:

$$|D\Phi(\phi, F, h, H)| \leq c||h|| \text{ and } |R\Phi(\phi, F, h, H)| \leq c||h||^2.$$

By Lemma 3.8, $h(z)$ converges with rate $\sqrt{Nh}$, while by Theorem 3.2, $h_{Y,Z|X}(y, z)$ has speed $\sqrt{Nh^{L+2}}$. Thus,

$$\sqrt{Nh^{L+2}}D\Phi(\phi, F, h, H) \xrightarrow{d} N(0, 4\phi^2(y)\phi^2(z)V_{ayz}).$$

Now by Lemma B.3 in Newey (1994), $\sqrt{Nh^{L+2}}R\Phi(\phi, F, h, H) \xrightarrow{p} 0$ and it follows that

$$\sqrt{Nh^{L+2}}(\hat{f}_{\delta,c}(\theta) - f_{\delta,c}(\theta)) \xrightarrow{d} N(0, 4\phi^2(y)\phi^2(z)V_{ayz}).$$

□

Proof of Theorem 3.5. Note that by Taylor expansion,

$$\begin{aligned}
\hat{f}_{\delta,c}(\hat{\theta}) - f_{\delta,c}(\theta) &= \hat{f}_{\delta,c}(\hat{\theta}) - \hat{f}_{\delta,c}(\theta) + \hat{f}_{\delta,c}(\theta) - f_{\delta,c}(\theta) \\
&= \frac{\partial \hat{f}_{\delta,c}(\theta)}{\partial \theta_1}(\hat{\theta}_1 - \theta_1) + \frac{\partial \hat{f}_{\delta,c}(\theta)}{\partial \theta_2}(\hat{\theta}_2 - \theta_2) + \hat{f}_{\delta,c}(\theta) - f_{\delta,c}(\theta) + Res \\
&= \frac{\partial f_{\delta,c}(\theta)}{\partial \theta_1}(\hat{\theta}_1 - \theta_1) + \frac{\partial f_{\delta,c}(\theta)}{\partial \theta_2}(\hat{\theta}_2 - \theta_2) + \hat{f}_{\delta,c}(\theta) - f_{\delta,c}(\theta) \\
&\quad + \frac{\partial h_{\delta,c}(\theta)}{\partial \theta_1}(\hat{\theta}_1 - \theta_1) + \frac{\partial h_{\delta,c}(\theta)}{\partial \theta_2}(\hat{\theta}_2 - \theta_2) + Res,
\end{aligned}$$

where $\theta_1 = -0.5\left(2\sum_{i=1}^{L}\beta_i X^i + G^{-1}(y) + G^{-1}(z)\right)$ and $\theta_2 = 0.5(G^{-1}(z) - G^{-1}(y))$. Denote $R(f_{\delta,c}, h_{\delta,c}) = \frac{\partial h_{\delta,c}(\theta)}{\partial \theta_1}(\hat{\theta}_1 - \theta_1) + \frac{\partial h_{\delta,c}(\theta)}{\partial \theta_2}(\hat{\theta}_2 - \theta_2) + Res.$

Now if we show that $\left|\frac{\partial h_{\delta,c}(\theta)}{\partial \theta_1}\right|$ and $\left|\frac{\partial h_{\delta,c}(\theta)}{\partial \theta_2}\right|$ are bounded by $||h||$, then $|Res|$

and $\left|\frac{\partial h_{\delta,c}(\theta)}{\partial \theta_i}(\hat{\theta}_i - \theta_i)\right|$ will be bounded by $||h||^2$ due to the Taylor Theorem. As a result, we will obtain uniform convergence of $\hat{f}_{\delta,c}(\hat{\theta}) - f_{\delta,c}(\theta)$.

First, notice that

$$\frac{\partial \theta_1}{\partial y} = \frac{\partial \theta_2}{\partial y} = -\frac{1}{2\phi(y)}; \ \frac{\partial \theta_1}{\partial z} = -\frac{1}{2\phi(z)}; \ \frac{\partial \theta_2}{\partial z} = \frac{1}{2\phi(z)}.$$

Moreover,

$$\frac{\partial h_{\delta,c}(\theta)}{\partial y} = \frac{\partial h_{\delta,c}(\theta)}{\partial \theta_1}\frac{\partial \theta_1}{\partial y} + \frac{\partial h_{\delta,c}(\theta)}{\partial \theta_2}\frac{\partial \theta_2}{\partial y} = -\frac{1}{2\phi(y)}\left[\frac{\partial h_{\delta,c}(\theta)}{\partial \theta_1} + \frac{\partial h_{\delta,c}(\theta)}{\partial \theta_2}\right]$$

and

$$\frac{\partial h_{\delta,c}(\theta)}{\partial z} = \frac{\partial h_{\delta,c}(\theta)}{\partial \theta_1}\frac{\partial \theta_1}{\partial z} + \frac{\partial h_{\delta,c}(\theta)}{\partial \theta_2}\frac{\partial \theta_2}{\partial z} = -\frac{1}{2\phi(z)}\left[\frac{\partial h_{\delta,c}(\theta)}{\partial \theta_1} - \frac{\partial h_{\delta,c}(\theta)}{\partial \theta_2}\right].$$

Hence,

$$\frac{\partial h_{\delta,c}(\theta)}{\partial \theta_1} = -\phi(y)\frac{\partial h_{\delta,c}(\theta)}{\partial y} - \phi(z)\frac{\partial h_{\delta,c}(\theta)}{\partial z}$$

and

$$\frac{\partial h_{\delta,c}(\theta)}{\partial \theta_2} = -\phi(y)\frac{\partial h_{\delta,c}(\theta)}{\partial y} + \phi(z)\frac{\partial h_{\delta,c}(\theta)}{\partial z}.$$

In addition, $f_{\delta,c}(\theta) = 2\phi(y)\phi(z)f_{Y,Z|X}(y,z)$, implying that

$$\frac{\partial h_{\delta,c}(\theta)}{\partial y} = 2f_{Y,Z|X}(y,z)\phi'(y)h_{\phi} + 2f_{Y,Z|X}(y,z)\phi(z)h_{\phi'} + 2\phi'(y)\phi(z)h_{Y,Z|X}(y,z)$$
$$+2\phi(y)\phi(z)\frac{\partial h_{Y,Z|X}(y,z)}{\partial y} + 2\phi(y)\frac{\partial f_{Y,Z|X}(y,z)}{\partial y}h_{\phi} + 2\phi(z)\frac{\partial f_{Y,Z|X}(y,z)}{\partial y}h_{\phi} + Res_2.$$

Notice that for sufficiently small $||h||$ and some $0 < a < \infty$: $\left|\frac{\partial h_{\delta,c}(\theta)}{\partial y}\right| \leq a||h|| + a||h||^2$ by previous results and Taylor Theorem. The same result holds for $\left|\frac{\partial h_{\delta,c}(\theta)}{\partial y}\right|$

by analogy, and as a result, for $\frac{\partial h_{\delta,c}(\theta)}{\partial \theta_i}$. Thus, $\sup \left| \hat{f}_{\delta,c}(\hat{\theta}) - f_{\delta,c}(\theta) \right| \xrightarrow{p} 0$.

Now recall that by Theorems 3.3 and 3.4, $(\hat{\theta}_1 - \theta_1)$ and $(\hat{\theta}_2 - \theta_2)$ converge with speed $\sqrt{N}$, while by Lemma 3.13, $\hat{f}_{\delta,c}(\theta) - f_{\delta,c}(\theta)$ has rate $\sqrt{Nh^{L+2}}$. Additionally, $\sqrt{Nh^{L+2}} R(f_{\delta,c}, h_{\delta,c}) \xrightarrow{p} 0$ as before by Lemma B3 in Newey (1994). Thus, the result follows:

$$\sqrt{Nh^{L+2}} \left(\hat{f}_{\delta,c}(\hat{\theta}) - f_{\delta,c}(\theta) \right) \xrightarrow{d} N(0, 4\phi^2(y)\phi^2(z)V_{ayz}).$$

□

References

[1] Afriat, S. *"The Construction of Utility Functions from Expenditure Data"*, International Economic Review 8 (1), 67-77, 1967.

[2] Ahn, D., and H. Ergin *"Framing Contingencies"*, Econometrica 78 (2), 655-695, 2010.

[3] Brown, D., and R. Matzkin *"Testable Restrictions on the Equilibrium Manifold"*, Econometrica 64 (6), 1249-1262, 1996.

[4] Coate, S. and M. Conlin *"A Group Rule-Utilitarian Approach to Voter Turnout: Theory and Evidence"*, American Economic Review 94 (5), 1476-1504, 2004.

[5] De Paula, A. and A. Merlo *"Identification and Estimation of Preference Distributions When Voters are Ideological"*, CEPR Discussion Paper No. DP10821, 2015.

[6] Degan, A. *"Candidate Valence: Evidence from Consecutive Presidential Elections"*, International Economic Review 48 (2), 457-482, 2007.

[7] Degan, A. and A. Merlo *"A Structural Model of Turnout and Voting in Multiple Elections"*, Journal of the European Economic Association 9 (2), 209-245, 2011.

[8] Dekel, E., B. Lipman, and A. Rustichini *"Representing Preferences With a Unique Subjective State Space"*, Econometrica 69 (4), 891-934, 2001.

[9] Diewert, W. *"Afriat's Theorem and some Extensions to Choice under Uncertainty"*, Economic Journal 122 (560), 305-331, 2012.

[10] Downs, A. *"An Economic Theory of Political Action in a Democracy"*, Journal of Political Economy 65 (2), 135-150, 1957.

[11] El-Gamal, M., and D. Grether *"Are People Bayesian? Uncovering Behavioral Strategies"*, Journal of the American Statistical Association 90, 1137-1145, 1995.

[12] Ellsberg, D. *"Risk, Ambiguity, and the Savage Axioms"*, Quarterly Journal of Economics 75, 643-669, 1961.

[13] Epstein, L., and M. LeBreton *"Dynamically Consistent Beliefs Must Be Bayesian"*, Journal of Economic Theory 61(1), 1-22, 1993.

[14] Epstein, L., M. Marinacci, and K. Seo *"Coarse Contingencies and Ambiguity"*, Theoretical Economics 2, 355-394, 2007.

[15] Epstein, L., and M. Schneider *"Recursive Multiple-Priors"*, Journal of Economic Theory 113, 1-31, 2003.

[16] Epstein, L., and S. Zin *"Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework"*, Econometrica 57 (4), 937-969, 1989.

[17] Epstein, L., and S. Zin *"Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: An Empirical Analysis"*, Journal of Political Economy 99 (2), 263-286, 1991.

[18] Echenique, F., and K. Saito *"Savage in the Market"*, Econometrica 83(4), 1467-1495, 2015.

[19] Ghirardato, P. *"Coping with Ignorance: Unforeseen Contingencies and Non-Additive Uncertainty"*, Economic Theory 17, 247-276, 2001.

[20] Ghirardato, P. *"Revisiting Savage in a Conditional World"*, Economic Theory 20, 83-92, 2002.

[21] Ghirardato, P., F. Maccheroni, and M. Marinacci *"Differentiating Ambiguity and Ambiguity Attitude"*, Journal of Economic Theory 118, 133-173, 2004.

[22] Gilboa, I., and D. Schmeidler *"Maxmin Expected Utility with Non-Unique Prior"*, Journal of Mathematical Economics 18 (2), 141-153, 1989.

[23] Grant, S., A. Kajii, and B. Polak *"Intrinsic Preference for Informations"*, Journal of Economic Theory 83, 233-259, 1998.

[24] Grant, S., A. Kajii, and B. Polak *"Temporal Resolution of Uncertainty and Recursive Non-Expected Utility Models"*, Econometrica 68 (2), 425-434, 2000.

[25] Halevy, Y. *"Ellsberg Revisited: An Experimental Study"*, Econometrica 75 (2), 503-536, 2007.

[26] Hayashi, T. *"Intertemporal Substitution, Risk Aversion and Ambiguity Aversion"*, Economic Theory 25, 933-956, 2005.

[27] Hinich, M.J. and M.C. Munger *"Ideology and the Theory of Public Choice"*, Ann Arbor: University of Michigan Press, 1996.

[28] Houser, D., M. Keane, and K. McCabe *"Behavior in a Dynamic Decision Problem: An Analysis of Experimental Evidence Using a Bayesian Type Classification Algorithm"*, Econometrica 72 (3), 781-822, 2004.

[29] Kahneman, D., and A. Tversky *"Prospect Theory: An Analysis of Decision under Risk"*, Econometrica 47 (2), 263-292, 1979.

[30] Karni, E., and D. Schmeidler *"Atemporal Dynamic Consistency and Expected Utility Theory"*, Journal of Economic Theory 54, 401-408, 1991.

[31] Kernell, G. *"Giving Order to Districts: Estimating Voter Distributions with National Election Returns"*, Political Analysis 17 (3), 215-235, 2009.

[32] Klibanoff, P., M. Marinacci, and S. Mukerji *"Recursive Smooth Ambiguity Preferences"*, Journal of Economic Theory 144 (3), 930-976, 2009.

[33] Kreps, D. *"Static Choice and Unforeseen Contingencies"*, Economic Analysis of Markets and Games: Essays in Honor of Frank Hahn, ed. by P. Dasgupta, D. Gale, O. Hart, and E. Maskin, Cambridge, MA: MIT Press, 259-281, 1992.

[34] Kreps, D., and E. Porteus *"Temporal Resolution of Uncertainty and Dynamic Choice Theory"*, Econometrica 46, 185-200, 1978.

[35] Kubler, F., L. Selden, and X. Wei *"Asset Demand Based Tests of Expected Utility Maximization"*, American Economic Review 104(11), 3459-3480, 2014.

[36] Li, J. *"Preferences for Information and Ambiguity"*, job market paper, 2011.

[37] Li, Q. and J.S. Racine *"Nonparametric Econometrics"*, Princeton University Press, 2007.

[38] Lindbeck, A. and J.W. Weibull, *"Balanced-Budget Redistribution as the Outcome of Political Competition"*, Public Choice 52, 273-297, 1987.

[39] Machina, M. *"Dynamic Consistency and Non-Expected Utility Models of Choice under Uncertainty"*, Journal of Economic Literature 27, 1622-1668, 1989.

[40] McMurray, J. *"Aggregating Information by Voting: The Wisdom of the Experts versus the Wisdom of the Masses"*, Review of Economic Studies 80, 277-312, 2013.

[41] Mukerji, S. *"Understanding the Nonadditive Probability Decision Model"*, Economic Theory 9, 23-46, 1997.

[42] Newey, W. *"Kernel Estimation of Partial Means and a General Variance Estimator"*, Econometric Theory 10 (2), 1-21, 1994.

[43] Persson, T. and G. Tabellini *"Political Economics: Explaining Economic Policy"*, MIT Press, 2000.

[44] Polisson, M., J. Quah, and L. Renou *"Revealed Preferences over Risk and Uncertainty"*, Working Paper, Department of Economics, Oxford, No. 740, 2015.

[45] Riker, W. and P. Ordeshook *"A Theory of the Calculus of Voting"*, The American Political Science Review 62 (1), 25-42, 1968.

[46] Simon, H. *"A Behavioral Model of Rational Choice"*, Quarterly Journal of Economics 69 (1), 99-118, 1955.

[47] Tallarini, T. *"Risk-sensitive Real Business Cycles"*, Journal of Monetary Economics 45, 507-532, 2000.

[48] Varian, H. *"Non-Parametric Tests of Consumer Behavior"*, Review of Economic Studies 50 (1), 99-110, 1983.

[49] Wang, T. *"Conditional Preferences and Updating"*, Journal of Economic Theory 108, 286-321, 2003.