# REVENUE MANAGEMENT WITH CUSTOMER CHOICE AND SELLERS COMPETITION

A Thesis
Presented to
The Academic Faculty

by

Xinchang Wang

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
H. Milton Stewart School of Industrial and Systems Engineering

Georgia Institute of Technology
August 2015

# REVENUE MANAGEMENT WITH CUSTOMER CHOICE AND SELLERS COMPETITION

Approved by:

Professor Anton Kleywegt, Advisor
H. Milton Stewart School of Industrial and Systems Engineering
*Georgia Institute of Technology*

Professor Sigrún Andradóttir
H. Milton Stewart School of Industrial and Systems Engineering
*Georgia Institute of Technology*

Professor Hayriye Ayhan
H. Milton Stewart School of Industrial and Systems Engineering
*Georgia Institute of Technology*

Professor Jim Dai
School of Operations Research and Information Engineering
*Cornell University*

Professor Fabian Bastin
Computing Science and Operational Research Department
*University of Montreal*

Date Approved: May 11, 2015

## ACKNOWLEDGEMENTS

First and foremost, my deepest gratitude goes to my Ph.D. supervisor, Professor Anton Kleywegt, for his wise supervision, invaluable assistance and guidance during my studies at Georgia Tech. This thesis is a result of fives years of interesting work with Dr. Klweygt. I have been profoundly infected with his passion, enthusiasm and rigorous attitude for research, and I have benefited from his perspectives, experiences and broad knowledge of diverse fields.

I am deeply grateful to Professor Sigrún Andradóttir and Professor Hayriye Ayhan for advising me for the first two years and for their continuous encouragement during the rest of my Ph.D. study. I have learned a lot and will benefit for life from their instructions.

Many thanks go to Professor Jim Dai for his generous and kind helps and support. He is a great model, not only for being a successful professor but also for being a good man.

Special thanks go to Professor Fabian Bastin for our insightful and detailed discussions about the stochastic truest region algorithm and for sharing with me AMLET and the MobiDrive dataset.

Last but not least, I give my sincere thanks to my wife for her quiet patience and unwavering love. Thank you for always standing by me and being willing to help me whenever there is a need.

## TABLE OF CONTENTS

# LIST OF TABLES

## LIST OF FIGURES

# SUMMARY

Revenue management is concerned with managing demands of customers and has been found successful in broad areas such as airline, hotel and retailing industries. In revenue management, decisions of sellers such as designing a product portfolio or choosing prices of products are often made based on customer choice. It is thus important to understand customer choice behavior and analyze how it affects sellers' decisions, especially when customers' choice exhibits specific behavioral phenomena that deviate from axioms of rational choice (e.g., Luce's axiom of choice) and sellers compete.

My thesis is focused on revenue management problems, with particular emphasis on customer choice behavior, and it consists of three essential chapters.

In the first chapter, we build a variety of customer booking choice models for a major airline that operates in a very competitive origin-destination market, including the multinomial logit (MNL) models, nested logit (NL) models, mixed-logit (ML) models and latent logit class (LCL) models. The latter three types of models are aimed at incorporating unobserved heterogeneous customer preferences for different departure times of flights and identifying latent customer types. More interestingly, we incorporate in all our models the context effect that the attractiveness of a fare class is influenced by the other fare classes offered in the same assortment, which is not standard in the literature of discrete choice models. The estimation results show that including these factors into choice models dramatically affects price sensitivity estimates, and therefore matters.

Previously available algorithms are inefficient for estimating choice models from large sets of data (observations), especially for estimating advanced choice models that usually involve high-dimensional integrals, such as the ML-type models. In the second chapter, we present a stochastic trust region algorithm for ML-type model estimations. The algorithm embeds two sampling processes: (i) a data sampling process and (ii) a Monte Carlo sampling process. The second process is employed to compute the sample average approximation of a high-dimensional integral. The algorithm dynamically controls the sample sizes based on the magnitude of the errors incurred due to the two sampling processes. First, the algorithm controls the size of Monte Carlo samples for each observation in the dataset to minimize the total sample size subject to a constraint on the variance of the objective estimate. Second, the algorithm controls sampling from the dataset according to the magnitude of data sampling error relative to the Monte Carlo sampling error. The first-order convergence is proved based on generalized uniform law of large numbers theories for both the average log-likelihood function and its gradient. The efficiency of the algorithm is tested with real data and compared with existing algorithms.

In the third chapter, we study how a specific behavioral phenomenon, called the decoy effect, affects the decisions of sellers in product assortment competition in a duopoly. We propose a discrete choice model to capture decoy effects, and we use the model to provide a complete characterization of the Nash equilibria and their dependence on choice model parameters. For the cases in which there are multiple equilibria, we consider dynamical systems models of the sellers responding to their competitors using Cournot adjustment or fictitious play to study the evolution of the assortment competition and the stability of the equilibria. Our results show that all pure-strategy Nash equilibria can provide reliable forecasts of the outcome of the competition in the sense that they have large domains of attraction. In contrast, mixed-strategy Nash equilibria have negligible domains of attraction, except for a special case, and thus we conclude that mixed-strategy Nash equilibria do not provide reliable forecasts of the outcome of the competition. Our results also provide a simple geometric characterization of the dynamics of fictitious play for general $2 \times 2$ games that is more complete than previous characterizations.

# CHAPTER I

# DISCRETE CHOICE MODELING AND ESTIMATION

In this chapter we describe discrete choice modeling and estimation with applications to airline revenue management. We consider a revenue management problem of a major airline that operates in a fiercely competitive market involving two major hubs and having more than 30 parallel daily flights. We build a variety of booking choice models to incorporate unobserved heterogeneous customer preferences for different departure times. The way departure time preferences are modeled dramatically affects price sensitivity estimates, and therefore the modeling of heterogeneous departure time preferences matters. We also show that customer choice behavior exhibits the context effects, with much greater demand for the cheapest alternative than for the second cheapest alternative even when the price difference is small, and much greater demand for fully refundable tickets than almost fully refundable tickets.

## *1.1 The Airline Ticket Booking Choice*

Consider a number of airlines, indexed by $i \in I$, selling tickets for travel on parallel flights in a single origin-destination (O-D) market. In this section, for ease of notation, airline XX for whom we estimate customers' booking choice models is indexed by $i = 1$, and $-i$ denotes the competitors {YY, ZZ}.

We now specify the "products" that airlines offer to the market. Airline $i$ sells tickets for a set $F_i$ of flights. We consider all the flights for the O-D pair that depart on a particular day of week (say, Monday, Tuesday, Saturday etc.). Each airline maintains a list of fare classes each associated with a fixed ticket price, and the airline can sell tickets with the price associated with each fare class for each of the flights. For each airline, we refer to a flight-fare class combination as the product of the airline. Let $J_i$ be the set of products that airline $i$ offers to the market. The selling horizon is denoted with $[0, T]$, where $T$ denotes the scheduled departure time of the last flight during the time horizon. There is a set $K$ of sales channels that can be used to sell airlines' products, for example, an airline's own web site, an airline's call center, various third-party web sites, and independent travel agents. Some channels, such as an airline's own web site or call center, are used by only one airline, and some, such as third party web sites, are used by multiple airlines.

Customer booking requests arrive in each channel $k$ according to independent nonhomogeneous Poisson processes with rates $\lambda_k(t)$. For each time $t \in [0, T]$ and each sales channel $k \in K$, each airline $i$ chooses a set $A_{i,k}(t) \subset J_i$ of products to offer; $A_{i,k}(t)$ is called the assortment offered by airline $i$ through channel $k$ at time $t$. Let $K_i$ denote the set of channels used by airline $i$. If airline $i$ does not use channel $k$, i.e. $k \notin K_i$, then $A_{i,k}(t) = \varnothing$. Also, products in $A_{i,k}(t)$ cannot belong to flights that depart before time $t$. Let $\mathcal{A}_{i,k}(t) \subset 2^{J_i}$ denote the collection of assortments that airline $i$ considers at time $t$ for channel $k$. Let $A_k(t) := \cup_{i \in I} A_{i,k}(t)$ denote the assortment offered in channel $k$ at time $t$, by all the airlines in the market. Given the assortment $A_k(t) = A$, a customer who arrives at time $t$ using channel $k$ books alternative $j$ with probability $q_{j:A}(k, t)$ ($q_{j:A}(k, t) = 0$ if $j \notin A$).

Booking requests are indexed by $n \in \mathcal{N}$. The corresponding customer is referred to as customer $n$. Let $t_n$ denote the arrival time of booking request $n$, let $k_n$ denote the channel used, and let $j_n$ denote the alternative chosen (booked) by customer $n$. Thus, for a customer $n$ using channel $k_n$ at booking time $t_n$, $A_{k_n}(t_n)$ denotes the assortment offered by all airlines to customer $n$. For ease of notation, let $A_n := A_{k_n}(t_n)$ denote the assortment for customer $n$, and let $q_{j:A_n}$ denote the probability of customer $n$ choosing $j$ from $A_n$. In addition, each customer $n$ has a consideration set $C_n \subset \cup_{i \in I} J_i$ of products that the customer would consider. Thus, each customer $n$ chooses from products in the customer's choice set $S_n := A_n \cap C_n$. Note that the customer's observed choice $j_n$ must be in $S_n$. Typically, the consideration set $C_n$ and the choice set $S_n$ of customer $n$ is not observed — this is one of the challenges encountered in discrete choice modeling.

## 1.2 Literature Review

Our study falls within a vast literature on discrete choice modeling and estimation. Discrete choice models have found broad applications to predicting travelers' choices in transportation [9] and customers' choices from a set of products in revenue management [63]. The classic multinomial logit (MNL) model has been widely used due to its tractability, but it has a number of shortcomings, including (i) the independence from irrelevant alternatives (IIA) property, (ii) the assumption that each customer's choice set is known, and (iii) the assumption that all customers have the same preferences or taste coefficients. To address these shortcomings, a variety of other discrete choice models have been developed, including the nested logit (NL), mixed logit (ML), latent class logit (LCL) models, and probit models [24]. There is a significant amount of work, including [54, 55, 1], that incorporate preference heterogeneity in consumer choice models. Interested readers are referred to [14] for a detailed review of the MNL model and its many kinds of variants.

The revenue management problem discussed in our study is similar to that of [70] in the sense that we also use discrete choice models for parallel flights calibrated with airline data and we also use simulation to evaluate the performance of our policies. However, our choice models address some issues not addressed in [70]: (i) we develop models that incorporate the idea that different customers have different preferences (taste heterogeneity) for different departure times, (ii) we allow differences in price sensitivity depending on when the customer books and what channel the customer uses, and (iii) we identified and modeled context effects for the cheapest available fare classes as well as for fully refundable fare classes. [71] also considered discrete choice models for parallel flights. They studied structural properties of a Markov decision process formulation, and they compare a number of heuristics for their model.

## 1.3 The 2011 and 2012 Airline Data

There are three major airlines that we call XX, YY, and ZZ, in the market, and we consider to build and estimate customers' booking choice models for airline XX. For airline XX, we have booking data. The booking data contain the values of various factors that are important for the estimation of booking choice models discussed in Section 1.4.

We also have availability data that show snapshots, typically once per day, of the assortment being offered by each airline at that time. The assortment sometimes changes during a day, and we also use customers' booking data to identify when such changes took place, and to construct the historical assortments $A_{i,k}(t)$ for each airline $i$ in channel $k$ as a function of time $t$.

In Sections 1.4.1–1.4.4 we introduce four booking choice models for applications of general purpose. The four booking choice models differ in the way that they incorporate heterogeneous customer preferences for different departure times. Then we describe the factors, encoded attributes and our choice models in Section 1.5. The estimation results are compared and discussed in Section 1.6.

## *1.4 Various Discrete Choice Models*

Discrete choice models predict the probability of customers choosing a specific product from among an assortment of products offered at market and formulate customer choice probabilities as functions of utilities of the alternatives in the offered assortment. The utility of an alternative is further formulated as a function of the alternative's attributes that are often encoded into numeral values using the factors of the alternative such as fare price, departure time, and booking channel (discussed in detail in Section 1.5.1).

Let $x_{n,j,m}$ denote the value of attribute $m \in \{1, 2, \ldots, \bar{m}\}$ for customer $n \in \mathcal{N}$ and alternative $j \in A_n$, and let $x_{n,j} := (x_{n,j,1}, \ldots, x_{n,j,\bar{m}}) \in \mathbb{R}^{\bar{m}}$ denote the attribute vector for customer $n$ and alternative $j$. The systematic utility, $v_{n,j}$, of alternative $j$ for customer $n$ is represented in terms of the following linear function,

$$v_{n,j} := \beta^{\intercal} x_{n,j}, \tag{1.1}$$

where $\beta := (\beta_1, \ldots, \beta_{\bar{m}}) \in \mathbb{R}^{\bar{m}}$ denotes the parameter vector and $\beta_m$ denotes the coefficient or weight of attribute $m$. Let $q_{j:A_n}$ denote the probability of customer $n$ choosing product $j$ from $A_n$. We describe four customers' booking choice models that all use (1.1) either in the original form or an enhanced form to capture the heterogeneity in customer preferences.

### 1.4.1 The Multinomial Logit (MNL) Model

One of the most popular discrete choice models is the multinomial logit model. For basic properties of the MNL model, see for example [9] and [64]. The probability that customer $n$ chooses alternative $j \in A_n$ is given by

$$q_{j:A_n} = \frac{\exp(v_{n,j})}{\sum_{j' \in S_n} \exp\left(v_{n,j'}\right)} = \frac{\exp(\beta^\intercal x_{n,j})}{\sum_{j' \in S_n} \exp\left(\beta^\intercal x_{n,j'}\right)}. \tag{1.2}$$

In the context of airline demand (and many other applications), different customers consider different sets of alternatives, but the consideration sets $C_n$ and choice sets $S_n := A_n \cap C_n$ are not observed (but some data related to the consideration sets may be observed). For example, different customers consider different sets of departure times to be reasonable for their purposes. Some customers are flexible and may consider all flights in a wide time window, whereas other customers have tight schedules and want to depart as close as possible to a specific time. These time preferences are not observed.

The following modeler's selection of $S_n$ was suggested in [70]: Given that customer $n$ booked a ticket from $A$ to $B$ for a flight departing on a particular day, it is assumed that $S_n = A_n$ is the set of all flights from $A$ to $B$ on the same day. We used the same selection of $S_n$ for the MNL model results discussed in Section 1.6. However, the following intuitive argument suggests that such a selection may produce biased parameter estimates. Suppose that the price of an alternative is an important attribute of the alternative. More specifically, suppose that each customer chooses the cheapest ticket for a flight that departs in the customer's preferred time window. Thus customers are quite price sensitive, with attention restricted to a subset of alternatives. Now suppose that flights departing at different times of the day have different cheapest available fares (which is often the case). In a data set of bookings, a significant fraction of customers do not choose one of the cheapest tickets over all flights departing on the particular day (because none of the cheapest tickets were for a flight departing in the customers' time windows). If it is assumed that each customer chooses from the set of all flights on the same day, then it appears that customers are not very price sensitive, and as a result the estimated price coefficients will be biased. As shown in Section 1.6, our results were consistent with this intuition. Next we discuss a number of models that attempt to incorporate heterogeneity in customer preferences.

### 1.4.2 The Nested Logit (NL) Model

In the nested logit model, the set of alternatives is partitioned into subsets called nests, indexed by $l \in \{1, 2, \ldots, L\}$. For example, different nests contain tickets for flights departing during different time windows. Correspondingly, for each customer $n$, $A_n$ is partitioned into $L$ nests denoted with $A_{n,l}$. In the NL model, different alternatives in the same nest have positively correlated utilities. Thus, by choosing different nests to contain tickets for flights departing during different time windows, the NL model can capture heterogeneous preferences for different departure times. A restriction of the NL model is that the set of alternatives has to be partitioned, for example, the NL model does not capture a setting in which customers either prefer departure times between $t_1$ and $t_3$ or departure times between $t_2$ and $t_4$, where $t_1 < t_2 < t_3 < t_4$. For more detail of the NL model, see for example [9] and [64]. The systematic utility of customer $n$ for alternative $j \in A_{n,l}$ is given by $v_{n,j} := \beta^\intercal x_{n,j}/\alpha_l$, where $\alpha_l \in [0, 1/\alpha]$ is the parameter that represents the variation of preferences for alternatives in $A_{n,l}$, and $\alpha > 0$ is a scaling factor. Then the probability that customer $n$ chooses alternative $j \in A_{n,l}$ is given by

$$
\begin{aligned}
q_{j:A_n} &= \frac{\exp(v_{n,j})}{\sum_{j' \in A_{n,l}} \exp(v_{n,j'})} \frac{\exp\left(\alpha\alpha_l \bar{v}_{n,l}\right)}{\sum_{l'=1}^{L} \exp\left(\alpha\alpha_{l'} \bar{v}_{n,l'}\right)} \\
&= \frac{\exp(\beta^\intercal x_{n,j}/\alpha_l)}{\sum_{j' \in A_{n,l}} \exp(\beta^\intercal x_{n,j'}/\alpha_l)} \frac{\exp\left(\alpha\alpha_l \bar{v}_{n,l}\right)}{\sum_{l'=1}^{L} \exp\left(\alpha\alpha_{l'} \bar{v}_{n,l'}\right)},
\end{aligned}
$$

where

$$
\bar{v}_{n,l} := \ln\left(\sum_{j \in A_{n,l}} \exp(\beta^\intercal x_{n,j}/\alpha_l)\right), \quad \forall\, l \in \{1, \ldots, L\}.
$$

### 1.4.3 The Mixed Logit (ML) Model

Let $\theta$ denote the attribute coefficients that reflect the tastes of customers in evaluating attributes and $\zeta_{n,j}$ represent the vector of attribute values. A natural way for a choice model to capture heterogeneous tastes is to allow variation in the values of $\theta$. Let $\pi_\theta$ denote the probability distribution of a customer's parameter vector $\theta$. Thus, the systematic utility of customer $n$ for alternative $j$, given by $v_{n,j} := \theta^\intercal \zeta_{n,j}$, is random (given the vector $\zeta_{n,j}$ of attribute values) with distribution determined by $\pi_\theta$. Then the probability that customer $n$ chooses alternative $j \in A_n$ is given by

$$q_{j:A_n}(k_n, t_n) \;=\; \mathbb{E}_{\pi_\theta}\left[\frac{\exp\left(v_{n,j}\right)}{\sum_{j' \in A_n} \exp\left(v_{n,j'}\right)}\right] \;=\; \mathbb{E}_{\pi_\theta}\left[\frac{\exp\left(\theta^\intercal \zeta_{n,j}\right)}{\sum_{j' \in A_n} \exp\left(\theta^\intercal \zeta_{n,j'}\right)}\right]. \tag{1.3}$$

ML models can approximate heterogeneous consideration sets by including random coefficients $\theta_w$ for product subsets $w \subset \cup_{i \in I} J_i$, where a value of $\theta_w < -M$ for large $M$ in effect removes alternatives $j \in w$ from the customer's consideration set. Specifically, to model departure time preferences, we partition the departure times into hourly time windows indexed by $w = 1, \ldots, 14$. For each customer $n$, alternative $j \in A_n$, and time window $w$, let $y_{n,j,w} = 1$ if the flight for alternative $j$ departs in time window $w$, let $y_{n,j,w} = 0$ otherwise, and let $y_{n,j} := (y_{n,j,1}, \ldots, y_{n,j,14})$. The corresponding parameter vector $\gamma := (\gamma_1, \ldots, \gamma_{14})$ is random. We estimated a model in which the values of $\gamma$ of different customers are independent normally distributed with mean $\mu \in \mathbb{R}^{14}$ and covariance matrix $\Sigma \in \mathbb{R}^{14 \times 14}$. A large mean $\mu_w$ indicates a time window $w$ that is on average more popular, a large variance $\Sigma_{w,w}$ indicates a time window $w$ that some customers strongly like and other customers strongly dislike, and a large positive covariance $\Sigma_{w,w'}$ indicates a pair of time windows $(w, w')$ with similar preferences — some customers like both and other customers dislike both. We can represent $\gamma = \mu + \sigma\xi$, where $\xi \in \mathbb{R}^{14}$ has independent standard normal components, and $\sigma \in \mathbb{R}^{14 \times 14}$ is the lower-triangular Cholesky factor such that $\Sigma = \sigma\sigma^\intercal$. Let $\beta$ denote the deterministic parameters, that is, the values of $\beta$ are the same across the customer population, and let $x_{n,j}$ denote the corresponding vector of attribute values for customer $n$ and alternative $j$. Then $\theta = (\beta, \gamma)$, $\zeta_{n,j} = (x_{n,j}, y_{n,j})$, and the systematic utility is $v_{n,j} = \theta^\intercal \zeta_{n,j} = \beta^\intercal x_{n,j} + \gamma^\intercal y_{n,j} = \beta^\intercal x_{n,j} + \mu^\intercal y_{n,j} + \xi^\intercal \sigma^\intercal y_{n,j}$. The parameters $(\beta, \mu, \sigma)$ are estimated by solving a maximum likelihood problem. For more detail of the ML model, see for example [64].

### 1.4.4 The Latent Class Logit (LCL) Model

In the LCL model there are discrete customer classes, and different customer classes have different consideration sets and/or different values of the parameter vector $\beta$, but the class of each customer is not observed. We consider the case in which all customer classes have the same value of the parameter vector $\beta$, but different customer classes have different consideration sets. In general, the modeler enumerates a collection $\mathcal{C} \subset 2^{\cup_{i \in I} J_i}$ of sets of products that a customer may consider. Let $\pi_C$ denote the probability that a customer's consideration set is $C \in \mathcal{C}$. As before, the systematic utility of customer $n$ for alternative $j$ is given by $v_{n,j} := \beta^\intercal x_{n,j}$. Then the probability that customer $n$ chooses alternative $j \in A_n$ is given by

$$q_{j:A_n} = \mathbb{E}_{\pi_C}\left[\frac{\mathbf{1}_{[j \in C]} \exp(v_{n,j})}{\sum_{j' \in A_n \cap C} \exp(v_{n,j'})}\right] = \sum_{\{C \in \mathcal{C} \,:\, j \in C\}} \pi_C \frac{\exp(v_{n,j})}{\sum_{j' \in A_n \cap C} \exp(v_{n,j'})}.$$

The parameters $(\beta, \pi)$ are estimated by solving a maximum likelihood problem subject to the constraints that $\pi_C \geq 0$ for all $C \in \mathcal{C}$ and $\sum_{C \in \mathcal{C}} \pi_C = 1$. For example, to model departure time preferences, we construct the collection $\mathcal{C}$ of consideration sets $C$ as follows: Each $C$ contains all flight-fare class combinations with departure times within the same time window $[t_1, t_2]$, and $\mathcal{C}$ is constructed by taking all combinations of $t_1$ and $t_2$ with hourly increments such that $t_1 < t_2$. Note that $\cup_{C \in \mathcal{C}} C = \cup_{i \in I} J_i$, but, unlike the nests of the nested logit model, $\mathcal{C}$ is not a partition of $\cup_{i \in I} J_i$. For example, some customer types have narrow time windows and other customer types have wider time windows that intersect multiple narrow time windows.

## *1.5 Estimation of Airline Ticket Booking Choice Models*

### 1.5.1 Encoded Attributes using Factors

For all the choice models, attributes are constructed to encode observed factors such as fare price, departure time, and booking channel. We also consider that that affect the choice probabilities $q_{j:A_n}$ in such a way that the choice probabilities depend on the factors through a linear function of the attributes only, as follows.

Table 1 lists the alternative-specific factors (1–5) and the customer-specific factors (6–7) for which we obtained data, and that affected the booking choice probabilities.

We first discuss the use of the factors in Table 1 for booking choice models.

1. It is natural for booking choices to be affected by ticket prices – everything else being the same, the lower the price, the greater the probability that the customer chooses

Table 1: Alternative-specific and customer-specific factors used to encode attributes and estimate the discrete choice models.

| | Factor | Description |
|---|---|---|
| 1 | Ticket price | the ticket fare, e.g., $1350 |
| 2 | Departure time | the time when a flight takes off, e.g., 09:00 |
| 3 | Ticket change fee | the fee charged for changing to another flight, e.g., $75 |
| 4 | Mileage gain | the mileage credits earned by a customer if the customer buys the ticket, e.g., 1140 points |
| 5 | Carrier | the airline that sells tickets |
| 6 | Booking time | the date, hour, minute at which the booking was made, e.g., Tuesday 2011-06-07 09:20 |
| 7 | Booking channel | the channel via which a ticket is booked, e.g., airline web site, call center |

the alternative. Ticket prices used in the models were the total prices paid by the customers, including taxes and fees.

2. Customers have preferences regarding departure times. No flights on the schedule departed between 00:00 and 07:00. We partitioned the departure times from 07:00 to 21:00 into 14 hourly time windows. The time window $[21:00, 07:00)$ represents the late night flights typically around 21:30 and 22:00. The estimated MNL model captured the different popularity of different departure times by terms $\beta_w x_{n,j,w}$, where $\beta_w$ is an element of vector $\beta$ and $x_{n,j,w}$ is the corresponding attribute value of vector $x_{n,j}$ in expression (1.2) for $q_{j:A_n}$. It has that $x_{n,j,w}$ is 1 if alternative $j$ departs in time window $w$, and 0 otherwise, and $\beta_w$ represents the contribution of alternatives in time window $w$ to customers' systematic utility (relative to one of the time windows).

   As discussed before, not only are some departure times more popular than other ones, but different customers have different preferences regarding departure times. The NL model partitions departure times into three subsets or nests; nest $l = 1$ contains the alternatives with departure times between 07:00 and 11:00, nest $l = 2$ contains the alternatives with departure times between 11:00 and 17:00, and nest $l = 3$ consists of the alternatives with departure times between 17:00 and 07:00. Thus, customers are modeled as having random preferences for departure times in these three nests. Customer preferences for departure times within each nest are modeled as in the MNL model. In the ML model, customers' preferences for departure times are modeled by terms $\gamma_w y_{n,j,w}$ in (1.3) of Section 1.4.3, where $(\gamma_w, w = 1, \ldots, 14)$ is a random vector, with a multivariate normal $N(\mu, \Sigma)$ distribution. A large value of $\mu_w$ represents a departure time window that is popular (on average), and a large value of $\Sigma_{w,w'}$ means that customers tend to prefer both time windows $w$ and $w'$, or neither.

In the LCL model, consideration sets determined by departure times were constructed as follows: Each consideration set $C$ contains all tickets with departure times within the same time window $[t_1, t_2]$, where $t_1, t_2 \in \{07:00, 08:00, \ldots, 21:00\}$, $t_1 < t_2$ (with the interpretation of $t_2 = 07:00$ as the largest of the times). Customer preferences for departure times within each consideration set $C$ was modeled as in the MNL model.

3. Different fare classes have different ticket change fees. The effect of change fees on customers' preferences is captured by terms $\beta_m x_{n,j,m}$ in the expressions for $q_{j:A_n}$, where $x_{n,j,m}$ denotes the amount of the change fee for alternative $j$, and $\beta_m$ represents the contribution of a unit of change fee to customers' systematic utility.

4. Although the distance flown from O (origin) to D (destination) is the same for all alternatives, not all tickets contribute the same number of credits to customers' frequent flyer balances. Some customers' preferences are influenced by this, and this effect is captured by terms $\beta_m x_{n,j,m}$ in the expressions for $q_{j:A_n}$, where $x_{n,j,m}$ denotes the frequent flyer credit if a customer purchases alternative $j$, and $\beta_m$ represents the contribution of a unit of frequent flyer credit to customers' systematic utility.

5. The estimated models capture the different popularity of different airlines by terms $\beta_a x_{n,j,a}$ in expression (1.2) for $q_{j:A_n}$, where $x_{n,j,a}$ is 1 if alternative $j$ is sold by airline $a$, and 0 otherwise, and $\beta_a$ represents the contribution of alternatives sold by airline $i$ to customers' systematic utility (relative to one of the airlines).

6. The time at which a customer makes a booking is expected to be correlated with the customer's price sensitivity. The use of booking time for price coefficients is discussed in the later provided examples.

7. The channel that a customer uses to search for a ticket and make a booking affects the alternatives that are displayed to the customer, and thus the customer's choice set. It is assumed that if customer $n$ uses the web site or the call center of airline $i$, then the customer considers only alternatives sold by airline $i$. In addition, the booking channel is also expected to be correlated with the customer's price sensitivity and the use of booking channels for price coefficients was discussed in the later given examples.

Next we give an example. It is to be expected that choice probability $q_{j:A_n}$ depends on the price of alternative $j$. However, we also suspected that customers who book at different times and who use different booking channels have different price sensitivities. For example, we suspected that customers who book long in advance of departure time are more price sensitive than customers who book close to departure time, that customers who book during work hours are less price sensitive than customers who book outside work hours, and that customers who book using third-party web sites are more price sensitive than customers who book using the airline's call center. To capture the effect of booking time on price sensitivity, we partitioned the booking horizon as follows. First, the number of days until departure was partitioned into 3 intervals: $[0, 6]$ days before departure, $[7, 13]$ days before departure, and more than 13 days before departure. Second, the booking day-of-week is partitioned into 2 subsets: weekdays and weekends. Third, the time-of-day when customers make their booking requests was partitioned into 3 intervals: $[00:00, 09:00)$, $[09:00, 18:00)$, and $[18:00, 24:00)$. The channels that customers use to make bookings were partitioned into 5 subsets: airline web sites, other well-known web sites, other lesser-known web sites, airline call centers, and other channels including travel agents. For each of the 90 combinations of subsets of number of days until departure, booking day-of-week, booking time-of-day, and booking channel, a separate price coefficient was estimated. Thus, the three factors of ticket price, booking time, and booking channel were encoded into $3 \times 2 \times 3 \times 5 = 90$ attributes $x_{n,j,m}$, $m = 1, \ldots, 90$, where $x_{n,j,m}$ is equal to the price of alternative $j$ if customer $n$ booked in the time interval and used the booking channel represented by index $m$, and $x_{n,j,m}$ is equal to zero otherwise. The corresponding coefficient $\beta_m$ represents the estimate of price sensitivity given that a customer books in the time interval and uses the booking channel represented by index $m$. Hence, data on customer-specific factors such as booking time and booking channel allow us to study the effects of these factors on price sensitivity. The results are summarized in Section 1.6. Due to lack of good data, other studies such as [22] and [70] estimated a single price coefficient for all customers.

Table 2 shows the encoded values we assign to the three booking time factors: days to departure, booking time and booking day, and the factor of booking channel according to their original values.

Table 2: Encoded Values for Booking Times and Channels.

| Days to Dep. | # | Time-of-day | # | Day-of-week | # | Channel | # |
|---|---|---|---|---|---|---|---|
| [0, 6] | 1 | [00:00, 09:00) | 1 | weekday | 1 | others | 1 |
| [7, 13] | 2 | [09:00, 18:00) | 2 | weekend | 2 | airline website | 2 |
| [14, $\infty$) | 3 | [19:00, 24:00) | 3 | | | call center | 3 |
| | | | | | | large websites | 4 |
| | | | | | | small websites | 5 |

Next we discuss a context effect that seems to influence the booking choices of customers, and that we incorporated in all our models. We use Figure 1 to facilitate our explanations as below.

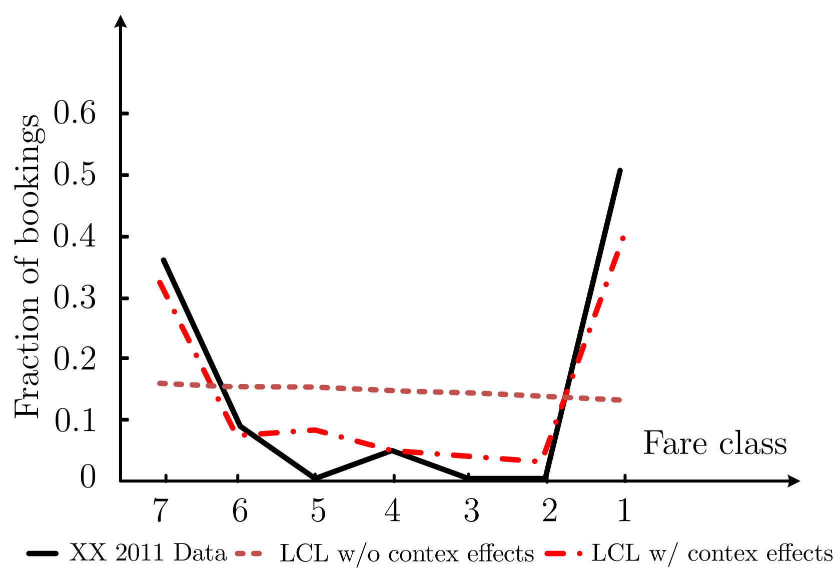



Figure 1: Comparison of fraction of XX bookings in each fare class given that Channel 1 was used and given that fare class 7 was the cheapest available fare class for the chosen flight and channel.

In Figure 1, the curve labeled “XX 2011 Data” shows the observed fraction of XX bookings in each fare class given that a particular booking channel 1 was used and given that a specific fare class 7 was the cheapest available fare class for the chosen flight and booking channel. Figure 1 shows that when fare class 7 is the cheapest available fare class for a flight, then about 36.0% of the customers who make a booking for that flight book in fare class 7, and about 50.4% of the customers who make a booking for that flight book in fare class 1. It can be seen that a large fraction of customers either book the cheapest ticket or the most expensive ticket (fare class 1), apparently because it is the only (and thus cheapest) fully refundable ticket. These effects are called context effects because the attractiveness of a fare class is influenced by the other fare classes offered (because the other fare classes offered determine whether a particular fare class is the cheapest available or the most fully refundable). We also refer to these context effects as “spikes”.

Figure 1 also shows that a LCL model that does not explicitly incorporate coefficients for spikes does not match the observed fare class distribution well. The curves obtained with MNL, NL, and ML models that do not explicitly incorporate coefficients for the context effect are similar to the curves for the LCL model shown in Figure 1, and are thus omitted.

Also, we note that the relative sizes of the spikes are different depending on what fare class is the cheapest available fare class and depending on what booking channel is used. The relative sizes of the spikes are also different for different airlines. We were able to capture the spikes by using the following attributes and corresponding coefficients: (i) For each combination of airline and booking channel there is an attribute that is equal to one if customer $n$ uses the channel and alternative $j$ belongs to the airline and it is the cheapest available fully refundable fare class on its flight. (ii) For each combination of airline, channel, and fare class, there is an attribute that is equal to one if customer $n$ uses the channel and alternative $j$ belongs to the airline and it is the cheapest available fare class on its flight. Figure 1 also shows that a LCL model that explicitly incorporates coefficients for spikes as described above matches the observed spikes quite well. The curves obtained with MNL, NL, and ML models that incorporate coefficients for spikes are similar to the curves for the MNL model shown in Figure 1, and are omitted.

The phenomenon of spikes reflects the competition among fare classes for each combination of airline and booking channel. Other researchers have considered competition among fare classes. For example, [21] emphasized that the competition among fare classes for a flight and across different flights on the same departure day may be important. [21] did not model this competition due to limited data.

In the standard MNL model, the relative choice probabilities of two alternatives do not depend on the presence of other alternatives in the choice set (the independence from irrelevant alternatives (IIA) property). Similarly, in standard NL, ML and LCL models, the relative choice probabilities of two alternatives in the same nest or class do not depend on the presence of other alternatives in that same nest or class.

The introduction of the spike coefficient for the cheapest available fare class destroys this property. As a result, the choice models with the spike coefficients for the cheapest available fare class are not standard and the estimation problems may lose some nice properties that have been established for the standard versions of these models. For the estimation problems, we can rectify this potential shortcoming of models with spike coefficients by extending the set of alternatives in the following way. For each product (i.e., a flight-fare class combination) $j$, add a spike counterpart, say $j'$, that is, for each product, there are two copies: one copy without a spike coefficient, and one copy with a spike coefficient. If a fare class is the cheapest in a customer's choice set, then the choice set contains the spike counterpart, but not the non-spike counterpart, and if a fare class is not the cheapest in a customer's choice set, then the choice set contains the non-spike counterpart, but not the spike counterpart. With such product representation, the relative choice probabilities of two (extended) alternatives do not depend on the presence of other alternatives (for MNL) or the presence of other alternatives in the same nest or class (for NL, ML, and LCL). Thus, for estimation purposes, the choice models with spike coefficients retain the nice properties of standard models.

### 1.5.2 Description of Estimated Choice Models

In this section we describe in detail various choice models that were estimated from the dataset of 2011. There are 90 price sensitivity parameters, 11 parameters for each of combinations (i), 102 parameters for each of combinations (ii), two parameters for carriers, one parameter for mileage gain and one for cancel fee, which gives 207 common parameters to all the four choice models. Combinations (i) and (ii) were introduced at the end previous section. For all the following three choice models, we take departure times in time window [21:00, 07:00) as the base case.

For the MNL model, there are 14 parameters to be estimated for attributes of departure times, which gives 221 parameters to be estimated.

For the NL model, we partition all the alternatives in $\cup A_n$, where $n \in \mathcal{N}$, into $L = 3$ nests. Nest $l = 1$ contains the alternatives with departure times between 7:00am and 10:00am, nest $l = 2$ contains the alternatives with departure times between 11:00am and 17:00pm, and nest $l = 3$ consists of the alternatives with departure times between 17:00pm and 7:00am. We set the scaling factor $\alpha = 10^{-4}$ and need to estimate dissimilarity factors $\alpha_l$, $l = 1, 2, 3$, 207 common parameters and 14 parameters for departure times.

For the ML model, to capture the variation of customer preferences in evaluating departure times, we consider a random parameter vector $\gamma_n := (\gamma_{n,1}, \gamma_{n,2}, \ldots, \gamma_{n,14})$ for departure times for each customer $n \in \mathcal{N}$, where $\{\gamma_n\}_{n \in \mathcal{N}}$ is a sequence of i.i.d. Gaussian vectors with mean vector $\mu \in \mathbb{R}^{14}$ and covariance matrix $\Sigma \in \mathbb{R}^{14\times 14}$. We can represent $\gamma_n = \mu + \sigma\xi_n$, where $\xi_n \in \mathbb{R}^{14}$ is a standard Gaussian vector and $\sigma \in \mathbb{R}^{14\times 14}$ is the lower-triangular C-holesky factor such that $\Sigma = \sigma\sigma^\intercal$. Let $\beta \in \mathbb{R}^{207}$ denote the 207 common parameters that are assumed to be deterministic and the same across the customer population $\mathcal{N}$. Let $x_{n,j} \in \mathbb{R}^{207}$ denote the vector of attribute values except departure times and $y_{n,j} \in \mathbb{R}^{14}$ denote the vector of attribute values for departure times for alternative $j \in \cup A_n$. The systematic utility is written as $v_{n,j} = \beta^\intercal x_{n,j} + \gamma_n^\intercal y_{n,j} = \beta^\intercal x_{n,j} + \mu^\intercal y_{n,j} + \xi_n^\intercal \sigma^\intercal y_{n,j}$. We need to estimate $(\beta, \mu, \sigma)$ by maximizing the simulated log-likelihood function

$$\max_{\beta,\mu,\sigma} \frac{1}{|\mathcal{N}|} \sum_{n\in\mathcal{N}} \ln \left( \frac{1}{|I_n|} \sum_{i=1}^{|I_n|} \frac{\exp\left(\beta^\intercal x_{n,j_n} + \mu^\intercal y_{n,j_n} + (\xi_n^i)^\intercal \sigma^\intercal y_{n,j_n}\right)}{\sum_{j\in\cup A_n} \exp\left(\beta^\intercal x_{n,j} + \mu^\intercal y_{n,j} + (\xi_n^i)^\intercal \sigma^\intercal y_{n,j}\right)} \right),$$

where $|I_n|$ is the Monte Carlo integration sample size and $\{\xi_n^i\}_{i=1}^{|I_n|}$ is a sequence of i.i.d. standard Gaussian variates and independent for each $n$.

For the LCL model, define $C[t_1, t_2] := \{j \in \cup_{i=1}^{3} P_i : j \text{ has departure time in } [t_1, t_2]\}$ as the set of tickets sold by airlines XX, YY and ZZ with departure times in time window $[t_1, t_2]$, where $t_1, t_2 \in \{07:00, 08:00, \ldots, 21:00\}$. Let $\mathcal{C} := \{C[t_1, t_2] : t_1, t_2 \in \{07:00, 08:00, \ldots, 21:00\}\}$ denote the collection of consideration sets (with the interpretation of $t_2 = 07:00$ as the largest of the times). For the LCL model, it assumed that customers purchasing via different channels have the same collection $\mathcal{C}$ of consideration sets and have the same type distribution $\pi_{(\cdot)}$ defined on the support $\mathcal{C}$. We need to estimate parameter vector $(\beta, (\pi_C, C \in \mathcal{C}))$ by solving the following optimization problem,

$$\begin{aligned} \max \quad & \frac{1}{|\mathcal{N}|} \sum_{n \in \mathcal{N}} \ln \left( \sum_{C \in \mathcal{C}: j_n \in C} \frac{\pi_C \mathbf{1}_{[\cup A_n \cap C \neq \varnothing]}}{\sum_{C' \in \mathcal{C}} \pi_{C'} \mathbf{1}_{[\cup A_n \cap C' \neq \varnothing]}} \frac{\exp(\beta^\intercal x_{n,j_n})}{\sum_{j \in \cup A_n \cap C} \exp(\beta^\intercal x_{n,j})} \right) \\ \text{s.t.} \quad & \sum_{C \in \mathcal{C}} \pi_C = 1, \\ & \pi_C \in [0,1], \quad \forall C \in \mathcal{C}. \end{aligned}$$

where $\beta \in \mathbb{R}^{221}$ consists of the 207 common parameters and 14 parameters for departure times, $x_{n,j} \in \mathbb{R}^{221}$ is the vector of encoded attribute values for $j \in \cup A_n$, where $n \in \mathcal{N}$, and the choice probability inside the logarithm is calculated conditioning on $\cup A_n \cap C_n \neq \varnothing$.

## *1.6 Estimation Results*

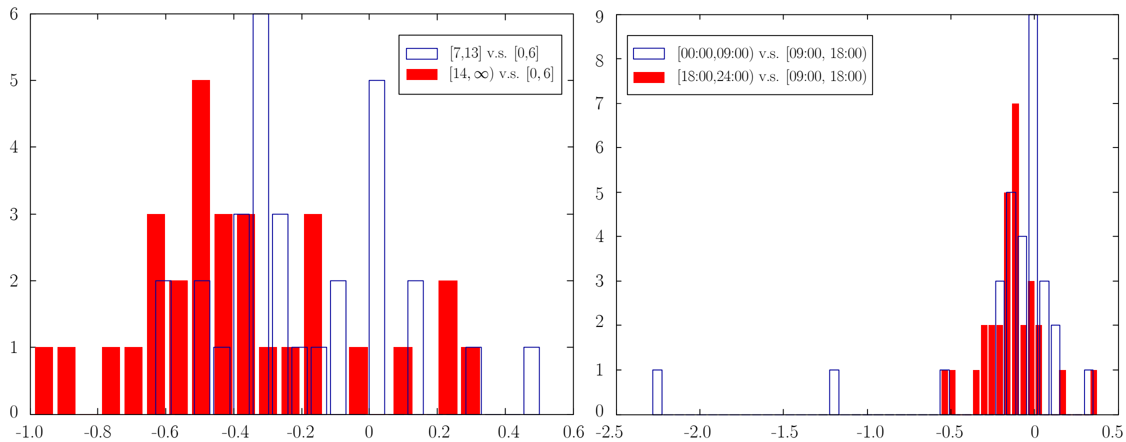
In this section we describe the estimation results for the four choice models with context effects, including the price coefficients that reflect customers' price sensitivity as well as the parameters for departure times that reflect the popularity of different departure time windows. Each of the four models is estimated with the 2011 data, 2012 data, 1-6/2012 (January-June, 2012) data and 7-12/2012 (July-December, 2012) data. Thus, there are 16 choice models in total.

### 1.6.1 Price Sensitivity

All the price coefficients are negative for all four choice models, consistent with the intuition that the more expensive the ticket is, everything else being the same, the less likely it is that a customer will choose it. Of particular interest is the behavior of the price coefficients as the values of various factors are varied. Recall that for each combination of number of days from booking until departure ($[0, 6]$, $[7, 13]$, or $[14, \infty)$ days), booking day-of-week (weekday or weekend), booking time-of-day ($[00:00, 09:00)$, $[09:00, 18:00)$, or $[18:00, 24:00)$), and booking channel (airline websites, other well-known websites, other lesser-known websites, airline call centers, or other channels including travel agents) there is a price coefficient. It is part of revenue management folk wisdom that customers who book longer in advance of departure times tend to be more price sensitive. To illustrate the effect of number of days until departure on the price coefficients, we compare the values of the price coefficients for $[7, 13]$ or $[14, \infty)$ days before departure with the values of the price coefficients for $[0, 6]$ days before departure, for each combination of the other factors (booking day-of-week, booking time-of-day, and booking channel). For example, the price coefficient of ([7,13], [00:00,09:00), weekday, others) is -7.2397 and the price coefficient of ([0,6], [00:00,09:00), weekday, others) is -4.8802. Then, the relative price coefficient of $[7, 13]$ relative to $[0, 6]$ for ([00:00,09:00), weekday, others) is $[-7.2397-(-4.8802)]/7.2397 = -0.3259$. Figure 2a shows a histogram of the MNL relative price coefficients for $[7, 13]$ relative to $[0, 6]$, and for $[14, \infty)$ relative to $[0, 6]$, for all combinations of booking day-of-the-week, booking time-of-day, and booking channel. As Figure 2a shows, for most combinations of booking day-of-the-week, booking time-of-day, and booking channel, the price coefficients for $[7, 13]$ are smaller (more negative) than the price coefficients for $[0, 6]$, and the price coefficients for $[14, \infty)$ are even more negative relative to the price coefficients for $[0, 6]$. Thus most price coefficients are consistent with revenue management folk wisdom.

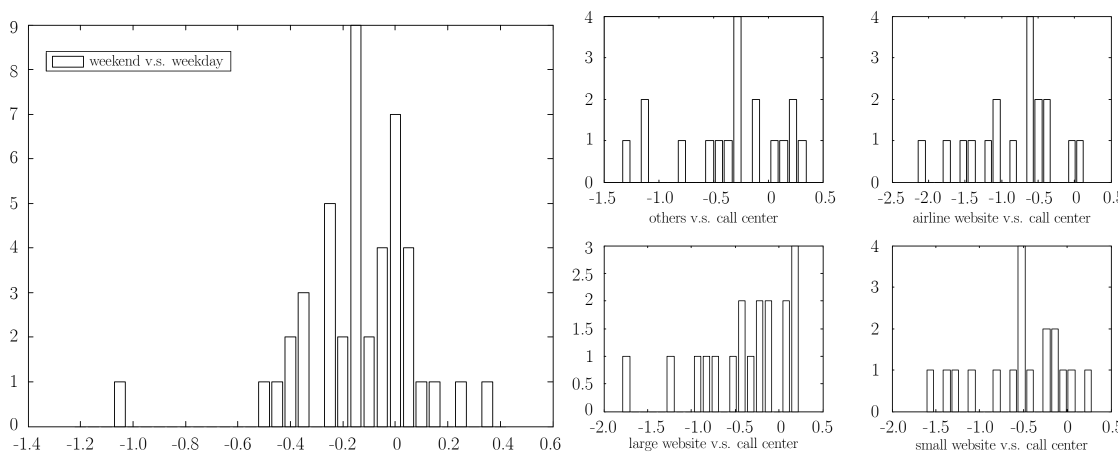
Similarly, Figure 2b shows a histogram of the MNL relative price coefficients for booking time-of-day $[00:00, 09:00)$ relative to $[09:00, 18:00)$, and for $[18:00, 24:00)$ relative to $[09:00, 18:00)$, for all combinations of booking days until departure, booking day-of-the-week, and booking channel. As Figure 2b shows, customers who book outside work hours tend to be more price sensitive than customers who book during work hours.

Figure 3a shows a histogram of the MNL relative price coefficients for weekend bookings




(a) Booking days until departure (b) Booking time-of-day

Figure 2: Histograms of relative price coefficients for booking days until departure and booking time-of-day.

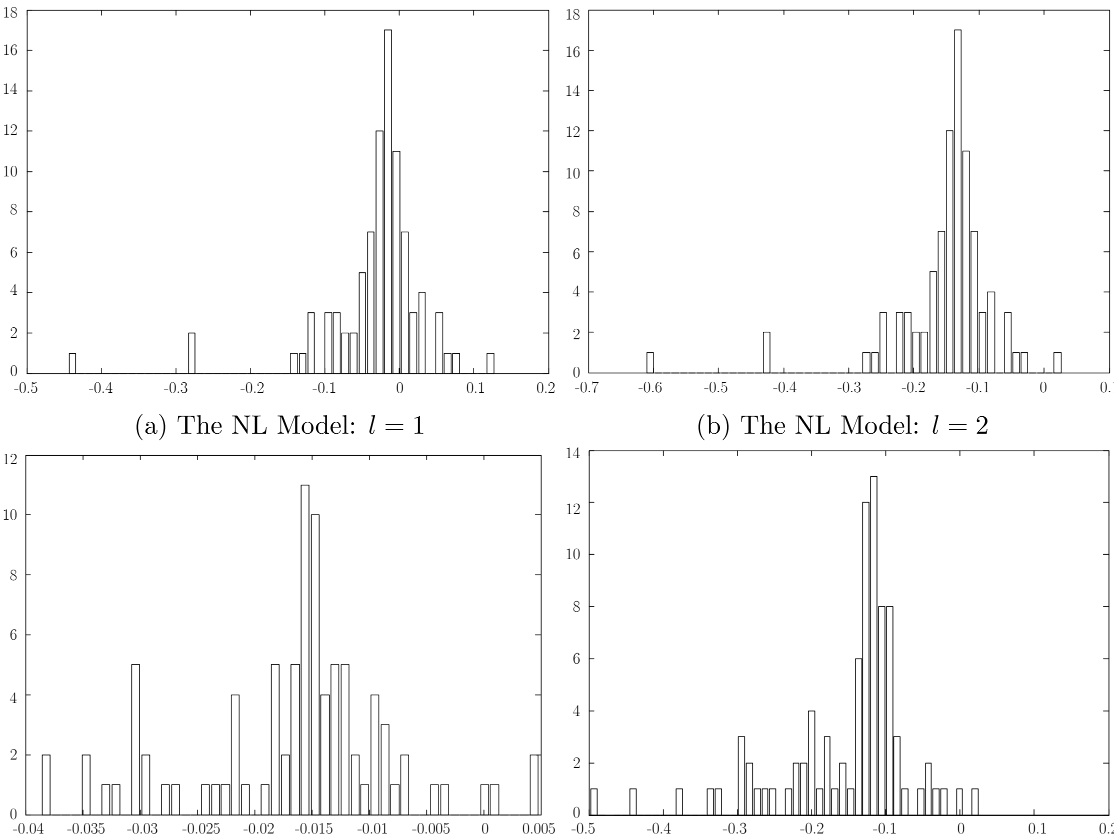



(a) Booking day-of-week (b) Booking channel

Figure 3: Histograms of relative price coefficients for booking day-of-week and channel.

relative to weekday bookings, for all combinations of booking days until departure, booking time-of-day, and booking channel. As Figure 3a shows, customers who book during weekends tend to be more price sensitive than customers who book during weekdays.

Figure 3b shows histograms of the MNL relative price coefficients for airline websites, large websites, small websites, and other channels, relative to airline call centers, for all combinations of booking days until departure, booking day-of-week, and booking time-of-day. As Figure 3b shows, customers who book through airline websites, large websites,




(a) The NL Model: $l = 1$

(b) The NL Model: $l = 2$

(c) The ML Model

(d) The LCL Model

Figure 4: Histograms of price coefficients for the NL, ML, and LCL models relative to the price coefficients of the MNL model.

small websites, and other channels, tend to be more price sensitive than customers who book through airline call centers.

Figure 4 shows histograms of the relative price coefficients for the NL, ML, and LCL models, relative to the MNL model, for all combinations of booking days until departure, booking day-of-week, booking time-of-day, and booking channel. As Figure 4 shows, almost all the price coefficients for the NL, ML, and LCL models are more negative than the corresponding price coefficients for the MNL model, which suggests that the MNL model underestimates the price sensitivity of customers because of its assumption that all customers consider all flights for the origin-destination pair and departure date.

Table 3: Choice probabilities for different departure time windows according to the MNL, NL, ML, and LCL models.

| Index | Time window | MNL | NL | ML | LCL |
|---|---|---|---|---|---|
| 1 | [07:00, 08:00) | 0.09387 | 0.09224 | 0.08903 | 0.09675 |
| 2 | [08:00, 09:00) | 0.07107 | 0.07693 | 0.06811 | 0.08046 |
| 3 | [09:00, 10:00) | 0.08454 | 0.09352 | 0.08141 | 0.09089 |
| 4 | [10:00, 11:00) | 0.08062 | 0.09185 | 0.07783 | 0.08820 |
| 5 | [11:00, 12:00) | 0.06409 | 0.09689 | 0.06630 | 0.07384 |
| 6 | [12:00, 13:00) | 0.05851 | 0.08519 | 0.06086 | 0.06602 |
| 7 | [13:00, 14:00) | 0.04853 | 0.07015 | 0.05041 | 0.04774 |
| 8 | [14:00, 15:00) | 0.06919 | 0.10720 | 0.07294 | 0.07112 |
| 9 | [15:00, 16:00) | 0.07958 | 0.12569 | 0.08279 | 0.07511 |
| 10 | [16:00, 17:00) | 0.07838 | 0.12127 | 0.08239 | 0.07038 |
| 11 | [17:00, 18:00) | 0.07062 | 0.00982 | 0.07357 | 0.05964 |
| 12 | [18:00, 19:00) | 0.05552 | 0.00853 | 0.05337 | 0.04727 |
| 13 | [19:00, 20:00) | 0.06705 | 0.00946 | 0.06789 | 0.06061 |
| 14 | [20:00, 21:00) | 0.04709 | 0.00688 | 0.04439 | 0.04232 |
| 15 | [21:00, 07:00) | 0.03136 | 0.00438 | 0.02870 | 0.02966 |

### 1.6.2 Departure Time Popularity

Table 3 shows the probability that a customer chooses a departure time if all other attribute values (such as price) is the same for all departure times, according to each of the booking choice models. As Table 3 shows, the estimation results indicate that the flights that depart in the morning before 11:00 and in the afternoon between 15:00 and 18:00 are more popular, and flights that depart in the middle of the day and in the late evening are less popular. Although the choice models incorporate departure time preferences in different ways, the resulting departure time choice probabilities are quite similar (but, as pointed out before, these differences have a dramatic impact on price coefficient estimates).

Figure 5 shows a decreasing trend between correlation coefficients of departure time windows and the distances between indices of departure time windows for the estimated ML model. As the figure indicates, as two flights depart with a bigger time gap, the less correlated the two departure times are. In other words, customers who choose a particular flight would prefer the flights with departure times closer to the chosen flight to those with departure times more separated from the chosen one.

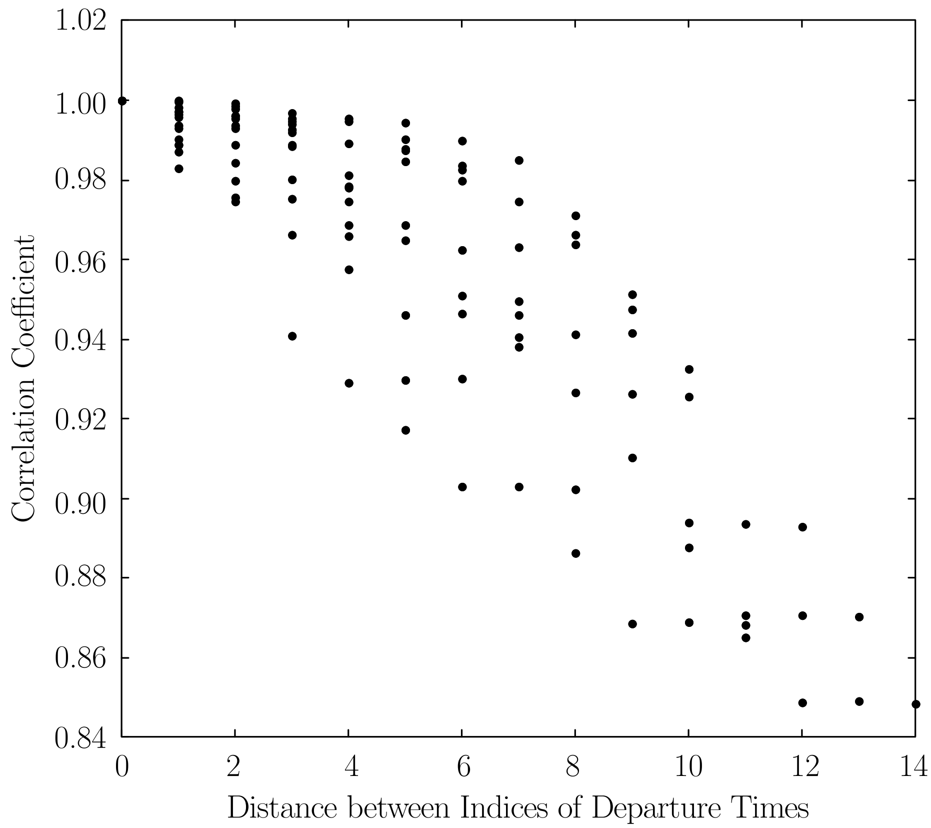



Figure 5: Correlation coefficient v.s. distance between indices of departure times for the ML model.

## *1.7 Statistical Tests*

### 1.7.1 Likelihood Ratio Tests

The likelihood ratio tests are commonly used to test the estimated models. We give two example hypotheses and the corresponding likelihood ratio test statistics. Let $\beta_p$ denote the vector of 90 price coefficients, and let $\ell^*$ denote the optimal log likelihood value for each of the 16 choice models. First consider the simple null hypothesis $H_0$ that $\beta_p = 0$. Let $\ell^*_{[\beta_p=0]}$ denote the optimal log likelihood value under $H_0$. Table 4 shows the likelihood ratio test statistic $-2(\ell^*_{[\beta_p=0]} - \ell^*)$ for each of the 16 choice models. Under $H_0$, $-2(\ell^*_{[\beta_p=0]} - \ell^*)$ is $\chi^2$ distributed with 90 degrees of freedom. Let $\chi^2(\alpha, \nu)$ denote the $\alpha$-quantile of the $\chi^2$ distribution with $\nu$ degrees of freedom. Note that $\chi^2(0.99, 90) = 61.754 \approx 62$ is the critical value of the $\chi^2$ distribution with 90 degrees of freedom at a significance level of 99%. As can be seen in Table 4, $H_0$ can be rejected for each of the choice models.

Table 4: Likelihood ratio test results: All price coefficients are zero.

| Choice model | $-\ell^*$ | $-\ell^*_{[\beta_p=0]}$ | $-2(\ell^*_{[\beta_p=0]} - \ell^*)$ | $\chi^2(0.99, 90)$ | Reject $H_0$? |
|---|---|---|---|---|---|
| MNL 2011 Data | 2,368,900 | 2,415,530 | 93,260 | 62 | Yes |
| MNL 2012 Data | 2,842,400 | 2,950,690 | 216,580 | 62 | Yes |
| MNL 1-6/2012 Data | 1,278,900 | 1,324,900 | 92,000 | 62 | Yes |
| MNL 7-12/2012 Data | 1,555,100 | 1,616,470 | 122,740 | 62 | Yes |
| NL 2011 Data | 2,353,500 | 2,404,720 | 102,440 | 62 | Yes |
| NL 2012 Data | 2,817,271 | 2,940,290 | 246,037 | 62 | Yes |
| NL 1-6/2012 Data | 1,265,800 | 1,318,040 | 104,480 | 62 | Yes |
| NL 7-12/2012 Data | 1,543,427 | 1,612,710 | 138,565 | 62 | Yes |
| ML 2011 Data | 2,368,400 | 2,415,221 | 93,641 | 62 | Yes |
| ML 2012 Data | 2,842,000 | 2,934,024 | 184,048 | 62 | Yes |
| ML 1-6/2012 Data | 1,270,706 | 1,286,439 | 31,467 | 62 | Yes |
| ML 7-12/2012 Data | 1,554,760 | 1,604,251 | 98,982 | 62 | Yes |
| LCL 2011 Data | 2,305,800 | 2,358,150 | 104,700 | 62 | Yes |
| LCL 2012 Data | 2,771,300 | 2,842,980 | 143,360 | 62 | Yes |
| LCL 1-6/2012 Data | 1,240,800 | 1,288,500 | 95,400 | 62 | Yes |
| LCL 7-12/2012 Data | 1,522,500 | 1,584,300 | 123,600 | 62 | Yes |

Next consider the null hypothesis $H_0$ that all 90 price coefficients are equal (but not necessarily equal to 0). Let $\ell^*_{[\beta_p==]}$ denote the optimal log likelihood value under $H_0$. Table 5 shows the likelihood ratio test statistic $-2(\ell^*_{[\beta_p==]} - \ell^*)$ for each of the 16 choice models. Under $H_0$, $-2(\ell^*_{[\beta_p==]} - \ell^*)$ is $\chi^2$ distributed with 89 degrees of freedom. Note that $\chi^2(0.99, 89) = 60.928 \approx 61$ is the critical value of the $\chi^2$ distribution with 89 degrees of freedom at a significance level of 99%. As can be seen in Table 5, $H_0$ can be rejected for each of the choice models.

### 1.7.2 The Significance of Price Differences between Choice Models

In Section 1.6.1, we point out that most of the price coefficients of the NL, ML, and LCL models are smaller (more negative) than the corresponding price coefficients of the MNL models, consistent with the intuition that the MNL model will tend to underestimate customers' price sensitivity. One may wonder whether these differences in price coefficients are statistically significant. We use the following approach. We have estimated 16 choice models. For a model $i \in \{1, 2, \ldots, 16\}$, let $\theta^i$ denote any value of the parameter vector, let $\theta^{i*}$ denote the (population) optimal parameter vector, let $\widehat{\theta}^i$ denote the (random) parameter vector estimated with a finite data set, and let $\mathcal{L}^i$ denote the (population expected)

Table 5: Likelihood ratio test results: All price coefficients are equal.

| Choice model | $-\ell^*$ | $-\ell^*_{[\beta_p==]}$ | $-2(\ell^*_{[\beta_p==]} - \ell^*)$ | $\chi^2(0.99, 89)$ | Reject $H_0$? |
|---|---|---|---|---|---|
| MNL 2011 Data | 2,368,900 | 2,386,200 | 34,600 | 61 | Yes |
| MNL 2012 Data | 2,842,400 | 2,873,210 | 61,620 | 61 | Yes |
| MNL 1-6/2012 Data | 1,278,900 | 1,291,080 | 24,360 | 61 | Yes |
| MNL 7-12/2012 Data | 1,555,100 | 1,572,820 | 35,440 | 61 | Yes |
| NL 2011 Data | 2,353,500 | 2,372,700 | 38,400 | 61 | Yes |
| NL 2012 Data | 2,817,271 | 2,849,500 | 64,457 | 61 | Yes |
| NL 1-6/2012 Data | 1,265,800 | 1,279,410 | 27,220 | 61 | Yes |
| NL 7-12/2012 Data | 1,543,427 | 1,560,940 | 35,025 | 61 | Yes |
| ML 2011 Data | 2,368,400 | 2,385,900 | 35,000 | 61 | Yes |
| ML 2012 Data | 2,842,000 | 2,873,000 | 62,000 | 61 | Yes |
| ML 1-6/2012 Data | 1,270,706 | 1,290,400 | 39,389 | 61 | Yes |
| ML 7-12/2012 Data | 1,554,760 | 1,571,430 | 33,340 | 61 | Yes |
| LCL 2011 Data | 2,305,800 | 2,324,600 | 37,600 | 61 | Yes |
| LCL 2012 Data | 2,771,300 | 2,801,100 | 59,600 | 61 | Yes |
| LCL 1-6/2012 Data | 1,240,800 | 1,252,700 | 23,800 | 61 | Yes |
| LCL 7-12/2012 Data | 1,522,500 | 1,539,300 | 33,600 | 61 | Yes |

log-likelihood function. Consider the following second-order Taylor expansion of $\mathcal{L}^i$:

$$\begin{aligned}
\mathcal{L}^i(\theta^i) &\approx \mathcal{L}^i(\theta^{i*}) + \nabla \mathcal{L}^i(\theta^{i*})^\intercal \left(\theta^i - \theta^{i*}\right) + \frac{1}{2}\left(\theta^i - \theta^{i*}\right)^\intercal \nabla^2 \mathcal{L}^i(\theta^{i*}) \left(\theta^i - \theta^{i*}\right) \\
&= \mathcal{L}^i(\theta^{i*}) + \frac{1}{2}\left(\theta^i - \theta^{i*}\right)^\intercal \nabla^2 \mathcal{L}^i(\theta^{i*}) \left(\theta^i - \theta^{i*}\right) \\
\Rightarrow \quad \nabla \mathcal{L}^i(\theta^i) &\approx \nabla^2 \mathcal{L}^i(\theta^{i*}) \left(\theta^i - \theta^{i*}\right) \\
\Rightarrow \quad \theta^i &\approx \theta^{i*} + \left[\nabla^2 \mathcal{L}^i(\theta^{i*})\right]^{-1} \nabla \mathcal{L}^i(\theta^i)
\end{aligned}$$

The first equality followed from $\nabla \mathcal{L}^i(\theta^{i*}) = 0$. Specifically,

$$\widehat{\theta^i} \approx \theta^{i*} + \left[\nabla^2 \mathcal{L}^i(\theta^{i*})\right]^{-1} \nabla \mathcal{L}^i(\widehat{\theta^i})$$

To simplify writing, let $M^i := \left[\nabla^2 \mathcal{L}^i(\theta^{i*})\right]^{-1}$ and let $\widehat{Z}^i := \nabla \mathcal{L}^i(\widehat{\theta^i})$, so that $\widehat{\theta^i} \approx \theta^{i*} + M^i \widehat{Z}^i$. Note that $M^i$ is deterministic and $\widehat{Z}^i$ is random. Next, for any two models, say $i = 1, 2$, let

$$\widehat{\theta} := \begin{bmatrix} \widehat{\theta}^1 \\ \widehat{\theta}^2 \end{bmatrix}, \quad \theta^* := \begin{bmatrix} \theta^{1*} \\ \theta^{2*} \end{bmatrix}, \quad M := \begin{bmatrix} M^1 & 0 \\ 0 & M^2 \end{bmatrix}, \quad \widehat{Z} := \begin{bmatrix} \widehat{Z}^1 \\ \widehat{Z}^2 \end{bmatrix}$$

Then

$$\widehat{\theta} \approx \theta^* + M\widehat{Z}$$

Thus,

$$\operatorname{Cov}(\widehat{\theta}) \approx \operatorname{Cov}(M\widehat{Z})$$

$$
\begin{aligned}
&= \mathbb{E}[M\widehat{Z}\widehat{Z}^{\intercal}M^{\intercal}] - \mathbb{E}[M\widehat{Z}]\mathbb{E}[\widehat{Z}^{\intercal}M^{\intercal}] \\
&= M\mathbb{E}[\widehat{Z}\widehat{Z}^{\intercal}]M - M\mathbb{E}[\widehat{Z}]\mathbb{E}[\widehat{Z}^{\intercal}]M \\
&= M\text{Cov}(\widehat{Z})M.
\end{aligned}
$$

Note that, for any two parameters $\theta_k^1$ and $\theta_l^2$, the variance of the difference $\widehat{\theta}_l^2 - \widehat{\theta}_k^1$ can be calculated from the entries of $\text{Cov}(\widehat{\theta})$ as follows:

$$
\text{Var}(\widehat{\theta}_l^2 - \widehat{\theta}_k^1) = \text{Var}(\widehat{\theta}_l^2) + \text{Var}(\widehat{\theta}_k^1) - 2\text{Cov}(\widehat{\theta}_k^1, \widehat{\theta}_l^2)
$$

In our calculations, $M$ and

$$
\text{Cov}(\widehat{Z}) = \begin{bmatrix} \text{Cov}(\widehat{Z}^1) & \text{Cov}(\widehat{Z}^1, \widehat{Z}^2) \\ \text{Cov}(\widehat{Z}^2, \widehat{Z}^1) & \text{Cov}(\widehat{Z}^2) \end{bmatrix}
$$

are calculated as explained next. Let $N$ denote the number of observations in the data set, let $n$ be the observation index. Let $\widehat{\mathcal{L}}^i$ denote the finite sample average log-likelihood function. Note that for any model, $\widehat{\mathcal{L}}^i$ can be written in the following form:

$$
\widehat{\mathcal{L}}^i(\theta^i) = \frac{1}{N}\sum_{n=1}^{N} \widehat{\mathcal{L}}_n^i(\theta^i)
$$

for an appropriate log-likelihood function $\widehat{\mathcal{L}}_n^i$ for observation $n$ that depends on the model. Thus,

$$
\begin{aligned}
\nabla\widehat{\mathcal{L}}^i(\theta^i) &= \frac{1}{N}\sum_{n=1}^{N} \nabla\widehat{\mathcal{L}}_n^i(\theta^i) \\
\nabla^2\widehat{\mathcal{L}}^i(\theta^i) &= \frac{1}{N}\sum_{n=1}^{N} \nabla^2\widehat{\mathcal{L}}_n^i(\theta^i)
\end{aligned}
$$

To simplify writing, let

$$
\widehat{Z}_n^i := \nabla\widehat{\mathcal{L}}_n^i(\widehat{\theta}^i)
$$

and note that

$$
\frac{1}{N}\sum_{n=1}^{N} \widehat{Z}_n^i = \frac{1}{N}\sum_{n=1}^{N} \nabla\widehat{\mathcal{L}}_n^i(\widehat{\theta}^i) = \nabla\widehat{\mathcal{L}}^i(\widehat{\theta}^i) = 0
$$

Then,

$$
\text{Cov}(\widehat{Z}^1, \widehat{Z}^2) = \text{Cov}(\nabla\mathcal{L}^1(\widehat{\theta}^1), \nabla\mathcal{L}^2(\widehat{\theta}^2)) \approx \text{Cov}(\nabla\widehat{\mathcal{L}}^1(\widehat{\theta}^1), \nabla\widehat{\mathcal{L}}^2(\widehat{\theta}^2))
$$

$$
\begin{aligned}
&= \operatorname{Cov}\left(\frac{1}{N}\sum_{n=1}^{N}\widehat{Z}_n^1, \frac{1}{N}\sum_{n=1}^{N}\widehat{Z}_n^2\right) \\
&= \frac{1}{N^2}\sum_{n=1}^{N}\operatorname{Cov}(\widehat{Z}_n^1, \widehat{Z}_n^2) \\
&\approx \frac{1}{N(N-1)}\sum_{n=1}^{N}\left(\widehat{Z}_n^1 - \frac{1}{N}\sum_{n=1}^{N}\widehat{Z}_n^1\right)\left(\widehat{Z}_n^2 - \frac{1}{N}\sum_{n=1}^{N}\widehat{Z}_n^2\right)^\intercal \\
&= \frac{1}{N(N-1)}\sum_{n=1}^{N}\widehat{Z}_n^1(\widehat{Z}_n^2)^\intercal.
\end{aligned}
$$

Similarly,

$$
\operatorname{Cov}(\widehat{Z}^i) \;=\; \operatorname{Cov}(\nabla\mathcal{L}^i(\widehat{\theta}^i)) \;\approx\; \frac{1}{N(N-1)}\sum_{n=1}^{N}\widehat{Z}_n^i(\widehat{Z}_n^i)^\intercal.
$$

Also, $M^i := \left[\nabla^2\mathcal{L}^i(\theta^{i*})\right]^{-1}$ is approximated by $\left[\nabla^2\widehat{\mathcal{L}}^i(\widehat{\theta}^i)\right]^{-1}$.

The expressions of the gradient and Hessian of each of the four choice models are given in Appendix B. Let $\beta_p^{\mathrm{MNL}}, \beta_p^{\mathrm{NL}} \in \mathbb{R}^{90}$ denote the estimated price coefficients for the MNL and NL models respectively, and let $t_m^{\mathrm{NL,MNL}}$ denote the estimated $t$-statistic of the price coefficient difference $\left[(\beta_p^{\mathrm{MNL}})_m - (\beta_p^{\mathrm{NL}})_m\right]$ for each $m \in \{1, 2, \ldots, 90\}$. We say that $(\beta_p^{\mathrm{NL}})_m$ is statistically significantly less than $(\beta_p^{\mathrm{MNL}})_m$ at the 95% confidence level if

$$
t_m^{\mathrm{NL,MNL}} \;:=\; \frac{(\beta_p^{\mathrm{MNL}})_m - (\beta_p^{\mathrm{NL}})_m}{\sqrt{\operatorname{Var}\left[(\beta_p^{\mathrm{MNL}})_m - (\beta_p^{\mathrm{NL}})_m\right]}} \;>\; 1.645
$$

The same test is also applied to the price coefficients of the ML and LCL models versus the MNL model. Table 6 shows the number of price coefficients of the NL, ML, and LCL models that are significantly less than the corresponding price coefficients of the MNL model.

Table 6: The number of price coefficients out of 90 price coefficients of the NL, ML, and LCL models that are statistically significantly less than the corresponding price coefficients of the MNL model at the 95% confidence level.

| Choice model | 2011 data | 2012 data | 1-6/2012 data | 7-12/2012 data |
|---|---|---|---|---|
| NL $l = 1$ | 89 | 89 | 89 | 90 |
| NL $l = 2$ | 89 | 89 | 89 | 90 |
| NL $l = 3$ | 89 | 89 | 89 | 90 |
| ML | 25 | 69 | 44 | 35 |
| LCL | 72 | 72 | 65 | 75 |

# CHAPTER II

# A STOCHASTIC TRUST REGION ALGORITHM FOR ESTIMATING MIXED LOGIT TYPE MODELS

Motivated by mixed logit estimation problems, we consider stochastic optimization problems of the form $\min_\theta \sum_n r(\mathbb{E}_\xi[F_n(\theta, \xi)])$, where $\theta$ is the decision variable, and $\xi$ is a random variable with chosen distribution. In the case of mixed logit estimation, the sum involves observations in a data set, $r$ is a negative logarithm, and $\theta$ includes parameters of the systematic utility as well as parameters of the probability distribution. In many applications, the dimension of $\xi$ is sufficiently high to exclude calculation of the expectation using quadrature methods. Thus we propose an algorithm that embeds a sample average approximation of the expectation. The algorithm controls the sample size for each observation $n$ in the data set to minimize the total sample size subject to a constraint on the variance of the objective estimate. In addition, the algorithm controls sampling from the data set. We provide sufficient conditions for convergence of a trust region based algorithm.

## 2.1 Introduction

Consider a set $\mathcal{N}$ of customers, where $|\mathcal{N}| < +\infty$. Customer $n \in \mathcal{N}$ makes a choice from a set of alternatives, $S_n$, available to her/him. For each alternative $j \in S_n$, denote by $u_{n,j}$ its utility. We consider that the utility $u_{n,j}$ can be represented by

$$u_{n,j}(x_{n,j}, y_{n,j}, \beta, \gamma_n) \quad = \quad v_{n,j}(x_{n,j}, y_{n,j}, \beta, \gamma_n) + \varepsilon_{n,j}, \tag{2.1}$$

where $(x_{n,j}, y_{n,j})$ is the vector of attribute values that characterize alternative $j$, $\beta$ and $\gamma_n$ are the weights (tastes) the customer has on the corresponding attributes, $\varepsilon_{n,j}$ is the random error term, and $v_{n,j}(x_{n,j}, y_{n,j}, \beta, \gamma_n)$ is the systematic utility. The error term $\varepsilon_{n,j}$ is usually assumed to follow an extreme-value type distribution, e.g., Gumbel distribution with parameters (0,1). It is also assumed that $\varepsilon_{n,j}$'s are identically and independent distributed (i.i.d.) Gumbel random variables for each customer and across customers. The weights reflect the tastes of a customer in making choice given the attributes of the alternatives, where $\beta$ represents the taste coefficients that are deterministic and identical over the entire customer population and $\gamma_n$ denotes the random taste coefficients. It is assumed that $\gamma_n$'s are i.i.d. with a certain distribution for $n \in \mathcal{N}$. The random taste coefficients are most widely modeled as Gaussian vectors or log-normal random vectors. Assume furthermore that $\varepsilon_{n,j}$'s and $\gamma_n$'s are independent.

In general, the distribution of $\gamma_n$ is characterized by a parameter vector $\phi$ such as mean value and variance. By using the Monte Carlo (MC) sampling technique, we can represent $\gamma_n = t(\phi, \xi_n)$, where $\xi_n$ is a basic random vector with support $\Xi$, which is used by the sample generator such as standard uniform or standard normal random vector with distribution $P_{\text{in}}$. For instance, if $\gamma_n$ is a Gaussian vector with mean $\mu$ and covariance matrix $\Sigma = \sigma\sigma^\intercal$, where $\sigma$ is the lower-triangular Cholesky factor of $\Sigma$ and $\phi = (\mu, \sigma)$, we can represent $\gamma_n = \mu + \sigma\xi_n$, where $\xi_n$ is a standard Gaussian vector. Let $\theta = (\beta, \phi)$ be the parameter values that need to be estimated. We consider that $\theta$ is constrained in a nonempty compact convex set $\mathcal{C} \subseteq \mathbb{R}^d$ with $\ell_2-$norm $\|\cdot\| := \|\cdot\|_2$ and we assume that $\mathcal{C}$ can be described by a finite set of smooth equality or inequality constraints, i.e., $\mathcal{C} := \cap_{i=1}^{\bar{m}} \{x \in \mathbb{R}^d : c_i(x) \geq 0\}$, where $c_i : \mathbb{R}^d \mapsto \mathbb{R}$ is a twice continuously differentiable function and $\bar{m} \in \mathbb{N}$.

Let $z_{n,j} = 1$ if $j \in S_n$ is chosen by customer $n$ from $S_n$ and $z_{n,j} = 0$ otherwise. For customer $n$, let $j_n$ be the chosen alternative such that $z_{j_n} = 1$. Each customer is associated with one observation of data, $(x_{n,j}, y_{n,j}, z_{n,j}, j \in S_n)$. Since we have a finite set of customers, the dataset can be represented by the customer set $\mathcal{N}$. The likelihood function of parameter $\theta$ is equal to the (joint) probability of the observed customer choices, which is computed as

$$
\begin{aligned}
& \mathbb{P}\left(u_{n,j_n} > u_{n,j}, j \in S_n, j \neq j_n, \forall n \in \mathcal{N} | \theta\right) \\
= & \int_{\mathbb{R}^{d\times|\mathcal{N}|}} \mathbb{P}\left[u_{n,j_n}(\xi_n) > u_{n,j}(\xi_n), j \in S_n, j \neq j_n, \forall n \in \mathcal{N} \middle| \xi_n, n \in \mathcal{N}\right] \prod_{n\in\mathcal{N}} dP_{\text{in}}(\xi_n) \\
= & \int_{\mathbb{R}^{d\times|\mathcal{N}|}} \prod_{n\in\mathcal{N}} \left[\frac{\exp\{v_{n,j_n}(x_{n,j_n}, y_{n,j_n}, \beta, t(\phi,\xi_n))\}}{\sum_{j\in S_n} \exp\{v_{n,j}(x_{n,j}, y_{n,j}, \beta, t(\phi,\xi_n))\}} dP_{\text{in}}(\xi_n)\right] \\
= & \prod_{n\in\mathcal{N}} \mathbb{E}_\xi[F_n(\theta,\xi)],
\end{aligned}
$$

where $\xi$ has distribution $P_{\text{in}}$ and

$$F_n(\theta,\xi) \quad = \quad \frac{\exp\{v_{n,j_n}(x_{n,j_n}, y_{n,j_n}, \beta, t(\phi,\xi))\}}{\sum_{j\in S_n}\exp\{v_{n,j}(x_{n,j}, y_{n,j}, \beta, t(\phi,\xi)\}}$$

is the probability of customer $n \in \mathcal{N}$ choosing alternative $j_n \in S_n$ conditioning on $\xi$, which has a form of the multinomial logit (MNL) model based on the assumption that $\varepsilon_{n,j}$'s are i.i.d. Gumbel variables. Thus, the likelihood function can be written as

$$\mathcal{L}(\theta|\mathcal{N}) \quad := \quad \prod_{n\in\mathcal{N}} p_n(\theta),$$

where $p_n(\theta)$ is the choice probability of customer $n$ that is given by the mixed logit (ML) model,

$$p_n(\theta) \quad = \quad \mathbb{E}_\xi[F_n(\theta,\xi)].$$

The average log-likelihood function of the ML model is written as,

$$\ell(\theta|\mathcal{N}) \quad = \quad \frac{1}{|\mathcal{N}|}\sum_{n\in\mathcal{N}} \ln\left(\mathbb{E}_\xi[F_n(\theta,\xi)]\right). \tag{2.2}$$

Our goal is to estimate an optimal $\widehat{\theta} \in \mathcal{C}$ by solving the maximum likelihood estimation (MLE) of the ML model, i.e., to find $\widehat{\theta}$ that maximizes the average log-likelihood function (2.2). The problem can be considered as a special case of the following minimization problem,

$$\min\left\{f(\theta) \quad := \quad \frac{1}{|\mathcal{N}|}\sum_{n\in\mathcal{N}} r\left(\mathbb{E}_\xi[F_n(\theta,\xi)]\right)] : \theta\in\mathcal{C}\subset\mathbb{R}^d\right\}, \tag{2.3}$$

where $F_n(\cdot,\xi) : \mathcal{C} \mapsto \mathcal{X} \subset \mathbb{R}^m$ is a function for almost every $\xi \in \Xi$, $\mathcal{X}$ is a convex set with norm $\|\cdot\|_{\mathcal{X}}$, $F_n(\theta,\cdot) : \Xi \mapsto \mathcal{X}$ is measurable for all $\theta \in \mathcal{C}$ and $r : \mathcal{X} \mapsto \mathbb{R}$ is a function. In the average log-likelihood function (2.2), $r$ is the negative logarithm. We refer to (2.3) as the "true" problem and $f(\theta)$ as the "true" objective function. Let $\|\cdot\|_{m\times d}$ denote the norm for matrices in $\mathbb{R}^{m\times d}$ for any $m, d \in \mathbb{N}$.

## *2.2 Literature Review*

The ML model is one of the most popular variants of the MNL model and it assumes random taste/preference coefficients to capture the heterogeneity in customer preference, while the MNL assumes a deterministic and identical vector of preference coefficients across the customer population. [43] showed that any RUM can be approximated to any degree of accuracy by a ML model with appropriate choice of variables and mixing distribution. One one hand, the ML model has a flexible structure that enables users to model customer heterogeneity; on the other hand, estimating it can be computationally challenging, because it involves the computation of the high-dimensional integral incorporated in the model (e.g., the expectation in (2.2)). The estimation is essentially solving the MLE problem with a set of observations, and it adds into the estimation another dimension of difficulty if the number of observations is large. The simulation technique that solves the simulated/approximate MLE problem seems one of the only a few methods, in real applications, to solve the problem. There are two branches of studies that all use Monte Carlo (MC) related simulation techniques to compute multi-variate integrals involved in choice probabilities, but have different ideas to save the computational time.

### 2.2.1 The MC and Quasi-MC Simulation Methods

The idea to use the MC simulation technique to estimate multi-dimensional integrals in choice probabilities is not new; one may refer to Daganzo's monograph [24] in 1979 for a reference, which covers a comprehensive list of topics related to estimating the probit choice probability that is an integral based on multi-variate Gaussian distributed random errors of utilities (which are assumed to be standard Gumbel variables in the MNL model). Some other studies [16, 15] focused on deriving unbiased estimators of the likelihood in choice models.

The ML model estimation involves the expectation as the high-dimensional integral (see (2.2)). The MC simulation is usually used to approximate the expectation by using its sample average approximation (SAA), resulting in a simulated average log-likelihood function. [7] showed that the solution of maximizing the simulated average log-likelihood function of the ML model converges to the true MLE estimators almost surely, in terms of both the first- and second-order criticality conditions when sample size goes to infinity, which extends the results about the statistical inference of stochastic programming in [58]. The MC simulation technique usually includes a pseudo-random sequence generating process, which can be slow when the sample size increases. [33] emphasized that faster MC simulations are needed in practice.

Thus, a trend of studies on using quasi-MC techniques have risen since the end of 1990s. Bhat wrote a series of papers [14, 12, 13] to advocate using the Halton sequence for estimating ML models and reported a faster estimation results than using standard MC random samples. [34] proposed a modified Latin hypercube sampling (MLHS) method as an alternative to the Halton sequence and showed by numerical studies that the MLHS method performs better than the Halton sequence. However, [47] reported that the quasi-MC technique outperforms the standard MC sampling technique when the integration dimension is low, but the advantage of the quasi-MC technique compared to the standard MC technique is still unclear in computing high-dimensional integrals.

The MC and quasi-MC methods reviewed above concentrate on generating high performance sampling sequences to approximate high-dimensional integrals, but use a given optimization algorithm or software tool to solve the MLE estimation problem with a fixed number of sample size to compute the choice probability for each observation.

### 2.2.2 Optimization Algorithms to Estimate ML-type Models

The ML model estimation problem is a special case of stochastic programming. Another branch of studies are focused on developing efficient optimization algorithms to solve a stochastic programming model of such type.

Our work falls within this category of studies and [6] is one of such studies close to ours, which embeds an adaptive sampling process into the trust region algorithm to control the sample size for approximating the choice probability of each observation according to the sampling error incurred in the sampling process. The idea behind this method is that only a small number of random variables are needed for approximation when the iterate of the algorithm is not mature. The same idea was also adopted in [59] to solve a two-stage stochastic programming model. While our work adopts the same idea and also uses the trust region framework, it differs from [6] in four aspects: (1) our algorithm adaptively controls a different sample size to compute the SAA of the choice probability for each observation, according to the approximation error incurred in computing the SAA of that observation, while [6] always uses the sample size for all the observations, (2) [6] requires a maximum sample size for all the observations, while our algorithm does not have such a cap, (3) We solve a constrained optimization model by using the projected gradient method, while [6] solves a unconstrained optimization problem, and (4) Our algorithm also embeds a data sampling process in order to handle large-scale datasets, which has not been addressed in [6].

There is another trend of studies [31, 51] that address the problem of choosing the optimal sequence of sample sizes (effort) adaptively during solving a stochastic recursion that approximates a deterministic recursion with random samples and showed the convergence of the iterates to the true solution of the deterministic recursion in a rigorous sense. These studies pre-assume an algorithm framework and obtain convergence results under the assumption that the objective function is strongly convex, which points out an interesting direction of incorporating the mechanism of choosing an optimal sequence of sample sizes for our problem.

## *2.3 The Simulated Objective Function*

There are two major challenges in solving (2.3).

(1) The population size of observations is often very large (the size can be hundreds of thousands) so that the size of input data can be enormous. It makes computations of the objective function, gradient and Hessian inefficient.

(2) The objective function involves an expectation operator that is hard to compute since it may require multidimensional integral. When the dimension of the integral is greater than 5, we can hardly expect a high computational accuracy [49]. These challenges often make it intolerably expensive to evaluate the function value, gradient or Hessian. The algorithms that are typically used to solve deterministic nonlinear programming models are not quite applicable to solve (2.3), since the function value, gradient or Hessian needs to be evaluated at each iteration of the algorithm.

One idea is to use the sampling techniques to solve the approximation of problem (2.3). We have two types of samples to generate: (1) the sample from the data set which we call the *data sample*, and (2) the sample from the distribution of $\xi$, which is used to compute the expectation in the objective function and we refer to as *integration sample*.

Let $N \subset \mathcal{N}$ denote the set of customers associated with the sampled data such that $|N| \geq 2$, which we will later refer to as data sample. Let $P_{\text{out}}$ denote the probability that we sample any $N \subset \mathcal{N}$ with

$$P_{\text{out}}(N) = \frac{|N|!(|\mathcal{N}| - |N|)!}{|\mathcal{N}|!}, \quad \forall N \subset \mathcal{N}.$$

Note that the elements of $N$ are dependent if we sample without replacements. For each data $n \in N$, let $I_n$ be a set (sequence) of i.i.d. samples of $\xi$, i.e., $I_n = \{\xi_n^i : i = 1, 2, \ldots, |I_n|\}$, and $\mathcal{I} = \{I_n, n \in N\}$ be the set of integration samples associated with all the sample customers. It is worthwhile to notice that $\mathcal{I}$ is dependent on the set of sample customers $N$ and $|\mathcal{I}| = \sum_{n \in N} |I_n|$. Assume further that the components of integration samples are independent for each customer and cross customers. Thus, for each $n \in N$, the sample average approximation (SAA) of $p_n(\theta)$ can be represented as

$$p_{I_n}(\theta) \quad = \quad \frac{1}{|I_n|} \sum_{i=1}^{|I_n|} F_n(\theta, \xi_n^i).$$

For data sample $N$ and integration sample $\mathcal{I}$, the *approximate objective function* is written as

$$f_{\mathcal{I}}^N(\theta) \quad := \quad \frac{1}{|N|} \sum_{n \in N} r\left(p_{I_n}(\theta)\right). \tag{2.4}$$

We can view the approximate simulated function $f_{\mathcal{I}}^N(\theta)$ as defined on a common probability space $(\Omega, \mathcal{F}, P)$, where $\Omega := \bigcup_{N \in \mathcal{N}} \prod_{n \in N} \Xi^\infty$ and $P = P_{\text{out}} \prod_{n \in \mathcal{N}} P_{\text{in}}^\infty$ is a product measure on $\Omega$. By the statement "an event happens w.p. 1 for $K$ large enough" we mean that for $P-$almost every realization $\omega := \{N \subset \mathcal{N}, \xi_1^1, \xi_1^2, \ldots, \xi_2^1, \xi_2^2, \ldots, \ldots, \xi_{|N|}^1, \xi_{|N|}^2, \ldots\} \in \Omega$ of the random sequence, there exists integer $K(\omega)$ such that the considered event happens for all samples $\{N \subset \mathcal{N}, \xi_1^1, \xi_1^2, \ldots, \xi_1^k, \xi_2^1, \xi_2^2, \ldots, \xi_2^k, \ldots, \xi_{|N|}^1, \xi_{|N|}^2, \ldots, \xi_{|N|}^k\}$ from $\omega$ with $k \geq K(\omega)$. We use $x \wedge y := \min\{x, y\}$ and $x \vee y := \max\{x, y\}$ for any $x, y \in \mathbb{R}$. We also use $X \sim P_X$ to represent that $X$ is a random variable or random sample with distribution $P_X$.

## 2.4 Projections on a Convex Set

We consider $\mathcal{C}$ to be a nonempty closed convex set on which projections are fairly easy to compute, such as box-constrained domains or spheres. For any $x \in \mathbb{R}^d$, define the projection $\Pi_{\mathcal{C}}[x]$ of $x$ on $\mathcal{C}$ as

$$\Pi_{\mathcal{C}}[x] \quad := \quad \arg\min_{y \in \mathcal{C}} \|x - y\|.$$

It is well-known that the above projection exists and is uniquely defined on $\mathcal{C}$. The following lemma gives the properties of the projection on a convex set.

The following three properties of the projection will be useful and follow from Proposition 2.2.1 in [10] and its proof.

Let $\mathcal{C} \subset \mathbb{R}^d$ be a nonempty closed convex set and $\Pi_{\mathcal{C}}$ be the projection operator onto $\mathcal{C}$.

P.1 For any $y \in \mathcal{C}$, $\langle x - \Pi_{\mathcal{C}}[x], y - \Pi_{\mathcal{C}}[x] \rangle \leq 0$ for all $x \in \mathbb{R}^d$.

P.2 (Monotonicity) $\langle \Pi_{\mathcal{C}}[x] - \Pi_{\mathcal{C}}[y], x - y \rangle \geq 0$ for all $x, y \in \mathbb{R}^d$. If $\Pi_{\mathcal{C}}[x] \neq \Pi_{\mathcal{C}}[y]$, the strict inequality holds.

P.3 (Nonexpansiveness) $\|\Pi_{\mathcal{C}}[x] - \Pi_{\mathcal{C}}[y]\| \leq \|x - y\|$ for all $x, y \in \mathbb{R}^d$.

If $\mathcal{C}$ is a box-constrained domain defined by

$$\mathcal{C} \quad = \quad \{x \in \mathbb{R}^d : l_i \leq [x]_i \leq u_i, i = 1, 2, \ldots, d\},$$

the projection $\Pi_{\mathcal{C}}[x]$ of $x$ can be conveniently computed by,

$$[\Pi_{\mathcal{C}}[x]]_i \quad = \quad \begin{cases} l_i & \text{if} \quad [x]_i \in (-\infty, l_i], \\ x_i & \text{if} \quad [x]_i \in (l_i, u_i), \\ u_i & \text{if} \quad [x]_i \in [u_i, +\infty). \end{cases}$$

Let $\mathcal{H}(x)$ be the normal cone at $x \in \mathcal{C}$ with respect to $\mathcal{C}$ defined by

$$\mathcal{H}(x) \quad := \quad \{\theta \in \mathbb{R}^d : \theta^\intercal (y - x) \leq 0, \quad \forall y \in \mathcal{C}\},$$

and $\mathcal{T}(x)$ be the tangent cone at $x \in \mathcal{C}$ with respect to $\mathcal{C}$, which is defined by,

$$\mathcal{T}(x) \quad := \quad \text{cl}\{\lambda(y - x) : \lambda \geq 0, \quad y \in \mathcal{C}\},$$

where $\mathrm{cl}(C)$ denotes the closure of any set $C$.

For any set $\mathcal{C} \in \mathbb{R}^d$, we organize a list of assumptions that we may use in later results.

A.1. For each $n \in \mathcal{N}$, $F_n(\cdot, \xi) : \mathcal{C} \mapsto \mathcal{X}$ is continuous for almost every $\xi \in \Xi$ and $F_n(\theta, \cdot) : \Xi \mapsto \mathcal{X}$ is measurable for all $\theta \in \mathcal{C}$.

A.2. For any $\theta \in \mathcal{C}$, there exist a $\delta > 0$ and $K : \Xi \mapsto \mathbb{R}$ such that $\mathbb{E}(K(\xi)) < \infty$, and for almost every $\xi \in \Xi$ and every $n \in \mathcal{N}$,

$$\|F_n(\theta', \xi)\|_{\mathcal{X}} \leq K(\xi)$$

for all $\theta' \in \mathcal{C}$ such that $\|\theta' - \theta\| \leq \delta$.

A.3. For each $n \in \mathcal{N}$, $F_n(\cdot, \xi) : \mathcal{C} \mapsto \mathcal{X}$ is continuously differentiable for almost every $\xi \in \Xi$.

A.4. For any $\theta \in \mathcal{C}$, there exists $\delta > 0$ and $L : \Xi \mapsto \mathbb{R}$ such that $\mathbb{E}(L(\xi)) < \infty$, and for almost every $\xi \in \Xi$ and every $n \in \mathcal{N}$,

$$\|F_n(\theta'', \xi) - F_n(\theta', \xi)\|_{\mathcal{X}} \quad \leq \quad L(\xi)\|\theta'' - \theta'\|$$

for all $\theta'', \theta' \in B(\theta) := \{\theta' \in \mathcal{C} : \|\theta' - \theta\| \leq \delta\}$.

A.5. For any $\theta \in \mathcal{C}$, there exist a $\delta > 0$ and $K : \Xi \mapsto \mathbb{R}$ such that $\mathbb{E}(K(\xi)) < \infty$, and for almost every $\xi \in \Xi$ and every $n \in \mathcal{N}$,

$$\|F_n(\theta', \xi)F_n(\theta', \xi)^\intercal\|_{m \times m} \leq K(\xi)$$

for all $\theta' \in \mathcal{C}$ such that $\|\theta' - \theta\| \leq \delta$.

A.6. $r : \mathcal{X} \mapsto \mathbb{R}$ is continuous.

A.7. $r : \mathcal{X} \mapsto \mathbb{R}$ is continuously differentiable.

A.8. $r : \mathcal{X} \mapsto \mathbb{R}_-$ is a continuously differentiable concave function,

A.9. $\mathbb{E}\left[\sup_{\theta \in \mathcal{C}} \left[r(F_n(\theta, \xi))\right]^4\right] < \infty$ for every $n \in \mathcal{N}$.

A.10. $\xi_n^1, \xi_n^2, \ldots, \xi_n^I, \ldots$ is a sequence of i.i.d. observations with distribution $P_{\text{in}}$ for all $n \in \mathcal{N}$.

## *2.5 The Stochastic Trust-Region Algorithm*

### 2.5.1 The Model Function

In general, an algorithm for solving nonlinear programming models exploits an iterative search strategy. We propose a stochastic trust-region algorithm to solve the true problem (2.3), where a sequence of points in $\mathcal{C}$ will be iteratively generated until a stopping criterion is satisfied. At step $k$, let $\theta_k \in \mathcal{C}$ be the current point, $N_k$ be data sample, $I_n^k$ be the integration sample for $n \in N_k$ and $\mathcal{I}_k$ be the integration samples for all data points in $N_k$.

At step $k$, we define a quadratic model function that approximates the objective function in (2.4) within a neighborhood of $\theta_k$, which is often referred to as the *trust region.* The trust region is defined as a ball centered at $\theta_k$,

$$\mathcal{B}_k \quad := \quad \{\theta \in \mathbb{R}^d : \|\theta - \theta_k\| \leq \Delta_k\},$$

where $\Delta_k$ is the trust-region radius. Under Assumptions A.3 and A.7, the model function $m_k(\theta_k + s)$ is defined as

$$m_k(\theta_k + s) \quad := \quad f_k(\theta_k) + g_k^\intercal s + \frac{1}{2} s^\intercal H_k s, \tag{2.5}$$

where

$$\begin{aligned} f_k(\cdot) \quad &:= \quad f_{\mathcal{I}_k}^{N_k}(\cdot), \\ g_k \quad &:= \quad \nabla f_k(\theta_k), \end{aligned}$$

and $H_k$ represents the Hessian or the approximation of the Hessian (e.g., the BFGS or SR1 approximations) associated with the approximate function $f_k$.

We then attempt to find a *trial step* $s_k$ to sufficiently reduce the model function while maintaining the search within the trust region and the feasible set, i.e., we aim to find

$$s_k \in \arg\min_{s \in \mathbb{R}^d} \left\{ m_k(\theta_k + s) : \theta_k + s \in \mathcal{C} \cap \mathcal{B}_k \right\}. \tag{2.6}$$

Exactly solving (2.6) is far from easy and is not necessary. We can approximately solve it by searching along the projected gradient path and applying the Goldstein-type line search rule, where the *projected gradient path* (with respect to data sample $N_k$ and the integral sample $\mathcal{I}_k$) is defined by

$$v_k(t, \theta_k) \quad = \quad \Pi_{\mathcal{C}}[\theta_k - t g_k] \quad \forall t \geq 0.$$

Define

$$s_k(t) \quad := \quad v_k(t, \theta_k) - \theta_k, \quad t \geq 0,$$

and we aim to find a $t \geq 0$ such that the following two conditions

$$\|s_k(t)\| \quad \leq \quad \Delta_k, \tag{2.7}$$

$$m_k(\theta_k) - m_k(v_k(t, \theta_k)) \quad \geq \quad -\kappa_1 g_k^\intercal s_k(t), \tag{2.8}$$

hold and one of the following conditions

$$m_k(\theta_k) - m_k(v_k(t, \theta_k)) \quad < \quad -\kappa_2 g_k^\intercal s_k(t) \tag{2.9}$$

$$\|s_k(t)\| \quad \geq \quad \kappa_3 \Delta_k \tag{2.10}$$

$$\left\|\Pi_{\mathcal{T}(v_k(t,\theta_k))}[-g_k]\right\| \quad \leq \quad \kappa_4 \frac{|g_k^\intercal s_k(t)|}{\Delta_k} \tag{2.11}$$

is satisfied, where

$$\kappa_1 \in (0,1), \kappa_2 \in (\kappa_1, 1), \kappa_3 \in (0,1), \kappa_4 \in (0, 1/2). \tag{2.12}$$

If such $t$ is found, we define the point $\theta_k^{\mathrm{GC}} := v_k(t, \theta_k)$ and consider it as a candidate for $\theta_{k+1}$. Several test criteria need to be satisfied before we formally accept the candidate as the next point. One test is to evaluate how well the model function approximates the true objective function and adjust the trust-region radius according the test result, as the basic trust-region method does. We call

$$\Delta f_k(\theta_k, \theta_k^{\mathrm{GC}}) \quad := \quad f_k(\theta_k) - f_k(\theta_k^{\mathrm{GC}})$$

the *actual improvement* and

$$\Delta m_k(\theta_k, \theta_k^{\mathrm{GC}}) \quad := \quad m_k(\theta_k) - m_k(\theta_k^{\mathrm{GC}})$$

and the *predicted improvement.* Let

$$\rho_k \quad := \quad \frac{\Delta f_k(\theta_k, \theta_k^{\mathrm{GC}})}{\Delta m_k(\theta_k, \theta_k^{\mathrm{GC}})}$$

be the ratio that measures the agreement between the model function and the simulated objective function.

For any $g \in \mathbb{R}^d$, $\theta \in \mathcal{C}$ and $\delta > 0$, define

$$\chi(g, \theta, \delta) \quad := \quad |\min\{g^\intercal s : \theta + s \in \mathcal{C}, \|s\| \leq \delta\}|.$$

In the following we lists a variety of properties of $\chi(g_k, \theta_k, \delta)$, which follow from Theorem 12.1.3, Theorem 12.1.4 and Theorem 12.1.5 (i) and (ii) in [23].

Under Assumptions A.3 and A.7, the following holds for each $k$:

P.4 Both $s_k(t)$ and $\|s_k(t)\|$ are continuous in $t$ and $\|s_k(t)\|$ is nondecreasing for all $t \geq 0$.

P.5 The limit $\lim_{t\to\infty} \|s_k(t)\| < \infty$ implies $\lim_{t\to\infty} \|\Pi_{\mathcal{T}(v_k(t,\theta_k))}[-g_k]\| = 0$.

P.6 $s_k(t)$ is a solution of the problem $\min\{g_k^\intercal s : \theta_k + s \in \mathcal{C}, \|s\| \leq \|s_k(t)\|\}$ for all $t \geq 0$.

P.7 $\chi(g_k, \theta_k, \delta)$ is continuous and nondecreasing as a function of $\delta$ for all $\delta \geq 0$.

P.8 The function $\chi(g_k, \theta_k, \delta)/\delta$ is nonincreasing as a function of $\delta$ for all $\delta > 0$.

It follows from Theorem 12.1.6 of [23] that $\chi_k(\theta) := \chi(\nabla f_k(\theta), \theta, 1)$ is the first-order criticality measure for the approximate problem (2.4), i.e., it is nonnegative, continuous and vanishes at $\theta \in \mathcal{C}$ if and only if $-\nabla f_k(\theta) \in \mathcal{H}(\theta)$. Due to the randomness involved, there exists noise in the actual improvement. We thus need to evaluate the errors incurred in the sampling processes and compare the errors with the actual improvement before we calculate $\rho_k$. By a similar idea, we define the first-order criticality measure function as $\chi(\theta) := \chi(\nabla f(\theta), \theta, 1)$.

### 2.5.2 Sampling Errors

For any point $\theta \in \mathcal{C}$, data point $n \in \mathcal{N}$, and integration sample $I_n$, it holds that $\mathbb{E}[p_{I_n}(\theta)] = p_n(\theta)$. Under A.7, it follows from Taylor's first-order expansion and the differentiability of $r$ that

$$r\left(p_{I_n}(\theta)\right) \quad \approx \quad r\left(p_n(\theta)\right) + \nabla r\left(p_n(\theta)\right)^\intercal [p_{I_n}(\theta) - p_n(\theta)],$$

which yields that

$$\mathbb{E}[r(p_{I_n}(\theta))] \approx r(p_n(\theta)) \approx r(p_{I_n}(\theta)), \tag{2.13}$$

$$\mathrm{Var}[r(p_{I_n}(\theta))] \approx \nabla r(p_n(\theta))^{\intercal}\,\mathrm{Var}[p_{I_n}(\theta)]\nabla r(p_n(\theta))]. \tag{2.14}$$

It follows from the law of total variance that the variance of the simulated objective functions evaluated at $\theta$ is computed as,

$$\mathrm{Var}(f_{\mathcal{I}}^{N}(\theta)) = \sigma_1^2(\theta, |N|, \mathbf{I}) + \sigma_2^2(\theta, |N|, \mathbf{I}),$$

where

$$\begin{aligned}\mathbf{I} &= (|I_n|, n \in \mathcal{N}),\\ \sigma_1^2(\theta, |N|, \mathbf{I}) &:= \mathrm{Var}\left(\mathbb{E}[f_{\mathcal{I}}^{N}(\theta)|N]\right),\\ \sigma_2^2(\theta, |N|, \mathbf{I}) &:= \mathbb{E}\left(\mathrm{Var}[f_{\mathcal{I}}^{N}(\theta)|N]\right).\end{aligned}$$

Now, consider any two points $\theta_1, \theta_2 \in \mathcal{C}$. The variance of the difference between the simulated objective function values evaluated at $\theta_1, \theta_2 \in \mathcal{C}$ is computed as,

$$\mathrm{Var}(f_{\mathcal{I}}^{N}(\theta_1) - f_{\mathcal{I}}^{N}(\theta_2)) = \sigma_1^2(\theta_1, \theta_2, N, \mathcal{I}) + \sigma_2^2(\theta_1, \theta_2, N, \mathcal{I}),$$

where

$$\begin{aligned}\sigma_1^2(\theta_1, \theta_2, |N|, \mathbf{I}) &= \mathrm{Var}\left(\mathbb{E}[f_{\mathcal{I}}^{N}(\theta_1) - f_{\mathcal{I}}^{N}(\theta_2)|N]\right),\\ \sigma_2^2(\theta_1, \theta_2, |N|, \mathbf{I}) &= \mathbb{E}\left(\mathrm{Var}[f_{\mathcal{I}}^{N}(\theta_1) - f_{\mathcal{I}}^{N}(\theta_2)|N]\right),\end{aligned}$$

and we refer to $\sigma_1(\theta_1, \theta_2, |N|, \mathbf{I})$ as the data sampling error and $\sigma_2(\theta_1, \theta_2, |N|, \mathbf{I})$ as the integration sampling error in the change of approximate function values associated with $\theta_1$ and $\theta_2$.

#### *2.5.2.1 Data Sampling Error*

For any data sample $N$ and integration sample $\mathcal{I}$, the data sampling error is measured by the standard deviation of the difference between the approximate objective function values with respect to two points $\theta_1, \theta_2 \in \mathcal{C}$. Thus,

$$\sigma_1^2(\theta_1, \theta_2, |N|, \mathbf{I}) := \mathrm{Var}\left(\mathbb{E}[f_{\mathcal{I}}^{N}(\theta_1) - f_{\mathcal{I}}^{N}(\theta_2)|N]\right)$$

$$
\begin{aligned}
&= \mathrm{Var}_N \left( \frac{1}{|N|} \sum_{n \in N} \mathbb{E}[\Delta r_{I_n}(\theta_1, \theta_2)] \right) \\
&= \frac{\kappa_c}{|N||\mathcal{N}|} \sum_{n \in \mathcal{N}} \left( \mathbb{E}[\Delta r_{I_n}(\theta_1, \theta_2)] - \frac{\sum_{n \in \mathcal{N}} \mathbb{E}[\Delta r_{I_n}(\theta_1, \theta_2)]}{|\mathcal{N}|} \right)^2,
\end{aligned}
$$

where

$$
\begin{aligned}
\kappa_c &= \frac{|\mathcal{N}| - |N|}{|\mathcal{N}| - 1}, \\
\Delta r_{I_n}(\theta_1, \theta_2) &= r(p_{I_n}(\theta_1)) - r(p_{I_n}(\theta_2)).
\end{aligned}
$$

is the correction factor for sampling without replacements from a finite population. It follows from (2.13) that the data sampling error can be further approximated by the (random) approximate data sampling error, $\widehat{\sigma}_1(\theta_1, \theta_2, N, \mathcal{I})$, which is defined as,

$$
\widehat{\sigma}_1(\theta_1, \theta_2, N, \mathcal{I}) := \sqrt{\frac{|\mathcal{N}| - |N|}{|N||\mathcal{N}|} \bar{\nu}^2(\theta_1, \theta_2, N, \mathcal{I})},
$$

where

$$
\bar{\nu}^2(\theta_1, \theta_2, N, \mathcal{I}) = \frac{1}{|N| - 1} \sum_{n \in N} \left( \Delta r_{I_n}(\theta_1, \theta_2) - \frac{\sum_{n \in N} \Delta r_{I_n}(\theta_1, \theta_2)}{|N|} \right)^2
$$

is the sample approximation of the variance of a data point sampled from the dataset.

#### *2.5.2.2 Monte-Carlo Integration Sampling Error*

Note that $I_1, I_2, \ldots, I_{\mathcal{N}}$ are independent. The squared integration sampling error at $\theta_1, \theta_2 \in \mathcal{C}$ is computed by,

$$
\begin{aligned}
\sigma_2^2(\theta_1, \theta_2, |N|, \mathbf{I}) &= \mathbb{E}(\mathrm{Var}\left[ f_{\mathcal{I}}^N(\theta_1) - f_{\mathcal{I}}^N(\theta_2) | N \right]) \\
&= \mathbb{E} \left( \mathrm{Var} \left[ \frac{1}{|N|} \sum_{n \in N} \Delta r_{I_n}(\theta_1, \theta_2) \middle| N \right] \right) \\
&= \mathbb{E}_N \left( \frac{1}{|N|^2} \sum_{n \in N} \mathrm{Var}\left[ \Delta r_{I_n}(\theta_1, \theta_2) \right] \right) \\
&= \frac{1}{|N||\mathcal{N}|} \sum_{n \in \mathcal{N}} \mathrm{Var}[\Delta r_{I_n}(\theta_1, \theta_2)].
\end{aligned}
$$

It follows from (2.14) that the integration sampling error can be approximated by the following (random) approximate integration sampling error,

$$
\widehat{\sigma}_2(\theta_1, \theta_2, N, \mathcal{I}) := \sqrt{\frac{1}{|N||\mathcal{N}|} \sum_{n \in \mathcal{N}} \mathrm{Var}[\nabla r(p_n(\theta_1))^\intercal p_{I_n}(\theta_1) - \nabla r(p_n(\theta_2))^\intercal p_{I_n}(\theta_2)]}
$$

$$\approx \sqrt{\frac{1}{|N|^2}\sum_{n\in N}\frac{1}{|I_n|}\mathrm{Var}[\nabla r(p_n(\theta_1))^\intercal F_n(\theta_1,\xi) - \nabla r(p_n(\theta_2))^\intercal F_n(\theta_2,\xi)]}$$
$$\approx \frac{1}{|N|}\sqrt{\sum_{n\in N}\frac{1}{|I_n|}\nu_n^2(\theta_1,\theta_2,I_n)}$$

where

$$S_n^2(\theta) = \frac{1}{|I_n|-1}\sum_{i=1}^{|I_n|}(F_n(\theta,\xi_n^i) - p_{I_n}(\theta))(F_n(\theta,\xi_n^i) - p_{I_n}(\theta))^\intercal,$$
$$S_n(\theta_1,\theta_2) = \frac{1}{|I_n|-1}\sum_{i=1}^{|I_n|}(F_n(\theta_1,\xi_n^i) - p_{I_n}(\theta_1))(F_n(\theta_2,\xi_n^i) - p_{I_n}(\theta_2))^\intercal,$$

are the sample approximations of $\mathrm{Var}(F_n(\theta,\xi))$ and $\mathrm{Cov}(F_n(\theta_1,\xi), F_n(\theta_2,\xi))$, respectively, and

$$\begin{aligned}\nu_n^2(\theta_1,\theta_2,I_n) &= \nabla r(p_{I_n}(\theta_1))^\intercal S_n^2(\theta_1)\nabla r(p_{I_n}(\theta_1)) + \nabla r(p_{I_n}(\theta_2))^\intercal S_n^2(\theta_2)\nabla r(p_{I_n}(\theta_2)) \\ &- 2\nabla r(p_{I_n}(\theta_1))^\intercal S_n(\theta_1,\theta_2)\nabla r(p_{I_n}(\theta_2)).\end{aligned}$$

Note that $\nu_n^2(\theta_1,\theta_2,I_n) \geq 0$ is a sample approximation of $\mathrm{Var}[r'(p_n(\theta_1))p_{I_n}(\theta_1) - r'(p_n(\theta_2))p_{I_n}(\theta_2)]$ and will not change much when $|I_n|$ becomes large enough. The following notations for any iteration $k$ during the execution of Algorithm 1 will be useful. Define $\widehat{\sigma}_{k,i} := \widehat{\sigma}_i(\theta_k, \theta_k^{\mathrm{GC}}, N_k, \mathcal{I}_k)$ for $i \in \{1, 2\}$.

### 2.5.3 The Algorithm

The formal stochastic trust-region algorithm (STRA) is described in Algorithm 1. The algorithm for search for the generalized Cauchy point is described in Algorithm 2 and the algorithm for sample update is described in Algorithm 3.

### 2.5.4 Optimal allocation of integration samples

It is desirable to minimize the computational efforts while increasing the integration sample size to reduce the integration sample error. We describe in this section the method to allocate Monte Carlo simulation samples to different customers based on the sampling error of each customer. Consider an iteration with index $k$. It follows that

$$\widehat{\sigma}_{k,2}^2 = \frac{1}{|N_k|^2}\sum_{n\in N_k}\frac{\nu_n^2(\theta_k,\theta_k^{\mathrm{GC}},I_n^k)}{|I_n^k|}.$$

**Algorithm 1:** STOCHASTIC TRUST-REGION ALGORITHM (STRA)

**Step 0. Initialization.** An initial point $\theta_0 \in \mathcal{C}$ and trust-region radius $\Delta_0 \in (0, \infty)$ are given. Choose $\eta_1$, $\eta_2$, $\eta_3$ and $\tau_1, \tau_2$ such that $0 < \eta_1 < \eta_2 < \eta_3 < 1$ and $0 < \tau_1 < 1 < \tau_2$ together with $\eta_0 \in (0, 1)$. Choose the initial data $N_0 \subset \mathcal{N}$ and integration sample $\mathcal{I}_0$. Set $k \leftarrow 0$.

**Step 1. Model function definition.** Compute the simulated objective function $f_k(\theta_k)$, gradient $g_k$ and Hessian matrix $H_k$ using the sample sets $N_k$ and $\mathcal{I}_k$. Obtain a model function $m_k$ defined in (2.5) over $\mathcal{B}_k \cap \mathcal{C}$.

**Step 2. The generalized Cauchy point calculation.** If $s_k(1) = 0$, set $\alpha_k = 0$; otherwise, find $\alpha_k > 0$ such that $s_k(\alpha_k) = v_k(\alpha_k, \theta_k) - \theta_k$ and $v_k(\alpha_k, \theta_k)$ satisfy both the trust region constraint (2.7) and (2.8), and either of (2.9), (2.10) and (2.11) (See **Algorithm** 2). Set $\theta_k^{\mathrm{GC}} \leftarrow v_k(\alpha_k, \theta_k)$ to be the generalized Cauchy point.

**Step 3. Sample update.** Compute the predicted improvement $\Delta m_k(\theta_k, \theta_k^{\mathrm{GC}})$, data sampling error $\widehat{\sigma}_{k,1}$ and integration sampling error $\widehat{\sigma}_{k,2}$. Generate $N_k^+$ and $\mathcal{I}_k^+$ by using **Algorithm** 3. Set $N_{k+1} \leftarrow N_k^+$, $\mathcal{I}_{k+1} \leftarrow \mathcal{I}_k^+$

**Step 4. Acceptance of the trial point.** If

$$\widehat{\sigma}_{k,1} \vee \widehat{\sigma}_{k,2} \quad \leq \quad \eta_0 \left( \Delta m_k(\theta_k, \theta_k^{\mathrm{GC}}) \right) \tag{2.15}$$

holds, compute $\Delta f_k(\theta_k, \theta_k^{\mathrm{GC}})$ and the ratio $\rho_k$. If $\rho_k \geq \eta_1$, then define $\theta_{k+1} \leftarrow \theta_k^{\mathrm{GC}}$; otherwise define $\theta_{k+1} \leftarrow \theta_k$ and go to **Step 5**; If (2.15) fails, define $\theta_{k+1} \leftarrow \theta_k$, set $k \leftarrow k + 1$, and go to **Step 1**.

**Step 5. Trust-region radius update.** Set

$$\Delta_{k+1} = \begin{cases} \tau_1 \Delta_k & \text{if} \quad \rho_k \in (0, \eta_2), \\ \Delta_k & \text{if} \quad \rho_k \in [\eta_2, \eta_3), \\ \tau_2 \Delta_k & \text{if} \quad \rho_k \in [\eta_3, \infty). \end{cases}$$

Set $k \leftarrow k + 1$ and go to **Step 1**.

**Algorithm 2:** SEARCH FOR THE GENERALIZED CAUCHY POINT

**Input**: Trust region radius $\Delta_k$, the iterate $\theta_k \in \mathcal{C}$ and the model $m_k(\theta_k + s)$. The constants $\kappa_1$, $\kappa_2$ and $\kappa_3$ satisfying condition (2.12). Set $t_{\min} = 0$, $t_1 = \Delta_k / \|g_k\|$ and $j = 1$.

**Output**: $\alpha_k$

**for** $j = 1$ *to* $\infty$ **do**
- $v_k(t_j, \theta_k) \leftarrow \Pi_{\mathcal{C}}[\theta_k - t_j g_k]$;
- $s_k(t_j) \leftarrow v_k(t_j, \theta_k) - \theta_k$;
- Compute $m_k(v_k(t_j, \theta_k))$;
- **if** *(2.7) or (2.8) is violated* **then**
  - $t_{\max} \leftarrow t_j$;
  - $t_{j+1} \leftarrow \frac{1}{2}(t_{\min} + t_{\max})$;
  - $i \leftarrow j + 1$ and stop;
- **else if** *all the (2.9),(2.10) and (2.11) are violated* **then**
  - $t_{\min} = t_j$;
  - $t_{j+1} = 2t_j$;
- **else**
  - $\alpha_k \leftarrow t_j$ and stop;

**for** $j = i$ *to* $\infty$ **do**
- $v_k(t_j, \theta_k) \leftarrow \Pi_{\mathcal{C}}[\theta_k - t_j g_k]$;
- $s_k(t_j) \leftarrow v_k(t_j, \theta_k) - \theta_k$;
- Compute $m_k(v_k(t_j, \theta_k))$;
- **if** *(2.7) or (2.8) is violated* **then**
  - $t_{\max} \leftarrow t_j$;
  - $t_{j+1} \leftarrow \frac{1}{2}(t_{\min} + t_{\max})$;
- **else if** *all the (2.9),(2.10) and (2.11) are violated* **then**
  - $t_{\min} \leftarrow t_j$;
  - $t_{j+1} \leftarrow \frac{1}{2}(t_{\min} + t_{\max})$;
- **else**
  - $\alpha_k \leftarrow t_j$ and stop;

**Algorithm 3:** Sample update mechanism

**Input**: $N_k$, $\mathcal{I}_k$, $\widehat{\sigma}_{k,1}$, $\widehat{\sigma}_{k,2}$, $\Delta m_k(\theta_k, \theta_k^{\mathrm{GC}})$, $n_k$, $n_a \in \mathbb{N}$, $i_a \in \mathbb{N}$ and $i_a \geq 2$, $\eta_0$, $P_{k,i}^j$, where $i \in \{1,2\}$ and $j \in \{0,1\}$, $P_{\mathrm{in}}$

**Output**: $N_k^+$, $\mathcal{I}_k^+$

**if** *(2.15) holds* **then**
  Generate two independent random variables, $X_{k,1} \sim P_{k,1}^1$ and $X_{k,2} \sim P_{k,2}^1$;
  **case** *1:* $X_{k,1} = 0$ *and* $X_{k,2} = 0$
    $N_k^+ \leftarrow N_k$, $I_n^{k+} \leftarrow I_n^k$ for all $n \in N_k^+$;
  **case** *2:* $X_{k,1} = 1$ *and* $X_{k,2} = 0$
    $I_n^{k+} \leftarrow I_n^k$ for all $n \in N_k$;
    Sample a set $N^{\mathrm{add}}$ of $n_a \wedge (|\mathcal{N}| - |N_k|)$ data points from $\mathcal{N} \setminus N_k$ and $N_k^+ \leftarrow N_k \cup N^{\mathrm{add}}$;
    Generate a set $I_n^{k+}$ of $i_a$ observations of $\xi \sim P_{\mathrm{in}}$ for $n \in N^{\mathrm{add}}$;
  **case** *3:* $X_{k,1} = 0$ *and* $X_{k,2} = 1$
    $N_k^+ \leftarrow N_k$;
    Generate a set $I_n^{\mathrm{add}}$ of $i_a$ observations of $\xi \sim P_{\mathrm{in}}$ and $I_n^{k+} \leftarrow I_n^k \cup I_n^{\mathrm{add}}$ for $n \in N_k^+$;
  **case** *4:* $X_{k,1} = 1$ *and* $X_{k,2} = 1$
    Generate a set $I_n^{\mathrm{add}}$ of $i_a$ observations of $\xi \sim P_{\mathrm{in}}$ and $I_n^{k+} \leftarrow I_n^k \cup I_n^{\mathrm{add}}$ for $n \in N_k$;
    Sample a set $N^{\mathrm{add}}$ of $n_a \wedge (|\mathcal{N}| - |N_k|)$ data points from $\mathcal{N} \setminus N_k$ and $N_k^+ \leftarrow N_k \cup N^{\mathrm{add}}$;
    Generate a set $I_n^{k+}$ of $i_a$ observations of $\xi \sim P_{\mathrm{in}}$ for $n \in N^{\mathrm{add}}$;
**else**
  **case** *5:* $\widehat{\sigma}_{k,1} > \widehat{\sigma}_{k,2} \geq 0$
    Generate a random variable $X_{k,2} \sim P_{k,2}^0$;
    **if** $X_{k,2} = 1$ **then**
      Generate a set $I_n^{\mathrm{add}}$ of $i_a$ observations of $\xi \sim P_{\mathrm{in}}$ and $I_n^{k+} \leftarrow I_n^k \cup I_n^{\mathrm{add}}$ for $n \in N_k$;
    **else**
      $I_n^{k+} \leftarrow I_n^k$ for $n \in N_k$;
    Sample a set $N^{\mathrm{add}}$ of $n_k$ data points from $\mathcal{N} \setminus N_k$ and $N_k^+ \leftarrow N_k \cup N^{\mathrm{add}}$;
    Generate a set $I_n^{k+}$ of $i_a$ observations of $\xi \sim P_{\mathrm{in}}$ for $n \in N^{\mathrm{add}}$;
  **case** *6:* $\widehat{\sigma}_{k,2} \geq \widehat{\sigma}_{k,1} \geq 0$
    Generate $\{I_n^{k+} : n \in N_k\}$ by using sample allocation Algorithm 4;
    Generate a random variable $X_{k,1} \sim P_{k,1}^0$;
    **if** $X_{k,1} = 1$ **then**
      Sample a set $N^{\mathrm{add}}$ of $n_a \wedge (|\mathcal{N}| - |N_k|)$ data points from $\mathcal{N} \setminus N_k$ and $N_k^+ \leftarrow N_k \cup N^{\mathrm{add}}$;
      Generate a set $I_n^{k+}$ of $i_a$ observations of $\xi \sim P_{\mathrm{in}}$ for $n \in N^{\mathrm{add}}$;
    **else**
      $N_k^+ \leftarrow N_k$;
  $\mathcal{I}_k^+ \leftarrow \{I_n^{k+} : n \in N_k^+\}$;

According to Algorithm 3, the sample allocation Algorithm 4 is executed for the current sampled set of customers, $N_k$, and we might need to expand the Monte-Carlo integration sample size for each $n \in N_k$ to reduce integration sampling error. To save the computational expense, we allocate samples to each customer with the objective of minimizing the total number of samples needed subject to decreasing the integration sampling error to a fractional level. Let $y_n$ denote the number of samples allocated to customer $n \in N_k$. Thus, we need to solve the optimization model,

$$\begin{aligned} \min \quad & \sum_{n \in N_k} y_n \\ \text{s.t.} \quad & \sum_{n \in N_k} \frac{\nu_n^2(\theta_k, \theta_k^{\text{GC}}, I_n^k)}{y_n} \leq (\eta_0 |N_k| \Delta m(\theta_k, \theta_k^{\text{GC}}))^2. \end{aligned}$$

Let $\lambda \geq 0$ be the Lagrangian multiplier. The Lagrangian function is written as,

$$L(\lambda, y_n, n \in N_k) = \sum_{n \in N_k} y_n + \lambda \left( \sum_{n \in N_k} \frac{\nu_n^2(\theta_k, \theta_k^{\text{GC}}, I_n^k)}{y_n} - (\eta_0 |N_k| \Delta m(\theta_k, \theta_k^{\text{GC}}))^2 \right),$$

which yields

$$y_n = \frac{\nu_n(\theta_k, \theta_k^{\text{GC}}, I_n^k) \left( \sum_{m \in N_k} \nu_m(\theta_k, \theta_k^{\text{GC}}, I_m^k) \right)}{(\eta_0 |N_k| \Delta m(\theta_k, \theta_k^{\text{GC}}))^2}, \quad \forall n \in N_k,$$

and we define

$$i_n = \lceil y_n \rceil, \quad \forall n \in N_k. \tag{2.16}$$

**Algorithm 4:** SAMPLE ALLOCATION

**Input**: $N_k$, $\mathcal{I}_k$, $(i_n, n \in N_k)$ defined in (2.16), $\widehat{\sigma}_{k,2}$ and $\eta_4$
**Output**: $I_n^{k+}$ for all $n \in N_k$
1 Generate a set $I_n^{\text{add}}$ of $i_a \vee (i_n - |I_n^k|)$ samples of $\xi \sim P_{\text{in}}$;
2 $I_n^{k+} \leftarrow I_n^k \cup I_n^{\text{add}}$ for $n \in N_k$;

Thus, we have the sample allocation algorithm 4. The following quantities are also defined for each $k$ in Algorithm 3.

$$n_k = \left\lceil \frac{|\mathcal{N}|}{1 + |\mathcal{N}| \frac{(\eta_0 \Delta m(\theta_k, \theta_k^{\text{GC}}))^2}{\bar{\nu}^2(\theta_k, \theta_k^{\text{GC}}, N_k, \mathcal{I}_k)}} \right\rceil - |N_k|, \tag{2.17}$$

$$
\begin{aligned}
q^1_{k,i} &= \frac{\epsilon + h(\widehat{\sigma}_{k,i})}{\epsilon + h(\eta_0 \Delta m(\theta_k, \theta_k^{\mathrm{GC}}))} \quad \forall i \in \{1,2\}, \\
q^0_{k,i} &= \frac{\epsilon + h(\widehat{\sigma}_{k,i})}{\epsilon + h(\widehat{\sigma}_{k,1}) + h(\widehat{\sigma}_{k,2})} \quad \forall i \in \{1,2\},
\end{aligned}
$$

where $\epsilon > 0$ and $h : \mathbb{R} \mapsto [0, \infty)$ is a nondecreasing continuous function. Let $P^j_{k,i}$ denote the distribution on a random variable $X$ such that $\mathbb{P}(X = 1) = q^j_{k,i}$ and $\mathbb{P}(X = 0) = 1 - q^j_{k,i}$ for each $k$, $i \in \{1, 2\}$ and $j \in \{0, 1\}$.

## 2.6 Convergence Analysis

Since $\mathcal{N}$ is a finite set, w.l.o.g., we show Theorems 2.1 and 2.2 for a generic data point $n \in \mathcal{N}$ and the results remain valid when we consider the entire dataset $\mathcal{N}$. For notational convenience we suppress data index $n$ in the statements of the Theorems 2.1 and 2.2 and their proofs.

**Theorem 2.1.** *Let $\mathcal{C}$ be a nonempty compact subset of $\mathbb{R}^d$ with norm $\|\cdot\|_{\mathcal{C}}$. Assume A.1, A.2, A.6, and A.10. Let $p(\theta) := \mathbb{E}(F(\theta, \xi))$ and $p_I(\theta) := \frac{1}{I}\sum_{i=1}^{I} F(\theta, \xi^i)$ for $I \in \mathbb{N}$ and $\theta \in \mathcal{C}$. Then, $p(\theta) \in \mathcal{X}$, $\varphi(\theta) := r(p(\theta))$ is continuous, and w.p. 1,*

$$
\sup_{\theta \in \mathcal{C}} |r(p_I(\theta)) - r(p(\theta))| \to 0 \text{ as } I \to \infty.
$$

**Proof:** It follows from Assumption A.2 that $\mathbb{E}[\|F(\theta,\xi)\|_{\mathcal{X}}] < \mathbb{E}[K(\xi)] < \infty$ for all $\theta \in \mathcal{C}$. Since $\mathcal{X}$ is a convex set, it follows from [17, p. 25] that $p(\theta) := \mathbb{E}[F(\theta,\xi)] \in \mathcal{X}$ for all $\theta \in \mathcal{C}$.

First we show continuity of $\varphi$. Choose any $\theta \in \mathcal{C}$ and a sequence $\{\theta_k\}_{k=1}^{\infty} \subset \mathcal{C}$ such that $\theta_k \to \theta$ as $k \to \infty$. It follows from continuity of $r$ and $F$ and the Dominated Convergence Theorem (DCT) that

$$
\begin{aligned}
\lim_{k\to\infty} r(\mathbb{E}(F(\theta_k,\xi))) &= r\left(\lim_{k\to\infty} \mathbb{E}[F(\theta_k,\xi)]\right) \\
&= r\left(\mathbb{E}\left[\lim_{k\to\infty} F(\theta_k,\xi)\right]\right) \\
&= r\left(\mathbb{E}[F(\theta,\xi)]\right) \\
&= r(p(\theta))
\end{aligned}
\tag{2.18}
$$

which implies that $\varphi$ is continuous.

Next we show uniform convergence of $r(p_I(\theta))$ to $r(p(\theta))$ as $I \to \infty$. Consider any point $\theta_0 \in \mathcal{C}$. Consider any $\varepsilon > 0$. There exists $\widehat{\delta} > 0$ such that

$$|r(x) - r(y)| \leq \varepsilon \text{ for all } x, y \in B_{\tilde{\delta}}(p(\theta_0)) \text{ such that } \|x - y\|_{\mathcal{X}} < \widehat{\delta}. \tag{2.19}$$

Note that

$$|r(p_I(\theta)) - r(p(\theta))| \leq |r(p_I(\theta)) - r(p_I(\theta_0))| + |r(p_I(\theta_0)) - r(p(\theta_0))| + |r(p(\theta_0)) - r(p(\theta))|$$

We bound each term on the right, and then we address the choice of $\theta_0$.

Let $\{\delta_k\}_{k=1}^{\infty} \subset (0,1)$ be a decreasing sequence such that $\lim_{k\to\infty} \delta_k = 0$. Consider the sequence of balls $B_k := B_k(\theta_0) = \{\theta \in \mathcal{C} : \|\theta - \theta_0\|_{\mathcal{C}} < \delta_k\}$. Define

$$d_k(\xi) := \sup_{\theta \in B_k} \|F(\theta, \xi) - F(\theta_0, \xi)\|_{\mathcal{X}}.$$

By Assumption A.1, $d_k(\xi) \to 0$ for almost every $\xi \in \Xi$ as $k \to \infty$. By Assumption A.2, $d_k(\xi) \leq 2K(\xi)$ for all $k$ and, by the DCT,

$$\mathbb{E}(d_k(\xi)) \to 0 \text{ as } k \to \infty \tag{2.20}$$

For all $\theta \in B_k$, it holds that

$$\begin{aligned} \|p_I(\theta) - p_I(\theta_0)\|_{\mathcal{X}} &\leq \frac{1}{I}\sum_{i=1}^{I} \|F(\theta, \xi^i) - F(\theta_0, \xi^i)\|_{\mathcal{X}} \\ &\leq \frac{1}{I}\sum_{i=1}^{I} \sup_{\theta \in B_k} \|F(\theta, \xi^i) - F(\theta_0, \xi^i)\|_{\mathcal{X}} \\ &= \frac{1}{I}\sum_{i=1}^{I} d_k(\xi^i), \end{aligned}$$

which implies

$$\sup_{\theta \in B_k} \|p_I(\theta) - p_I(\theta_0)\|_{\mathcal{X}} \leq \frac{1}{I}\sum_{i=1}^{I} d_k(\xi^i). \tag{2.21}$$

Since $r$ is continuous on $\mathcal{X} \subset \mathbb{R}^n$ by A.6, it follows from [38] that $r$ is locally uniformly continuous. Thus, there exists some ball $B_{\tilde{\delta}}(p(\theta_0)) := \{x \in \mathcal{X} : \|x - p(\theta_0)\|_{\mathcal{X}} < \tilde{\delta}\}$ such that $r$ is uniformly continuous on $B_{\tilde{\delta}}(p(\theta_0))$.

It follows from (2.20) that there exists $k_1 \in \mathbb{N}$ such that $\mathbb{E}(d_k(\xi)) < (\widehat{\delta}/2) \wedge (\tilde{\delta}/4)$ for all $k \geq k_1$. By the SLLN, there exists $\Omega_1^* \subset \Omega$ with $\mathbb{P}(\Omega_1^*) = 1$ such that for any $\omega \in \Omega_1^*$, there exists $I_1(\omega) \in \mathbb{N}$ such that for all $I \geq I_1(\omega)$ it holds that

$$\left| \frac{1}{I} \sum_{i=1}^{I} d_{k_1}(\xi^i) - \mathbb{E}(d_{k_1}(\xi)) \right| \quad < \quad (\widehat{\delta}/2) \wedge (\tilde{\delta}/4).$$

Thus, it follows from (2.21) that for all $I \geq I_1(\omega)$ it holds that

$$\begin{aligned} \sup_{\theta \in B_{k_1}} \|p_I(\theta) - p_I(\theta_0)\|_{\mathcal{X}} &\leq \frac{1}{I} \sum_{i=1}^{I} d_{k_1}(\xi^i) \\ &< \mathbb{E}(d_{k_1}(\xi)) + (\widehat{\delta}/2) \wedge (\tilde{\delta}/4) \\ &< \widehat{\delta} \wedge (\tilde{\delta}/2). \end{aligned}$$

Since $p_I(\theta_0) \to p(\theta_0)$ w.p.1 by the SLLN, there exists $\Omega_2^* \subset \Omega$ with $\mathbb{P}(\Omega_2^*) = 1$ such that for any $\omega \in \Omega_2^*$, there exists $I_2(\omega)$ such that for all $I \geq I_2(\omega)$ it holds that

$$\|p_I(\theta_0) - p(\theta_0)\|_{\mathcal{X}} \quad < \quad \widehat{\delta} \wedge (\tilde{\delta}/2).$$

Thus, for the given $\omega \in \Omega_2^*$, it holds that $p_I(\theta_0) \in B_{\tilde{\delta}}(p(\theta_0))$ for all $I \geq I_2(\omega)$.

Let $\Omega_3^* := \Omega_1^* \cap \Omega_2^*$. Then, $\mathbb{P}(\Omega_3^*) = 1$. Thus, for any $\omega \in \Omega_3^*$ and all $I \geq I_3(\omega) := \max\{I_1(\omega), I_2(\omega)\}$, it holds for all $\theta \in B_{k_1}$ that

$$\|p_I(\theta) - p(\theta_0)\|_{\mathcal{X}} \quad \leq \quad \|p_I(\theta) - p_I(\theta_0)\|_{\mathcal{X}} + \|p_I(\theta_0) - p(\theta_0)\|_{\mathcal{X}} \quad < \quad \tilde{\delta}.$$

Thus, for the given $\omega \in \Omega_3^*$, it holds that $p_I(\theta) \in B_{\tilde{\delta}}(p(\theta_0))$ for all $I \geq I_3(\omega)$ and all $\theta \in B_{k_1}$. To summarize, for all $\omega \in \Omega_3^*$, there exists $I_3(\omega)$ such that for all $I \geq I_3(\omega)$ and all $\theta \in B_{k_1}$,

$$p_I(\theta_0), p_I(\theta) \in B_{\tilde{\delta}}(p(\theta_0)) \text{ and } \|p_I(\theta) - p_I(\theta_0)\|_{\mathcal{X}} \quad < \quad \widehat{\delta}.$$

Hence, it follows from (2.19) that for all $k \geq k_1$,

$$\sup_{\theta \in B_k} |r(p_I(\theta)) - r(p_I(\theta_0))| \quad \leq \quad \sup_{\theta \in B_{k_1}} |r(p_I(\theta)) - r(p_I(\theta_0))| \quad \leq \quad \varepsilon$$

and

$$|r(p_I(\theta_0)) - r(p(\theta_0))| \quad \leq \quad \varepsilon$$

Since $\varphi(\theta) := r(p(\theta))$ is continuous, there exists $k_2 \in \mathbb{N}$ such that for all $k \geq k_2$,

$$\sup_{\theta \in B_k} |r(p(\theta)) - r(p(\theta_0))| \leq \sup_{\theta \in B_{k_2}} |r(p(\theta)) - r(p(\theta_0))| \leq \varepsilon.$$

In summary, for each $\theta_0 \in \mathcal{C}$, there exists $B(\theta_0) := B_{k'}$ where $k' := \max\{k_1, k_2\}$ and $\Omega^*(\theta_0) := \Omega_3^*$ with $\mathbb{P}(\Omega^*(\theta_0)) = 1$ such that for all $\omega \in \Omega^*(\theta_0)$, there exists $I(\omega, \theta_0) := I_3(\omega) \in \mathbb{N}$ such that

$$\begin{aligned} \sup_{\theta \in B(\theta_0)} |r(p_I(\theta)) - r(p_I(\theta_0))| &\leq \varepsilon \\ |r(p_I(\theta_0)) - r(p(\theta_0))| &\leq \varepsilon \\ \sup_{\theta \in B(\theta_0)} |r(p(\theta)) - r(p(\theta_0))| &\leq \varepsilon \end{aligned}$$

for all $I \geq I(\omega, \theta_0)$. Note that $\mathcal{C} \subset \bigcup_{\theta \in \mathcal{C}} B(\theta)$. Since $\mathcal{C}$ is compact, there is a finite number of points $\theta_1, \theta_2, \ldots, \theta_m \in \mathcal{C}$ such that $\mathcal{C} \subset \bigcup_{i=1}^m B(\theta_i)$. Let $\Omega^* := \bigcap_{i=1}^m \Omega^*(\theta_i)$. Then $\mathbb{P}(\Omega^*) = 1$. For any $\omega \in \Omega^*$, let $I(\omega) := \max_{i=1,\ldots,m}\{I(\omega, \theta_i)\}$.

Consider any $\theta \in \mathcal{C}$. Then there exists $B(\theta_i)$ such that $\theta \in B(\theta_i)$. Thus, for all $\omega \in \Omega^*$ and $I \geq I(\omega)$ it holds that

$$\begin{aligned} |r(p_I(\theta)) - r(p_I(\theta_i))| &\leq \varepsilon \\ |r(p_I(\theta_i)) - r(p(\theta_i))| &\leq \varepsilon \\ |r(p(\theta)) - r(p(\theta_i))| &\leq \varepsilon \end{aligned}$$

Therefore

$$\sup_{\theta \in \mathcal{C}} |r(p_I(\theta)) - r(p(\theta))| \leq 3\varepsilon.$$

□

**Remark 2.1.** *Theorem 2.1 holds if $r : \mathcal{X} \mapsto \mathbb{R}^{m \times d}$ for any $m, d \in \mathbb{N}$ is a continuous function.*

**Theorem 2.2.** *Let $\mathcal{C}$ be a nonempty compact subset of $\mathbb{R}^d$. Assume A.2, A.3, A.4, A.7, and A.10. Then,*

*(1) $p(\theta) \in \mathcal{X}$,*

*(2)* $\varphi(\theta) := r(p(\theta))$ *is continuously differentiable,*

*(3) for each* $\theta \in \mathcal{C}$*, it holds that*

$$\nabla\varphi(\theta) \quad = \quad \nabla p(\theta)^\intercal \nabla r(p(\theta)) \quad = \quad \mathbb{E}(\nabla F(\theta,\xi))^\intercal \nabla r(p(\theta)),$$

*(4) w.p. 1,*

$$\sup_{\theta\in\mathcal{C}} \|\nabla p_I(\theta)^\intercal \nabla r(p_I(\theta)) - \nabla p(\theta)^\intercal \nabla r(p(\theta))\| \to 0 \text{ as } I \to \infty.$$

**Proof**: It follows from Theorem 2.1 that $p(\theta) \in \mathcal{X}$. We first show that $\varphi(\theta)$ is differentiable at all $\theta \in \mathcal{C}$. By A.2, the convexity of $\|\|_{\mathcal{X}}$, and Jensen's inequality that $\|p(\theta)\|_{\mathcal{X}} = \|\mathbb{E}(F_n(\theta,\xi))\|_{\mathcal{X}} \le \mathbb{E}[\|F_n(\theta,\xi)\|_{\mathcal{X}}] < \mathbb{E}_\xi[K(\xi)] < \infty$ is finite-valued for all $\theta \in \mathcal{C}$. By Theorem 7.44 in [58], it follows from assumptions A.3 and A.4 that $p(\theta)$ is differentiable and $\nabla p(\theta) = \mathbb{E}(\nabla F(\theta,\xi))$ for each $\theta \in \mathcal{C}$. By assumption A.7 and the chain rule, we have $\varphi(\theta)$ is differentiable and result (3) holds.

Next, we show the continuity of $\nabla p(\theta)$ and $\nabla\varphi(\theta)$. Consider any $\theta \in \mathcal{C}$ and a sequence $\{\theta_k\} \subset \mathcal{C}$ such that $\theta_k \to \theta$ as $k \to \infty$. By assumption A.4, it follows that $\|\nabla F(\theta,\xi)\|_{m\times d} \le L(\xi)$ for all $\theta \in \mathcal{C}$. By the DCT and A.3,

$$\lim_{k\to\infty} \nabla p(\theta_k) \quad = \quad \lim_{k\to\infty} \mathbb{E}(\nabla F(\theta_k,\xi)) \quad = \quad \mathbb{E}(\nabla F(\lim_{k\to\infty}\theta_k,\xi)) \quad = \quad \nabla p(\theta),$$

which shows the continuity of $\nabla p(\theta)$. By the continuity of $p(\theta)$ (See (2.18) in the proof of Theorem 2.1), A.7 and result (3),

$$\lim_{k\to\infty} \nabla\varphi(\theta_k) \quad = \quad \lim_{k\to\infty} \nabla p(\theta_k)^\intercal \lim_{k\to\infty} \nabla r(p(\theta_k)) \quad = \quad \nabla p(\theta)^\intercal \nabla r(p(\lim_{k\to\infty}\theta_k)) \quad = \quad \nabla\varphi(\theta),$$

which implies (2).

By A.7, $\nabla r$ is continuous. Since $p(\theta)$ is continuous, we have $\|\nabla r(p(\theta))\|_{m\times 1}$ is continuous on $\mathcal{C}$ and thus attains maximum due to the compactness of $\mathcal{C}$. Define

$$C_1 \quad := \quad \sup_{\theta\in\mathcal{C}} \|\nabla r(p(\theta))\|_{m\times 1} \quad < \quad \infty,$$

and

$$C_2 \quad := \quad \sup_{\theta\in\mathcal{C}} \|\nabla p(\theta)\|_{m\times d} \quad = \quad \sup_{\theta\in\mathcal{C}} \|\mathbb{E}(\nabla F(\theta,\xi))\|_{m\times d} \quad \le \quad \mathbb{E}(L(\xi)) \quad < \quad \infty.$$

Choose any $\varepsilon > 0$. Since $\nabla r$ is a continuous function under A.7. By replacing with $r$ replaced by $\nabla r$, it follows from A.2, A.3, Theorem 2.1 and Remark 2.1 that, there exists $\Omega_1$ with $\mathbb{P}(\Omega_1) = 1$ such that for every $\omega \in \Omega_1$, there exists $I_1(\omega) \in \mathbb{N}$ such that for all $I \geq I_1(\omega)$ and $\theta \in \mathcal{C}$,

$$\|\nabla r(p_I(\theta)) - \nabla r(p(\theta))\|_{m\times 1} \quad < \quad (\varepsilon/C_2) \wedge 1.$$

Thus, it follows that

$$\sup_{\theta\in\mathcal{C}} \|\nabla r(p_I(\theta))\|_{m\times 1} \quad \leq \quad \sup_{\theta\in\mathcal{C}} \|\nabla r(p(\theta))\|_{m\times 1} + 1 = C_1 + 1.$$

By A.3, A.4 and Theorem 2.1 again, there exists $\Omega_2$ with $\mathbb{P}(\Omega_2) = 1$ such that for every $\omega \in \Omega_2$ , there exists $I_2(\omega) \in \mathbb{N}$ such that for all $I \geq I_2(\omega)$,

$$\sup_{\theta\in\mathcal{C}} \|\nabla p_I(\theta)) - \nabla p(\theta)\|_{m\times d} \quad < \quad \varepsilon/(C_1 + 1).$$

Let $\Omega = \Omega_1 \cap \Omega_2$. It follows that for each $\omega \in \Omega$, where $\mathbb{P}(\Omega) = 1$, letting $I^*(\omega) := \max\{I_1(\omega), I_2(\omega)\}$ gives that for all $I \geq I^*(\omega)$

$$\begin{aligned}
& \sup_{\theta\in\mathcal{C}} \|\nabla p_I(\theta)^\intercal \nabla r(p_I(\theta)) - \nabla p(\theta)^\intercal r(p(\theta))\| \\
\leq\ & \sup_{\theta\in\mathcal{C}} \|\nabla p_I(\theta)^\intercal \nabla r(p_I(\theta)) - \nabla p(\theta)^\intercal \nabla r(p_I(\theta))\| \\
+\ & \sup_{\theta\in\mathcal{C}} \|\nabla p(\theta)^\intercal \nabla r(p_I(\theta)) - \nabla p(\theta)^\intercal \nabla r(p(\theta))\| \\
\leq\ & \sup_{\theta\in\mathcal{C}} \|\nabla r(p_I(\theta))\|_{m\times 1} \sup_{\theta\in\mathcal{C}} \|\nabla p_I(\theta)) - \nabla p(\theta)\|_{m\times d} \\
+\ & \sup_{\theta\in\mathcal{C}} \|\nabla p(\theta)\|_{m\times d} \sup_{\theta\in\mathcal{C}} \|\nabla r(p_I(\theta)) - \nabla r(p(\theta))\|_{m\times 1} \\
<\ & 2\varepsilon,
\end{aligned}$$

which prove (4). □

**Lemma 2.1.** *Consider any nonempty compact set $\mathcal{C} \subset \mathbb{R}^d$. Assume A.1, A.5, A.7 and A.10. Then, the following holds:*

*(1) Assume further A.8 and A.9. Then,* $\sup_{\theta_1,\theta_2\in\mathcal{C}} \sigma_1(\theta_1, \theta_2, |N|, \mathbf{I}) \to 0$ *as* $|N| \to |\mathcal{N}|$.

*(2)* $\sup_{\theta_1,\theta_2\in\mathcal{C}} \sigma_2(\theta_1, \theta_2, |N|, \mathbf{I}) \to 0$ *as* $|I_n| \to \infty$ *for each* $n \in \mathcal{N}$.

*(3) W.p. 1,* $\sup_{\theta_1,\theta_2\in\mathcal{C}} \widehat{\sigma}_1(\theta_1, \theta_2, N, \mathcal{I}) \to 0$ *as* $|N| \to |\mathcal{N}|$ *and* $|I_n| \to \infty$ *for all* $n \in \mathcal{N}$.

*(4) W.p. 1,* $\sup_{\theta_1,\theta_2\in\mathcal{C}} \widehat{\sigma}_2(\theta_1,\theta_2,N,\mathcal{I}) \to 0$ *as* $|N| \to |\mathcal{N}|$ *and* $|I_n| \to \infty$ *for all* $n \in \mathcal{N}$*.*

**Proof:**

(1) Define

$$\alpha(\theta_1,\theta_2) \quad := \quad \sum_{n\in\mathcal{N}} \left( \mathbb{E}[\Delta r_{I_n}(\theta_1,\theta_2)] - \frac{\sum_{n\in\mathcal{N}} \mathbb{E}[\Delta r_{I_n}(\theta_1,\theta_2)]}{|\mathcal{N}|} \right)^2.$$

Consider any customer $n \in \mathcal{N}$. We first show that,

$$\mathbb{E}\left( \sup_{|I_n|\in\mathbb{N}} \sup_{\theta\in\mathcal{C}} |r(p_{I_n}(\theta))| \right) \quad < \quad \infty.$$

We suppress $n$ for notational convenience. It follows from Jensen's inequality and condition (A.8) that

$$\begin{aligned}
\sup_{\theta\in\mathcal{C}} |r(p_I(\theta))| \quad &= \quad \sup_{\theta\in\mathcal{C}} \left| -r\left( \frac{1}{|I|} \sum_{i=1}^{|I|} F(\theta,\xi^i) \right) \right| \quad \le \quad \sup_{\theta\in\mathcal{C}} \left( \frac{1}{|I|} \sum_{i=1}^{|I|} |r\left(F(\theta,\xi^i)\right)| \right) \\
&\le \quad \frac{1}{|I|} \sum_{i=1}^{|I|} \sup_{\theta\in\mathcal{C}} |r\left(F(\theta,\xi^i)\right)|.
\end{aligned}$$

Let $\mu = \mathbb{E}\left(\sup_{\theta\in\mathcal{C}} |r\left(F(\theta,\xi)\right)|\right)$, where $\xi$ is identically distributed to $\xi^i$, $i = 1,\ldots,|I|$. By condition (A.9), $\mu < \infty$. Define $Z_i = \sup_{\theta\in\mathcal{C}} |r\left(F(\theta,\xi^i)\right)| - \mu$ for $i = 1,\ldots,|I|$ and $Z = \sup_{\theta\in\mathcal{C}} |r\left(F(\theta,\xi)\right)| - \mu$. Thus, $Z_i, i = 1,\ldots,|I|$ are i.i.d. with $\mathbb{E}(\eta_i) = 0$. For $|I| = 1,2,\ldots$, define

$$X_I \quad := \quad \frac{1}{|I|} \sum_{i=1}^{|I|} Z_i.$$

Let

$$C_1(\omega) \quad := \quad \sup_{|I|\in\mathbb{N}} \left( \frac{1}{|I|} \sum_{i=1}^{|I|} \sup_{\theta\in\mathcal{C}} |r\left(F(\theta,\xi^i)\right)| \right) \quad = \quad \sup_{|I|\in\mathbb{N}} |X_I(\omega)| + \mu$$

for $\omega \in \Omega$. It follows that

$$\begin{aligned}
\mathbb{E}(C_1^2) \quad &= \quad \mathbb{E}\left( \sup_{|I|\in\mathbb{N}} |X_I(\omega)| + \mu \right)^2 \\
&\le \quad 2\mathbb{E}\left( \sup_{|I|\in\mathbb{N}} |X_I(\omega)| \right)^2 + 2\mu^2 \\
&\le \quad 2\int_0^\infty \mathbb{P}\left( \sup_{|I|\in\mathbb{N}} |X_I| > \sqrt{x} \right) dx + 2\mu^2
\end{aligned}$$

$$
\begin{aligned}
&= 2\int_0^1 \mathbb{P}\left(\sup_{|I|\in\mathbb{N}} |X_I| > \sqrt{x}\right) dx + 2\int_1^\infty \mathbb{P}\left(\sup_{|I|\in\mathbb{N}} |X_I| > \sqrt{x}\right) dx + 2\mu^2 \\
&\leq 2\int_1^\infty \mathbb{P}\left(\omega\in\Omega : \sup_{|I|\in\mathbb{N}} |X_I(\omega)| > \sqrt{x}\right) dx + 2 + 2\mu^2 \\
&\leq 2\int_1^\infty \mathbb{P}\left(\omega\in\Omega : \bigcup_{|I|\in\mathbb{N}} \{|X_I(\omega)| > \sqrt{x}\}\right) dx + 2 + 2\mu^2 \\
&\leq 2\int_1^\infty \sum_{|I|=1}^\infty \mathbb{P}(X_I^4 > x^2) dx + 2 + 2\mu^2 \\
&\leq 2\int_1^\infty \sum_{|I|=1}^\infty \frac{\mathbb{E}(X_I^4)}{x^2} dx + 2 + 2\mu^2,
\end{aligned}
$$

and

$$
\mathbb{E}(X_I^4) = \mathbb{E}\left(\frac{1}{|I|}\sum_{i=1}^{|I|} Z_i\right)^4 = \frac{\mathbb{E}(Z^4)}{|I|^3} + \frac{|I|^2-|I|}{|I|^4}(\mathbb{E}(Z^2))^2.
$$

By condition (A.9), $\mathbb{E}\left(\sup_{\theta\in\mathcal{C}}[r(F(\theta,\xi))]^4\right) < \infty$, thus,

$$
\begin{aligned}
\mathbb{E}(Z^4) &= \mathbb{E}\left(\sup_{\theta\in\mathcal{C}} |r\left(F(\theta,\xi)\right)| - \mu\right)^4 \\
&\leq 8\mathbb{E}\left(\sup_{\theta\in\mathcal{C}} |r\left(F(\theta,\xi)\right)|\right)^4 + 8\mu^4 \\
&= 8\mathbb{E}\left(\sup_{\theta\in\mathcal{C}} [r(F(\theta,\xi)]^4\right) + 8\mu^4 < \infty,
\end{aligned}
$$

which also implies $\mathbb{E}(Z^2) < \infty$. Since

$$
C_2 := \mathbb{E}(Z^4)\sum_{|I|=1}^\infty \frac{1}{|I|^3} + (\mathbb{E}(Z^2))^2 \sum_{|I|=1}^\infty \frac{|I|^2-|I|}{|I|^4} < \infty,
$$

it follows that

$$
\begin{aligned}
\mathbb{E}(C_1^2) &\leq 2 + 2\int_1^\infty \frac{\mathbb{E}(Z^4)\sum_{|I|=1}^\infty \frac{1}{|I|^3} + (\mathbb{E}(Z^2))^2 \sum_{|I|=1}^\infty \frac{|I|^2-|I|}{|I|^4}}{x^2} dx + 2\mu^2 \\
&= 2 + 2C_2 \int_1^\infty \frac{1}{x^2} dx + 2\mu^2 \qquad (2.22) \\
&= 2 + 2\mu^2 + 2C_2 < \infty,
\end{aligned}
$$

which implies,

$$
\mathbb{E}\left(\sup_{|I|\in\mathbb{N}} \sup_{\theta\in\mathcal{C}} |r(p_I(\theta))|\right) \leq \mathbb{E}(C_1) < \infty, \qquad (2.23)
$$

which is the desired result.

We next show that $\varphi_n(\theta) := \mathbb{E}[r(p_{I_n}(\theta))]$ is continuous on $\theta \in \mathcal{C}$. Choose any $\theta \in \mathcal{C}$ and a sequence $\{\theta_k\}_{k=1}^{\infty} \subseteq \mathcal{C}$ such that $\theta_k \to \theta$ as $k \to \infty$. It follows from (2.23), the continuity of $r$ and $F$ and the DCT that

$$\begin{aligned}\lim_{k\to\infty} \varphi_n(\theta_k) &= \lim_{k\to\infty} \mathbb{E}[r(p_{I_n}(\theta_k))] \\ &= \mathbb{E}\left(\lim_{k\to\infty} r(p_{I_n}(\theta_k))\right) \\ &= \mathbb{E}(r(p_{I_n}(\theta))).\end{aligned}$$

which implies that $\varphi_n$ is continuous.

Thus, $\Delta r_{I_n}(\theta_1, \theta_2)$ is continuous and $\alpha(\theta_1, \theta_2)$ is continuous on $\mathcal{C} \times \mathcal{C}$, and since $\mathcal{C}$ is compact, $\alpha(\theta_1, \theta_2)$ attains its maximality, i.e., $\sup_{\theta_1,\theta_2 \in \mathcal{C}} \alpha(\theta_1, \theta_2) < \infty$. Thus,

$$\sup_{\theta_1,\theta_2\in\mathcal{C}} \sigma_1^2(\theta_1, \theta_2, |N|, \mathbf{I}) = \frac{|\mathcal{N}| - |N|}{|N||\mathcal{N}|(|\mathcal{N}| - 1)} \sup_{\theta_1,\theta_2\in\mathcal{C}} \alpha(\theta_1, \theta_2),$$

which further implies that the result holds.

(2) For each $n \in \mathcal{N}$, define $Y_{I_n}(\theta) := r\left(p_{I_n}(\theta)\right)$. Note that

$$\begin{aligned}&\sigma_2^2(\theta_1, \theta_2, |N|, \mathbf{I}) \\ =& \frac{1}{|N||\mathcal{N}|} \sum_{n\in\mathcal{N}} \text{Var}[\Delta r_{I_n}(\theta_1, \theta_2)] \\ \leq& \frac{1}{|N||\mathcal{N}|} \sum_{n\in\mathcal{N}} 2\left(\mathbb{E}\left[Y_{I_n}(\theta_1) - \mathbb{E}\left(Y_{I_n}(\theta_1)\right)\right]^2 + \mathbb{E}\left[Y_{I_n}(\theta_2) - \mathbb{E}\left(Y_{I_n}(\theta_2)\right)\right]^2\right) \\ \leq& \frac{4}{|N||\mathcal{N}|} \sum_{n\in\mathcal{N}} \sup_{\theta\in\mathcal{C}} \left(\mathbb{E}\left[Y_{I_n}(\theta) - \mathbb{E}\left(Y_{I_n}(\theta)\right)\right]^2\right)\end{aligned}$$

Since $\mathcal{N}$ is a finite set, it suffices to show for each $n \in \mathcal{N}$ that

$$\lim_{|I_n|\to\infty} \sup_{\theta\in\mathcal{C}} \mathbb{E}\left[Y_{I_n}(\theta) - \mathbb{E}\left(Y_{I_n}(\theta)\right)\right]^2 = 0.$$

We suppress $n$ for notational simplicity. It follows from (2.22) that,

$$\mathbb{E}\left(\sup_{|I|\in\mathbb{N}} \left(\sup_{\theta\in\mathcal{C}} |r(p_I(\theta))|\right)^2\right) \leq \mathbb{E}(C_1^2) < \infty. \tag{2.24}$$

Define $\tilde{C} = \sup_{|I|\in\mathbb{N}} \left(\sup_{\theta\in\mathcal{C}} |Y_I(\theta) - \mathbb{E}\left(Y_I(\theta)\right)|\right)^2$. It follows from (2.24) that

$$
\begin{aligned}
\mathbb{E}(\tilde{C}) &= \mathbb{E}\left(\sup_{|I|\in\mathbb{N}} \left(\sup_{\theta\in\mathcal{C}} |Y_I(\theta) - \mathbb{E}\left(Y_I(\theta)\right)|\right)^2\right) \\
&\leq 2\mathbb{E}\left(\sup_{|I|\in\mathbb{N}} \left(\sup_{\theta\in\mathcal{C}} |Y_I(\theta)|\right)^2 + \sup_{|I|\in\mathbb{N}} \left(\sup_{\theta\in\mathcal{C}} \mathbb{E}|Y_I(\theta)|\right)^2\right) \\
&\leq 2\mathbb{E}\left(\sup_{|I|\in\mathbb{N}} \left(\sup_{\theta\in\mathcal{C}} |Y_I(\theta)|\right)^2 + \sup_{|I|\in\mathbb{N}} \left(\mathbb{E}\sup_{\theta\in\mathcal{C}} |Y_I(\theta)|\right)^2\right) \\
&\leq 2\mathbb{E}\left(\sup_{|I|\in\mathbb{N}} \left(\sup_{\theta\in\mathcal{C}} |Y_I(\theta)|\right)^2\right) + 2\left(\mathbb{E}\sup_{|I|\in\mathbb{N}}\sup_{\theta\in\mathcal{C}} |Y_I(\theta)|\right)^2 \\
&= 2\mathbb{E}(C_1^2) + 2(\mathbb{E}(C_1))^2 \quad < \quad \infty.
\end{aligned}
$$

Note that

$$
\sup_{\theta\in\mathcal{C}} \mathbb{E}\left[Y_I(\theta) - \mathbb{E}\left(Y_I(\theta)\right)\right]^2 \quad \leq \quad \mathbb{E}\left[\sup_{\theta\in\mathcal{C}} |Y_I(\theta) - \mathbb{E}\left(Y_I(\theta)\right)|\right]^2. \tag{2.25}
$$

By the DCT, it follows from (2.25) and $\mathbb{E}(\tilde{C}) < \infty$ that

$$
\begin{aligned}
& \lim_{|I|\to\infty} \sup_{\theta\in\mathcal{C}} \mathbb{E}\left[Y_I(\theta) - \mathbb{E}\left(Y_I(\theta)\right)\right]^2 \\
\leq\; & \lim_{|I|\to\infty} \mathbb{E}\left[\sup_{\theta\in\mathcal{C}} |Y_I(\theta) - \mathbb{E}\left(Y_I(\theta)\right)|\right]^2 \\
=\; & \mathbb{E}\left[\lim_{|I|\to\infty} \sup_{\theta\in\mathcal{C}} |Y_I(\theta) - \mathbb{E}\left(Y_I(\theta)\right)|\right]^2 \\
=\; & \mathbb{E}\left[\lim_{|I|\to\infty} \sup_{\theta\in\mathcal{C}} |Y_I(\theta) - r(p(\theta)) + r(p(\theta)) - \mathbb{E}\left(Y_I(\theta)\right)|\right]^2 \\
\leq\; & \mathbb{E}\left[\lim_{|I|\to\infty} \sup_{\theta\in\mathcal{C}} |Y_I(\theta) - r(p(\theta)| + \lim_{|I|\to\infty} \sup_{\theta\in\mathcal{C}} |r(p(\theta)) - \mathbb{E}\left(Y_I(\theta)\right)|\right]^2.
\end{aligned}
$$

By Theorem 2.1, it follows w.p.1 that $\lim_{|I|\to\infty} \sup_{\theta\in\mathcal{C}} |Y_I(\theta) - r(p(\theta))| = 0$. It also follows from (2.23), the DCT and Theorem 2.1 that

$$
\begin{aligned}
\lim_{|I|\to\infty} \sup_{\theta\in\mathcal{C}} |r(p(\theta)) - \mathbb{E}\left(Y_I(\theta)\right)| &= \lim_{|I|\to\infty} \sup_{\theta\in\mathcal{C}} |\mathbb{E}(r(p_I(\theta)) - r(p(\theta)))| \\
&\leq \lim_{|I|\to\infty} \mathbb{E}\left(\sup_{\theta\in\mathcal{C}} |r(p_I(\theta)) - r(p(\theta))|\right) \\
&= \mathbb{E}\left(\lim_{|I|\to\infty} \sup_{\theta\in\mathcal{C}} |r(p_I(\theta)) - r(p(\theta))|\right) = 0,
\end{aligned}
$$

which implies

$$
\lim_{|I|\to\infty} \sup_{\theta\in\mathcal{C}} \mathbb{E}\left[Y_I(\theta) - \mathbb{E}\left(Y_I(\theta)\right)\right]^2 = 0
$$

and this completes the proof. $\square$

(3) Consider any data sample $N \subset \mathcal{N}$ with $|N| \geq 2$ and integration sample $\mathcal{I}$. Since $\widehat{\sigma}_1(\theta_1, \theta_2, N, \mathcal{I}) \geq 0$ for all $\theta_1, \theta_2 \in \mathcal{C}$, it holds that

$$\left( \sup_{\theta_1, \theta_2 \in \mathcal{C}} \widehat{\sigma}_1(\theta_1, \theta_2, N, \mathcal{I}) \right)^2 = \sup_{\theta_1, \theta_2 \in \mathcal{C}} \widehat{\sigma}_1^2(\theta_1, \theta_2, N, \mathcal{I}).$$

Thus, it suffices to show that w.p. 1 $\sup_{\theta_1, \theta_2 \in \mathcal{C}} \widehat{\sigma}_1^2(\theta_1, \theta_2, N, \mathcal{I}) \to 0$ as $|N| \to |\mathcal{N}|$ and $|I_n| \to \infty$ for all $n \in \mathcal{N}$.

Consider any $n \in \mathcal{N}$. It follows from A.1, A.5, A.7 and Theorem 2.1 that $r(p_n(\theta))$ is continuous on $\mathcal{C}$. Then, $|r(p_n(\theta))|$ attains its maximum on $\mathcal{C}$ due to the compactness of $\mathcal{C}$ and $C_n := \sup_{\theta \in \mathcal{C}} |r(p_n(\theta))| + 1 < \infty$. Thus $C := \max_{n \in \mathcal{N}} C_n < \infty$. It also follows from Theorem 2.1 there exists $\Omega_n \subset \Omega$ with $\mathbb{P}(\Omega_n) = 1$ such that for any $\omega \in \Omega_n$, there exists $M_n(\omega) \in \mathbb{N}$ such that for all $|I_n| \geq M_n(\omega)$ it holds that

$$\sup_{\theta \in \mathcal{C}} |r(p_{I_n}(\theta)) - r(p_n(\theta))| \leq 1,$$

which implies that

$$\begin{aligned}
\sup_{\theta \in \mathcal{C}} |r(p_{I_n}(\theta))| &= \sup_{\theta \in \mathcal{C}} |r(p_{I_n}(\theta)) - r(p_n(\theta)) + r(p_n(\theta))| \\
&\leq \sup_{\theta \in \mathcal{C}} |r(p_{I_n}(\theta)) - r(p_n(\theta))| + \sup_{\theta \in \mathcal{C}} |r(p_n(\theta))| \\
&\leq C.
\end{aligned}$$

Let $\Omega^* := \cap_{n \in \mathcal{N}} \Omega_n$. Then, $\mathbb{P}(\Omega^*) = 1$. Thus, for any $\omega \in \Omega^*$,

$$\begin{aligned}
\sup_{\theta_1, \theta_2 \in \mathcal{C}} |\Delta r_{I_n}(\theta_1, \theta_2)| &= \sup_{\theta_1, \theta_2 \in \mathcal{C}} |r(p_{I_n}(\theta_1)) - r(p_{I_n}(\theta_2))| \\
&\leq 2 \sup_{\theta \in \mathcal{C}} |r(p_{I_n}(\theta))| \\
&\leq 2C,
\end{aligned}$$

holds for all $|I_n| \geq M(\omega) := \max_{n \in \mathcal{N}} M_n(\omega)$ and for all $n \in \mathcal{N}$. Thus,

$$\bar{\nu}^2(\theta_1, \theta_2 N, \mathcal{I}) = \frac{1}{|N| - 1} \sum_{n \in N} \left( \Delta r_{I_n}(\theta_1, \theta_2) - \frac{\sum_{n \in N} \Delta r_{I_n}(\theta_1, \theta_2)}{|N|} \right)^2$$

$$\leq \quad 16\frac{|N|}{|N|-1}C^2$$

holds for all $\theta_1, \theta_2 \in \mathcal{C}$, which implies that,

$$\sup_{\theta_1,\theta_2 \in \mathcal{C}} \widehat{\sigma}_1^2(\theta_1, \theta_2, N, \mathcal{I}) \quad = \quad \frac{|\mathcal{N}|-|N|}{|N||\mathcal{N}|}\bar{\nu}^2(\theta_1, \theta_2, N, \mathcal{I}) \quad \leq \quad 16C^2\frac{(|\mathcal{N}|-|N|)|N|}{|N||\mathcal{N}|(|N|-1)}.$$

Thus, $\sup_{\theta_1,\theta_2\in\mathcal{C}} \widehat{\sigma}_1^2(\theta_1, \theta_2, N, \mathcal{I}) \to 0$ as $|N| \to |\mathcal{N}|$ and $|I_n| \to \infty$ for all $n \in \mathcal{N}$, which completes the proof. □

(4) By the same argument in Result (3), it suffices to show that w.p. 1

$$\sup_{\theta_1,\theta_2 \in \mathcal{C}} \widehat{\sigma}_2^2(\theta_1, \theta_2, N, \mathcal{I}) \to 0$$

as $|N| \to |\mathcal{N}|$ and $|I_n| \to \infty$ for all $n \in \mathcal{N}$. Since $\mathcal{N}$ is finite, it further reduces to show that w.p. 1 $\frac{1}{|I_n|}\sup_{\theta_1,\theta_2\in\mathcal{C}} \nu_n^2(\theta_1, \theta_2, I_n) \to 0$ as $|I_n| \to \infty$ for each $n \in \mathcal{N}$. We suppress index $n$ and will use $I$ as both a set of integration samples (as its original definition) and the cardinality of that set for notational convenience.

It follows from Result (3) that $p(\theta)$ is continuous. Thus, $C_0 := \sup_{\theta\in\mathcal{C}} \|p(\theta)\|_{\mathcal{X}} < \infty$. It follows from A.7 that $\|\nabla r(p(\theta))\|_{m\times 1}$ is continuous on $\mathcal{C}$. Since $\mathcal{C}$ is compact, $\|\nabla r(p(\theta))\|_{m\times 1}$ attains maximum. Thus, $C_1 := \sup_{\theta\in\mathcal{C}} \|\nabla r(p(\theta)\|_{m\times 1} + 1 < \infty$. It follows from A.1 that $F(\theta,\xi)F(\theta,\xi)^\intercal$ is continuous on $\mathcal{C}$ for almost all $\xi \in \Xi$. It follows from A.5, the DCT and Theorem 2.1 that $\mathbb{E}[F(\theta,\xi)F(\theta,\xi)^\intercal]$ is continuous on $\mathcal{C}$. Thus, $\mathrm{Cov}[F(\theta,\xi), F(\theta,\xi)] = \mathbb{E}[F(\theta,\xi)F(\theta,\xi)^\intercal] - p(\theta)p(\theta)^\intercal$ is continuous. Thus, $\|\mathrm{Cov}[F(\theta,\xi), F(\theta,\xi)]\|_{m\times m}$ is continuous and attains its maximum. Thus, $C_2 := \sup_{\theta\in\mathcal{C}} \|\mathrm{Cov}[F(\theta,\xi), F(\theta,\xi)]\|_{m\times m} + 1 < \infty$.

We now show that w.p. 1 $\sup_{\theta\in\mathcal{C}} \|S^2(\theta)\|_{m\times m}$ is bounded for sufficient large $I$. Since $(\nabla r(\cdot))^2$ is continuous, it follows from A.1, A.5, A.7 and Theorem 2.1 that there exists $\Omega_1 \in \Omega$ with $\mathbb{P}(\Omega_1) = 1$ such that for every $\omega \in \Omega_1$ there exists $M_1(\omega) \in \mathbb{N}$ such that, for all $I \geq M_1(\omega)$, it holds that

$$\sup_{\theta\in\mathcal{C}} \|\nabla r(p_I(\theta)) - \nabla r(p(\theta))\|_{m\times 1} \quad < \quad 1,$$

which implies

$$\sup_{\theta\in\mathcal{C}} \|\nabla r(p_I(\theta))\|_{m\times 1} \quad = \quad \sup_{\theta\in\mathcal{C}} \|\nabla r(p_I(\theta)) - \nabla r(p(\theta)) + \nabla r(p(\theta))\|_{m\times 1}$$

$$
\begin{aligned}
&\le \sup_{\theta\in\mathcal{C}} \|\nabla r(p_I(\theta)) - \nabla r(p(\theta))\|_{m\times 1} + \sup_{\theta\in\mathcal{C}} \|\nabla r(p(\theta))\|_{m\times 1} \\
&\le C_1.
\end{aligned}
$$

Note that A.5 implies A.2. Thus, it follows from A.1 and Theorem 2.1 that there exists $\Omega_2 \subset \Omega$ with $\mathbb{P}(\Omega_2) = 1$ such that for every $\omega \in \Omega_2$ there exists $M_2(\omega) \in \mathbb{N}$ such that for all $I \ge M_2(\omega)$,

$$
\sup_{\theta\in\mathcal{C}} \|p_I(\theta) - p(\theta)\|_{\mathcal{X}} \le (1/(4C_0)) \wedge (3/C_0) \wedge (C_0/3),
$$

which yields

$$
\sup_{\theta\in\mathcal{C}} \|p_I(\theta)\|_{\mathcal{X}} \le \sup_{\theta\in\mathcal{C}} \|p_I(\theta) - p(\theta)\|_{\mathcal{X}} + \sup_{\theta\in\mathcal{C}} \|p(\theta)\|_{\mathcal{X}} \le C_0 + C_0/3 = 4C_0/3.
$$

Since $F(\theta,\xi)F(\theta,\xi)^\intercal$ is continuous and A.5 holds, Theorem 2.1 ensures that there exists $\Omega_3 \subset \Omega$ with $\mathbb{P}(\Omega_3) = 1$ such that for every $\omega \in \Omega_3$ there exists $M_3(\omega) \in \mathbb{N}$ such that for all $I \ge M_2(\omega)$,

$$
\sup_{\theta\in\mathcal{C}} \left\| \frac{1}{I}\sum_{i=1}^{I} F(\theta,\xi^i)F(\theta,\xi^i)^\intercal - \mathbb{E}[F(\theta,\xi)F(\theta,\xi)^\intercal] \right\|_{m\times m} \le 1/3.
$$

Consider any $\varepsilon > 0$. Let $\Omega^* := \Omega_1 \cap \Omega_2 \cap \Omega_3$. Then, $\mathbb{P}(\Omega^*) = 1$. Thus, for any $\omega \in \Omega^*$ and all $I \ge M(\omega) := \max\{M_1(\omega), M_2(\omega), M_3(\omega)\}$,

$$
\begin{aligned}
& \sup_{\theta\in\mathcal{C}} \left\| \frac{I-1}{I} S^2(\theta) - \mathrm{Cov}[F(\theta,\xi), F(\theta,\xi)] \right\|_{m\times m} \\
= & \sup_{\theta\in\mathcal{C}} \left\| \frac{1}{I}\sum_{i=1}^{I} F(\theta,\xi^i)F(\theta,\xi^i)^\intercal - p_I(\theta)p_I(\theta)^\intercal - \mathrm{Cov}[F(\theta,\xi), F(\theta,\xi)] \right\|_{m\times m} \\
= & \sup_{\theta,\theta\in\mathcal{C}} \left\| \frac{1}{I}\sum_{i=1}^{I} F(\theta,\xi^i)F(\theta,\xi^i)^\intercal - \mathbb{E}[F(\theta,\xi)F(\theta,\xi)^\intercal] - [p_I(\theta_1)p_I(\theta)^\intercal - p(\theta)p(\theta_2)^\intercal] \right\|_{m\times m} \\
\le & \sup_{\theta\in\mathcal{C}} \left\| \frac{1}{I}\sum_{i=1}^{I} F(\theta,\xi^i)F(\theta,\xi^i)^\intercal - \mathbb{E}[F(\theta,\xi)F(\theta,\xi)^\intercal] \right\|_{m\times m} \\
& + \sup_{\theta\in\mathcal{C}} \|p_I(\theta)p_I(\theta)^\intercal - p_I(\theta)p(\theta)^\intercal\|_{m\times m} + \sup_{\theta\in\mathcal{C}} \|p_I(\theta)p(\theta)^\intercal - p(\theta)p(\theta)^\intercal]\|_{m\times m} \\
\le & 1/3 + \sup_{\theta\in\mathcal{C}} \|p_I(\theta)\|_{\mathcal{X}} \sup_{\theta\in\mathcal{C}} \|p_I(\theta) - p(\theta)\|_{\mathcal{X}} + \sup_{\theta\in\mathcal{C}} \|p(\theta)\|_{\mathcal{X}} \sup_{\theta\in\mathcal{C}} \|p_I(\theta) - p(\theta)\|_{\mathcal{X}} \\
\le & \frac{1}{3} + \frac{4C_0}{3}\frac{1}{4C_0} + C_0\frac{3}{C_0} = 1,
\end{aligned}
$$

which yields that

$$\sup_{\theta\in\mathcal{C}} \|S^2(\theta)\|_{m\times m} \quad \leq \quad 2C_2 = 2\sup_{\theta_1,\theta_2\in\mathcal{C}} \|\mathrm{Cov}[F(\theta_1,\xi),F(\theta_2,\xi)\|_{m\times m} + 2 \quad < \quad \infty.$$

when $I \geq 2$. Note that for any $\theta_1, \theta_2 \in \mathcal{C}$

$$\begin{aligned} & 2|\nabla r(p_I(\theta_1))^\intercal S_n(\theta_1,\theta_2)\nabla r(p_I(\theta_2))| \\ \leq \;& \nabla r(p_I(\theta_1))^\intercal S^2(\theta_1)\nabla r(p_I(\theta_1)) + \nabla r(p_I(\theta_2))^\intercal S^2(\theta_2)\nabla r(p_I(\theta_2)) \end{aligned}$$

Thus, for all $I \geq \max\left\{M(\omega), 8C_1^2C_2/\varepsilon, 2\right\}$, it follows that

$$\begin{aligned} & \frac{1}{I}\sup_{\theta_1,\theta_2\in\mathcal{C}} \nu^2(\theta_1,\theta_2,I) \\ \leq \;& \frac{2}{I}\sup_{\theta_1,\theta_2\in\mathcal{C}} \left(\nabla r(p_I(\theta_1))^\intercal S^2(\theta_1)\nabla r(p_I(\theta_1)) + \nabla r(p_I(\theta_2))^\intercal S^2(\theta_2)\nabla r(p_I(\theta_2))\right) \\ \leq \;& \frac{4}{I}\left(\sup_{\theta\in\mathcal{C}}\|S^2(\theta)\|_{m\times m}\right)\left(\sup_{\theta\in\mathcal{C}}\|\nabla r(p_I(\theta))\|_{m\times 1}\right)^2 \\ \leq \;& \varepsilon. \end{aligned}$$

which completes the proof. □

**Lemma 2.2.** *Assume A.3, A.7 and $\|H_k\|_{d\times d} < C < \infty$ for all $k$. At iteration $k$, suppose that $-g_k \notin \mathcal{H}(\theta_k)$. If (2.8) is ever violated for $\widehat{t} > 0$ in Algorithm 2, there exists an interval of $t$ between $(0, \widehat{t})$ strictly satisfying (2.8) and (2.9). If (2.7) is ever violated for $\widehat{t}$, then*

*(i) there exists an interval of $t$ between $(0, \widehat{t})$ satisfying (2.7), (2.8) and (2.10) or*

*(ii) there exists an interval of $t$ between $(0, \widehat{t})$ strictly satisfying (2.8) and (2.9).*

**Proof:** We first show that $g_k^\intercal s_k(t) < 0$ for all $t > 0$. Assume by contradiction that $g_k^\intercal s_k(t_j) \geq 0$ for some $t_j > 0$. By P.1,

$$\langle \Pi_{\mathcal{C}}[\theta_k - t_j g_k] - (\theta_k - t_j g_k), y - \Pi_{\mathcal{C}}[\theta_k - t_j g_k]\rangle \quad \geq \quad 0 \quad \forall y \in \mathcal{C}.$$

If $\|s_k(t_j)\| := \|\Pi_{\mathcal{C}}[\theta_k - t_j g_k] - \theta_k\| = 0$, it has $\Pi_{\mathcal{C}}[\theta_k - t_j g_k] = \theta_k$ and that

$$g_k^\intercal(y - \theta_k) \quad \geq \quad 0 \quad \forall y \in \mathcal{C},$$

which is $-g_k \in \mathcal{H}(\theta_k)$ and contradicts $-g_k \notin \mathcal{H}(\theta_k)$. Thus, $\|s_k(t_j)\| > 0$. It follows from Property P.6 that

$$s_k(t_j) \in \arg\min \left\{ g_k^\intercal s : \theta_k + s \in \mathcal{C}, \|s\| \leq \|s_k(t_j)\| \right\}.$$

Consider any $y \in \mathcal{C}$. There exists some $\lambda > 0$ such that $\|\lambda(y-\theta_k)\| = \lambda\|y-\theta_k\| \leq \|s_k(t_j)\|$. Thus, $\lambda g_k^\intercal (y-\theta_k) \geq g_k^\intercal s_k(t_j) \geq 0$, which implies that $-g_k \in \mathcal{H}(\theta_k)$ and contradicts $-g_k \notin \mathcal{H}(\theta_k)$.

Second, we show that there exists $\alpha > 0$ such that for all $t \in (0, \alpha]$, (2.9) is strictly violated. By the continuity of the solution of convex optimization problem (Theorem 3.2.8, [23, pp. 44 – 45]), $g_k^\intercal s_k(t)$ is continuous in $t$ and $g_k^\intercal s_k(t) \to 0$ as $t \downarrow 0$. By Property P.4, both $s_k(t)$ and $\|s_k(t)\|$ are continuous in $t$ and $\|s_k(t)\| \to 0$ as $t \downarrow 0$. Define

$$q(t) \quad := \quad m_k(v_k(t,\theta_k)) - m_k(\theta_k) - \kappa_2 g_k^\intercal s_k(t) \quad = \quad (1-\kappa_2) g_k^\intercal s_k(t) + \frac{1}{2}(s_k(t))^\intercal H_k s_k(t).$$

Thus, $q(t)$ is continuous in $t$ and $q(0) = 0$, since $s_k(0) = 0$. Thus, since $\kappa_2$ satisfies (2.12) and $g_k^\intercal s_k(t) < 0$ for all $t > 0$, it follows

$$\begin{aligned} \liminf_{t\downarrow 0} \frac{q(t)}{\|s_k(t)\|} &= \liminf_{t\downarrow 0} \frac{m_k(\theta_k + s_k(t)) - m_k(\theta) - \kappa_2 g_k^\intercal s_k(t)}{\|s_k(t)\|} \\ &= \liminf_{t\downarrow 0} \frac{(1-\kappa_2) g_k^\intercal s_k(t)}{\|s_k(t)\|} + \lim_{t\downarrow 0} \frac{(s_k(t))^\intercal H_k s_k(t)}{2\|s_k(t)\|} \\ &= \liminf_{t\downarrow 0} \frac{(1-\kappa_2) g_k^\intercal s_k(t)}{\|s_k(t)\|} \\ &< 0, \end{aligned}$$

which implies there exists some $\alpha > 0$ such that $q(t) < 0$ for all $t \in (0, \alpha]$, i.e., (2.9) is strictly violated for all $t \in (0, \alpha]$. Also, condition (2.8) strictly holds for all $t \in (0, \alpha]$ since $\kappa_1 < \kappa_2$, i.e.,

$$\begin{aligned} m_k(v_k(t,\theta_k)) - m_k(\theta_k) - \kappa_1 g_k^\intercal s_k(t) &< 0, \quad \forall t \leq \alpha, \\ m_k(v_k(t,\theta_k)) - m_k(\theta_k) - \kappa_2 g_k^\intercal s_k(t) &< 0, \quad \forall t \leq \alpha. \end{aligned}$$

If condition (2.8) is violated for $\widehat{t} > \alpha$, i.e.,

$$m_k(v_k(\widehat{t},\theta_k)) - m_k(\theta_k) - \kappa_1 g_k^\intercal s_k(\widehat{t}) \quad > \quad 0,$$

there exists $\widehat{t}_1 \in (\alpha, \widehat{t})$ such that

$$\begin{aligned} m_k(v_k(\widehat{t}_1, \theta_k)) - m_k(\theta_k) - \kappa_1 g_k^\intercal s_k(\widehat{t}_1) &= 0, \\ m_k(v_k(\widehat{t}_1, \theta_k)) - m_k(\theta_k) - \kappa_2 g_k^\intercal s_k(\widehat{t}_1) &> 0. \end{aligned}$$

By a similar argument, there exists $\widehat{t}_2 \in (\alpha, \widehat{t}_1)$ such that

$$\begin{aligned} m_k(v_k(\widehat{t}_2, \theta_k)) - m_k(\theta_k) - \kappa_1 g_k^\intercal s_k(\widehat{t}_2) &< 0, \\ m_k(v_k(\widehat{t}_2, \theta_k)) - m_k(\theta_k) - \kappa_2 g_k^\intercal s_k(\widehat{t}_2) &= 0. \end{aligned}$$

Consider the following optimization problem:

$$\widehat{t}_3 \;:=\; \inf_{t \in A} \; t,$$

where

$$A \;:=\; \{t \in [\widehat{t}_2, \widehat{t}_1] : m_k(v_k(t, \theta_k)) - m_k(\theta_k) - \kappa_1 g_k^\intercal s_k(t) \;\geq\; 0\}.$$

Since $\widehat{t}_1$ is a feasible point, it holds that $A \neq \varnothing$. Due to the continuity of $m_k(v_k(t, \theta_k)) := m_k(\theta_k + s_k(t))$, we have $A$ is compact and the optimization problem attains the minimum at $\widehat{t}_3 > \widehat{t}_2$. Thus, (2.8) strictly holds for all $[\widehat{t}_2, \widehat{t}_3)$, and that

$$m_k(v_k(\widehat{t}_3, \theta_k)) - m_k(\theta_k) - \kappa_2 g_k^\intercal s_k(\widehat{t}_3) \;>\; 0.$$

Consider the following optimization problem:

$$\widehat{t}_4 \;:=\; \sup_{t \in B} \; t,$$

where

$$B \;:=\; \{t \in [\widehat{t}_2, \widehat{t}_3] : m_k(v_k(t, \theta_k)) - m_k(\theta_k) - \kappa_2 g_k^\intercal s_k(t) \;\leq\; 0\}.$$

Since $\widehat{t}_2$ is a feasible point, it holds that $B \neq \varnothing$. Due to the continuity of $m_k(v_k(t, \theta_k)) := m_k(\theta_k + s_k(t))$, $B$ is compact, and thus the optimization problem attains the maximum at $\widehat{t}_4 < \widehat{t}_3$. Thus, (2.8) and (2.9) strictly hold for all $(\widehat{t}_4, \widehat{t}_3)$. Thus, there exists an interval between $(\widehat{t}_2, \widehat{t}_1] \subset (\alpha, \widehat{t})$ such that (2.8) and (2.9) strictly hold for all $t_j$ in the interval.

If condition (2.7) is violated for $\widehat{t}$, it has $\|s_k(\widehat{t})\| > \Delta_k$. By Property P.4, $\|s_k(t)\|$ is continuous and nondecreasing in $t \geq 0$. Thus, there exists some interval of $t$ between $(0, \widehat{t})$ such that

$$\kappa_3 \Delta_k \;\leq\; \|s_k(t)\| \;<\; \Delta_k$$

for all $t$ in the interval. If all points in the interval satisfy (2.8), the interval satisfies (2.7), (2.8) and (2.10), which is (i). If there exists some point $\widehat{t}_1$ in the interval that violates (2.8), there exists an interval between $(\alpha, \widehat{t}_1) \subset (\alpha, \widehat{t})$ that strictly satisfy (2.8) and (2.9) according to the above argument, which is (ii). This completes the proof. □

**Theorem 2.3.** *Assume A.3, A.7 and* $\|H_k\|_{d\times d} < C < \infty$ *for all* $k$*. Suppose* $-g_k \notin \mathcal{H}(\theta_k)$ *for each* $k$*. Algorithm 2 terminates in finite number of iterations.*

**Proof:** If both (2.7) and (2.8) hold but (2.9), (2.10) and (2.11) are violated for an infinite number of iterations. There exists an increasing sequence of numbers $\{t_j\}_{j=1}^{\infty}$ such that $\|s_k(t_j)\| < \kappa_3 \Delta_k$ for all $j$. It follows from Properties P.4 and P.5 that

$$\lim_{j\to\infty} \|\Pi_{\mathcal{T}(v_k(t_j,\theta_k))}[-g_k]\| \quad = \quad 0.$$

Since $g_k^\intercal s_k(t) < 0$ for all $t > 0$ and $|g_k^\intercal s_k(t_j)|$ is nondecreasing as $j \to \infty$ by Properties P.6 and P.7 and $t_j \geq t_0 > 0$,

$$|g_k^\intercal s_k(t_j)| \quad \geq \quad |g_k^\intercal s_k(t_0)| \quad > \quad 0,$$

which yields (2.11) holds after a finite number of iterations. In this case $\alpha_k \leftarrow t_j$ for the first $t_j$ that satisfies (2.11) and Algorithm 2 terminates in a finite number of iterations.

Suppose that (2.7) or (2.8) is ever violated for some $\widehat{t}$. It follows from Lemma 2.2 that, if (2.7) or (2.8) is violated for $\widehat{t}$, there exists an interval of $t$ that satisfies case (i) (2.7), (2.8) and (2.10) or that strictly satisfy case (ii) (2.8) and (2.9). For case (i), we have showed that there exists an interval of $t$ that satisfies the stopping condition of Algorithm 2, i.e., (2.7), (2.8) and (2.10) are satisfied for that interval.

We next consider case (ii), i.e., there exists an interval of $t$ between $(\alpha, \widehat{t})$ such that the interval strictly satisfies (2.8) and (2.9). Let $I_1$ be the set of all such $t$ between $(\alpha, \widehat{t})$ that satisfies (2.8) and (2.9), where $\alpha > 0$ is such that (2.9) is violated for all $t \in (0, \alpha]$. We have the following two cases.

**C.1.** There exists a $t_0 \in I_1$ such that $\|s_k(t_0)\| \leq \Delta_k$, i.e., (2.7) holds for $t_0$. It follows from Lemma 2.2 that $g_k^\intercal s_k(t) < 0$ for all $t > 0$. Note $t_0 > 0$. Since $t_0 \in I_1$, it follows that

$$\kappa_1 \quad \leq \quad q(t_0) \quad := \quad \frac{m_k(v_k(t_0, \theta_k)) - m_k(\theta_k)}{g_k^\intercal s_k(t_0)} \quad < \quad \kappa_2.$$

If $q(t_0) = \kappa_1$, i.e., (2.8) holds as equality at $t_0$, it follows from (the proof of) Lemma 2.2 that there exists an interval $A \subset (\alpha, t_0] \subset (\alpha, \widehat{t})$ satisfying (2.8) and (2.9). For all $t \in A$, it holds that $t \leq t_0$. It follows from P.4 that $\|s_k(t)\| \leq \|s_k(t_0)\| \leq \Delta_k$. Thus, (2.7) holds for all points in $A$. Then there exists an interval $A$ that satisfies the stopping condition of Algorithm 2, i.e., (2.7), (2.8) and (2.9) are satisfied for $A$.

Now consider $q(t_0) > \kappa_1$. Property P.4 asserts that $s_k(t)$ is continuous. Thus, $q(t)$ is continuous, and there exists $\delta > 0$ such that

$$\kappa_1 \quad < \quad q(t) \quad < \quad \kappa_2, \quad \forall t \in B := [t_0 - \delta, t_0] \subset (\alpha, t_0] \subset (\alpha, \widehat{t}),$$

which implies that (2.8) and (2.9) are satisfied over $B$. It follows from P.4 that (2.7) holds for all points in $B$. Then there exists an interval $B$ that satisfies the stopping condition of Algorithm 2, i.e., (2.7), (2.8) and (2.9) are satisfied for $B$.

**C.2.** Suppose all the points in $I_1$ violate (2.7). Let $\widehat{t}_1$ be such that $\|s_k(\widehat{t}_1)\| = \Delta_k$. Thus, the interval $(0, \widehat{t}_1]$ satisfy (2.7). Thus, $I_1 \cap (0, \widehat{t}_1] = \varnothing$. We claim that all points in $(0, \widehat{t}_1]$ violate (2.9). Assume by contradiction that there exists a point $t_2 \in (0, \widehat{t}_1]$ that satisfies (2.9). If $t_2$ satisfies (2.8) too, it gives $t_2 \in I_1 \cap (0, \widehat{t}_1]$, contradicting $I_1 \cap (0, \widehat{t}_1] = \varnothing$. If $t_2$ violates (2.8), it follows from (the proof of) Lemma 2.2 that there exists an interval $D \subset (0, t_2] \subset (0, \widehat{t}_1]$ satisfying (2.8) and (2.9), contradicting $I_1 \cap (0, \widehat{t}_1] = \varnothing$. Let $\widehat{t}_2$ be such that $\|s_k(\widehat{t}_2)\| = \kappa_3 \Delta_k$. Thus, $(\widehat{t}_2, \widehat{t}_1]$ satisfies (2.7) and (2.10) and violates (2.9) and thus satisfies (2.8).

Thus, there exists some interval $I \subset (\alpha, \widehat{t})$ that satisfies (2.7), (2.8) and (2.9) or (2.7), (2.8) and (2.10). In this case, the algorithm sets $t_{\max} := \widehat{t}$.

If both (2.7) and (2.8) hold but (2.9), (2.10) and (2.11) are violated for some $\widehat{t} < t_{\max}$. Note that $t_{\max}$ either violates (2.7) or (2.8). We have the following two cases:

**C.3.** $t_{\max}$ violates (2.8). Since $\widehat{t}$ violates (2.9), it follows from the same argument in Lemma 2.2 and the argument for Case **C.1** that there exists an interval within $(\widehat{t}, t_{\max})$ that satisfies (2.7), (2.8) and (2.9) by replacing $\widehat{t}$ as $\alpha$. Since $\widehat{t}$ violates (2.10), it follows from Lemma 2.2 and the argument for Case **C.2** that, there exists an interval within $(\widehat{t}, t_{\max})$ satisfying (2.7), (2.8) and (2.10).

**C.4.** $t_{\max}$ violates (2.7). Let $t_1$ be such that $\|s_k(t_1)\| = \Delta_k$ and $t_2$ be such that $\|s_k(t_2)\| = \kappa_3\Delta_k$. Since $\widehat{t}$ violates (2.10), $[t_2, t_1] \subset (\widehat{t}, t_{\max})$. Thus, it follows Lemma 2.2, Case **C.1** and Case **C.2** that there exists an interval within $(\widehat{t}, t_{\max})$ satisfying (2.7), (2.8) and (2.9) or (2.7), (2.8) and (2.10).

In this case, we set $t_{\min} := \widehat{t}$.

Thus, the interval that satisfies (2.7), (2.8) and (2.9) or (2.7), (2.8) and (2.10) is contained in $[t_{\min}, t_{\max}]$ after each iteration. Assume by contradiction that (2.7) or (2.8) is violated for infinitely many iterations. Let $\{\bar{t}_j\}_{j=1}^{\infty}$ be the decreasing sequence of upper bounds of the interval and $\{\underline{t}_j\}_{j=1}^{\infty}$ be the increasing sequence of lower bounds of the interval. Since the search is bisection, it follows that $\bar{t}_j \downarrow t^*$ and $\underline{t}_j \uparrow t^*$ as $j \to \infty$. By the definition of the upper bounds, it obtains that

$$\|s_k(\bar{t}_j)\| \quad > \quad \Delta_k \quad \text{or} \quad m_k(\theta_k + s_k(\bar{t}_j)) - m_k(\theta_k) - \kappa_1 g_k^{\intercal} s_k(\bar{t}_j) \quad > \quad 0.$$

Taking limit on both sides of the above inequalities with respect to $j \to \infty$ gives that

$$\|s_k(t^*)\| \quad \geq \quad \Delta_k \quad \text{or} \quad m_k(\theta_k + s_k(t^*)) - m_k(\theta_k) - \kappa_1 g_k^{\intercal} s_k(t^*) \quad \geq \quad 0. \tag{2.26}$$

By the definition of lower bonds, it obtains that

$$\|s_k(\underline{t}_j)\| \quad < \quad \kappa_3\Delta_k \quad \text{and} \quad m_k(\theta_k + s_k(\underline{t}_j)) - m_k(\theta_k) - \kappa_2 g_k^{\intercal} s_k(\underline{t}_j) \quad < \quad 0.$$

Taking limit on both sides of the above inequalities with respect to $j \to \infty$ gives that

$$\|s_k(t^*)\| \quad \leq \quad \kappa_3\Delta_k \quad \text{and} \quad m_k(\theta_k + s_k(t^*)) - m_k(\theta_k) - \kappa_2 g_k^{\intercal} s_k(t^*) \quad \leq \quad 0,$$

which contradicts (2.26). Thus, (2.7) or (2.8) is violated only in a finite number of iterations.

It follows from the argument in the first paragrah that, Algorithm 2 terminates in a finite number of iterations. □

**Lemma 2.3.** *For each $k \in \{0, 1, \ldots\}$, if (2.15) fails and $\widehat{\sigma}_{k,1} > \widehat{\sigma}_{k,2}$, then $n_k \geq 1$, where $n_k$ is defined in (2.17).*

**Proof**: Since (2.15) fails for $k$ and $\widehat{\sigma}_{k,1} > \widehat{\sigma}_{k,2}$, it follows that

$$\widehat{\sigma}_1^2(\theta_k, \theta_k^{\mathrm{GC}}, N_k, \mathcal{I}_k) \quad = \quad \frac{|\mathcal{N}| - |N_k|}{|\mathcal{N}||N_k|}\bar{\nu}^2(\theta_k, \theta_k^{\mathrm{GC}}, N_k, \mathcal{I}_k) \quad > \quad (\eta_0 \Delta m(\theta_k, \theta_k^{\mathrm{GC}}))^2,$$

which indicates that $|N_k| < |\mathcal{N}|$; $\widehat{\sigma}_1^2(\theta_k, \theta_k^{\mathrm{GC}}, N_k, \mathcal{I}_k) = 0$ otherwise, and yields that

$$|N_k| \quad < \quad \frac{|\mathcal{N}|}{1 + |\mathcal{N}| \frac{(\eta_0 \Delta m(\theta_k, \theta_k^{\mathrm{GC}}))^2}{\bar{\nu}^2(\theta_k, \theta_k^{\mathrm{GC}}, N_k, \mathcal{I}_k)}} \quad \leq \quad \left\lceil \frac{|\mathcal{N}|}{1 + |\mathcal{N}| \frac{(\eta_0 \Delta m(\theta_k, \theta_k^{\mathrm{GC}}))^2}{\bar{\nu}^2(\theta_k, \theta_k^{\mathrm{GC}}, N_k, \mathcal{I}_k)}} \right\rceil.$$

Thus,

$$n_k \quad = \quad \left\lceil \frac{|\mathcal{N}|}{1 + |\mathcal{N}| \frac{(\eta_0 \Delta m(\theta_k, \theta_k^{\mathrm{GC}}))^2}{\bar{\nu}^2(\theta_k, \theta_k^{\mathrm{GC}}, N_k, \mathcal{I}_k)}} \right\rceil - |N_k| \quad \geq \quad 1,$$

which is the desired result. □

**Lemma 2.4.** *Assume A.3, A.5, A.7 and A.10. For any iteration $k$ in Algorithm 1, if (2.15) fails for all $k, k+1, \ldots,$ then $|N_k| \to |\mathcal{N}|$ and $|I_n^k| \to \infty$ for all $n \in \mathcal{N}$ as $k \to \infty$.*

**Proof**: Consider any iteration $k$. If (2.15) fails for all $k+j$, where $j \in \mathbb{Z}_+$, then Algorithm 3 will be executed for $k+j$, where $j \in \mathbb{Z}_+$, under the condition that (2.15) fails. We suppress index $k$ for notational convenience.

We first show that for any $j \in \mathbb{Z}_+ := \mathbb{N} \cup \{0\}$ such that $|N_j| < |\mathcal{N}|$, there exists $j' \in \mathbb{Z}_+$ with $j' > j$ such that $|N_{j'}| > |N_j|$. Assume by contradiction that there exists some $J \in \mathbb{Z}_+$ such that $|N_J| < |\mathcal{N}|$,

$$|N_j| \quad = \quad |N_J| \quad \forall\, j \geq J.$$

We now analyze Case 5 and Case 6 in Algorithm 3 and conclude that these cases are impossible.

**C.1.** Suppose Case 5 is ever executed at some $j \geq J$. Note that $|N_j| = |N_J| < |\mathcal{N}|$. It follows from Lemma 2.3 that $n_k \geq 1$ and that $|N_{j+1}| = |N_j| + n_k > |N_j| = |N_J|$, contradicting $|N_j| = |N_J|$ for all $j \geq J$.

**C.2.** Suppose Case 5 is not executed at all $j \in \mathbb{Z}_+$. Then Case 6 is executed for all $j \in \mathbb{Z}_+$. Suppose there exists some $j \geq J$ such that $X_{j,1} = 1$. Note that $|N_j| < |\mathcal{N}|$. then $|N_{j+1}| = |N_j| + n_a \wedge (|\mathcal{N}| - |N_j|)$. Since $n_a \geq 1$ and $|\mathcal{N}| - |N_j| \geq 1$, it holds that $|N_{j+1}| \geq |N_j| + 1$, contradicting $|N_j| = |N_J|$ for all $j \geq J$. Thus, $X_{j,1} = 0$ for all $j \geq J$.

Consider the event $A_6 := [$ Case 6 holds and $X_{J+j,1} = 0$ for $j \in \{0, 1, \dots\}]$. Note that for any $j \in \mathbb{Z}_+$,

$$1 - q^0_{j,i} \quad = \quad \frac{h(\widehat{\sigma}_{j,2})}{\epsilon + h(\widehat{\sigma}_{j,1}) + h(\widehat{\sigma}_{j,2})} \quad \forall i \in \{1, 2\}.$$

Assume by contradiction that

$$\begin{aligned} \mathbb{P}(A_6) \quad = \quad \prod_{j=J}^{\infty}(1 - q^0_{j,1}) \quad > \quad 0 \quad \Rightarrow \quad 0 \quad > \quad \sum_{j=J}^{\infty} \log(1 - q^0_{j,1}) \quad > \quad -\infty \\ \Rightarrow \quad 0 \quad < \quad \sum_{j=J}^{\infty} \log\left(\frac{\epsilon + h(\widehat{\sigma}_{j,1}) + h(\widehat{\sigma}_{j,2})}{h(\widehat{\sigma}_{j,2})}\right) \quad < \quad \infty, \end{aligned}$$

which implies

$$\frac{\epsilon + h(\widehat{\sigma}_{j,1}) + h(\widehat{\sigma}_{j,2})}{h(\widehat{\sigma}_{j,2})} \to 1, \quad \text{as} \quad j \to \infty. \tag{2.27}$$

For event $A_6$, $|N_j| = |N_J|$ for all $j \geq J$ and $|I^{j+1}_n| = |I^j_n| + i_a \vee (i_n - |I^j_n|) \geq |I^j_n| + 1$ since $i_a \geq 1$. Thus, $|I^j_n| \to \infty$ as $j \to \infty$ for all $n \in |N_J|$. Since the trust region is a compact set and A.3, A.5, A.7 and A.10 hold, it follows from Lemma 2.1(4) that $\widehat{\sigma}_{j,2} \to 0$ as $j \to \infty$. Since $h$ is a nondecreasing continuous function, it follows that $h(\widehat{\sigma}_{j,2}) \to h(0)$ as $j \to \infty$. Note that under Case 6 we have $\widehat{\sigma}_{j,2} \geq \widehat{\sigma}_{j,1}$ for all $j \geq J$, which implies that $\widehat{\sigma}_{j,1} \leq \widehat{\sigma}_{j,2}$ and $\widehat{\sigma}_{j,1} \to 0$ as $j \to \infty$. Thus, since $0 \leq h(0) < \infty$.

$$\frac{\epsilon + h(\widehat{\sigma}_{j,1}) + h(\widehat{\sigma}_{j,2})}{h(\widehat{\sigma}_{j,2})} \to \frac{\epsilon + 2h(0)}{h(0)} \quad \neq \quad 1, \quad \text{as} \quad j \to \infty,$$

contradicting (2.27). Thus, $\mathbb{P}(A_6) = 0$ and this case cannot happen and $|N_k| \to |\mathcal{N}|$ as $k \to \infty$.

Next, we show that for any $j \in \mathbb{Z}_+$, there exists some $j' \in \mathbb{Z}_+$ with $j' > j$ such that $|I^{j'}_n| > |I^j_n|$ for all $n \in \mathcal{N}$. Assume by contradiction that there exists some $n \in \mathcal{N}$ and $J \in \mathbb{Z}_+$ such that ,

$$|I^j_n| \quad = \quad |I^J_n| \;\; \forall\, j \geq J.$$

We now analyze Case 5 and Case 6 in Algorithm 3 and conclude that these cases are impossible.

**C.3.** Suppose Case 6 is ever executed at some $j \geq J$. It follows that $|I_n^{j+}| = |I_n^j| + \max\{i_a, i_n - |I_n^j|\} \geq |I_n^j| + 1$ and $|I_n^{j+1}|$ is updated as $|I_n^{j+}|$. Thus, $|I_n^{j+1}| \geq |I_n^j| + 1$ which contradicts $|I_n^j| = |I_n^J|$ for all $j \geq J$. This case cannot happen.

**C.4.** Suppose Case 6 is not executed at all $j \in \mathbb{Z}_+$. Then Case 5 is executed for all $j \in \mathbb{Z}_+$. Suppose there exists some $j \geq J$ such that $X_{j,2} = 1$. Then $|I_n^{j+1}| = |I_n^j| + i_a$. Since $i_a \geq 1$, it holds that $|I_n^{j+1}| \geq |I_n^j| + 1$, contradicting $|I_n^j| = |I_n^J|$ for all $j \geq J$. Thus, $X_{j,2} = 0$ for all $j \geq J$.

Consider the event $A_5 := [$ Case 5 holds and $X_{J+j,2} = 0$ for $j \in \{0, 1, \ldots\}]$. By a similar argument in **C.2**, $\mathbb{P}(A_5) = 0$ and this case cannot happen and $|I_n^k| \to \infty$ for all $n \in \mathcal{N}$ as $k \to \infty$. □

**Theorem 2.4.** *Assume A.3, A.4, A.5, A.7 and A.10. Suppose that* $\|H_k\|_{d\times d} \leq C < \infty$ *for each* $k$ *and* $-\nabla f(\theta_k) \notin \mathcal{H}(\theta_k)$ *at iteration* $k$*. Then, w.p. 1, for each iteration* $k$ *at* ***Step 4*** *in Algorithm 1,* ***Step 5*** *will be executed after a finite number of iterations.*

**Proof**: Assume by contradiction that there exists $\Omega^* \subset \Omega$ with $\mathbb{P}(\Omega^*) = \alpha > 0$ such that for each sampling process $\omega \in \Omega^*$, (2.15) fails for all $k+j$, where $j \in \{0, 1, \ldots\}$, i.e., **Step 5** will not be executed after $k$. We suppress index $k$ for convenience. Fix a sampling process $\omega \in \Omega^*$. It follows from Lemma 2.4 that the data and Monte-Carlo sample sizes are updated such that

$$|N_j| \uparrow |\mathcal{N}| \text{ and } |I_n^j| \uparrow \infty \quad \forall n \in \mathcal{N}, \text{ as } j \to \infty.$$

Since the sample update process does not terminate, condition (2.15) is violated after each sample update $j$, $j = 0, 1, \ldots$. Denote by $\theta$ the current position $\theta_k$ and $\Delta$ the current trust region radius $\Delta_k$. Recall that $\theta_j^{\mathrm{GC}} = \theta + s_j(\alpha_j)$. Thus, it follows from Lemma 2.1 that w.p.1,

$$\Delta m_j(\theta, \theta_j^{\mathrm{GC}}) \quad \leq \quad \frac{1}{\eta_0} \max\{\widehat{\sigma}_{k,1}, \widehat{\sigma}_{k,2}\} \to 0. \tag{2.28}$$

Since $s_j(\alpha_j)$ satisfies (2.7) and (2.8) for each $j$, we have that

$$-\kappa_1 g_j^\intercal s_j(\alpha_j) \quad \leq \quad \Delta m_k(\theta, \theta_j^{\mathrm{GC}}) \to 0 \text{ as } j \to \infty,$$

$$\|s_j(\alpha_j)\| \quad \leq \quad \Delta,$$

which implies $|g_j^\intercal s_j(\alpha_j)| \to 0$ as $j \to \infty$. We consider the following two cases.

**C.1.** $\|s_j(\alpha_j)\| \to 0$ as $j \to \infty$. Thus, we can assume $\|s_j(\alpha_j)\| \leq \min\{\Delta, 1\}$ for all $j$. Since condition (2.10) will be eventually violated as $\|s_j(\alpha_j)\| \to 0$, it follows that there exists a subsequence $\{j_k\}_{k=0}^\infty$ such that either (2.9) or (2.11) is satisfied for each $j_k$, where $k = 1, 2, \ldots$. Without loss of generality, we use $\{j\}_{j=1}^\infty$ as the subsequence.

Suppose (2.9) is satisfied for each $j$ of the subsequence. It follows $s_j(\alpha_j) \neq 0$ since $s_j(\alpha_j) = 0$ violates (2.9). Moreover, we have $g_j^\intercal s_j(\alpha_j) < 0$, since $g_j^\intercal s_j(\alpha_j) \geq 0$ does not satisfy both (2.8) and (2.9). Thus,

$$\begin{aligned} \Delta m_j(\theta, \theta + s_j(\alpha_j)) \quad &= \quad -g_j^\intercal s_j(\alpha_j) - \frac{1}{2}(s_j(\alpha_j))^\intercal H_j s_j(\alpha_j) \\ &< \quad -\kappa_2 g_j^\intercal s_j(\alpha_j) \ \ \forall j, \end{aligned}$$

which yields,

$$-(1-\kappa_2)\frac{g_j^\intercal s_j(\alpha_j)}{\|s_j(\alpha_j)\|} - \frac{s_j(\alpha_j))^\intercal H_j s_j(\alpha_j)}{2\|s_j(\alpha_j)\|} \quad < \quad 0 \ \ \forall j \tag{2.29}$$

By Property P.7 and $\|s_j(\alpha_j)\| \leq 1$, we have $\frac{-g_j^\intercal s_j(\alpha_j)}{\|s_j(\alpha_j)\|} \geq |g_j^\intercal d_j|$, where $d_j = v_j(\theta, \widehat{\alpha}_j) - \theta$ such that $\|d_j\| = 1$ for some $\widehat{\alpha}_j > 0$. Suppose $|g_j^\intercal d_j| \to 0$ as $j \to \infty$. By Property P.6,

$$|g_j^\intercal d_j| \quad = \quad |\chi(g_j, \theta, 1)| \quad := \quad |\min\{g_j^\intercal s : \theta + s \in \mathcal{C}, \|s\| \leq 1\}| \to 0.$$

By Theorem 3.2.8 in [23, pp. 44–45], $\chi(g_j, \theta, 1)$ is continuous in $g_j$. By ULLN, $\chi(\theta) = \lim_{j\to\infty} \chi(g_j, \theta, 1) = 0$ and $-\nabla f \in \mathcal{H}(\theta)$ contradicting $-\nabla f \notin \mathcal{H}(\theta)$. Suppose there exists a subsequence (w.l.o.g, we can still use sequence $\{j\}_{j=0}^\infty$ as the subsequence) and some $\varepsilon > 0$ such that $|g_j^\intercal d_j| > \varepsilon$ for all $j$. Thus, $\frac{-g_j^\intercal s_j(\alpha_j)}{\|s_j(\alpha_j)\|} \geq |g_j^\intercal d_j| \geq \varepsilon$ for all $j$. Since $s_j(\alpha_j) \to 0$ and $\|H_j\| \leq C$, there exists $N \in \mathbb{N}$ such that for all $j \geq N$,

$$\left| \frac{s_j(\alpha_j))^\intercal H_j s_j(\alpha_j)}{2\|s_j(\alpha_j)\|} \right| \quad \leq \quad \frac{(1-\kappa_2)\varepsilon}{2},$$

which combines (2.29) to imply that

$$-(1-\kappa_2)\frac{g_j^\intercal s_j(\alpha_j)}{\|s_j(\alpha_j)\|} \quad \leq \quad \frac{(1-\kappa_2)\varepsilon}{2} \quad \Rightarrow \quad -\frac{g_j^\intercal s_j(\alpha_j)}{\|s_j(\alpha_j)\|} \quad \leq \quad \frac{\varepsilon}{2},$$

contradicting $\frac{-g_j^\intercal s_j(\alpha_j)}{\|s_j(\alpha_j)\|} \geq |g_j^\intercal d_j| \geq \varepsilon$ for all $j$. Thus (2.9) cannot be satisfied infinitely often.

Suppose (2.11) is satisfied for each $j$ of the subsequence. By ULLN, $g_j \to \nabla f(\theta)$ as $j \to \infty$. Thus for any $\varepsilon > 0$, there exists $N_1 \in \mathbb{N}$ such that $\|g_j - \nabla f(\theta)\| < \varepsilon$ for all $j \geq N$. Thus, $\|g_j\| \leq \|\nabla f(\theta)\| + \varepsilon < \infty$ for all $j \geq N$. Since $\|s_j(\alpha_j)\| \to 0$ as $j \to \infty$, we have

$$\Pi_{\mathcal{T}(\theta+s_j(\alpha_j))}[-g_j] \quad \leq \quad \frac{\kappa_4 |g_j^\intercal s_j(\alpha_j)|}{\Delta} \to 0 \text{ as } j \to \infty \tag{2.30}$$

Since tangent cone is a closed convex cone, it follows from Property P.3 that

$$\begin{aligned}
& \|\Pi_{\mathcal{T}(\theta+s_j(\alpha_j))}[-\nabla f(\theta+s_j)]\| \\
\leq \; & \left\|\Pi_{\mathcal{T}(\theta+s_j(\alpha_j))}[-\nabla f(\theta+s_j)] - \Pi_{\mathcal{T}(\theta+s_j(\alpha_j))}[-g_j(\theta+s_j)]\right\| \\
+ \; & \left\|\Pi_{\mathcal{T}(\theta+s_j(\alpha_j))}[-g_j(\theta+s_j)] - \Pi_{\mathcal{T}(\theta+s_j(\alpha_j))}[-g_j]\right\| \\
+ \; & \left\|\Pi_{\mathcal{T}(\theta+s_j(\alpha_j))}[-g_j]\right\| \\
\leq \; & \|\nabla f(\theta+s_j) - g_j(\theta+s_j)\| + \|g_j(\theta+s_j) - g_j\| \\
+ \; & \|\Pi_{\mathcal{T}(\theta+s_j(\alpha_j))}[-g_j]\|
\end{aligned}$$

By Theorem 2.2, as $j \to \infty$,

$$\|\nabla f(\theta+s_j) - g_j(\theta+s_j)\| \to 0, \tag{2.31}$$

and

$$\begin{aligned}
\|g_j(\theta+s_j) - g_j\| \quad \leq \quad & \|g_j(\theta+s_j) - \nabla f(\theta+s_j)\| \\
+ \quad & \|\nabla f(\theta+s_j) - \nabla f(\theta)\| + \|\nabla f(\theta) - g_j\| \to 0,
\end{aligned} \tag{2.32}$$

where $\|g_j(\theta+s_j) - \nabla f(\theta+s_j)\| \to 0$ by Theorem 2.2, $\|\nabla f(\theta+s_j) - \nabla f(\theta)\| \to 0$ by the continuity of $\nabla f$ and $\|\nabla f(\theta) - g_j\| \to 0$ by ULLN. Combining (2.30), (2.31) and (2.32) gives that

$$\|\Pi_{\mathcal{T}(\theta+s_j(\alpha_j))}[-\nabla f(\theta+s_j)]\| \to 0 \text{ as } j \to \infty.$$

By [19, Lemma 3.3], since $f$ is continuously differentiable by Theorem 2.2, it follows that $\|\Pi_{\mathcal{T}(\theta+s_j(\alpha_j))}[-\nabla f(\theta+s_j)]\|$ is lower semicontinuous in $s_j$. Thus,

$$\|\Pi_{\mathcal{T}(\theta)}[-\nabla f(\theta)]\| \quad \leq \quad \liminf_{j\to\infty} \|\Pi_{\mathcal{T}(\theta+s_j(\alpha_j))}[-\nabla f(\theta+s_j)]\| \quad = \quad 0.$$

By [19, Lemma 3.1c], $-\nabla f(\theta) \in \mathcal{H}(\theta)$ if and only if $\|\Pi_{\mathcal{T}(\theta)}[-\nabla f(\theta)]\| = 0$, which contradicts $-\nabla f(\theta) \notin \mathcal{H}(\theta)$. Thus, **C.1** cannot happen.

**C.2.** There exists a subsequence of $\{j\}_{j=1}^{\infty}$ (we use $\{j\}_{j=1}^{\infty}$ as the subsequence) and $\varepsilon > 0$ such that $\|s_j(\alpha_j)\| > \varepsilon$ for all $j$. Under this case, we have $|g_j^\intercal s_j(\alpha_j)| \to 0$ as $j \to \infty$. It follows that

$$\begin{aligned} |\min\{g_j^\intercal s : \theta + s \in \mathcal{C}, \|s\| \leq \varepsilon\}| \quad &\leq \quad |\min\{g_j^\intercal s : \theta + s \in \mathcal{C}, \|s\| \leq \|s_j(\alpha_j)\|\}| \\ &= \quad |g_j^\intercal s_j(\alpha_j)| \to 0, \end{aligned}$$

since the two minimization problems obtain nonpositive optimal solutions. Thus, as $j \to \infty$

$$|\chi(g_j, \theta, \varepsilon)| \quad := \quad |\min\{g_j^\intercal s : \theta + s \in \mathcal{C}, \|s\| \leq \varepsilon\|\}| \to 0$$

By Theorem 3.2.8, pp.44 – 45, [23], $\chi(g_j, \theta, \varepsilon)$ is continuous in $g_j$. By ULLN,

$$\chi(\nabla f(\theta), \theta, \varepsilon) \quad = \quad |\min\{\nabla f(\theta)^\intercal s : \theta + s \in \mathcal{C}, \|s\| \leq \varepsilon\|\}| \quad = \quad 0,$$

which implies that if $\varepsilon \leq 1$,

$$\chi(\theta) \quad \leq \quad \frac{\chi(\nabla f(\theta), \theta, \varepsilon)}{\varepsilon} \quad = \quad 0$$

and if $\varepsilon > 1$

$$\chi(\theta) \quad \leq \quad \chi(\nabla f(\theta), \theta, \varepsilon) \quad = \quad 0,$$

which implies $-\nabla f(\theta) \in \mathcal{H}(\theta)$ contradicting $-\nabla f(\theta) \notin \mathcal{H}(\theta)$. Thus, **C.2** cannot happen.

This completes the proof. □

**Lemma 2.5.** *Assume A.3, A.5, A.8 and A.9 and A.10. It holds that w.p. 1* $|N_k| \to |\mathcal{N}|$ *as* $k \to \infty$ *and* $|I_n^k| \to \infty$ *as* $k \to \infty$ *for all* $n \in \mathcal{N}$.

**Proof**: We first show that for any iteration $k \in \mathbb{Z}_+$, if $|N_k| < |\mathcal{N}|$, then there exists $m \in \mathbb{Z}_+$ with $m > k$ such that $|N_m| > |N_k|$. Assume by contradiction that there exists some $k_0 \in \mathbb{N}$ such that $|N_m| = |N_{k_0}| < |\mathcal{N}|$ for all $m \geq k_0$. We now analyze each of the cases in Algorithm 3 and conclude that these cases are impossible.

**B.1.** Suppose that Case 2 or Case 4 holds for some $k \geq k_0$. It follows from the hypothesis assumption that $|N_{k_0}| = |N_k|$ and $|N_{k+1}| = |N_k| + n_a \wedge (|\mathcal{N}| - |N_k|)$. Since $n_a \geq 1$ and $|N_k| = |N_{k_0}| < |\mathcal{N}|$, it follows that $|N_{k+1}| \geq |N_k| + 1 = |N_{k_0}| + 1$, which contradicts $|N_m| = |N_{k_0}|$ for all $m \geq k_0$.

**B.2.** Suppose that Case 5 holds for some $k \geq k_0$. It follows that $|N_k^+| = |N_k| + n_k$. It follows from Lemma 2.3 that $n_k \geq 1$. Then $|N_k^+| \geq |N_k| + 1 = |N_{k_0}| + 1$ and $|N_{k+1}|$ is updated as $|N_k^+|$. Then, $|N_{k+1}| \geq |N_k| + 1 = |N_{k_0}| + 1$ which contradicts $|N_m| = |N_{k_0}|$ for all $m \geq k_0$. Thus, either Case 1, Case 3 or Case 6 in Algorithm 3 holds for an infinite sequence of iterations.

**B.3.** Consider the event $A_1 :=$ [Case 1 holds for $k \in \{k_1, k_2, \ldots\}$]. Note that for any $k \in \mathbb{Z}_+$,

$$q_{k,i}^1 = \frac{\epsilon + h(\widehat{\sigma}_{k,i})}{\epsilon + h(\eta_0 \Delta m_k(\theta_k, \theta_k^{\mathrm{GC}}))} \quad \forall i \in \{1, 2\}.$$

The probability of this event is

$$\mathbb{P}(A_1) = \prod_{j=1}^{\infty} (1 - q_{k_j,1}^1)(1 - q_{k_j,2}^1).$$

Assume by contradiction that $\mathbb{P}(A_1) > 0$. Then,

$$\sum_{j=1}^{\infty} \log(1 - q_{k_j,1}^1) + \sum_{j=1}^{\infty} \log(1 - q_{k_j,2}^1) > -\infty,$$

which implies that

$$\frac{1}{(\prod_{i=1}^{2}(1 - q_{k_j,i}^1))} = \frac{(\epsilon + h(\eta_0 \Delta m_{k_j}(\theta_{k_j}, \theta_{k_j}^{\mathrm{GC}})))^2}{\prod_{i=1}^{2}[h(\eta_0 \Delta m_{k_j}(\theta_{k_j}, \theta_{k_j}^{\mathrm{GC}})) - h(\widehat{\sigma}_{k_j,i})]} \to 1 \text{ as } j \to \infty.$$

Since both the data and integration sample sizes do not change. It follows from the first-order convergence result of the basic trust-region algorithm, and from P.7 and $\|H_k\| \leq C < \infty$ that

$$\|s_{k_j}(\alpha_{k_j})\| \to 0, \quad |g_{k_j}^{\mathsf{T}} s_{k_j}(\alpha_{k_j})| \leq \chi_{k_j}(\theta_{k_j}) \to 0 \quad \text{as } j \to \infty,$$

which implies that $\Delta m_{k_j}(\theta_{k_j}, \theta_{k_j}^{\text{GC}}) \to 0$. Note that $\eta_0 \Delta m_{k_j}(\theta_{k_j}, \theta_{k_j}^{\text{GC}}) \geq \widehat{\sigma}_{k_j,i}$ for $i \in \{1,2\}$ under this case. Since $h$ is a nondecreasing continuous function, it follows that $h(\eta_0 \Delta m_{k_j}(\theta_{k_j}, \theta_{k_j}^{\text{GC}})) \to h(0)$, $h(\widehat{\sigma}_{k_j,i}) \leq h(\eta_0 \Delta m_{k_j}(\theta_{k_j}, \theta_{k_j}^{\text{GC}}))$ and $h(\widehat{\sigma}_{k_j,i}) \to h(0)$ as $j \to \infty$. Thus,

$$\frac{1}{(\prod_{i=1}^{2}(1 - q_{k_j,i}^1))} \to \infty \text{ as } j \to \infty$$

contradicting $1/(\prod_{i=1}^{2}(1 - q_{k_j,i}^1)) \to 1$ as $j \to \infty$. Thus, $\mathbb{P}(A_1) = 0$.

**B.4.** Consider the event $A_3 := $ [Case 3 holds for $k \in \{k_1, k_2, \ldots\}$]. The probability of this event is

$$\mathbb{P}(A_3) \quad = \quad \prod_{j=1}^{\infty}(1 - q_{k_j,1}^1)q_{k_j,2}^1.$$

Assume by contradiction that $\mathbb{P}(A_1) > 0$. Then,

$$\sum_{j=1}^{\infty} \log[q_{k_j,2}^1(1 - q_{k_j,1}^1)] \quad > \quad -\infty,$$

which implies that

$$\frac{1}{q_{k_j,2}^1(1 - q_{k_j,1}^1)} \quad = \quad \frac{(\epsilon + h(\eta_0 \Delta m_{k_j}(\theta_{k_j}, \theta_{k_j}^{\text{GC}})))^2}{[\epsilon + h(\widehat{\sigma}_{k_j,2})][h(\eta_0 \Delta m_{k_j}(\theta_{k_j}, \theta_{k_j}^{\text{GC}})) - h(\widehat{\sigma}_{k_j,1})]} \to 1 \text{ as } j \to \infty.$$

Since data sample does not change and integration sample size increases up to infinity, it follows from Lemma 2.1(4) that $\widehat{\sigma}_{k_j,2} \to 0$ as $j \to \infty$. Thus, $h(\widehat{\sigma}_{k_j,2}) \to h(0)$, and it follows from the DCT, Theorem 2.1, and (2.23) that

$$h(\widehat{\sigma}_{k_j,1}) \to h\left(\sigma_1\right), \quad \text{as} \quad j \to \infty.$$

where $\sigma_1$ is given by,

$$\sigma_1^2 \quad = \quad \frac{\kappa_c}{|N||\mathcal{N}|} \sum_{n \in N} \left( \Delta r_n(\theta_1, \theta_2) - \frac{\sum_{n \in N} \Delta r_n(\theta_1, \theta_2)}{N|} \right)^2,$$

$$\Delta r_n(\theta_1, \theta_2) \quad := \quad r(p_n(\theta_1)) - r(p_n(\theta_2)).$$

Suppose that $h(\eta_0 \Delta m_{k_j}(\theta_{k_j}, \theta_{k_j}^{\text{GC}})) \to h(x_0)$ as $j \to \infty$. Then,

$$(\epsilon + h(\eta_0 \Delta m_{k_j}(\theta_{k_j}, \theta_{k_j}^{\text{GC}})))^2 \quad = \quad \epsilon[h(\eta_0 \Delta m_{k_j}(\theta_{k_j}, \theta_{k_j}^{\text{GC}})) - h(\sigma_1)],$$

which implies that

$$\epsilon^2 + h^2(\eta_0 \Delta m_{k_j}(\theta_{k_j}, \theta_{k_j}^{\text{GC}})) + \epsilon h(\eta_0 \Delta m_{k_j}(\theta_{k_j}, \theta_{k_j}^{\text{GC}})) \quad = \quad -\epsilon h(\sigma_2) \quad \leq \quad 0,$$

which contradicts $\epsilon > 0$. Thus, $\mathbb{P}(A_3) = 0$.

**B.5.** Consider the event $A_6 :=$ [ Case 6 holds and $X_{k,1} = 0$ for $k \in \{k_1, k_2, \ldots\}$]. It follows from **C.2** in Lemma 2.4 that $\mathbb{P}(A_6) = 0$.

Thus, w.p.1 $|N_k| \to |\mathcal{N}|$ as $k \to \infty$.

Next, we show that for any iteration $k \in \mathbb{Z}_+$ and each customer $n \in N_k$, there exits $m \in \mathbb{Z}_+$ with $m > k$ such that $|I_n^m| > |I_n^k|$. Assume by contradiction that, there exists $k_0 \in \mathbb{Z}_+$ and $n \in N_k$ such that $|I_n^k| = |I_n^{k_0}|$ for all $k \geq k_0$. By a similar argument as above, we analyze each of the 6 cases in Algorithm 3 and make the claim that these cases are impossible.

**C.1.** Suppose that Case 3 or Case 4 holds for some $k \geq k_0$. It follows from the hypothesis assumption that $|I_n^{k_0}| = |I_n^k|$ and $|I_n^{k+1}| = |I_n^k| + i_a$. Since $i_a \geq 1$, it follows that $|I_n^{k+1}| \geq |I_n^k| + 1 = |N_{k_0}| + 1$, which contradicts $|I_n^k| = |I_n^{k_0}|$ for all $k \geq k_0$.

**C.2.** Suppose that Case 6 holds for some $k \geq k_0$. It follows that $|I_n^{k+}| = |I_n^k| + \max\{i_a, i_n - |I_n^k|\} \geq |I_n^k| + 1$ and $|I_n^{k+1}|$ is updated as $|I_n^{k+}|$. It follows that $|I_n^{k+1}| \geq |I_n^k| + 1 = |I_n^{k_0}| + 1$ which contradicts $|I_n^k| = |I_n^{k_0}|$ for all $k \geq k_0$. Thus, either Case 1, Case 2 or Case 5 in Algorithm 3 holds for an infinite sequence of iterations.

**C.3.** By using the same argument as in **B.3**, the event $A_1$ happens with probability zero.

**C.4.** By using a similar proof as in **B.4**, $\mathbb{P}(A_2) = 0$.

**C.5.** It follows from **C.4** in Lemma 2.4 that $\mathbb{P}(A_5) = 0$.

Thus, w.p.1 $|I_n^k| \to \infty$ as $k \to \infty$ for all $n \in \mathcal{N}$. □

**Theorem 2.5.** *Assume A.3, A.4, A.5, A.8 and A.9 and A.10. Let $\mathcal{C}$ be the nonempty compact convex set. Suppose furthermore that,* *(1)* $\inf_{\theta \in \mathcal{C}} f(\theta) = f^* > -\infty$,

*(2)* $\|H_k\|_{d \times d} \leq \max\{C, 1\} < \infty$ *for each* $k$,

*Then, w.p. 1*

$$\liminf_{k \to \infty} \chi(\theta_k) \quad = \quad 0.$$

**Proof**: Let $k$ denote the index of an iteration. We first show that w.p.1, for any $\varepsilon > 0$ there exists $K$ such that

$$f_k(\theta_k) \quad \geq \quad f^* - \varepsilon, \quad \forall k \geq K. \tag{2.33}$$

Fix $\varepsilon > 0$. It follows from Lemma 2.5 that $|N_k| \to |\mathcal{N}|$ and $|I_n^k| \to \infty$ for all $n \in \mathcal{N}$ as $k \to \infty$, and thus follows from Theorem 2.1 and condition (1) that there exists $K \in \mathbb{N}$ such that

$$\begin{aligned} f_k(\theta_k) \quad &= \quad f_k(\theta_k) - f(\theta_k) + f(\theta_k) \quad \geq \quad -|f_k(\theta_k) - f(\theta_k)| + f(\theta_k) \\ &\geq \quad -\sup_{\theta \in \mathcal{C}} |f_k(\theta) - f(\theta)| + f(\theta_k) \\ &\geq \quad f^* - \varepsilon, \end{aligned}$$

for all $k \geq K$. This proves that $f_k(\theta_k)$ is lower bounded for sufficiently large $K$.

Next, note that all the assumptions for Theorem 12.2.2 in [23] hold. It follows from Theorem 12.2.2 in [23] and condition 2 that there exits some constant $\kappa_{\text{mdc}} \in (0, 1)$ such that

$$\Delta m_k(\theta_k, \theta_k^{\text{GC}}) \quad \geq \quad \kappa_{\text{mdc}} \min\{\chi_k(\theta_k), 1\} \min\left\{\frac{\min\{\chi_k(\theta_k), 1\}}{C}, \Delta_k\right\}.$$

Thus, it ensures, with Theorem 6.4.5 in [23] and the lower boundedness of $f_k(\theta_k)$, that w.p.1

$$\liminf_{k \to \infty} \chi_k(\theta_k) \quad = \quad 0.$$

Consider any $\varepsilon > 0$. Theorem 3.2.8 of [23, pp.44–45] ensures that $\chi(g, \theta, 1)$ is continuous in $g = g_k$. Thus, there exists $\delta > 0$ such that $\|\nabla f(\theta_k) - g_k\| < \delta$ implies

$$|\chi(\theta_k) - \chi_k(\theta_k)| \quad = \quad |\chi(\nabla f(\theta_k), \theta_k, 1) - \chi(g_k, \theta_k, 1)| \quad < \quad \varepsilon.$$

Theorem 2.2 implies that, $\sup_{\theta \in \mathcal{C}} \|g_k(\theta) - \nabla f(\theta)\| \to 0$ as $k \to \infty$. Thus, there exists $K \in \mathbb{N}$ such that $\|\nabla f(\theta_k) - g_k\| < \delta$ holds for all $k \geq K$, and that

$$|\chi(\theta_k) - \chi_k(\theta_k)| \quad \leq \quad \varepsilon.$$

holds for all $k \geq K$, which implies $\liminf_{k \to \infty} |\chi(\theta_k) - \chi_k(\theta_k)| = 0$. It thus follows that

$$\chi(\theta_k) \quad = \quad |\chi(\theta_k) - \chi_k(\theta_k) + \chi_k(\theta_k)| \quad \leq \quad |\chi(\theta_k) - \chi_k(\theta_k)| + |\chi_k(\theta_k)||.$$

Taking liminf on both sides of the above inequality with respect to $k \to \infty$ gives that

$$\liminf_{k \to \infty} \chi(\theta_k) \quad = \quad 0.$$

This completes the proof. □

## *2.7 Numerical Studies: The ML Model Estimation*

The ML model estimation that involves the computation of high-dimensional integrals is one of the important applications of the STRA, as motivated in Section 2.1. In this section we use the STRA to estimate ML models with real choice datasets and compare it with existing algorithms.

In this section, we consider that the utility $u_{n,j}(x_{n,j}, y_{n,j}, \beta, \gamma_n)$ in (2.1) is of the following linear form,

$$u_{n,j}(x_{n,j}, y_{n,j}, \beta, \gamma_n) \quad = \quad \beta^\intercal x_{n,j} + \gamma_n^\intercal y_{n,j} + \varepsilon_{n,j},$$

where $n \in \mathcal{N}$ is the index of the data point corresponding to observation (customer) $n$, $j \in S_n$ is an alternative in the choice set $S_n$ of $n$, $x_{n,j} \in \mathbb{R}^{m_1}$ represents the vector of attribute values corresponding to the deterministic preference coefficient vector $\beta \in \mathbb{R}^{m_1}$, and $y_{n,j} \in \mathbb{R}^{m_2}$ is the vector of attribute values corresponding to the random preference coefficient vector $\gamma_n \in \mathbb{R}^{m_2}$.

We assume that $\gamma_n$ is the Gaussian vector with mean vector $\mu$ and covariance matrix $\Sigma$. By decomposition, we have

$$\Sigma \quad = \quad \sigma\sigma^T,$$

where $\sigma$ is a lower triangular matrix. Note that $\sigma$ is not unique for this the decomposition of such kind, unless we require all its diagonal entries to be positive and such a $\sigma$ will be unique and is called the Cholesky factor. Let $\xi_n$ be the $m_2-$dimensional standard Gaussian vector. We can represent the random coefficient $\gamma_n$ by

$$\gamma_n \quad = \quad \mu + \sigma\xi_n,$$

and $\theta := (\beta, \mu, \sigma)$ is the vector of parameter values to be estimated. Thus, the probability, $p_n(\theta)$, of customer $n \in \mathcal{N}$ choosing $j \in S_n$is given by the following ML model,

$$p_n(\theta) \quad := \quad \mathbb{E}_{\xi_n}[F_n(\theta, \xi)] \quad = \quad \mathbb{E}_{\xi_n}\left[\frac{\exp\{\beta^\intercal x_{n,j_n} + \mu^\intercal y_{n,j_n} + \xi_n^\intercal \sigma^\intercal y_{n,j_n}\}}{\sum_{j \in S_n} \exp\{\beta^\intercal x_{n,j} + \mu^\intercal y_{n,j} + \xi_n^\intercal \sigma^\intercal y_{n,j}\}}\right],$$

where $j_n \in S_n$ is the chosen alternative of customer $n$. The average log-likelihood of $\theta$ is defined in (2.2). As explained in Section 2.1, it is of great interest to us to estimate $\theta$ that maximizes the average log-likelihood, and the estimation problem (i.e., the MLE) can be considered as an optimization problem, with the objective function $f(\theta)$ defined in (2.3) and approximate objective function $f_{\mathcal{I}}^N$ defined in (2.4) and with $r(\cdot) = -\ln(\cdot)$.

In addition, we incorporate the constraint that the diagonal entries of $\sigma$ are nonnegative; otherwise, we will have $2^{m_2}$ estimates of $\sigma$ which all give the same $p_n(\theta)$, which means we have $2^{m_2}$ indistinguishable solutions. Thus, we consider to solve the constrained optimization model:

$$\begin{aligned} \min_{\theta} \quad & f(\theta) \\ \text{s.t.} \quad & 0 \leq \sigma_{i,i} \leq M, \quad \forall i \in \{1, 2, \ldots, m_2\}, \end{aligned} \tag{2.34}$$

where $M$ is a big number that we assume to be $2^{64}$ in our numerical studies. It follows from [7] that the ML model defined above satisfies A.3, A.4, A.5, A.8 and A.9 and A.10.

The STRA is tested with two real datasets: (i) MobiDrive data and (ii) the Airline data (described in Section 1.3). We first discuss the numerical results with the MobiDrive data.

### 2.7.1 The MobiDrive Data

In this section, we test the STRA using the MobiDrive data, and compare the estimation results with the estimation results obtained under AMLET (version 0.11.0) with the same data, where AMLET is a software tool developed by [6] to estimate ML and MNL models. We highlight below the two differences between the STRA and AMLET for the ML model estimation with the MobiDrive data.

(1) The STRA uses the sample allocation Algorithm 4 to adaptively update the sample size for each $n$ according to the integration sampling error incurred in computing $p_{I_n}(\theta)$ of observation $n$, while AMLET introduces the same maximum integration sample size $I_{\max}$ for all $n$.

(2) The STRA computes the Hessian of the approximate objective function (2.2) using the expression provided in Appendix B.3 for the ML model, while AMLET uses the approximation of the Hessian based on the BFGS method [50].

For the computational test with the MobiDrive data, the STRA is coded with C and both the STRA and AMLET are executed under CentOS Linux 7.0 on a PC with Intel i5 2.50GHz CPU and 8GB RAM.

The MobiDrive dataset describes the travel mode choice of travelers each of which could choose among from five possible alternatives: (i) car driver, (ii) car passenger, (iii) public transport, (iv) walk and (v) bike. The dataset involves 10 alternative-specific and traveler-specific factors, including urban household location, suburban household location, full-time worker, female and part-time, married with children, annual mileage by car, number of stops, time, cost and time budget. The data were originally collected from the six-week travel dairy in Karlsruhe and Halle (Germany) by [5], and later cleaned to only concentrate on Karlsruhe for a better data quality by [6]. The dataset contains a set $\mathcal{N}$ of 5799 observations and each observation corresponds to a traveler making travel mode choice, but it might be the case that a traveler cannot access the full set of five alternatives, i.e., a traveler's choice set could be a subset of the five alternatives. Readers can refer to [6] for more details about the MobiDrive data.

The ML model calibrated with this dataset has 14 parameters with three random members, time, cost and time budget, and the other 11 deterministic parameters of which there are four alternative-specific constants with car driver being the base. It thus follows that

$m_1 = 11$, $m_2 = 3$ and $\sigma \in \mathbb{R}^{3\times 3}$. We suppose furthermore that $\sigma$ is a diagonal covariance matrix. Thus, $\theta := (\beta, \mu, \sigma_{1,1}, \sigma_{2,2}, \sigma_{2,3}) \in \mathbb{R}^{17}$ is the vector of parameters to be estimated.

Both the STRA and AMLET are used to solve the optimization model (2.34) with the MobiDrive data and the same initial point. As the MobiDrive contains only a fair amount of data (5799 observations), we load the full set of data at the initial step of the STRA and do not resort to the data sampling process. Let $\widehat{\theta}^{k,j}_{\text{STRA}}$ denote the $k$-th last solution that is accepted at **Step 4** during executing Algorithm 1 (i.e., the STRA) with random seed $\#j$, where $k \in \{1, 2, \ldots\}$ and $j \in \{1, 2, \ldots, 10\}$. For all $k \geq 2$, we call $\widehat{\theta}^{k,j}_{\text{STRA}}$ the $k$-th last *intermediate estimate* with seed $\#j$. In addition, we refer $\widehat{\theta}^{1,j}_i$ to as the *stopping estimate* with seed $\#j$ obtained at the termination of algorithm $i \in \{\text{STRA, AMLET}\}$. To evaluate an estimate $\widehat{\theta}$, we feed $\widehat{\theta}$ into an approximate function defined in (2.4) with $N = \mathcal{N}$ and $|I_n| = 10,000$ for each $n \in \mathcal{N}$ and refer

$$f^*(\widehat{\theta}) \quad = \quad -\frac{1}{5799} \sum_{n \in \mathcal{N}} \ln\left(p_{I_n}(\widehat{\theta})\right)$$

to as the "true" log-likelihood of $\widehat{\theta}$. For any $k \in \{2, \ldots, \}$, define

$$\bar{f}^k_{\text{STRA}} \quad := \quad \frac{1}{10} \sum_{j=1}^{10} f^*(\widehat{\theta}^{k,j}_{\text{STRA}})$$

as the average true log-likelihood of the $k$-th last intermediate estimate obtained using the STRA, and define

$$\bar{f}^1_i \quad := \quad \frac{1}{10} \sum_{j=1}^{10} f^*(\widehat{\theta}^{1,j}_i)$$

as the average true log-likelihood of the stopping estimates computed using algorithm $i \in \{\text{STRA}, \text{AMLET}\}$.

The AMLET introduces a maximum integration sample size, $I_{\max}$, as input for computing $p_{I_n}(\theta)$ for all $n$, by which that the number of integration samples, $|I_n|$, is constrained to be bounded above by $I_{\max}$ for all $n \in \mathcal{N}$. Table 7 shows the true log-likelihoods of the stopping estimates with 10 seeds, computed using AMLET under nine maximum integration sample sizes, $I_{\max} = 100, 500, 1000, 2000, 3000, 4000, 6000, 8000, 10000$, and the corresponding computational cputimes.

The STRA, however, does not have a restriction on the number of integration samples used for computing $p_{I_n}(\theta)$, the SAA of the choice probability of each observation and, it adaptively controls the sample size for each observation according to the integration sampling error incurred in computing the SAA of that observation (see Algorithm 4). Table 8 shows the true log-likelihoods evaluated at both intermediate and stopping estimates with 10 seeds that are obtained during solving (2.34) by using the STRA, in which a starred entry denotes the stopping estimate with each of the seeds. Table 9 shows the cputime that has been consumed at obtaining each intermediate or stopping estimate with each of the 10 different seeds under the STRA (i.e., at **Step 4** in Algorithm 1). For example in Table 8, $f^*(\widehat{\theta}^{6,1}_{\mathrm{STRA}}) = 1.1654159$ in column two is the true log-likelihood of the 6-th last intermediate estimate with seed #1 obtained at **Step 4** in Algorithm 1 when running the STRA. Correspondingly, 112 in column two of Table 9 is the cputime spent when the 6-th last intermediate estimate, $\widehat{\theta}^{6,1}_{\mathrm{STRA}}$, is obtained.

Table 7: The true log-likelihoods of the stopping estimates with 10 different seeds estimated using AMLET and the corresponding estimation cputimes under nine different maximum integration sample sizes.

| Seed $\#j$ | $f^*(\widehat{\theta}^{1,j}_{\text{AMLET}})$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | $I_{\max} = 100$ | $I_{\max} = 500$ | $I_{\max} = 1000$ | $I_{\max} = 2000$ | $I_{\max} = 3000$ | $I_{\max} = 4000$ | $I_{\max} = 6000$ | $I_{\max} = 8000$ | $I_{\max} = 10000$ |
| 1 | 1.1646606 | 1.1646561 | 1.1646860 | 1.1646583 | 1.1646518 | 1.1646835 | 1.1646287 | 1.1646831 | 1.1646837 |
| 2 | 1.1646600 | 1.1647249 | 1.1646520 | 1.1646853 | 1.1646295 | 1.1646849 | 1.1646828 | 1.1646838 | 1.1646832 |
| 3 | 1.1646584 | 1.1646317 | 1.1646841 | 1.1646520 | 1.1647203 | 1.1646834 | 1.1646828 | 1.1646506 | 1.1646512 |
| 4 | 1.1646831 | 1.1646914 | 1.1646295 | 1.1647201 | 1.1646845 | 1.1646830 | 1.1646832 | 1.1646841 | 1.1646577 |
| 5 | 1.1648068 | 1.1646338 | 1.1646573 | 1.1646888 | 1.1646505 | 1.1646503 | 1.1646510 | 1.1647194 | 1.1647251 |
| 6 | 1.1646594 | 1.1646979 | 1.1646587 | 1.1646846 | 1.1646512 | 1.1646514 | 1.1646828 | 1.1646829 | 1.1646829 |
| 7 | 1.1647037 | 1.1646891 | 1.1646839 | 1.1647256 | 1.1646847 | 1.1646837 | 1.1646971 | 1.1646828 | 1.1647199 |
| 8 | 1.1647024 | 1.1646867 | 1.1646293 | 1.1646505 | 1.1646859 | 1.1646970 | 1.1646835 | 1.1646836 | 1.1646829 |
| 9 | 1.1647160 | 1.1647221 | 1.1646881 | 1.1646840 | 1.1647195 | 1.1646829 | 1.1646830 | 1.1646501 | 1.1646828 |
| 10 | 1.1647012 | 1.1646845 | 1.1647198 | 1.1647269 | 1.1647195 | 1.1646507 | 1.1646845 | 1.1646831 | 1.1646571 |
| Mean | 1.1646952 | 1.1646818 | 1.1646689 | 1.1646876 | 1.1646797 | 1.1646751 | 1.1646759 | 1.1646804 | 1.1646826 |

| Seed $\#j$ | Cputimes (seconds) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | $I_{\max} = 100$ | $I_{\max} = 500$ | $I_{\max} = 1000$ | $I_{\max} = 2000$ | $I_{\max} = 3000$ | $I_{\max} = 4000$ | $I_{\max} = 6000$ | $I_{\max} = 8000$ | $I_{\max} = 10000$ |
| 1 | 40.7 | 179.9 | 336.5 | 781.3 | 1047.0 | 1330.4 | 1969.8 | 2790.6 | 2970.4 |
| 2 | 39.7 | 156.7 | 341.6 | 709.4 | 935.0 | 1228.4 | 1728.7 | 2392.8 | 3310.7 |
| 3 | 41.8 | 173.7 | 338.8 | 664.9 | 956.0 | 1303.8 | 1957.4 | 2574.7 | 3410.1 |
| 4 | 41.6 | 177.4 | 372.2 | 626.6 | 1039.0 | 1329.9 | 1896.2 | 2544.8 | 3531.5 |
| 5 | 40.4 | 177.2 | 393.4 | 631.7 | 1001.0 | 1338.1 | 1905.4 | 2425.3 | 3425.9 |
| 6 | 42.5 | 188.7 | 328.5 | 680.1 | 953.0 | 1146.5 | 1906.9 | 2939.1 | 3255.0 |
| 7 | 41.5 | 168.6 | 351.8 | 632.7 | 983.0 | 1236.4 | 1934.6 | 2530.5 | 3356.3 |
| 8 | 41.7 | 165.1 | 352.3 | 668.6 | 982.0 | 1169.9 | 1889.2 | 2836.4 | 3276.6 |
| 9 | 42.5 | 170.3 | 340.8 | 662.1 | 941.0 | 1237.8 | 1979.3 | 2785.6 | 3182.7 |
| 10 | 41.4 | 183.0 | 399.5 | 779.6 | 1087.0 | 1206.6 | 1754.4 | 2684.5 | 3155.7 |
| Mean | 41.4 | 174.1 | 355.5 | 683.7 | 992.4 | 1252.8 | 1892.2 | 2650.4 | 3287.5 |

Table 8: The true log-likelihoods of intermediate and stopping estimates with 10 different seeds obtained using the STRA.

| $k$ | $f^*(\widehat{\theta}^{k,j}_{\text{STRA}})$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Seed #1 | Seed #2 | Seed #3 | Seed #4 | Seed #5 | Seed #6 | Seed #7 | Seed #8 | Seed #9 | Seed #10 |
| 24 | | | 1.4434307 | | | | | | | |
| 23 | | | 1.3830915 | | | | | | | 1.4424436 |
| 22 | | | 1.3306520 | | | | | | 1.4402711 | 1.3993964 |
| 21 | | | 1.3278011 | | | | | | 1.3924635 | 1.3403742 |
| 20 | | | 1.3350273 | | | 1.4564232 | | | 1.3339743 | 1.3061398 |
| 19 | 1.4420490 | 1.4436005 | 1.3539639 | 1.4479875 | | 1.4181882 | 1.4463727 | | 1.2998245 | 1.2876708 |
| 18 | 1.3845811 | 1.3861841 | 1.2600153 | 1.3892042 | 1.4419940 | 1.3594266 | 1.3848309 | 1.4472181 | 1.2932008 | 1.2790997 |
| 17 | 1.3280002 | 1.3302305 | 1.2429765 | 1.3361796 | 1.3618695 | 1.2748794 | 1.3321110 | 1.3820014 | 1.2278026 | 1.2547842 |
| 16 | 1.2970632 | 1.2993861 | 1.2249016 | 1.3132325 | 1.3171126 | 1.2517412 | 1.3101085 | 1.3282880 | 1.2239293 | 1.2498217 |
| 15 | 1.2453337 | 1.2384648 | 1.2167488 | 1.3170046 | 1.3153829 | 1.2731659 | 1.3124659 | 1.2989548 | 1.2078435 | 1.2345927 |
| 14 | 1.2298357 | 1.2231231 | 1.2070005 | 1.2447191 | 1.2903494 | 1.2171652 | 1.2270208 | 1.2398747 | 1.2044371 | 1.2186494 |
| 13 | 1.2247841 | 1.2026231 | 1.2018212 | 1.2234420 | 1.2607025 | 1.2071878 | 1.2163320 | 1.2560537 | 1.2033205 | 1.2047902 |
| 12 | 1.1968835 | 1.1725106 | 1.1823905 | 1.2043396 | 1.2119205 | 1.1730567 | 1.1680212 | 1.2024212 | 1.2028677 | 1.1795713 |
| 11 | 1.1685137 | 1.1668329 | 1.1674190 | 1.1713074 | 1.1860605 | 1.1669783 | 1.1674760 | 1.1713323 | 1.1948178 | 1.1678133 |
| 10 | 1.1665735 | 1.1666464 | 1.1669378 | 1.1675286 | 1.1684976 | 1.1662696 | 1.1673757 | 1.1669171 | 1.1719421 | 1.1672914 |
| 9 | 1.1661721 | 1.1665659 | 1.1664113 | 1.1669753 | 1.1666890 | 1.1661016 | 1.1666212 | 1.1659960 | 1.1670917 | 1.1668637 |
| 8 | 1.1657596 | 1.1661061 | 1.1663192 | 1.1666886 | 1.1661155 | 1.1657028 | 1.1661166 | 1.1660784 | 1.1665084 | 1.1663946 |
| 7 | 1.1656494 | 1.1658513 | 1.1662166 | 1.1660147 | 1.1657546 | 1.1652994 | 1.1659115 | 1.1656241 | 1.1663296 | 1.1658571 |
| 6 | 1.1654159 | 1.1657332 | 1.1655922 | 1.1651200 | 1.1654226 | 1.1652914 | 1.1657408 | 1.1654871 | 1.1659731 | 1.1657737 |
| 5 | 1.1651504 | 1.1651830 | 1.1652340 | 1.1650246 | 1.1650700 | 1.1649971 | 1.1650390 | 1.1649813 | 1.1654646 | 1.1654595 |
| 4 | 1.1646320 | 1.1645854 | 1.1646880 | 1.1649466 | 1.1646272 | 1.1646837 | 1.1646130 | 1.1646529 | 1.1653877 | 1.1647709 |
| 3 | 1.1646245 | 1.1645200 | 1.1646516 | 1.1647444 | 1.1646144 | 1.1646091 | 1.1646090 | 1.1646486 | 1.1647742 | 1.1647490 |
| 2 | 1.1646240 | 1.1645138 | 1.1646499 | 1.1647355 | 1.1646139 | 1.1646091 | 1.1646077 | 1.1646488 | 1.1647515 | 1.1647449 |
| 1 | 1.1646240* | 1.1645139* | 1.1646499* | 1.1647355* | 1.1646139* | 1.1646091* | 1.1646077* | 1.1646488* | 1.1647516* | 1.1647449* |

*: the true log-likelihood of a stopping estimate

Table 9: The cputimes for computing the intermediate and stopping estimates with 10 different seeds obtained using the STRA.

| $k$ | Cputime (seconds) | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Seed #1 | Seed #2 | Seed #3 | Seed #4 | Seed #5 | Seed #6 | Seed #7 | Seed #8 | Seed #9 | Seed #10 | Mean |
| 24 | | | 0 | | | | | | | | |
| 23 | | | 0 | | | | | | | 0 | |
| 22 | | | 0 | | | | | | 0 | 0 | |
| 21 | | | 0 | | | | | | 0 | 0 | |
| 20 | | | 0 | | | 0 | | | 0 | 0 | |
| 19 | 0 | 0 | 0 | 0 | | 0 | 0 | | 0 | 0 | |
| 18 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.0 |
| 17 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.0 |
| 16 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.1 |
| 15 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0.2 |
| 14 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0.3 |
| 13 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0.3 |
| 12 | 0 | 1 | 1 | 0 | 0 | 1 | 2 | 0 | 1 | 1 | 0.7 |
| 11 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 0.8 |
| 10 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1.0 |
| 9 | 1 | 49 | 2 | 2 | 1 | 1 | 287 | 1 | 1 | 1 | 34.6 |
| 8 | 2 | 71 | 2 | 2 | 1 | 1 | 427 | 1 | 2 | 2 | 51.1 |
| 7 | 2 | 93 | 2 | 3 | 2 | 3 | 566 | 493 | 2 | 22 | 118.8 |
| 6 | 112 | 116 | 482 | 3 | 2 | 4 | 706 | 734 | 2 | 33 | 219.4 |
| 5 | 168 | 138 | 728 | 4 | 32 | 572 | 845 | 972 | 2 | 44 | 350.5 |
| 4 | 224 | 160 | 966 | 597 | 46 | 850 | 984 | 1209 | 3 | 54 | 509.3 |
| 3 | 280 | 182 | 1196 | 900 | 563 | 1130 | 1122 | 1444 | 618 | 64 | 749.9 |
| 2 | 942 | 783 | 2941 | 1200 | 2300 | 2878 | 1729 | 3203 | 925 | 673 | 1757.4 |
| 1 | 2842 | 2532 | 4672 | 2918 | 4043 | 4635 | 3525 | 4974 | 2672 | 2424 | 3523.7 |

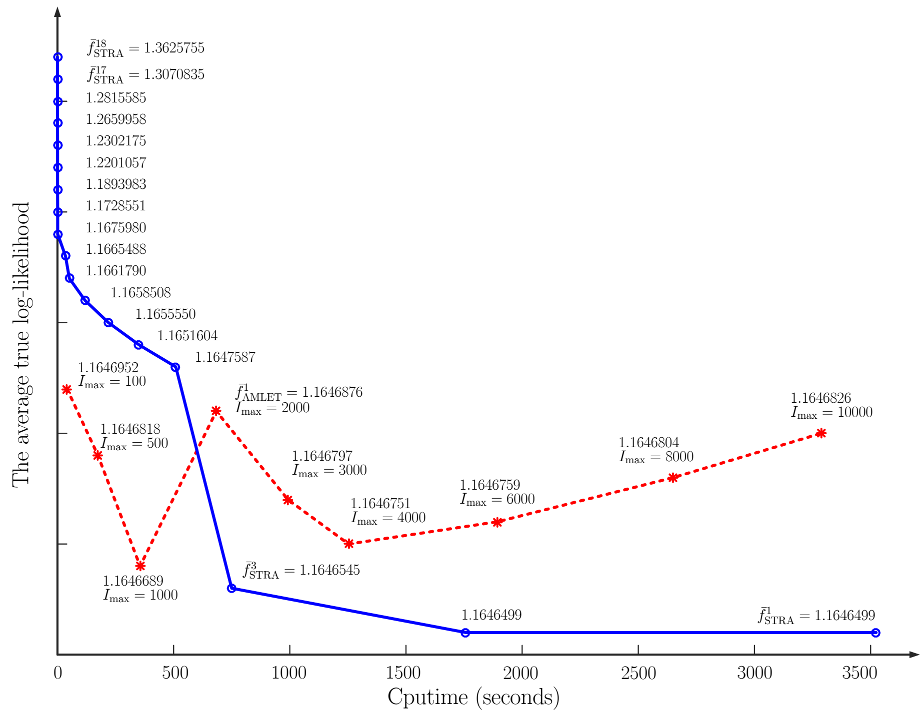



Figure 6: The STRA v.s. AMLET with the MobiDrive dataset.

We next compare the estimation results obtained using the STRA and AMLET. Figure 6 shows the average true log-likelihoods of the stopping estimates for both the STRA and AMLET, and the average true log-likelihoods of intermediate estimates for the STRA. The nine stars on the dashed line denote the average true log-likelihoods of the stopping estimates for AMLET under the nine different maximum integration sample sizes versus their corresponding mean cputimes shown in Table 7. The 18 circles located on the solid line represent average true log-likelihoods, $\bar{f}^{k}_{\text{STRA}}$, for $k \in \{1, 2, \dots, 18\}$ versus their corresponding mean cputimes shown in Table 9. Note that $\bar{f}^{1}_{\text{STRA}} = 1.1646499$ is the average true log-likelihood of the stopping estimates for the STRA.

As Figure 6 shows, the average true log-likelihood for the STRA decreases as the algorithm progresses until the stopping estimate is approached. The average true log-likelihood of the stopping estimates of the STRA is smaller than that of AMLET no matter which one of the nine maximum integration sample sizes is adopted in AMLET. This indicates that the STRA averagely generates a better stopping estimate than AMLET does as far as the stopping estimates are concerned. More interestingly, the STRA generates the estimates that give rise to smaller average true log-likelihoods than AMLET does as the computational time tends to be longer. In particular after 749.9s at which the third last intermediate estimate is obtained (in an average sense) under the STRA, the estimates of the STRA perform better than the stopping estimates that are obtained under AMLET. This shows that the STRA tends to produce a better estimate than AMLET when the two algorithms are terminated at the same stopping time no earlier than 749.9s.

### 2.7.2 The Airline Data

An additional difference between the STRA and AMLET is that the STRA embeds a data sampling process to handle large-scale datasets, which is not, however, implemented in AMLET. To test the STRA on a larger dataset, we use the 2011 airline data that we discussed earlier in Sections 1.3 and 1.5.1. The dataset has 326,148 observations and each observation has a choice set that could include from one to hundreds of alternatives. The ML model calibrated with this dataset is described in Section 1.5.2

For the large-scale application, both the STRA and SAA-50 are coded with Matlab and executed under Windows 7 on a PC with Intel i5 2.50GHz CPU and 8GB RAM.

For the large-scale data test, the STRA is stopped if the following criterion is satisfied,

$$\max\{\|g_k\|_2, \sigma_{k,1}, \sigma_{k,2}\} \quad \leq \quad \varepsilon \text{ and } |N_k| = |\mathcal{N}|$$

where $\varepsilon = 10^{-6}$.

The estimated parameter coefficients, $t$-statistics and $p$-values are summarized in Table 16 in Appendix C. In the table, $\widehat{\theta}_{\text{STRA}}$ represents the vector of estimated coefficients using the STRA and $\widehat{\theta}_{\text{SAA-50}}$ is the vector of estimated coefficients using the traditional SAA method [58] with a fixed integration sample size $|I_n| = 50$ for each $n$ with 2011 data, which we refer to as the SAA-50.

Note that there are $3 \times 3 \times 2 \times 5 = 90$ combinations of encoded number of days to departure, booking time-of-day, booking day-of-week, and booking channel. As explained in Section 1.5.1, we expect the price sensitivity of customers to depend on number of days to departure, booking time-of-day, booking day-of-week, and booking channel. Therefore, we estimate a separate price coefficient for each of the 90 combinations. One may refer to Section 1.5.1 for detailed discussions of the factors and attributes used for the estimation. In Table 16, each of the 90 combinations is represented by a combination of four codes given in Table 2, listed in the column headed "Attribute", with the associated price coefficient listed in the column headed "$\widehat{\theta}_{\text{STRA}}$" of Table 16. For example, the estimated coefficient $-12.05400$ for combination "$1, 1, 1, 2$" represents the price coefficient for customers who book [0,6] days before departure, between $00:00$ and $09:00$ on a weekday, and through the airline website. Also, the estimated coefficient 3.45670 for "XX–1–3 is the most expensive" represents the coefficient for a fully refundable ticket of airline XX, fare class 1 (the only fully refundable fare class), and airline call center channel. The estimated coefficient 2.94450 for "XX–13–2 is the cheapest" represents the coefficient for the cheapest ticket for an airline XX flight, fare class 13, and airline website channel. In addition, the estimated coefficient 0.41643 for "$\sigma_{2,1}$" represents entry $(2, 1)$ of the lower-triangular Cholesky factor of the covariance matrix for the ML model. The estimate parameter coefficients obtained using the SAA-50 are listed in the column headed "$\widehat{\theta}_{\text{SAA-50}}$". It can be observed from Table 16 that the price coefficients estimated using the STRA are more negative than the coefficients estimated using the SAA-50, thus rendering the estimation results to capture more price sensitivity and preference heterogeneity among customers.

Table 10 shows the computational times and approximate average log-likelihoods evaluated at the stopping coefficients estimated using the STRA and SAA-50 with the same initial point. The column headed "$n_{\text{sig}/90}$" shows the number of price coefficients out of 90 price coefficients of the ML model that are statistically significantly less than the corresponding price coefficients of the MNL model estimated with the same 2011 airline data (see Sections 1.4.1 and 1.6) at the 95% confidence level. The value $n_{\text{sig}/90}$ is obtained using the method described in Section 1.7.2. The results show that, compared with the SAA-50, the STRA achieves a much better solution by using significantly less computational time, and generates more price coefficient estimates that are statistically significantly less than the corresponding price coefficients of the MNL model. This is consistent with our intuition that the ML model should have more negative price coefficients than the MNL model since the ML model is structured to captures customer preference heterogeneity.

Table 10: Comparison between the STRA and the SAA-50 with the 2011 airline data.

| Algorithm | Cputime | Approximate average log-likelihood | $n_{\text{sig}/90}$ |
|---|---|---|---|
| STRA | 2.7075 days | 7.2172756 | 68 |
| SAA-50 | $>$ 30 days | 7.2617339 | 25 |

## *2.8 Conclusions*

The ML-type model estimation usually involves computing high-dimensional integrals, which excludes the possibility of using quadrature methods. In this chapter, we present an s-tochastic trust region algorithm (i.e., the STRA) to estimate ML-type choice models. The algorithm embeds a data sampling process and an integration sampling process under the framework of the trust region algorithm. The first process is used to sample from a large set of observations (data), and the second process is used to compute the SAA of the choice probability associated with each observation, which is expressed in terms of a high-dimensional integral.

During the sampling processes, the algorithm adaptively controls the data sample size and integration sample size, according to the magnitude of the sampling errors compared with the structural error between the approximate average log-likelihood and its model. We show that the algorithm converges to the first-order criticality points w.p. 1 and test the algorithm with two real datasets, the small-size MobiDrive dataset and the large-size Airline dataset. The numerical studies show that the STRA exhibits a competitive performance compared with AMLET, and it outperforms the traditional SAA-50 when it is applied to large-size estimation problems.

# CHAPTER III

# PRODUCT ASSORTMENT COMPETITION WITH THE DECOY EFFECT

The fraction of customers who choose a particular item from among a set of available items can be increased significantly by the inclusion of a related inferior (and apparently irrelevant) item in the choice set. This violation of the independence from irrelevant alternatives and the regularity properties is called the *decoy effect*, *dominance effect*, or *attraction effect*. The decoy effect is one of the robust cognitive biases in the decision-making processes of customers. We propose a discrete choice model that is simple and that captures decoy effects. A monopolist may take advantage of the decoy effect to increase profit. However, exploitation of the decoy effect in a competitive setting requires closer investigation. To understand the effect of decoys on competition, we study product assortment competition in a duopoly in which each seller may choose whether to include a decoy in the seller's product assortment. We provide a complete characterization of the Nash equilibria and their dependence on choice model parameters. We study the evolution of assortment competition and we evaluate the stability of the equilibria in the context of sellers learning about the behavior of their competitors. Our results indicate under what conditions it is beneficial for a seller to include a decoy into the seller's assortment, and under what conditions the seller obtains a free ride from the competitor's decoy. Our results also show that every pure-strategy Nash equilibrium is stable and every mixed-strategy Nash equilibrium is unstable.

## *3.1 Introduction*

It has been observed in many settings that human decision making deviates from axioms of rational choice [65, 48]. Some of these deviations are sufficiently widespread and predictable to be useful in forecasting aggregate choice outcomes, for example, in forecasting market shares. Therefore it may be a good idea for a seller to take such behavioral phenomena into account when designing a product portfolio or when choosing prices of products. In this paper we focus on one such behavioral phenomenon, called the *decoy effect*, *dominance effect*, or *attraction effect*, and we investigate the possible outcomes if two competing sellers both try to take the decoy effect into account when selecting their product portfolios.

The decoy effect refers to the phenomenon that the addition of an item to decision makers' choice sets significantly increases the market shares of other, usually similar but superior, items in the choice sets, while getting minimal market share itself. The item that serves this purpose to increase the market shares of target items is called a "decoy". The decoy effect is one of the robust cognitive biases in the decision-making processes of customers. It has been widely observed and demonstrated in both real-life and experimental choice situations. The following two examples are excerpted from [62], and [27].

*Example 1*: [62] provided an experimental example. In one setting, 106 people were each offered a choice between $6 and a Cross pen. In this setting, 36% of the people chose the pen and the remaining 64% chose the cash. In another setting, 115 people were each offered a choice among $6, a Cross pen, and another less attractive pen. In this setting, 46% of the people chose the Cross pen and 52% of them chose the cash. Only 2% of the people chose the less attractive pen.

*Example 2*: [27] investigated the sales of baked beans. Initially, the following two brands were put on the shelves of a local grocery store: 420-g Heinz baked beans for 29 pence each, and 420-g Spar baked beans for 21 pence each. After one week, Spar baked beans accounted for only 19% of sales, even though it was offered at a lower price. Then a third product, 220-g Spar baked beans for 21 pence each, was added (as a decoy). After another week, the market share of 420-g Spar baked beans increased to 33%. The authors finally concluded that the decoy effect "is robust, has a wide scope, is quite sizable and is of practical significance".

The decoy effect is one of the ways in which human decision making deviates from axioms of rational choice, such as Luce's choice axiom [42, p.6]. Here we show how the decoy effect violates some corollaries of Luce's choice axiom. One of the corollaries, called "independence from irrelevant alternatives (IIA)" or "proportionality", asserts that the ratio of choice probabilities of any two alternatives is independent of the presence or absence of a third alternative in the choice set [42, p.9]. Another corollary is that for each alternative there exists a nonnegative response strength such that for every choice set, the choice probability of each alternative in the choice set is equal to the ratio of the response strength of the alternative to the sum of the response strengths of all the alternatives in the choice set [42, p.23]. This corollary further implies the so-called "regularity" property in a choice context. The regularity property states that the choice probability of an alternative from a choice set cannot be increased by adding more alternatives to the choice set. Widely used choice models such as attraction models, including the multinomial logit (MNL) model, satisfy the IIA and regularity properties. However, the IIA and regularity properties are violated in choice settings with the decoy effect [67, 53], because if the decoy alternative is included in the choice set, then the choice probability of the target alternative increases, clearly violating the IIA and regularity properties.

An important decision made by retailers or revenue managers is to choose the set or assortment of products to offer to customers. As illustrated by Example 2, a revenue manager might choose an assortment that includes a decoy to increase the market share and/or revenue of a target item. The objective of this paper is to understand how decoy effects together with other demand characteristics impact equilibria of product assortment competition. In a competitive setting, there is quite a rich variety of possible outcomes resulting from the choices of sellers to include decoys in their assortments or not. For example, as illustrated by Example 2 above, if one seller includes a decoy and the other does not, then the first seller may gain revenue and the second seller may lose revenue. In such a setting a Nash equilibrium may be for both sellers to offer decoys. In a setting in which sellers have more products, it is also possible that a decoy introduced by one seller shifts demand from one product offered by that seller to another product (the target product) offered by that seller, and the same decoy also shifts demand among the other seller's products. In such a setting a Nash equilibrium may be for one seller to offer a decoy and the other seller not to offer decoys (but rather get a "free ride" on the first seller's decoy). It is also possible that the overall effect of a decoy, through a shift of demand among a seller's products as well as through its effect on demand for the competitor's products, is such that none of the sellers offers decoys in equilibrium.

We consider a duopoly in which both sellers have knowledge of the decoy effect and each one has to decide whether or not to add a decoy into the assortment offered to customers. To introduce the model and develop intuition, we first characterize the Nash equilibria for a setting with simple product sets in which each seller's product set contains only a target product and a decoy product. Thereafter we extend the results to assortment competition with general product sets including decoys. We characterize the conditions under which different possible outcomes hold. Under some conditions there are multiple Nash equilibria. To develop a better understanding regarding which of these equilibria are more reliable as predictors of the outcome of the product assortment competition, we study dynamical systems models of learning by the competitors, and establish which of the equilibria these systems converge to. Two widely used learning models, Cournot adjustment and fictitious play, are employed to analyze the dynamic behavior of sellers' decisions when the sellers try to learn the strategies of their competitors.

### 3.1.1 Contributions

This paper makes the following contributions:

(1) First, we propose a modified attraction discrete choice model that captures the effect of a seller's decoy on the market shares of the seller's own products (the intra-decoy effect) as well as the market shares of other sellers' products (the inter-decoy effect). To the best of our knowledge, this is the first choice model to explicitly incorporate context-dependent behavioral effects such as the decoy effect, the similarity effect, and the compromise effect [53] in a competitive setting.

(2) Second, we use the modified attraction discrete choice model to provide a complete characterization of the pure- and mixed-strategy Nash equilibria for assortment competition in a duopoly.

(3) Third, to evaluate the stability of the equilibria, we consider two learning processes, Cournot adjustment and fictitious play, of the sellers in assortment competition, and

characterize the behavior of the resulting dynamical systems. Part of this analysis also gives a geometric characterization of the dynamics of fictitious play for general $2 \times 2$ games that is more complete and easier to follow than previous results for such games.

### 3.1.2 Managerial Insights

This paper also contributes the following managerial insights:

(1) In the setting with simple product sets, if the intra-decoy effects dominate the inter-decoy effects (in a sense specified in Section 3.4.1), then both sellers include decoys in their assortments. In other words, if introducing a decoy benefits a seller much more than its competitor, then the seller offers an assortment with the decoy. On the other hand, if the inter-decoy effects dominate the intra-decoy effects, then no seller includes a decoy in its assortment. If the intra- and inter-decoy effects are "approximately equal", then the equilibrium consists of one seller offering a decoy and the other seller not offering a decoy.

(2) In some cases a mixed-strategy Nash equilibrium coexists with two pure-strategy Nash equilibria. In these cases, each pure-strategy Nash equilibrium is a steady state of the Cournot adjustment process (but a mixed-strategy Nash equilibrium cannot be a steady state of a Cournot adjustment process). If the two sellers choose an initial strategy profile that is not a pure-strategy Nash equilibrium, then the Cournot adjustment process cycles. The long-run decision frequencies of the Cournot adjustment process do not correspond to any mixed-strategy Nash equilibrium, nor any correlated equilibrium, nor any coarse correlated equilibrium.

(3) In the cases where a mixed-strategy Nash equilibrium coexists with two pure-strategy Nash equilibria, each pure-strategy Nash equilibrium is a steady state of the fictitious play process, but the mixed-strategy Nash equilibrium is not a steady state of the fictitious play process. In cases with simple action sets (either include a decoy or not), the fictitious play process always converges regardless of the initial strategy profile

chosen by the sellers. In the case of symmetric sellers, and for special initial strategy profiles that we explicitly give, the fictitious play process converges to the mixed-strategy Nash equilibrium. In all other cases the fictitious play process converges to one of the pure-strategy Nash equilibria.

(4) In the sense outlined above, the pure-strategy Nash equilibria are reasonable predictors of the outcome of assortment competition with decoys, but it is unlikely that sellers will settle on one of the mixed-strategy Nash equilibria.

The remainder of this paper is organized as follows. The related literature is reviewed in Section 3.2. In Section 3.3, we present a modified attraction model of consumer choice that includes the decoy effect, and the duopoly model for product assortment competition with the decoy effect. We characterize the Nash equilibria and study the dynamics of two dynamical systems models of learning, for competition with simple product sets in Section 3.4, for competition with general product sets and simple actions in Section 3.5, and for competition with general product sets and general actions in Section 3.6. Conclusions are summarized in Section 3.7. All the proofs and supporting material are provided in the Appendix A to this paper.

## *3.2 Literature Review*

We classify the literature related to our research into four branches: (1) empirical studies that identify and investigate the nature of the decoy effect, (2) choice models that explicitly incorporate the decoy effect, (3) models of assortment planning and competition, and (4) models of learning in games.

### 3.2.1 Empirical Studies of the Decoy Effect

Since [36] and [37] identified the decoy effect, and recognized that it violated the IIA and regularity properties that are implied by some rational choice models, there have been many experimental studies that identified the decoy effect in consumer product choice [61, 62, 67, 32]. The decoy effect is deemed to be robust in the sense that it has been observed in a variety of choice settings, ranging from in-store grocery purchases [27] to on-line subscriptions [4], regardless of whether the products in the choice sets are traded in a market or not [8], or whether the decision makers are humans or honey bees [57]. [2] used experimental studies to show that the decoy effect can be used to facilitate coordination among players to reach efficient outcomes.

### 3.2.2 Models Capturing the Decoy Effect

Compared with the large number of empirical studies, relatively few papers have proposed choice models that incorporate the decoy effect. [67] modeled the decoy effect (called the local context effect) by adding the relative advantages of the target product over all other alternatives in the same choice set to the context-free utility of the target. It is assumed that the context-free utility of a product has an additive representation as the sum of functions measuring the contribution of each attribute to the total utility of the product. Then the advantage of a product over another product is given by the sum of the nonnegative differences of the attribute function values of the two products. It is shown that the model is able to account for the decoy effect and some other context-dependent effects. [53] proposed a unified utility model to incorporate context effects including the compromise, decoy (attraction), and similarity effects. The contribution of the decoy effect to the utility of the target is modeled as the distance between the attribute points of the (dominated) decoy and the (dominating) target parallel to a preference vector in the attribute space, where the preference vector [66] is chosen to point from the least desirable attribute point to the most desirable attribute point in the attribute space. These studies model the total utility of an alternative as the sum of a context-free utility of the alternative and various utility increments contributed by context effects such as the decoy effect, by using pairwise comparisons of attribute values.

### 3.2.3 Assortment Decision Models

In recent years various studies of both static and dynamic assortment decision problems have appeared. [39] provide a review of static assortment problems and [68] provide a review of dynamic assortment problems. A number of recent papers, such as [56, 41, 25, 26] and [29], have addressed assortment optimization problems under a variety of discrete choice models.

Relatively few papers on assortment competition have appeared. [35] considered price, service quality, and product assortment competition based on the logit model that satisfies the IIA property. [18] studied how the adoption of search-facilitating technologies, such as the internet, affects equilibrium prices and assortments in a game with competing sellers. [11] considered both assortment-only competition and joint price and assortment competition between two retailers subject to a constraint that each retailer can offer at most a certain number of products, and characterized conditions for existence and uniqueness of a Nash equilibrium. [40] considered assortment and price competition under nested logit models.

We are particularly interested in assortment decision models that explicitly incorporate specific features of consumer choice behavior. One such feature is the satiation effect, that is the phenomenon that consumers' marginal utilities for a product tend to decrease as more of the product is consumed. [20] incorporated the satiation effect into an attraction model of demand, and used the model to analyze price and assortment competition among sellers. Their paper highlighted the importance of incorporating this feature of consumer behavior into assortment decision problems. Our paper focuses on assortment competition with the decoy effect, one of the context-dependent effects in consumer choice. The papers reviewed here focus on describing equilibrium behavior. In addition to this, we also evaluate the stability of the equilibria by studying dynamical systems describing sequences of decisions made by competing sellers under dynamic learning.

### 3.2.4 Learning Processes in Games

There is a large literature on learning in games that is related to our research. See [28] for an overview of earlier literature on learning in games. We specifically mention the results of [52], [44], and [46], who established the convergence of discrete-time fictitious play for respectively any two-person finite-action zero-sum game, any non-degenerate nonzero-sum $2 \times 2$ game, and any common-interest game. [45] showed that fictitious play for a $2 \times 2$ game converges to a Nash equilibrium, given that each player starts with a degenerate probability distribution for the other player's actions. [30] showed that continuous-time fictitious play for a two-person zero-sum game converges uniformly at rate $1/t$.

The dynamic behavior of fictitious play for general $3 \times 3$ (or $m \times n$ with $m, n \geq 3$) games can be much more complicated. Fully characterizing it remains a challenge. [60] argued that fictitious play for a $3 \times 3$ game with payoff matrices that satisfy certain conditions cycle without convergence to any Nash equilibrium. More recently, [69] considered a $3 \times 3$ bimatrix game with a particular structure and showed that fictitious play can exhibit periodic or even chaotic behavior.

Most of these studies are aimed at verifying whether or not fictitious play converges to a Nash equilibrium of the game. In addition, our paper provides a complete, but simple geometric, characterization of the dynamics of fictitious play in $2 \times 2$ games. In the process we also extend the results of [45] to the more general setting in which each player starts with an arbitrary probability distribution for the other player's actions. Also, our proofs cover some cases that were missed in the proof in [45]. We also provide an example of assortment competition with the decoy effect, in which each seller chooses among three assortments, and fictitious play cycles without converging to any Nash equilibrium.

## *3.3 Model of Buyer Choice and Seller Competition*

First we present a general choice model that incorporates contexts effects in Section 3.3.1, and then we present a particular version of it for the decoy effect in Section 3.3.2.

### 3.3.1 A General Choice Model Incoporating Context Effects

Let $S$ denote a set of alternatives of interest, for example, $S$ may represent a set of products that can be offered to customers (e.g., airline tickets for different fare classes). If a no-purchase alternative is of interest, then it is denoted with 0, and is understood to be included in $S$. For any assortment $A \subset S$ that can be offered to customers, the fraction of customers choosing (or the choice probability of) $i \in A$ can be written as

$$q_i(A) \quad = \quad \frac{\gamma_i(A) v_i}{\sum_{k \in A} \gamma_k(A) v_k} \tag{3.1}$$

Choice model (3.1) is general, because for any given choice probabilities $q_i(A)$ and $v_i > 0$ for all $A$ and $i$, one can set $\gamma_i(A) = q_i(A)/v_i$. The idea behind this model is to interpret $v_i > 0$

as the intrinsic attractiveness (the context-effect-free attractiveness) of alternative $i \in S$, and $\gamma_i(A)$ as factors that capture context effects (such as the decoy effect). The intrinsic attractiveness can be thought of as a function of the attribute values of alternative $i$ and the decision maker (but not of the other alternatives in the choice set). One can assume, without loss of generality, that intrinsic attractiveness is scaled in such a way that the intrinsic attractiveness of one of the alternatives, such as the no-purchase alternative, is normalized to be unity, i.e., $v_0 = 1$. An important special case of choice model (3.1) is the attraction demand model with $\gamma_i(A) = 1$ for all $A \subset S$ and $i \in A$. Attraction demand models satisfy the IIA and regularity properties. The attraction demand model includes the multinomial logit (MNL) model and the multiplicative competitive interaction (MCI) model as special cases. An example of a multinomial logit model with price $p_i$ as the only attribute is choice model (3.1) with $\gamma_i(A) = 1$ and $v_i = e^{\theta_0 - \theta p_i}$ for all $A \subset S$ and $i \in A \setminus \{0\}$, where $\theta_0$ is a constant parameter and $\theta \geq 0$ represents the price sensitivity of customers.

### 3.3.2 A Buyer Choice Model Incorporating the Decoy Effect

First, consider the decoy effect in a monopoly setting. Of particular interest is a target product $t \in S$ and an associated decoy product $d \in S$ ($d \neq t$). The target product dominates the decoy product in terms of all attributes. The target product may be of particular interest to the seller for many reasons, for example because it is a lucrative product, or because the seller wants to increase its market share, or because it is a new product and the seller needs to quickly build its brand reputation. To study the decoy effect, choice model (3.1) with

$$\gamma_i(A) \begin{cases} > 1 & \text{if} \quad i = t \in A, d \in A, \\ = 0 & \text{if} \quad i = d \in A, t \in A, \\ = 1 & \text{for} \quad \text{all other cases} \end{cases}$$

for all $A \subset S$ and $i \in A$, is of particular interest. Note that the IIA and regularity properties do not hold for the resulting choice model.

Next we present a choice model that incorporates the decoy effect for a duopoly of two sellers who offer substitutable products (e.g., two airlines); the extension to more than two sellers is straightforward with additional notation. The sellers are indexed by $m \in \{-1, 1\}$. Each seller $m$ has a set $S_m$ of products that can be offered to customers. Each seller $m$ has a target product $t_m \in S_m$ and an associated decoy product $d_m \in S_m$ ($d_m \neq t_m$). As before, the intrinsic attractiveness of product $i$ is denoted by $v_i > 0$, and the no-purchase alternative is denoted by 0. Let $\cup S := \{0\} \cup S_{-1} \cup S_1$, and for any given $A_m \subset S_m$, $m = \pm 1$, let $\cup A := \{0\} \cup A_{-1} \cup A_1$. For any $\cup A$ and $i \in \cup A$, let $\gamma_i(\cup A)$ denote the factor that represents the context effect of choice set $\cup A$ on product $i$. Thus, given an assortment $A_m \subseteq S_m$ for each seller $m$, the probability of a customer choosing $i \in \cup A$ is given by

$$q_i(A_{-1}, A_1) \quad = \quad \frac{\gamma_i(\cup A) v_i}{\sum_{k \in \cup A} \gamma_k(\cup A) v_k} \,. \tag{3.2}$$

In addition to the effect of each seller's decoy on the demand for the seller's target, we are also interested in the effect of each seller's decoy on the demand for the other seller's target, since in competitive applications the target products of the sellers may be similar. Thus, of particular interest are the effect of the decoy of seller $m$ on the demand for the target of seller $m$, which we call the intra-decoy effect, and the effect of the decoy of seller $-m$ on the demand for the target of seller $m$, which we call the inter-decoy effect. Therefore, to study the decoy effect, we will consider choice model (3.2) with

$$\gamma_i(\cup A) \quad = \quad \begin{cases} \alpha_m > 1 & \text{if} \quad i = t_m \in \cup A, d_m \in \cup A, d_{-m} \notin \cup A, \\ \beta_{-m} \geq 1 & \text{if} \quad i = t_m \in \cup A, d_m \notin \cup A, d_{-m} \in \cup A, \\ \alpha_m \beta_{-m} > 1 & \text{if} \quad i = t_m \in \cup A, d_m \in \cup A, d_{-m} \in \cup A, \\ 0 & \text{if} \quad i = d_m \in \cup A, (t_m \in \cup A \text{ or } t_{-m} \in \cup A), \\ 1 & \text{for} \quad \text{all other } i \in \cup A, \end{cases}$$

where $\alpha_m$, $m = \pm 1$, are called the intra-decoy factors, and $\beta_m$, $m = \pm 1$, are called the inter-decoy factors.

### 3.3.3 Model of Assortment Competition

In this section we formulate each seller's assortment problem. We assume that each product $i$ has a given excess $p_i > 0$ of price over marginal cost, and that each seller's objective is to maximize the seller's total profit. (Similar to many airlines, one may choose multiple price classes for essentially the same product, and consider each product-price combination as a "product" in the model.) Then, the objective function $\pi_m$ of each seller $m$ is a function of the assortments $(A_m, A_{-m})$ chosen by the sellers, and is given by

$$\pi_m(A_m, A_{-m}) \quad := \quad \sum_{i \in A_m} p_i \, q_i(A_{-1}, A_1) \quad = \quad \sum_{i \in A_m} p_i \, \frac{\gamma_i(\cup A) v_i}{\sum_{k \in \cup A} \gamma_k(\cup A) v_k}.$$

For each seller $m$, let $C_m \subset 2^{S_m}$ denote the set of feasible actions (pure strategies) of seller $m$, where $2^S$ denotes the collection of subsets of a set $S$. We assume that each seller always offers its target product, i.e., we consider action sets of seller $m$ that satisfy $C_m \subset \{A_m \in 2^{S_m} : t_m \in A_m\}$. Each action $A_m \in C_m$ denotes an assortment that seller $m$ can choose to offer to the market. Let $\Delta(C_m)$ denote the set of probability distributions on set $C_m$, that is, $\Delta(C_m)$ denotes the set of mixed-strategies over action set $C_m$. Thus, for each $\bar{x}_m \in \Delta(C_m)$ and $A_m \in C_m$, $\bar{x}_m(A_m)$ denotes seller $m$'s probability of choosing action $A_m$. Let $C := C_{-1} \times C_1$ denote the set of joint actions, and let $\Delta(C)$ denote the set of (joint) distributions over $C$.

Let $\mathtt{BR}_m : \Delta(C_{-m}) \mapsto 2^{\Delta(C_m)}$ denote the mixed best response correspondence of seller $m$ given by

$$\mathtt{BR}_m(\bar{x}_{-m}) \quad := \quad \arg\max_{\bar{x}_m \in \Delta(C_m)} \sum_{(A_{-1}, A_1) \in C} \pi_m(A_m, A_{-m}) \bar{x}_{-1}(A_{-1}) \bar{x}_1(A_1)$$

and $\mathtt{PBR}_m : \Delta(C_{-m}) \mapsto 2^{C_m}$ denote the pure best response correspondence of seller $m$ given by

$$\mathtt{PBR}_m(\bar{x}_{-m}) \quad := \quad \arg\max_{A_m \in C_m} \sum_{A_{-m} \in C_{-m}} \pi_m(A_m, A_{-m}) \bar{x}_{-m}(A_{-m}).$$

If $\bar{x}_{-m}(A_{-m}) = 1$ for some $A_{-m} \in C_{-m}$, then we also write $\mathtt{PBR}_m(A_{-m})$ for $\mathtt{PBR}_m(\bar{x}_{-m})$. A mixed-strategy profile $\bar{x}^* = (\bar{x}^*_{-1}, \bar{x}^*_1) \in \Delta(C_{-1}) \times \Delta(C_1)$ is called a mixed-strategy Nash equilibrium if $\bar{x}^*_m \in \mathtt{BR}_m(\bar{x}^*_{-m})$ for all $m$, and a pure-strategy profile $(A^*_{-1}, A^*_1) \in C$ is called a pure-strategy Nash equilibrium if $A^*_m \in \mathtt{PBR}_m(A^*_{-m})$ for all $m$. Note that for any $m$, any $\bar{x}_{-m} \in \Delta(C_{-m})$, any $\bar{x}^*_m \in \mathtt{BR}_m(\bar{x}_{-m})$, and any $A^*_m$ such that $\bar{x}^*_m(A^*_m) > 0$, it holds that $A^*_m \in \arg\max_{A_m \in C_m} \sum_{A_{-m} \in C_{-m}} \pi_m(A_m, A_{-m}) \bar{x}_{-m}(A_{-m})$, that is, each action with positive probability under a mixed best response must be a pure best response to $\bar{x}_{-m}$.

It follows that all actions $A_m^*$ such that $\bar{x}_m^*(A_m^*) > 0$ have the same payoff for seller $m$, and $\mathtt{BR}_m(\bar{x}_{-m})$ is the convex hull of $\arg\max_{A_m \in C_m} \sum_{A_{-m} \in C_{-m}} \pi_m(A_m, A_{-m}) \bar{x}_{-m}(A_{-m})$ in $\Delta(C_m)$. If there exists joint actions $(A_{-1}^*, A_1^*) \in C$ such that $\bar{x}_m^*(A_m^*) = 1$ for all $m$, then $A_m^* \in \mathtt{PBR}_m(A_{-m}^*)$ for all $m$, i.e., $(A_{-1}^*, A_1^*)$ is a pure-strategy Nash equilibrium.

When we consider convergence of strategy profiles, we will also consider the notions of correlated and coarse correlated Nash equilibria. A probability distribution $\bar{x}^* \in \Delta(C)$ is called a correlated equilibrium if for all $m$ and all $A_m, A_m' \in C_m$ it holds that

$$\sum_{A_{-m} \in C_{-m}} \pi_m(A_m, A_{-m}) \bar{x}^*(A_{-1}, A_1) \quad \geq \sum_{A_{-m} \in C_{-m}} \pi_m(A_m', A_{-m}) \bar{x}^*(A_{-1}, A_1). \qquad (3.3)$$

A probability distribution $\bar{x}^* \in \Delta(C)$ is called a coarse correlated equilibrium if for all $m$ and all $A_m' \in C_m$ it holds that

$$\sum_{(A_{-1}, A_1) \in C} \pi_m(A_m, A_{-m}) \bar{x}^*(A_{-1}, A_1) \quad \geq \sum_{(A_{-1}, A_1) \in C} \pi_m(A_m', A_{-m}) \bar{x}^*(A_{-1}, A_1). \qquad (3.4)$$

It is easy to verify that each Nash equilibrium is a correlated equilibrium, and each correlated equilibrium is a coarse correlated equilibrium.

### 3.3.4 Models of Seller Learning

It is customary in much of the literature on applications of non-cooperative game theory to identify equilibria, but not to address the question whether there is reason to be confident that the players will settle on a specific equilibrium. This question is relevant even when there is a unique equilibrium, and is especially pertinent when there are multiple equilibria. In this paper we approach this question by considering processes in which the sellers repeatedly make assortment decisions while they learn about the other sellers' assortment decisions. The idea is that if the sellers' decisions converge to an equilibrium, then the equilibrium may be a reasonable prediction of the sellers' long-run behavior, and if the sellers' decisions do not converge to an equilibrium, then the equilibrium is a questionable prediction of the sellers' decisions.

In general, sellers may learn about many things, including the behavior of their customers (i.e., their demand) and their competitors, and their costs. In this paper, we restrict attention to sellers who learn about the other sellers' assortment choices. We consider a discrete-time process with time indices $t \in \mathbb{N} := \{1, 2, \ldots\}$. At each time $t$, each seller $m$ has observed all previous actions $A_{-m}(0), \ldots, A_{-m}(t-1)$ of the other seller. Each seller $m$ then chooses an action $A_m(t)$ that is a best response to the seller's forecast of the action that the other seller is about to take, where the forecast is based on the observed data $A_{-m}(0), \ldots, A_{-m}(t-1)$. Thereafter these steps repeat at the next time $t+1$. We are interested in answering the following questions for the resulting process:

Q.1. What are the steady states (fixed points) of the process? Specifically, do these steady states coincide with the equilibria?

Q.2. Are these steady states stable? Specifically, does the process converge, and if so, does it converge to an equilibrium?

Q.3. How does the initial state affect the long-run behavior of the process?

## *3.4 Assortment Competition with Simple Product Sets*

To facilitate explanation of the results, we first consider the setting with *simple product sets* $S_m = \{t_m, d_m\}$ in this section, and thereafter we consider the setting with general product sets in Sections 3.5 and 3.6.

### 3.4.1 Characterization of Equilibria

In this section we consider assortment competition with the decoy effect for the setting in which each seller has a simple product set $S_m = \{t_m, d_m\}$. The set of actions is $C_m = \{A_m^0, A_m^1\}$, where $A_m^0 = \{t_m\}$ and $A_m^1 = \{t_m, d_m\}$. The four possible action pairs are $(A_{-1}^0, A_1^0)$, $(A_{-1}^1, A_1^1)$, $(A_{-1}^0, A_1^1)$, and $(A_{-1}^1, A_1^0)$. Next we give necessary and sufficient conditions in terms of the decoy factors for each of these action pairs to be a pure-strategy Nash equilibrium. Intuitively, if a seller's inter-decoy factor is small relative to the seller's intra-decoy factor, then it is attractive to the seller to use the decoy, and if the seller's inter-decoy factor is relatively large, then it is attractive not to use the decoy. The following thresholds for the inter-decoy factors determine what is relatively small and relatively large:

$$\underline{\beta}_m \;:=\; \alpha_m + \frac{\alpha_m - 1}{\alpha_{-m} v_{t_{-m}}} \quad \text{and} \quad \bar{\beta}_m \;:=\; \alpha_m + \frac{\alpha_m - 1}{v_{t_{-m}}}.$$

Note that the thresholds are well defined since $\alpha_m, \alpha_{-m} > 1$ and $v_i > 0$ for all $i$. Also, note that $1 < \underline{\beta}_m < \bar{\beta}_m$. Proposition 3.1 gives necessary and sufficient conditions for the different action pairs to be pure-strategy Nash equilibria.

**Proposition 3.1.** *For assortment competition with simple product sets, it holds that*

*(1)* $(A^0_{-1}, A^0_1)$ *is a pure-strategy Nash equilibrium iff* $\beta_m \geq \bar{\beta}_m$ *for* $m = \pm 1$,

*(2)* $(A^1_{-1}, A^1_1)$ *is a pure-strategy Nash equilibrium iff* $\beta_m \leq \underline{\beta}_m$ *for* $m = \pm 1$,

*(3)* $(A^0_{-1}, A^1_1)$ *is a pure-strategy Nash equilibrium iff* $\beta_{-1} \geq \underline{\beta}_{-1}$ *and* $\beta_1 \leq \bar{\beta}_1$,

*(4)* $(A^1_{-1}, A^0_1)$ *is a pure-strategy Nash equilibrium iff* $\beta_{-1} \leq \bar{\beta}_{-1}$ *and* $\beta_1 \geq \underline{\beta}_1$.

Each seller has a finite action set in assortment competition, and thus there always exists a mixed-strategy Nash equilibrium. Next, Proposition 3.2 provides a sufficient condition for existence of a strict mixed-strategy Nash equilibrium. For assortment competition with simple product sets, a probability distribution $\bar{x}_m = (\bar{x}_m(A^0_m), \bar{x}_m(A^1_m))$ is specified by $x_m := \bar{x}_m(A^1_m)$.

**Proposition 3.2.** *For assortment competition with simple product sets, there exists a mixed-strategy Nash equilibrium* $(\bar{x}^*_{-1}, \bar{x}^*_1) \in (0,1)^2$ *if and only if* $\underline{\beta}_m < \beta_m < \bar{\beta}_m$ *for* $m = \pm 1$. *The unique such Nash equilibrium* $(\bar{x}^*_{-1}, \bar{x}^*_1) \in (0,1)^2$ *is given by* $(x^*_{-1}, x^*_1) := (1/(1+\Gamma_{-1}), 1/(1+\Gamma_1))$, *where*

$$\begin{aligned} \Gamma_m &:= \frac{\alpha_m \beta_m (1 + v_{t_{-m}} + v_{t_m})(1 + \alpha_{-m} v_{t_{-m}} + \beta_{-m} v_{t_m})(\underline{\beta}_{-m} - \beta_{-m})}{(1 + \alpha_{-m}\beta_m v_{t_{-m}} + \alpha_m \beta_{-m} v_{t_m})(1 + \beta_m v_{t_{-m}} + \alpha_m v_{t_m})(\beta_{-m} - \bar{\beta}_{-m})} \\ &\in (0, \infty) \end{aligned} \tag{3.5}$$

*for* $m = \pm 1$.

We summarize the results of Propositions 3.1 and 3.2 in Table 11. Table 11 and Remark 3.1 give a complete characterization of the Nash equilibrium for simple product sets. Table 11 shows 9 cases specified by comparing $\beta_m$ versus $\underline{\beta}_m$ and $\bar{\beta}_m$ for $m = \pm 1$, and the resulting equilibria. In principle, the four comparisons give 16 cases. The cases that are not listed are not possible because $\underline{\beta}_m < \bar{\beta}_m$ for $m = \pm 1$.

| # | Sufficient condition | | | | Nash equilibrium |
|---|---|---|---|---|---|
| | $\underline{\beta}_{-1}$ v.s. $\beta_{-1}$ | $\bar{\beta}_{-1}$ v.s. $\beta_{-1}$ | $\underline{\beta}_{1}$ v.s. $\beta_{1}$ | $\bar{\beta}_{1}$ v.s. $\beta_{1}$ | |
| 1 | $<$ | $>$ | $<$ | $>$ | $(x^*_{-1}, x^*_1)$, $(A^0_{-1}, A^1_1)$, $(A^1_{-1}, A^0_1)$ |
| 2 | $>$ | $>$ | $>$ | $>$ | $(A^1_{-1}, A^1_1)$ |
| 3 | $<$ | $>$ | $>$ | $>$ | $(A^0_{-1}, A^1_1)$ |
| 4 | $<$ | $<$ | $>$ | $>$ | |
| 5 | $<$ | $<$ | $<$ | $>$ | |
| 6 | $>$ | $>$ | $<$ | $>$ | $(A^1_{-1}, A^0_1)$ |
| 7 | $>$ | $>$ | $<$ | $<$ | |
| 8 | $<$ | $>$ | $<$ | $<$ | |
| 9 | $<$ | $<$ | $<$ | $<$ | $(A^0_{-1}, A^0_1)$ |

Table 11: Characterization of the Nash equilibria for assortment competition with simple product sets.

**Remark 3.1.** *For each case in Table 11, the equilibrium/equilibria are the only equilibria if the four strict inequalities hold. If we change one or two of the strict inequalities to equalities, then one of the following cases holds:*

1. *The resulting conditions are impossible. For example, in Case 9, if the first strict inequality $\underline{\beta}_{-1} < \beta_{-1}$ is changed to the equality $\underline{\beta}_{-1} = \beta_{-1}$, then the resulting conditions are impossible, because $\underline{\beta}_{-1} < \bar{\beta}_{-1}$ and the second inequality in Case 9 is $\bar{\beta}_{-1} < \beta_{-1}$. Also, since $\underline{\beta}_m < \bar{\beta}_m$ for $m = \pm 1$, if for any of the cases both the first inequality and the second inequality are changed to equalities or both the third inequality and the fourth inequality are changed to equalities, then the resulting conditions are impossible.*

2. *The resulting conditions can be obtained from more than one of the cases in Table 11 by changing one or two of the strict inequalities in each case to equalities. For example, the same conditions are obtained from Case 1 by changing the first strict inequality $\underline{\beta}_{-1} < \beta_{-1}$ to the equality $\underline{\beta}_{-1} = \beta_{-1}$ and from Case 6 by changing the first strict inequality $\underline{\beta}_{-1} > \beta_{-1}$ to the equality $\underline{\beta}_{-1} = \beta_{-1}$. Similarly, the same conditions are obtained from Cases 1, 2, 3, and 6 by changing for each case both the first and the third inequalities to equalities. Under the resulting conditions, the set of equilibria is the union of the equilibria for the cases from which the conditions can be obtained. For example, if the first inequalities in Cases 1 and 6 are changed to equality, then the resulting set of equilibria is given by $\{(x^*_{-1}, x^*_1), (A^0_{-1}, A^1_1), (A^1_{-1}, A^0_1)\}$,*

*where* $x^*_{-1} = 1/(1+\Gamma_{-1})$ *and* $\Gamma_{-1}$ *is given by (3.5), and* $x^*_1 = 1$*. Similarly, if both the first and the third inequalities in Cases 1, 2, 3, and 6 are changed to equalities, then the resulting set of equilibria is given by* $\{(x^*_{-1}, x^*_1), (A^0_{-1}, A^1_1), (A^1_{-1}, A^0_1), (A^1_{-1}, A^1_1)\}$*. In this case,* $x^*_m = 1$ *for* $m = \pm 1$*, and thus the resulting set of equilibria is equal to* $\{(A^0_{-1}, A^1_1), (A^1_{-1}, A^0_1), (A^1_{-1}, A^1_1)\}$*.*

*If, for any case, three or more inequalities are changed to equalities, then the resulting conditions are impossible, because* $\underline{\beta}_m < \bar{\beta}_m$ *for* $m = \pm 1$*.*

**Economic Implications.** We say that the inter-decoy effect $\beta_m$ dominates the intra-decoy effect $\alpha_m$ if $\beta_m > \bar{\beta}_m (:= \alpha_m + [\alpha_m - 1]/v_{t_{-m}})$, that $\alpha_m$ dominates $\beta_m$ if $\beta_m < \underline{\beta}_m (:= \alpha_m + [\alpha_m - 1]/[\alpha_{-m} v_{t_{-m}}])$, and that $\alpha_m$ and $\beta_m$ are similar if $\underline{\beta}_m \le \beta_m \le \bar{\beta}_m$. If $\alpha_m$ dominates $\beta_m$ for $m = \pm 1$, then $(A^1_{-1}, A^1_1)$ is the Nash equilibrium, i.e., both sellers use decoys. If $\beta_m$ dominates $\alpha_m$ for $m = \pm 1$, then $(A^0_{-1}, A^0_1)$ is the Nash equilibrium, i.e., neither seller uses a decoy. We next consider settings in which $\alpha_m$ and $\beta_m$ are similar. If $\underline{\beta}_m \le \beta_m$ and $\beta_{-m} \le \bar{\beta}_{-m}$, then $(A^1_{-m}, A^0_m)$ is a Nash equilibrium. In this case, seller $-m$ uses a decoy but seller $m$ takes a free ride to take advantage of seller $-m$'s decoy. Hence, if $\underline{\beta}_m \le \beta_m \le \bar{\beta}_m$ for $m = \pm 1$, then both $(A^0_{-1}, A^1_1)$ and $(A^1_{-1}, A^0_1)$ are Nash equilibria. In that case, there is also a mixed-strategy Nash equilibrium $(x^*_{-1}, x^*_1)$, but it will be shown that such a mixed-strategy Nash equilibrium $(x^*_{-1}, x^*_1)$ is unstable, whereas the pure-strategy Nash equilibria are stable.

### 3.4.2 Cournot Adjustment Process

Next we study the dynamics of assortment competition when sellers learn about each other's decisions, and we infer from it the stability of the equilibria given in Table 11. We start with a very simple process called Cournot adjustment, in which, at each time $t \in \mathbb{N}$, each seller $m$ chooses a best response $A_m(t)$ to the other seller's previous action $A_{-m}(t-1)$. Let $A(0) := (A_{-1}(0), A_1(0))$ denote the initial state of the Cournot adjustment process, and let $A(t) := (A_{-1}(t), A_1(t))$ denote the state of the Cournot adjustment process at time $t$, where

$$A_m(t) \in \arg\max_{A_m \in C_m} \pi_m(A_m, A_{-m}(t-1)) . \tag{3.6}$$

#### *3.4.2.1 Case 1*

Recall from Table 11 that Case 1 holds if $\underline{\beta}_m < \beta_m < \bar{\beta}_m$ for $m = \pm 1$, and that under Case 1, one mixed-strategy Nash equilibrium $x^* = (x^*_{-1}, x^*_1)$ coexists with two pure-strategy Nash equilibria, $(A^0_{-1}, A^1_1)$ and $(A^1_{-1}, A^0_1)$. Theorem 3.1 characterizes the dynamics of the Cournot adjustment process.

**Theorem 3.1.** *(Behavior of the Cournot adjustment process under Case 1.)*

*(1) If* $A(0) \in \{(A^0_{-1}, A^1_1), (A^1_{-1}, A^0_1)\}$, *then* $A(t) = A(0)$ *for all* $t \in \mathbb{N}_0 := \mathbb{N} \cup \{0\}$, *i.e.,* $(A^0_{-1}, A^1_1)$ *and* $(A^1_{-1}, A^0_1)$ *are steady states of the Cournot adjustment process.*

*(2) If* $A(0) \in \{(A^0_{-1}, A^0_1), (A^1_{-1}, A^1_1)\}$, *then* $A(t)$ *cycles. Specifically, if* $A(0) = (A^1_{-1}, A^1_1)$, *then*

$$A(t) = \begin{cases} (A^1_{-1}, A^1_1) & \text{if } t \geq 1 \text{ and } t \text{ is even,} \\ (A^0_{-1}, A^0_1) & \text{if } t \geq 1 \text{ and } t \text{ is odd.} \end{cases}$$

*and if* $A(0) = (A^0_{-1}, A^0_1)$, *then*

$$A(t) = \begin{cases} (A^0_{-1}, A^0_1) & \text{if} \quad t \geq 1 \text{ and } t \text{ is even,} \\ (A^1_{-1}, A^1_1) & \text{if} \quad t \geq 1 \text{ and } t \text{ is odd.} \end{cases}$$

Thus, if $A(0) \in \{(A^1_{-1}, A^1_1), (A^0_{-1}, A^0_1)\}$, then $A(t)$ cycles as $(A^1_{-1}, A^1_1) \to (A^0_{-1}, A^0_1) \to (A^1_{-1}, A^1_1) \to (A^0_{-1}, A^0_1) \to \ldots$ or $(A^0_{-1}, A^0_1) \to (A^1_{-1}, A^1_1) \to (A^0_{-1}, A^0_1) \to (A^1_{-1}, A^1_1) \to \ldots$. Clearly, neither trajectory converges to a pure-strategy Nash equilibrium. The empirical frequencies of $(A^1_{-1}, A^1_1)$ and $(A^0_{-1}, A^0_1)$ converge to a (joint) probability distribution $\bar{x}^* \in \Delta(C)$ as $t \to \infty$, given by $\bar{x}^*((A^1_{-1}, A^1_1)) = 1/2$ and $\bar{x}^*((A^0_{-1}, A^0_1)) = 1/2$. One may wonder whether or not $\bar{x}^*$ is a correlated equilibrium or coarse correlated equilibrium. Proposition 3.3 gives the answer.

**Proposition 3.3.** *The limit empirical distribution* $\bar{x}^*$ *given by* $\bar{x}^*((A^1_{-1}, A^1_1)) = 1/2$ *and* $\bar{x}^*((A^0_{-1}, A^0_1)) = 1/2$ *is neither a correlated equilibrium nor a coarse correlated equilibrium.*

#### *3.4.2.2 Cases 2–9*

First consider Case 2, which holds under condition $\beta_m < \bar{\beta}_m$ and $\beta_m < \underline{\beta}_m$ for $m = \pm 1$.

**Proposition 3.4.** *(Behavior of the Cournot adjustment process under Case 2.) For any initial condition $A(0) \in C$, it holds that $A(t) = (A^1_{-1}, A^1_1)$ for all $t \in \mathbb{N}$.*

The Cournot adjustment process under Cases 3–9 behave in a similar way to the process under Case 2. Each seller's action in the unique pure-strategy Nash equilibrium under each of the Cases 3–9 dominates the other action and as a result the Cournot adjustment process stays at the equilibrium after the first step.

### 3.4.3 Fictitious Play Process

The fictitious play process works as follows: At the beginning of period $t \in \mathbb{N}$ each seller $m$ constructs an empirical distribution of the decisions of the other seller using the available data $A_{-m}(0), \ldots, A_{-m}(t-1)$. Then seller $m$ chooses $A_m(t)$ that optimizes the expected objective value of seller $m$ with respect to the empirical distribution of the decisions of the other seller. Thereafter each seller observes the decision of the other seller, the empirical distributions are updated, and the steps repeat.

Let

$$x_m(t) \quad := \quad \frac{M_m x_m(0) + \sum_{\tau=1}^{t} \mathbf{1}_{[A_m(\tau)=A^1_m]}}{M_m + t}$$

denote the empirical probability based on data $A_m(0), \ldots, A_m(t)$ that seller $m$ chooses action $A^1_m$, where $\mathbf{1}_{[\cdot]}$ denotes the indicator function. One can think of $M_m$ as the assessment of seller $-m$ of the number of observations that the initial value $x_m(0)$ is based on. We allow any initial value $x_m(0) \in [0,1]$ with weight $M_m \geq 0$. However, some notation will be simplified if $M_m > 0$, and therefore some later notation is based on the assumption that $M_m > 0$; equivalently, one may set time index $t = 0$ after an observation has been made. The vector $x(t) := (x_{-1}(t), x_1(t))$ is called the state (of fictitious play) in period $t$.

Let

$$\mathtt{PBR}_m(x) \quad := \quad \arg\max_{A_m \in C_m} x\, \pi_m(A_m, A^1_{-m}) + (1-x)\, \pi_m(A_m, A^0_{-m})$$

denote the optimal assortments for seller $m$ given that seller $m$ assesses the probability that seller $-m$ chooses action $A^1_{-m}$ to be $x$. In period $t$, each seller $m$ chooses an action $A_m(t) \in \texttt{PBR}_m(x_{-m}(t-1))$. In some cases we want to show how the long-run behavior of $x(t)$ depends on an initial condition $x(t_0)$ for some specific $t_0 \in \mathbb{N}_0$. Therefore it is convenient to use notation

$$\begin{aligned} \phi_m(t_1, t_0, x) &:= x_m(t_1) \quad \text{given that } x(t_0) = x \\ \phi(t_1, t_0, x) &:= (\phi_{-1}(t_1, t_0, x), \phi_1(t_1, t_0, x)) \end{aligned}$$

to explicitly denote the dependence of $x(t_1)$ on the initial value $x$ at time $t_0$, for $t_1 \geq t_0$.

#### *3.4.3.1 Case 1*

Recall that Case 1 holds if $\underline{\beta}_m < \beta_m < \bar{\beta}_m$ for $m = \pm 1$, in which case there are the following three equilibria: $(x^*_{-1}, x^*_1)$, $(0, 1)$ (i.e., $(A^0_{-1}, A^1_1)$) and $(1, 0)$ (i.e., $(A^1_{-1}, A^0_1)$).

It follows from $\underline{\beta}_m < \beta_m < \bar{\beta}_m$ that $A^0_m \in \texttt{PBR}_m(x_{-m}(t-1))$ iff $x_{-m}(t-1) \geq x^*_{-m}$, and $A^1_m \in \texttt{PBR}_m(x_{-m}(t-1))$ iff $x_{-m}(t-1) \leq x^*_{-m}$. Note that the best response of seller $m$ in period $t$ is not unique iff $x_{-m}(t-1) = x^*_{-m}$. Therefore, for the dynamics of $x(t)$ to be well-defined, we choose a tie-breaking rule to be used whenever $x_{-m}(t-1) = x^*_{-m}$. The choice of tie-breaking rule affects the notation, but it does not substantially affect the results. Specifically, we choose

$$A_m(t) = \begin{cases} A^0_m & \text{if} \quad x_{-m}(t-1) > x^*_{-m} \\ A^1_m & \text{if} \quad x_{-m}(t-1) \leq x^*_{-m} \end{cases} \tag{3.7}$$

Thereafter, each seller $m$ observes $A_{-m}(t)$, and updates the empirical distribution of the decisions of the other seller as follows:

$$x_{-m}(t) = \frac{(M_{-m} + t - 1)x_{-m}(t-1) + \mathbf{1}_{[A_{-m}(t) = A^1_{-m}]}}{M_{-m} + t}. \tag{3.8}$$

Note that

$$x(t) = \begin{cases} \left(\frac{(M_{-1}+t-1)x_{-1}(t-1)+1}{M_{-1}+t}, \frac{(M_1+t-1)x_1(t-1)+1}{M_1+t}\right) & \text{if} \quad x(t-1) \in P_0, \\ \left(\frac{(M_{-1}+t-1)x_{-1}(t-1)}{M_{-1}+t}, \frac{(M_1+t-1)x_1(t-1)}{M_1+t}\right) & \text{if} \quad x(t-1) \in P_1, \\ \left(\frac{(M_{-1}+t-1)x_{-1}(t-1)}{M_{-1}+t}, \frac{(M_1+t-1)x_1(t-1)+1}{M_1+t}\right) & \text{if} \quad x(t-1) \in P_2, \\ \left(\frac{(M_{-1}+t-1)x_{-1}(t-1)+1}{M_{-1}+t}, \frac{(M_1+t-1)x_1(t-1)}{M_1+t}\right) & \text{if} \quad x(t-1) \in P_3, \end{cases} \tag{3.9}$$

where $P_0 := [0, x^*_{-1}] \times [0, x^*_1]$, $P_1 := (x^*_{-1}, 1] \times (x^*_1, 1]$, $P_2 := [0, x^*_{-1}] \times (x^*_1, 1]$, $P_3 := (x^*_{-1}, 1] \times [0, x^*_1]$, and $P := [0,1]^2 = P_0 \cup P_1 \cup P_2 \cup P_3$. To simplify notation, we define $\phi(t_1, t_0, \cdot)$ on an extended domain $\hat{P} := \hat{P}_0 \cup \hat{P}_1 \cup P_2 \cup P_3$, where $\hat{P}_0 := (-\infty, x^*_{-1}] \times (-\infty, x^*_1]$, $\hat{P}_1 := (x^*_{-1}, \infty) \times (x^*_1, \infty)$, as follows. For any $t \in \mathbb{N}_0$, let

$$\phi(t+1, t, x) \quad = \quad \begin{cases} \left(\frac{(M_{-1}+t)x_{-1}+1}{M_{-1}+t+1}, \frac{(M_1+t)x_1+1}{M_1+t+1}\right) & \text{if} \quad x \in \hat{P}_0, \\ \left(\frac{(M_{-1}+t)x_{-1}}{M_{-1}+t+1}, \frac{(M_1+t)x_1}{M_1+t+1}\right) & \text{if} \quad x \in \hat{P}_1, \\ \left(\frac{(M_{-1}+t)x_{-1}}{M_{-1}+t+1}, \frac{(M_1+t)x_1+1}{M_1+t+1}\right) & \text{if} \quad x \in P_2, \\ \left(\frac{(M_{-1}+t)x_{-1}+1}{M_{-1}+t+1}, \frac{(M_1+t)x_1}{M_1+t+1}\right) & \text{if} \quad x \in P_3, \end{cases} \tag{3.10}$$

We first study the dynamics of fictitious play with initial points in the extended domain $\hat{P}$, and the results obtained for $\hat{P}$ will describe the dynamics of fictitious play on $P$. It follows from (3.10) that the directions of movement from $x(t)$ to $x(t+1)$ are as given by Lemma 3.1.

**Lemma 3.1.** *The following holds at any time $t \in \mathbb{N}_0$:*

*(1) If $x(t) \in \hat{P}_0$, then $x_m(t+1) > x_m(t)$ for $m = \pm 1$.*

*(2) If $x(t) \in \hat{P}_1$, then $x_m(t+1) < x_m(t)$ for $m = \pm 1$.*

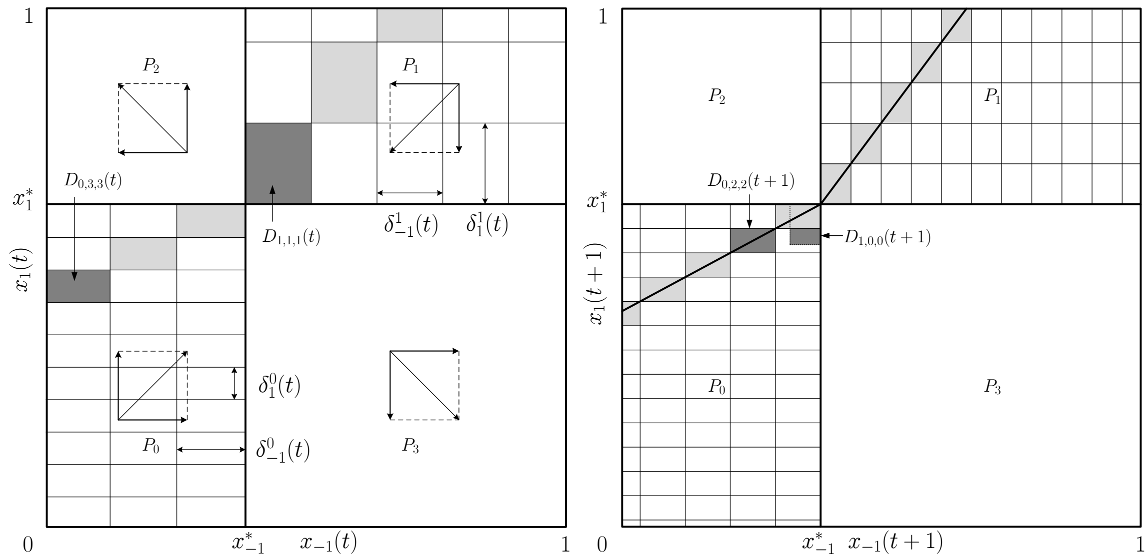
*(3) If $x(t) \in P_2$, then $x_{-1}(t+1) \leq x_{-1}(t)$ and $x_1(t+1) \geq x_1(t)$.*

*(4) If $x(t) \in P_3$, then $x_{-1}(t+1) \geq x_{-1}(t)$ and $x_1(t+1) \leq x_1(t)$.*

Figure 7 will be used to describe a generic step of the state $x(t)$ from time $t$ to time $t+1$. First, Theorem 3.2 uses (3.10) and Lemma 3.1 to establish convergence for initial points $x \in P_2 \cup P_3$. Thereafter we consider the more complicated dynamics for initial points $x \in \hat{P}_0 \cup \hat{P}_1$.

**Theorem 3.2.** *For any $x \in P_2$ and $t \in \mathbb{N}_0$, it holds that $\phi(t+\tau, t, x) \to (0,1)$ (i.e., $(A_{-1}(\tau), A_1(\tau)) \to (A^0_{-1}, A^1_1))$ as $\tau \to \infty$. For any $x \in P_3$, it holds that $\phi(t+\tau, t, x) \to (1,0)$ (i.e., $(A_{-1}(\tau), A_1(\tau)) \to (A^1_{-1}, A^0_1))$ as $\tau \to \infty$.*

**One-step Analysis**




(a) At time $t$. (b) At time $t+1$.

Figure 7: One-step evolution of the state $x(t)$ of the discrete-time fictitious play process.

To analyze the dynamics when fictitious play starts from a point in $\hat{P}_0$ or $\hat{P}_1$, we first characterize one-step changes. For any $t, \tau \in \mathbb{N}_0$ and any $D \subset \hat{P}$, let $\phi(t+\tau, t, D) := \{\phi(t+\tau, t, x) : x \in D\}$ denote the image of $D$ under $\phi(t+\tau, t, \cdot)$. Consider any $k \in \{0, 1\}$. Note from (3.10) that $\phi(t+1, t, \cdot) : \hat{P}_k \mapsto \phi(t+1, t, \hat{P}_k)$ is an increasing separable affine mapping, that is, there are $\ell_m^k(t) > 0$ and $a_m^k(t) \in \mathbb{R}$ such that $\phi_m(t+1, t, x) = \ell_m^k(t) x_m + a_m^k(t)$ for any $x \in \hat{P}_k$. For example, for $k = 0$ and $m = 1$, $\ell_1^0(t) = (M_1 + t)/(M_1 + t + 1)$ and $a_1^0(t) = 1/(M_1 + t + 1)$.

For any set $D \subset \mathbb{R}^n$, we say that $D$ walks to $\tilde{D} \subset \mathbb{R}^n$ if and only if there exists an increasing separable affine mapping $f$ such that $\tilde{D} = f(D)$. Lemma 3.2 gives a useful result that rectangles walk to rectangles under increasing separable affine mappings such as $\phi(t+1, t, \cdot) : \hat{P}_k \mapsto \phi(t+1, t, \hat{P}_k)$.

**Lemma 3.2.** *For any $t \in \mathbb{N}_0$, any rectangle $D \subset \hat{P}_k$, where $k \in \{0, 1\}$, or any rectangle $D \subset P_k$, where $k \in \{2, 3\}$, it holds that $D$ walks to $\phi(t+1, t, D)$, and $\phi(t+1, t, D)$ is a rectangle.*

We will be particularly interested in special rectangles called *cells* that are defined next. For $k \in \{0, 1\}$ and $m \in \{-1, 1\}$, let $\delta_m^k(t) := |k - 1 + x_m^*| / (M_m + t)$ denote the length parallel to the axis for $x_m$ of the cells in $\hat{P}_k$ at time step $t \in \mathbb{N}_0$; see Figure 7a. (If $M_m > 0$ then $\delta_m^k(0)$ is well defined.) The cells are indexed starting at $x^* = (x_{-1}^*, x_1^*)$, using indices $i, j$. Specifically, for $\hat{P}_0$, the index sets at time $t$ are $\hat{I}_0(t) := \{0, 1, \ldots, \lfloor x_1^*/\delta_1^0(t)\rfloor + 1\}$, and $\hat{J}_0(t) := \{0, 1, \ldots, \lfloor x_{-1}^*/\delta_{-1}^0(t)\rfloor + 1\}$. For $\hat{P}_1$, the index sets at time $t$ are $\hat{I}_1(t) := \{0, 1, \ldots, \lceil (1 - x_1^*)/\delta_1^1(t)\rceil\}$ and $\hat{J}_1(t) := \{0, 1, \ldots, \lceil (1 - x_{-1}^*)/\delta_{-1}^1(t)\rceil\}$. Then cell $D_{0,i,j}(t)$ is given by

$$D_{0,i,j}(t) \quad := \quad \big(x_{-1}^* - j\delta_{-1}^0(t),\, x_{-1}^* - (j-1)\delta_{-1}^0(t)\big] \times \big(x_1^* - i\delta_1^0(t),\, x_1^* - (i-1)\delta_1^0(t)\big]$$

for indices $(i, j) \in \hat{I}_0(t) \times \hat{J}_0(t)$, and cell $D_{1,i,j}(t)$ is given by

$$D_{1,i,j}(t) \quad := \quad \big(x_{-1}^* + (j-1)\delta_{-1}^1(t),\, x_{-1}^* + j\delta_{-1}^1(t)\big] \times \big(x_1^* + (i-1)\delta_1^1(t),\, x_1^* + i\delta_1^1(t)\big]$$

for indices $(i, j) \in \hat{I}_1(t) \times \hat{J}_1(t)$. Note that $D_{0,0,0}(t) \subset \hat{P}_1$ and $D_{0,i,j}(t) \subset \hat{P}_0$ for $i, j \geq 1$; and $D_{1,0,0}(t) \subset \hat{P}_0$ and $D_{1,i,j}(t) \subset \hat{P}_1$ for $i, j \geq 1$; see Figure 7b. Let $I_k(t) := \hat{I}_k(t) \setminus \{0\}$, $J_k(t) := \hat{J}_k(t) \setminus \{0\}$. A cell $D_{k,i,i}(t)$ for $i \in I_k(t) \cap J_k(t)$ is called a *diagonal cell.* The diagonal cells are shown in gray in Figure 7.

**Proposition 3.5.** *Consider any $t \in \mathbb{N}_0$, and any cell $D_{k,i,j}(t)$, where $k \in \{0, 1\}$, $i \in I_k(t)$ and $j \in J_k(t)$. Then $\phi(t+1, t, D_{k,i,j}(t)) = D_{k,i-1,j-1}(t+1)$, that is, $D_{k,i,j}(t)$ walks to $D_{k,i-1,j-1}(t+1)$ from time $t$ to time $t+1$.*

**Multi-step Analysis**

As an extension of single-step walking defined before, a set $D \subset \mathbb{R}^n$ is said to walk to $\tilde{D} \subset \mathbb{R}^n$ from time $t$ to time $t + \tau$, where $t, \tau \in \mathbb{N}_0$, if and only if there exists a sequence $\{f_{t+s}\}_{s=1}^{\tau}$ of increasing separable affine mappings $f_{t+s} : \mathbb{R}^n \mapsto \mathbb{R}^n$ such that $\tilde{D} = f_{t+1}^{t+\tau}(D)$, where $f_{t+1}^{t+\tau} := f_{t+\tau} \circ \cdots \circ f_{t+1}$. We will use the property that if a set walks, then all its subsets walk too, and state it as Lemma 3.3.

**Lemma 3.3.** *Suppose that a set $D$ walks to a set $\tilde{D}$ from time $t$ to time $t + \tau$ under $f_{t+1}^{t+\tau}$. Then any $E \subset D$ walks to $f_{t+1}^{t+\tau}(E) \subset \tilde{D}$ from time $t$ to time $t + \tau$.*

Note that the one-step results of Lemma 3.2 and Proposition 3.5 can be applied repeatedly to obtain a multi-step characterization of the evolution of fictitious play. For example, if rectangle $D \subset \hat{P}_k$, where $k \in \{0,1\}$, and $\phi(t+1,t,D) \subset \hat{P}_k$, then it follows from a repeated application of Lemma 3.2 that $D$ walks to $\phi(t+2,t,D)$ from time $t$ to time $t+2$, and $\phi(t+2,t,D)$ is a rectangle. Similarly, if $i, i-1 \in I_k(t)$ and $j, j-1 \in J_k(t)$, then it follows from a repeated application of Proposition 3.5 that $\phi(t+2,t,D_{k,i,j}(t)) = D_{k,i-2,j-2}(t+2)$, that is, $D_{k,i,j}(t)$ walks to $D_{k,i-2,j-2}(t+2)$ from time $t$ to time $t+2$.

Next we discuss what happens once $\phi(t+\tau,t,D) \not\subset \hat{P}_k$, or $i-\tau \notin I_k(t)$, or $j-\tau \notin J_k(t)$. Theorem 3.3 gives a complete characterization of the evolution of the off-diagonal cells. Let $\mathcal{D}_2(t) := \{D_{0,i,j}(t) : i \in I_0(t), j \in J_0(t), i < j\} \cup \{D_{1,i,j}(t) : i \in I_1(t), j \in J_1(t), i > j\}$ denote the cells in $\hat{P}_0 \cup \hat{P}_1$ above the diagonal, and let $\mathcal{D}_3(t) := \{D_{0,i,j}(t) : i \in I_0(t), j \in J_0(t), i > j\} \cup \{D_{1,i,j}(t) : i \in I_1(t), j \in J_1(t), i < j\}$ denote the cells in $\hat{P}_0 \cup \hat{P}_1$ below the diagonal.

**Theorem 3.3.** *Consider any $t \in \mathbb{N}_0$. Then, the following holds:*

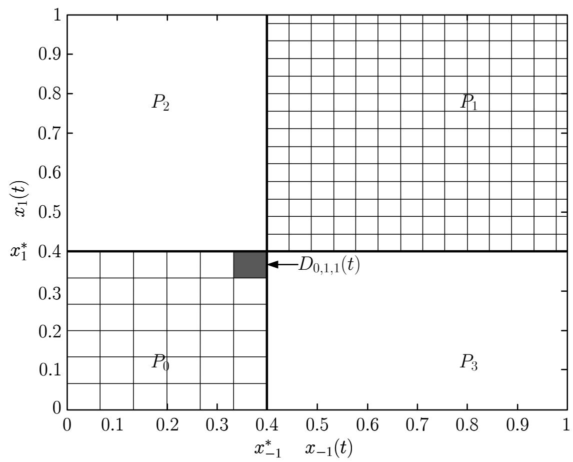
*(1) If $x \in D_{k,i,j}(t) \in \mathcal{D}_2(t)$, where $k \in \{0,1\}$, $i \in I_k(t)$, $j \in J_k(t)$, then $\phi(t+\tau,t,x) \in \hat{P}_k \cap \cup_{D \in \mathcal{D}_2(t+\tau)} D$ for $\tau \in \{0,1,\ldots,\min\{i,j\}-1\}$, $\phi(t+\tau,t,x)) \in P_2$ for $\tau \geq \min\{i,j\}$, and $\phi(t+\tau,t,x) \to (0,1)$ (i.e., $(A_{-1}(\tau), A_1(\tau)) \to (A^0_{-1}, A^1_1)$) as $\tau \to \infty$.*

*(2) If $x \in D_{k,i,j}(t) \in \mathcal{D}_3(t)$, where $k \in \{0,1\}$, $i \in I_k(t)$, $j \in J_k(t)$, then $\phi(t+\tau,t,x) \in \hat{P}_k \cap \cup_{D \in \mathcal{D}_3(t+\tau)} D$ for $\tau \in \{0,1,\ldots,\min\{i,j\}-1\}$, $\phi(t+\tau,t,x)) \in P_3$ for $\tau \geq \min\{i,j\}$, and $\phi(t+\tau,t,x)) \to (1,0)$ (i.e., $(A_{-1}(\tau), A_1(\tau)) \to (A^1_{-1}, A^0_1)$) as $\tau \to \infty$.*

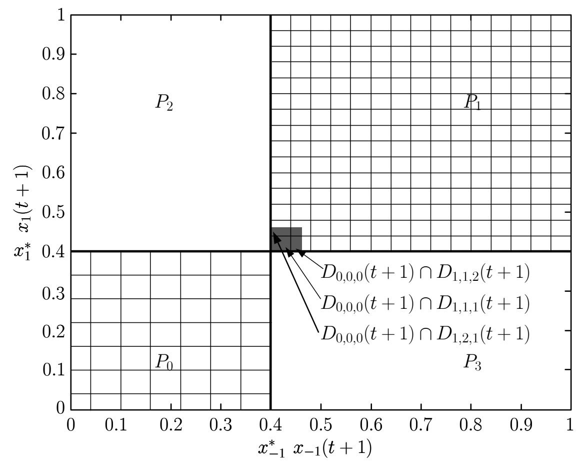
It remains to describe the evolution of the diagonal cells. If a rectangle $D \subset \hat{P}_k$ walks to rectangle $D' \subset \hat{P}_{k'}$ from time $t$ to time $t+1$, where $k, k' \in \{0,1\}, k \neq k'$, then we say that $D$ *jumps* to $D'$.

**Proposition 3.6.** *Consider any diagonal cell $D_{k,i,i}(t)$, where $t \in \mathbb{N}_0$, $k \in \{0,1\}$, and $i \in I_k(t) \cap J_k(t)$. Then $D_{k,i,i}(t)$ walks to $D_{k,1,1}(t+i-1)$ from time $t$ to time $t+i-1$, and then jumps to $D_{k,0,0}(t+i) \subset \hat{P}_{k'}$, where $k' \in \{0,1\}, k' \neq k$.*

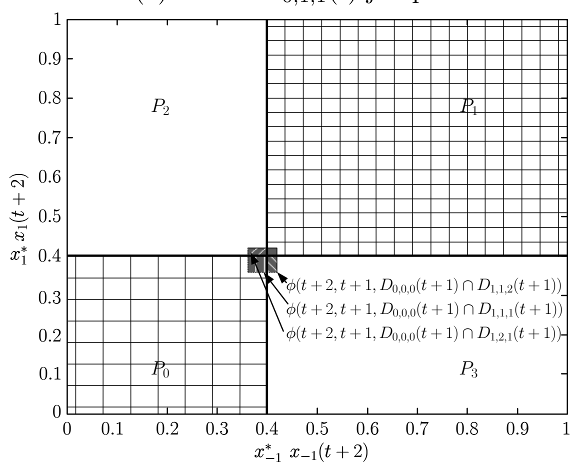
Figure 8a shows a diagonal cell $D_{0,1,1}(t) \subset \hat{P}_0$ right before a jump, and Figure 8b shows the corresponding diagonal cell $D_{0,0,0}(t+1) \subset \hat{P}_1$ right after the jump. The evolution of a diagonal cell $D_{0,0,0}(t+1)$ right after a jump depends on how $D_{0,0,0}(t+1)$ intersects with $D_{1,i,j}(t+1)$, $i \in I_1(t+1)$, $j \in J_1(t+1)$. Specifically, it follows from Lemma 3.3,



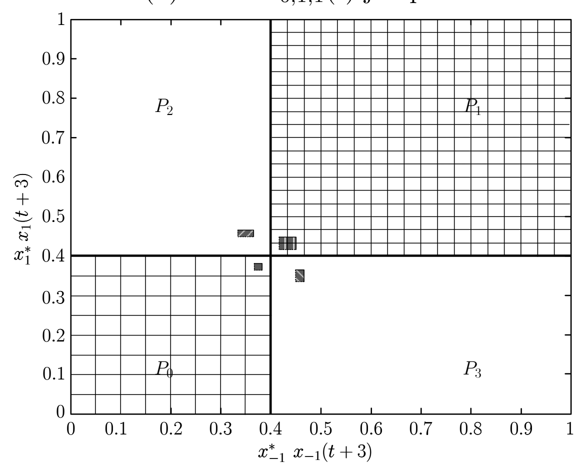

(a) Before $D_{0,1,1}(t)$ jumps.




(b) After $D_{0,1,1}(t)$ jumps.




(c) One step after the jump.




(d) Two steps after the jump of $D_{0,1,1}(t)$.

Figure 8: Images of diagonal cell $D_{0,1,1}(t)$ before and after a jump.

Theorem 3.3, and Proposition 3.6 that $D_{0,0,0}(t+1) \cap D_{1,i,j}(t+1)$ for $i > j$ walks to a rectangle $\tilde{D} \subset D_{1,i-j,0}(t+1+j) \subset P_2$ and then $\phi(t+1+j+\tau, t+1+j, \tilde{D})$ converges in $P_2$ to $(0,1)$ as $\tau \to \infty$, $D_{0,0,0}(t+1) \cap D_{1,i,j}(t+1)$ for $i < j$ walks to a rectangle $\tilde{D} \subset D_{1,0,j-i}(t+1+i) \subset P_3$ and then $\phi(t+1+i+\tau, t+1+i, \tilde{D})$ converges in $P_3$ to $(1,0)$ as $\tau \to \infty$, and $D_{0,0,0}(t+1) \cap D_{1,i,i}(t+1)$ walks to a rectangle $\tilde{D} \subset D_{1,0,0}(t+1+i) \subset \hat{P}_0$. Figure 8c shows the image of $D_{0,0,0}(t+1) \cap D_{1,i,j}(t+1)$ after one step for $(i,j) = (1,1)$, $(i,j) = (1,2)$, and $(i,j) = (2,1)$. The question remains whether every point of a diagonal cell eventually either converges to $(0,1)$ or $(1,0)$, or whether some points of a diagonal cell remain in diagonal cells forever. The answer to this question depends on whether $x_{-1}^* = x_1^*$ or $x_{-1}^* \neq x_1^*$. We will consider these two cases separately. First, the next result points out that for every initial point $x \in P$, $\phi(t,0,x)$ converges as $t \to \infty$, to $(0,1)$, or to $(1,0)$, or to $x^*$. Thus, if the sequence $\{\phi(t,0,x)\}_{t=0}^{\infty}$ remain in diagonal cells, then $\phi(t,0,x) \to x^*$ as $t \to \infty$.

Let $\mathcal{D}_=(t) := \{D_{k,i,i}(t) : k \in \{0,1\},\ i \in I_k(t) \cap J_k(t)\}$ denote the collection of diagonal cells at time $t$. Let $P_k(t) := \cup_{i \in I_k(t)} \cup_{j \in J_k(t)} D_{k,i,j}(t)$ for $k \in \{0,1\}$. Note that $P_k \subset P_k(t) \subset \hat{P}_k$ for all $t \in \mathbb{N}_0$ and $k \in \{0,1\}$, $P \subset P(t) := P_0(t) \cup P_1(t) \cup P_2 \cup P_3 \subset \hat{P}$, and that $P_0(t) \cup P_1(t) = [\cup_{D \in \mathcal{D}_2(t)} D] \cup [\cup_{D \in \mathcal{D}_3(t)} D] \cup [\cup_{D \in \mathcal{D}_=(t)} D]$.

**Proposition 3.7.** *Consider any $x \in P(0)$. One of the following three cases holds:*

*(1) $\phi(t,0,x) \to (0,1)$ as $t \to \infty$,*

*(2) $\phi(t,0,x) \to (1,0)$ as $t \to \infty$,*

*(3) $\phi(t,0,x) \to x^*$ as $t \to \infty$.*

Lemma 3.4 provides a necessary condition for $\phi(t,0,x)$ to converge to $x^*$.

**Lemma 3.4.** *Consider any $x \in P$. If $\phi(t,0,x) \to x^*$ as $t \to \infty$, then $x^*_{-1} = x^*_1$.*

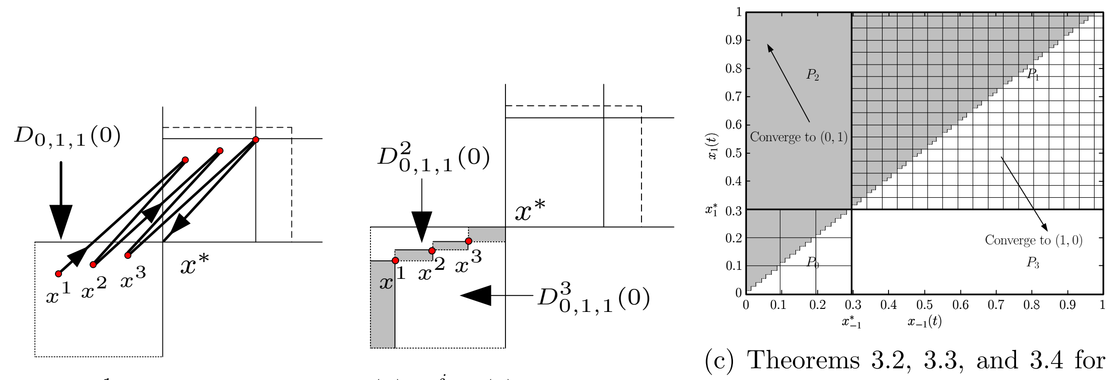
**Characterization of Convergence for $x^*_{-1} \neq x^*_1$.**

For any $x \in \hat{P}$, let $\Omega^2(x) := \{y \in \hat{P} : y_{-1} \leq x_{-1}, y_1 > x_1\}$ and $\Omega^3(x) := \{y \in \hat{P} : y_{-1} > x_{-1}, y_1 \leq x_1\}$. For any $t \in \mathbb{N}_0$ and $D \subset P(t)$, let

$$\phi^{-1}(t,D) \quad := \quad \{x \in D : \text{there exists } \tau \in \mathbb{N}_0 \text{ such that } \phi(t+\tau,t,x) = x^*\}$$

denote the set of pre-images of $x^*$ in $D$ at time $t$. For any $k \in \{0,1\}$, $i \in I_k(t) \cap J_k(t)$, and $j \in \{2,3\}$, let $D^j_{k,i,i}(t) := D_{k,i,i}(t) \cap \left(\cup_{x \in \phi^{-1}(t,D_{k,i,i}(t))} \Omega^j(x)\right)$ denote the set of points in $D_{k,i,i}(t)$ that will be cut to $P_j$ by a pre-image of $x^*$ in $D_{k,i,i}(t)$. Figures 9a and 9b show the set of pre-images $\phi^{-1}(0, D_{0,1,1}(0)) = \{x^1, x^2, x^3\}$, as well as the sets $D^2_{0,1,1}(0)$ and $D^3_{0,1,1}(0)$ that will be cut to $P_2$ and $P_3$ respectively by these pre-images.

Theorem 3.4 provides a complete geometric characterization of the convergence of fictitious play for the case with $x^*_{-1} \neq x^*_1$. Figure 9c illustrates the results of Theorems 3.2, 3.3, and 3.4 for the case with $x^*_{-1} \neq x^*_1$.




(a) $\phi^{-1}(0, D_{0,1,1}(0))$.

(b) $D^j_{0,1,1}(0)$ for $j = 2, 3$.

(c) Theorems 3.2, 3.3, and 3.4 for $x^*_{-1} \neq x^*_1$.

Figure 9: Illustrations of the set of preimages in $D_{0,1,1}(0)$, $D^j_{0,1,1}(0)$, $j = 2, 3$, and Theorems 3.2, 3.3, and 3.4 for the case with $x^*_{-1} \neq x^*_1$.

**Theorem 3.4.** *Consider any $k \in \{0, 1\}$ and $i \in I_k(0) \cap J_k(0)$. Then, the following hold:*

*(1) If $x \in D^2_{k,i,i}(0)$, then $\phi(t, 0, x) \to (0, 1)$ as $t \to \infty$.*

*(2) If $x \in D^3_{k,i,i}(0)$, then $\phi(t, 0, x) \to (1, 0)$ as $t \to \infty$.*

*(3) If $x \in D_{k,i,i}(0) \setminus (D^2_{k,i,i}(0) \cup D^3_{k,i,i}(0))$ and $x^*_{-1} < x^*_1$, then $\phi(t, 0, x) \to (1, 0)$ as $t \to \infty$.*

*(4) If $x \in D_{k,i,i}(0) \setminus (D^2_{k,i,i}(0) \cup D^3_{k,i,i}(0))$ and $x^*_{-1} > x^*_1$, then $\phi(t, 0, x) \to (0, 1)$ as $t \to \infty$.*

**Characterization of Convergence for $x^*_{-1} = x^*_1$.**

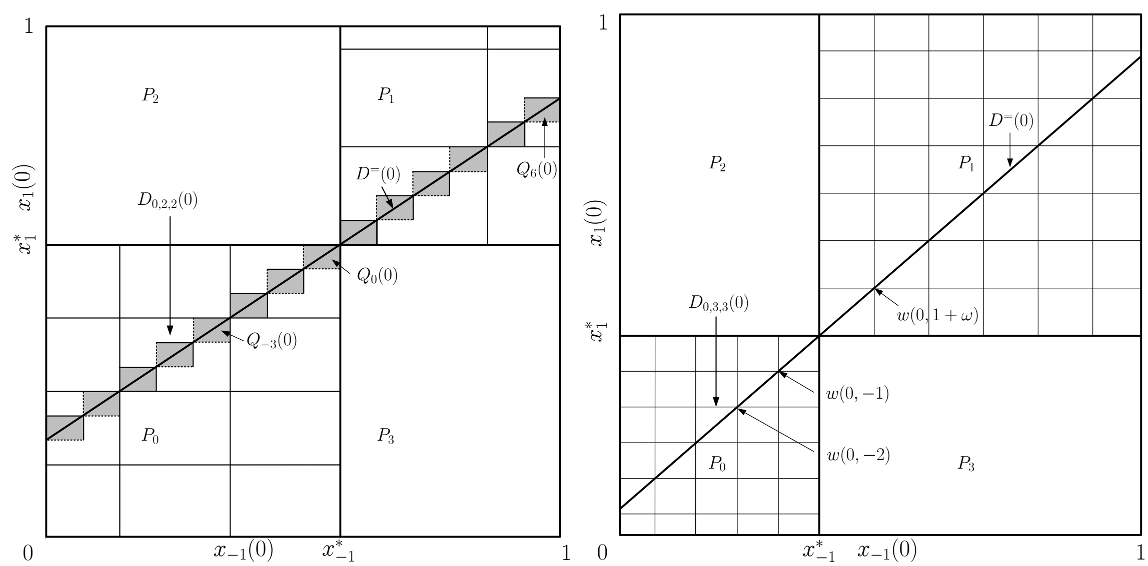
Theorems 3.5 and 3.6 completely characterize how the convergence of $\phi(t, 0, x)$ depends on the initial value $x \in P(0)$ for the case with $x^*_{-1} = x^*_1$. First, consider the rational case in which $x^*_{-1} = x^*_1 \in (0, 1)$ is a rational number. Note that $x^*_{-1} = x^*_1 \in (0, 1)$ is a rational number if and only if $x^*_1/(1 - x^*_1) = K + p/q$ for some $K \in \mathbb{N}_0$, $p, q \in \mathbb{N}$, $p/q < 1$, and $\gcd(p, q) = 1$, or $K \in \mathbb{N}$, $p = 0$, and $q = 1$. It will be shown that the images of initial points in certain small rectangles inside the diagonal cells converge to $x^*$, the images of initial points above the small rectangles converge to $(0, 1)$, and the images of initial points below the small rectangles converge to $(1, 0)$. See Figure 10a for an example of this case. For any $j \in \mathbb{Z}$, let

$$Q_j(t) \;\; := \;\; \left(x^*_{-1} + (j-1)\frac{\delta^0_{-1}(t)}{q},\; x^*_{-1} + j\frac{\delta^0_{-1}(t)}{q}\right] \times \left(x^*_1 + (j-1)\frac{\delta^0_1(t)}{q}, x^*_1 + j\frac{\delta^0_1(t)}{q}\right]$$

denote a small rectangle indexed by $j$. Note that $Q_j(t) \cap Q_{j'}(t) = \varnothing$ if $j \neq j'$. Also note that $Q_j(t) \subset \hat{P}_0$ iff $j \leq 0$, and $Q_j(t) \subset \hat{P}_1$ iff $j > 0$. Each diagonal cell $D_{0,i,i}(t)$ contains exactly $q$ small rectangles, and each diagonal cell $D_{1,i,i}(t)$ contains exactly $qx_1^*/(1-x_1^*) = Kq+p$ small rectangles. Let $\mathcal{I}_{k,i}$ denote the set of indices of small rectangles that are contained in diagonal cell $D_{k,i,i}(t)$, for $k \in \{0,1\}$ and $i \in I_k(t) \cap J_k(t)$, that is,

$$\mathcal{I}_{k,i} \quad := \quad \begin{cases} \{-iq+1, -iq+2, \ldots, -(i-1)q\} & \text{for} \quad k=0, \\ \{(i-1)(Kq+p)+1, (i-1)(Kq+p)+2, \ldots, i(Kq+p)\} & \text{for} \quad k=1. \end{cases}$$

Let $\mathcal{I}(t) := \cup_{k\in\{0,1\}} \cup_{i \in I_k(t)\cap J_k(t)} \mathcal{I}_{k,i}$, and let $Q(t) := \cup_{j\in\mathcal{I}(t)} Q_j(t)$.




(a) $K=1$, $q=3$, $p=1$ (b) $K=1$, $\omega \in (0,1)\setminus\mathbb{Q}$

Figure 10: Examples to illustrate Theorems 3.5 and 3.6.

**Theorem 3.5.** *Suppose that $x_{-1}^* = x_1^*$ and that $x_1^*/(1-x_1^*) = K + p/q$, where $K \in \mathbb{N}_0$, $p, q \in \mathbb{N}$, $p/q < 1$, and $\gcd(p,q) = 1$, or $K \in \mathbb{N}$, $p = 0$ and $q = 1$. Then, the following hold:*

*(1) If $x \in D^2_{k,i,i}(0)$ for some $k \in \{0,1\}$ and $i \in I_k(0) \cap J_k(0)$, then $\phi(t,0,x) \to (0,1)$ as $t \to \infty$.*

*(2) If $x \in D^3_{k,i,i}(0)$ for some $k \in \{0,1\}$ and $i \in I_k(0) \cap J_k(0)$, then $\phi(t,0,x) \to (1,0)$ as $t \to \infty$.*

*(3) If $x \in D_{k,i,i}(0) \setminus (D^2_{k,i,i}(0) \cup D^3_{k,i,i}(0))$ for some $k \in \{0, 1\}$ and $i \in I_k(0) \cap J_k(0)$, that is, if $x \in Q(0)$, then $\phi(t, 0, x) \to x^*$ as $t \to \infty$.*

Next, consider the irrational case in which $x^*_{-1} = x^*_1 \in (0, 1)$ is an irrational number, which is the case iff $x^*_1/(1 - x^*_1)$ is an irrational number. If $x^*_{-1} = x^*_1$, the diagonal line at time $t$ is given by $D^{=}(t) := \{x \in P(t) : (x_1 - x^*_1)(M_1 + t) = (x_{-1} - x^*_{-1})(M_{-1} + t)\}$. Also, let $D^{>}(t) := \{x \in P(t) : (x_1 - x^*_1)(M_1 + t) > (x_{-1} - x^*_{-1})(M_{-1} + t)\}$ denote points above $D^{=}(t)$, and let $D^{<}(t) := \{x \in P(t) : (x_1 - x^*_1)(M_1 + t) < (x_{-1} - x^*_{-1})(M_{-1} + t)\}$ denote points below $D^{=}(t)$.

**Theorem 3.6.** *Suppose that $x^*_{-1} = x^*_1$ and $x^*_1/(1 - x^*_1) = K + \omega$, where $K \in \mathbb{N}_0$ and $\omega \in (0, 1) \setminus \mathbb{Q}$. Then, the following hold:*

*(1) If $x \in D^{>}(0)$, then $\phi(t, 0, x) \to (0, 1)$ as $t \to \infty$.*

*(2) If $x \in D^{<}(0)$, then $\phi(t, 0, x) \to (1, 0)$ as $t \to \infty$.*

*(3) If $x \in D^{=}(0)$, then $\phi(t, 0, x) \to x^*$ as $t \to \infty$.*

Proposition 3.8 summarizes the implication of the previous results for the stabilities of the Nash equilibria for Case 1.

**Proposition 3.8.** *(Stability of the Nash equilibria under Case 1) Under fictitious play dynamics, $(0, 1)$ and $(1, 0)$ are stable equilibria, but $x^*$ is an unstable equilibrium.*

#### *3.4.3.2 Cases 2–9*

First consider Case 2, that holds under condition $\beta_m < \bar{\beta}_m$ and $\beta_m < \underline{\beta}_m$ for $m = \pm 1$.

**Proposition 3.9.** *Under Case 2 in Table 11, it holds for all $x \in P$ that $\phi(t, 0, x) \to (1, 1)$, i.e., $(A_{-1}(t), A_1(t)) \to (A^1_{-1}, A^1_1)$, as $t \to \infty$.*

For each of Cases 3–9, a similar result holds, i.e., for any initial value $x \in P$, fictitious play converges to the unique pure-strategy Nash equilibrium under that case.

## 3.5 *Assortment Competition with General Product Sets and Simple Actions*

In this section we consider assortment competition with general product sets and simple actions. Product set $S_m$ may contain other products in addition to the target product $t_m$ and the decoy product $d_m$. Each seller $m$ makes a choice between including or excluding the decoy $d_m$ from the seller's assortment. Thus this case holds when sellers consider whether or not to add the decoy products into their current assortments without redesigning the remainder of their assortments. Let $C_m = \{A_m^0, A_m^1\}$ denote the action set of seller $m$, where $A_m^1 = S_m$ and $A_m^0 = S_m \setminus \{d_m\}$.

### 3.5.1 Characterization of Equilibria

For assortment competition with simple actions, a mixed strategy $\bar{x}_m = (\bar{x}_m(A_m^0), \bar{x}_m(A_m^1))$ is specified by $x_m := \bar{x}_m(A_m^1)$. Let

$$
\begin{aligned}
b_m &:= \sum_{i \in S_m \setminus \{t_m, d_m\}} v_i p_i \\
c &:= \sum_{i \in (S_{-1} \cup S_1) \setminus \{t_{-1}, d_{-1}, t_1, d_1\}} v_i \\
\underline{\beta}_m &:= \frac{\beta_{-m} v_{t_m} p_{t_m}}{b_m + \beta_{-m} v_{t_m} p_{t_m}} \left( \alpha_m + \frac{\alpha_m - 1}{\alpha_{-m} v_{t_{-m}}} + \frac{(\alpha_m - 1)c}{\alpha_{-m} v_{t_{-m}}} \right) + \frac{b_m \left[ \alpha_{-m} v_{t_{-m}} - (\alpha_m - 1)\beta_{-m} v_{t_m} \right]}{(b_m + \beta_{-m} v_{t_m} p_{t_m})\, \alpha_{-m} v_{t_{-m}}} \\
\bar{\beta}_m &:= \frac{v_{t_m} p_{t_m}}{b_m + v_{t_m} p_{t_m}} \left( \alpha_m + \frac{\alpha_m - 1}{v_{t_{-m}}} + \frac{(\alpha_m - 1)c}{v_{t_{-m}}} \right) + \frac{b_m \left[ v_{t_{-m}} - (\alpha_m - 1) v_{t_m} \right]}{(b_m + v_{t_m} p_{t_m})\, v_{t_{-m}}} \\
\underline{\lambda}_m &:= \frac{\alpha_{-m} v_{t_{-m}} \left( \beta_{-m} v_{t_m} p_{t_m} + b_m \right)}{\left( 1 + \beta_{-m} v_{t_m} + \alpha_{-m} v_{t_{-m}} + c \right) \left( 1 + \alpha_m \beta_{-m} v_{t_m} + \alpha_{-m} \beta_m v_{t_{-m}} + c \right)} \\
\bar{\lambda}_m &:= \frac{v_{t_{-m}} \left( v_{t_m} p_{t_m} + b_m \right)}{\left( 1 + v_{t_m} + v_{t_{-m}} + c \right) \left( 1 + \alpha_m v_{t_m} + \beta_m v_{t_{-m}} + c \right)} \\
\Gamma_{-m} &:= \frac{\underline{\lambda}_m \left( \underline{\beta}_m - \beta_m \right)}{\bar{\lambda}_m \left( \beta_m - \bar{\beta}_m \right)}
\end{aligned}
$$

Note that, unlike the case with simple product sets, it does not necessarily hold that $\underline{\beta}_m < \bar{\beta}_m$. Proposition 3.10 characterizes the equilibria for the case with general product sets and simple actions.

**Proposition 3.10.** *For assortment competition with general product sets and simple actions, the set of equilibria is completely characterized by the 16 cases in Table 12, with* $x_m^* = 1/(1 + \Gamma_m)$.

| # | Sufficient condition | | | | Nash equilibrium |
|---|---|---|---|---|---|
| | $\underline{\beta}_{-1}$ v.s. $\beta_{-1}$ | $\bar{\beta}_{-1}$ v.s. $\beta_{-1}$ | $\underline{\beta}_1$ v.s. $\beta_1$ | $\bar{\beta}_1$ v.s. $\beta_1$ | |
| 1 | $>$ | $<$ | $>$ | $<$ | $(x^*_{-1}, x^*_1)$,$(A^1_{-1}, A^1_1)$, $(A^0_{-1}, A^0_1)$ |
| 2 | $<$ | $>$ | $>$ | $<$ | $(x^*_{-1}, x^*_1)$ |
| 3 | $>$ | $<$ | $<$ | $>$ | $(x^*_{-1}, x^*_1)$ |
| 4 | $<$ | $>$ | $<$ | $>$ | $(x^*_{-1}, x^*_1)$, $(A^0_{-1}, A^1_1)$, $(A^1_{-1}, A^0_1)$ |
| 5 | $>$ | $>$ | $>$ | $>$ | $(A^1_{-1}, A^1_1)$ |
| 6 | $>$ | $>$ | $>$ | $<$ | |
| 7 | $>$ | $<$ | $>$ | $>$ | |
| 8 | $>$ | $>$ | $<$ | $>$ | $(A^0_{-1}, A^1_1)$ |
| 9 | $>$ | $>$ | $<$ | $<$ | |
| 10 | $<$ | $>$ | $<$ | $<$ | |
| 11 | $<$ | $>$ | $>$ | $>$ | $(A^1_{-1}, A^0_1)$ |
| 12 | $<$ | $<$ | $>$ | $>$ | |
| 13 | $<$ | $<$ | $<$ | $>$ | |
| 14 | $>$ | $<$ | $<$ | $<$ | $(A^0_{-1}, A^0_1)$ |
| 15 | $<$ | $<$ | $>$ | $<$ | |
| 16 | $<$ | $<$ | $<$ | $<$ | |

Table 12: Characterization of the Nash equilibria for general product sets and simple actions.

Note that not all 16 cases can occur for all values of $\underline{\beta}_m$ and $\bar{\beta}_m$. For example, Case 1 can occur only if $\underline{\beta}_m > \bar{\beta}_m$ for $m = \pm 1$, Case 5 can occur for all values of $\underline{\beta}_m$ and $\bar{\beta}_m$, and Case 6 can occur only if $\underline{\beta}_{-1} > \bar{\beta}_{-1}$. For each of the four settings (1) $\underline{\beta}_m < \bar{\beta}_m$ for $m = \pm 1$, (2) $\underline{\beta}_m > \bar{\beta}_m$ for $m = \pm 1$, (3) $\underline{\beta}_1 < \bar{\beta}_1$ and $\underline{\beta}_{-1} > \bar{\beta}_{-1}$, and (4) $\underline{\beta}_1 > \bar{\beta}_1$ and $\underline{\beta}_{-1} < \bar{\beta}_{-1}$, exactly nine cases in Table 12 can occur.

Cases in which some of the inequalities in Table 12 are replaced with equalities can be resolved as explained in Remark 3.1.

### 3.5.2 Cournot Adjustment Process

#### *3.5.2.1 Cases 1 and 4*

The behavior of the Cournot adjustment process under Case 4 in Table 12 is the same as the behavior described in Theorem 3.1. For Case 1 in Table 12, the result of Theorem 3.1 holds after interchanging $(A^1_{-1}, A^0_1)$ and $(A^1_{-1}, A^1_1)$ and interchanging $(A^0_{-1}, A^1_1)$ and $(A^0_{-1}, A^0_1)$.

#### *3.5.2.2 Cases 2 and 3*

We first study Case 3 in Table 12, that holds if $\bar{\beta}_{-1} < \beta_{-1} < \underline{\beta}_{-1}$ and $\underline{\beta}_1 < \beta_1 < \bar{\beta}_1$. Proposition 3.11 asserts that the Cournot adjustment process cycles.

**Proposition 3.11.** *Under Case 3 in Table 12, the Cournot adjustment process $A(t)$ cycles as follows:* $\cdots \to (A^1_{-1}, A^1_1) \to (A^1_{-1}, A^0_1) \to (A^0_{-1}, A^0_1) \to (A^0_{-1}, A^1_1) \to (A^1_{-1}, A^1_1) \to \cdots$.

By a similar argument, the Cournot adjustment process under Case 2 in Table 12 cycles as follows: $\cdots \to (A^1_{-1}, A^1_1) \to (A^0_{-1}, A^1_1) \to (A^0_{-1}, A^0_1) \to (A^1_{-1}, A^0_1) \to (A^1_{-1}, A^1_1) \to \cdots$.

#### *3.5.2.3 Cases 5–16*

Under Cases 5–16 in Table 12, each seller's action in the unique pure-strategy Nash equilibrium for each case dominates the other action, and as a result the Cournot adjustment process stays at the equilibrium after the first step.

### 3.5.3 Fictitious Play Process

#### *3.5.3.1 Cases 1 and 4*

The dynamics of the fictitious play process under Case 4 in Table 12 is the same as described in Section 3.4.3.1. By changing variables $y_{-1} = 1 - x_1$ and $y_1 = x_{-1}$, the analysis of the dynamics of the fictitious play process under Case 4 applies to Case 1 on the $(y_{-1}, y_1)$-plane with $y^* = (y^*_{-1}, y^*_1)$ as the mixed-strategy equilibrium, where $y^*_{-1} = 1 - x^*_1$ and $y^*_1 = x^*_{-1}$.

#### *3.5.3.2 Cases 2 and 3*

Since the mixed-strategy Nash equilibrium $x^*$ is the unique Nash equilibrium for Case 2 or Case 3, it follows from [44] that, for any initial $x \in P$, the fictitious play process converges to $x^*$ as $t \to \infty$.

#### *3.5.3.3 Cases 5–16*

Under Cases 5–16 in Table 12, each seller's action in the unique pure-strategy Nash equilibrium for each case dominates the other action, and as a result the fictitious play process converges to the pure-strategy Nash equilibrium for that case.

## 3.6 *Assortment Competition with General Product Sets and General Actions*

In this section we consider the setting in which each seller has a general product set ($S_m$ may contain other products in addition to the target and decoy), and each seller chooses which subset of the product set to offer.

### 3.6.1 Characterization of Equilibria

Let $C_m^0 := \{A_m \in C_m : d_m \notin A_m\}$ denote the set of assortments of seller $m$ excluding the decoy, and let $C_m^1 := \{A_m \in C_m : d_m \in A_m\}$ denote the set of assortments of seller $m$ including the decoy. For any mixed strategy $\bar{x}_m \in \Delta(C_m)$, let $C_m^+(\bar{x}_m) := \{A_m \in C_m : \bar{x}_m(A_m) > 0\}$ denote the assortments chosen with positive probability by $\bar{x}_m$. For $i \in \{0, 1\}$, define the restricted pure best response correspondences $\mathtt{PBR}_m^i : \Delta(C_{-m}) \mapsto 2^{C_m^i}$ as

$$\mathtt{PBR}_m^i(\bar{x}_{-m}) \quad := \quad \operatorname*{argmax}_{A_m^i \in C_m^i} \sum_{A_{-m} \in C_{-m}} \pi_m(A_m^i, A_{-m})\bar{x}_{-m}(A_{-m}),$$

that is, $\mathtt{PBR}_m^0(\bar{x}_{-m})$ ($\mathtt{PBR}_m^1(\bar{x}_{-m})$) is the set of best responses of seller $m$ to assortment distribution $\bar{x}_{-m}$ among the assortments that exclude (include) the decoy. If $\bar{x}_{-m}(A_{-m}) = 1$ for some $A_{-m} \in C_{-m}$, then we also write $\mathtt{PBR}_m^i(A_{-m})$ for $\mathtt{PBR}_m^i(\bar{x}_{-m})$.

For any $A, A' \subset S$, let $b_m(A) := \sum_{j \in A \setminus \{t_m, d_m\}} v_j p_j$ and $c(A, A') := \sum_{j \in A \cup A' \setminus \{t_1, d_1, t_{-1}, d_{-1}\}} v_j$. For any $A_{-m} \in C_{-m}$, and for any (arbitrarily) chosen $A_m^i \in \mathtt{PBR}_m^i(A_{-m})$, let

$$\begin{aligned}
\bar{\beta}_m(A_{-m}) \quad &:= \quad \frac{v_{t_m} p_{t_m}}{v_{t_m} p_{t_m} + b_m(A_m^0)} \left( \alpha_m + \frac{\alpha_m - 1}{v_{t_{-m}}} + \frac{\alpha_m c(A_{-m}, A_m^0) - c(A_{-m}, A_m^1)}{v_{t_{-m}}} \right) \\
&+ \quad \frac{b_m(A_m^1)\left[1 + v_{t_m} + v_{t_{-m}} + c(A_{-m}, A_m^0)\right] - b_m(A_m^0)\left[1 + \alpha_m v_{t_m} + c(A_{-m}, A_m^1)\right]}{v_{t_{-m}}\left(v_{t_m} p_{t_m} + b_m(A_m^0)\right)} \\
\underline{\beta}_m(A_{-m}) \quad &:= \quad \frac{\beta_{-m} v_{t_m} p_{t_m}}{\beta_{-m} v_{t_m} p_{t_m} + b_m(A_m^0)} \left( \alpha_m + \frac{\alpha_m - 1}{\alpha_{-m} v_{t_{-m}}} + \frac{\alpha_m c(A_{-m}, A_m^0) - c(A_{-m}, A_m^1)}{\alpha_{-m} v_{t_{-m}}} \right) \\
&+ \quad \frac{b_m(A_m^1)\left[1 + \beta_{-m} v_{t_m} + \alpha_{-m} v_{t_{-m}} + c(A_{-m}, A_m^0)\right]}{\alpha_{-m} v_{t_{-m}}\left(\beta_{-m} v_{t_m} p_{t_m} + b_m(A_m^0)\right)} \\
&- \quad \frac{b_m(A_m^0)\left[1 + \alpha_m \beta_{-m} v_{t_m} + c(A_{-m}, A_m^1)\right]}{\alpha_{-m} v_{t_{-m}}\left(\beta_{-m} v_{t_m} p_{t_m} + b_m(A_m^0)\right)}.
\end{aligned}$$

Proposition 3.12 shows that for each seller $m$ a best response to the competitor's assortment contains a decoy iff seller $m$'s inter-decoy factor $\beta_m$ is small relative to a threshold.

**Proposition 3.12.** *The following holds:*

*(1) For any $A^0_{-m} \in C^0_{-m}$ and any $A^0_m \in \texttt{PBR}^0_m(A^0_{-m})$, it holds that $A^0_m \in \texttt{PBR}_m(A^0_{-m})$ iff $\beta_m \geq \bar{\beta}_m(A^0_{-m})$. That is, a no-decoy best response for seller $m$ to $A^0_{-m} \in C^0_{-m}$ is an overall best response iff seller $m$'s inter-decoy factor $\beta_m$ is greater than threshold $\bar{\beta}_m(A^0_{-m})$.*

*(2) For any $A^1_{-m} \in C^1_{-m}$ and any $A^1_m \in \texttt{PBR}^1_m(A^1_{-m})$, it holds that $A^1_m \in \texttt{PBR}_m(A^1_{-m})$ iff $\beta_m \leq \underline{\beta}_m(A^1_{-m})$. That is, a decoy best response for seller $m$ to $A^1_{-m} \in C^1_{-m}$ is an overall best response iff seller $m$'s inter-decoy factor $\beta_m$ is less than threshold $\underline{\beta}_m(A^1_{-m})$.*

*(3) For any $A^0_{-m} \in C^0_{-m}$ and any $A^1_m \in \texttt{PBR}^1_{-m}(A^0_{-m})$, it holds that $A^1_m \in \texttt{PBR}_{-m}(A^0_{-m})$ iff $\beta_m \leq \bar{\beta}_m(A^0_{-m})$. That is, a decoy best response for seller $m$ to $A^0_{-m} \in C^0_{-m}$ is an overall best response iff seller $m$'s inter-decoy factor $\beta_m$ is less than threshold $\bar{\beta}_m(A^0_{-m})$.*

*(4) For any $A^1_{-m} \in C^1_{-m}$ and any $A^0_m \in \texttt{PBR}^0_m(A^1_{-m})$, it holds that $A^0_m \in \texttt{PBR}_m(A^1_{-m})$ iff $\beta_m \geq \underline{\beta}_m(A^1_{-m})$. That is, a no-decoy best response for seller $m$ to $A^1_{-m} \in C^1_{-m}$ is an overall best response iff seller $m$'s inter-decoy factor $\beta_m$ is greater than threshold $\underline{\beta}_m(A^1_{-m})$.*

For $i \in \{0, 1\}$, and any $\bar{x}_{-m} \in \Delta(C_{-m})$, define the restricted mixed best response correspondences $\texttt{BR}^i_m : \Delta(C_{-m}) \mapsto 2^{\Delta(C^i_m)}$ as

$$\texttt{BR}^i_m(\bar{x}_{-m}) \quad := \quad \operatorname{argmax} \left\{ \begin{array}{c} \sum_{(A_{-1}, A_1) \in C} \pi_m(A_m, A_{-m}) \bar{x}_{-m}(A_{-m}) \bar{x}_m(A_m) \\ \text{s.t. } \bar{x}_m \in \Delta(C_m),\ \sum_{A^i_m \in C^i_m} \bar{x}_m(A^i_m) = 1 \end{array} \right\},$$

that is, $\texttt{BR}^0_m(\bar{x}_{-m})$ ($\texttt{BR}^1_m(\bar{x}_{-m})$) is the set of best responses of seller $m$ to assortment distribution $\bar{x}_{-m}$ among the assortment distributions that exclude (include) the decoy w.p.1. Also, define the restricted mixed best response correspondences $\texttt{BR}^2_m : \Delta(C_{-m}) \mapsto 2^{\Delta(C_m)}$ as

$$\texttt{BR}^2_m(\bar{x}_{-m}) \quad := \quad \operatorname{argmax} \left\{ \begin{array}{c} \sum_{(A_{-1}, A_1) \in C} \pi_m(A_m, A_{-m}) \bar{x}_{-m}(A_{-m}) \bar{x}_m(A_m) \\ \text{s.t. } \bar{x}_m \in \Delta(C_m),\ \sum_{A^i_m \in C^i_m} \bar{x}_m(A^i_m) > 0,\ i = 0, 1 \end{array} \right\},$$

that is, $\mathtt{BR}^2_m(\bar{x}_{-m})$ is the set of best responses of seller $m$ to assortment distribution $\bar{x}_{-m}$ among the assortment distributions that both exclude and include the decoy with positive probability.

For any $\bar{x}_{-m} \in \Delta(C_{-m})$ and $A_{-m} \in C^{+}_{-m}(\bar{x}_{-m})$, and any $A^i_m \in \mathtt{PBR}^i_m(\bar{x}_{-m})$, let

$$\bar{\beta}_m(A_{-m}, A^0_m, A^1_m) := \frac{v_{t_m} p_{t_m}}{v_{t_m} p_{t_m} + b_m(A^0_m)} \left( \alpha_m + \frac{\alpha_m - 1}{v_{t_{-m}}} + \frac{\alpha_m c(A_{-m}, A^0_m) - c(A_{-m}, A^1_m)}{v_{t_{-m}}} \right) + \frac{b_m(A^1_m)\left[1 + v_{t_m} + v_{t_{-m}} + c(A_{-m}, A^0_m)\right] - b_m(A^0_m)\left[1 + \alpha_m v_{t_m} + c(A_{-m}, A^1_m)\right]}{v_{t_{-m}}\left(v_{t_m} p_{t_m} + b_m(A^0_m)\right)}$$

$$\underline{\beta}_m(A_{-m}, A^0_m, A^1_m) := \frac{\beta_{-m} v_{t_m} p_{t_m}}{\beta_{-m} v_{t_m} p_{t_m} + b_m(A^0_m)} \left( \alpha_m + \frac{\alpha_m - 1}{\alpha_{-m} v_{t_{-m}}} + \frac{\alpha_m c(A_{-m}, A^0_m) - c(A_{-m}, A^1_m)}{\alpha_{-m} v_{t_{-m}}} \right) + \frac{b_m(A^1_m)\left[1 + \beta_{-m} v_{t_m} + \alpha_{-m} v_{t_{-m}} + c(A_{-m}, A^0_m)\right]}{\alpha_{-m} v_{t_{-m}}\left(\beta_{-m} v_{t_m} p_{t_m} + b_m(A^0_m)\right)} - \frac{b_m(A^0_m)\left[1 + \alpha_m \beta_{-m} v_{t_m} + c(A_{-m}, A^1_m)\right]}{\alpha_{-m} v_{t_{-m}}\left(\beta_{-m} v_{t_m} p_{t_m} + b_m(A^0_m)\right)}$$

$$\bar{\lambda}_m(A_{-m}, A^0_m, A^1_m) := \frac{v_{t_{-m}}\left(v_{t_m} p_{t_m} + b_m(A^0_m)\right)}{\left(1 + v_{t_m} + v_{t_{-m}} + c(A_{-m}, A^0_m)\right)\left(1 + \alpha_m v_{t_m} + \beta_m v_{t_{-m}} + c(A_{-m}, A^1_m)\right)}$$

$$\underline{\lambda}_m(A_{-m}, A^0_m, A^1_m) := \frac{\alpha_{-m} v_{t_{-m}}\left(\beta_{-m} v_{t_m} p_{t_m} + b_m(A^0_m)\right)}{\left(1 + \beta_{-m} v_{t_m} + \alpha_{-m} v_{t_{-m}} + c(A_{-m}, A^0_m)\right)\left(1 + \alpha_m \beta_{-m} v_{t_m} + \alpha_{-m} \beta_m v_{t_{-m}} + c(A_{-m}, A^1_m)\right)}$$

$$\beta_m(\bar{x}_{-m}) := \frac{\sum_{A_{-m} \in C^0_{-m}} \bar{\lambda}_m(A_{-m}, A^0_m, A^1_m) \bar{\beta}_m(A_{-m}, A^0_m, A^1_m) \bar{x}_{-m}(A_{-m})}{\sum_{A_{-m} \in C^0_{-m}} \bar{\lambda}_m(A_{-m}, A^0_m, A^1_m) \bar{x}_{-m}(A_{-m}) + \sum_{A_{-m} \in C^1_{-m}} \underline{\lambda}_m(A_{-m}, A^0_m, A^1_m) \bar{x}_{-m}(A_{-m})} + \frac{\sum_{A_{-m} \in C^1_{-m}} \underline{\lambda}_m(A_{-m}, A^0_m, A^1_m) \underline{\beta}_m(A_{-m}, A^0_m, A^1_m) \bar{x}_{-m}(A_{-m})}{\sum_{A_{-m} \in C^0_{-m}} \bar{\lambda}_m(A_{-m}, A^0_m, A^1_m) \bar{x}_{-m}(A_{-m}) + \sum_{A_{-m} \in C^1_{-m}} \underline{\lambda}_m(A_{-m}, A^0_m, A^1_m) \bar{x}_{-m}(A_{-m})}.$$

Proposition 3.13 shows that for each seller $m$ there is a best response strategy that always contains a decoy iff seller $m$'s inter-decoy factor $\beta_m$ is small relative to a threshold. In addition, there is a best response strategy that both contains a decoy and does not contain a decoy with positive probabilities iff seller $m$'s inter-decoy factor $\beta_m$ is equal to a threshold.

**Proposition 3.13.** *The following holds:*

*(1) For any $\bar{x}_{-m} \in \Delta(C_{-m})$ and any $\bar{x}^0_m \in \mathtt{BR}^0_m(\bar{x}_{-m})$, it holds that $\bar{x}^0_m \in \mathtt{BR}_m(\bar{x}_{-m})$ iff $\beta_m \geq \beta_m(\bar{x}^0_{-m})$. That is, a no-decoy best response for seller $m$ to $\bar{x}_{-m}$ is an overall best response iff seller $m$'s inter-decoy factor $\beta_m$ is greater than threshold $\beta_m(\bar{x}_{-m})$.*

*(2) For any $\bar{x}_{-m} \in \Delta(C_{-m})$ and any $\bar{x}^1_m \in \mathtt{BR}^1_m(\bar{x}_{-m})$, it holds that $\bar{x}^1_m \in \mathtt{BR}_m(\bar{x}_{-m})$ iff $\beta_m \leq \beta_m(\bar{x}_{-m})$. That is, a decoy best response for seller $m$ to $\bar{x}_{-m}$ is an overall best response iff seller $m$'s inter-decoy factor $\beta_m$ is less than threshold $\beta_m(\bar{x}_{-m})$.*

*(3) For any $\bar{x}_{-m} \in \Delta(C_{-m})$ and any $\bar{x}^2_m \in \mathtt{BR}^2_m(\bar{x}_{-m})$, it holds that $\bar{x}^2_m \in \mathtt{BR}_m(\bar{x}_{-m})$. Also, $\mathtt{BR}^2_m(\bar{x}_{-m}) \neq \varnothing$ iff $\beta_m = \beta_m(\bar{x}_{-m})$. That is, a mixed decoy/no-decoy best response for seller $m$ to $\bar{x}_{-m}$ is an overall best response, and a mixed decoy/no-decoy best response exists iff seller $m$'s inter-decoy factor $\beta_m$ is equal to the threshold $\beta_m(\bar{x}_{-m})$.*

### 3.6.2 Fictitious Play that Cycles

As we showed, the dynamic behavior of fictitious play with simple product sets and simple actions or general product sets and simple actions can be completely characterized with quite simple geometry. The dynamic behavior of fictitious play with general product sets and general actions is qualitatively more complicated, and does not allow a characterization as simple as that with two actions. Here we illustrate this point by example. We show that fictitious play for a duopoly in which each seller chooses among three assortments can cycle without convergence to any equilibrium. The example was constructed to satisfy the sufficient conditions specified in [60, p.25] for fictitious play to cycle without convergence to any Nash equilibrium.

Seller -1 has product set $S_{-1} = \{1, 2, \ldots, 5\}$ and seller 1 has product set $S_1 = \{6, 7, \ldots, 10\}$, where products 1 and 2 are respectively the target and decoy products of seller -1, and products 6 and 7 are respectively the target and decoy products of seller 1. Table 13 gives the attractiveness parameter $v_i$ of each product $i$. Note that $v_1 > v_2$ and $v_6 > v_7$, consistent with the idea that each seller's target dominates the seller's decoy in terms of attractiveness to buyers.

Table 13: The attractiveness parameter $v_i$ of each product $i$.

| Product # $i$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Attractiveness $v_i$ | 12 | 1.2 | $4.2 \times 10^5$ | $10^{-3}$ | $10^{-3}$ | 12 | 6.8 | $1.6 \times 10^5$ | $3.7 \times 10^3$ | $3.2 \times 10^5$ |

The decoy factors are $\alpha_{-1} = 18$, $\alpha_1 = 1.5 \times 10^3$, $\beta_{-1} = 660$ and $\beta_1 = 820$. Each seller chooses among three assortments to offer. Specifically, the sellers' action sets are $C_{-1} = \{\{1, 2, 4\}, \{1, 5\}, \{1, 3\}\}$ and $C_1 = \{\{6, 9\}, \{6, 7, 10\}, \{6, 8\}\}$. Table 14 gives the profit margins $p_i$ (i.e., the unit prices minus unit costs) of each product $i$.

Table 14: The profit margin $p_i$ of each product $i$.

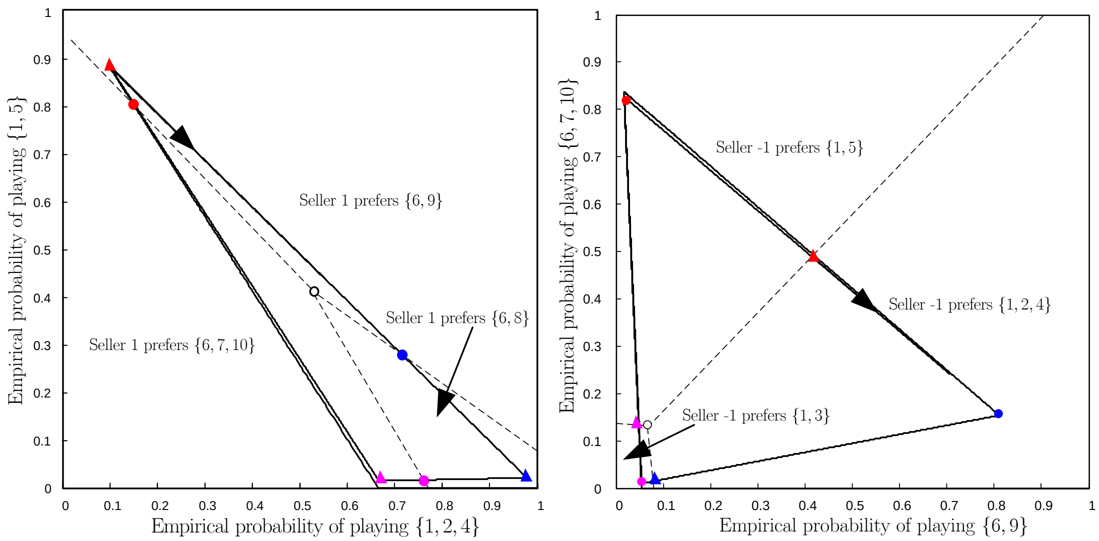
| Product # $i$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Profit margin $p_i$ | $1.8 \times 10^5$ | $2.1 \times 10^5$ | 570 | $8 \times 10^4$ | $1.5 \times 10^5$ | 350 | 570 | 370 | 400 | 400 |

The resulting objective functions $\pi_m(A_m, A_{-m})$ of the two sellers are given in Table 15. The unique equilibrium is the mixed-strategy equilibrium $\bar{x}^*_{-1} = (0.520, 0.411, 0.069)$ and $\bar{x}^*_1 = (0.057, 0.132, 0.811)$.

Table 15: The objective functions $\pi_m(A_m, A_{-m})$ of the two sellers.

| | | seller 1 | | |
|---|---|---|---|---|
| | | {6, 9} | {6, 7, 10} | {6, 8} |
| seller -1 | {1,2,4} | 3285, 359.1 | 2576, 346.2 | 231.2, 368.6 |
| | {1,5} | 579.9, 398.4 | 5092, 386.1 | 13.50, 369.9 |
| | {1,3} | 570.1, 3.503 | 2619, 175.0 | 416.5, 102.1 |

The initial conditions are $x_{-1}(0) = (1, 0, 0)$, $x_1(0) = (1, 0, 0)$, $M_{-1} = 5$, and $M_1 = 3$. Figure 11a shows the trajectory of empirical probabilities of seller $-1$ choosing assortments $\{1, 2, 4\}$ and $\{1, 5\}$, and Figure 11b shows the trajectory of empirical probabilities of seller 1 choosing assortments $\{6, 9\}$ and $\{6, 7, 10\}$, for $t = 0, 1, \ldots, 10^7$. The trajectories quickly converge to triangular limit cycles. Figure 11a also shows the regions (with dotted boundaries) in which each of the assortments of seller 1 is preferred by that seller, and Figure 11b shows similar preference regions for seller $-1$. The unique (mixed-strategy) equilibrium corresponds to the intersection points of the three regions for each seller. Also note that when the trajectory of one seller's empirical probabilities crosses a boundary between two regions, then the chosen assortment of the other seller changes, and thus the trajectory of the other seller's empirical probabilities changes direction. For example, when the trajectory of seller $-1$ crosses the boundary at the blue dot in Figure 11a, then the chosen assortment of seller 1 changes from $\{6, 9\}$ to $\{6, 8\}$, and thus the trajectory of seller 1 changes at the blue dot in Figure 11b from moving in the direction of $(1, 0)$ to moving in the direction of $(0, 0)$.




(a) seller -1 (b) seller 1

Figure 11: A trajectory of the fictitious play process for two sellers with three actions each.

## *3.7 Conclusion*

The decoy effect has been observed in a variety of choice settings, both experimental settings as well as real-life choice settings. We proposed a modified attraction choice model that is simple and that captures the decoy effect. We also studied assortment competition between two sellers who take the decoy effect into account. It was found that every type of pure-strategy Nash equilibrium — with neither seller offering a decoy, with one seller offering a decoy, and with both sellers offering a decoy — can occur in such a duopoly, and we characterized the conditions under which each type of equilibrium occurs. In short, it was found that if the effect of a seller's decoy on the attractiveness of the other seller's target (the inter-decoy effect) is small relative to the effect of the seller's decoy on the attractiveness of the seller's own target (the intra-decoy effect), then the seller chooses to offer the decoy.

We also studied the stability of the Nash equilibria under learning dynamics, to obtain a sense of whether the equilibria provide a potentially trustworthy forecast of the outcome of the competition. This is especially interesting and relevant in the settings with multiple pure-strategy Nash equilibria and a mixed-strategy Nash equilibrium. This type of investigation is not very common in the supply chain literature, but we think that questions and study of equilibrium stability should be more standard. In short, it was found that all pure-strategy Nash equilibria can provide reliable forecasts of the outcome of the competition in the sense that they have large domains of attraction. In contrast, mixed-strategy Nash equilibria have negligible domains of attraction, except for a special case, and thus we conclude that mixed-strategy Nash equilibria do not provide reliable forecasts of the outcome of the competition. Our results also provide a simple geometric characterization of the dynamics of fictitious play for general $2 \times 2$ games that is more complete than previous characterizations. The dynamics of fictitious play for more general games, including $3 \times 3$ games, has been shown to be qualitatively more complicated, and remains to be studied further.

# APPENDIX A

# PROOFS AND SUPPLEMENTARY MATERIAL FOR CHAPTER III

In this section, we provide proofs for the results in Chapter III that characterize Nash Equilibira and that characterize the dynamics of the Cournot adjustment and fictitious play processes. We also provide supporting material, including Lemmas and additional explanations.

## *A.1 Assortment Competition with Simple Product Sets*

### A.1.1 Proofs for the Characterization of Equilibria

**Proof of Proposition 3.1:** Consider the following four cases.

(1) $(A^0_{-1}, A^0_1)$ is a pure-strategy Nash equilibrium iff, for $m = \pm 1$,

$$\begin{aligned} & \pi_m(A^0_m, A^0_{-m}) && \geq && \pi_m(A^1_m, A^0_{-m}) \\ \Leftrightarrow \quad & \frac{v_{t_m} p_{t_m}}{1 + v_{t_m} + v_{t_{-m}}} && \geq && \frac{\alpha_m v_{t_m} p_{t_m}}{1 + \alpha_m v_{t_m} + \beta_m v_{t_{-m}}} \\ \Leftrightarrow \quad & \beta_m && \geq && \alpha_m + \frac{\alpha_m - 1}{v_{t_{-m}}} \quad = \quad \bar{\beta}_m\,. \end{aligned}$$

(2) $(A^1_{-1}, A^1_1)$ is a pure-strategy Nash equilibrium iff, for $m = \pm 1$,

$$\begin{aligned} & \pi_m(A^1_m, A^1_{-m}) && \geq && \pi_m(A^0_m, A^1_{-m}) \\ \Leftrightarrow \quad & \frac{\alpha_m \beta_{-m} v_{t_m} p_{t_m}}{1 + \alpha_m \beta_{-m} v_{t_m} + \alpha_{-m} \beta_m v_{t_{-m}}} && \geq && \frac{\beta_{-m} v_{t_m} p_{t_m}}{1 + \beta_{-m} v_{t_m} + \alpha_{-m} v_{t_{-m}}} \\ \Leftrightarrow \quad & \beta_m && \leq && \alpha_m + \frac{\alpha_m - 1}{\alpha_{-m} v_{t_{-m}}} \quad = \quad \underline{\beta}_m\,. \end{aligned}$$

(3) $(A^0_{-1}, A^1_1)$ is a pure-strategy Nash equilibrium iff

$$\begin{aligned} & \pi_{-1}(A^0_{-1}, A^1_1) && \geq && \pi_{-1}(A^1_{-1}, A^1_1) \\ \Leftrightarrow \quad & \frac{\beta_1 v_{t_{-1}} p_{t_{-1}}}{1 + \beta_1 v_{t_{-1}} + \alpha_1 v_{t_1}} && \geq && \frac{\alpha_{-1} \beta_1 v_{t_{-1}} p_{t_{-1}}}{1 + \alpha_{-1} \beta_1 v_{t_{-1}} + \alpha_1 \beta_{-1} v_{t_1}} \\ \Leftrightarrow \quad & 1 + \alpha_{-1} \beta_1 v_{t_{-1}} + \alpha_1 \beta_{-1} v_{t_1} && \geq && \alpha_{-1}(1 + \beta_1 v_{t_{-1}} + \alpha_1 v_{t_1}) \end{aligned}$$

$$\Leftrightarrow \quad 1+\alpha_1\beta_{-1}v_{t_1} \quad \geq \quad \alpha_{-1}(1+\alpha_1 v_{t_1})$$

$$\Leftrightarrow \quad \beta_{-1} \quad \geq \quad \alpha_{-1}+\frac{\alpha_{-1}-1}{\alpha_1 v_{t_1}} \quad = \quad \underline{\beta}_{-1}$$

and

$$\begin{aligned}
\pi_1(A_1^1, A_{-1}^0) &\geq \pi_1(A_1^0, A_{-1}^0)\\
\Leftrightarrow \quad \frac{\alpha_1 v_{t_1} p_{t_1}}{1+\beta_1 v_{t_{-1}}+\alpha_1 v_{t_1}} &\geq \frac{v_{t_1} p_{t_1}}{1+v_{t_{-1}}+v_{t_1}}\\
\Leftrightarrow \quad \alpha_1(1+v_{t_{-1}}+v_{t_1}) &\geq 1+\beta_1 v_{t_{-1}}+\alpha_1 v_{t_1}\\
\Leftrightarrow \quad \alpha_1(1+v_{t_{-1}}) &\geq 1+\beta_1 v_{t_{-1}}\\
\Leftrightarrow \quad \beta_1 &\leq \alpha_1+\frac{\alpha_1-1}{v_{t_{-1}}} = \bar{\beta}_1\,.
\end{aligned}$$

(4) Case (4) follows from Case (3) by interchanging -1 and 1. □

**Proof of Proposition 3.2:** The best response problem of seller $m$ in response to $x_{-m}$ is

$$\max_{x_m\in[0,1]} \quad \big\{x_m\left[x_{-m}\pi_m(A_m^1, A_{-m}^1)+(1-x_{-m})\pi_m(A_m^1, A_{-m}^0)\right]$$
$$+(1-x_m)\left[x_{-m}\pi_m(A_m^0, A_{-m}^1)+(1-x_{-m})\pi_m(A_m^0, A_{-m}^0)\right]\big\}.$$

A necessary and sufficient condition for $(x_{-1}^*, x_1^*)\in(0,1)^2$ to be a mixed-strategy Nash equilibrium is that the objective function of each seller $m$ is invariant in $x_m$ given $x_{-m}^*$, and thus $x_{-m}^*\pi_m(A_m^1, A_{-m}^1)+(1-x_{-m}^*)\pi_m(A_m^1, A_{-m}^0) = x_{-m}^*\pi_m(A_m^0, A_{-m}^1)+(1-x_{-m}^*)\pi_m(A_m^0, A_{-m}^0)$. It follows that

$$x_{-m}^* \quad = \quad \frac{\pi_m(A_m^0, A_{-m}^0)-\pi_m(A_m^1, A_{-m}^0)}{\pi_m(A_m^1, A_{-m}^1)-\pi_m(A_m^1, A_{-m}^0)-\pi_m(A_m^0, A_{-m}^1)+\pi_m(A_m^0, A_{-m}^0)} \quad = \quad \frac{1}{1+\Gamma_{-m}}$$

where

$$\begin{aligned}
\Gamma_{-m} &:= \frac{\pi_m(A_m^1, A_{-m}^1)-\pi_m(A_m^0, A_{-m}^1)}{\pi_m(A_m^0, A_{-m}^0)-\pi_m(A_m^1, A_{-m}^0)}\\
&= \frac{\frac{\alpha_m\beta_{-m}v_{t_m}p_{t_m}}{1+\alpha_m\beta_{-m}v_{t_m}+\alpha_{-m}\beta_m v_{t_{-m}}}-\frac{\beta_{-m}v_{t_m}p_{t_m}}{1+\beta_{-m}v_{t_m}+\alpha_{-m}v_{t_{-m}}}}{\frac{v_{t_m}p_{t_m}}{1+v_{t_m}+v_{t_{-m}}}-\frac{\alpha_m v_{t_m}p_{t_m}}{1+\alpha_m v_{t_m}+\beta_m v_{t_{-m}}}}\\
&= \frac{\beta_{-m}\alpha_{-m}(1+v_{t_m}+v_{t_{-m}})(1+\alpha_m v_{t_m}+\beta_m v_{t_{-m}})(\underline{\beta}_m-\beta_m)}{(1+\alpha_m\beta_{-m}v_{t_m}+\alpha_{-m}\beta_m v_{t_{-m}})(1+\beta_{-m}v_{t_m}+\alpha_{-m}v_{t_{-m}})(\beta_m-\bar{\beta}_m)}.
\end{aligned}$$

Note that if $\underline{\beta}_m<\beta_m<\bar{\beta}_m$ then $\Gamma_{-m}\in(0,\infty)$. □

### A.1.2 Proofs for the Cournot Adjustment Process

*A.1.2.1 Case 1*

Recall from Table 11 that Case 1 holds under condition

$$\underline{\beta}_m \quad < \quad \beta_m \quad < \quad \bar{\beta}_m, \quad m = \pm 1. \tag{A.1}$$

We first state Lemma A.1 that will be used to prove Theorem 3.1 and Proposition 3.3.

**Lemma A.1.** *Condition (A.1) holds if and only if*

$$\pi_m(A^1_m, A^0_{-m}) \quad > \quad \pi_m(A^0_m, A^0_{-m}) \quad \textit{and} \quad \pi_m(A^0_m, A^1_{-m}) \quad > \quad \pi_m(A^1_m, A^1_{-m}), \quad m = \pm 1.$$

**Proof**: Note that

$$\begin{aligned} \pi_m(A^1_m, A^0_{-m}) \quad &> \quad \pi_m(A^0_m, A^0_{-m}) \\ \Leftrightarrow \quad \frac{\alpha_m v_{t_m} p_{t_m}}{1 + \alpha_m v_{t_m} + \beta_m v_{t_{-m}}} \quad &> \quad \frac{v_{t_m} p_{t_m}}{1 + v_{t_m} + v_{t_{-m}}} \\ \Leftrightarrow \quad \beta_m \quad &< \quad \bar{\beta}_m, \end{aligned}$$

$$\begin{aligned} \pi_m(A^0_m, A^1_{-m}) \quad &> \quad \pi_m(A^1_m, A^1_{-m}) \\ \Leftrightarrow \quad \frac{\beta_{-m} v_{t_m} p_{t_m}}{1 + \beta_{-m} v_{t_m} + \alpha_{-m} v_{t_{-m}}} \quad &> \quad \frac{\alpha_m \beta_{-m} v_{t_m} p_{t_m}}{1 + \alpha_m \beta_{-m} v_{t_m} + \alpha_{-m} \beta_m v_{t_{-m}}} \\ \Leftrightarrow \quad \underline{\beta}_m \quad &< \quad \beta_m. \end{aligned}$$

□

**Proof of Theorem 3.1:**

(1) For any $t \in \mathbb{N}_0$ and $m = \pm 1$, if $A(t) = (A^0_{-m}, A^1_m)$, i.e., $A_{-m}(t) = A^0_{-m}$ and $A_m(t) = A^1_m$, it follows from Lemma A.1 that $\pi_m(A^1_m, A^0_{-m}) > \pi_m(A^0_m, A^0_{-m})$ and that $\pi_{-m}(A^0_{-m}, A^1_m) > \pi_{-m}(A^1_{-m}, A^1_m)$. Thus it follows from (3.6) that $A_{-m}(t+1) = A^0_{-m}$ and $A_m(t+1) = A^1_m$, which implies that $(A^0_{-m}, A^1_m)$ is a steady state.

(2) For any $t \in \mathbb{N}_0$, it follows from Lemma A.1 that, if $A(t) = (A^1_{-1}, A^1_1)$, then $A(t+1) = (A^0_{-1}, A^0_1)$, and if $A(t) = (A^0_{-1}, A^0_1)$, then $A(t+1) = (A^1_{-1}, A^1_1)$. Thus, if $A(0) =$

$(A^1_{-1}, A^1_1)$,

$$A(t) = \begin{cases} (A^1_{-1}, A^1_1) & \text{if} \quad t \geq 1 \text{ and } t \text{ is even}, \\ (A^0_{-1}, A^0_1) & \text{if} \quad t \geq 1 \text{ and } t \text{ is odd}. \end{cases}$$

and if $A(0) = (A^0_1, A^0_{-1})$, it follows

$$A(t) = \begin{cases} (A^0_{-1}, A^0_1) & \text{if} \quad t \geq 1 \text{ and } t \text{ is even}, \\ (A^1_{-1}, A^1_1) & \text{if} \quad t \geq 1 \text{ and } t \text{ is odd}. \end{cases}$$

□

**Proof of Proposition 3.3:** It follows from Lemma A.1 that

$$\frac{1}{2}\pi_m(A^1_m, A^1_{-m}) \quad < \quad \frac{1}{2}\pi_m(A^0_m, A^1_{-m}) \text{ and } \frac{1}{2}\pi_m(A^0_m, A^0_{-m}) \quad < \quad \frac{1}{2}\pi_m(A^1_m, A^0_{-m}).$$

Thus, condition (3.3) fails to hold and the limit empirical joint distribution $\bar{x}^*(A^1_{-1}, A^1_1) = 1/2$ and $\bar{x}^*(A^0_{-1}, A^0_1) = 1/2$ is not a correlated equilibrium. Since $0.5\pi_m(A^1_m, A^1_{-m}) + 0.5\pi_m(A^0_m, A^0_{-m}) < 0.5\pi_m(A^1_m, A^1_{-m}) + 0.5\pi_m(A^1_m, A^0_{-m})$, condition (3.4) fails to hold and the limit empirical joint distribution $\bar{x}^*(A^1_{-1}, A^1_1) = 1/2$ and $\bar{x}^*(A^0_{-1}, A^0_1) = 1/2$ is not a coarse correlated equilibrium. □

*A.1.2.2 Case 2*

Under Case 2, it holds that

$$\beta_m \quad < \quad \bar{\beta}_m \quad \text{and} \quad \beta_m \quad < \quad \underline{\beta}_m, \quad m = \pm 1. \tag{A.2}$$

**Lemma A.2.** *Suppose that (A.2) holds. Then, $A^1_m$ dominates $A^0_m$ for $m = \pm 1$.*

**Proof**: Note that, for $m = \pm 1$

$$\begin{aligned} \pi_m(A^1_m, A^1_{-m}) \quad &> \quad \pi_m(A^0_m, A^1_{-m}) \\ \Leftrightarrow \quad \frac{\alpha_m \beta_{-m} v_{t_m} p_{t_m}}{1 + \alpha_m \beta_{-m} v_{t_m} + \alpha_{-m} \beta_m v_{t_{-m}}} \quad &> \quad \frac{\beta_{-m} v_{t_m} p_{t_m}}{1 + \beta_{-m} v_{t_m} + \alpha_{-m} v_{t_{-m}}} \\ \Leftrightarrow \quad \beta_m \quad &< \quad \underline{\beta}_m, \end{aligned}$$

and

$$\pi_m(A^1_m, A^0_{-m}) \quad > \quad \pi_m(A^0_m, A^0_{-m})$$

$$\Leftrightarrow \quad \frac{\alpha_m v_{t_m} p_{t_m}}{1+\alpha_m v_{t_m}+\beta_m v_{t_{-m}}} \quad > \quad \frac{v_{t_m} p_{t_m}}{1+v_{t_m}+v_{t_{-m}}}$$

$$\Leftrightarrow \quad \beta_m \quad < \quad \bar{\beta}_m.$$

This shows that $A_m^1$ dominates $A_m^0$ for $m = \pm 1$. □

**Proof of Proposition 3.4:** The result follows from Lemma A.2. □

### A.1.3 Continuous-time Fictitious Play under Case 1

Before investigating the more complicated discrete-time fictitious play process (3.9), we consider its simpler continuous-time analogue. Similar to (3.7) and (3.9), let $x(t) := (x_{-1}(t), x_1(t))$ denote the state at time $t \in \mathbb{R}_+$ with specified initial condition $x(0)$, and let the dynamics be given by

$$\dot{x}_m(t) \quad = \quad \mathbf{1}_{[x_{-m}(t)\le x^*_{-m}]} - x_m(t), \qquad m = \pm 1. \tag{A.3}$$

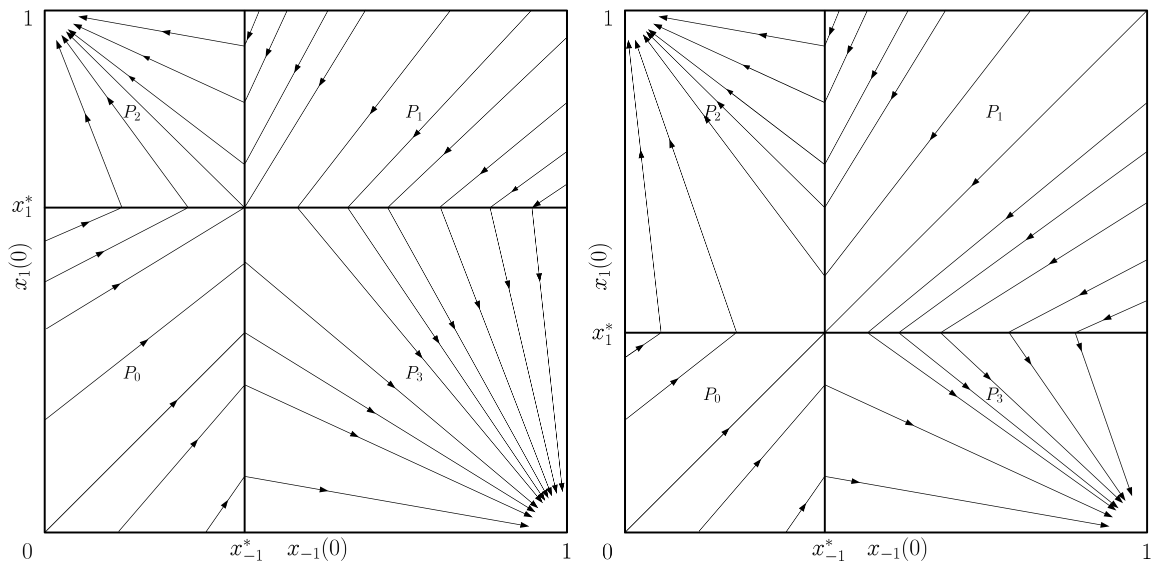
The resulting trajectory $x(t)$ is given by

$$x(t) \quad = \quad \begin{cases} \left(1-e^{-t}[1-x_{-1}(0)], 1-e^{-t}[1-x_1(0)]\right) & \text{for} \quad x(0)\in P_0,\ t\in[0,\bar{t}_1], \\ \left(e^{-t}x_{-1}(0), e^{-t}x_1(0)\right) & \text{for} \quad x(0)\in P_1,\ t\in[0,\bar{t}_2], \\ \left(e^{-t}x_{-1}(0), 1-e^{-t}[1-x_1(0)]\right) & \text{for} \quad x(0)\in P_2,\ t\ge 0, \\ \left(1-e^{-t}[1-x_{-1}(0)], e^{-t}x_1(0)\right) & \text{for} \quad x(0)\in P_3,\ t\ge 0, \end{cases} \tag{A.4}$$

where $\bar{t}_1 := \min_{m=\pm1}\left\{\ln\left(\frac{1-x_m(0)}{1-x^*_m}\right)\right\}$ and $\bar{t}_2 =:= \min_{m=\pm1}\left\{\ln\left(\frac{x_m(0)}{x^*_m}\right)\right\}$.

Trajectories $x(t)$ starting from various initial points $x(0)$ are shown in Figure 12a for $x^*_{-1} \neq x^*_1$ and in Figure 12b for $x^*_{-1} = x^*_1$. Note that if $x(0) \in P_0$, then $x(t)$ reaches the boundary of $P_0$ at $t_0 := \min_{m=\pm1}\{\ln([1-x_m(0)]/[1-x^*_m])\}$. If $\ln\left([1-x_{-1}(0)]/[1-x^*_{-1}]\right) < \ln([1-x_1(0)]/[1-x^*_1])$, then $x_{-1}(t_0) = x^*_{-1}$, and at time $t_0$ the trajectory enters $P_3$, and thus $x(t) = \left(1-e^{-(t-t_0)}[1-x_{-1}(t_0)], e^{-(t-t_0)}x_1(t_0)\right)$ for $t > t_0$. Similarly, if $x(0) \in P_0$ and $\ln\left([1-x_{-1}(0)]/[1-x^*_{-1}]\right) > \ln([1-x_1(0)]/[1-x^*_1])$, then $x_1(t_0) = x^*_1$, and at time $t_0$ the trajectory enters $P_2$, and thus $x(t) = \left(e^{-(t-t_0)}x_{-1}(t_0), 1-e^{-(t-t_0)}[1-x_1(t_0)]\right)$ for $t > t_0$. However, if $x(0) \in P_0$ and $\ln\left([1-x_{-1}(0)]/[1-x^*_{-1}]\right) = \ln([1-x_1(0)]/[1-x^*_1])$, then (A.3) allows three solutions:

$$x(t) \quad = \quad \left(1-e^{-(t-t_0)}[1-x_{-1}(t_0)], e^{-(t-t_0)}x_1(t_0)\right),$$




(a) $x_1^* \neq x_{-1}^*$ (b) $x_1^* = x_{-1}^*$

Figure 12: Trajectories of the continuous-time fictitious play process starting from various initial points $x(0)$.

or

$$x(t) \quad = \quad \left(e^{-(t-t_0)}x_{-1}(t_0), 1 - e^{-(t-t_0)}[1 - x_1(t_0)]\right),$$

or

$$x(t) \quad = \quad x^*$$

for $t > t_0$. Similar comments apply to the case with $x(0) \in P_1$. Note that $(A_{-1}^0, A_1^1)$ and $(A_{-1}^1, A_1^0)$ are both attracting equilibria, each with an easily identified domain of attraction, whereas $x^*$ is an unstable equilibrium. The solution given in (A.4) for continuous-time fictitious play can be regarded as a simplified approximation of the trajectories of the (discrete-time) fictitious play process; the discrepancies being caused by the discrete steps taken in the latter process.

### A.1.4 Proofs and Additional Results for Discrete-time Fictitious Play under Case 1

*A.1.4.1 Proof of Convergence from $P_2$ and $P_3$*

**Proof of Theorem 3.2:** Consider any $x \in P_2$ and $t \in \mathbb{N}_0$. It follows from Lemma 3.1 that $\phi(t+\tau,t,x) \in P_2$ for all $\tau \in \mathbb{N}_0$. Thus, it follows from (3.10) that

$$\begin{aligned}
\phi_m(t+\tau,t,x) &= \frac{(M_m+t)x_m(t) + \sum_{i=1}^{\tau} \mathbf{1}_{[A_m(t+i)=A_m^1]}}{M_m+t+\tau} \\
&= \begin{cases} \frac{(M_m+t)x_m(t)}{M_m+t+\tau} & \text{if} \quad m=-1, \\ \frac{(M_m+t)x_m(t)+\tau}{M_m+t+\tau} & \text{if} \quad m=1, \end{cases}
\end{aligned}$$

for all $\tau \in \mathbb{N}_0$, and thus $\phi_{-1}(t+\tau,t,x) \to 0$ and $\phi_1(t+\tau,t,x) \to 1$ as $\tau \to \infty$, that is, $\phi(t+\tau,t,x) \to (0,1)$ as $\tau \to \infty$.

Consider any $x \in P_3$ and $t \in \mathbb{N}_0$. It follows from Lemma 3.1 that $\phi(t+\tau,t,x) \in P_3$ for all $\tau \in \mathbb{N}_0$. Thus it follows from (3.10) that

$$\phi_m(t+\tau,t,x) = \begin{cases} \frac{(M_m+t)x_m(t)+\tau}{M_m+t+\tau} & \text{if} \quad m=-1, \\ \frac{(M_m+t)x_m(t)}{M_m+t+\tau} & \text{if} \quad m=1, \end{cases}$$

for all $\tau \in \mathbb{N}_0$, and hence $\phi(t+\tau,t,x) \to (1,0)$ as $\tau \to \infty$. □

*A.1.4.2 Properties of an Increasing Separable Affine Mapping*

The following properties of an increasing separable affine mapping will be important. For any $D \subset \mathbb{R}^n$, let $\bar{D}$ denote the closure of $D$ in $\mathbb{R}^n$, and let $\partial D$ denote the boundary of $D$ in $\mathbb{R}^n$.

**Lemma A.3.** *Consider any increasing separable affine mapping $f : \mathbb{R}^n \mapsto \mathbb{R}^n$. Then, the following properties hold:*

*(1) There exists $\ell \in (0,\infty)^n$ such that $f_m(x) - f_m(y) = \ell_m(x_m - y_m)$ for all $x, y \in \mathbb{R}^n$ and $m \in \{1,\dots,n\}$.*

*(2) For any $D \subset \mathbb{R}^n$, $f : D \mapsto f(D) := \{f(x) : x \in D\}$ is a bijection, and $f : \mathbb{R}^n \mapsto \mathbb{R}^n$ is a bijection.*

*(3) For any $D \subset \mathbb{R}^n$, it holds that $E \subset D$ if and only if $f(E) \subset f(D)$, $E \subset \partial D$ if and only if $f(E) \subset \partial(f(D))$, and $E \cap D = \varnothing$ if and only if $f(E) \cap f(D) = \varnothing$.*

*(4) For any $D \subset \mathbb{R}^n$ it holds that $f(conv(D)) = conv(f(D))$.*

**Proof:** Since $f$ is an increasing separable affine mapping, there exists $\ell_m > 0$ and $a_m \in \mathbb{R}$ such that $f_m(x) = \ell_m x_m + a_m$ for all $x \in \mathbb{R}^n$ and $m \in \{1, 2, \ldots, n\}$. Let $\ell_{\max} := \max_{m \in \{1,2,\ldots,n\}} \ell_m$ and $\ell_{\min} := \min_{m \in \{1,2,\ldots,n\}} \ell_m$.

(1) For any $x, y \in \mathbb{R}^n$ and $m \in \{1, 2, \ldots, n\}$, it holds that $f_m(x) - f_m(y) = \ell_m(x_m - y_m)$ for all $m$.

(2) By the definition of $f(D)$, $f : D \mapsto f(D)$ is a surjection. Consider any $y \in \mathbb{R}^n$. Choose $x \in \mathbb{R}^n$ such that $x_m = (y_m - a_m)/\ell_m$ for all $m$, and note that $f(x) = y$. Thus $f : \mathbb{R}^n \mapsto \mathbb{R}^n$ is a surjection. Consider any $x^1, x^2 \in D$. Note that $f(x^1) = f(x^2)$ implies that $\ell_m x^1_m = \ell_m x^2_m$ for all $m$, and thus, $x^1 = x^2$. Hence $f$ is a bijection.

(3) It follows from $f : \mathbb{R}^n \mapsto \mathbb{R}^n$ being a bijection that $E \subset D$ if and only if $f(E) \subset f(D)$ and $E \cap D = \varnothing$ if and only if $f(E) \cap f(D) = \varnothing$.

Suppose that $E \subset \partial D$. Choose any $y \in f(E)$ and any neighborhood $B(y, \varepsilon) := \{y' \in \mathbb{R}^n : \|y' - y\|_\infty < \varepsilon\}$, where $\varepsilon > 0$. There exists $x \in E$ such that $f(x) = y$. Since $x \in E \subset \partial D$, there exists $x^1, x^2 \in B(x, \varepsilon/\ell_{\max})$ such that $x^1 \in D$ and $x^2 \notin D$. For $i \in \{1, 2\}$, it holds that $\|f(x^i) - y\|_\infty = \|f(x^i) - f(x)\|_\infty \leq \ell_{\max}\|x^i - x\|_\infty < \varepsilon$. Thus, $f(x^i) \in B(y, \varepsilon)$ for $i \in \{1, 2\}$. Note that $f(x^1) \in f(D)$ but $f(x^2) \notin f(D)$, since $f : \mathbb{R}^n \mapsto \mathbb{R}^n$ is a bijection by Lemma A.3(2). Thus, $y \in \partial(f(D))$, and hence $f(E) \subset \partial(f(D))$.

Suppose that $f(E) \subset \partial(f(D))$. Choose any $x \in E$ and any neighborhood $B(x, \varepsilon)$. Since $f(x) \in f(E) \subset \partial(f(D))$, there exists $y^1, y^2 \in B(f(x), \varepsilon\ell_{\min})$ such that $y^1 \in f(D)$ and $y^2 \notin f(D)$. Since $f$ is a bijection, there exists $x^1 \in D$ and $x^2 \notin D$ such that $f(x^i) = y^i$ for $i \in \{1, 2\}$. For $i \in \{1, 2\}$, it holds that $\|x^i - x\|_\infty \leq \|f(x^i) - f(x)\|_\infty/\ell_{\min} = \|y^i - f(x)\|_\infty/\ell_{\min} < \varepsilon$. Thus, $x^1, x^2 \in B(x, \varepsilon)$. Thus, $x \in \partial(D)$, and hence $E \subset \partial D$.

(4) Result (4) holds for any affine mapping $f$. $\square$

We will be interested in cases in which $D$ is a rectangle of the form $\prod_{m=1}^{n}[x_m^1, x_m^2]$ or $\prod_{m=1}^{n}(x_m^1, x_m^2]$, or the boundaries or vertices of such a rectangle. The following results will be useful.

**Lemma A.4.** *Consider any increasing separable affine mapping* $f : \mathbb{R}^n \mapsto \mathbb{R}^n$, $x^1, x^2 \in \mathbb{R}^n$ *such that* $x^1 \leq x^2$, *and* $D := \prod_{m=1}^{n}[x_m^1, x_m^2] \subset \mathbb{R}^n$. *Then,* $f(D) = \prod_{m=1}^{n}[f_m(x^1), f_m(x^2)]$.

**Proof:** Let $f$ be given by $f_m(x) = \ell_m x_m + a_m$, where $\ell_m > 0$. Note that, for any $m \in \{1, 2, \ldots, n\}$, $x_m^1 \leq x_m^2$ if and only if $f_m(x^1) \leq f_m(x^2)$. First consider any $x \in D$. Then $x_m \in [x_m^1, x_m^2]$, and thus $f_m(x) = \ell_m x_m + a_m \in [\ell_m x_m^1 + a_m, \ell_m x_m^2 + a_m] = [f_m(x^1), f_m(x^2)]$. Hence, $f(D) \subset \prod_{m=1}^{n}[f_m(x^1), f_m(x^2)]$. Next consider any $y \in \prod_{m=1}^{n}[f_m(x^1), f_m(x^2)]$. Then $y_m \in [f_m(x^1), f_m(x^2)] = [\ell_m x_m^1 + a_m, \ell_m x_m^2 + a_m]$, and thus $x_m := (y_m - a_m)/\ell_m \in [x_m^1, x_m^2]$. Note that $x := (x_1, \ldots, x_n) \in D$ and $f(x) = y$, and thus $y \in f(D)$. Hence, $\prod_{m=1}^{n}[f_m(x^1), f_m(x^2)] \subset f(D)$. Therefore, $f(D) = \prod_{m=1}^{n}[f_m(x^1), f_m(x^2)]$. $\square$

Lemma A.4 has several useful implications that we point out next.

**Remark A.1.** *Let* $D := [x_{-1}^1, x_{-1}^2] \times [x_1^1, x_1^2]$ *where* $x^1 < x^2$, *and let* $E := (x_{-1}^1, x_{-1}^2] \times (x_1^1, x_1^2]$.

*(1)* $f(D) = [f_{-1}(x^1), f_{-1}(x^2)] \times [f_1(x^1), f_1(x^2)]$ *and* $f(x^1) < f(x^2)$.

*(2)* $f$ *maps a boundary of* $D$ *to the corresponding boundary of* $f(D)$. *For example,*

$$f(\{x_{-1}^i\} \times [x_1^1, x_1^2]) \quad = \quad \{f_{-1}(x^i)\} \times [f_1(x^1), f_1(x^2)]$$

*and*

$$f([x_{-1}^1, x_{-1}^2] \times \{x_1^i\}) = [f_{-1}(x^1), f_{-1}(x^2)] \times \{f_1(x^i)\}$$

*for* $i \in \{1, 2\}$.

*(3)* $f$ *maps a vertex of* $D$ *into the corresponding vertex of* $f(D)$. *For example,* $f((x_{-1}^i, x_1^j)) = (f_{-1}(x^i), f_1(x^j))$ *for* $i, j \in \{1, 2\}$.

*(4) It follows that if the left/right/upper/lower edge of a rectangle is included/excluded in/from the rectangle, then the left/right/upper/lower edge of the image of the rectangle is included/excluded in/from the image of the rectangle. For example,* $f(E) = (f_{-1}(x^1), f_{-1}(x^2)] \times (f_1(x^1), f_1(x^2)]$.

*A.1.4.3 One-step Analysis*

**Proof of Lemma 3.2:** Because $\phi(t+1,t,\cdot) : \hat{P}_k \mapsto \phi(t+1,t,\hat{P}_k)$, where $k \in \{0,1\}$, and $\phi(t+1,t,\cdot) : P_k \mapsto \phi(t+1,t,P_k)$, where $k \in \{2,3\}$, are increasing separable affine mappings, the result follows from Remark A.1. □

**Proof of Proposition 3.5:** Consider $k = 0$. Consider any

$$x = (x_{-1}, x_1) \in D_{0,i,j}(t) := (x^*_{-1} - j\delta^0_{-1}(t),\, x^*_{-1} - (j-1)\delta^0_{-1}(t)] \times (x^*_1 - i\delta^0_1(t),\, x^*_1 - (i-1)\delta^0_1(t)].$$

Let $i_1 := i$ and $i_{-1} := j$. It follows from (3.10) that

$$\begin{aligned}
\phi_m(t+1,t,x) &= \frac{(M_m+t)x_m + 1}{M_m+t+1} \\
&\in \frac{M_m+t}{M_m+t+1}\left(x^*_m - i_m\delta^0_m(t),\, x^*_m - (i_m-1)\delta^0_m(t)\right] + \frac{1}{M_m+t+1} \\
&= (x^*_m - (i_m-1)\delta^0_m(t+1),\, x^*_m - (i_m-2)\delta^0_m(t+1)],
\end{aligned}$$

for $m = \pm 1$, and thus $\phi(t+1, D_{0,i,j}(t)) \subset D_{0,i-1,j-1}(t+1)$.

Next we show that $D_{0,i-1,j-1}(t+1) \subset \phi(t+1, D_{0,i,j}(t))$. Consider any $\tilde{x} \in D_{0,i-1,j-1}(t+1)$, that is,

$$\tilde{x}_m \in \left(x^*_m - (i_m-1)\delta^0_m(t+1),\, x^*_m - (i_m-2)\delta^0_m(t+1)\right], \quad m = \pm 1.$$

Then consider $x = (x_{-1}, x_1)$, where

$$x_m = \frac{(M_m+t+1)\tilde{x}_m - 1}{M_m+t}$$

for $m = \pm 1$. It is easy to verify that $\phi(t+1,t,x) = \tilde{x}$. Next we show that $x \in D_{0,i,j}(t)$. Note that

$$x_m = \frac{(M_m+t+1)\tilde{x}_m - 1}{M_m+t}$$

$$\in \frac{M_m+t+1}{M_m+t}\left(x_m^*-(i_m-1)\delta_m^0(t+1),\, x_m^*-(i_m-2)\delta_m^0(t+1)\right]-\frac{1}{M_m+t}$$
$$= \left(x_m^*-i_m\delta_m^0(t),\, x_m^*-(i_m-1)\delta_m^0(t)\right].$$

Thus, $x \in D_{0,i,j}(t)$. Hence, $D_{0,i-1,j-1}(t+1) = \phi(t+1,t,D_{0,i,j}(t))$. A similar argument can be used for the case $k = 1$. □

*A.1.4.4 Multi-step Analysis*

**Proof of Lemma 3.3:** Let $f := f_{t+1}^{t+\tau}$. Since $D$ walks to $\tilde{D}$, $f$ is an increasing separable affine mapping such that $\tilde{D} = f(D)$. For any $E \subset D$, it holds that $f(E) \subset f(D)$ and that $E$ walks to $f(E)$. □

Recall from Lemma 3.1 that if $x(t) \le x^*$ (and thus $x(t) \in \hat{P}_0$), then $x(t+1) > x(t)$, and thus $x(t+\tau)$ is increasing in $\tau$ until $x_m(t+\tau) > x_m^*$ for some $m \in \{-1,1\}$ and some $\tau \in \mathbb{N}_0$. Similarly, if $x(t) > x^*$ (and thus $x(t) \in \hat{P}_1$), then $x(t+1) < x(t)$, and thus $x(t+\tau)$ is decreasing in $\tau$ until $x_m(t+\tau) \le x_m^*$ for some $m \in \{-1,1\}$ and some $\tau \in \mathbb{N}_0$. Given $x(t) \in \hat{P}_0$ ($x(t) \in \hat{P}_1$), we are interested in the first time $t+\tau$ such that $x(t+\tau) \notin \hat{P}_0$ ($x(t+\tau) \notin \hat{P}_1$). For any $t \in \mathbb{N}_0$ and $x \in \hat{P}_0 \cup \hat{P}_1$, let

$$T_m(t,x) \quad := \quad \begin{cases} \inf\{\tau \ge 1 \;:\; \phi_m(t+\tau,t,x) > x_m^*\} & \text{if} \quad x \in \hat{P}_0 \\ \inf\{\tau \ge 1 \;:\; \phi_m(t+\tau,t,x) \le x_m^*\} & \text{if} \quad x \in \hat{P}_1 \end{cases}$$

and let $T(t,x) := \min\{T_{-1}(t,x), T_1(t,x)\}$ denote the first time at which $x(t+\tau)$ leaves the region of $x(t)$. Lemma A.5 relates the time $T(t,x)$ with the cell where $x(t)$ resides.

**Lemma A.5.** *Consider any $x \in D_{k,i,j}(t)$, where $t \in \mathbb{N}_0$, $k \in \{0,1\}$, $i \in I_k(t)$, and $j \in J_k(t)$. If $i \le j$, then $T_1(t,x) = i \le T_{-1}(t,x)$, and if $i \ge j$, then $T_{-1}(t,x) = j \le T_1(t,x)$. Thus, $T(t,x) = \min\{i,j\}$.*

**Proof**: Consider the case with $k = 0$ and $i < j$. Then it follows from Proposition 3.5 that $D_{0,i,j}(t)$ walks to $D_{0,(i-\tau),(j-\tau)}(t+\tau) \subset \hat{P}_0$ from time $t$ to time $t+\tau$ for all $\tau \in \{0,1,\ldots,i-1\}$. It also follows that $D_{0,i,j}(t)$ walks to $D_{0,0,(j-i)}(t+i) \subset P_2$ from time $t$ to time $t+i$. Thus, $\phi(t+\tau,t,x) \in D_{0,(i-\tau),(j-\tau)}(t+\tau) \subset \hat{P}_0$ for all $\tau \in \{0,1,\ldots,i-1\}$, and $\phi(t+i,t,x) \in D_{0,0,(j-i)}(t+i) \subset P_2$. Hence, $\phi_1(t+\tau,t,x) \le x_1^*$ for all $\tau \in \{0,1,\ldots,i-1\}$ and $\phi_1(t+i,t,x) > x_1^*$, which implies that $T_1(t,x) = i$. Also, $\phi_{-1}(t+\tau,t,x) \le x_{-1}^*$ for all

$\tau \in \{0, 1, \ldots, i\}$, which implies that $T_{-1}(t, x) > i = T_1(t, x)$. The other cases can be proved by a similar argument. □

**Proof of Theorem 3.3:** Theorem 3.3 follows from Theorem 3.2 and Lemma A.5. □

**Proof of Proposition 3.6:** Proposition 3.6 follows from Proposition 3.5 and Lemma A.5. □

**Lemma A.6.** *For any $t \in \mathbb{N}_0$, it holds that $\phi(t+1, t, P(t)) \subset P \subset P(t+1)$.*

**Proof**: Consider any $x \in P(t)$. If $x \in P_2 \cup P_3$, then it follows from Lemma 3.1 that $\phi(t+1, t, x) \in P_2 \cup P_3 \subset P$. If $x \in P_0(t)$, then $-\delta_m^0(t) \leq x_m \leq x_m^*$ and $\phi_m(t+1, t, x) = [(M_m + t)x_m + 1]/(M_m + t + 1)$. Thus,

$$\begin{aligned} \phi_m(t+1, t, x) &\in \left[\frac{(M_m + t)(-\delta_m^0(t)) + 1}{M_m + t + 1}, \frac{(M_m + t)x_m^* + 1}{M_m + t + 1}\right] \\ &= \left[\frac{-(1 - x_m^*) + 1}{M_m + t + 1}, \frac{(M_m + t)x_m^* + 1}{M_m + t + 1}\right] \\ &= \left[\frac{x_m^*}{M_m + t + 1}, \frac{(M_m + t)x_m^* + 1}{M_m + t + 1}\right] \subset [0, 1] \end{aligned}$$

for $m = \pm 1$. If $x \in P_1(t)$, then $x_m^* < x_m < 1 + \delta_m^1(t)$ and $\phi_m(t+1, t, x) = (M_m + t)x_m/(M_m + t + 1)$. Thus,

$$\begin{aligned} \phi_m(t+1, t, x) &\in \left(\frac{(M_m + t)x_m^*}{M_m + t + 1}, \frac{(M_m + t)[1 + \delta_m^1(t)]}{M_m + t + 1}\right) \\ &= \left(\frac{(M_m + t)x_m^*}{M_m + t + 1}, \frac{(M_m + t)1 + x_m^*}{M_m + t + 1}\right) \subset [0, 1] \end{aligned}$$

for $m = \pm 1$. Therefore, $\phi(t+1, t, P(t)) \subset P \subset P(t+1)$. □

**Proof of Proposition 3.7:** It follows from Theorems 3.2 and 3.3 that if $\phi(t, 0, x) \in \cup_{D \in \mathcal{D}_2(t)} D \cup P_2$ at any time $t \geq 0$ (that is, $\phi(t, 0, x)$ is above the diagonal cells at time $t$), then $\phi(\tau, 0, x) \to (0, 1)$ as $\tau \to \infty$, and if $\phi(t, 0, x) \in \cup_{D \in \mathcal{D}_3(t)} D \cup P_3$ at any time $t \geq 0$ (that is, $\phi(t, 0, x)$ is below the diagonal cells at time $t$), then $\phi(\tau, 0, x) \to (1, 0)$ as $\tau \to \infty$.

Next, suppose that $\phi(t, 0, x) \in \cup_{D \in \mathcal{D}_=(t)} D$ for all $t$. We show that $\phi(t, 0, x) \to x^*$ as $t \to \infty$. Since $\phi(t, 0, x)$ is in a diagonal cell for all $t$, it follows from Lemma A.5 that $T(t, \phi(t, 0, x)) = T_{-1}(t, \phi(t, 0, x)) = T_1(t, \phi(t, 0, x)) < \infty$ for all $t$. Let $\Delta T_{n+1}$ denote the number of steps between the $n^{\text{th}}$ and the $(n+1)^{\text{th}}$ jumps of the sequence $\{\phi(t, 0, x)\}_{t=0}^{\infty}$.

Thus, $\{\Delta T_n\}_{n=1}^{\infty}$ is given by

$$\begin{array}{rclcrcl}
\Delta T_1 & := & T(0,x), & & \tau_1 & := & \Delta T_1, \\
\Delta T_2 & := & T(\tau_1, \phi(\tau_1,0,x)), & & \tau_2 & := & \tau_1 + \Delta T_2, \\
\Delta T_3 & := & T(\tau_2, \phi(\tau_2,0,x)), & & \tau_3 & := & \tau_2 + \Delta T_3, \\
 & \vdots & & & & \vdots & \\
\Delta T_n & := & T(\tau_{n-1}, \phi(\tau_{n-1},0,x)), & & \tau_n & := & \tau_{n-1} + \Delta T_n.
\end{array}$$

Suppose that $x \in P_0(0)$. Then $\phi(\tau_{2n-1},0,x) \in D_{0,0,0}(\tau_{2n-1}) \subset P_1$ and $\phi(\tau_{2n},0,x) \in D_{1,0,0}(\tau_{2n}) \subset P_0$ for all $n$. Note that $x_m^* - \delta_m^1(\tau_{2n}) \leq \phi_m(t,0,x) \leq x_m^* + \delta_m^0(\tau_{2n-1})$ for all $t \in \{\tau_{2n-1},\ldots,\tau_{2n}\}$, and $x_m^* - \delta_m^1(\tau_{2n}) \leq \phi_m(t,0,x) \leq x_m^* + \delta_m^0(\tau_{2n+1})$ for all $t \in \{\tau_{2n},\ldots,\tau_{2n+1}\}$, and for all $n$. Thus,

$$\begin{array}{rclcll}
x_m^* - \dfrac{x_m^*}{M_m+\tau_{2n}} & \leq & \phi_m(t,0,x) & \leq & x_m^* + \dfrac{1-x_m^*}{M_m+\tau_{2n-1}} & \forall\, t \in \{\tau_{2n-1},\ldots,\tau_{2n}\}, \\
x_m^* - \dfrac{x_m^*}{M_m+\tau_{2n}} & \leq & \phi_m(t,0,x) & \leq & x_m^* + \dfrac{1-x_m^*}{M_m+\tau_{2n+1}} & \forall\, t \in \{\tau_{2n},\ldots,\tau_{2n+1}\},
\end{array}$$

for all $n$. Since $T(t,x) \geq 1$ for all $t$ and all $x$, it follows that $\Delta T_n \geq 1$ for all $n$ and $\tau_n \to \infty$ as $n \to \infty$. Therefore, $\phi_m(t,0,x) \to x_m^*$ as $t \to \infty$. A similar argument applies if $x \in P_1(0)$.

□

**Proof of Lemma 3.4:** Since $\phi(t,0,x) \to x^*$ as $t \to \infty$, it follows that $\phi(t,0,x)$ is in a diagonal cell for all $t$. Let $\{\Delta T_n\}_{n=1}^{\infty}$ and $\{\tau_n\}_{n=1}^{\infty}$ be defined as in the proof of Proposition 3.7. Suppose that $x \in P_0$. Then

$$\phi_m(\tau_{2n-1},0,x) = \frac{M_m x_m + \sum_{k=1}^{n} \Delta T_{2k-1}}{M_m + \tau_{2n-1}}.$$

Since $\tau_n \to \infty$ as $n \to \infty$, it follows that

$$\begin{aligned}
x_{-1}^* &= \lim_{n\to\infty} \phi_{-1}(\tau_{2n-1},0,x) = \lim_{n\to\infty} \frac{M_{-1}x_{-1} + \sum_{k=1}^{n} \Delta T_{2k-1}}{M_{-1}+\tau_{2n-1}} \\
&= \lim_{n\to\infty} \frac{M_1 x_1 + \sum_{k=1}^{n} \Delta T_{2k-1}}{M_1+\tau_{2n-1}} = \lim_{n\to\infty} \phi_1(\tau_{2n-1},0,x) = x_1^*.
\end{aligned}$$

A similar argument applies if $x \in P_1$. □

*A.1.4.5 Characterization of Convergence for $x^*_{-1} \neq x^*_1$*

Let

$$\begin{aligned} D^{=}(t) \quad := \quad & \left\{x \in P_0(t) \,:\, x_1 - x_1^* = \frac{\delta_1^0(t)}{\delta_{-1}^0(t)}(x_{-1} - x_{-1}^*)\right\} \cup \\ & \left\{x \in P_1(t) \,:\, x_1 - x_1^* = \frac{\delta_1^1(t)}{\delta_{-1}^1(t)}(x_{-1} - x_{-1}^*)\right\} \end{aligned}$$

denote the *diagonal line*, that is, the line that connects $x^*$ with the vertices of the diagonal cells, at time $t$. Let $\rho(t,x) := (\rho_{-1}(t,x), \rho_1(t,x))$ be given by $\rho_m(t,x) := (x_m - x_m^*)(M_m + t)$ for $m = \pm 1$. Then the diagonal line at time $t$ is given by

$$\begin{aligned} D^{=}(t) \quad = \quad & \left\{x \in P_0(t) \,:\, \rho_{-1}(t,x)(1 - x_1^*) = \rho_1(t,x)(1 - x_{-1}^*)\right\} \cup \\ & \left\{x \in P_1(t) \,:\, \rho_{-1}(t,x) x_1^* = \rho_1(t,x) x_{-1}^*\right\}. \end{aligned}$$

Also, let $D^{\geq}(t)$ and $D^{\leq}(t)$ denote the sets of points above/on and below/on the diagonal line respectively, i.e.,

$$\begin{aligned} D^{\geq}(t) \quad := \quad & \left\{x \in P_0(t) \,:\, \rho_1(t,x)(1 - x_{-1}^*) \geq \rho_{-1}(t,x)(1 - x_1^*)\right\} \cup \\ & \left\{x \in P_1(t) \,:\, \rho_1(t,x) x_{-1}^* \geq \rho_{-1}(t,x) x_1^*\right\} \cup P_2, \\ D^{\leq}(t) \quad := \quad & \left\{x \in P_0(t) \,:\, \rho_1(t,x)(1 - x_{-1}^*) \leq \rho_{-1}(t,x)(1 - x_1^*)\right\} \cup \\ & \left\{x \in P_1(t) \,:\, \rho_1(t,x) x_{-1}^* \leq \rho_{-1}(t,x) x_1^*\right\} \;\cup\; P_3. \end{aligned}$$

Let $D^{>}$ and $D^{<}$ denote the sets of points above and below the diagonal line respectively, i.e.,

$$\begin{aligned} D^{>}(t) \quad := \quad & \left\{x \in P_0(t) \,:\, \rho_1(t,x)(1 - x_{-1}^*) > \rho_{-1}(t,x)(1 - x_1^*)\right\} \cup \\ & \left\{x \in P_1(t) \,:\, \rho_1(t,x) x_{-1}^* > \rho_{-1}(t,x) x_1^*\right\} \;\cup\; P_2, \\ D^{<}(t) \quad := \quad & \left\{x \in P_0(t) \,:\, \rho_1(t,x)(1 - x_{-1}^*) < \rho_{-1}(t,x)(1 - x_1^*)\right\} \cup \\ & \left\{x \in P_1(t) \,:\, \rho_1(t,x) x_{-1}^* < \rho_{-1}(t,x) x_1^*\right\} \;\cup\; P_3. \end{aligned}$$

**Proposition A.1.** *Consider any $t \in \mathbb{N}_0$. Then, the following holds:*

*(1) If $x_1^* \geq x_{-1}^*$, then $\phi(t+1, t, D^{\leq}(t)) \subset D^{\leq}(t+1)$ and $\phi(t+1, t, D^{<}(t)) \subset D^{<}(t+1)$.*

*(2) If $x_1^* \le x_{-1}^*$, then $\phi(t+1, t, D^{\geq}(t)) \subset D^{\geq}(t+1)$ and $\phi(t+1, t, D^{>}(t)) \subset D^{>}(t+1)$.*

**Proof**: Suppose that $x_1^* \ge x_{-1}^*$. Consider any $x \in D^{\leq}(t) \subset \cup_{D \in \mathcal{D}_=(t) \cup \mathcal{D}_3(t)} D \cup P_3$. If $x \in P_3$, then it follows from Lemma 3.1 that $\phi(t+1, t, x) \in P_3$. If $x \in \cup_{D \in \mathcal{D}_3(t)} D$, then it follows from Theorem 3.3 that $\phi(t+1, t, x) \in \cup_{D \in \mathcal{D}_3(t+1)} D \cup P_3 \subset D^{\leq}(t+1)$.

Next, suppose that $x \in \cup_{D \in \mathcal{D}_=(t)} D \cap D^{\leq}(t)$. First consider the case in which $x \in P_0(t)$, and thus $x \in D_{0,i,i}(t)$ for $i \ge 1$. It follows from (3.10) that

$$\begin{aligned} \phi_m(t+1, t, x) &= \frac{(M_m + t)x_m + 1}{M_m + t + 1} \\ &= \frac{(M_m + t)x_m^* + \rho_m(t, x) + 1}{M_m + t + 1} \\ &= x_m^* + \frac{\rho_m(t, x) + 1 - x_m^*}{M_m + t + 1}, \end{aligned}$$

and thus $\rho_m(t+1, \phi(t+1, t, x)) = \rho_m(t, x) + 1 - x_m^*$. Since $x \in D^{\leq}(t) \cap P_0(t)$, it follows that

$$\begin{aligned} \rho_{-1}(t, x)(1 - x_1^*) &\ge \rho_1(t, x)(1 - x_{-1}^*) \\ \Leftrightarrow \quad \rho_{-1}(t, x)(1 - x_1^*) + (1 - x_{-1}^*)(1 - x_1^*) &\ge \rho_1(t, x)(1 - x_{-1}^*) + (1 - x_{-1}^*)(1 - x_1^*) \\ \Leftrightarrow \quad (\rho_{-1}(t, x) + 1 - x_{-1}^*)(1 - x_1^*) &\ge (\rho_1(t, x) + 1 - x_1^*)(1 - x_{-1}^*) \\ \Leftrightarrow \quad \rho_{-1}(t+1, \phi(t+1, t, x))(1 - x_1^*) &\ge \rho_1(t+1, \phi(t+1, t, x))(1 - x_{-1}^*). \end{aligned} \quad \text{(A.5)}$$

Recall from Proposition 3.5 that $\phi(t+1, t, x) \in D_{0,i-1,i-1}(t+1)$. Thus, if $i > 1$, then $\phi(t+1, t, x) \in \left\{x \in P_0(t+1) \,:\, \rho_{-1}(t+1, x)(1 - x_1^*) \ge \rho_1(t+1, x)(1 - x_{-1}^*)\right\} \subset D^{\leq}(t+1)$. Next, suppose $i = 1$. Then $\phi(t+1, t, x) \in D_{0,0,0}(t+1) \subset P_1 \subset P_1(t+1)$. Note that $\rho_m(t+1, \phi(t+1, t, x)) > 0$. Also, note that since $x_1^* \ge x_{-1}^*$ and $x_m^* \in (0, 1)$, it follows that $x_1^*/(1 - x_1^*) \ge x_{-1}^*/(1 - x_{-1}^*) > 0$. Thus it follows from (A.5) that $\rho_{-1}(t+1, \phi(t+1, t, x))x_1^* \ge \rho_1(t+1, \phi(t+1, t, x))x_{-1}^*$. Therefore, $\phi(t+1, t, x) \in \left\{x \in P_1(t+1) \,:\, \rho_{-1}(t+1, x)x_1^* \ge \rho_1(t+1, x)x_{-1}^*\right\} \subset D^{\leq}(t+1)$.

Next consider the case in which $x \in P_1(t)$, and thus $x \in D_{1,i,i}(t)$ for $i \ge 1$. It follows from (3.10) that

$$\phi_m(t+1, t, x) = \frac{(M_m + t)x_m}{M_m + t + 1} = \frac{(M_m + t)x_m^* + \rho_m(t, x)}{M_m + t + 1} = x_m^* + \frac{\rho_m(t, x) - x_m^*}{M_m + t + 1},$$

and thus $\rho_m(t+1, \phi(t+1,t,x)) = \rho_m(t,x) - x_m^*$. Since $x \in D^{\leq}(t) \cap P_1(t)$, it follows that

$$\begin{aligned} & \rho_{-1}(t,x)x_1^* && \geq \rho_1(t,x)x_{-1}^* \\ \Leftrightarrow \quad & \rho_{-1}(t,x)x_1^* - x_{-1}^*x_1^* && \geq \rho_1(t,x)x_{-1}^* - x_{-1}^*x_1^* \\ \Leftrightarrow \quad & (\rho_{-1}(t,x) - x_{-1}^*)x_1^* && \geq (\rho_1(t,x) - x_1^*)x_{-1}^* \\ \Leftrightarrow \quad & \rho_{-1}(t+1, \phi(t+1,t,x))x_1^* && \geq \rho_1(t+1, \phi(t+1,t,x))x_{-1}^*. \end{aligned} \tag{A.6}$$

If $i > 1$, then $\phi(t+1,t,x) \in \{x \in P_1(t+1) : \rho_{-1}(t+1,x)x_1^* \geq \rho_1(t+1,x)x_{-1}^*\} \subset D^{\leq}(t+1)$. If $i = 1$, then $\phi(t+1,t,x) \in D_{1,0,0}(t+1) \subset P_0 \subset P_0(t+1)$. Note that $\rho_m(t+1, \phi(t+1,t,x)) < 0$ and that $0 < (1-x_1^*)/x_1^* \leq (1-x_{-1}^*)/x_{-1}^*$. Thus it follows from (A.6) that $\rho_{-1}(t+1, \phi(t+1,t,x))(1-x_1^*) \geq \rho_1(t+1, \phi(t+1,t,x))(1-x_{-1}^*)$. Therefore $\phi(t+1,t,x) \in \{x \in P_0(t+1) : \rho_{-1}(t+1,x)(1-x_1^*) \geq \rho_1(t+1,x)(1-x_{-1}^*)\} \subset D^{\leq}(t+1)$.

By changing the inequalities in (A.5) and (A.6) to strict inequalities, it follows that $\phi(t+1,t,D^{<}(t)) \subset D^{<}(t+1)$. This completes the proof for (1). Result (2) follows by a similar argument. □

**Corollary A.1.** *Consider any $t \in \mathbb{N}_0$. Then, the following holds:*

*(1) If $x_1^* > x_{-1}^*$, then for any $x \in D^{\leq}(t)$, it holds that $\phi(t+\tau,t,x) \to (1,0)$ as $\tau \to \infty$.*

*(2) If $x_1^* < x_{-1}^*$, then for any $x \in D^{\geq}(t)$, it holds that $\phi(t+\tau,t,x) \to (0,1)$ as $\tau \to \infty$.*

**Proof**: Suppose that $x_1^* > x_{-1}^*$. It follows from Proposition A.1 that $\phi(t+\tau,t,x) \in D^{\leq}(t+\tau)$ for all $\tau \in \mathbb{N}_0$, thus $\phi(t+\tau,t,x) \notin P_2$ for all $\tau$. Thus $\phi(t+\tau,t,x)$ does not converge to $(0,1)$ as $\tau \to \infty$. Also, it follows from Lemma 3.4 that $\phi(t+\tau,t,x)$ does not converge to $x^*$ as $\tau \to \infty$. Hence it follows from Proposition 3.7 that $\phi(t+\tau,t,x) \to (1,0)$ as $\tau \to \infty$. This completes the proof for (1). Result (2) follows by a similar argument. □

Note that, since $x^* \in D^{\leq}(t)$ and $x^* \in D^{\geq}(t)$, it follows from Corollary A.1 that if $x_1^* > x_{-1}^*$, then for all $t$ it holds that $\phi(t+\tau,t,x^*) \to (1,0)$ as $\tau \to \infty$, and if $x_1^* < x_{-1}^*$, then for all $t$ it holds that $\phi(t+\tau,t,x^*) \to (0,1)$ as $\tau \to \infty$.

For any $t \in \mathbb{N}_0$ and $x \in P(t)$, define the *cutting time*

$$\chi(t,x) \quad := \quad \inf\{\tau \in \mathbb{N}_0 : \phi(t+\tau,t,x) \in P_2 \cup P_3\}$$

as the first time (after time $t$) that the image $\phi(t+\tau,t,x)$ of $x$ is in $P_2 \cup P_3$, with the understanding that $\chi(t,x) = \infty$ if $\phi(t+\tau,t,x)$ is in $P_0(t+\tau) \cup P_1(t+\tau)$ for all $\tau \geq 0$. We also say that a point $x(t) \in P_0(t) \cup P_1(t)$ is *cut off* at time $\chi(t,x(t))$. Suppose that $\phi(t+\chi(t,y),t,y) \in P_2$. Then there is a point $x$ such that $\phi(t+\chi(t,y),t,x) = x^*$ and $y_{-1} \leq x_{-1}, y_1 > x_1$. We will write that $y$ will be cut to $P_2$ by $x$.

For any $x \in \phi^{-1}(t,P(t))$, let

$$\zeta(t,x) \quad := \quad \inf\{\tau \in \mathbb{N}_0 \ : \ \phi(t+\tau,t,x) = x^*\}$$

denote the *hitting time* of $x$, i.e., the first time when $\phi(t+\tau,t,x)$ hits $x^*$. Note that for any $x \in \phi^{-1}(t,P(t))$, it holds that $\zeta(t,x) < \infty$ and $\zeta(t,x) < \chi(t,x)$.

Recall that for any $x \in \hat{P}$, $\Omega^2(x) := \{y \in \hat{P} \ : \ y_{-1} \leq x_{-1}, y_1 > x_1\}$ and $\Omega^3(x) := \{y \in \hat{P} \ : \ y_{-1} > x_{-1}, y_1 \leq x_1\}$. Also, let $\Omega(x) := \Omega^2(x) \cup \Omega^3(x)$ denote the *cut set* of $x$, let $\Omega^0(x) := \{y \in \hat{P} \ : \ y_{-1} \leq x_{-1}, y_1 \leq x_1\}$, and let $\Omega^1(x) := \{y \in \hat{P} \ : \ y_{-1} > x_{-1}, y_1 > x_1\}$.

**Lemma A.7.** *Consider any $t \in \mathbb{N}_0$ and $x, y \in P_k(t)$, where $k \in \{0,1\}$, such that $y \in \Omega(x)$. Let $S(t) := \{x, y, (x_{-1}, y_1), (y_{-1}, x_1)\}$, $S(t+\tau) := \phi(t+\tau,t,S(t))$, and $D(t+\tau) := \mathrm{conv}(S(t+\tau))$ for $\tau \in \mathbb{N}_0$. Then, rectangle $D(t)$ walks to rectangle $D(t+\tau)$ from time $t$ to time $t+\tau$ for all $0 \leq \tau \leq \min\{\chi(t,x), \chi(t,y)\}$.*

**Proof**: Note that $x, y \in P_k(t)$ implies that $D(t) \subset P_k(t)$. Recall from Lemma 3.2 that rectangle $D(t) \subset P_k(t)$ walks to a rectangle $\phi(t+1,t,D(t))$ from time $t$ to time $t+1$. In general, if $\phi(t+\tau,t,x), \phi(t+\tau,t,y) \in P_{k'}(t+\tau)$ for $k' \in \{0,1\}$, then $D(t+\tau) \subset P_{k'}(t+\tau)$, and rectangle $D(t+\tau)$ walks to a rectangle $\phi(t+\tau+1,t+\tau,D(t+\tau))$ from time $t+\tau$ to time $t+\tau+1$. Let $T := \inf\{\tau \in \mathbb{N}_0 \ : \ \phi(t+\tau,t,x) \in P_k(t+\tau), \phi(t+\tau,t,y) \in P_{k'}(t+\tau) \text{ for } k \neq k'\}$. If $\phi(t+T,t,x) \in P_2 \cup P_3$, then the result holds. Otherwise, $\phi(t+T,t,x) \in P_0(t+T) \cup P_1(t+T)$, and then $\phi(t+T,t,y) \in P_2 \cup P_3$ (since $y \in \Omega(x)$), and the result holds. □

Lemma A.8 follows from Corollary A.1.

**Lemma A.8.**

*(1) If $x_1^* > x_{-1}^*$, then for any $k \in \{0,1\}$, $t \in \mathbb{N}_0$, and $x \in \phi^{-1}(t, D_{k,1,1}(t))$, it holds that $\phi(t+\tau,t,x) \to (1,0)$ as $\tau \to \infty$.*

*(2) If $x^*_1 < x^*_{-1}$, then for any $k \in \{0,1\}$, $t \in \mathbb{N}_0$, and $x \in \phi^{-1}(t, D_{k,1,1}(t))$, it holds that $\phi(t+\tau, t, x) \to (0,1)$ as $\tau \to \infty$.*

**Lemma A.9.** *For any $k \in \{0,1\}$, $t \in \mathbb{N}_0$ and $D \subset P_k(t)$, the points in $\phi^{-1}(t, D)$ are nondecreasing, that is, for any two pre-images $x^1, x^2 \in \phi^{-1}(t, D)$, either $x^1 \leq x^2$ or $x^2 \leq x^1$.*

**Proof**: Without loss of generality, suppose that $x^1_{-1} < x^2_{-1}$. We show by contradiction that $x^1_1 \leq x^2_1$. Suppose that $x^1_1 > x^2_1$. Then, $x^1 \in \Omega^2(x^2)$ and $x^2 \in \Omega^3(x^1)$.

If $\zeta(t, x^1) \leq \zeta(t, x^2)$, then it holds that $\zeta(t, x^1) \leq \zeta(t, x^2) < \chi(t, x^2)$ and $\zeta(t, x^1) < \chi(t, x^1)$. Thus, $\zeta(t, x^1) < \min\left\{\chi(t, x^1), \chi(t, x^2)\right\}$. It follows from Lemma A.7 that $\phi(t + \zeta(t, x^1), t, x^2) \in \Omega^3(\phi(t+\zeta(t, x^1), t, x^1)) = \Omega^3(x^*) = P_3$. Thus, $\zeta(t, x^2) < \chi(t, x^2) \leq \zeta(t, x^1)$, contradicting $\zeta(t, x^1) \leq \zeta(t, x^2)$.

If $\zeta(t, x^2) \leq \zeta(t, x^1)$, then it holds that $\zeta(t, x^2) \leq \zeta(t, x^1) < \chi(t, x^1)$ and $\zeta(t, x^2) < \chi(t, x^2)$. Thus, $\zeta(t, x^2) < \min\left\{\chi(t, x^1), \chi(t, x^2)\right\}$. It follows from Lemma A.7 that $\phi(t + \zeta(t, x^2), t, x^1) \in \Omega^2(\phi(t+\zeta(t, x^2), t, x^2)) = \Omega^2(x^*) = P_2$. Thus, $\zeta(t, x^1) < \chi(t, x^1) \leq \zeta(t, x^2)$, contradicting $\zeta(t, x^2) \leq \zeta(t, x^1)$. □

For any $k \in \{0,1\}$, $t, T \in \mathbb{N}_0$, $D \subset P(t)$, let

$$\phi^{-1}_{\leq}(t, D, T) \quad := \quad \left\{x \in \phi^{-1}(t, D) \ : \ \zeta(t, x) \leq T\right\}$$

denote the set of pre-images in $D$ at time $t$ with hitting time no later than $T$.

**Lemma A.10.** *Consider any $k \in \{0,1\}$, $t, T \in \mathbb{N}_0$, $D \subset P(t)$. It holds that $\phi^{-1}_{\leq}(t, D, T)$ is a finite set with cardinality $|\phi^{-1}_{\leq}(t, D, T)| \leq 2^{T+1} - 1$.*

**Proof**: If $T = 0$, the result holds since $x^*$ is the only point $x \in P(t)$ with $\zeta(t, x) = 0$. Next suppose that $T > 0$. If $x \in P_2 \cup P_3$, then it follows from Lemma 3.1 that $\phi(t+1, t, x) \in P_2 \cup P_3$, and thus $\phi(t+\tau, t, x) \neq x^*$ for all $\tau \in \mathbb{N}_0$. If $x \in P_0(t) \cup P_1(t)$, then it follows from $\phi(t+1, t, \cdot) : P_k(t) \mapsto \phi(t+1, t, P_k(t)) \subset \hat{P}$ being a separable affine mapping for $k \in \{0,1\}$, that for any $y \in \hat{P}$ there exists at most one point $x \in P_k(t)$ such that $\phi(t+1, t, x) = y$ for $k \in \{0,1\}$. Thus, for any set $Y \subset \hat{P}_0 \cup \hat{P}_1$, it holds that $|\{x \in P_0(t) \cup P_1(t) \ : \ \phi(t+1, t, x) \in Y\}| \ \leq \ 2|Y|$. Next we show by induction on $\tau$ that

$|\{x \in P_0(t) \cup P_1(t) \,:\, \phi(t+\tau, t, x) \in Y\}| \leq 2^\tau |Y|$. Suppose that it holds for some $\tau \in \mathbb{N}_0$. Then

$$\begin{aligned} & |\{x \in P_0(t) \cup P_1(t) \,:\, \phi(t+\tau+1, t, x) \in Y\}| \\ = \; & \left|\{x \in P_0(t) \cup P_1(t) \,:\, \phi(t+\tau, t, x) \in P_{0,1}(t+\tau)\right| \\ \leq \; & 2^\tau \,|\{y \in P_0(t+\tau) \cup P_1(t+\tau) \,:\, \phi(t+\tau+1, t+\tau, y) \in Y\}| \\ \leq \; & 2^\tau 2|Y| \;=\; 2^{\tau+1}|Y|, \end{aligned}$$

where

$$P_{0,1}(t+\tau) \;:=\; \{y \in P_0(t+\tau) \cup P_1(t+\tau) \,:\, \phi(t+\tau+1, t+\tau, y) \in Y\}\}.$$

Therefore

$$\begin{aligned} \left|\phi_{\leq}^{-1}(t, D, T)\right| &= \sum_{\tau=0}^{T} \left|\{x \in \phi^{-1}(t, D) \,:\, \zeta(t, x) = \tau\}\right| \\ &\leq \sum_{\tau=0}^{T} |\{x \in P_0(t) \cup P_1(t) \,:\, \phi(t+\tau, t, x) = x^*\}| \\ &\leq \sum_{\tau=0}^{T} 2^\tau \;=\; 2^{T+1} - 1. \end{aligned}$$

This completes the proof. □

Let $\text{cv}(D_{k,i,j}(t))$ denote the top right corner vertex of cell $D_{k,i,j}(t)$, given by

$$\text{cv}(D_{k,i,j}(t)) \;:=\; \begin{cases} (x^*_{-1} - (j-1)\delta^0_{-1}(t),\, x^*_1 - (i-1)\delta^0_1(t)) & \text{if} \quad k = 0,\ (i,j) \in \hat{I}_0(t) \times \hat{J}_0(t), \\ (x^*_{-1} + j\delta^1_{-1}(t),\, x^*_1 + i\delta^1_1(t)) & \text{if} \quad k = 1,\ (i,j) \in \hat{I}_1(t) \times \hat{J}_1(t). \end{cases}$$

For example, in Figure 7b, $D_{0,2,2}(t+1)$ is indicated by the gray diagonal rectangle with the top right corner vertex $\text{cv}(D_{0,2,2}(t+1)) = \left(x^*_{-1} - \delta^0_{-1}(t+1), x^*_1 - \delta^0_1(t+1)\right)$, and $D_{1,0,0}(t+1)$ is represented by the gray rectangle with the top right corner $\text{cv}(D_{1,0,0}(t+1)) = x^*$.

**Lemma A.11.** *Suppose that $x^*_1 \neq x^*_{-1}$. Consider any $k \in \{0,1\}$, $t \in \mathbb{N}_0$, and $x \in D_{k,1,1}(t) \setminus (D^2_{k,1,1}(t) \cup D^3_{k,1,1}(t))$. Then, the following hold:*

*(1) There exist $\tilde{x}^1 \in \phi^{-1}(t, \bar{D}_{k,1,1}(t))$ and $\tilde{x}^2 \in \phi^{-1}(t, D_{k,1,1}(t))$, such that $\tilde{x}^1 < \tilde{x}^2$, $x \in E := (\tilde{x}^1_{-1}, \tilde{x}^2_{-1}] \times (\tilde{x}^1_1, \tilde{x}^2_1] \subset D_{k,1,1}(t)$, $\max\{\zeta(t, \tilde{x}^1), \zeta(t, \tilde{x}^2)\} \leq \chi(t, x)$, and $\phi_{\leq}^{-1}(t, \bar{D}_{k,1,1}(t), \chi(t,x)) \cap (\bar{E} \setminus \{\tilde{x}^1, \tilde{x}^2\}) = \varnothing$.*

*(2) $\bar{E}$ walks to $\phi(t+\tau,t,\bar{E})$ from time $t$ to time $t+\tau$, and $E$ walks to $\phi(t+\tau,t,E)$ from time $t$ to time $t+\tau$, for all $0 \le \tau \le \min\{\zeta(t,\tilde{x}^1),\zeta(t,\tilde{x}^2)\}$.*

**Proof**:

(1) Note that $\mathrm{cv}(D_{k,1,1}(t)) \in \phi_{\le}^{-1}(t,\bar{D}_{k,1,1}(t),\chi(t,x))$, and thus $\phi_{\le}^{-1}(t,\bar{D}_{k,1,1}(t),\chi(t,x)) \neq \varnothing$. It follows from Lemmas A.9 and A.10 that $\phi_{\le}^{-1}(t,\bar{D}_{k,1,1}(t),\chi(t,x)) \subset \phi^{-1}(t,\bar{D}_{k,1,1}(t))$ is a finite nondecreasing set. Thus we can represent

$$\phi_{\le}^{-1}(t,\bar{D}_{k,1,1}(t),\chi(t,x)) \quad = \quad \{x^n\}_{n=0}^{N} \quad \subset \quad \bar{D}_{k,1,1}(t)$$

such that $x^n \neq x^{n+1}$ for $n \in \{0,1,\ldots,N-1\}$, and $x^0 \le x^1 \le \cdots \le x^N = \mathrm{cv}(D_{k,1,1}(t))$, where if $k=0$, then $x^0 = \mathrm{cv}(D_{0,2,2}(t))$, and if $k=1$, then $x^0 = x^*$. Note that $\zeta(t,x^0) \le 1 \le \chi(t,x)$.

Consider the case with $x_1^* > x_{-1}^*$. Next we show by contradiction that $x_{-1}^0 < x_{-1}^1 < \cdots < x_{-1}^N$. Suppose that $x_{-1}^n = x_{-1}^{n+1}$ for some $n \in \{0,1,\ldots,N-1\}$. Then $x_1^n < x_1^{n+1}$. First we show that $\chi(t,x^n) \le \chi(t,x^{n+1})$. Since $\phi(t+\tau,t,x^{n+1}) \to (1,0)$ as $\tau \to \infty$ from Lemma A.8, it follows that $\phi(t+\chi(t,x^{n+1}),t,x^{n+1}) \in P_3$. If $\chi(t,x^n) \ge \chi(t,x^{n+1})$, then it follows from Lemma A.7 that $\phi_{-1}(t+\chi(t,x^{n+1}),t,x^n) = \phi_{-1}(t+\chi(t,x^{n+1}),t,x^{n+1})$ and $\phi_1(t+\chi(t,x^{n+1}),t,x^n) < \phi_1(t+\chi(t,x^{n+1}),t,x^{n+1})$. Thus $\phi(t+\chi(t,x^{n+1}),t,x^n) \in P_3$, and hence $\chi(t,x^n) \le \chi(t,x^{n+1})$. Then, since $\zeta(t,x^n) < \chi(t,x^n) \le \chi(t,x^{n+1})$, it follows from $x^{n+1} \in \Omega^2(x^n)$ that $\phi(t+\zeta(t,x^n),t,x^{n+1}) \in P_2$ and thus $\chi(t,x^{n+1}) \le \zeta(t,x^n)$, giving a contradiction.

Since $x \notin D_{k,1,1}^2(t) \cup D_{k,1,1}^3(t)$, it follows that for each $x^n$ it holds that $x \notin \Omega^2(x_n) \cup \Omega^3(x_n)$, and thus $x \in \Omega^0(x_n) \cup \Omega^1(x_n)$. Note that there exists $n \in \{0,1,\ldots,N-1\}$ such that $x_{-1}^n < x_{-1} \le x_{-1}^{n+1}$. Then $x \in \Omega^1(x^n) \cap \Omega^0(x^{n+1})$, and thus $x_1^n < x_1 \le x_1^{n+1}$. Then choose $\tilde{x}^1 = x^n$ and $\tilde{x}^2 = x^{n+1}$.

Note that $\tilde{x}^1 \in \phi^{-1}(t,\bar{D}_{k,1,1}(t))$, $\tilde{x}^2 \in \phi^{-1}(t,D_{k,1,1}(t))$, $\tilde{x}^1 < \tilde{x}^2$, $x \in E := (\tilde{x}_{-1}^1,\tilde{x}_{-1}^2] \times (\tilde{x}_1^1,\tilde{x}_1^2] \subset D_{k,1,1}(t)$, $\max\{\zeta(t,\tilde{x}^1),\zeta(t,\tilde{x}^2)\} \le \chi(t,x)$, and $\phi_{\le}^{-1}(t,\bar{D}_{k,1,1}(t),\chi(t,x)) \cap (\bar{E} \setminus \{\tilde{x}^1,\tilde{x}^2\}) = \varnothing$. The proof for the case with $x_1^* < x_{-1}^*$ is similar.

(2) If $k = 0$ and $\tilde{x}^1 = \mathrm{cv}(D_{0,2,2}(t))$, then $\min\{\zeta(t,\tilde{x}^1), \zeta(t,\tilde{x}^2)\} \leq 1$, and $\bar{E}$ walks to $\phi(t+\tau, t, \bar{E})$ from time $t$ to time $t+\tau$ for all $0 \leq \tau \leq \min\{\zeta(t,\tilde{x}^1), \zeta(t,\tilde{x}^2)\}$. If $k = 1$ and $\tilde{x}^1 = x^*$, then $\min\{\zeta(t,\tilde{x}^1), \zeta(t,\tilde{x}^2)\} = 0$, and $\bar{E}$ walks to $\phi(t+\tau, t, \bar{E})$ from time $t$ to time $t+\tau$ for all $0 \leq \tau \leq \min\{\zeta(t,\tilde{x}^1), \zeta(t,\tilde{x}^2)\}$. Otherwise, $\tilde{x}^1, \tilde{x}^2 \in D_{k,1,1}(t)$ and $\bar{E} \subset D_{k,1,1}(t) \subset P_k(t)$. Recall from Lemma 3.2 that rectangle $\bar{E} \subset P_k(t) \subset \hat{P}_k$ walks to a rectangle $\phi(t+1, t, \bar{E})$ from time $t$ to time $t+1$. In general, if $\phi(t+\tau, t, \tilde{x}^1), \phi(t+\tau, t, \tilde{x}^2) \in P_{k'}(t+\tau)$ for $k' \in \{0, 1\}$, then $\phi(t+\tau, t, \bar{E}) \subset P_{k'}(t+\tau)$, and rectangle $\phi(t+\tau, t, \bar{E})$ walks to a rectangle $\phi(t+\tau+1, t+\tau, \phi(t+\tau, t, \bar{E})) = \phi(t+\tau+1, t, \bar{E})$ from time $t+\tau$ to time $t+\tau+1$. Let $T := \inf\left\{\tau \in \mathbb{N}_0 : \phi(t+\tau, t, \tilde{x}^1) \in P_k(t+\tau), \phi(t+\tau, t, \tilde{x}^2) \in P_{k'}(t+\tau) \text{ for } k \neq k'\right\}$. Then $\bar{E}$ walks to $\phi(t+\tau, t, \bar{E})$ from time $t$ to time $t+\tau$ for all $0 \leq \tau \leq T$. Next we show by contradiction that $T \geq \min\{\zeta(t,\tilde{x}^1), \zeta(t,\tilde{x}^2)\}$. Suppose that $T < \min\{\zeta(t,\tilde{x}^1), \zeta(t,\tilde{x}^2)\}$. Then, since $T < \min\{\zeta(t,\tilde{x}^1), \zeta(t,\tilde{x}^2)\} < \min\{\chi(t,\tilde{x}^1), \chi(t,\tilde{x}^2)\}$, it follows that $\phi(t+T, t, \tilde{x}^1), \phi(t+T, t, \tilde{x}^2) \in P_0(t+T) \cup P_1(t+T)$. Also, since $\tilde{x}^1 < \tilde{x}^2$, and $\bar{E}$ walks to $\phi(t+T, t, \bar{E})$ from time $t$ to time $t+T$, it follows that $\phi(t+T, t, \tilde{x}^1) < \phi(t+T, t, \tilde{x}^2)$. Thus, $\phi(t+T, t, \tilde{x}^1) \in P_0(t+T) \setminus \{x^*\}$ (since $T < \zeta(t,\tilde{x}^1)$), and $\phi(t+T, t, \tilde{x}^2) \in P_1(t+T)$. Thus, $x^* \in \phi(t+T, t, \bar{E}) \setminus \{\phi(t+T, t, \tilde{x}^1), \phi(t+T, t, \tilde{x}^2)\}$. Since $\bar{E}$ walks to $\phi(t+T, t, \bar{E})$ from time $t$ to time $t+T$, it follows that there exists $\hat{x} \in \bar{E} \setminus \{\tilde{x}^1, \tilde{x}^2\}$ such that $\phi(t+T, t, \hat{x}) = x^*$. Since $\bar{E} \subset D_{k,1,1}(t)$ and $T < \min\{\zeta(t,\tilde{x}^1), \zeta(t,\tilde{x}^2)\} \leq \chi(t,x)$, it follows that $\hat{x} \in \phi_{\leq}^{-1}(t, \bar{D}_{k,1,1}(t), \chi(t,x)) \cap (\bar{E} \setminus \{\tilde{x}^1, \tilde{x}^2\})$, contradicting $\phi_{\leq}^{-1}(t, \bar{D}_{k,1,1}(t), \chi(t,x)) \cap (\bar{E} \setminus \{\tilde{x}^1, \tilde{x}^2\}) = \varnothing$.

Thus $T \geq \min\{\zeta(t,\tilde{x}^1), \zeta(t,\tilde{x}^2)\}$, and $\bar{E}$ walks to $\phi(t+\tau, t, \bar{E})$ from time $t$ to time $t+\tau$ for all $0 \leq \tau \leq \min\{\zeta(t,\tilde{x}^1), \zeta(t,\tilde{x}^2)\}$. This also implies that $E$ walks to $\phi(t+\tau, t, E)$ from time $t$ to time $t+\tau$ for all $0 \leq \tau \leq \min\{\zeta(t,\tilde{x}^1), \zeta(t,\tilde{x}^2)\}$. □

**Lemma A.12.** *Suppose that $x_1^* > x_{-1}^*$. Consider any time $t \in \mathbb{N}_0$ and any $x^1, x^2 \in P(t)$ such that*

*(a)* $x^1 < x^2$,

*(b)* $x^* \notin E \setminus \{x^2\}$, *where* $E := (x_{-1}^1, x_{-1}^2] \times (x_1^1, x_1^2]$,

*(c)* $x^1 \in D^{\leq}(t)$, *and*

*(d)* $\phi(t+\tau, t, x^2) \to (1,0)$ *as* $\tau \to \infty$.

*Then* $E \cap P_2 = \varnothing$. *Also, either* $\tilde{E} := E \cap (P_0(t) \cup P_1(t)) = \varnothing$, *or there are* $\tilde{x}^1, \tilde{x}^2 \in P_0(t) \cup P_1(t)$ *such that*

*(i)* $\tilde{x}^1 < \tilde{x}^2$,

*(ii)* $\tilde{E} = (\tilde{x}^1_{-1}, \tilde{x}^2_{-1}] \times (\tilde{x}^1_1, \tilde{x}^2_1]$,

*(iii)* *either* $x^2 \in P_3$, *or* $\tilde{x}^2 = x^2$,

*(iv)* $\tilde{E} \subset P_0(t)$ *or* $\tilde{E} \subset P_1(t)$,

*(v)* $\tilde{x}^1 \in D^{\leq}(t)$,

*(vi)* $\phi(t+\tau, t, \tilde{x}^2) \to (1,0)$ *as* $\tau \to \infty$, *and*

*(vii)* $x^* \notin \tilde{E} \setminus \{\tilde{x}^2\}$.

**Proof**: Note that the property $\tilde{E} = (\tilde{x}^1_{-1}, \tilde{x}^2_{-1}] \times (\tilde{x}^1_1, \tilde{x}^2_1]$ and the property $\tilde{E} \subset P_0(t)$ or $\tilde{E} \subset P_1(t)$ imply the property $x^* \notin \tilde{E} \setminus \{\tilde{x}^2\}$. Consider the following 9 cases regarding the position of $x^*$ relative to $E$:

(1) $x^*_{-1} \leq x^1_{-1}$ and $x^*_1 \geq x^2_1$: Then $E \subset P_3$, and thus $E \cap P_2 = \varnothing$ and $\tilde{E} = \varnothing$.

(2) $x^*_{-1} \leq x^1_{-1}$ and $x^1_1 \leq x^*_1 < x^2_1$: Then $E \cap P_2 = \varnothing$. Also, $\tilde{E} = E \cap P_1(t) = (\tilde{x}^1_{-1}, \tilde{x}^2_{-1}] \times (\tilde{x}^1_1, \tilde{x}^2_1] \subset P_1(t)$, where $\tilde{x}^1 = (x^1_{-1}, x^*_1)$ and $\tilde{x}^2 = x^2$. Note that $\tilde{x}^1 < \tilde{x}^2$ and that $\tilde{x}^1 \in D^{\leq}(t)$. Since $\tilde{x}^2 = x^2$, it follows that $\phi(t+\tau, t, \tilde{x}^2) \to (1,0)$ as $\tau \to \infty$.

(3) $x^*_{-1} \leq x^1_{-1}$ and $x^*_1 < x^1_1$: Then $E \subset P_1(t)$ and $\tilde{E} = E$. All the results hold.

(4) $x^1_{-1} < x^*_{-1} \leq x^2_{-1}$ and $x^*_1 \geq x^2_1$: Then $E \cap P_2 = \varnothing$. Also, $\tilde{E} = E \cap P_0(t) = (\tilde{x}^1_{-1}, \tilde{x}^2_{-1}] \times (\tilde{x}^1_1, \tilde{x}^2_1] \subset P_0(t)$, where $\tilde{x}^1 = x^1$ and $\tilde{x}^2 = (x^*_{-1}, x^2_1)$. Note that $\tilde{x}^1 < \tilde{x}^2$. If $x^*_{-1} < x^2_{-1}$, then $x^2 \in P_3$, and if $x^*_{-1} = x^2_{-1}$, then $\tilde{x}^2 = x^2$. Since $\tilde{x}^1 = x^1$, it follows that $\tilde{x}^1 \in D^{\leq}(t)$. Note that $\rho_{-1}(t, \tilde{x}^2)(1 - x^*_1) = 0 \geq \rho_1(t, \tilde{x}^2)(1 - x^*_{-1})$, and thus $\tilde{x}^2 \in D^{\leq}(t)$. It follows from Corollary A.1 that $\phi(t+\tau, t, \tilde{x}^2) \to (1,0)$ as $\tau \to \infty$.

(5) $x^1_{-1} < x^*_{-1} \le x^2_{-1}$ and $x^1_1 \le x^*_1 < x^2_1$: If $x^1_{-1} < x^*_{-1} \le x^2_{-1}$ and $x^1_1 < x^*_1 < x^2_1$, then $x^* \in E\backslash\{x^2\}$, which cannot happen. If $x^1_{-1} < x^*_{-1} \le x^2_{-1}$ and $x^1_1 = x^*_1$, then $x^1 \in P_0(t)$ and $\rho_{-1}(t, x^1)(1 - x^*_1) < 0 = \rho_1(t, x^1)(1 - x^*_{-1})$, contradicting $x^1 \in D^{\le}(t)$. Hence this case cannot happen.

(6) $x^1_{-1} < x^*_{-1} \le x^2_{-1}$ and $x^*_1 < x^1_1$: Then $x^1 \in P_2$, contradicting $x^1 \in D^{\le}(t)$. Hence this case cannot happen.

(7) $x^2_{-1} < x^*_{-1}$ and $x^2_1 \le x^*_1$: Then $E \subset P_0(t)$ and $\tilde{E} = E$. All the results hold.

(8) $x^2_{-1} < x^*_{-1}$ and $x^1_1 \le x^*_1 < x^2_1$: Then $x^2 \in P_2$ and it follows from Theorem 3.2 that $\phi(t + \tau, t, x^2) \to (0, 1)$ as $\tau \to \infty$, contradicting $\phi(t + \tau, t, x^2) \to (1, 0)$ as $\tau \to \infty$. Hence this case cannot happen.

(9) $x^2_{-1} < x^*_{-1}$ and $x^*_1 < x^1_1$: Then $x^2 \in P_2$, contradicting $\phi(t+\tau, t, x^2) \to (1, 0)$ as $\tau \to \infty$. Hence this case cannot happen. □

**Lemma A.13.** *Suppose that $x^*_1 > x^*_{-1}$. Consider any time $t \in \mathbb{N}_0$ and $E := (x^1_{-1}, x^2_{-1}] \times (x^1_1, x^2_1] \subset P_1(t)$ with $x^1 < x^2$ and $x^1_1 = x^*_1$. Suppose that $\phi(t + \tau, t, x^2) \to (1, 0)$ as $\tau \to \infty$. Then, there are $\tilde{x}^1, \tilde{x}^2 \in P(t + 1)$ such that $\tilde{x}^1 < \tilde{x}^2$ and $E$ walks to $\tilde{E} := \phi(t + 1, t, E) = (\tilde{x}^1_{-1}, \tilde{x}^2_{-1}] \times (\tilde{x}^1_1, \tilde{x}^2_1]$ from time $t$ to time $t + 1$, and the following hold:*

*(i)* $\tilde{x}^1 \in D^{\le}(t + 1)$,

*(ii)* $\phi(t + 1 + \tau, t + 1, \tilde{x}^2) \to (1, 0)$ *as* $\tau \to \infty$.

**Proof**: Since $\phi(t + 1, t, \cdot) : P_1(t) \mapsto P(t + 1)$ is an increasing separable affine mapping, it follows from Remark A.1(4) that $E$ walks to $\tilde{E} := \phi(t + 1, t, E)$, and that $\tilde{E}$ is a rectangle of the form $(\tilde{x}^1_{-1}, \tilde{x}^2_{-1}] \times (\tilde{x}^1_1, \tilde{x}^2_1]$ with $\tilde{x}^1 < \tilde{x}^2$. It follows from (3.10) that

$$\tilde{x}^1 \quad = \quad \left(\frac{(M_{-1} + t)x^1_{-1}}{M_{-1} + t + 1}, \frac{(M_1 + t)x^1_1}{M_1 + t + 1}\right) \quad = \quad \left(x^1_{-1} - \frac{x^1_{-1}}{M_{-1} + t + 1}, x^*_1 - \frac{x^*_1}{M_1 + t + 1}\right).$$

Note that $\tilde{x}^1_1 < x^*_1$. Thus, if $\tilde{x}^1_{-1} > x^*_{-1}$, then $\tilde{x}^1 \in P_3$, and hence $\tilde{x}^1 \in D^{\le}(t + 1)$. If $\tilde{x}^1_{-1} \le x^*_{-1}$, then $\tilde{x}^1 \in P_0(t + 1)$. Then

$$\rho_{-1}(t + 1, \tilde{x}^1) \quad = \quad \left[\frac{(M_{-1} + t)x^1_{-1}}{M_{-1} + t + 1} - x^*_{-1}\right](M_{-1} + t + 1)$$

$$\geq \left[\frac{(M_{-1}+t)x^*_{-1}}{M_{-1}+t+1} - x^*_{-1}\right](M_{-1}+t+1)$$
$$= -x^*_{-1}$$

and $\rho_1(t+1,\tilde{x}^1) = \left[(M_1+t)x^1_1/(M_1+t+1) - x^*_1\right](M_1+t+1) = -x^*_1$. Thus $\rho_{-1}(t+1,\tilde{x}^1)(1-x^*_1) \geq -x^*_{-1}+x^*_{-1}x^*_1 > -x^*_1+x^*_{-1}x^*_1 = \rho_1(t+1,\tilde{x}^1)(1-x^*_{-1})$. Hence, $\tilde{x}^1 \in D^{\leq}(t+1)$. Also, since $x^2 \in E \subset P_1(t)$, it follows that $\tilde{x}^2 = \phi(t+1,t,x^2)$. Since $\phi(t+\tau,t,x^2) \to (1,0)$ as $\tau \to \infty$, it holds that $\phi(t+1+\tau,t+1,\tilde{x}^2) \to (1,0)$ as $\tau \to \infty$. □

**Lemma A.14.** *Consider any time $t \in \mathbb{N}_0$ and any $x^1, x^2 \in P(t)$ such that*

*(a)* $x^1 < x^2$,

*(b)* $x^* \notin \bar{E} \setminus \{x^1, x^2\}$, *where* $E := (x^1_{-1}, x^2_{-1}] \times (x^1_1, x^2_1]$,

*(c)* $x^1 \in P_0(t) \cup P_1(t)$ *and* $x^1 \notin D^{<}(t)$, *and*

*(d)* $x^2 \in D^{\leq}(t)$.

*Then* $E \cap P_2 = \varnothing$. *Also, there are* $\tilde{x}^1, \tilde{x}^2 \in P_0(t) \cup P_1(t)$ *such that*

*(i)* $x^1 = \tilde{x}^1 < \tilde{x}^2$,

*(ii)* $\tilde{E} := E \cap (P_0(t) \cup P_1(t)) = (\tilde{x}^1_{-1}, \tilde{x}^2_{-1}] \times (\tilde{x}^1_1, \tilde{x}^2_1]$,

*(iii)* $x^* = x^1$ *or* $\bar{\tilde{E}} \subset P_k(t)$ *for some* $k \in \{0,1\}$,

*(iv)* $x^2 \in P_3$ *or* $\tilde{x}^2 = x^2$, *and*

*(v)* $\tilde{x}^2 \in D^{\leq}(t)$.

**Proof**: Consider the following 15 cases:

(1) $x^*_{-1} < x^1_{-1}$ and $x^*_1 \geq x^2_1$: Then $E \subset P_3$, contradicting $x^1 \in P_0(t) \cup P_1(t)$. Hence this case cannot happen.

(2) $x^*_{-1} < x^1_{-1}$ and $x^1_1 \leq x^*_1 < x^2_1$: Then $x^1 \in P_3$, contradicting $x^1 \in P_0(t) \cup P_1(t)$. Hence this case cannot happen.

(3) $x^*_{-1} < x^1_{-1}$ and $x^*_1 < x^1_1$: Then $\bar{\tilde{E}} = \bar{E} \subset P_1(t)$ and all the results hold.

(4) $x^*_{-1} = x^1_{-1}$ and $x^*_1 > x^1_1$: Then $x^1 \in P_0(t)$ and $\rho_{-1}(t, x^1)(1 - x^*_1) = 0 > \rho_1(t, x^1)(1 - x^*_{-1})$, which implies that $x^1 \in D^<(t)$, contradicting $x^1 \notin D^<(t)$. Hence this case cannot happen.

(5) $x^*_{-1} = x^1_{-1}$ and $x^*_1 = x^1_1$: Then $x^1 = x^*$ and $\tilde{E} = E \subset P_1(t)$. Thus, $x^1 = \tilde{x}^1$ and $\tilde{x}^2 = x^2 \in D^{\leq}(t)$.

(6) $x^*_{-1} = x^1_{-1}$ and $x^*_1 < x^1_1$: Then $x^1 \in P_2$, contradicting $x^1 \in P_0(t) \cup P_1(t)$. Hence this case cannot happen.

(7) $x^1_{-1} < x^*_{-1} < x^2_{-1}$ and $x^*_1 \geq x^2_1$: Then $E \cap P_2 = \varnothing$. Also, $\tilde{E} = E \cap P_0(t) = (\tilde{x}^1_{-1}, \tilde{x}^2_{-1}] \times (\tilde{x}^1_1, \tilde{x}^2_1] \subset P_0(t)$, where $\tilde{x}^1 = x^1$ and $\tilde{x}^2 = (x^*_{-1}, x^2_1)$. Thus (i) and (ii) hold. Since $\tilde{x}^1, \tilde{x}^2 \in P_0(t)$, it holds that $\bar{\tilde{E}} \subset P_0(t)$, and thus (iii) holds. Since $x^*_{-1} < x^2_{-1}$ and $x^*_1 \geq x^2_1$, it holds that $x^2 \in P_3$, and thus (iv) holds. Also note that $\tilde{x}^2 \in P_0(t)$ and $\rho_{-1}(t, \tilde{x}^2)(1 - x^*_1) = 0 \geq \rho_1(t, \tilde{x}^2)(1 - x^*_{-1})$, and thus (v) holds.

(8) $x^1_{-1} < x^*_{-1} < x^2_{-1}$ and $x^1_1 \leq x^*_1 < x^2_1$: Then $x^* \in \bar{E} \setminus \{x^1, x^2\}$, contradicting $x^* \notin \bar{E} \setminus \{x^1, x^2\}$. Hence this case cannot happen.

(9) $x^1_{-1} < x^*_{-1} < x^2_{-1}$ and $x^*_1 < x^1_1$: Then $x^1 \in P_2$, contradicting $x^1 \in P_0(t) \cup P_1(t)$. Hence this case cannot happen.

(10) $x^*_{-1} = x^2_{-1}$ and $x^*_1 \geq x^2_1$: Then $\tilde{E} = E \subset P_0(t)$ and $\bar{\tilde{E}} = \bar{E} \subset P_0(t)$. Thus $\tilde{x}^1 = x^1$ and $\tilde{x}^2 = x^2 \in D^{\leq}(t)$.

(11) $x^*_{-1} = x^2_{-1}$ and $x^1_1 \leq x^*_1 < x^2_1$: Then $x^* \in \bar{E} \setminus \{x^1, x^2\}$, contradicting $x^* \notin \bar{E} \setminus \{x^1, x^2\}$. Hence this case cannot happen.

(12) $x^*_{-1} = x^2_{-1}$ and $x^*_1 < x^1_1$: Then $x^1 \in P_2$, contradicting $x^1 \in P_0(t) \cup P_1(t)$. Hence this case cannot happen.

(13) $x^2_{-1} < x^*_{-1}$ and $x^*_1 \geq x^2_1$: Then $\tilde{E} = E \subset P_0(t)$ and $\bar{\tilde{E}} = \bar{E} \subset P_0(t)$. Thus $\tilde{x}^1 = x^1$ and $\tilde{x}^2 = x^2 \in D^{\leq}(t)$.

(14) $x^2_{-1} < x^*_{-1}$ and $x^1_1 \le x^*_1 < x^2_1$: Then $x^2 \in P_2$, contradicting $x^2 \in D^{\le}(t)$. Hence this case cannot happen.

(15) $x^2_{-1} < x^*_{-1}$ and $x^*_1 < x^1_1$: Then $x^1 \in P_2$, contradicting $x^1 \in P_0(t) \cup P_1(t)$. Hence this case cannot happen. □

**Proof of Theorem 3.4:** Consider any $k \in \{0,1\}$ and $i \in I_k(0) \cap J_k(0)$.

(1) Consider any $x \in D^2_{k,i,i}(0) := D_{k,i,i}(0) \cap (\cup_{y \in \phi^{-1}(0, D_{k,i,i}(0))} \Omega^2(y))$. Thus, there exists $y \in \phi^{-1}(0, D_{k,i,i}(0)) \subset D_{k,i,i}(0)$ such that $x \in \Omega^2(y) \cap D_{k,i,i}(0)$. Note that $\phi(\zeta(0,y), 0, y) = x^*$. If $\zeta(0,y) \le \chi(0,x)$, then it follows from Lemma A.7 that

$$\phi(\zeta(0,y), 0, x) \in \Omega^2(\phi(\zeta(0,y), 0, y)) = \Omega^2(x^*).$$

Thus, $\phi(\zeta(0,y), 0, x) \in P_2$ (and $\zeta(0,y) = \chi(0,x)$), and it follows from Theorem 3.2 that $\phi(t, 0, x) \to (0,1)$ as $t \to \infty$. If $\zeta(0,y) > \chi(0,x)$, then $\chi(0,x) < \zeta(0,y) < \chi(0,y)$. We show by contradiction that $\phi(\chi(0,x), 0, x) \in P_2$. Suppose that $\phi(\chi(0,x), 0, x) \in P_3$. It follows from Lemma A.7 that $\phi(\chi(0,x), 0, x) \in \Omega^2(\phi(\chi(0,x), 0, y))$, and thus $\phi_{-1}(\chi(0,x), 0, y) \ge \phi_{-1}(\chi(0,x), 0, x) > x^*_{-1}$ and $\phi_1(\chi(0,x), 0, y) < \phi_1(\chi(0,x), 0, x) \le x^*_1$. Thus, $\phi_{-1}(\chi(0,x), 0, y) \in P_3$, which implies that $\chi(0,y) \le \chi(0,x)$, contradicting $\chi(0,x) < \chi(0,y)$. Hence it follows from Theorem 3.2 that $\phi(t, 0, x) \to (0,1)$ as $t \to \infty$.

(2) The proof of (2) is similar to the proof of (1).

(3) Consider any $x \in D_{k,i,i}(0) \setminus (D^2_{k,i,i}(0) \cup D^3_{k,i,i}(0))$. Thus, for each $y' \in \phi^{-1}(0, D_{k,i,i}(0))$, it holds that $x \in \Omega^0(y') \cup \Omega^1(y')$. It follows from Proposition 3.6 that $D_{k,i,i}(0)$ walks to $D_{k,1,1}(i-1)$ from time 0 to time $i-1$, and $\tilde{x} := \phi(i-1, 0, x) \in D_{k,1,1}(i-1)$. Next we show by contradiction that $\tilde{x} \in D_{k,1,1}(i-1) \setminus (D^2_{k,1,1}(i-1) \cup D^3_{k,1,1}(i-1))$. Suppose that $\tilde{x} \in D^2_{k,1,1}(i-1) \cup D^3_{k,1,1}(i-1)$. Then there exists $\tilde{y} \in \phi^{-1}(i-1, D_{k,1,1}(i-1)) \subset D_{k,1,1}(i-1)$ such that $\tilde{x} \in \Omega(\tilde{y})$. Since $D_{k,i,i}(0)$ walks to $D_{k,1,1}(i-1)$, there exists $y \in D_{k,i,i}(0)$ such that $\phi(i-1, 0, y) = \tilde{y}$. Then it follows that $y \in \phi^{-1}(0, D_{k,i,i}(0))$ (and $\zeta(0,y) \le i - 1 + \zeta(i-1, \tilde{y})$). Also, it follows from Proposition 3.6 that $x \in \Omega(y)$. This contradicts $x \in \Omega^0(y') \cup \Omega^1(y')$ for all $y' \in \phi^{-1}(0, D_{k,i,i}(0))$.

Next, consider any $t \in \mathbb{N}_0$ and any $x \in D_{k,1,1}(t) \setminus (D^2_{k,1,1}(t) \cup D^3_{k,1,1}(t))$. It follows from Proposition 3.7 and Lemma 3.4 that $\phi(t+\tau, t, x) \to (0,1)$ or $\phi(t+\tau, t, x) \to (1,0)$ as $\tau \to \infty$. Thus $\chi(t,x) < \infty$.

If $x \in \phi^{-1}(t, D_{k,1,1}(t))$, then it follows from Lemma A.8 that $\phi(t+\tau, t, x) \to (1,0)$ as $\tau \to \infty$.

Next consider the case in which $x \notin \phi^{-1}(t, D_{k,1,1}(t))$. It follows from Lemma A.11((1)) that there exist $x^1 \in \phi^{-1}(t, \bar{D}_{k,1,1}(t))$, $x^2 \in \phi^{-1}(t, D_{k,1,1}(t))$, such that $x^1 < x^2$, $x \in E := (x^1_{-1}, x^2_{-1}] \times (x^1_1, x^2_1] \subset D_{k,1,1}(t)$, $\max\{\zeta(t,x^1), \zeta(t,x^2)\} \leq \chi(t,x)$, and $\phi^{-1}_{\leq}(t, \bar{D}_{k,1,1}(t), \chi(t,x)) \cap (\bar{E} \setminus \{x^1, x^2\}) = \varnothing$. Inductively define $E(0) := E$, $\tilde{E}(\tau) := E(\tau) \cap (P_0(t+\tau) \cup P_1(t+\tau))$, and $E(\tau+1) := \phi(t+\tau+1, t+\tau, \tilde{E}(\tau))$ for $\tau = 0, 1, \ldots, \chi(t,x)$. For any $\tau \in \{0, 1, \ldots, \chi(t,x)\}$, let $x^1(\tau)$ and $x^2(\tau)$ denote respectively the left-bottom and right-top vertices of $E(\tau)$, and let $\tilde{x}^1(\tau)$ and $\tilde{x}^2(\tau)$ denote respectively the left-bottom and right-top vertices of $\tilde{E}(\tau)$. Note that $x^1 = x^1(0)$ and $x^2 = x^2(0)$.

Let $\tau_0 := \min\{\zeta(t,x^1), \zeta(t,x^2)\}$. It follows from Lemma A.11((2)) that $E$ walks to $\phi(t+\tau_0, t, E)$ from time $t$ to time $t+\tau_0$. Since $\phi(t+\tau_0, t, x^1) < \phi(t+\tau_0, t, x^2)$, it follows that either $\zeta(t,x^1) < \zeta(t,x^2)$ or $\zeta(t,x^2) < \zeta(t,x^1)$. Hence, consider the following two cases:

1 **Case 1**. $\zeta(t,x^1) < \zeta(t,x^2) \leq \chi(t,x)$:

   Then $\tau_0 = \zeta(t,x^1)$. It follows from Lemma A.11((2)) that $E$ walks to $\phi(t+\tau_0, t, E)$ from time $t$ to time $t+\tau_0$, and for all $0 \leq \tau < \tau_0$, it holds that $\phi(t+\tau, t, E) \subset P_0(t+\tau)$ or $\phi(t+\tau, t, E) \subset P_1(t+\tau)$. Thus, $\tilde{E}(\tau) = E(\tau) \subset P_0(t+\tau)$ or $\tilde{E}(\tau) = E(\tau) \subset P_1(t+\tau)$ for all $0 \leq \tau < \tau_0$.

   Next we show by induction on $\tau \in \{\tau_0, \ldots, \chi(t,x)\}$ that

   (a) there exist $x^1(\tau), x^2(\tau) \in P(t+\tau)$ s.t. $x^1(\tau) < x^2(\tau)$ and

   $$E(\tau) = (x^1_{-1}(\tau), x^2_{-1}(\tau)] \times (x^1_1(\tau), x^2_1(\tau)],$$

   (b) $\phi(t+\tau, t, x) \in E(\tau)$,

(c) $E(\tau) \subset \phi(t+\tau, t, E)$,

(d) $x^* \notin E(\tau) \setminus \{x^2(\tau)\}$,

(e) $x^1(\tau) \in D^{\leq}(t+\tau)$,

(f) $\phi(t+\tau+\tau', t+\tau, x^2(\tau)) \to (1,0)$ as $\tau' \to \infty$,

(g) $E(\tau) \cap P_2 = \varnothing$,

(h) $\tilde{E}(\tau) \subset P_0(t+\tau)$ or $\tilde{E}(\tau) \subset P_1(t+\tau)$, and

(i) $\tilde{E}(\tau) = \varnothing$, or there exist $\tilde{x}^1(\tau), \tilde{x}^2(\tau) \in P_0(t+\tau) \cup P_1(t+\tau)$ such that $\tilde{x}^1(\tau) < \tilde{x}^2(\tau)$ and $\tilde{E}(\tau) = (\tilde{x}^1_{-1}(\tau), \tilde{x}^2_{-1}(\tau)] \times (\tilde{x}^1_1(\tau), \tilde{x}^2_1(\tau)]$, and $\phi(t+\tau, t, x^2) = \tilde{x}^2(\tau)$ or $\phi(t+\tau, t, x^2) \in P_3$.

Consider $\tau = \tau_0$. Since $E$ walks to $E(\tau_0)$, it follows from Remark A.1(3) that $x^1(\tau_0) = \phi(t+\tau_0, t, x^1) < \phi(t+\tau_0, t, x^2) = x^2(\tau_0)$ and $E(\tau_0) = \phi(t+\tau_0, t, E) = (x^1_{-1}(\tau_0), x^2_{-1}(\tau_0)] \times (x^1_1(\tau_0), x^2_1(\tau_0)]$. Since $x \in E$, it follows that $\phi(t+\tau_0, t, x) \in \phi(t+\tau_0, t, E) = E(\tau_0)$. Since $\tau_0 = \zeta(t, x^1)$, it holds that $x^1(\tau_0) = \phi(t+\tau_0, t, x^1) = x^* \in D^{\leq}(t+\tau_0)$, and thus $\tilde{E}(\tau_0) = E(\tau_0) \subset P_1(t+\tau_0)$, $x^* \notin E(\tau_0)$, $E(\tau_0) \cap P_2 = \varnothing$, and $\tilde{x}^1(\tau_0) = x^1(\tau_0) < \tilde{x}^2(\tau_0) = x^2(\tau_0) = \phi(t+\tau_0, t, x^2)$. Since $x^2 \in \phi^{-1}(t, D_{k,1,1}(t))$, it follows from Lemma A.8 that $\phi(t+\tau', t, x^2) \to (1,0)$ as $\tau' \to \infty$. It follows that $\phi(t+\tau_0+\tau', t+\tau_0, x^2(\tau_0)) = \phi(t+\tau_0+\tau', t, x^2) \to (1,0)$ as $\tau' \to \infty$. Thus (a)–(i) hold for $\tau = \tau_0$.

Assume that (a)–(i) hold for some $\tau < \chi(t,x)$. Since $\phi(t+\tau, t, x) \in E(\tau)$ and $\tau < \chi(t,x)$, it holds that $\phi(t+\tau, t, x) \in E(\tau) \cap (P_0(t+\tau) \cup P_1(t+\tau)) = \tilde{E}(\tau)$. Thus, $\tilde{E}(\tau) \neq \varnothing$. It follows from Lemma A.12 that there are $\tilde{x}^1(\tau), \tilde{x}^2(\tau) \in P_0(t+\tau) \cup P_1(t+\tau)$ such that $\tilde{x}^1(\tau) < \tilde{x}^2(\tau)$ and $\tilde{E}(\tau) = (\tilde{x}^1_{-1}(\tau), \tilde{x}^2_{-1}(\tau)] \times (\tilde{x}^1_1(\tau), \tilde{x}^2_1(\tau)]$, $x^2(\tau) \in P_3$ or $\tilde{x}^2(\tau) = x^2(\tau)$, $\tilde{E}(\tau) \subset P_0(t+\tau)$ or $\tilde{E}(\tau) \subset P_1(t+\tau)$, $\tilde{x}^1(\tau) \in D^{\leq}(t+\tau)$, and $\phi(t+\tau+\tau', t+\tau, \tilde{x}^2(\tau)) \to (1,0)$ as $\tau' \to \infty$. Since $\tilde{E}(\tau) \subset P_k(t+\tau)$, where $k \in \{0,1\}$, and $\phi(t+\tau+1, t+\tau, \cdot) : \hat{P}_k \mapsto \hat{P}$ is an increasing separable affine mapping, it follows that $\tilde{E}(\tau)$ walks to $E(\tau+1) := \phi(t+\tau+1, t+\tau, \tilde{E}(\tau))$ from time $t+\tau$ to time $t+\tau+1$.

Next we show that (a)–(i) hold for $\tau+1$.

(a) Since $\tilde{x}^1(\tau) < \tilde{x}^2(\tau)$, it follows from Remark A.1(1) that $E(\tau+1) = (x^1_{-1}(\tau+1), x^2_{-1}(\tau+1)] \times (x^1_1(\tau+1), x^2_1(\tau+1)]$, where $x^1(\tau+1) < x^2(\tau+1)$.

(b) Since $\tilde{E}(\tau)$ walks to $E(\tau+1)$ from time $t+\tau$ to time $t+\tau+1$, and $\phi(t+\tau,t,x) \in \tilde{E}(\tau)$, it follows that $\phi(t+\tau+1,t,x) \in E(\tau+1)$.

(c) It holds that $E(\tau+1) = \phi(t+\tau+1, t+\tau, \tilde{E}(\tau)) \subset \phi(t+\tau+1, t+\tau, E(\tau)) \subset \phi(t+\tau+1, t+\tau, \phi(t+\tau, t, E)) = \phi(t+\tau+1, t, E)$.

(d) We show by contradiction that $x^* \notin E(\tau+1) \setminus \{x^2(\tau+1)\}$. Suppose that $x^* \in E(\tau+1) \setminus \{x^2(\tau+1)\}$. Since $E(\tau+1) \subset \phi(t+\tau+1,t,E)$, there exists $x' \in E$ such that $\phi(t+\tau+1,t,x') = x^*$. By property (i) for time $t+\tau$, $\phi(t+\tau,t,x^2) = \tilde{x}^2(\tau)$ or $\phi(t+\tau,t,x^2) \in P_3$. If $\phi(t+\tau,t,x^2) = \tilde{x}^2(\tau)$, then since $\tilde{E}(\tau)$ walks to $E(\tau+1)$ from time $t+\tau$ to time $t+\tau+1$, it follows from Remark A.1(3) that $x^2(\tau+1) = \phi(t+\tau+1, t+\tau, \tilde{x}^2(\tau)) = \phi(t+\tau+1, t+\tau, \phi(t+\tau, t, x^2)) = \phi(t+\tau+1, t, x^2) \neq x^*$, and thus $x' \neq x^2$. If $\phi(t+\tau,t,x^2) \in P_3$, then it follows from Lemma 3.1 that $\phi(t+\tau+1,t,x^2) \in P_3$, and thus $x' \neq x^2$. Hence, either way $x' \neq x^2$. Thus, $x' \in \bar{E} \setminus \{x^1, x^2\}$. Note that $\tau + 1 \leq \chi(t,x)$. Thus, $x' \in \phi^{-1}_{\leq}(t, \bar{D}_{k,1,1}(t), \chi(t,x)) \cap \left(\bar{E} \setminus \{x^1, x^2\}\right)$, contradicting $\phi^{-1}_{\leq}(t, \bar{D}_{k,1,1}(t), \chi(t,x)) \cap (\bar{E} \setminus \{x_1, x_2\}) = \varnothing$.

(e) Recall that $\tilde{E}(\tau) \subset P_0(t+\tau)$ or $\tilde{E}(\tau) \subset P_1(t+\tau)$. First, note that if $\tilde{x}^1_{-1}(\tau) = x^*_{-1}$ and $\tilde{x}^1_1(\tau) > x^*_1$, then $\tilde{x}^1(\tau) \in P_2$, which contradicts $\tilde{x}^1(\tau) \in D^{\leq}(t+\tau)$. Thus, it cannot hold that $\tilde{x}^1_{-1}(\tau) = x^*_{-1}$ and $\tilde{x}^1_1(\tau) > x^*_1$. If $\tilde{E}(\tau) \subset P_1(t+\tau)$ and $\tilde{x}^1_1(\tau) = x^*_1$, then since $\tilde{E}(\tau)$ walks to $E(\tau+1)$, it follows from Lemma A.13 that $x^1(\tau+1) \in D^{\leq}(t+\tau+1)$. If $\tilde{E}(\tau) \subset P_1(t+\tau)$ and $\tilde{x}^1(\tau) > x^*$, then $\tilde{x}^1(\tau) \in P_1(t+\tau)$. Then it follows from Remark A.1(3) that $x^1(\tau+1) = \phi(t+\tau+1, t+\tau, \tilde{x}^1(\tau))$. Thus, it follows from Proposition A.1(1) that $x^1(\tau+1) \in D^{\leq}(t+\tau+1)$. If $\tilde{E}(\tau) \subset P_0(t+\tau)$, then $\tilde{x}^1(\tau) \in P_0(t+\tau)$, and it follows from Remark A.1(3) that $x^1(\tau+1) = \phi(t+\tau+1, t+\tau, \tilde{x}^1(\tau))$. Thus, it follows from Proposition A.1(1) that $x^1(\tau+1) \in D^{\leq}(t+\tau+1)$.

(f) Since $\phi(t+\tau+\tau', t+\tau, \tilde{x}^2(\tau)) \to (1,0)$ as $\tau' \to \infty$ and $x^2(\tau+1) = \phi(t+\tau+1, t+\tau, \tilde{x}^2(\tau))$, it follows that $\phi(t+\tau+1+\tau', t+\tau+1, x^2(\tau+1)) = \phi(t+\tau+$

$1+\tau', t+\tau+1, \phi(t+\tau+1, t+\tau, \tilde{x}^2(\tau))) = \phi(t+\tau+1+\tau', t+\tau, \tilde{x}^2(\tau)) \to (1,0)$ as $\tau' \to \infty$.

(g) It follows from Lemma A.12 that $E(\tau+1) \cap P_2 = \varnothing$.

(h) It follows from Lemma A.12 that $\tilde{E}(\tau+1) \subset P_0(t+\tau+1)$ or $\tilde{E}(\tau+1) \subset P_1(t+\tau+1)$.

(i) If $\tilde{E}(\tau+1) = \varnothing$, then result (i) holds. Otherwise, since $\tilde{E}(\tau) \neq \varnothing$, it follows from the induction hypothesis that $\phi(t+\tau, t, x^2) = \tilde{x}^2(\tau)$ or $\phi(t+\tau, t, x^2) \in P_3$. If $\phi(t+\tau, t, x^2) \in P_3$, then it follows from Lemma 3.1 that $\phi(t+\tau+1, t, x^2) \in P_3$. If $\phi(t+\tau, t, x^2) = \tilde{x}^2(\tau)$, then since $\tilde{E}(\tau)$ walks to $E(\tau+1)$, it follows that $x^2(\tau+1) = \phi(t+\tau+1, t+\tau, \tilde{x}^2(\tau)) = \phi(t+\tau+1, t+\tau, \phi(t+\tau, t, x^2)) = \phi(t+\tau+1, t, x^2)$. Then, since $\tilde{E}(\tau+1) \neq \varnothing$, it follows from Lemma A.12 that $x^2(\tau+1) = \phi(t+\tau+1, t, x^2) \in P_3$ or $\tilde{x}^2(\tau+1) = x^2(\tau+1) = \phi(t+\tau+1, t, x^2)$.

2 **Case 2**. $\zeta(t, x^2) < \zeta(t, x^1) \leq \chi(t, x)$:

Then $\tau_0 = \zeta(t, x^2)$. It follows from Lemma A.11((2)) that $\bar{E}$ walks to $\phi(t+\tau_0, t, \bar{E})$ from time $t$ to time $t+\tau_0$, and for all $0 \leq \tau < \tau_0$, it holds that $\phi(t+\tau, t, \bar{E}) \subset P_0(t+\tau)$ or $\phi(t+\tau, t, \bar{E}) \subset P_1(t+\tau)$. Thus, $\tilde{\bar{E}}(\tau) = \bar{E}(\tau) \subset P_0(t+\tau)$ or $\tilde{\bar{E}}(\tau) = \bar{E}(\tau) \subset P_1(t+\tau)$ for all $0 \leq \tau < \tau_0$.

We show by induction on $\tau \in \{\tau_0, \ldots, \zeta(t, x^1)\}$ that

(a) there exist $x^1(\tau), x^2(\tau) \in P(t+\tau)$ such that $x^1(\tau) < x^2(\tau)$ and $E(\tau) = (x^1_{-1}(\tau), x^2_{-1}(\tau)] \times (x^1_1(\tau), x^2_1(\tau)]$,

(b) $\phi(t+\tau, t, x) \in E(\tau)$,

(c) $\bar{E}(\tau) \subset \phi(t+\tau, t, \bar{E})$ and $E(\tau) \subset \phi(t+\tau, t, E)$,

(d) $\phi(t+\tau, t, x^1) = x^1(\tau) \in P_0(t+\tau) \cup P_1(t+\tau)$ and $\phi(t+\tau, t, x^1) \notin D^{<}(t+\tau)$.

(e) $x^* \notin \bar{E}(\tau) \setminus \{x^1(\tau), x^2(\tau)\}$,

(f) $x^2(\tau) \in D^{\leq}(t+\tau)$,

(g) $E(\tau) \cap P_2 = \varnothing$,

(h) $x^1(\tau) = x^*$ or $\tilde{\bar{E}}(\tau) \subset P_k(t+\tau)$ for some $k \in \{0, 1\}$, and

(i) $\phi(t+\tau, t, x^2) = x^2(\tau)$ or $\phi(t+\tau, t, x^2) \in P_3$.

Consider $\tau = \tau_0$. Since $\bar{E}$ walks to $\bar{E}(\tau_0)$, it follows from Remark A.1(3) that $x^1(\tau_0) = \phi(t+\tau_0, t, x^1) < \phi(t+\tau_0, t, x^2) = x^2(\tau_0)$, $\bar{E}(\tau_0) = \phi(t+\tau_0, t, \bar{E}) = [x^1_{-1}(\tau_0), x^2_{-1}(\tau_0)] \times [x^1_1(\tau_0), x^2_1(\tau_0)]$, where $x^2(\tau_0) = \phi(t+\tau_0, t, x^2) = x^*$. Thus it holds that $x^* \notin \bar{E}(\tau_0) \setminus \{x^1(\tau_0), x^2(\tau_0)\}$, $\phi(t+\tau_0, t, x^1) = x^1(\tau_0) \in P_0(t+\tau_0)$, $x^2(\tau_0) = x^* \in D^{\leq}(t+\tau_0)$, $\bar{E}(\tau_0) \cap P_2 = \varnothing$, $\bar{E}(\tau_0) \subset P_0(t+\tau_0)$ and $\bar{\tilde{E}}(\tau_0) = \bar{E}(\tau_0) \subset P_0(t+\tau_0)$, and $\phi(t+\tau_0, t, x^2) = x^2(\tau_0)$. Also, since $E$ walks to $E(\tau_0)$, it follows that $E(\tau_0) = \phi(t+\tau_0, t, E)$. Since $x \in E$, it follows that $\phi(t+\tau_0, t, x) \in E(\tau_0)$. It follows from $\tau_0 < \zeta(t, x^1)$ that $\phi(t+\tau_0, t, x^1) \notin D^{<}(t+\tau_0)$; since if $\phi(t+\tau_0, t, x^1) \in D^{<}(t+\tau_0)$ then it would follow from Proposition A.1(1) that $\phi(t+\tau, t, x^1) \in D^{<}(t+\tau)$ and thus $\phi(t+\tau, t, x^1) \neq x^*$ for all $\tau \geq \tau_0$, contradicting $\zeta(t, x^1) > \tau_0$. Thus (a)–(i) hold for $\tau = \tau_0$.

Assume that (a)–(i) hold for some $\tau < \zeta(t, x^1)$. Since $\phi(t+\tau, t, x) \in E(\tau)$ and $\tau < \zeta(t, x^1) \leq \chi(t, x)$, it holds that $\phi(t+\tau, t, x) \in P_0(t+\tau) \cup P_1(t+\tau)$, and thus $\phi(t+\tau, t, x) \in E(\tau) \cap (P_0(t+\tau) \cup P_1(t+\tau)) = \tilde{E}(\tau)$. Since $\tilde{E}(\tau) \subset E(\tau)$, it holds that $\bar{\tilde{E}}(\tau) \subset \bar{E}(\tau)$. It follows from Lemma A.14 that there are $\tilde{x}^1(\tau), \tilde{x}^2(\tau)$ such that $x^1(\tau) = \tilde{x}^1(\tau) < \tilde{x}^2(\tau)$ and $\tilde{E}(\tau) = (\tilde{x}^1_{-1}(\tau), \tilde{x}^2_{-1}(\tau)] \times (\tilde{x}^1_1(\tau), \tilde{x}^2_1(\tau)]$, and $\tilde{x}^2(\tau) \in D^{\leq}(t+\tau)$. Since $\tau < \zeta(t, x^1)$, it also follows that $x^1(\tau) = \phi(t+\tau, t, x^1) \neq x^*$, and thus $\bar{\tilde{E}}(\tau) \in P_k(t+\tau)$ for some $k \in \{0, 1\}$. It follows that $\bar{\tilde{E}}(\tau)$ walks to $\bar{E}(\tau+1)$ and $\tilde{E}(\tau)$ walks to $E(\tau+1)$. It also follows from Lemma A.14 that $x^2(\tau) \in P_3$ or $x^2(\tau) = \tilde{x}^2(\tau)$.

Next we show that (a)–(i) hold for $\tau+1$.

(a) Since $\tilde{E}(\tau) = (\tilde{x}^1_{-1}(\tau), \tilde{x}^2_{-1}(\tau)] \times (\tilde{x}^1_1(\tau), \tilde{x}^2_1(\tau)]$ walks to $E(\tau+1)$, it follows that there exist $x^1(\tau+1), x^2(\tau+1) \in P(t+\tau)$ such that $x^1(\tau+1) < x^2(\tau+1)$ and $E(\tau+1) = (x^1_{-1}(\tau+1), x^2_{-1}(\tau+1)] \times (x^1_1(\tau+1), x^2_1(\tau+1)]$.

(b) Since $\phi(t+\tau, t, x) \in \tilde{E}(\tau)$, it follows that $\phi(t+\tau+1, t, x) \in E(t+\tau+1)$.

(c) It follows that $\bar{E}(\tau+1) = \phi(t+\tau+1, t+\tau, \bar{\tilde{E}}(\tau)) \subset \phi(t+\tau+1, t+\tau, \bar{E}(\tau)) \subset \phi(t+\tau+1, t+\tau, \phi(t+\tau, t, \bar{E})) = \phi(t+\tau+1, t, \bar{E})$, and $E(\tau+1) = \phi(t+\tau+$

$1, t+\tau, \tilde{E}(\tau)) \subset \phi(t+\tau+1, t+\tau, E(\tau)) \subset \phi(t+\tau+1, t+\tau, \phi(t+\tau, t, E)) = \phi(t+\tau+1, t, E)$.

(d) Since $\tilde{\bar{E}}(\tau)$ walks to $\bar{E}(\tau+1)$ and $\tilde{x}^1(\tau) = x^1(\tau) = \phi(t+\tau, t, x^1)$, it follows from Remark A.1(3) that $x^1(\tau+1) = \phi(t+\tau+1, t+\tau, \tilde{x}^1(\tau)) = \phi(t+\tau+1, t+\tau, \phi(t+\tau, t, x^1)) = \phi(t+\tau+1, t, x^1)$. Since $\tau+1 \leq \zeta(t, x^1)$, it holds that $\phi(t+\tau+1, t, x^1) \in P_0(t+\tau) \cup P_1(t+\tau)$. It also holds that $\phi(t+\tau+1, t, x^1) \notin D^<(t+\tau+1)$; since otherwise it follows from Proposition A.1(1) that $\phi(t+\tau', t, x^1) \in D^<(t+\tau')$ for all $\tau' \geq \tau+1$, contradicting $\tau+1 \leq \zeta(t, x^1)$.

(e) We show by contradiction that $x^* \notin \bar{E}(\tau+1) \setminus \{x^1(\tau+1), x^2(\tau+1)\}$. Suppose that $x^* \in \bar{E}(\tau+1) \setminus \{x^1(\tau+1), x^2(\tau+1)\}$. Since $\bar{E}(\tau+1) \subset \phi(t+\tau+1, t, \bar{E})$, there exists $x' \in \bar{E}$ such that $\phi(t+\tau+1, t, x') = x^*$. Note that $\phi(t+\tau+1, t, x^1) = x^1(\tau+1) \neq x^*$, and thus $x' \neq x^1$. Next, note that $\phi(t+\tau_0+1, t, x^2) = \phi(t+\tau_0+1, t+\tau_0, x^*) = (x^*_{-1}+\delta^0_{-1}(t+\tau_0+1), x^*_1+\delta^0_1(t+\tau_0+1)) \in P_1(t+\tau)$, and thus $\rho_m(t+\tau_0+1, \phi(t+\tau_0+1, t, x^2)) = \delta^0_m(t+\tau_0+1)(M_m+t+1) = 1-x^*_m$ for $m = \pm 1$. Since $x^*_1 > x^*_{-1}$, it follows that $\rho_{-1}(t+\tau_0+1, \phi(t+\tau_0+1, t, x^2))x^*_1 = (1-x^*_{-1})x^*_1 = x^*_1 - x^*_{-1}x^*_1 > x^*_{-1} - x^*_{-1}x^*_1 = (1-x^*_1)x^*_{-1} = \rho_1(t+\tau_0+1, \phi(t+\tau_0+1, t, x^2))x^*_{-1}$, and hence $\phi(t+\tau_0+1, t, x^2) \in D^<(t+\tau_0+1)$. Then it follows from Proposition A.1(1) that $\phi(t+\tau', t, x^2) \in D^<(t+\tau')$ for all $\tau' \geq \tau_0+1$. Since $\tau+1 \geq \tau_0+1$, it follows that $\phi(t+\tau+1, t, x^2) \in D^<(t+\tau+1)$, and thus $\phi(t+\tau+1, t, x^2) \neq x^*$. Hence $x' \neq x^2$. Note that $\tau+1 \leq \chi(t, x)$, and thus, $x' \in \phi^{-1}_{\leq}(t, \bar{D}_{k,1,1}(t), \chi(t, x)) \cap (\bar{E} \setminus \{x^1, x^2\})$, contradicting $\phi^{-1}_{\leq}(t, \bar{D}_{k,1,1}(t), \chi(t, x)) \cap (\bar{E} \setminus \{x_1, x_2\}) = \varnothing$.

(f) Since $\tilde{x}^2(\tau) \in D^{\leq}(t+\tau)$ and $\tilde{E}(\tau)$ walks to $E(\tau+1)$, it follows from Proposition A.1(1) that $x^2(\tau+1) = \phi(t+\tau+1, t+\tau, \tilde{x}^2(\tau)) \in D^{\leq}(t+\tau+1)$.

(g) It follows from Lemma A.14 that $E(\tau+1) \cap P_2 = \varnothing$.

(h) It follows from Lemma A.14 that $x^1(\tau+1) = x^*$ or $\tilde{\bar{E}}(\tau+1) \subset P_k(t+\tau+1)$ for some $k \in \{0, 1\}$.

(i) By property (i) for time $t+\tau$, $\phi(t+\tau, t, x^2) = x^2(\tau)$ or $\phi(t+\tau, t, x^2) \in P_3$. Also, recall that $x^2(\tau) \in P_3$ or $x^2(\tau) = \tilde{x}^2(\tau)$. If $\phi(t+\tau, t, x^2) \in P_3$, then it follows from Lemma 3.1 that $\phi(t+\tau+1,t,x^2) \in P_3$. If $\phi(t+\tau,t,x^2) = x^2(\tau)$ and $x^2(\tau) \in P_3$, then it follows from Lemma 3.1 that $\phi(t+\tau+1,t,x^2) \in P_3$. If $\phi(t+\tau,t,x^2) = x^2(\tau)$ and $x^2(\tau) = \tilde{x}^2(\tau)$, then $x^2(\tau+1) = \phi(t+\tau+1,t+\tau,\tilde{x}^2(\tau)) = \phi(t+\tau+1,t+\tau,\phi(t+\tau,t,x^2)) = \phi(t+\tau+1,t,x^2)$.

Note that at time $\tau = \zeta(t,x^1)$, it holds that $x^1(\tau) = \phi(t+\tau,t,x^1) = x^* \in D^{\leq}(t+\tau)$. Also, $x^* = x^1(\tau) \notin E(\tau) \setminus \{x^2(\tau)\}$. In addition, $x^2(\tau) \in D^{\leq}(t+\tau)$ and thus it follows from Corollary A.1 that $\phi(t+\tau+\tau',t+\tau,x^2(\tau)) \to (1,0)$ as $\tau' \to \infty$. It also follows that $\tilde{E}(\tau) = E(\tau) \in P_1(t+\tau)$ and thus $\tilde{x}^2(\tau) = x^2(\tau)$. Hence $\phi(t+\tau,t,x^2) = \tilde{x}^2(\tau)$ or $\phi(t+\tau,t,x^2) \in P_3$. Therefore the induction hypothesis for Case 1 holds at time $\tau = \zeta(t,x^1)$. Then the same induction argument as in Case 1 shows that the induction hypothesis for Case 1 holds at times $\tau \in \{\zeta(t,x^1),\ldots,\chi(t,x)\}$.

Thus, $\phi(t+\tau,t,x) \in E(\tau)$ and $E(\tau) \cap P_2 = \varnothing$, and hence $\phi(t+\tau,t,x) \notin P_2$ for all $\tau \in \{t,\ldots,\chi(t,x)\}$. Thus, $\phi(t+\chi(t,x),t,x) \in P_3$. It follows from Theorem 3.2 that $\phi(t+\tau,t,x) \to (1,0)$ as $\tau \to \infty$.

(4) The proof of (4) is similar to the proof of (3). □

*A.1.4.6 Characterization of Convergence for $x^*_{-1} = x^*_1$*

**The Rational Case**

For $j \in \mathbb{Z}$, let $v_j(t)$ denote the top right vertex of $Q_j(t)$, i.e., let

$$v_j(t) \quad := \quad \left(x^*_{-1} + j\delta^0_{-1}(t)/q, x^*_1 + j\delta^0_1(t)/q\right).$$

**Lemma A.15.** *Consider any $k \in \{0,1\}$, $i \in I_k(t) \cap J_k(t)$, and $t \in \mathbb{N}_0$. Then, $\cup_{j\in\mathcal{I}_{k,i}} Q_j(t) = D_{k,i,i}(t) \setminus \left(\cup_{j\in\mathcal{I}_{k,i}} \Omega(v_j(t))\right)$.*

**Proof**: Note that for each $j \in \mathcal{I}_{k,i}$, it holds that $Q_j(t) = \Omega^1(v_{j-1}(t)) \cap \Omega^0(v_j(t))$. Consider any $x \in \cup_{j\in\mathcal{I}_{k,i}} Q_j(t) \subset D_{k,i,i}(t)$. Let $j \in \mathcal{I}_{k,i}$ be such that $x \in Q_j(t)$. Thus, $x \in \Omega^1(v_{j-1}(t)) \cap \Omega^0(v_j(t)) = \left(\cap_{\{j'\in\mathcal{I}_{k,i}\,:\,j'\leq j-1\}} \Omega^1(v_{j'}(t))\right) \cap \left(\cap_{\{j'\in\mathcal{I}_{k,i}\,:\,j'\geq j\}} \Omega^0(v_{j'}(t))\right)$. Thus, $x \notin \cup_{j\in\mathcal{I}_{k,i}} \Omega(v_j(t))$. Hence, $\cup_{j\in\mathcal{I}_{k,i}} Q_j(t) \subset D_{k,i,i}(t) \setminus \left(\cup_{j\in\mathcal{I}_{k,i}} \Omega(v_j(t))\right)$.

Next, consider any $x \in D_{k,i,i}(t) \setminus \left(\cup_{j \in \mathcal{I}_{k,i}} \Omega(v_j(t))\right)$. Thus, $x \in D_{k,i,i}(t)$ and $x \in \cap_{j \in \mathcal{I}_{k,i}} \left(\Omega^0(v_j(t)) \cup \Omega^1(v_j(t))\right)$. Note that $x \in \Omega^0(\mathrm{cv}(D_{k,i,i}(t)))$ and thus $\{j' \in \mathcal{I}_{k,i} : x \in \Omega^0(v_{j'}(t))\} \neq \varnothing$. Let $j := \min\{j' \in \mathcal{I}_{k,i} : x \in \Omega^0(v_{j'}(t))\}$. Then $x \in \Omega^1(v_{j-1}(t)) \cap \Omega^0(v_j(t)) = Q_j(t)$. Hence, $D_{k,i,i}(t) \setminus \left(\cup_{j \in \mathcal{I}_{k,i}} \Omega(v_j(t))\right) \subset \cup_{j \in \mathcal{I}_{k,i}} Q_j(t)$. Therefore, $\cup_{j \in \mathcal{I}_{k,i}} Q_j(t) = D_{k,i,i}(t) \setminus \left(\cup_{j \in \mathcal{I}_{k,i}} \Omega(v_j(t))\right)$. □

**Lemma A.16.** *Suppose that $x_1^* = x_{-1}^*$ and that $x_1^*/(1 - x_1^*) = K + p/q$, where $K \in \mathbb{N}_0$, $p, q \in \mathbb{N}$, $p/q < 1$, and gcd$(p, q) = 1$, or $K \in \mathbb{N}$, $p = 0$, and $q = 1$. Consider any $t \in \mathbb{N}_0$. Then, the following hold:*

*(1) If $j \leq 0$, then $\phi(t+1, t, v_j(t)) = v_{j+q}(t+1)$, and if $j > 0$, then $\phi(t+1, t, v_j(t)) = v_{j-(Kq+p)}(t+1)$.*

*(2) If $j \leq 0$, then $\phi(t+1, t, Q_j(t)) = Q_{j+q}(t+1)$, and if $j > 0$, then $\phi(t+1, t, Q_j(t)) = Q_{j-(Kq+p)}(t+1)$.*

*(3) For all $x \in Q(t)$ it holds that $\phi(t+\tau, t, x) \to x^*$ as $\tau \to \infty$.*

**Proof**:

(1) If $j \leq 0$, then it follows from (3.10) that

$$\phi_m(t+1, t, v_j(t)) = \frac{(M_m + t)v_j(t) + 1}{M_m + t + 1} = x_m^* + (j+q)\frac{\delta_m^0(t+1)}{q} = v_{j+q}(t+1).$$

If $j > 0$, then it follows from (3.10) that

$$\phi_m(t+1, t, v_j(t)) = \frac{(M_m + t)v_j(t)}{M_m + t + 1} = x_m^* + (j - (Kq+p))\frac{\delta_m^0(t+1)}{q} = v_{j-(Kq+p)}(t+1).$$

(2) Consider any $x \in Q_j(t)$, that is,

$$x_m \in \left(x_m^* + (j-1)\frac{\delta_m^0(t)}{q}, x_m^* + j\frac{\delta_m^0(t)}{q}\right] \quad \text{for } m = \pm 1. \tag{A.7}$$

If $j \leq 0$, then $x \in \hat{P}_0$, and it follows from (3.10) that

$$\begin{aligned} \phi_m(t+1, t, x) &= \frac{(M_m + t)x_m + 1}{M_m + t + 1} \\ &\in \left(x_m^* + (j+q-1)\frac{\delta_m^0(t+1)}{q}, x_m^* + (j+q)\frac{\delta_m^0(t+1)}{q}\right] \end{aligned}$$

for $m = \pm 1$. Thus, $\phi(t+1, t, Q_j(t)) \subset Q_{j+q}(t+1)$. If $j > 0$, then $x \in \hat{P}_1$, and it follows from (3.10) that

$$\begin{aligned} \phi_m(t+1, t, x) &= \frac{(M_m + t)x_m}{M_m + t + 1} \\ &\in \left(x^*_m + (j - (Kq+p) - 1)\frac{\delta^0_m(t+1)}{q}, x^*_m + (j - (Kq+p))\frac{\delta^0_m(t+1)}{q}\right] \end{aligned}$$

for $m = \pm 1$. Thus, $\phi(t+1, t, Q_j(t)) \subset Q_{j-(Kq+p)}(t+1)$.

Conversely, consider any $j \le 0$, and any $y \in Q_{j+q}(t+1)$, that is,

$$y_m \in \left(x^*_m + (j+q-1)\delta^0_m(t+1)/q, x^*_m + (j+q)\delta^0_m(t+1)/q\right]$$

for $m = \pm 1$. Note that $x$ with $x_m = [y_m (M_m + t + 1) - 1] / (M_m + t)$ satisfies $x \in Q_j(t) \subset \hat{P}_0$ and $\phi_m(t+1, t, x) = y$. Thus, $Q_{j+q}(t+1) \subset \phi(t+1, t, Q_j(t))$. Similarly, consider any $j > 0$, and any $y \in Q_{j-(Kq+p)}(t+1)$, that is,

$$y_m \in \left(x^*_m + (j - (Kq+p) - 1)\frac{\delta^0_m(t+1)}{q}, x^*_m + (j - (Kq+p))\frac{\delta^0_m(t+1)}{q}\right]$$

for $m = \pm 1$. Note that $x$ with $x_m = y_m (M_m + t + 1) / (M_m + t)$ satisfies $x \in Q_j(t) \subset \hat{P}_1$ and $\phi_m(t+1, t, x) = y$. Thus, $Q_{j-(Kq+p)}(t+1) \subset \phi(t+1, t, Q_j(t))$.

(3) If $x \in Q(t)$, then it follows from a repeated application of (2) that $\phi(t+\tau, t, x) \in Q(t+\tau) \subset \cup_{D \in \mathcal{D}_=(t+\tau)} D$ for all $\tau \in \mathbb{N}_0$, and hence it follows from Proposition 3.7 that $\phi(t+\tau, t, x) \to x^*$ as $\tau \to \infty$. □

For any $t \in \mathbb{N}_0$, note that $\mathrm{cv}(D_{0,i,i}(t)) = v_{-(i-1)q}(t)$ for $i \in I_0(t) \cap J_0(t)$, and $\mathrm{cv}(D_{1,i,i}(t)) = v_{i(Kq+p)}(t)$ for $i \in I_1(t) \cap J_1(t)$.

**Proposition A.2.** *Suppose that $x^*_1 = x^*_{-1}$ and that $x^*_1/(1 - x^*_1) = K + p/q$, where $K \in \mathbb{N}_0$, $p, q \in \mathbb{N}$, $p/q < 1$, and $\gcd(p, q) = 1$, or $K \in \mathbb{N}$, $p = 0$ and $q = 1$. Consider any $t \in \mathbb{N}_0$, $k \in \{0, 1\}$, $i \in I_k(t) \cap J_k(t)$. Then, $\phi^{-1}(t, D_{k,i,i}(t)) = \{v_j(t) : j \in \mathcal{I}_{k,i}\}$.*

**Proof**: First we show that $\{v_j(t) : j \in \mathcal{I}_{k,i}\} \subset \phi^{-1}(t, D_{k,i,i}(t))$. Consider any $v_j(t)$ with $j \in \mathcal{I}_{k,i}$. We consider 3 cases.

Case 1: $K \in \mathbb{N}$, $p = 0$, $q = 1$: Then, $x^*_1/(1 - x^*_1) = K$. First suppose that $k = 0$, that is, $j \le 0$. Note that for any $i \in I_0(t) \cap J_0(t)$, it holds that $Q_{1-i}(t) = D_{0,i,i}(t)$ and

$v_{1-i}(t) = \mathrm{cv}(D_{0,i,i}(t))$, and for any $j \le 0$, it holds that $v_j(t) = \mathrm{cv}(D_{0,1-j,1-j}(t))$. It follows from Proposition 3.6 that $D_{0,1-j,1-j}(t)$ walks to $D_{0,1,1}(t-j)$ from time $t$ to time $t-j$. Thus, $\phi(t-j, t, v_j(t)) = x^*$, and thus $v_j(t) \in \phi^{-1}(t, D_{k,i,i}(t))$. Next suppose that $k = 1$, that is, $j \in \{(i-1)K+1, (i-1)K+2, \ldots, iK\}$. Since $j > 0$, it follows from Lemma A.16(1) that $\phi(t+\tau, t, v_j(t)) = v_{j-\tau K}(t+\tau)$ for all $0 \le \tau \le i$. Since $j - iK \le 0$, it follows that $\phi(t+i, t, v_j(t)) = v_{j-iK}(t+i) = \mathrm{cv}(D_{0,1-j+iK,1-j+iK}(t+i))$. It follows from Proposition 3.6 that $\phi(t+i-j+iK, t, v_j(t)) = x^*$. Thus, $v_j(t) \in \phi^{-1}(t, D_{k,i,i}(t))$.

Case 2: $K \in \mathbb{N}$, $p, q \in \mathbb{N}$, $p/q < 1$, and $\gcd(p,q) = 1$: First suppose that $k = 0$, that is, $j \le 0$. Then $j = -lq - n$, where $l \in \mathbb{N}_0$ and $n \in \{0, 1, \ldots, q-1\}$. It follows from Lemma A.16(1) that $\phi(t+l, t, v_j(t)) = v_{j+lq}(t+l)$. Note that $j + lq = -n \in \mathcal{I}_- := \{-q+1, \ldots, -1, 0\}$. Next suppose that $k = 1$, that is, $j > 0$. Then $j = l'(Kq+p) + n'$, where $l' \in \mathbb{N}_0$ and $n' \in \{1, \ldots, Kq+p\}$. It follows from Lemma A.16(1) that $\phi(t+l'+1, t, v_j(t)) = v_{j-(l'+1)(Kq+p)}(t+l'+1)$. Note that $j - (l'+1)(Kq+p) = n' - (Kq+p) \le 0$. Then $j - (l'+1)(Kq+p) = -lq - n$, where $l \in \mathbb{N}_0$ and $n \in \{0, 1, \ldots, q-1\}$. It follows from Lemma A.16(1) that $\phi(t+l'+1+l, t, v_j(t)) = v_{j-(l'+1)(Kq+p)+lq}(t+l'+1+l)$. Note that $j - (l'+1)(Kq+p) + lq = -n \in \mathcal{I}_-$.

Next, consider $v_j(t)$ with $j \in \mathcal{I}_-$. We show that there exists $\tau \in \mathbb{N}_0$ such that $\phi(t+\tau, t, v_j(t)) = x^*$. For any $\tau \in \mathbb{N}_0$, let $f(\tau)$ denote the negative of the index of $\phi(t+\tau, t, v_j(t))$, that is, $\phi(t+\tau, t, v_j(t)) = v_{-f(\tau)}(t+\tau)$. Thus, we will show that there exists $\tau \in \mathbb{N}_0$ such that $f(\tau) = 0$.

Inductively define $\{\tau_n\}_{n=0}^{\infty}$ as follows. Let $\tau_0 = 0$. Note that $-f(\tau_0) = j \in \mathcal{I}_-$. For $n = 0, 1, \ldots$, let $\tau_{n+1} := \inf\{\tau > \tau_n : -f(\tau) \in \mathcal{I}_-\}$. Suppose that $-f(\tau_n) \in \mathcal{I}_-$. Then, $f(\tau_n) + p \in \{p, \ldots, p+q-1\}$. It follows from Lemma A.16(1) that $f(\tau_n + 1) = f(\tau_n) - q \in \{-q, \ldots, -1\}$ and that $f(\tau_n + 2) = f(\tau_n) - q + Kq + p = f(\tau_n) + (K-1)q + p \ge 0$ since $K \ge 1$. It follows from Lemma A.16(1) that, if $f(\tau_n) + p < q$, then $\tau_{n+1} = \tau_n + 2 + K - 1$, and $f(\tau_{n+1}) = f(\tau_n) + p \in \{p, \ldots, q-1\} \subset -\mathcal{I}_-$. If $f(\tau_n) + p \ge q$, then $\tau_{n+1} = \tau_n + 2 + K$, and $f(\tau_{n+1}) = f(\tau_n) + p - q \in \{0, \ldots, p-1\} \subset -\mathcal{I}_-$. Thus,

$$f(\tau_{n+1}) = \begin{cases} f(\tau_n) + p & \text{if } f(\tau_n) + p < q, \\ f(\tau_n) + p - q & \text{if } f(\tau_n) + p \ge q, \end{cases}$$

which implies that

$$f(\tau_n) = (-j + np) \bmod q$$

for all $n \in \mathbb{N}_0$.

Since $p \geq 1$ and $\gcd(p, q) = 1$, there exists $p^* \in \mathbb{N}$ such that $(p^* p) \bmod q = 1$. Thus, there exists $N \in \mathbb{N}$ such that $p^* p = Nq + 1$. Note that

$$f(\tau_{(j+q)p^*}) = (-j + (j+q)p^* p) \bmod q = (-j + (j+q)(Nq+1)) \bmod q = 0.$$

Thus, there exists $\tau \in \mathbb{N}_0$ such that $f(\tau) = 0$. Hence, $\phi(t+\tau, t, v_j(t)) = v_0(t+\tau) = x^*$, which implies that $v_j(t) \in \phi^{-1}(t, D_{k,i,i}(t))$.

Case 3: $K = 0$, $p, q \in \mathbb{N}$, $p/q < 1$, and $\gcd(p, q) = 1$: Then $(1 - x^*)/x^* = \tilde{K} + m/p$, where $\tilde{K} = q \in \mathbb{N}$, $m = 0$, $p = 1$, or $\tilde{K} \in \mathbb{N}$, $m, p \in \mathbb{N}$, $m/p < 1$, and $\gcd(m, p) = 1$. By changing variables $y^*_m := 1 - x^*_m$, similar arguments as in Cases 1 and 2 apply to this case.

Next we show that $\phi^{-1}(t, D_{k,i,i}(t)) \subset \{v_j(t) : j \in \mathcal{I}_{k,i}\}$. Consider any $x \in \phi^{-1}(t, D_{k,i,i}(t))$. Then there exists $\tau \in \mathbb{N}_0$ such that $\phi(t+\tau, t, x) = x^* \in D^{=}(t+\tau)$. It follows from Proposition A.1 that $\phi(t+\tau+\tau', t, x) \in D^{=}(t+\tau+\tau')$, and thus it follows from Proposition 3.7 that $\phi(t+\tau+\tau', t, x) \to x^*$ as $\tau' \to \infty$. We show by contradiction that $x \in \{v_j(t) : j \in \mathcal{I}_{k,i}\}$. Since $\{v_j(t) : j \in \mathcal{I}_{k,i}\} \subset \phi^{-1}(t, D_{k,i,i}(t))$, it follows that if $x \in \cup_{j \in \mathcal{I}_{k,i}} \Omega(v_j(t))$, then $x \in D^2_{k,i,i}(t) \cup D^3_{k,i,i}(t)$. Then it follows from Theorem 3.4((1)) and 3.4((2)) that $\phi(t+\tau+\tau', t, x) \to (1, 0)$ or $\phi(t+\tau+\tau', t, x) \to (0, 1)$ as $\tau' \to \infty$, contradicting $\phi(t+\tau+\tau', t, x) \to x^*$ as $\tau' \to \infty$. If $x \in D_{k,i,i}(t) \setminus \left[\cup_{j \in \mathcal{I}_{k,i}} \Omega(v_j(t))\right]$, then it follows from Lemma A.15 that $x \in \cup_{j \in \mathcal{I}_{k,i}} Q_j(t)$. Thus, there exists some $j \in \mathcal{I}_{k,i}$ such that $x \in Q_j(t)$. Since $x \notin \{v_j(t) : j \in \mathcal{I}_{k,i}\}$, it holds that $\Omega(x) \cap Q_j(t) \neq \varnothing$. Consider any $y \in \Omega(x) \cap Q_j(t)$. Since $x \in \phi^{-1}(t, D_{k,i,i}(t))$, it follows that $y \in D^2_{k,i,i}(t) \cup D^3_{k,i,i}(t)$. It follows from Theorem 3.4((1)) and 3.4((2)) that $\phi(t+\tau, t, y) \to (1, 0)$ or $\phi(t+\tau, t, y) \to (0, 1)$ as $\tau \to \infty$, contradicting $\phi(t+\tau, t, y) \to x^*$ by Lemma A.16(3). Thus, $x \in \{v_j(t) : j \in \mathcal{I}_{k,i}\}$, and hence $\phi^{-1}(t, D_{k,i,i}(t)) \subset \{v_j(t) : j \in \mathcal{I}_{k,i}\}$. □

**Proof of Theorem 3.5**: Results (1) and (2) follow from Theorem 3.4((1)) and Theorem 3.4((2)). It follows from Proposition A.2 that $\phi^{-1}(0, D_{k,i,i}(0)) = \{v_j(0) : j \in \mathcal{I}_{k,i}\}$. Thus $D^l_{k,i,i}(0) := D_{k,i,i}(0) \cap \left[\cup_{x \in \phi^{-1}(0, D_{k,i,i}(0))} \Omega^l(x)\right] = D_{k,i,i}(0) \cap \left[\cup_{j \in \mathcal{I}_{k,i}} \Omega^l(v_j(0))\right]$.

Hence $D_{k,i,i}(0) \setminus (D^2_{k,i,i}(0) \cup D^3_{k,i,i}(0)) = D_{k,i,i}(0) \cap D^2_{k,i,i}(0)^c \cap D^3_{k,i,i}(0)^c = D_{k,i,i}(0) \cap \left[\cap_{j\in\mathcal{I}_{k,i}}\Omega^2(v_j(0))^c\right] \cap \left[\cap_{j\in\mathcal{I}_{k,i}}\Omega^3(v_j(0))^c\right] = D_{k,i,i}(0) \cap \left[\cap_{j\in\mathcal{I}_{k,i}} \left(\Omega^2(v_j(0)) \cup \Omega^3(v_j(0))\right)^c\right] = D_{k,i,i}(0) \cap \left[\cap_{j\in\mathcal{I}_{k,i}}\Omega(v_j(0))^c\right] = D_{k,i,i}(0) \cap \left[\cup_{j\in\mathcal{I}_{k,i}}\Omega(v_j(0))\right]^c = D_{k,i,i}(0) \setminus \left[\cup_{j\in\mathcal{I}_{k,i}}\Omega(v_j(0))\right] = \cup_{j\in\mathcal{I}_{k,i}}Q_j(0)$, where the last equality follows from Lemma A.15. Therefore $\cup_{k\in\{0,1\}}\cup_{i\in I_k(0)\cap J_k(0)} \left[D_{k,i,i}(0) \setminus (D^2_{k,i,i}(0) \cup D^3_{k,i,i}(0))\right] = \cup_{k\in\{0,1\}} \cup_{i\in I_k(0)\cap J_k(0)} \cup_{j\in\mathcal{I}_{k,i}} Q_j(0) = Q(0)$. It follows from Lemma A.16(3) that $\phi(t, 0, x) \to x^*$ as $t \to \infty$ for all $x \in Q(0)$. □

**The Irrational Case**

Let $x^*_1/(1-x^*_1) = K+\omega$, where $K \in \mathbb{N}_0$ and $\omega \in (0,1)\setminus\mathbb{Q}$. We show that in the irrational case, $\phi(t+\tau, t, x) \to x^*$ as $\tau \to \infty$ if and only if $x \in D^=(t)$. For each $t \in \mathbb{N}_0$ and $r \in \mathbb{R}$, let $w(t, r) = (w_{-1}(t, r), w_1(t, r)) := (x^*_{-1}+r\delta^0_{-1}(t), x^*_1+r\delta^0_1(t))$. If $x^*_1 = x^*_{-1}$, the diagonal line at time $t$ can be written as $D^=(t) = \left\{x \in P(t) : (x_1 - x^*_1)(M_1 + t) = (x_{-1} - x^*_{-1})(M_{-1} + t)\right\} = \{w(t, r) \in P(t) : r \in \mathbb{R}\}$. For any $t, \tau \in \mathbb{N}_0$ and $x \in D^=(t)$, let $g(\tau, x) = -r$ iff $\phi(t+\tau, t, x) = w(t + \tau, r)$. That is, for any $t, \tau \in \mathbb{N}_0$ and $x \in D^=(t)$, $\phi(t + \tau, t, x) = w(t + \tau, -g(\tau, x))$.

**Lemma A.17.** *Suppose that $x^*_{-1} = x^*_1$. For any $t, \tau \in \mathbb{N}_0$ it holds that $x \in D^=(t)$ iff $\phi(t + \tau, t, x) \in D^=(t + \tau)$. If $x \in D^=(t)$, then $\phi(t + \tau, t, x) \to x^*$ as $\tau \to \infty$.*

**Proof**: It follows from Proposition A.1 that if $x \in D^=(t)$, then $\phi(t + \tau, t, x) \in D^=(t + \tau)$ for all $\tau \in \mathbb{N}_0$, if $x \in D^<(t)$ then $\phi(t + \tau, t, x) \in D^<(t + \tau)$ for all $\tau \in \mathbb{N}_0$, and if $x \in D^>(t)$ then $\phi(t+\tau, t, x) \in D^>(t+\tau)$ for all $\tau \in \mathbb{N}_0$. Hence $x \in D^=(t)$ iff $\phi(t+\tau, t, x) \in D^=(t+\tau)$. It follows from Proposition 3.7 that if $x \in D^=(t)$, then $\phi(t + \tau, t, x) \to x^*$ as $\tau \to \infty$. □

**Lemma A.18.** *Suppose that $x^*_{-1} = x^*_1$ and $x^*_1/(1 - x^*_1) = K + \omega$. Consider any $t, \tau \in \mathbb{N}_0$ and $x \in D^=(t)$. If $g(\tau, x) \geq 0$, then $g(\tau + 1, x) = g(\tau, x) - 1$, and if $g(\tau, x) < 0$, then $g(\tau + 1, x) = g(\tau, x) + K + \omega$.*

**Proof**: If $g(\tau, x) \geq 0$, then $\phi(t + \tau, t, x) \in P_0(t)$ and

$$\phi_m(t+\tau+1, t, x) = \frac{(M_m + t + \tau)\phi_m(t + \tau, t, x) + 1}{M_m + t + \tau + 1} = x^*_m - [g(\tau, x) - 1]\delta^0_m(t+\tau+1)$$

for $m = \pm1$, and thus $g(\tau+1, x) = g(\tau, x)-1$. If $g(\tau, x) < 0$, then $\phi(t+\tau, t, x) \in P_1(t)$ and

$$\phi_m(t+\tau+1, t, x) = \frac{(M_m + t + \tau)\phi_m(t + \tau, t, x)}{M_m + t + \tau + 1} = x^*_m - [g(\tau, x)+K+\omega]\delta^0_m(t+\tau+1)$$

for $m = \pm 1$, and thus $g(\tau + 1, x) = g(\tau, x) + K + \omega$. $\square$

For any $r \in \mathbb{R}$, let $\text{frac}(r) := r \bmod 1$ denote the fractional part of $r$. For any $i \in \mathbb{N}$, let $\mathcal{R}_i := \{-i + \text{frac}(n\omega) : n \in \mathbb{N}\}$. It follows from Kronecker's Approximation Theorem for the one-dimensional case [3] that if $\omega$ is irrational, then $\{\text{frac}(n\omega) : n \in \mathbb{N}\}$ is dense in (0,1). Thus $\mathcal{R}_i$ is dense in $(-i, -i+1]$. Note that if $r \in \mathcal{R}_i$, then $w_m(t,r) \in (x_m^* - i\delta_m^0(t), x_m^* - (i-1)\delta_m^0(t))$, and thus $w(t,r) \in D_{0,i,i}(t)$.

**Lemma A.19.** *Suppose that $x_{-1}^* = x_1^*$ and $x_1^*/(1 - x_1^*) = K + \omega$, where $K \in \mathbb{N}_0$ and $\omega \in (0,1) \setminus \mathbb{Q}$. Consider any $t \in \mathbb{N}_0$, $i \in I_0(t) \cap J_0(t)$. Then, $D_{0,i,i}(t) \setminus D^{=}(t) = (\cup_{r \in \mathcal{R}_i} \Omega(w(t,r))) \cap D_{0,i,i}(t)$.*

**Proof**: Consider any $x \in D_{0,i,i}(t) \setminus D^{=}(t)$. Thus, $x_m \in (x_m^* - i\delta_m^0(t), x_m^* - (i-1)\delta_m^0(t)]$ for $m = \pm 1$. Thus, there exist $r_{-1}, r_1 \in (-i, -i+1]$ such that $r_{-1} \neq r_1$, $x_{-1} = x_{-1}^* + r_{-1}\delta_{-1}^0(t)$, and $x_1 = x_1^* + r_1\delta_1^0(t)$. Recall that $\mathcal{R}_i$ is dense in $(-i, -i+1]$. If $r_{-1} > r_1$, then there exists $r \in (r_1, r_{-1}) \cap \mathcal{R}_i$. Then, $x_{-1} > w_{-1}(t,r)$ and $x_1 < w_1(t,r)$, which implies that $x \in \Omega^3(w(t,r))$. If $r_{-1} < r_1$, then there exists $r \in (r_{-1}, r_1) \cap \mathcal{R}_i$. Then, $x_{-1} < w_{-1}(t,r)$ and $x_1 > w_1(t,r)$, which implies $x \in \Omega^2(w(t,r))$. Thus, $x \in (\cup_{r \in \mathcal{R}_i} \Omega(w(t,r))) \cap D_{0,i,i}(t)$, and hence $D_{0,i,i}(t) \setminus D^{=}(t) \subset (\cup_{r \in \mathcal{R}_i} \Omega(w(t,r))) \cap D_{0,i,i}(t)$.

Consider any $x \in (\cup_{r \in \mathcal{R}_i} \Omega(w(t,r))) \cap D_{0,i,i}(t)$. There exists $r \in \mathcal{R}_i$ such that $x \in \Omega(w(t,r))$. If $x \in \Omega^2(w(t,r))$, then $x_{-1} \leq w_{-1}(t,r) = x_{-1}^* + r(1 - x_{-1}^*)/(M_{-1} + t)$ and $x_1 > w_1(t,r) = x_1^* + r(1 - x_1^*)/(M_1 + t)$, and thus $(x_{-1} - x_{-1}^*)(M_{-1} + t) \leq r(1 - x_{-1}^*) = r(1 - x_1^*) < (x_1 - x_1^*)(M_1 + t)$, which implies that $x \notin D^{=}(t)$. If $x \in \Omega^3(w(t,r))$, then $x_{-1} > w_{-1}(t,r) = x_{-1}^* + r(1 - x_{-1}^*)/(M_{-1} + t)$ and $x_1 \leq w_1(t,r) = x_1^* + r(1 - x_1^*)/(M_1 + t)$, and thus $(x_{-1} - x_{-1}^*)(M_{-1} + t) > r(1 - x_{-1}^*) = r(1 - x_1^*) \geq (x_1 - x_1^*)(M_1 + t)$, which implies that $x \notin D^{=}(t)$. Thus, $x \in D_{0,i,i}(t) \setminus D^{=}(t)$, and hence $(\cup_{r \in \mathcal{R}_i} \Omega(w(t,r))) \cap D_{0,i,i}(t) \subset D_{0,i,i}(t) \setminus D^{=}(t)$. $\square$

**Proposition A.3.** *Suppose that $x_{-1}^* = x_1^*$ and $x_1^*/(1 - x_1^*) = K + \omega$, where $K \in \mathbb{N}$ and $\omega \in (0,1) \setminus \mathbb{Q}$. Consider any $t \in \mathbb{N}_0$ and $i \in I_0(t) \cap J_0(t)$. Then, $\{w(t,r) : r \in \mathcal{R}_i\} \subset \phi^{-1}(t, D_{0,i,i}(t))$.*

**Proof**: Consider any $r \in \mathcal{R}_i$. Thus, there exists $n^* \in \mathbb{N}$ such that $r = -i + \text{frac}(n^*\omega)$.

Note that $g(0, w(t,r)) = -r \in [i-1, i)$. It follows from Lemma A.18 that $g(i-1, w(t,r)) = -r - (i-1) \in [0,1)$. Define $\tau_0 = i-1$ and inductively define $\tau_{n+1} := \inf\{\tau > \tau_n : g(\tau, w(t,r)) \in [0,1)\}$.

Thus, $g(\tau_0, w(t,r)) \in [0,1)$. Consider any $n \in \mathbb{N}_0$, and assume that $g(\tau_n, w(t,r)) \in [0,1)$. It follows from Lemma A.18 that $g(\tau_n + 1, w(t,r)) = g(\tau_n, w(t,r)) - 1 \in [-1, 0)$, and that $g(\tau_n + 2, w(t,r)) = g(\tau_n, w(t,r)) - 1 + K + \omega > 0$ since $K \geq 1$ and $\omega > 0$. Note that $g(\tau_n, w(t,r)) + \omega \in (0,2)$. If $g(\tau_n, w(t,r)) + \omega < 1$, then $\tau_{n+1} = \tau_n + 2 + K - 1$ and $g(\tau_{n+1}, w(t,r)) = g(\tau_n, w(t,r)) + \omega \in (0,1)$. If $g(\tau_n, w(t,r)) + \omega \geq 1$, then $\tau_{n+1} = \tau_n + 2 + K$ and $g(\tau_{n+1}, w(t,r)) = g(\tau_n, w(t,r)) + \omega - 1 \in [0,1)$. Thus

$$g(\tau_{n+1}, w(t,r)) = \begin{cases} g(\tau_n, w(t,r)) + \omega & \text{if } g(\tau_n, w(t,r)) + \omega < 1, \\ g(\tau_n, w(t,r)) + \omega - 1 & \text{if } g(\tau_n, w(t,r)) + \omega \geq 1. \end{cases} \tag{A.8}$$

which is equivalent to

$$g(\tau_n, w(t,r)) = \mathrm{frac}(-r - (i-1) + n\omega) = \mathrm{frac}(1 - \mathrm{frac}(n^*\omega) + n\omega)$$

for $n = 0, 1, \ldots$. Thus, $g(\tau_{n^*}, w(t,r)) = \mathrm{frac}(1 - \mathrm{frac}(n^*\omega) + n^*\omega) = 0$.

Hence, $\phi(t + \tau_{n^*}, t, w(t,r)) = w(t + \tau_{n^*}, 0) = x^*$. Thus, $w(t,r) \in \phi^{-1}(t, D_{0,i,i}(t))$. □

**Proof of Theorem 3.6:**

(1) Consider any $x \in D^>(0) \subset \cup_{D \in \mathcal{D}_2(0) \cup \mathcal{D}_=(0)} D \cup P_2$. Suppose that $K \in \mathbb{N}$. If $x \in \cup_{D \in \mathcal{D}_2(0)} D \cup P_2$, then it follows from Theorems 3.2 and 3.3 that $\phi(t, 0, x) \to (0,1)$ as $t \to \infty$. If $x \in \cup_{D \in \mathcal{D}_=(0)} D$, then there exists $k \in \{0,1\}$ and $i \in I_k(0) \cap J_k(0)$ such that $x \in D_{k,i,i}(0) \setminus D^=(0)$. First suppose that $k = 0$. Thus, $x \in D_{0,i,i}(0) \setminus D^=(0)$. It follows from Lemma A.19 and Proposition A.3 that $x \in (\cup_{r \in \mathcal{R}_i} \Omega(w(0,r))) \cap D_{0,i,i}(0) \subset \left(\cup_{y \in \phi^{-1}(0, D_{0,i,i}(0))} \Omega(y)\right) \cap D_{0,i,i}(0) = D^2_{0,i,i}(0) \cup D^3_{0,i,i}(0)$. Next suppose that $k = 1$. Then it follows from Proposition 3.6 that $\phi(i, 0, x) \in P_0$. If $\phi(i, 0, x) \in \cup_{D \in \mathcal{D}_2(i) \cup \mathcal{D}_3(i)} D$, then it follows from Theorem 3.3 that $\phi(t, 0, x) \to (0,1)$ or $\phi(t, 0, x) \to (1,0)$ as $t \to \infty$. Otherwise, there exists $i' \in I_0(i) \cap J_0(i)$ such that $\phi(i, 0, x) \in D_{0,i',i'}(i) \setminus D^=(i)$. Then it follows from Lemma A.19 and Proposition A.3 that $\phi(i, 0, x) \in \left(\cup_{r \in \mathcal{R}_{i'}} \Omega(w(i,r))\right) \cap D_{0,i',i'}(i) \subset \left(\cup_{y \in \phi^{-1}(i, D_{0,i',i'}(i))} \Omega(y)\right) \cap D_{0,i',i'}(i) =$

$D^2_{0,i',i'}(i) \cup D^3_{0,i',i'}(i)$. Thus it follows from Theorem 3.4((1)) and 3.4((2)) that $\phi(t,0,x) \to (0,1)$ or $\phi(t,0,x) \to (1,0)$ as $t \to \infty$. Since $x \in D^>(0)$ and $x^*_{-1} = x^*_1$, it follows from Proposition A.1(2) that $\phi(t,0,x) \in D^>(t)$ for all $t \in \mathbb{N}_0$. Thus, $\phi(t,0,x) \notin P_3$ for all $t \in \mathbb{N}_0$ and hence $\phi(t,0,x) \to (0,1)$ as $t \to \infty$.

If $K = 0$, then $x^*_1/(1-x^*_1) = \omega$. Then there exist $\tilde{K} \in \mathbb{N}$ and $\tilde{\omega} \in (0,1) \setminus \mathbb{Q}$ such that $1/\omega = \tilde{K} + \tilde{\omega}$ and $(1-x^*_1)/x^*_1 = \tilde{K} + \tilde{\omega}$. After changing variables to $y^*_1 := 1 - x^*_1$, a similar argument as for the case with $K \in \mathbb{N}$ can be used.

(2) Result (2) follows from a similar argument for Result (1).

(3) If $x \in D^=(0)$, then it follows from Lemma A.17 that $\phi(t,0,x) \to x^*$ as $t \to \infty$. □

**Proof of Proposition 3.8:** Consider any $\varepsilon > 0$. Choose $\delta = \min\{\varepsilon, x^*_{-1}, 1-x^*_1\}$. Consider any $x \in P$ such that $\|x - (0,1)\|_\infty < \delta$. Note that $x \in P_2$. By Lemma 3.1, $\phi(t,0,x) \in P_2$ for all $t \in \mathbb{N}_0$. Thus, it follows from (3.10) that

$$\begin{aligned} \|\phi(t,0,x) - (0,1)\|_\infty &= \max\left\{\left|\frac{M_{-1}x_{-1}}{M_{-1}+t}\right|, \left|\frac{M_1x_1 + t}{M_1+t} - 1\right|\right\} \\ &= \max\left\{\frac{M_{-1}\,|x_{-1}|}{M_{-1}+t}, \frac{M_1\,|x_1 - 1|}{M_1+t}\right\} < \delta \leq \varepsilon \end{aligned}$$

for all $t$, and hence $(0,1)$ is stable. By a similar argument, $(1,0)$ is also stable.

Consider any $\varepsilon \in (0, \min\{x^*_{-1}, 1-x^*_1\}/2)$ and any $\delta > 0$. There exists $x \in P_2$ such that $\|x - x^*\|_\infty < \delta$. Since $\phi(t,0,x) \to (0,1)$ as $t \to \infty$, there exists $t \in \mathbb{N}$ such that $\|\phi(t,0,x) - x^*\|_\infty > \varepsilon$. Thus $x^*$ is unstable. □

### A.1.5 Proofs for Discrete-time Fictitious Play under Case 2

**Lemma A.20.** *Consider any $x \in [0,1]$ and $m = \pm 1$. Then,* $\texttt{PBR}_m(x) = \{A^1_m\}$.

**Proof:** It follows from Lemma A.2 that

$$\begin{aligned} & x_{-m}\pi_m(A^0_m, A^1_{-m}) + (1-x_{-m})\pi_m(A^0_m, A^0_{-m}) \\ & \qquad < x_{-m}\pi_m(A^1_m, A^1_{-m}) + (1-x_{-m})\pi_m(A^1_m, A^0_{-m}), \end{aligned}$$

which implies $\texttt{PBR}_m(x) = \{A^1_m\}$. □

**Proof of Proposition 3.9:** It follows from Lemma A.20 that for each $m = \pm 1$ and $\tau \in \mathbb{N}_0$,

$$\phi_m(t+\tau, t, x) \quad = \quad \frac{(M_m + t)x_m + \sum_{i=1}^{\tau} \mathbf{1}_{[A_m(t+i)=A_m^1]}}{M_m + t + \tau} \quad = \quad \frac{(M_m + t)x_m + \tau}{M_m + t + \tau},$$

which implies that $\phi(t+\tau, t, x) \to (1,1)$ as $\tau \to \infty$. □

## *A.2 Assortment Competition with General Product sets and Simple Actions*

### A.2.1 Proof of Proposition 3.10:

The proof follows from similar arguments as those for Propositions 3.1 and 3.2. □

### A.2.2 Cournot Adjustment Process under Case 3

We provide a proof for Proposition 3.11 that describes the behavior of the Cournot adjustment process under Case 3 in Table 12. Under Case 3, it holds that $\bar{\beta}_{-1} < \beta_{-1} < \underline{\beta}_{-1}$ and $\underline{\beta}_1 < \beta_1 < \bar{\beta}_1$.

**Lemma A.21.** *Under Case 3 in Table 12, it holds that* $\pi_1(A_1^1, A_{-1}^1) < \pi_1(A_1^0, A_{-1}^1)$ *and* $\pi_1(A_1^0, A_{-1}^0) < \pi_1(A_1^1, A_{-1}^0)$ *and* $\pi_{-1}(A_{-1}^1, A_1^1) > \pi_{-1}(A_{-1}^0, A_1^1)$ *and* $\pi_{-1}(A_{-1}^0, A_1^0) > \pi_{-1}(A_{-1}^1, A_1^0)$.

**Proof**: For $m = \pm 1$, note that, $\bar{\lambda}_m > 0$, $\underline{\lambda}_m > 0$ and that

$$\pi_m(A_m^1, A_{-m}^1) - \pi_m(A_m^0, A_{-m}^1) \quad = \quad \underline{\lambda}_m(\underline{\beta}_m - \beta_m),$$
$$\pi_m(A_m^0, A_{-m}^0) - \pi_m(A_m^1, A_{-m}^0) \quad = \quad \bar{\lambda}_m(\beta_m - \bar{\beta}_m).$$

Thus, the result follows from $\bar{\beta}_{-1} < \beta_{-1} < \underline{\beta}_{-1}$ and $\underline{\beta}_1 < \beta_1 < \bar{\beta}_1$. □

**Proof of Proposition 3.11:** The result follows from Lemma A.21. □

## *A.3 Assortment Competition with General Product Sets and General Actions*

In this section we provide proofs for Propositions 3.12 and 3.13. We also present two corollaries that follow from the propositions.

**Proof of Proposition 3.12:**

(1) Consider any $A_{-m}^0 \in C_{-m}^0$ and any $A_m^0 \in \mathtt{PBR}_m^0(A_{-m}^0)$. Then $A_m^0 \in \mathtt{PBR}_m(A_{-m}^0)$ iff for all $A_m^1 \in \mathtt{PBR}_m^1(A_{-m}^0)$ it holds that

$$\pi_m(A_m^0, A_{-m}^0) \quad \geq \quad \pi_m(A_m^1, A_{-m}^0) \Leftrightarrow$$

$$\frac{v_{t_m}p_{t_m} + b_m(A^0_{-m}, A^0_m)}{1 + v_{t_m} + v_{t_{-m}} + c(A^0_{-m}, A^0_m)} \geq \frac{\alpha_m v_{t_m} p_{t_m} + b_m(A^0_{-m}, A^1_m)}{1 + \alpha_m v_{t_m} + \beta_m v_{t_{-m}} + c(A^0_{-m}, A^1_m)} \tag{A.9}$$
$$\Leftrightarrow \quad \beta_m \geq \bar{\beta}_m(A^0_{-m}).$$

(2) Consider any $A^1_{-m} \in C^1_{-m}$ and any $A^1_m \in \texttt{PBR}^1_m(A^1_{-m})$. Then $A^1_m \in \texttt{PBR}_m(A^1_{-m})$ iff for all $A^0_m \in \texttt{PBR}^0_m(A^1_{-m})$ it holds that

$$\begin{aligned} &\pi_m(A^1_m, A^1_{-m}) \geq \pi_m(A^0_m, A^1_{-m}) \Leftrightarrow \\ &\frac{\alpha_m \beta_{-m} v_{t_m} p_{t_m} + b_m(A^1_{-m}, A^1_m)}{1 + \alpha_m \beta_{-m} v_{t_m} + \alpha_{-m}\beta_m v_{t_{-m}} + c(A^1_{-m}, A^1_m)} \\ &\geq \frac{\beta_{-m} v_{t_m} p_{t_m} + b_m(A^1_{-m}, A^0_m)}{1 + \beta_{-m} v_{t_m} + \alpha_{-m} v_{t_{-m}} + c(A^1_{-m}, A^0_m)} \Leftrightarrow \\ &\beta_m \leq \underline{\beta}_m(A^1_{-m}). \end{aligned} \tag{A.10}$$

(3) Consider any $A^0_{-m} \in C^0_{-m}$ and any $A^1_m \in \texttt{PBR}^1_m(A^0_{-m})$. Then $A^1_m \in \texttt{PBR}_m(A^0_{-m})$ iff for all $A^0_m \in \texttt{PBR}^0_m(A^0_{-m})$ it holds that

$$\pi_m(A^1_m, A^0_{-m}) \geq \pi_m(A^0_m, A^0_{-m})$$
$$\Leftrightarrow \frac{\alpha_m v_{t_m} p_{t_m} + b_m(A^0_{-m}, A^1_m)}{1 + \alpha_m v_{t_m} + \beta_m v_{t_{-m}} + c(A^0_{-m}, A^1_m)} \geq \frac{v_{t_m} p_{t_m} + b_m(A^0_{-m}, A^0_m)}{1 + v_{t_m} + v_{t_{-m}} + c(A^0_{-m}, A^0_m)} \tag{A.11}$$
$$\Leftrightarrow \quad \beta_m \leq \bar{\beta}_m(A^0_{-m}).$$

(4) Consider any $A^1_{-m} \in C^1_{-m}$ and any $A^0_m \in \texttt{PBR}^0_m(A^1_{-m})$. Then $A^0_m \in \texttt{PBR}_m(A^1_{-m})$ iff for all $A^1_m \in \texttt{PBR}^1_m(A^1_{-m})$ it holds that

$$\begin{aligned} &\pi_m(A^0_m, A^1_{-m}) \geq \pi_m(A^1_m, A^1_{-m}) \Leftrightarrow \\ &\frac{\beta_{-m} v_{t_m} p_{t_m} + b_m(A^1_{-m}, A^0_m)}{1 + \beta_{-m} v_{t_m} + \alpha_{-m} v_{t_{-m}} + c(A^1_{-m}, A^0_m)} \\ &\geq \frac{\alpha_m \beta_{-m} v_{t_m} p_{t_m} + b_m(A^1_{-m}, A^1_m)}{1 + \alpha_m \beta_{-m} v_{t_m} + \alpha_{-m}\beta_m v_{t_{-m}} + c(A^1_{-m}, A^1_m)} \Leftrightarrow \\ &\beta_m \geq \underline{\beta}_m(A^1_{-m}). \end{aligned} \tag{A.12}$$

□

**Remark A.2.** *Note that the left sides of (A.9)–(A.11) do not depend on the choice of* $A^i_m \in$ *PBR*$^i_m(A_{-m})$*, and the right sides of (A.9)–(A.11) are the objective values* $\pi_m(A^i_m, A_{-m})$*, and* *are by definition of* $\mathtt{PBR}^i_m(A_{-m})$ *the same for all choices of* $A^i_m \in \mathtt{PBR}^i_m(A_{-m})$*. It follows that* $\bar{\beta}_m(A_{-m})$ *and* $\underline{\beta}_m(A_{-m})$ *do not depend on the choice of* $A^i_m \in \mathtt{PBR}^i_m(A_{-m})$*, which is why* $A^i_m$ *was omitted from the notation for* $\bar{\beta}_m(A_{-m})$ *and* $\underline{\beta}_m(A_{-m})$*.*

Corollary A.2 characterizes the conditions for the existence of pure-strategy Nash equilibria.

**Corollary A.2.** *The following holds:*

(1) *A pair of assortments* $(A^0_{-1}, A^0_1) \in C^0_{-1} \times C^0_1$ *is a Nash equilibrium iff* $A^0_m \in \mathtt{PBR}^0_m(A^0_{-m})$ *and* $\beta_m \geq \bar{\beta}_m(A^0_{-m})$ *for* $m = \pm 1$*.*

(2) *A pair of assortments* $(A^1_{-1}, A^1_1) \in C^1_{-1} \times C^1_1$ *is a Nash equilibrium iff* $A^1_m \in \mathtt{PBR}^1_m(A^1_{-m})$ *and* $\beta_m \leq \underline{\beta}_m(A^1_{-m})$ *for* $m = \pm 1$*.*

(3) *A pair of assortments* $(A^0_{-m}, A^1_m) \in C^0_{-m} \times C^1_m$ *is a Nash equilibrium iff* $A^0_{-m} \in \mathtt{PBR}^0_{-m}(A^1_m)$*,* $A^1_m \in \mathtt{PBR}^1_m(A^0_{-m})$*,* $\beta_{-m} \geq \underline{\beta}_{-m}(A^1_m)$*, and* $\beta_m \leq \bar{\beta}_m(A^0_{-m})$*.*

Let $C^{0+}_m(\bar{x}_m) := C^0_m \cap C^+_m(\bar{x}_m)$, and $C^{1+}_m(\bar{x}_m) := C^1_m \cap C^+_m(\bar{x}_m)$. Thus, $C^+_m(\bar{x}_m) = C^{0+}_m(\bar{x}_m) \cup C^{1+}_m(\bar{x}_m)$

**Proof of Proposition 3.13:**

(1) If $\bar{x}^0_m \in \mathtt{BR}^0_m(\bar{x}_{-m})$, then $\bar{x}^0_m \in \mathtt{BR}_m(\bar{x}_{-m})$ iff for all $A^0_m \in C^{0+}_m(\bar{x}^0_m)$ (note $C^{0+}_m(\bar{x}^0_m) \subset \mathtt{PBR}^0_m(\bar{x}_{-m})$) and all $A^1_m \in \mathtt{PBR}^1_m(\bar{x}_{-m})$ it holds that

$$\begin{aligned} \sum_{A_{-m} \in C_{-m}} \pi_m(A^0_m, A_{-m})\bar{x}_{-m}(A_{-m}) \quad &\geq \quad \sum_{A_{-m} \in C_{-m}} \pi_m(A^1_m, A_{-m})\bar{x}_{-m}(A_{-m}) \quad \text{(A.13)} \\ \Leftrightarrow \quad \sum_{A^0_{-m} \in C^{0+}_{-m}(\bar{x}_{-m})} &\left[\pi_m(A^0_m, A^0_{-m}) - \pi_m(A^1_m, A^0_{-m})\right] \bar{x}_{-m}(A^0_{-m}) \\ + \sum_{A^1_{-m} \in C^{1+}_{-m}(\bar{x}_{-m})} &\left[\pi_m(A^0_m, A^1_{-m}) - \pi_m(A^1_m, A^1_{-m})\right] \bar{x}_{-m}(A^1_{-m}) \quad \geq \quad 0. \end{aligned}$$

Since $A^0_m \in C^0_m$, $A^0_{-m} \in C^0_{-m}$, and $A^1_m \in C^1_m$, it follows that

$$\begin{aligned} &\pi_m(A^0_m, A^0_{-m}) - \pi_m(A^1_m, A^0_{-m}) \\ = \quad &\frac{v_{t_m} p_{t_m} + b_m(A^0_m)}{1 + v_{t_m} + v_{t_{-m}} + c(A^0_{-m}, A^0_m)} - \frac{\alpha_m v_{t_m} p_{t_m} + b_m(A^1_m)}{1 + \alpha_m v_{t_m} + \beta_m v_{t_{-m}} + c(A^0_{-m}, A^1_m)} \end{aligned}$$

$$= \bar{\lambda}_m(A^0_{-m}, A^0_m, A^1_m)\left[\beta_m - \bar{\beta}_m(A^0_{-m}, A^0_m, A^1_m)\right].$$

and since $A^0_m \in C^0_m$, $A^1_{-m} \in C^1_{-m}$, and $A^1_m \in C^1_m$, it follows that

$$\begin{aligned}
&\pi_m(A^0_m, A^1_{-m}) - \pi_m(A^1_m, A^1_{-m}) \\
= \ & \frac{\beta_{-m} v_{t_m} p_{t_m} + b_m(A^0_m)}{1 + \beta_{-m} v_{t_m} + \alpha_{-m} v_{t_{-m}} + c(A^1_{-m}, A^0_m)} - \frac{\alpha_m \beta_{-m} v_{t_m} p_{t_m} + b_m(A^1_m)}{1 + \alpha_m \beta_{-m} v_{t_m} + \alpha_{-m}\beta_m v_{t_{-m}} + c(A^1_{-m}, A^1_m)} \\
= \ & \underline{\lambda}_m(A^1_{-m}, A^0_m, A^1_m)\left[\beta_m - \underline{\beta}_m(A^1_{-m}, A^0_m, A^1_m)\right].
\end{aligned}$$

Then, (A.13) is equivalent to

$$\begin{aligned}
&\sum_{A^0_{-m} \in C^{0+}_{-m}(\bar{x}_{-m})} \bar{\lambda}_m(A^0_{-m}, A^0_m, A^1_m)\left[\beta_m - \bar{\beta}_m(A^0_{-m}, A^0_m, A^1_m)\right]\bar{x}_{-m}(A^0_{-m}) \\
&+ \sum_{A^1_{-m} \in C^{1+}_{-m}(\bar{x}_{-m})} \underline{\lambda}_m(A^1_{-m}, A^0_m, A^1_m)\left[\beta_m - \underline{\beta}_m(A^1_{-m}, A^0_m, A^1_m)\right]\bar{x}_{-m}(A^1_{-m}) \quad \geq \quad 0 \\
&\Leftrightarrow \quad \beta_m \quad \geq \quad \beta_m(\bar{x}_{-m}).
\end{aligned}$$

(2) If $\bar{x}^1_m \in \mathtt{BR}^1_m(\bar{x}_{-m})$, then $\bar{x}^1_m \in \mathtt{BR}_m(\bar{x}_{-m})$ iff for all $A^1_m \in C^{1+}_m(\bar{x}^1_m)$ (note $C^{1+}_m(\bar{x}^1_m) \subset \mathtt{PBR}^1_m(\bar{x}_{-m})$) and all $A^0_m \in \mathtt{PBR}^0_m(\bar{x}_{-m})$ it holds that

$$\begin{aligned}
&\sum_{A_{-m} \in C_{-m}} \pi_m(A^1_m, A_{-m})\bar{x}_{-m}(A_{-m}) \quad \geq \quad \sum_{A_{-m} \in C_{-m}} \pi_m(A^0_m, A_{-m})\bar{x}_{-m}(A_{-m}) \qquad \text{(A.14)} \\
\Leftrightarrow \quad &\sum_{A^0_{-m} \in C^{0+}_{-m}(\bar{x}_{-m})} \left[\pi_m(A^1_m, A^0_{-m}) - \pi_m(A^0_m, A^0_{-m})\right]\bar{x}_{-m}(A^0_{-m}) \\
&\quad + \sum_{A^1_{-m} \in C^{1+}_{-m}(\bar{x}_{-m})} \left[\pi_m(A^1_m, A^1_{-m}) - \pi_m(A^0_m, A^1_{-m})\right]\bar{x}_{-m}(A^1_{-m}) \quad \geq \quad 0.
\end{aligned}$$

Since $A^1_m \in C^1_m$, $A^0_{-m} \in C^0_{-m}$, and $A^0_m \in C^0_m$, it follows that

$$\begin{aligned}
&\pi_m(A^1_m, A^0_{-m}) - \pi_m(A^0_m, A^0_{-m}) \\
= \ & \frac{\alpha_m v_{t_m} p_{t_m} + b_m(A^1_m)}{1 + \alpha_m v_{t_m} + \beta_m v_{t_{-m}} + c(A^0_{-m}, A^1_m)} - \frac{v_{t_m} p_{t_m} + b_m(A^0_m)}{1 + v_{t_m} + v_{t_{-m}} + c(A^0_{-m}, A^0_m)} \\
= \ & \bar{\lambda}_m(A^0_{-m}, A^0_m, A^1_m)\left[\bar{\beta}_m(A^0_{-m}, A^0_m, A^1_m) - \beta_m\right].
\end{aligned}$$

and since $A^0_m \in C^0_m$, $A^1_{-m} \in C^1_{-m}$, and $A^1_m \in C^1_m$, it follows that

$$\begin{aligned}
&\pi_m(A^1_m, A^1_{-m}) - \pi_m(A^0_m, A^1_{-m}) \\
= \ & \frac{\alpha_m \beta_{-m} v_{t_m} p_{t_m} + b_m(A^1_m)}{1 + \alpha_m \beta_{-m} v_{t_m} + \alpha_{-m}\beta_m v_{t_{-m}} + c(A^1_{-m}, A^1_m)} - \frac{\beta_{-m} v_{t_m} p_{t_m} + b_m(A^0_m)}{1 + \beta_{-m} v_{t_m} + \alpha_{-m} v_{t_{-m}} + c(A^1_{-m}, A^0_m)} \\
= \ & \underline{\lambda}_m(A^1_{-m}, A^0_m, A^1_m)\left[\underline{\beta}_m(A^1_{-m}, A^0_m, A^1_m) - \beta_m\right].
\end{aligned}$$

Then, (A.14) is equivalent to

$$\begin{aligned}
&\sum_{A^0_{-m} \in C^{0+}_{-m}(\bar{x}_{-m})} \bar{\lambda}_m(A^0_{-m}, A^0_m, A^1_m) \left[\beta_m - \bar{\beta}_m(A^0_{-m}, A^0_m, A^1_m)\right] \bar{x}_{-m}(A^0_{-m}) \\
&+ \sum_{A^1_{-m} \in C^{1+}_{-m}(\bar{x}_{-m})} \underline{\lambda}_m(A^1_{-m}, A^0_m, A^1_m) \left[\beta_m - \underline{\beta}_m(A^1_{-m}, A^0_m, A^1_m)\right] \bar{x}_{-m}(A^1_{-m}) \quad \leq \quad 0 \\
&\Leftrightarrow \quad \beta_m \quad \leq \quad \beta_m(\bar{x}_{-m}).
\end{aligned}$$

(3) If $\bar{x}^2_m \in \; \mathtt{BR}^2_m(\bar{x}_{-m})$, then, for any $A^0_m \in C^{0+}_m(\bar{x}^2_m)$, $A^1_m \in C^{1+}_m(\bar{x}^2_m)$, and $A_m \in C_m$, it holds that

$$\begin{aligned}
\sum_{A_{-m} \in C_{-m}} \pi_m(A^0_m, A_{-m})\bar{x}_{-m}(A_{-m}) \quad &= \quad \sum_{A_{-m} \in C_{-m}} \pi_m(A^1_m, A_{-m})\bar{x}_{-m}(A_{-m}) \\
&\geq \quad \sum_{A_{-m} \in C_{-m}} \pi_m(A_m, A_{-m})\bar{x}_{-m}(A_{-m}),
\end{aligned}$$

which implies that $\bar{x}^2_m \in \mathtt{BR}_m(\bar{x}_{-m})$, and thus $\mathtt{BR}^2_m(\bar{x}_{-m}) \subset \mathtt{BR}_m(\bar{x}_{-m})$.

In addition, since $C^{0+}_m(\bar{x}^2_m) \subset \mathtt{PBR}^0_m(\bar{x}_{-m})$, it holds for all $\hat{A}^0_m \in \mathtt{PBR}^0_m(\bar{x}_{-m})$ that

$$\sum_{A_{-m} \in C_{-m}} \pi_m(\hat{A}^0_m, A_{-m})\bar{x}_{-m}(A_{-m}) \quad = \quad \sum_{A_{-m} \in C_{-m}} \pi_m(A^0_m, A_{-m})\bar{x}_{-m}(A_{-m}).$$

Similarly, since $C^{1+}_m(\bar{x}^2_m) \subset \mathtt{PBR}^1_m(\bar{x}_{-m})$, it holds for all $\hat{A}^1_m \in \mathtt{PBR}^1_m(\bar{x}_{-m})$ that

$$\sum_{A_{-m} \in C_{-m}} \pi_m(\hat{A}^1_m, A_{-m})\bar{x}_{-m}(A_{-m}) \quad = \quad \sum_{A_{-m} \in C_{-m}} \pi_m(A^1_m, A_{-m})\bar{x}_{-m}(A_{-m}).$$

Thus, it follows that

$$\sum_{A_{-m} \in C_{-m}} \pi_m(\hat{A}^0_m, A_{-m})\bar{x}_{-m}(A_{-m}) \quad = \quad \sum_{A_{-m} \in C_{-m}} \pi_m(\hat{A}^1_m, A_{-m})\bar{x}_{-m}(A_{-m}),$$

which is equivalent to

$$\begin{aligned}
&\sum_{A^0_{-m} \in C^0_{-m}} \left[\pi_m(\hat{A}^0_m, A^0_{-m}) - \pi_m(\hat{A}^1_m, A^0_{-m})\right] \bar{x}_{-m}(A^0_{-m}) \\
&+ \sum_{A^1_{-m} \in C^1_{-m}} \left[\pi_m(\hat{A}^0_m, A^1_{-m}) - \pi_m(\hat{A}^1_m, A^1_{-m})\right] \bar{x}_{-m}(A^1_{-m}) \quad = \quad 0,
\end{aligned}$$

that is,

$$\sum_{A^0_{-m} \in C^0_{-m}} \bar{\lambda}_m(A^0_{-m}, \hat{A}^0_m, \hat{A}^1_m) \left[\beta_m - \bar{\beta}_m(A^0_{-m}, \hat{A}^0_m, \hat{A}^1_m\right] \bar{x}_{-m}(A^0_{-m})$$

$$+ \sum_{A^1_{-m} \in C^1_{-m}} \underline{\lambda}_m(A^1_{-m}, \hat{A}^0_m, \hat{A}^1_m) \left[\beta_m - \underline{\beta}_m(A^1_{-m}, \hat{A}^0_m, \hat{A}^1_m)\right] \bar{x}_{-m}(A^1_{-m}) \quad = \quad 0,$$

which is equivalent to $\beta_m = \beta_m(\bar{x}_{-m})$. □

**Remark A.3.** *Note that the left side of (A.13) is the objective value*

$$\sum_{A_{-m} \in C_{-m}} \pi_m(A^0_m, A_{-m}) \bar{x}_{-m}(A_{-m}),$$

*and is by definition of* $\mathtt{PBR}^0_m(\bar{x}_{-m})$ *the same for all choices of* $A^0_m \in \mathtt{PBR}^0_m(\bar{x}_{-m})$*, and thus for all choices of* $A^0_m \in C^{0+}_m(\bar{x}^0_m) \subset \mathtt{PBR}^0_m(\bar{x}_{-m})$*. Similarly, the right side of (A.13) is the same for all choices of* $A^1_m \in \mathtt{PBR}^1_m(\bar{x}_{-m})$*. It follows that* $\beta_m(\bar{x}_{-m})$ *does not depend on the choice of* $A^0_m \in C^{0+}_m(\bar{x}^0_m)$ *and* $A^1_m \in \mathtt{PBR}^1_m(\bar{x}_{-m})$*, which is why* $(A^0_m, A^1_m)$ *was omitted from the argument of* $\beta_m(\bar{x}_{-m})$*. Similar comments apply to the proofs of Proposition 3.13((2)) and 3.13((3)).*

Corollary A.3 characterizes the existence of mixed-strategy Nash equilibria.

**Corollary A.3.** *The following holds:*

*(1) A pair of mixed strategies* $(\bar{x}^0_{-1}, \bar{x}^0_1) \in \Delta(C^0_{-1}) \times \Delta(C^0_1)$ *is a Nash equilibrium iff* $\bar{x}^0_m \in \mathtt{BR}^0_m(\bar{x}^0_{-m})$ *and* $\beta_m \geq \beta_m(\bar{x}^0_{-m})$ *for* $m = \pm 1$.

*(2) A pair of mixed strategies* $(\bar{x}^1_{-1}, \bar{x}^1_1) \in \Delta(C^1_{-1}) \times \Delta(C^1_1)$ *is a Nash equilibrium iff* $\bar{x}^1_m \in \mathtt{BR}^1_m(\bar{x}^1_{-m})$ *and* $\beta_m \leq \beta_m(\bar{x}^1_{-m})$ *for* $m = \pm 1$.

*(3) A pair of mixed strategies* $(\bar{x}^0_{-m}, \bar{x}^1_m) \in \Delta(C^0_{-m}) \times \Delta(C^1_m)$ *is a Nash equilibrium iff* $\bar{x}^0_{-m} \in \mathtt{BR}^0_{-m}(\bar{x}^1_m)$, $\bar{x}^1_m \in \mathtt{BR}^1_m(\bar{x}^0_{-m})$, $\beta_{-m} \geq \beta_{-m}(\bar{x}^1_m)$*, and* $\beta_m \leq \beta_m(\bar{x}^0_{-m})$.

*(4) There exists a pair of mixed strategies* $(\bar{x}^2_{-1}, \bar{x}^2_1) \in \Delta(C_{-1}) \times \Delta(C_1)$ *such that* $\bar{x}^2_m \in \mathtt{BR}^2_m(\bar{x}^2_{-m})$ *for* $m = \pm 1$ *iff* $\beta_m = \beta_m(\bar{x}^2_{-m})$ *for* $m = \pm 1$*. Any such pair of mixed strategies is a Nash equilibrium.*

*(5) There exists a pair of mixed strategies* $(\bar{x}^0_{-m}, \bar{x}^2_m) \in \Delta(C^0_{-m}) \times \Delta(C_m)$ *such that* $\bar{x}^2_m \in \mathtt{BR}^2_m(\bar{x}^0_{-m})$ *iff* $\beta_m = \beta_m(\bar{x}^0_{-m})$*. Such a pair of mixed strategies is a Nash equilibrium iff* $\bar{x}^0_{-m} \in \mathtt{BR}^0_{-m}(\bar{x}^2_m)$ *and* $\beta_{-m} \geq \beta_{-m}(\bar{x}^2_m)$.

*(6) There exists a pair of mixed strategies* $(\bar{x}^1_{-m}, \bar{x}^2_m) \in \Delta(C^1_{-m}) \times \Delta(C_m)$ *such that* $\bar{x}^2_m \in \mathtt{BR}^2_m(\bar{x}^1_{-m})$ *iff* $\beta_m = \beta_m(\bar{x}^1_{-m})$. *Such a pair of mixed strategies is a Nash equilibrium iff* $\bar{x}^1_{-m} \in \mathtt{BR}^1_{-m}(\bar{x}^2_m)$ *and* $\beta_{-m} \leq \beta_{-m}(\bar{x}^2_m)$.

# APPENDIX B

# THE GRADIENT AND HESSIAN EXPRESSIONS OF MNL, NL, ML AND LCL MODELS

In this appendix we give expressions for the calculation of $\nabla\hat{\mathcal{L}}^i(\theta^i)$ and $\nabla^2\hat{\mathcal{L}}^i(\theta^i)$ for each model $i \in \{\text{MNL, NL, ML, LCL}\}$. Let $j_n \in A_n$ denote the alternative chosen in observation $n$, and let $x_{n,j}$ denote the attribute vector for observation $n$ and alternative $j \in A_n$.

## *B.1 The Gradient and Hessian for the MNL Model*

Let $\bar{m}$ denote the number of considered attributes. Thus, $x_{n,j} := (x_{n,j,1}, x_{n,j,2}, \dots, x_{n,j,\bar{m}}) \in \mathbb{R}^{\bar{m}}$ denotes the vector of attribute values for observation $n$ and alternative $j \in A_n$, and $\beta := (\beta_1, \beta_2, \dots, \beta_{\bar{m}}) \in \mathbb{R}^{\bar{m}}$ denotes the vector of parameter values (thus, for the MNL model, $\theta = \beta$). Then the gradient of the sample average log-likelihood function for the MNL model can be written as

$$\nabla\hat{\mathcal{L}}(\beta) = \frac{1}{N}\sum_{n=1}^{N}\left(x_{n,j_n} - \frac{\sum_{j\in A_n}\exp(\beta^\intercal x_{n,j})x_{n,j}}{\sum_{j\in A_n}\exp(\beta^\intercal x_{n,j})}\right).$$

Hence, if model $i$ is the MNL model, then

$$\text{Cov}(\hat{Z}^i) \approx \frac{1}{N(N-1)}\sum_{n=1}^{N}\left(x_{n,j_n} - \frac{\sum_{j\in A_n}\exp(\hat{\beta}^{i\intercal} x_{n,j})x_{n,j}}{\sum_{j\in A_n}\exp(\hat{\beta}^{i\intercal} x_{n,j})}\right)\left(x_{n,j_n} - \frac{\sum_{j\in A_n}\exp(\hat{\beta}^{i\intercal} x_{n,j})x_{n,j}}{\sum_{j\in A_n}\exp(\hat{\beta}^{i\intercal} x_{n,j})}\right)^\intercal$$

Also, for $k, m \in \{1, 2, \dots, \bar{m}\}$, the second partial derivative of $\hat{\mathcal{L}}$ with respect to $\beta_k$ and $\beta_m$ is given by

$$\frac{\partial^2\hat{\mathcal{L}}(\beta)}{\partial\beta_k\partial\beta_m} = \frac{1}{N}\sum_{n=1}^{N}\left[-\frac{\sum_{j\in A_n}\exp(\beta^\intercal x_{n,j})x_{n,j,k}x_{n,j,m}}{\sum_{j\in A_n}\exp(\beta^\intercal x_{n,j})} + \frac{\left(\sum_{j\in A_n}\exp(\beta^\intercal x_{n,j})x_{n,j,k}\right)\left(\sum_{j\in A_n}\exp(\beta^\intercal x_{n,j})x_{n,j,m}\right)}{\left(\sum_{j\in A_n}\exp(\beta^\intercal x_{n,j})\right)^2}\right].$$

Let

$$X_n := \begin{bmatrix} x_{n,1}^\intercal \\ x_{n,2}^\intercal \\ \vdots \\ x_{n,|A_n|}^\intercal \end{bmatrix} \in \mathbb{R}^{|A_n|\times\bar{m}}$$

denote the data matrix for observation $n$, let

$$q_{n,i} \quad := \quad \frac{\exp(\beta^\intercal x_{n,i})}{\sum_{j \in A_n} \exp(\beta^\intercal x_{n,j})}$$

denote the choice probability for alternative $i \in A_n$, and let $\mathbf{q}_n := (q_{n,j}, j \in A_n) \in \mathbb{R}^{|A_n|}$ denote the vector of choice probabilities. Let $\operatorname{diag}\{\mathbf{q}_n\} \in \mathbb{R}^{|A_n| \times |A_n|}$ denote the diagonal matrix with $\operatorname{diag}\{\mathbf{q}_n\}_{j,j} = q_{n,j}$. Then one can write

$$\begin{aligned}
\nabla \hat{\mathcal{L}}(\beta) &= \frac{1}{N} \sum_{n=1}^{N} \left( x_{n,j_n} - X_n^\intercal \mathbf{q}_n \right), \\
\nabla^2 \hat{\mathcal{L}}(\beta) &= \frac{1}{N} \sum_{n=1}^{N} \left( X_n^\intercal \mathbf{q}_n \mathbf{q}_n^\intercal X_n - X_n^\intercal \operatorname{diag}\{\mathbf{q}_n\} X_n \right).
\end{aligned}$$

## *B.2 The Gradient and Hessian for the NL Model*

Let $A_n = \bigcup_{l=1}^{L} A_{n,l}$ be the partition of $A_n$, and $x_{n,l,j} \in \mathbb{R}^{\bar{m}}$ denote the vector of attribute values for alternative $j \in A_{n,l}$. According to the nested logit model, the probability that customer $n \in \mathcal{N}$ chooses alternative $j \in A_{n,l}$ is given by

$$q_{n,j}(A_n) \quad = \quad q_{j|n,l} q_{l|n},$$

where

$$\begin{aligned}
q_{j|n,l} &:= \frac{\exp(\beta^\intercal x_{n,l,j}/\alpha_l)}{\sum_{i \in A_{n,l}} \beta^\intercal x_{n,l,i}/\alpha_l)}, \\
q_{l|n} &:= \frac{\exp\left(\alpha \alpha_l \bar{v}_{n,l}\right)}{\sum_{l'=1}^{L} \exp\left(\alpha \alpha_{l'} \bar{v}_{n,l'}\right)}, \\
\bar{v}_{n,l} &:= \ln\left( \sum_{j \in A_{n,l}} \exp(\beta^\intercal x_{n,l,j}/\alpha_l) \right),
\end{aligned}$$

and $\alpha > 0$ is an arbitrary scaling factor, and $\alpha_l \in (0, 1/\alpha]$ is a parameter that can be thought of as representing the dissimilarity of alternatives in subset (nest) $l$. Let $l_n \in \{1, 2, \ldots, L\}$ denote the nest that contains $j_n$, and let $\theta := (\alpha_1, \alpha_2, \ldots, \alpha_L, \beta) \in \mathbb{R}^{L+\bar{m}}$ be the vector of parameters to be estimated. Then

$$\hat{\mathcal{L}}_n(\theta) \quad = \quad \ln\left(q_{j_n|n,l_n} q_{l_n|n}\right).$$

Let

$$
\begin{aligned}
X_{n,l} &:= \begin{bmatrix} x_{n,l,1}^{\intercal} \\ x_{n,l,2}^{\intercal} \\ \vdots \\ x_{n,l,|A_{n,l}|}^{\intercal} \end{bmatrix} \quad \in \quad \mathbb{R}^{|A_{n,l}| \times \bar{m}} \\
X_n &:= \begin{bmatrix} X_{n,1} \\ X_{n,2} \\ \vdots \\ X_{n,L} \end{bmatrix} \quad \in \quad \mathbb{R}^{|A_n| \times \bar{m}} \\
\mathbf{q}_{n,l} &:= \left(q_{j|n,l}, j \in A_{n,l}\right) \quad \in \quad \mathbb{R}^{|A_{n,l}|} \\
\mathbf{q}_n &:= \left(\mathbf{q}_{n,l}, l \in \{1, 2, \ldots, L\}\right) \quad \in \quad \mathbb{R}^{|A_n|} \\
J_{n,l} &:= \frac{1}{\alpha_l}\left(\operatorname{diag}\{\mathbf{q}_{n,l}\} X_{n,l} - \mathbf{q}_{n,l}\mathbf{q}_{n,l}^{\intercal} X_{n,l}\right) \quad \in \quad \mathbb{R}^{|A_{n,l}| \times \bar{m}} \\
\beta_l &:= \beta/\alpha_l \quad \in \quad \mathbb{R}^{\bar{m}}.
\end{aligned}
$$

Then, $\nabla \hat{\mathcal{L}}_n(\theta)$ is given by

$$
\begin{aligned}
\frac{\partial \hat{\mathcal{L}}_n(\theta)}{\partial \alpha_l} &= \begin{cases} \alpha(1 - q_{l|n})\left(\bar{v}_{n,l} - \mathbf{q}_{n,l}^{\intercal} X_{n,l}\beta_l\right) + \frac{1}{\alpha_l}\left(\mathbf{q}_{n,l}^{\intercal} X_{n,l}\beta_l - x_{n,l,j_n}^{\intercal}\beta_l\right) & \text{if} \quad l = l_n \\ -\alpha\left(\bar{v}_{n,l} - \mathbf{q}_{n,l}^{\intercal} X_{n,l}\beta_l\right) q_{l|n} & \text{if} \quad l \neq l_n \end{cases} \\
\nabla_\beta \hat{\mathcal{L}}_n(\theta) &= \frac{1}{\alpha_{l_n}}\left(x_{n,l_n,j_n} - X_{n,l_n}^{\intercal}\mathbf{q}_{n,l_n}\right) + \alpha X_{n,l_n}^{\intercal}\mathbf{q}_{n,l_n} - \alpha X_n^{\intercal}\mathbf{q}_n.
\end{aligned}
$$

Also, $\nabla^2 \hat{\mathcal{L}}_n(\theta)$ is given by the following expressions: For $i \neq l$,

$$
\frac{\partial^2 \hat{\mathcal{L}}_n(\theta)}{\partial \alpha_l \partial \alpha_i} = \alpha^2\left(\bar{v}_{n,l} - \mathbf{q}_{n,l}^{\intercal} X_{n,l}\beta_l\right)\left(\bar{v}_{n,i} - \mathbf{q}_{n,i}^{\intercal} X_{n,i}\beta_i\right) q_{l|n} q_{i|n}
$$

For $l \neq l_n$,

$$
\frac{\partial^2 \hat{\mathcal{L}}_n(\theta)}{\partial \alpha_l^2} = -\alpha\beta_l^{\intercal} J_{n,l}^{\intercal} X_{n,l}\beta_l q_{l|n} - \alpha^2\left(\bar{v}_{n,l} - \mathbf{q}_{n,l}^{\intercal} X_{n,l}\beta_l\right)^2 q_{l|n}(1 - q_{l|n})
$$

For $l = l_n$,

$$
\begin{aligned}
\frac{\partial^2 \hat{\mathcal{L}}_n(\theta)}{\partial \alpha_l^2} &= \left(\alpha(1 - q_{l|n}) - \frac{1}{\alpha_l}\right)\beta_l^{\intercal} J_{n,l}^{\intercal} X_{n,l}\beta_l - \alpha^2\left(\bar{v}_{n,l} - \mathbf{q}_{n,l}^{\intercal} X_{n,l}\beta_l\right)^2 q_{l|n}(1 - q_{l|n}) \\
&\quad - \frac{2}{\alpha_l^2}\left(\mathbf{q}_{n,l}^{\intercal} X_{n,l}\beta_l - x_{n,l,j_n}^{\intercal}\beta_l\right)
\end{aligned}
$$

For $l \neq l_n$,

$$\nabla^2_{\alpha_l,\beta}\hat{\mathcal{L}}_n(\theta) \quad = \quad \alpha q_{l|n}\left[J_{n,l}^\intercal X_{n,l}\beta_l - \alpha\left(\bar{v}_{n,l} - \mathbf{q}_{n,l}^\intercal X_{n,l}\beta_l\right)\left(X_{n,l}^\intercal \mathbf{q}_{n,l} - X_n^\intercal \mathbf{q}_n\right)\right]$$

For $l = l_n$,

$$\begin{aligned}\nabla^2_{\alpha_l,\beta}\hat{\mathcal{L}}_n(\theta) &= \alpha q_{l|n}\left[J_{n,l}^\intercal X_{n,l}\beta_l - \alpha\left(\bar{v}_{n,l} - \mathbf{q}_{n,l}^\intercal X_{n,l}\beta_l\right)\left(X_{n,l}^\intercal \mathbf{q}_{n,l} - X_n^\intercal \mathbf{q}_n\right)\right] - \alpha J_{n,l}^\intercal X_{n,l}\beta_l \\ &+ \frac{1}{\alpha_l}\left[\frac{X_{n,l}^\intercal \mathbf{q}_{n,l} - x_{n,l,j_n}}{\alpha_l} + J_{n,l}^\intercal X_{n,l}\beta_l\right]\end{aligned}$$

Also,

$$\nabla^2_{\beta}\hat{\mathcal{L}}_n(\theta) \quad = \quad \frac{\alpha\alpha_{l_n} - 1}{\alpha_{l_n}} X_{n,l_n}^\intercal J_{n,l_n} - \alpha\sum_{l=1}^{L} q_{l|n} X_{n,l}^\intercal\left[\alpha\mathbf{q}_{n,l}\left(\mathbf{q}_{n,l}^\intercal X_{n,l} - \mathbf{q}_n^\intercal X_n\right) + J_{n,l}\right].$$

### B.3 The Gradient and Hessian for the ML Model

Let $\beta \in \mathbb{R}^{m_1}$ denote the (deterministic) parameters that are the same across the customer population, and let $x_{n,j}$ denote the corresponding vector of attribute values for observation $n$ and alternative $j \in A_n$. In addition to $\beta$, there are also (random) parameters with values that vary across the customer population. Let $\gamma_n \in \mathbb{R}^{m_2}$ denote the random vector of parameter values for observation $n$. We do not estimate a value for $\gamma_n$ for each observation $n$; instead, we estimate a distribution for $\gamma_n$ over the customer population. Here we consider the Gaussian mixture model, that is, we assume that $\{\gamma_n\}_n$ are i.i.d. normally distributed random vectors with mean $\mu \in \mathbb{R}^r$ and covariance matrix $\Sigma$. Let $\sigma$ denote the lower-triangular Cholesky factor of $\Sigma$, and let $\xi \in \mathbb{R}^{m_2}$ denote a random vector of $m_2$ i.i.d. standard normally distributed components. Let $y_{n,j} \in \mathbb{R}^{m_2}$ denote the corresponding vector of attribute values for observation $n$ and alternative $j \in A_n$. Then, the systematic utility of alternative $j \in A_n$ for observation $n$ has the same distribution as

$$v_{n,j} \quad := \quad \theta^\intercal \zeta_{n,j} \quad := \quad \beta^\intercal x_{n,j} + \mu^\intercal y_{n,j} + \xi^\intercal \sigma^\intercal y_{n,j},$$

where

$$\begin{aligned}\zeta_{n,i,j} \quad := \quad &(x_{n,j}, y_{n,j}, \xi_{n,i,1}y_{n,j,1}, \xi_{n,i,1}y_{n,j,2}, \ldots, \\ &\xi_{n,i,1}y_{n,j,r}, \xi_{n,i,2}y_{n,j,2}, \ldots, \xi_{n,i,2}y_{n,j,r}, \ldots, \xi_{n,i,r}y_{n,j,r}) \in \mathbb{R}^{\bar{m}},\end{aligned}$$

$$\theta \quad := \quad (\alpha, \mu, \sigma_{1,1}, \sigma_{2,1}, \ldots, \sigma_{r,1}, \sigma_{2,2}, \ldots, \sigma_{r,2}, \ldots, \sigma_{r,r}) \in \mathbb{R}^{\bar{m}}.$$

Thus, the log-likelihood function for observation $n$ is given by

$$\hat{\mathcal{L}}_n(\theta) \quad := \quad \ln \left( \int_{\mathbb{R}^{m_2}} \frac{\exp(\beta^\intercal x_{n,j_n} + \mu^\intercal y_{n,j_n} + \xi^\intercal \sigma^\intercal y_{n,j_n})}{\sum_{j \in A_n} \exp(\beta^\intercal x_{n,j} + \mu^\intercal y_{n,j} + \xi^\intercal \sigma^\intercal y_{n,j})} \, d\Phi(\xi) \right)$$

where $\Phi$ denotes the standard normal distribution on $\mathbb{R}^{m_2}$. Computing $\hat{\mathcal{L}}_n(\theta)$ accurately is hard because it involves calculation of the $r$-dimensional integral above (In our model, $m_2 = 14$).

For each observation $n$, let $|I_n|$ denote the chosen sample size for the Monte Carlo approximation of the integral, and let $\{\xi_{n,i} := (\xi_{n,i,1}, \xi_{n,i,2}, \ldots, \xi_{n,i,m_2}), i = 1, \ldots, |I_n|\}$ denote the corresponding sample of i.i.d. standard normally distributed vectors.

Then, the simulated log-likelihood function for observation $n$ is given by

$$\begin{aligned}
\hat{\hat{\mathcal{L}}}_n(\theta) \quad &:= \quad \ln \left( \frac{1}{|I_n|} \sum_{i=1}^{|I_n|} \frac{\exp(\beta^\intercal x_{n,j_n} + \mu^\intercal y_{n,j_n} + \xi_{n,i}^\intercal \sigma^\intercal y_{n,j_n})}{\sum_{j \in A_n} \exp(\beta^\intercal x_{n,j} + \mu^\intercal y_{n,j} + \xi_{n,i}^\intercal \sigma^\intercal y_{n,j})} \right) \\
&= \quad \ln \left( \frac{1}{|I_n|} \sum_{i=1}^{|I_n|} \frac{\exp(\theta^\intercal \zeta_{n,i,j_n})}{\sum_{j \in A_n} \exp(\theta^\intercal \zeta_{n,i,j})} \right)
\end{aligned}$$

Let

$$\begin{aligned}
q_{n,i,j} \quad &:= \quad \frac{\exp(\theta^\intercal \zeta_{n,i,j})}{\sum_{j' \in A_n} \exp(\theta^\intercal \zeta_{n,i,j'})} \quad \in \quad \mathbb{R} \\
\mathbf{q}_{n,i} \quad &:= \quad (q_{n,i,j}, \ j \in A_n) \quad \in \quad \mathbb{R}^{|A_n|} \\
\tilde{q}_{n,i,j} \quad &:= \quad \frac{q_{n,i,j}}{\sum_{i'=1}^{I_n} q_{n,i',j}} \quad \in \quad \mathbb{R} \\
\tilde{\mathbf{q}}_{n,j} \quad &:= \quad (\tilde{q}_{n,i,j}, i \in \{1, 2, \ldots, I_n\}) \quad \in \quad \mathbb{R}^{|I_n|} \\
\tilde{W}_{n,j} \quad &:= \quad \begin{bmatrix} \zeta_{n,1,j}^\intercal \\ \zeta_{n,2,j}^\intercal \\ \vdots \\ \zeta_{n,I_n,j}^\intercal \end{bmatrix} \quad \in \quad \mathbb{R}^{|I_n| \times \bar{m}} \\
W_{n,i} \quad &:= \quad \begin{bmatrix} \zeta_{n,i,1}^\intercal \\ \zeta_{n,i,2}^\intercal \\ \vdots \\ \zeta_{n,i,|A_n|}^\intercal \end{bmatrix} \quad \in \quad \mathbb{R}^{|A_n| \times \bar{m}}
\end{aligned}$$

$$
\begin{aligned}
W_n &:= \begin{bmatrix} W_{n,1} \\ W_{n,2} \\ \vdots \\ W_{n,|I_n|} \end{bmatrix} \in \mathbb{R}^{|I_n||A_n| \times \bar{m}} \\
\text{diag}\{\mathbf{q}_{n,1}, \ldots, \mathbf{q}_{n,|I_n|}\} &:= \begin{bmatrix} \mathbf{q}_{n,1} & 0 & \cdots & 0 \\ 0 & \mathbf{q}_{n,2} & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \mathbf{q}_{n,|I_n|} \end{bmatrix} \in \mathbb{R}^{I_n|A_n| \times |I_n|} \\
\text{diag}\{\mathbf{q}^\intercal_{n,1}, \ldots, \mathbf{q}^\intercal_{n,|I_n|}\} &:= \begin{bmatrix} \mathbf{q}^\intercal_{n,1} & 0 & \cdots & 0 \\ 0 & \mathbf{q}^\intercal_{n,2} & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \mathbf{q}^\intercal_{n,|I_n|} \end{bmatrix} \in \mathbb{R}^{|I_n| \times |I_n||A_n|} \\
J_{n,j} &:= \left(\text{diag}\{\tilde{\mathbf{q}}_{n,j}\} - \tilde{\mathbf{q}}_{n,j}\tilde{\mathbf{q}}^\intercal_{n,j}\right)\left(\tilde{W}_{n,j} - \text{diag}\{\mathbf{q}^\intercal_{n,1}, \ldots, \mathbf{q}^\intercal_{n,I_n}\}W_n\right) \\
&\in \mathbb{R}^{|I_n| \times \bar{m}}.
\end{aligned}
$$

Then, the gradient $\nabla\hat{\hat{\mathcal{L}}}_n(\theta)$ and Hessian $\nabla^2\hat{\hat{\mathcal{L}}}_n(\theta)$ for each observation $n$ are given by

$$
\begin{aligned}
\nabla\hat{\hat{\mathcal{L}}}_n(\theta) &= \left(\tilde{W}^\intercal_{n,j_n} - W^\intercal_n \text{diag}\{\mathbf{q}_{n,1}, \ldots, \mathbf{q}_{n,|I_n|}\}\right)\tilde{\mathbf{q}}_{n,j_n} \\
\nabla^2\hat{\hat{\mathcal{L}}}_n(\theta) &= -\sum_{i=1}^{|I_n|} \tilde{q}_{n,i,j_n} W^\intercal_{n,i}\left(\text{diag}\{\mathbf{q}_{n,i}\} - \mathbf{q}_{n,i}\mathbf{q}^\intercal_{n,i}\right)W_{n,i} \\
&+ \left(\tilde{W}^\intercal_{n,j_n} - W^\intercal_n \text{diag}\{\mathbf{q}_{n,1}, \ldots, \mathbf{q}_{n,|I_n|}\}\right)J_{n,j_n}.
\end{aligned}
$$

## *B.4 The Gradient and Hessian for the LCL Model*

As before, let $x_{n,j} := (x_{n,j,1}, x_{n,j,2}, \ldots, x_{n,j,\bar{m}}) \in \mathbb{R}^{\bar{m}}$ denote the vector of attribute values for observation $n$ and alternative $j \in A_n$, and let $\beta := (\beta_1, \beta_2, \ldots, \beta_{\bar{m}}) \in \mathbb{R}^{\bar{m}}$ denote the corresponding vector of coefficients. One consideration set $\hat{C} \in \mathcal{C}$ is chosen to be the base consideration set, such that $\pi(\hat{C}) = 1 - \sum_{C \in \mathcal{C} \setminus \{\hat{C}\}} \pi(C)$. Let $\pi := (\pi(C), C \in \mathcal{C} \setminus \{\hat{C}\}) \in \mathbb{R}^{\bar{c}}$ denote the vector of consideration set probabilities to be estimated. Then $\theta := (\pi, \beta) \in$

$\mathbb{R}^{\bar{c}+\bar{m}}$ denotes the vector of parameters to be estimated. Let

$$
\begin{aligned}
X_n &:= \begin{bmatrix} x_{n,1}^\intercal \\ x_{n,2}^\intercal \\ \vdots \\ x_{n,|A_n|}^\intercal \end{bmatrix} \in \mathbb{R}^{|A_n| \times \bar{m}} \\
e &:= (1, 1, \ldots, 1) \in \mathbb{R}^{\bar{c}} \\
\mathbf{1}_n &:= \left(\mathbf{1}_{[C \cap A_n \neq \varnothing]}, C \in \mathcal{C} \setminus \{\hat{C}\}\right) \in \mathbb{R}^{\bar{c}} \\
q_{n,C,j} &:= \frac{\mathbf{1}_{[j \in C]} \exp\left(\beta^\intercal x_{n,j}\right)}{\sum_{j' \in C \cap A_n} \exp(\beta^\intercal x_{n,j'})} \in \mathbb{R} \\
\tilde{\pi}_n(C) &:= \frac{\mathbf{1}_{[C \cap A_n \neq \varnothing]} \pi(C)}{\mathbf{1}_n^\intercal \pi + (1 - e^\intercal \pi) \mathbf{1}_{[\hat{C} \cap A_n \neq \varnothing]}} \in \mathbb{R} \\
\tilde{\pi}_n &:= \left(\tilde{\pi}_n(C), C \in \mathcal{C} \setminus \{\hat{C}\}\right) \in \mathbb{R}^{\bar{c}} \\
\mathbf{q}_{n,C} &:= (q_{n,C,j}, j \in A_n) \in \mathbb{R}^{|A_n|} \\
\tilde{\mathbf{q}}_{n,j} &:= \left(q_{n,C,j}, C \in \mathcal{C} \setminus \{\hat{C}\}\right) \in \mathbb{R}^{\bar{c}} \\
Q_n &:= [\tilde{\mathbf{q}}_{n,1}, \tilde{\mathbf{q}}_{n,2}, \ldots, \tilde{\mathbf{q}}_{n,|A_n|}] \in \mathbb{R}^{\bar{c} \times |A_n|} \\
J_{n,C} &:= (\mathrm{diag}\{\mathbf{q}_{n,C}\} - \mathbf{q}_{n,C} \mathbf{q}_{n,C}^\intercal) X_n \in \mathbb{R}^{|A_n| \times \bar{m}} \\
\tilde{J}_{n,j} &:= \tilde{\mathbf{q}}_{n,j} x_{n,j}^\intercal - \mathrm{diag}\{\tilde{\mathbf{q}}_{n,j}\} Q_n X_n \in \mathbb{R}^{\bar{c} \times \bar{m}}
\end{aligned}
$$

Then,

$$
\hat{\mathcal{L}}_n(\theta) = \ln\left(\tilde{\pi}_n^\intercal \tilde{\mathbf{q}}_{n,j_n} + (1 - e^\intercal \tilde{\pi}_n) q_{n,\hat{C},j_n}\right).
$$

Let

$$
\begin{aligned}
a_n &:= \mathbf{1}_n^\intercal \pi + (1 - e^\intercal \pi) \mathbf{1}_{[\hat{C} \cap A_n \neq \varnothing]} \in \mathbb{R} \\
b_n &:= \tilde{\pi}_n^\intercal \tilde{\mathbf{q}}_{n,j_n} + (1 - e^\intercal \tilde{\pi}_n) q_{n,\hat{C},j_n} \in \mathbb{R} \\
c_n &:= \mathbf{1}_n - e \mathbf{1}_{[\hat{C} \cap A_n \neq \varnothing]} \in \mathbb{R}^{\bar{c}} \\
\bar{q}_n &:= \tilde{\mathbf{q}}_{n,j_n} - e q_{n,\hat{C},j_n} \in \mathbb{R}^{\bar{c}} \\
d_n &:= \tilde{J}_{n,j_n}^\intercal \tilde{\pi}_n + (1 - e^\intercal \tilde{\pi}_n) q_{n,\hat{C},j_n} (x_{n,j_n} - X_n^\intercal \mathbf{q}_{n,\hat{C}}) \in \mathbb{R}^{\bar{m}} \\
J_n &:= \frac{1}{a_n} \mathrm{diag}\{\mathbf{1}_n\} - \frac{1}{a_n^2} \mathrm{diag}\{\mathbf{1}_n\} \pi c_n^\intercal \in \mathbb{R}^{\bar{c} \times \bar{c}}.
\end{aligned}
$$

Then, the gradient $\nabla\hat{\mathcal{L}}_n(\theta)$ for each observation $n$ is given by

$$\begin{aligned}
\nabla_\pi \hat{\mathcal{L}}_n(\theta) &= \frac{\frac{1}{a_n}\text{diag}\{\mathbf{1}_n\}\bar{q}_n - \frac{1}{a_n^2}c_n\pi^\intercal\text{diag}\{\mathbf{1}_n\}\bar{q}_n}{b_n} \quad \in \quad \mathbb{R}^{\bar{c}} \\
\nabla_\beta \hat{\mathcal{L}}_n(\theta) &= \frac{1}{b_n}d_n \quad \in \quad \mathbb{R}^{\bar{m}}
\end{aligned}$$

and the Hessian $\nabla^2\hat{\mathcal{L}}_n(\theta)$ for each observation $n$ is given by

$$\begin{aligned}
\nabla_\pi^2 \hat{\mathcal{L}}_n(\theta) &= \frac{-\frac{1}{a_n^2}\text{diag}\{\mathbf{1}_n\}\bar{q}_n c_n^\intercal - \frac{1}{a_n^2}c_n\bar{q}_n^\intercal\text{diag}\{\mathbf{1}_n\} + \frac{2}{a_n^3}c_n c_n^\intercal\pi^\intercal\text{diag}\{\mathbf{1}_n\}\bar{q}_n}{b_n} - \frac{J_n^\intercal\bar{q}_n\bar{q}_n^\intercal J_n}{b_n^2} \\
&\in \mathbb{R}^{\bar{c}\times\bar{c}} \\
\nabla_{\pi,\beta} \hat{\mathcal{L}}_n(\theta) &= J_n^\intercal\left(\frac{\tilde{J}_{n,j_n} - eq_{n,\hat{C},j_n}(x_{n,j_n}^\intercal - \mathbf{q}_{n,\hat{C}}^\intercal X_n)}{b_n} - \frac{\bar{q}_n d_n^\intercal}{b_n^2}\right) \\
&\in \mathbb{R}^{\bar{c}\times\bar{m}} \\
\nabla_\beta^2 \hat{\mathcal{L}}_n(\theta) &= \frac{\sum_{C\in\mathcal{C}\setminus\{\hat{C}\}}\tilde{\pi}(C)q_{n,C,j_n}\left[(x_{n,j_n} - X_n^\intercal\mathbf{q}_{n,C})(x_{n,j_n}^\intercal - \mathbf{q}_{n,C}^\intercal X_n) - X_n^\intercal J_{n,C}\right]}{\tilde{\pi}^\intercal\tilde{\mathbf{q}}_{n,j_n} + (1-e^\intercal\tilde{\pi})q_{n,\hat{C},j_n}} \\
&+ \frac{(1-e^\intercal\tilde{\pi})q_{n,\hat{C},j_n}\left[(x_{n,j_n} - X_n^\intercal\mathbf{q}_{n,\hat{C}})(x_{n,j_n}^\intercal - \mathbf{q}_{n,\hat{C}}^\intercal X_n) - X_n^\intercal J_{n,\hat{C}}\right]}{\tilde{\pi}^\intercal\tilde{\mathbf{q}}_{n,j_n} + (1-e^\intercal\tilde{\pi})q_{n,\hat{C},j_n}} \\
&- \frac{d_n d_n^\intercal}{\left(\tilde{\pi}^\intercal\tilde{\mathbf{q}}_{n,j_n} + (1-e^\intercal\tilde{\pi})q_{n,\hat{C},j_n}\right)^2} \\
&\in \quad \mathbb{R}^{\bar{m}\times\bar{m}}
\end{aligned}$$

# APPENDIX C

# ESTIMATED COEFFICIENTS, $T$-STATISTICS AND $P$-VALUES FOR THE ML MODEL WITH THE 2011 AIRLINE DATA USING THE STRA

Table 16: Estimated coefficients, $t$-statistics and $p$-values for the ML model with 2011 airline data using the STRA.

| ML (2011 Data) | | | | |
|---|---|---|---|---|
| **Attribute** | $\widehat{\theta}_{\text{STRA}}$ | **$t$-Statistic** | **$p$-Value** | $\widehat{\theta}_{\text{SAA-50}}$ |
| 1,1,1,1 | -5.04610 | -101.15000 | 0.00000 | -4.91710 |
| 1,1,1,2 | -12.05400 | -32.32300 | 0.00000 | -10.37200 |
| 1,1,1,3 | -7.51530 | -7.15020 | 0.00000 | -6.38970 |
| 1,1,1,4 | -5.59620 | -80.65100 | 0.00000 | -5.34490 |
| 1,1,1,5 | -6.51300 | -28.88800 | 0.00000 | -6.06150 |
| 1,1,2,1 | -7.28580 | -73.54800 | 0.00000 | -6.87900 |
| 1,1,2,2 | -16.82700 | -33.37000 | 0.00000 | -14.08100 |
| 1,1,2,3 | -6.61210 | -3.35920 | 0.00078 | -5.43440 |
| 1,1,2,4 | -7.24450 | -62.21400 | 0.00000 | -6.79410 |
| 1,1,2,5 | -8.83990 | -25.34900 | 0.00000 | -8.15600 |
| 1,2,1,1 | -5.30580 | -189.87000 | 0.00000 | -4.89050 |
| 1,2,1,2 | -10.33600 | -55.33700 | 0.00000 | -9.15270 |
| 1,2,1,3 | -5.76640 | -10.28400 | 0.00000 | -5.51300 |
| 1,2,1,4 | -5.84380 | -164.43000 | 0.00000 | -5.36190 |
| 1,2,1,5 | -6.58450 | -65.53600 | 0.00000 | -5.96090 |
| 1,2,2,1 | -6.53490 | -169.37000 | 0.00000 | -6.13450 |
| 1,2,2,2 | -10.05800 | -42.76100 | 0.00000 | -8.99900 |
| 1,2,2,3 | -6.37460 | -8.81230 | 0.00000 | -5.64960 |
| 1,2,2,4 | -6.69640 | -137.78000 | 0.00000 | -6.31370 |
| 1,2,2,5 | -7.40750 | -50.37100 | 0.00000 | -6.85370 |
| 1,3,1,1 | -6.90210 | -179.70000 | 0.00000 | -6.44530 |
| 1,3,1,2 | -11.64400 | -49.48200 | 0.00000 | -10.14300 |
| 1,3,1,3 | -7.95940 | -9.97140 | 0.00000 | -7.18070 |
| 1,3,1,4 | -7.16480 | -130.70000 | 0.00000 | -6.61090 |
| 1,3,1,5 | -8.68420 | -55.07100 | 0.00000 | -7.93530 |
| 1,3,2,1 | -6.39880 | -125.06000 | 0.00000 | -6.00710 |
| 1,3,2,2 | -10.92200 | -38.67100 | 0.00000 | -9.91550 |
| 1,3,2,3 | -4.70570 | -4.11930 | 0.00004 | -4.70470 |
| 1,3,2,4 | -6.97150 | -118.46000 | 0.00000 | -6.54090 |
| 1,3,2,5 | -8.05710 | -45.19600 | 0.00000 | -7.47470 |
| 2,1,1,1 | -7.89300 | -65.13700 | 0.00000 | -7.35560 |

Table 16 (continued.)

| ML (2011 Data) | | | | |
|---|---|---|---|---|
| **Attribute** | $\widehat{\theta}_{\text{STRA}}$ | ***t*-Statistic** | ***p*-Value** | $\widehat{\theta}_{\text{SAA-50}}$ |
| 2,1,1,2 | -11.96100 | -20.25700 | 0.00000 | -10.06000 |
| 2,1,1,3 | -5.10180 | -1.70460 | 0.08827 | -3.15190 |
| 2,1,1,4 | -9.53740 | -57.56900 | 0.00000 | -8.80450 |
| 2,1,1,5 | -8.60500 | -19.95100 | 0.00000 | -8.22040 |
| 2,1,2,1 | -8.99200 | -35.93800 | 0.00000 | -8.34670 |
| 2,1,2,2 | -10.33200 | -12.93200 | 0.00000 | -9.73490 |
| 2,1,2,3 | -7.17790 | -4.29650 | 0.00002 | -6.38190 |
| 2,1,2,4 | -10.07500 | -36.49800 | 0.00000 | -9.04580 |
| 2,1,2,5 | -11.59100 | -12.73000 | 0.00000 | -10.53300 |
| 2,2,1,1 | -7.82070 | -229.80000 | 0.00000 | -7.23140 |
| 2,2,1,2 | -10.82200 | -52.44700 | 0.00000 | -9.20950 |
| 2,2,1,3 | -5.89570 | -8.78380 | 0.00000 | -4.90130 |
| 2,2,1,4 | -8.08310 | -166.32000 | 0.00000 | -7.41320 |
| 2,2,1,5 | -8.32270 | -54.45600 | 0.00000 | -7.65940 |
| 2,2,2,1 | -9.03820 | -127.72000 | 0.00000 | -8.47730 |
| 2,2,2,2 | -12.01300 | -35.41900 | 0.00000 | -10.06800 |
| 2,2,2,3 | -10.57300 | -4.82770 | 0.00000 | -7.51070 |
| 2,2,2,4 | -9.12470 | -85.02900 | 0.00000 | -8.34350 |
| 2,2,2,5 | -9.64450 | -27.94500 | 0.00000 | -9.13460 |
| 2,3,1,1 | -9.04300 | -142.41000 | 0.00000 | -8.41070 |
| 2,3,1,2 | -11.45600 | -43.68600 | 0.00000 | -9.86020 |
| 2,3,1,3 | -6.51770 | -7.18180 | 0.00000 | -5.98490 |
| 2,3,1,4 | -9.39650 | -110.68000 | 0.00000 | -8.71750 |
| 2,3,1,5 | -9.42410 | -39.21900 | 0.00000 | -8.76610 |
| 2,3,2,1 | -10.59400 | -68.38500 | 0.00000 | -9.87610 |
| 2,3,2,2 | -11.02500 | -25.47100 | 0.00000 | -9.88470 |
| 2,3,2,3 | -5.15810 | -4.57270 | 0.00000 | -4.66020 |
| 2,3,2,4 | -9.59480 | -58.42800 | 0.00000 | -8.84860 |
| 2,3,2,5 | -11.35500 | -19.71800 | 0.00000 | -10.86800 |
| 3,1,1,1 | -8.45200 | -79.05600 | 0.00000 | -7.89940 |
| 3,1,1,2 | -11.88900 | -28.57200 | 0.00000 | -10.69800 |
| 3,1,1,3 | -12.51100 | -2.18410 | 0.02895 | -12.10400 |
| 3,1,1,4 | -10.13100 | -59.02100 | 0.00000 | -9.37710 |
| 3,1,1,5 | -9.19290 | -22.10700 | 0.00000 | -8.69540 |
| 3,1,2,1 | -8.72220 | -39.94400 | 0.00000 | -8.14310 |
| 3,1,2,2 | -13.80200 | -13.46500 | 0.00000 | -11.29400 |
| 3,1,2,3 | -13.80200 | -13.46500 | 0.00000 | -11.29400 |
| 3,1,2,4 | -10.05500 | -39.14000 | 0.00000 | -9.27770 |
| 3,1,2,5 | -13.22300 | -11.02100 | 0.00000 | -12.36700 |
| 3,2,1,1 | -8.44750 | -239.28000 | 0.00000 | -7.89080 |
| 3,2,1,2 | -12.22500 | -58.73700 | 0.00000 | -10.56300 |
| 3,2,1,3 | -4.31790 | -6.88140 | 0.00000 | -3.70140 |
| 3,2,1,4 | -8.99980 | -165.42000 | 0.00000 | -8.40380 |
| 3,2,1,5 | -9.14720 | -57.05800 | 0.00000 | -8.51180 |

Table 16 (continued.)

| ML (2011 Data) | | | | |
|---|---|---|---|---|
| **Attribute** | $\widehat{\theta}_{\text{STRA}}$ | $t$**-Statistic** | $p$**-Value** | $\widehat{\theta}_{\text{SAA-50}}$ |
| 3,2,2,1 | -9.66080 | -118.13000 | 0.00000 | -9.01980 |
| 3,2,2,2 | -14.23800 | -36.94100 | 0.00000 | -12.17000 |
| 3,2,2,3 | -5.87720 | -6.99370 | 0.00000 | -5.07600 |
| 3,2,2,4 | -9.99720 | -83.68200 | 0.00000 | -9.32350 |
| 3,2,2,5 | -11.32700 | -28.80700 | 0.00000 | -10.40400 |
| 3,3,1,1 | -9.35700 | -145.41000 | 0.00000 | -8.68930 |
| 3,3,1,2 | -13.31500 | -49.05000 | 0.00000 | -11.84900 |
| 3,3,1,3 | -6.61240 | -8.89460 | 0.00000 | -5.65650 |
| 3,3,1,4 | -10.50500 | -110.72000 | 0.00000 | -9.83110 |
| 3,3,1,5 | -10.83300 | -43.38800 | 0.00000 | -10.09400 |
| 3,3,2,1 | -10.91600 | -64.73000 | 0.00000 | -10.15100 |
| 3,3,2,2 | -14.13600 | -26.88400 | 0.00000 | -12.31300 |
| 3,3,2,3 | -8.94990 | -3.90960 | 0.00009 | -7.91510 |
| 3,3,2,4 | -10.36800 | -64.12300 | 0.00000 | -9.38890 |
| 3,3,2,5 | -12.39100 | -21.59300 | 0.00000 | -11.67200 |
| $[07:00, 08:00)$ popularity | -0.10638 | -1.32710 | 0.18448 | 1.13200 |
| $[08:00, 09:00)$ popularity | -0.45690 | -5.84300 | 0.00000 | 0.86406 |
| $[09:00, 10:00)$ popularity | -0.03298 | -0.40349 | 0.68659 | 1.04250 |
| $[10:00, 11:00)$ popularity | 0.00344 | 0.04152 | 0.96688 | 0.99750 |
| $[11:00, 12:00)$ popularity | 2.26890 | 55.09800 | 0.00000 | 0.83717 |
| $[12:00, 13:00)$ popularity | 2.31610 | 56.86700 | 0.00000 | 0.75148 |
| $[13:00, 14:00)$ popularity | 2.15540 | 47.60000 | 0.00000 | 0.56319 |
| $[14:00, 15:00)$ popularity | 2.72060 | 68.84300 | 0.00000 | 0.93265 |
| $[15:00, 16:00)$ popularity | 2.41880 | 44.62300 | 0.00000 | 1.05930 |
| $[16:00, 17:00)$ popularity | 2.77160 | 65.83100 | 0.00000 | 1.05450 |
| $[17:00, 18:00)$ popularity | 2.27500 | 44.84900 | 0.00000 | 0.94122 |
| $[18:00, 19:00)$ popularity | 1.06630 | 22.77600 | 0.00000 | 0.62031 |
| $[19:00, 20:00)$ popularity | 1.54000 | 30.39400 | 0.00000 | 0.86083 |
| $[20:00, 21:00)$ popularity | 0.59979 | 15.19100 | 0.00000 | 0.43594 |
| Carrier XX | 0.02678 | 2.66790 | 0.00763 | -0.10668 |
| Carrier YY | 0.63769 | 78.90300 | 0.00000 | 0.52444 |
| Change fee | -7.06560 | -160.22000 | 0.00000 | -6.69260 |
| Mileage gain | 0.73647 | 90.17500 | 0.00000 | 0.71424 |
| XX–1–1 is the most expensive | 3.16950 | 280.21000 | 0.00000 | 3.12600 |
| XX–1–2 is the most expensive | 3.62990 | 45.85600 | 0.00000 | 3.45040 |
| XX–1–3 is the most expensive | 3.45670 | 17.80900 | 0.00000 | 3.41130 |
| XX–1–4 is the most expensive | 2.32030 | 132.14000 | 0.00000 | 2.27200 |
| XX–1–5 is the most expensive | 2.38040 | 41.34000 | 0.00000 | 2.32380 |
| YY–1–1 is the most expensive | 1.58870 | 202.30000 | 0.00000 | 1.64790 |
| YY–1–4 is the most expensive | 0.88152 | 70.22300 | 0.00000 | 0.96136 |
| YY–1–5 is the most expensive | 1.01220 | 25.05300 | 0.00000 | 1.06840 |
| ZZ–1–1 is the most expensive | 1.39530 | 72.96400 | 0.00000 | 1.33620 |
| ZZ–1–4 is the most expensive | 1.04420 | 35.33700 | 0.00000 | 1.00820 |
| ZZ–1–5 is the most expensive | 1.50210 | 17.72600 | 0.00000 | 1.39090 |

Table 16 (continued.)

| ML (2011 Data) | | | | |
|---|---|---|---|---|
| **Attribute** | $\widehat{\theta}_{\text{STRA}}$ | ***t*-Statistic** | ***p*-Value** | $\widehat{\theta}_{\text{SAA-50}}$ |
| XX–13–1 is the cheapest | 2.85900 | 186.74764 | 0.00000 | 2.91800 |
| XX–13–2 is the cheapest | 2.94450 | 40.11362 | 0.00000 | 3.34000 |
| XX–13–3 is the cheapest | 4.16370 | 14.50578 | 0.00000 | 4.24980 |
| XX–13–4 is the cheapest | 2.98000 | 160.44877 | 0.00000 | 2.99640 |
| XX–13–5 is the cheapest | 3.11190 | 66.72659 | 0.00000 | 3.14470 |
| XX–12–1 is the cheapest | 2.72350 | 138.86143 | 0.00000 | 2.76330 |
| XX–12–2 is the cheapest | 1.69530 | 21.95855 | 0.00000 | 1.73560 |
| XX–12–3 is the cheapest | 2.09240 | 24.48850 | 0.00000 | 2.17850 |
| XX–12–4 is the cheapest | 2.04540 | 66.03556 | 0.00000 | 2.07920 |
| XX–12–5 is the cheapest | 2.09240 | 24.48850 | 0.00000 | 2.17850 |
| XX–11–1 is the cheapest | 2.91550 | 54.86497 | 0.00000 | 3.01490 |
| XX–11–2 is the cheapest | -0.02197 | -0.07424 | 0.94171 | 0.08856 |
| XX–11–3 is the cheapest | -0.02197 | -0.07424 | 0.94171 | 0.08856 |
| XX–11–4 is the cheapest | 0.79318 | 4.52877 | 0.00001 | 0.87579 |
| XX–11–5 is the cheapest | -0.02197 | -0.07424 | 0.94171 | 0.08856 |
| XX–10–1 is the cheapest | 1.70000 | 75.29299 | 0.00000 | 1.71860 |
| XX–10–2 is the cheapest | 1.69180 | 25.18523 | 0.00000 | 1.77740 |
| XX–10–3 is the cheapest | 1.43800 | 16.24762 | 0.00000 | 1.50110 |
| XX–10–4 is the cheapest | 1.87050 | 70.10175 | 0.00000 | 1.88120 |
| XX–10–5 is the cheapest | 1.43800 | 16.24762 | 0.00000 | 1.50110 |
| XX–9–1 is the cheapest | 1.48590 | 78.45685 | 0.00000 | 1.55890 |
| XX–9–2 is the cheapest | 1.88340 | 37.98344 | 0.00000 | 2.11200 |
| XX–9–3 is the cheapest | 1.37360 | 18.93176 | 0.00000 | 1.51420 |
| XX–9–4 is the cheapest | 1.53750 | 65.29314 | 0.00000 | 1.61190 |
| XX–9–5 is the cheapest | 1.37360 | 18.93176 | 0.00000 | 1.51420 |
| XX–8–1 is the cheapest | 1.72760 | 51.68969 | 0.00000 | 1.84110 |
| XX–8–2 is the cheapest | 2.19010 | 30.15375 | 0.00000 | 2.51040 |
| XX–8–3 is the cheapest | 1.39780 | 9.81710 | 0.00000 | 1.53490 |
| XX–8–4 is the cheapest | 1.71780 | 40.54765 | 0.00000 | 1.84330 |
| XX–8–5 is the cheapest | 1.39780 | 9.81710 | 0.00000 | 1.53490 |
| XX–7–1 is the cheapest | 1.58370 | 19.29154 | 0.00000 | 1.79400 |
| XX–7–2 is the cheapest | 1.87350 | 11.79308 | 0.00000 | 2.08290 |
| XX–7–3 is the cheapest | 1.87350 | 11.79308 | 0.00000 | 2.08290 |
| XX–7–4 is the cheapest | 1.75550 | 17.83526 | 0.00000 | 1.91580 |
| XX–7–5 is the cheapest | 1.87350 | 11.79308 | 0.00000 | 2.08290 |
| XX–6–1 is the cheapest | 1.98410 | 123.89878 | 0.00000 | 2.07770 |
| XX–6–2 is the cheapest | 2.57500 | 53.80845 | 0.00000 | 2.84910 |
| XX–6–3 is the cheapest | 2.30550 | 12.87816 | 0.00000 | 2.46920 |
| XX–6–4 is the cheapest | 1.91730 | 90.61544 | 0.00000 | 2.03310 |
| XX–6–5 is the cheapest | 1.32130 | 15.33387 | 0.00000 | 1.50760 |
| XX–5–1 is the cheapest | 1.81650 | 34.40839 | 0.00000 | 1.91520 |
| XX–5–2 is the cheapest | 2.36800 | 19.38292 | 0.00000 | 2.55200 |
| XX–5–3 is the cheapest | 1.45090 | 5.92795 | 0.00000 | 1.63340 |
| XX–5–4 is the cheapest | 1.59580 | 21.36461 | 0.00000 | 1.74400 |

Table 16 (continued.)

| **ML (2011 Data)** | | | | |
|---|---|---|---|---|
| **Attribute** | $\widehat{\theta}_{\text{STRA}}$ | $t$**-Statistic** | $p$**-Value** | $\widehat{\theta}_{\text{SAA-50}}$ |
| XX–5–5 is the cheapest | 1.45090 | 5.92795 | 0.00000 | 1.63340 |
| XX–4–1 is the cheapest | 2.12060 | 216.51875 | 0.00000 | 2.20360 |
| XX–4–2 is the cheapest | 3.15040 | 72.69452 | 0.00000 | 3.35860 |
| XX–4–3 is the cheapest | 2.21250 | 16.66451 | 0.00000 | 2.35440 |
| XX–4–4 is the cheapest | 2.05690 | 156.99937 | 0.00000 | 2.16350 |
| XX–4–5 is the cheapest | 1.35150 | 25.56215 | 0.00000 | 1.48310 |
| XX–3–1 is the cheapest | 1.81210 | 50.71883 | 0.00000 | 1.91350 |
| XX–3–2 is the cheapest | 2.83050 | 32.20968 | 0.00000 | 3.03150 |
| XX–3–3 is the cheapest | 1.10100 | 5.21409 | 0.00000 | 1.27970 |
| XX–3–4 is the cheapest | 1.79470 | 37.38379 | 0.00000 | 1.94940 |
| XX–3–5 is the cheapest | 1.10100 | 5.21409 | 0.00000 | 1.27970 |
| XX–2–1 is the cheapest | 2.31230 | 80.70696 | 0.00000 | 2.50190 |
| XX–2–2 is the cheapest | 3.97220 | 46.65835 | 0.00000 | 4.28360 |
| XX–2–3 is the cheapest | 2.51390 | 20.59935 | 0.00000 | 2.72150 |
| XX–2–4 is the cheapest | 2.47060 | 66.96073 | 0.00000 | 2.70980 |
| XX–2–5 is the cheapest | 2.51390 | 20.59935 | 0.00000 | 2.72150 |
| XX–1–1 is the cheapest | 0.46726 | 64.64781 | 0.00000 | 0.57313 |
| XX–1–2 is the cheapest | 0.98914 | 26.28173 | 0.00000 | 1.29210 |
| XX–1–3 is the cheapest | 0.50596 | 5.93366 | 0.00000 | 0.72098 |
| XX–1–4 is the cheapest | 0.89137 | 61.99222 | 0.00000 | 1.02650 |
| XX–1–5 is the cheapest | 0.35990 | 6.53902 | 0.00000 | 0.45584 |
| YY–12–1 is the cheapest | -2.20900 | -54.25961 | 0.00000 | -2.19660 |
| YY–12–4 is the cheapest | -1.91570 | -39.82236 | 0.00000 | -1.91210 |
| YY–12–5 is the cheapest | -1.61760 | -13.61487 | 0.00000 | -1.64130 |
| YY–11–1 is the cheapest | -1.91440 | -40.77609 | 0.00000 | -1.90560 |
| YY–11–4 is the cheapest | -1.61080 | -30.62775 | 0.00000 | -1.57950 |
| YY–11–5 is the cheapest | -1.68370 | -11.13062 | 0.00000 | -1.61470 |
| YY–10–1 is the cheapest | -6.76480 | -26.89851 | 0.00000 | -6.56260 |
| YY–10–4 is the cheapest | -6.76480 | -26.89851 | 0.00000 | -6.56260 |
| YY–10–5 is the cheapest | -6.76480 | -26.89851 | 0.00000 | -6.56260 |
| YY–9–1 is the cheapest | 2.42770 | 41.06164 | 0.00000 | 2.44130 |
| YY–9–4 is the cheapest | 3.07310 | 47.33796 | 0.00000 | 2.91590 |
| YY–9–5 is the cheapest | 3.07310 | 47.33796 | 0.00000 | 2.91590 |
| YY–8–1 is the cheapest | 2.64190 | 41.76408 | 0.00000 | 2.54650 |
| YY–8–4 is the cheapest | 3.08270 | 41.45569 | 0.00000 | 2.90460 |
| YY–8–5 is the cheapest | 2.55830 | 12.59262 | 0.00000 | 2.65160 |
| YY–7–1 is the cheapest | 2.84120 | 76.38378 | 0.00000 | 2.92120 |
| YY–7–4 is the cheapest | 2.88610 | 57.65762 | 0.00000 | 2.94970 |
| YY–7–5 is the cheapest | 2.61680 | 16.51603 | 0.00000 | 2.73670 |
| YY–6–1 is the cheapest | 3.43140 | 68.25712 | 0.00000 | 3.48160 |
| YY–6–4 is the cheapest | 3.47280 | 57.61764 | 0.00000 | 3.52610 |
| YY–6–5 is the cheapest | 3.47280 | 57.61764 | 0.00000 | 3.52610 |
| YY–5–1 is the cheapest | 3.41080 | 45.65323 | 0.00000 | 3.60350 |
| YY–5–4 is the cheapest | 3.53210 | 39.26840 | 0.00000 | 3.69440 |

Table 16 (continued.)

| ML (2011 Data) | | | | |
|---|---|---|---|---|
| **Attribute** | $\widehat{\theta}_{\text{STRA}}$ | ***t*-Statistic** | ***p*-Value** | $\widehat{\theta}_{\text{SAA-50}}$ |
| YY–5–5 is the cheapest | 3.53210 | 39.26840 | 0.00000 | 3.69440 |
| YY–4–1 is the cheapest | 2.26440 | 56.90378 | 0.00000 | 2.45830 |
| YY–4–4 is the cheapest | 2.34850 | 47.20660 | 0.00000 | 2.54880 |
| YY–4–5 is the cheapest | 2.41740 | 15.79645 | 0.00000 | 2.65220 |
| YY–3–1 is the cheapest | 3.38260 | 41.60418 | 0.00000 | 3.72520 |
| YY–3–4 is the cheapest | 3.19500 | 30.24512 | 0.00000 | 3.53850 |
| YY–3–5 is the cheapest | 3.19500 | 30.24512 | 0.00000 | 3.53850 |
| YY–2–1 is the cheapest | 2.72130 | 36.27016 | 0.00000 | 3.07900 |
| YY–2–4 is the cheapest | 2.56910 | 24.16297 | 0.00000 | 2.98200 |
| YY–2–5 is the cheapest | 2.56910 | 24.16297 | 0.00000 | 2.98200 |
| YY–1–1 is the cheapest | 2.09530 | 81.14670 | 0.00000 | 2.23410 |
| YY–1–4 is the cheapest | 2.05470 | 46.30428 | 0.00000 | 2.20380 |
| YY–1–5 is the cheapest | 2.05470 | 46.30428 | 0.00000 | 2.20380 |
| ZZ–15–1 is the cheapest | -4.83860 | -34.96236 | 0.00000 | -4.68410 |
| ZZ–15–4 is the cheapest | -4.83860 | -34.96236 | 0.00000 | -4.68410 |
| ZZ–15–5 is the cheapest | -4.83860 | -34.96236 | 0.00000 | -4.68410 |
| ZZ–14–1 is the cheapest | -4.83860 | -34.96236 | 0.00000 | -4.68410 |
| ZZ–14–4 is the cheapest | 0.58874 | 7.87538 | 0.00000 | 0.16014 |
| ZZ–14–5 is the cheapest | 0.58874 | 7.87538 | 0.00000 | 0.16014 |
| ZZ–13–1 is the cheapest | -0.75881 | -7.15580 | 0.00000 | -1.17540 |
| ZZ–13–4 is the cheapest | 0.75430 | 10.25684 | 0.00000 | 0.32832 |
| ZZ–13–5 is the cheapest | 1.61980 | 11.19344 | 0.00000 | 1.12280 |
| ZZ–12–1 is the cheapest | 1.61980 | 11.19344 | 0.00000 | 1.12280 |
| ZZ–12–4 is the cheapest | 1.61980 | 11.19344 | 0.00000 | 1.12280 |
| ZZ–12–5 is the cheapest | 1.61980 | 11.19344 | 0.00000 | 1.12280 |
| ZZ–11–1 is the cheapest | 1.61980 | 11.19344 | 0.00000 | 1.12280 |
| ZZ–11–4 is the cheapest | 1.61980 | 11.19344 | 0.00000 | 1.12280 |
| ZZ–11–5 is the cheapest | 1.61980 | 11.19344 | 0.00000 | 1.12280 |
| ZZ–10–1 is the cheapest | 4.95000 | 42.62072 | 0.00000 | 4.97900 |
| ZZ–10–4 is the cheapest | 4.95000 | 42.62072 | 0.00000 | 4.97900 |
| ZZ–10–5 is the cheapest | 4.95000 | 42.62072 | 0.00000 | 4.97900 |
| ZZ–9–1 is the cheapest | 3.90880 | 20.50797 | 0.00000 | 2.99270 |
| ZZ–9–4 is the cheapest | 3.90880 | 20.50797 | 0.00000 | 2.99270 |
| ZZ–9–5 is the cheapest | 3.90880 | 20.50797 | 0.00000 | 2.99270 |
| ZZ–8–1 is the cheapest | 3.33510 | 45.16780 | 0.00000 | 3.26460 |
| ZZ–8–4 is the cheapest | 3.73350 | 46.02444 | 0.00000 | 3.51290 |
| ZZ–8–5 is the cheapest | 3.73350 | 46.02444 | 0.00000 | 3.51290 |
| ZZ–7–1 is the cheapest | 3.29980 | 48.54296 | 0.00000 | 3.44830 |
| ZZ–7–4 is the cheapest | 3.29980 | 48.54296 | 0.00000 | 3.44830 |
| ZZ–7–5 is the cheapest | 3.29980 | 48.54296 | 0.00000 | 3.44830 |
| ZZ–6–1 is the cheapest | 3.46890 | 48.58865 | 0.00000 | 3.60390 |
| ZZ–6–4 is the cheapest | 2.97590 | 29.77111 | 0.00000 | 3.08350 |
| ZZ–6–5 is the cheapest | 2.97590 | 29.77111 | 0.00000 | 3.08350 |
| ZZ–5–1 is the cheapest | 2.91690 | 27.57240 | 0.00000 | 2.88510 |

Table 16 (continued.)

| **ML (2011 Data)** | | | | |
|---|---|---|---|---|
| **Attribute** | $\widehat{\theta}_{\text{STRA}}$ | $t$**-Statistic** | $p$**-Value** | $\widehat{\theta}_{\text{SAA-50}}$ |
| ZZ–5–4 is the cheapest | 3.07150 | 27.30399 | 0.00000 | 3.13920 |
| ZZ–5–5 is the cheapest | 3.07150 | 27.30399 | 0.00000 | 3.13920 |
| ZZ–4–1 is the cheapest | 3.25340 | 31.85561 | 0.00000 | 3.42430 |
| ZZ–4–4 is the cheapest | 3.25340 | 31.85561 | 0.00000 | 3.42430 |
| ZZ–4–5 is the cheapest | 3.25340 | 31.85561 | 0.00000 | 3.42430 |
| ZZ–3–1 is the cheapest | 3.66900 | 50.97011 | 0.00000 | 3.88840 |
| ZZ–3–4 is the cheapest | 3.66900 | 50.97011 | 0.00000 | 3.88840 |
| ZZ–3–5 is the cheapest | 3.66900 | 50.97011 | 0.00000 | 3.88840 |
| ZZ–2–1 is the cheapest | 2.87180 | 25.03104 | 0.00000 | 3.08890 |
| ZZ–2–4 is the cheapest | 3.59130 | 31.10176 | 0.00000 | 3.75020 |
| ZZ–2–5 is the cheapest | 3.59130 | 31.10176 | 0.00000 | 3.75020 |
| ZZ–1–1 is the cheapest | 3.96940 | 66.17263 | 0.00000 | 3.95900 |
| ZZ–1–4 is the cheapest | 3.34040 | 31.10747 | 0.00000 | 3.40230 |
| ZZ–1–5 is the cheapest | 3.34040 | 31.10747 | 0.00000 | 3.40230 |
| $\sigma_{1,1}$ | 0.44762 | 0.00001 | 1.00000 | 1.08950 |
| $\sigma_{2,1}$ | 0.41643 | 0.00001 | 1.00000 | 1.06950 |
| $\sigma_{3,1}$ | 0.51663 | 0.00001 | 0.99999 | 1.06620 |
| $\sigma_{4,1}$ | 0.44788 | 0.00001 | 1.00000 | 1.01610 |
| $\sigma_{5,1}$ | 1.27040 | 0.00002 | 0.99999 | 1.25790 |
| $\sigma_{6,1}$ | 1.36880 | 0.00002 | 0.99999 | 1.29880 |
| $\sigma_{7,1}$ | 1.85380 | 0.00002 | 0.99998 | 1.42440 |
| $\sigma_{8,1}$ | 1.64300 | 0.00002 | 0.99998 | 1.28820 |
| $\sigma_{9,1}$ | 2.04070 | 0.00003 | 0.99998 | 1.37330 |
| $\sigma_{10,1}$ | 1.79670 | 0.00002 | 0.99998 | 1.34030 |
| $\sigma_{11,1}$ | 1.94720 | 0.00002 | 0.99998 | 1.42100 |
| $\sigma_{12,1}$ | 2.21660 | 0.00003 | 0.99998 | 1.54750 |
| $\sigma_{13,1}$ | 2.10410 | 0.00003 | 0.99998 | 1.43700 |
| $\sigma_{14,1}$ | 2.24960 | 0.00003 | 0.99998 | 1.59590 |
| $\sigma_{2,2}$ | 0.24981 | 6.14990 | 0.00000 | 0.01905 |
| $\sigma_{3,2}$ | 0.04706 | 1.12850 | 0.25911 | 0.03477 |
| $\sigma_{4,2}$ | -0.06011 | -1.40750 | 0.15928 | 0.05530 |
| $\sigma_{5,2}$ | -2.18980 | -25.89600 | 0.00000 | 0.05976 |
| $\sigma_{6,2}$ | -2.54650 | -26.89800 | 0.00000 | 0.08805 |
| $\sigma_{7,2}$ | -3.90150 | -33.24600 | 0.00000 | -0.00264 |
| $\sigma_{8,2}$ | -3.14170 | -30.09400 | 0.00000 | 0.04369 |
| $\sigma_{9,2}$ | -4.22180 | -33.57100 | 0.00000 | 0.06362 |
| $\sigma_{10,2}$ | -3.65530 | -32.62700 | 0.00000 | -0.00772 |
| $\sigma_{11,2}$ | -4.23540 | -34.29600 | 0.00000 | 0.08543 |
| $\sigma_{12,2}$ | -4.93470 | -36.99600 | 0.00000 | 0.08420 |
| $\sigma_{13,2}$ | -4.68470 | -35.72600 | 0.00000 | 0.09031 |
| $\sigma_{14,2}$ | -4.90570 | -36.34200 | 0.00000 | 0.06590 |
| $\sigma_{3,3}$ | 0.07905 | 2.05090 | 0.04028 | 0.00153 |
| $\sigma_{4,3}$ | 0.09637 | 2.28820 | 0.02213 | -0.05407 |
| $\sigma_{5,3}$ | 2.74290 | 30.24700 | 0.00000 | -0.19262 |

Table 16 (continued.)

| ML (2011 Data) | | | | |
|---|---|---|---|---|
| **Attribute** | $\widehat{\theta}_{\text{STRA}}$ | ***t*-Statistic** | ***p*-Value** | $\widehat{\theta}_{\text{SAA-50}}$ |
| $\sigma_{6,3}$ | 3.20240 | 30.46900 | 0.00000 | -0.20901 |
| $\sigma_{7,3}$ | 5.09350 | 39.70400 | 0.00000 | -0.36537 |
| $\sigma_{8,3}$ | 4.14340 | 36.47700 | 0.00000 | -0.30259 |
| $\sigma_{9,3}$ | 5.59600 | 41.52200 | 0.00000 | -0.39453 |
| $\sigma_{10,3}$ | 4.58620 | 37.99900 | 0.00000 | -0.30670 |
| $\sigma_{11,3}$ | 5.54610 | 42.20000 | 0.00000 | -0.32174 |
| $\sigma_{12,3}$ | 6.54740 | 47.16900 | 0.00000 | -0.56764 |
| $\sigma_{13,3}$ | 6.29110 | 45.62300 | 0.00000 | -0.47309 |
| $\sigma_{14,3}$ | 6.80610 | 48.94200 | 0.00000 | -0.57822 |
| $\sigma_{4,4}$ | 0.01287 | 0.38173 | 0.70266 | 0.00527 |
| $\sigma_{5,4}$ | 1.69200 | 19.91700 | 0.00000 | 0.14332 |
| $\sigma_{6,4}$ | 1.93150 | 19.83200 | 0.00000 | 0.16526 |
| $\sigma_{7,4}$ | 3.21530 | 23.55200 | 0.00000 | 0.32593 |
| $\sigma_{8,4}$ | 2.59800 | 21.85000 | 0.00000 | 0.22060 |
| $\sigma_{9,4}$ | 3.65330 | 24.23300 | 0.00000 | 0.35837 |
| $\sigma_{10,4}$ | 2.91880 | 23.03600 | 0.00000 | 0.25490 |
| $\sigma_{11,4}$ | 3.51800 | 24.47500 | 0.00000 | 0.30862 |
| $\sigma_{12,4}$ | 4.31260 | 26.72500 | 0.00000 | 0.47465 |
| $\sigma_{13,4}$ | 4.14320 | 26.36800 | 0.00000 | 0.37304 |
| $\sigma_{14,4}$ | 4.49510 | 27.26000 | 0.00000 | 0.44864 |
| $\sigma_{5,5}$ | 0.62690 | 10.12500 | 0.00000 | 0.14797 |
| $\sigma_{6,5}$ | 0.76620 | 10.88000 | 0.00000 | 0.18068 |
| $\sigma_{7,5}$ | 0.97971 | 9.56250 | 0.00000 | 0.24964 |
| $\sigma_{8,5}$ | 0.89908 | 10.11200 | 0.00000 | 0.20416 |
| $\sigma_{9,5}$ | 1.00580 | 8.92360 | 0.00000 | 0.29526 |
| $\sigma_{10,5}$ | 0.96998 | 9.89390 | 0.00000 | 0.22662 |
| $\sigma_{11,5}$ | 1.10860 | 9.81100 | 0.00000 | 0.32326 |
| $\sigma_{12,5}$ | 1.10060 | 8.40300 | 0.00000 | 0.37614 |
| $\sigma_{13,5}$ | 1.11860 | 8.87080 | 0.00000 | 0.33473 |
| $\sigma_{14,5}$ | 1.14070 | 8.55340 | 0.00000 | 0.43163 |
| $\sigma_{6,6}$ | 0.06162 | 1.81270 | 0.06988 | 0.11402 |
| $\sigma_{7,6}$ | 0.18280 | 3.23950 | 0.00120 | 0.11427 |
| $\sigma_{8,6}$ | 0.12762 | 2.82760 | 0.00469 | 0.10241 |
| $\sigma_{9,6}$ | 0.19566 | 2.86380 | 0.00419 | 0.20294 |
| $\sigma_{10,6}$ | 0.11102 | 2.06990 | 0.03846 | 0.06878 |
| $\sigma_{11,6}$ | 0.21729 | 3.33930 | 0.00084 | 0.14410 |
| $\sigma_{12,6}$ | 0.22438 | 2.78140 | 0.00541 | 0.23448 |
| $\sigma_{13,6}$ | 0.20685 | 2.64220 | 0.00824 | 0.13798 |
| $\sigma_{14,6}$ | 0.17425 | 2.12260 | 0.03379 | 0.24126 |
| $\sigma_{7,7}$ | 0.43995 | 8.01050 | 0.00000 | 0.03958 |
| $\sigma_{8,7}$ | 0.21586 | 5.16630 | 0.00000 | -0.00793 |
| $\sigma_{9,7}$ | 0.49250 | 7.23560 | 0.00000 | 0.04637 |
| $\sigma_{10,7}$ | 0.36250 | 7.40060 | 0.00000 | -0.04320 |

Table 16 (continued.)

| ML (2011 Data) | | | | |
|---|---|---|---|---|
| **Attribute** | $\widehat{\theta}_{\text{STRA}}$ | $t$**-Statistic** | $p$**-Value** | $\widehat{\theta}_{\text{SAA-50}}$ |
| $\sigma_{11,7}$ | 0.67164 | 10.14100 | 0.00000 | 0.03208 |
| $\sigma_{12,7}$ | 0.88037 | 10.29600 | 0.00000 | 0.03596 |
| $\sigma_{13,7}$ | 0.77327 | 9.54420 | 0.00000 | -0.00900 |
| $\sigma_{14,7}$ | 0.89689 | 9.75440 | 0.00000 | 0.01679 |
| $\sigma_{8,8}$ | 0.02940 | 0.89030 | 0.37330 | 0.00967 |
| $\sigma_{9,8}$ | 0.04801 | 1.08210 | 0.27921 | 0.08601 |
| $\sigma_{10,8}$ | 0.04774 | 1.24870 | 0.21177 | 0.06084 |
| $\sigma_{11,8}$ | 0.05814 | 1.36520 | 0.17219 | 0.04351 |
| $\sigma_{12,8}$ | 0.07501 | 1.50980 | 0.13109 | 0.05106 |
| $\sigma_{13,8}$ | 0.10334 | 2.14810 | 0.03171 | 0.04802 |
| $\sigma_{14,8}$ | 0.03445 | 0.68092 | 0.49592 | 0.05453 |
| $\sigma_{9,9}$ | 0.08749 | 1.92150 | 0.05467 | 0.05751 |
| $\sigma_{10,9}$ | 0.00453 | 0.12902 | 0.89734 | 0.09355 |
| $\sigma_{11,9}$ | -0.03393 | -0.81310 | 0.41616 | 0.04615 |
| $\sigma_{12,9}$ | -0.01430 | -0.30186 | 0.76276 | 0.14551 |
| $\sigma_{13,9}$ | 0.02819 | 0.58799 | 0.55654 | 0.10061 |
| $\sigma_{14,9}$ | 0.01041 | 0.21137 | 0.83260 | 0.13085 |
| $\sigma_{10,10}$ | 0.07082 | 2.08640 | 0.03694 | -0.07082 |
| $\sigma_{11,10}$ | -0.02595 | -0.64760 | 0.51724 | 0.02595 |
| $\sigma_{12,10}$ | 0.02615 | 0.59126 | 0.55435 | -0.02615 |
| $\sigma_{13,10}$ | 0.00956 | 0.20910 | 0.83437 | -0.00956 |
| $\sigma_{14,10}$ | 0.09193 | 1.94830 | 0.05138 | -0.09193 |
| $\sigma_{11,11}$ | 0.01888 | 0.48664 | 0.62651 | 0.02695 |
| $\sigma_{12,11}$ | -0.08898 | -2.15860 | 0.03088 | -0.02597 |
| $\sigma_{13,11}$ | -0.02330 | -0.56737 | 0.57046 | 0.00967 |
| $\sigma_{14,11}$ | -0.18179 | -4.03040 | 0.00006 | 0.00520 |
| $\sigma_{12,12}$ | 0.07040 | 1.77690 | 0.07558 | 0.01693 |
| $\sigma_{13,12}$ | -0.04401 | -1.14380 | 0.25271 | 0.02065 |
| $\sigma_{14,12}$ | 0.05558 | 1.24290 | 0.21390 | -0.00416 |
| $\sigma_{13,13}$ | 0.04380 | 1.37360 | 0.16957 | 0.02910 |
| $\sigma_{14,13}$ | -0.14322 | -4.60990 | 0.00000 | 0.00834 |
| $\sigma_{14,14}$ | 0.06736 | 2.26310 | 0.02363 | 0.03717 |

# REFERENCES


[1] Allenby, G. M. and Rossi, P. E., "Marketing models of consumer heterogeneity," *Journal of Econometrics*, vol. 89, no. 1-2, pp. 57–58, 1999.

[2] Amaldoss, W., Bettman, J. R., and Payne, J. W., "Biased but efficient: An investigation of coordination facilitated by asymmetric dominance," *Marketing Science*, vol. 27, no. 5, pp. 903–921, 2008.

[3] Apostol, T. M., *Modular Functions and Dirichlet Series in Number Theory.* New York: Springer-Verlag, 1989.

[4] Ariely, D., *Predictably Irrational: The Hidden Forces That Shape Our Decisions.* New York: HarperCollins, 2008.

[5] Axhausen, K. W., Zimmerman, A., Schönfelder, S., Rindsfüser, G., and Haupt, T., "Observing the rhythms of daily life: A six-week travel diary," *Transportation*, vol. 29, no. 2, pp. 95–124, 2002.

[6] Bastin, F., Cirillo, C., and Toint, P. L., "An adaptive monte carlo algorithm for computing mixed logit estimators," *Computational Management Science*, vol. 3, no. 1, pp. 55–79, 2006.

[7] Bastin, F., Cirillo, C., and Toint, P. L., "Convergence theory for nonconvex stochastic programming with an application to mixed logit," *Mathematical Programming Series B*, vol. 108, no. 2-3, pp. 207–234, 2006.

[8] Bateman, I. J., Munro, A., and Poe, G. L., "Decoy effects in choice experiments and contingent valuation: Asymmetric dominance," *Land Economics*, vol. 84, no. 1, pp. 115–127, 2008.

[9] Ben-Akiva, M. and Lerman, S. R., *Discrete Choice Analysis: Theory and Application to Travel Demand.* Cambridge, MA: The MIT Press, 1985.

[10] Bertsekas, D. P., Nedić, A., and Ozdaglar, A. E., *Convex Analysis and Optimization.* Belmont: Athena Scientific, 2003.

[11] Besbes, O. and Saure, D., "Product assortment and price competition with informed consumers," *Working paper*, 2010.

[12] Bhat, C. R., "Flexible model structures for discrete choice analysis," in *Handbook of Transport Modelling* (Hensher, D. A. and Button, K. J., eds.), pp. 71–90, Oxford: Pergamon Press, 2000.

[13] Bhat, C. R., "Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model," *Transportation Research*, vol. 35B, no. 7, pp. 677–695, 2001.

[14] Bhat, C. R. and Koppelman, F. S., "Activity-based modeling of travel demand," in *Handbook of Transportation Science* (Hall, R. W., ed.), pp. 35–61, Norwell: Kluwer Academic Publisher, 1999.

[15] Börsch-Supan, A. and Hajivassiliou, V., "Smooth unbiased multivariate probability simulators for maximum likelihood estimation of limited dependent variable models," *Journal of Econometrics*, vol. 58, no. 3, pp. 347–368, 1993.

[16] Börsch-Supan, A., Hajivassiliou, V., Kotlikoff, L. J., and Morris, J. N., "Health, children, and elderly living arrangements: A multiperiod-multinomial probit model with unobserved heterogeneity and autocorrelated errors," in *Topics in the economics of aging* (Wise, D., ed.), pp. 79–108, Chicago, IL: University of Chicago Press, 1992.

[17] Boyd, S. and Vandenberghe, L., *Convex Optimization*. New York: Cambridge University Press, 2004.

[18] Cachon, G. P., Terwiesch, C., and Xu, Y., "On the effects of consumer search and firm entry in a multiproduct competitive market," *Marketing Science*, vol. 27, no. 3, pp. 461–473, 2008.

[19] Calamai, P. H. and Moré, J. J., "Projected gradient methods for linearly constrained problems," *Mathematical Programming*, vol. 39, no. 1, pp. 93–116, 1987.

[20] Caro, F. and Martínez-de-Albéniz, V., "Product and price competition with satiation effects," *Management Science*, vol. 58, no. 7, pp. 1357–1373, 2012.

[21] Coldren, G. M. and Koppelman, F. S., "Modeling the competition among air-travel itinerary shares: GEV model development," *Transportation Research Part A: Policy and Practice*, vol. 39, no. 4, pp. 345–365, 2005.

[22] Coldren, G. M., Koppelman, F. S., Kasturirangan, K., and Mukherjee, A., "Modeling aggregate air-travel itinerary shares: Logit model development at a major US airline," *Journal of Air Transport Management*, vol. 9, no. 6, pp. 361–369, 2003.

[23] Conn, A. R., Gould, N. I. M., and Toint, P. L., *Trust-region Methods*. Philadelphia: Society for Industrial and Applied Mathematics, 2000.

[24] Daganzo, C., *Multinomial Probit: The Theory and Its Applications to Demand Forecasting*. New York: Academic Press, 1979.

[25] Davis, J., Gallego, G., and Topaloglu, H., "Assortment planning under the multinomial logit model with totally unimodular constraint structures." Working paper, School of Operations Research and Information Engineering, Cornell University, 2013.

[26] Davis, J., Gallego, G., and Topaloglu, H., "Assortment optimization under variants of the nested logit model," *Operations Research*, vol. 62, no. 2, pp. 250–273, 2014.

[27] Doyle, J. R., O'Connor, D. J., Reynolds, G. M., and Bottomley, P. A., "The robustness of the asymmetrically dominated effect: Buying frames, phantom alternatives, and in-store purchases," *Psychology and Marketing*, vol. 16, no. 3, pp. 225–243, 1999.

[28] Fudenberg, D. and Levine, D. K., *The Theory of Learning in Games*. Cambridge: MIT Press, 1998.

[29] Gallego, G. and Topaloglu, H., "Constrained assortment optimization for the nested logit model," *Management Science*, 2014. Articles in Advance.

[30] Harris, C., "On the rate of convergence of continuous-time fictitious play," *Games and Economic Behavior*, vol. 22, no. 2, pp. 238–259, 1998.

[31] Hashemi, F. S., Ghosh, S., and Pasupathy, R., "On adaptive sampling rules for stochastic recursions," in *Proceedings of the 2014 Winter Simulation Conference* (Tolk, A., Diallo, S. Y., Ryzhov, I. O., Yilmaz, L., Buckley, S., and Miller, J. A., eds.), (Savanah, GA), pp. 3959–3970, IEEE Press, Dec. 2014.

[32] Heath, T. and Chatterjee, S., "Asymmetric decoy effects on lower-quality versus higher-quality brands: Meta-analytic and experimental evidence," *Journal of Consumer Research*, vol. 22, no. 3, pp. 268–284, 1995.

[33] Hensher, D. A. and Greene, W. H., "The mixed logit model: The state of practice," *Transportation*, vol. 30, no. 2, pp. 133–176, 2003.

[34] Hess, S., Train, K. E., and Polak, J. W., "On the use of a modified latin hypercube sampling (mlhs) method in the estimation of a mixed logit model for vehicle choice," *Transportation Research Part B: Methodological*, vol. 40, no. 2, pp. 147–163, 2006.

[35] Hopp, W. J. and Xu, X., "A static approximation for dynamic demand substitution with applications in a competitive market," *Operations Research*, vol. 56, no. 3, pp. 630–645, 2008.

[36] Huber, J., Payne, J. W., and Puto, C., "Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis," *Journal of Consumer Research*, vol. 9, no. 1, pp. 90–98, 1982.

[37] Huber, J. and Puto, C., "Market boundaries and product choice: Illustrating attraction and subsititution effects," *Journal of Consumer Research*, vol. 10, no. 1, pp. 31–44, 1983.

[38] Izzo, A. J., "Locally uniformly continuous functions," in *Proceedings of The American Mathematical Society*, vol. 122, pp. 1095–1100, 1994.

[39] Kök, A. G., Fisher, M., and Vaidyanathan, R., "Assortment planning: Review of literature and industry practice," in *Retail Supply Chain Management: Quantitative Models and Empirical Studies* (Agrawal, N. and Smith, S. A., eds.), New York: Springer, 2009.

[40] Kök, A. G. and Xu, Y., "Optimal and competitive assortments with endogenous pricing under hierarchical consumer choice models," *Management Science*, vol. 57, no. 9, pp. 1546–1563, 2011.

[41] Li, G., Rusmevichientong, P., and Topaloglu, H., "The *d*-level nested logit model: Assortment and price optimization problems." Working paper, School of Operations Research and Information Engineering, Cornell University, 2013.

[42] Luce, R. D., *Individual Choice Behavior: A Theoretical Analysis.* New York: John Wiley & Sons, Inc., 1959.

[43] McFadden, D. and Train, K., "Mixed mnl models for discrete response," *Journal of Applied Econometrics*, vol. 15, no. 5, pp. 447–470, 2000.

[44] Metrick, A. and Polak, B., "Fictitious play in $2 \times 2$ games: A geometric proof of convergence," *Economic Theory*, vol. 4, no. 6, pp. 923–933, 1994.

[45] Miyasawa, K., "On the convergence of the learning process in a $2 \times 2$ nonzero sum two-person game, Research Memorandum No. 33," tech. rep., Princeton University, 1961.

[46] Monderer, D. and Shapley, L. S., "Fictitious-play property for games with identical interests," *Journal of Economic Theory*, vol. 68, no. 1, pp. 258–265, 1996.

[47] Morokoff, W. J. and Caflisch, R. E., "Quasi-monte carlo integration," *Journal of Computational Physics*, vol. 122, no. 2, pp. 218–230, 1995.

[48] Nau, R., "Arbitrage, incomplete models, and other people's brains," in *Beliefs, Interactions, and Preferences in Decision-Making* (Machina, M. and Munier, B., eds.), Boston: Kluwer Academic Publishers, 1999.

[49] Nemirovski, A., Juditsky, A., Lan, G., and Shapiro, A., "Robust stochastic approximation approach to stochastic programming," *SIAM Journal on Optimization*, vol. 19, no. 4, pp. 1574–1609, 2009.

[50] Nocedal, J. and Wright, S., *Numerical Optimization.* New York, NY: Springer, second ed., 2006.

[51] Pasupathy, R., Glynn, P. W., Ghosh, S., and Hahemi, F., "How much to sample in simulation-based stochastic recursions.," *SIAM Journal on Optimization*, 2015. Under revision.

[52] Robinson, J., "An iterative method of solving a game," *Annals of Mathematics*, vol. 54, no. 2, pp. 296–301, 1951.

[53] Rooderkerk, R. P., Heerde, H. J. V., and Bijmolt, T. H. A., "Incorporating context effects into a choice model," *Journal of Marketing Research*, vol. 48, no. 4, pp. 767–780, 2011.

[54] Rossi, P. E. and Allenby, G. M., "A Bayesian approach to estimating household parameters," *Journal of Marketing Research*, vol. 30, no. 2, pp. 171–182, 1993.

[55] Rossi, P. E., McCulloch, R. E., and Allenby, G. M., "Hierarchical modeling of consumer heterogeneity: an application to target marketing," in *Case Studies in Bayesian Statistics* (Gatsonis, C., Hodges, J. S., Kass, R. E., and Singpurwalla, N. D., eds.), pp. 323–349, New York: Springer Verlag, 1995.

[56] Rusmevichieong, P. and Topaloglu, H., "Robust assortment optimization in revenue management under the multinomial logit choice model," *Operations Research*, vol. 60, no. 4, pp. 865–882, 2012.

[57] Shafir, S., Waite, T. A., and Smith, B. H., "Context-dependent violations of rational choice in honey bees (apis mellifera) and gray jays (perisoreus canadensis)," *Behavioral Ecology and Sociobiology*, vol. 51, no. 2, pp. 180–187, 2002.

[58] Shapiro, A., Dentcheva, D., and Ruszczyński, A., *Lectures on Stochastic Programming: Modeling and Theory*. Philadelphia: SIAM, 2009.

[59] Shapiro, A. and Homem-de-Mello, T., "A simulation-based approach to two-stage stochastic programming with recourse," *Mathematical Programming: Series A and B*, vol. 81, no. 3, pp. 301–325, 1998.

[60] Shapley, L. S., "Some topics in two-person games," in *Advances in Game Theory* (Dresher, M., Shaply, L. S., and Tucker, A. W., eds.), pp. 1–28, Princeton: Princeton University Press, 1964.

[61] Simonson, I., "Choice based on reasons: The case of attraction and compromise effects," *Journal of Consumer Research*, vol. 16, no. 2, pp. 158–174, 1989.

[62] Simonson, I. and Tversky, A., "Choice in context: Tradeoff contrast and extremeness aversion," *Journal of Marketing Research*, vol. 29, no. 3, pp. 281–295, 1992.

[63] Talluri, K. T. and van Ryzin, G. J., "Revenue management under a general discrete choice model of consumer behavior," *Management Science*, vol. 50, no. 1, pp. 15–33, 2004.

[64] Train, K. E., *Discrete Choice Methods with Simulation*. Cambridge, UK: Cambridge University Press, 2003.

[65] Tversky, A. and Kahneman, D., "Judgment under uncertainty: Heuristics and biases," *Science*, vol. 185, no. 4157, pp. 1124–1131, 1974.

[66] Tversky, A., Sattah, S., and Slovic, P., "Contingent weighting in judgment and choice," *Psychological Review*, vol. 95, no. 3, pp. 371–384, 1988.

[67] Tversky, A. and Simonson, I., "Context-dependent preference," *Management Science*, vol. 39, no. 10, pp. 1179–1189, 1993.

[68] Ulu, C., Honhon, D., and Alptekinoğlu, A., "Learning consumer tastes through dynamic assortments," *Operations Research*, vol. 60, no. 4, pp. 833–849, 2012.

[69] Van Strien, S. and Sparrow, C., "Fictitious play in $3 \times 3$ games: Chaos and dithering behaviour," *Games and Economic Behavior*, vol. 73, no. 1, pp. 262–286, 2011.

[70] Vulcano, G., van Ryzin, G., and Chaar, W., "Choice-based revenue management: An empirical study of estimation and optimization," *Manufacturing & Service Operations Management*, vol. 12, no. 3, pp. 371–392, 2010.

[71] Zhang, D. and Cooper, W. L., "Revenue management for parallel flights with customer-choice behavior," *Operations Research*, vol. 53, no. 3, pp. 415–431, 2005.