Identification, Estimation and Inference in Empirical Games

by

Mathieu Marcoux

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Graduate Department of Economics
University of Toronto

© Copyright 2017 by Mathieu Marcoux
Abstract

Identification, Estimation and Inference in Empirical Games

Mathieu Marcoux
Doctor of Philosophy
Graduate Department of Economics
University of Toronto
2017

This thesis collects three papers studying topics related to the econometrics of empirical games. In Chapter 1, I investigate the identification and the estimation of empirical games of incomplete information with common-knowledge unobservable heterogeneity and potentially multiple equilibria realized in the data. I introduce pre-determined outcome variables to recover such unobserved heterogeneity. The recovered unobservables provide an extra source of exogenous variation that helps to identify the primitives of the model. I apply this method to study mobile telecommunications in Canada. I estimate a game in which national incumbents and new entrants choose the number of transceivers they install in different markets. The results highlight sizeable economies of density in transceivers location decisions. Counterfactual experiments shed light on the government’s attempt to increase competition in this industry.

In Chapter 2, I propose a test of an assumption commonly maintained when estimating discrete games of incomplete information, i.e. the assumption of equilibrium uniqueness in the data generating process. The test I propose is robust to player-specific common-knowledge unobservables. The main identifying assumption is the existence of an observable variable interpreted as a proxy for these unobservables. It must (i) have sufficient variation; (ii) be correlated with the common-knowledge unobservables; and (iii) provide only redundant information regarding the players’ decisions and the equilibrium selection, were these unobservables actually observed.

In Chapter 3, I study bias reduction when estimating dynamic discrete games. An iterative approach (the $K$-step estimator) is known to reduce finite sample bias, provided that some equilibrium stability conditions are satisfied. Modified versions of the $K$-step estimator have been proposed to deal with this stability issue. Alternatively, there exist other bias reduction techniques which do not rely on equilibrium’s stability, but have not received much attention in this class of models. Using a dynamic game of market entry and exit, I compare the finite sample properties of the $K$-step approach with alternative methods. The results show that, even when the $K$-step estimator does not converge to a single point after a large number of iterations, it still considerably reduces finite sample bias for small values of $K$. 
Acknowledgements

I will forever be indebted to the members of my supervisory committee. I thank my main advisor, Victor Aguirregabiria, for his support, comments and criticisms that have greatly improved the chapters of this thesis. I am grateful for everything I have learned from Victor and the way he has impacted my understanding of econometrics and empirical industrial organization. I also thank Christian Gouriéroux, Ismael Mourifié and Yuanyuan Wan for their help and for everything they taught me along the way. Each and every member of my committee has influenced my research in a significant way.

I have also greatly benefitted from insightful discussions with many faculty members at the University of Toronto. In particular, I must mention Gordon Anderson, Martin Burda, Jiaying Gu, Yao Luo, Frank Mathewson, Robert McMillan, Angelo Melino, Jennifer Murdock, Eduardo Souza-Rodrigues, Junichi Suzuki and Adonis Yatchew.

Special thanks go to my colleagues and friends for all the discussions, ideas sharing and support. I owe them a lot.

I also want to thank my parents Mario and Lucie, as well as my brother Jean-Michel for their continuing encouragement and motivation.

Finally, the list would be incomplete without thanking Penelope Oleksiak for inspiration.
# Contents

## 1 Games with unobservable heterogeneity and multiple equilibria: An application to mobile telecommunications

1.1 Introduction .............................................................................................................. 1  
1.2 Transceivers: a primer ............................................................................................... 3  
1.3 A game of simultaneous decisions ........................................................................... 6  
  1.3.1 The economic model .......................................................................................... 6  
  1.3.2 The econometric model ...................................................................................... 8  
1.4 Identification of payoff functions ........................................................................... 11  
  1.4.1 Controlling for common-knowledge unobservable heterogeneity ................. 12  
  1.4.2 Controlling for multiplicity of equilibria ......................................................... 13  
  1.4.3 Inverting the conditional choice probabilities ................................................. 14  
  1.4.4 Identification result ......................................................................................... 15  
1.5 Application: transceivers’ locations in Canada ..................................................... 16  
  1.5.1 Industry background ......................................................................................... 17  
  1.5.2 Empirical model ............................................................................................... 18  
  1.5.3 Data ................................................................................................................ 19  
  1.5.4 Estimation ........................................................................................................ 21  
  1.5.5 Counterfactual experiments: subsidizing transceivers .................................. 25  
1.6 Conclusion ............................................................................................................... 30  
1.7 Appendix ............................................................................................................... 31  
  1.7.1 Factor model estimation .................................................................................... 31  
  1.7.2 Proof of Lemma 1 ............................................................................................ 32

## 2 Sharp test for equilibrium uniqueness in discrete games with a flexible information structure

2.1 Introduction ............................................................................................................... 35  
2.2 Related literature ................................................................................................. 37  
2.3 A static discrete game with simultaneous decisions ........................................... 39  
  2.3.1 Economic model ............................................................................................... 39  
  2.3.2 Econometric model .......................................................................................... 42  
2.4 Nonparametric identification .............................................................................. 44  
  2.4.1 The need for more structure .......................................................................... 44  
  2.4.2 Exclusion restriction ....................................................................................... 44
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.4.3</td>
<td>Constructing the identified set</td>
<td>47</td>
</tr>
<tr>
<td>2.4.4</td>
<td>Properties of the identified set</td>
<td>49</td>
</tr>
<tr>
<td>2.4.5</td>
<td>Identifying power of $\Theta_l \neq \emptyset$</td>
<td>50</td>
</tr>
<tr>
<td>2.5</td>
<td>Statistical test</td>
<td>53</td>
</tr>
<tr>
<td>2.5.1</td>
<td>A two-stage minimum distance approach</td>
<td>53</td>
</tr>
<tr>
<td>2.5.2</td>
<td>Asymptotic level</td>
<td>56</td>
</tr>
<tr>
<td>2.5.3</td>
<td>Monte-Carlo experiments</td>
<td>56</td>
</tr>
<tr>
<td>2.6</td>
<td>Concluding remarks</td>
<td>57</td>
</tr>
<tr>
<td>2.7</td>
<td>Appendix</td>
<td>59</td>
</tr>
<tr>
<td>2.7.1</td>
<td>Proof of Proposition 2</td>
<td>59</td>
</tr>
<tr>
<td>2.7.2</td>
<td>Proof of Corollary 1</td>
<td>60</td>
</tr>
<tr>
<td>2.7.3</td>
<td>Proof of Corollary 2</td>
<td>61</td>
</tr>
<tr>
<td>3</td>
<td>Two-step semiparametric estimation of dynamic discrete games: A comparison of bias reduction techniques</td>
<td>64</td>
</tr>
<tr>
<td>3.1</td>
<td>Introduction</td>
<td>64</td>
</tr>
<tr>
<td>3.2</td>
<td>Literature review</td>
<td>66</td>
</tr>
<tr>
<td>3.3</td>
<td>Model and two-step estimation</td>
<td>67</td>
</tr>
<tr>
<td>3.3.1</td>
<td>General framework</td>
<td>67</td>
</tr>
<tr>
<td>3.3.2</td>
<td>Dynamic discrete game of market entry and exit</td>
<td>67</td>
</tr>
<tr>
<td>3.3.3</td>
<td>Two-step semiparametric estimator</td>
<td>69</td>
</tr>
<tr>
<td>3.4</td>
<td>Bias reduction techniques and alternative estimators</td>
<td>70</td>
</tr>
<tr>
<td>3.4.1</td>
<td>K-step estimator</td>
<td>71</td>
</tr>
<tr>
<td>3.4.2</td>
<td>Modified K-step estimators</td>
<td>72</td>
</tr>
<tr>
<td>3.4.3</td>
<td>Bootstrap</td>
<td>74</td>
</tr>
<tr>
<td>3.4.4</td>
<td>Jackknife</td>
<td>75</td>
</tr>
<tr>
<td>3.4.5</td>
<td>Single Newton full maximum likelihood iteration</td>
<td>76</td>
</tr>
<tr>
<td>3.5</td>
<td>Monte-Carlo experiments</td>
<td>76</td>
</tr>
<tr>
<td>3.5.1</td>
<td>Data generating processes</td>
<td>76</td>
</tr>
<tr>
<td>3.5.2</td>
<td>Comparison of alternative estimators</td>
<td>78</td>
</tr>
<tr>
<td>3.5.3</td>
<td>Coverage issues</td>
<td>86</td>
</tr>
<tr>
<td>3.6</td>
<td>Concluding remarks</td>
<td>93</td>
</tr>
<tr>
<td>3.7</td>
<td>Appendix</td>
<td>93</td>
</tr>
<tr>
<td>3.7.1</td>
<td>Proof of Lemma 4</td>
<td>93</td>
</tr>
<tr>
<td>3.7.2</td>
<td>Proof of Lemma 5</td>
<td>95</td>
</tr>
<tr>
<td>3.7.3</td>
<td>Two-step semiparametric estimation of the model</td>
<td>95</td>
</tr>
</tbody>
</table>

Bibliography 98
## List of Tables

1.1 Correlation between players’ estimated unobservable heterogeneity ........................................... 22
1.2 Ordered-response estimates for control variables and thresholds ............................................. 24
1.3 Ordered-response estimates for strategic interactions ............................................................... 25
1.4 Average markets ....................................................................................................................... 26

2.1 Probability of rejecting $H_0 : \Theta_1 \neq \emptyset$ ........................................................................ 58

3.1 Stable equilibrium: Monte-Carlo comparison of two-step estimators ........................................ 79
3.2 Unstable equilibrium: Monte-Carlo comparison of two-step estimators .................................... 80
3.3 Stable equilibrium: Monte-Carlo comparison of the estimators ............................................... 82
3.4 Unstable equilibrium: Monte-Carlo comparison of the estimators .......................................... 84
3.5 RMSE ratios (compared to the unfeasible two-step estimator) .................................................. 85
3.6 Proportions (in %) of Monte-Carlo samples that converged ..................................................... 86
3.7 Convergence of the $K$-step estimator ....................................................................................... 90
List of Figures

1.1 Locations of transceivers .................................................. 5
1.2 Effect of subsidy on probability of installing at least one transceiver – small market .... 26
1.3 Effect of subsidy on probability of installing at least one transceiver – medium market ... 27
1.4 Effect of subsidy on probability of installing at least one transceiver – large market .... 27
1.5 Effect of subsidy on probabilities of installing many transceivers – small market ....... 28
1.6 Effect of subsidy on probabilities of installing many transceivers – medium market ....... 28
1.7 Effect of subsidy on probabilities of installing many transceivers – large market ....... 29

2.1 Equilibria multiplicity in Example 2 ....................................... 42
2.2 Graphical representation of the identified set for the $2 \times 2 \times 2$ case ......................... 49
2.3 Emptiness of $\Theta_1$ – First data generating process .......................... 52
2.4 Emptiness of $\Theta_1$ – Second data generating process .......................... 52

3.1 Stable equilibrium: estimation of $\theta_{RN}$ .................................... 88
3.2 Unstable equilibrium: estimation of $\theta_{RN}$ .................................... 89
3.3 Stable equilibrium: estimation of $\theta_{RN}$, odd and even iterations ......................... 91
3.4 Unstable equilibrium: estimation of $\theta_{RN}$, odd and even iterations ......................... 92
Chapter 1

Games with unobservable heterogeneity and multiple equilibria: An application to mobile telecommunications

1.1 Introduction

Game theory suggests a variety of models for studying strategic interactions between economic agents. Taking these models to data helps to better explain players’ decisions and to evaluate the impact of counterfactual experiments involving such interactions. Typical applications of empirical games investigate questions related to market entry, advertising and investment decisions, among others.\(^1\) These models have drawn the attention of econometricians due to the identification challenges that they raise. The simultaneous nature of decisions studied through empirical games is an obvious concern for the identification of these models. Other complications may be due to multiplicity of equilibria and asymmetry in the information known to all players when compared with the information known to the econometrician.

Intuitively, one may hope to recover the effect of strategic interactions on players’ payoffs if there exists an exogenous source of variation that does not affect their own payoffs, but shifts their competitors’ decisions. In some sense, the problem is similar to the identification of simultaneous equations, such as models of supply and demand, which require cost and demand shifters. Many identification arguments available in the literature exploit exogenous variation in some player-specific observable regressors that are excluded from competitors’ payoffs.\(^2\) Examples of variables satisfying this type of exclusion in games of market entry could be firms’ incumbency status or distance to firms’ headquarters. In theory, it is easy to argue that these excluded variables do exist. Unfortunately, they may not be observable in practice. This occasional unavailability has encouraged researchers such as Sweeting (2009), De Paula and Tang

---


\(^2\)This type of exclusion restriction is used both in games of complete information (e.g., Tamer, 2003) and in games of incomplete information (e.g., Bajari, Hong, Krainer, and Nekipelov, 2010). It is also put forward in the few recent papers that are associated with a flexible information structure that encompasses these two groups of models (Grieco, 2014; Aguirregabiria and Mira, 2016; Magnolfi and Roncoroni, 2016; Xiao, 2017).
Chapter 1. Games with unobservable heterogeneity and multiple equilibria

(2012) and Aradillas-Lopez and Gandhi (2016) to exploit multiplicity of equilibria as a potential source of identification. The justification is the same: different equilibria shift players’ decisions, but not their payoffs. However, as it is currently framed in the literature, this alternative identification argument fails if players’ decisions are affected by both multiple equilibria and player-specific common-knowledge unobservables. To extend these results, one must separate the payoff-relevant common-knowledge unobservables from the payoff-irrelevant multiple equilibria.

In this chapter, I present a new identification argument for empirical games, which I carry through an application. Using predetermined outcome variables, I propose a method to recover player-specific common-knowledge unobservables. These unobservables, henceforth referred to as common-knowledge unobservable heterogeneity, typically arise in applications where players know more about each other’s payoffs than the econometrician does. For instance, in an industry with a relatively small number of players, some components of the costs of entering and operating in a market may be commonly known to all firms, but not observed by the researcher. There are two important advantages associated with recovering such common-knowledge unobservables. First, it allows the researcher to explicitly control for unobservables that would otherwise bias the estimation. Second, it offers an extra source of exogenous variation that helps to identify the primitives of the model separately from the variation generated by potentially multiple equilibria realized in the data. As far as I know, the identifying power associated with variation of these unobservables has not been exploited in the literature on empirical games. This approach is specially useful in cases where player-specific regressors satisfying the exclusion restriction frequently used in the literature are not available. I apply the proposed method to study mobile telecommunications in Canada. Despite the federal government’s efforts to increase competition in this industry, some of the new entrants fail to use parts of their spectrum or eventually go out of business. In order to explain these drawbacks, I study transceivers’ location decisions made by incumbents and new entrants. Bids for spectrum licenses are predetermined outcomes used to control for agreements between firms and other common-knowledge information not observable to the econometrician.

The key identifying assumption of the method I propose is that player-specific predetermined outcome variables must be related to common-knowledge unobservable heterogeneity, but be conditionally independent of which equilibrium is realized in the game of interest. As I argue below, predetermined variables realized before the game is played, such as bids for licenses, costs of franchises, market research, exploratory investments, etc., are likely to satisfy these conditions. The conditional independence is crucial to treat unobservable heterogeneity separately from multiplicity of equilibria.

The intuition behind the new identification argument is as follows. Player-specific common-knowledge unobservable heterogeneity does not enter competitors’ payoff functions. If such unobservables were actually observed by the researcher, they would satisfy the commonly-used exclusion restriction. I propose a menu of control functions and factor models to recover player-specific common-knowledge unobservables from observable predetermined outcome variables. Next, methods recently developed in the literature on non-parametric finite mixtures (in particular Kasahara and Shimotsu, 2014 and Bonhomme, Jochmans, and Robin, 2016) are used to identify the potentially multiple equilibria realized in the data. One can then exploit exogenous variation generated by common-knowledge unobservable heterogeneity and multiple equilibria to recover the primitives of the model.

---

3This statement is not entirely true for Sweeting (2009), who discusses how common-knowledge unobservable heterogeneity could be parametrically introduced in the model. Notice, however, that the suggested estimation method is based on a nested fixed point algorithm, which becomes computationally very costly when multiple players choose from a rich choice set.
The proposed identification argument is well-suited for studying mobile transceivers’ location decisions. An important feature of the Canadian telecommunications industry is that some firms have reached agreements (e.g., network sharing, tower sharing, domestic roaming, etc.) that affect their transceivers’ locations, but that are not fully observable even to the market regulator. Furthermore, because there is a relatively small number of firms in this industry and since they compete in many markets across the country, these firms are likely to know more about each other’s payoffs than the researcher does. In other words, common-knowledge unobservable heterogeneity is an important concern. Firms’ bids for spectrum licenses, which are realized before firms decide where to install new transceivers, should reflect those agreements and other common-knowledge unobservables. These bids are therefore used as predetermined outcomes to control for common-knowledge unobservable heterogeneity in transceivers’ location decisions.

Estimation results suggest that, for the incumbents, there are considerable economies of density with respect to previously accumulated stocks of transceivers. Since new entrants must build brand-new networks, they do not benefit from such economies of density, which may explain why they do not use some of their allocated spectrum or even leave the industry. Furthermore, I perform counterfactual experiments to evaluate the equilibrium impact of subsidizing the first few transceivers installed by new entrants. I find that this subsidy may have a positive impact on both new entrants’ and incumbents’ probabilities of installing new transceivers. In particular, some of the incumbents respond to the increase in new entrants’ probabilities of installing transceivers by improving the quality of their network via installing more transceivers. This insight is specially relevant since, in this industry, policy makers have been worried that encouraging investments from new entrants may discourage investments made by incumbents.

The rest of the chapter is structured as follows. In order to justify the need for a new methodological contribution, Section 1.2 briefly describes the basics of transceivers and their role in mobile telecommunications. The economic model and its econometric counterpart are introduced in Section 1.3. Section 1.4 presents the identification result. In Section 1.5, the proposed approach is applied to study transceivers’ locations in Canada. Section 1.6 concludes.

1.2 Transceivers: a primer

Before introducing a general model, I start by presenting basic information about transceivers, the main object of interest in this chapter’s empirical application. Besides briefly describing the industry, the current section highlights how common-knowledge unobservable heterogeneity is a first-order concern, which justifies the need to develop the identification argument proposed in Section 1.4. However, most of the relevant industry background and institutional details are treated in Section 1.5.

Simply put, transceivers are antennae which are installed by network operators to provide coverage in a given geographic area. Prior to installing transceivers, network operators must first be granted a spectrum license. In Canada, since the early 2000’s, these licenses have been allocated through auctions.

There are at least two reasons for installing a larger number of transceivers in a given region. First, each transceiver can only provide coverage in a limited surrounding area. More transceivers are therefore needed to cover larger regions. Second, there is a limit to the amount of information that can be carried

\[\text{\cite{Hausman2002, Gruber2005}}\]
simultaneously through a given frequency. As a result, more transceivers are needed in more densely populated areas.

Figure 1.1 shows transceivers’ locations\(^5\) for three incumbents (Rogers, Bell and Telus) and two new entrants (Vidéotron and Wind). Some regional patterns are noticeable. In particular, while Rogers’ transceivers are spread out across the country, Bell and Telus tend to cover different areas. For new entrants, Vidéotron covers most of Quebec’s inhabited regions and Wind focuses on relatively large cities. These patterns are, at least in part, due to different agreements between network operators. In fact, to provide coverage in a given area, operators may either install their own transceivers or use other firms’. For instance, subscribers may use their mobile device outside of their provider’s service area insofar as their provider has agreed to use another operator’s network. Such agreements are referred to as domestic roaming. In some cases, network operators may even agree to explicitly share parts of their networks with each other. The terms of these agreements are, however, unobservable to the researcher. In other words, domestic roaming and network-sharing agreements are an important source of common-knowledge unobservable heterogeneity in this industry. Since these agreements affect firms’ decisions, one cannot simply omit their existence when studying strategic interactions in transceivers’ locations.

It is worth mentioning that competition in mobile telecommunications is a topic that has drawn considerable attention in economics. Several papers have studied Federal Communications Commission (FCC) auctions in the United States focussing, among other topics, on collusion (e.g., Cramton and Schwartz, 2000, 2002) and the efficiency of the system in place (e.g., Fox and Bajari, 2013). Effects of changes in market structure on industry outcomes have also been investigated. For instance, Bajari, Fox, and Ryan (2008) estimate consumers’ valuation of larger coverage areas resulting from firms consolidation. Seim and Viard (2011) examine how new entries affect retail pricing and product availability.

These papers are far from being an exhaustive list of all work done on competition in mobile telecommunications. Nonetheless, to the best of my knowledge, the current chapter is the first attempt at modelling strategic interactions in transceivers’ location decisions, despite the fact that these decisions are important determinants of competition in the industry. As can be seen from Figure 1.1, there are some parts of Canada where only a small number of providers have actually installed transceivers. In fact, even if there exist regulatory conditions regarding domestic roaming and site sharing (Industry Canada, 2013a), both the Canadian Radio-television and Telecommunications Commission (the telecommunications regulating body in Canada) and the Competition Bureau have expressed concerns about the effect of potential market power on negotiated agreements between incumbents and new entrants (see Industry Canada, 2013b; Commissioner of Competition, 2014).\(^6\)

To sum up, the effects of transceivers’ locations on competition raise policy-relevant questions that require the researcher to take the problem of common-knowledge unobservable heterogeneity seriously. The identification argument proposed in the current chapter aims at controlling for these unobservables and exploiting their identifying power.

---

\(^5\)As will be clearly explained in Section 1.5, these are transceivers operating on frequencies allocated up to the Advanced Wireless Service 1 2008 spectrum auction (AWS1-2008).

\(^6\)See Church and Wilkins (2013) for an alternative view of the state of competition in the Canadian mobile telecommunications industry.
Chapter 1. Games with unobservable heterogeneity and multiple equilibria

Notes: Transceivers’ locations for three incumbents (Rogers, Bell, and Telus) and two new entrants (Vidéotron and Wind). Each point corresponds to one or more transceivers since several transceivers can be located at the same site. Maps include all transceivers operating on frequencies allocated up to the Advanced Wireless Service 1 2008 spectrum auction (AWS1-2008).

Figure 1.1: Locations of transceivers
1.3 A game of simultaneous decisions

In this section, I introduce a game in which each player simultaneously chooses one action from a finite and discrete set of possible decisions. I first present the game as it is played by the players before turning to the game observed by the econometrician.

1.3.1 The economic model

Consider \( N \) players indexed by \( i \in \mathbb{N} = \{1, 2, \ldots, N\} \), who simultaneously choose a single action from a finite and discrete set \( \mathcal{Y} = \{0, 1, \ldots, J\} \) called the choice set. Let \( \mathcal{Y}_i \) be the random variable with realization \( y_i \in \mathcal{Y} \) corresponding to player \( i \)'s decision. Let \( \mathcal{Y} \) be the vector of all players’ decisions with realizations \( \mathbf{y} \in \mathcal{Y}^N \). Throughout the chapter, the index \( -i \) will be used to refer to all players except player \( i \). For instance, \( \mathcal{Y}_{-i} \) with realizations \( \mathbf{y}_{-i} \in \mathcal{Y}^{N-1} \) are the decisions of player \( i \)'s opponents.

Decisions are contingent on two types of state variables. First, there is a vector of state variables \( \mathcal{S} \) with realizations \( \mathbf{s} \in \mathcal{S} \) that are common-knowledge to all players. Second, for each player, there is a vector of action-specific state variables \( \mathcal{E}_i = [\varepsilon_i(0), \ldots, \varepsilon_i(J)] \) with realizations \( \varepsilon_i = [\varepsilon_i(0), \ldots, \varepsilon_i(J)] \in \mathbb{R}^{J+1} \) that are player \( i \)'s private information. The introduction of such private information in the model makes this game of simultaneous decisions a game of incomplete information. Let \( G_{\mathcal{E}_i}(\cdot) \) be the cumulative distribution function of \( \mathcal{E}_i \).

Let the function \( \pi_i (\cdot) : \mathcal{Y}^N \times \mathcal{S} \times \mathbb{R} \to \mathbb{R} \) be referred to as player \( i \)'s payoff function. Therefore, \( \pi_i (y_i, \mathbf{y}_{-i}, \mathbf{s}, \varepsilon_i(y_i)) \) is player \( i \)'s payoff for choosing \( y_i \), when the other players choose \( \mathbf{y}_{-i} \) and the realized state is associated with \( (\mathbf{s}, \varepsilon_i) \). The following assumptions, which are commonly used in the literature, are maintained on state variables and payoff functions.

**Assumption 1** (State variables and payoffs). (i) \( \mathcal{S}, \mathcal{E}_1, \ldots, \mathcal{E}_N \) are mutually independent. (ii) \( G_{\mathcal{E}_i}(\cdot) \), \( \ldots, G_{\mathcal{E}_N}(\cdot) \) are common-knowledge to all players and they are absolutely continuous with respect to the Lebesgue measure on \( \mathbb{R}^{J+1} \). (iii) \( \pi_1(\cdot), \ldots, \pi_N(\cdot) \) are common-knowledge to all players. (iv) Private information enters the payoff function additively, i.e. \( \pi_i (\mathbf{y}, \mathbf{s}, \varepsilon_i(y_i)) = \pi_i (\mathbf{y}, \mathbf{s}) + \varepsilon_i (y_i) \).

The timing of the game should be understood as follows. First, state variables are realized. Then, players simultaneously choose an action from their choice set. By Assumption 1, player \( i \)'s information set at the time of simultaneous decisions is given by:

\[
\mathcal{I}_i = \left\{ \mathbf{s}, \varepsilon_i, \{\pi_i(\cdot)\}_{i \in \mathbb{N}}, \{G_{\mathcal{E}_i}(\cdot)\}_{i \in \mathbb{N}} \right\}.
\]

In particular, players use their knowledge of the payoff functions and the distributions of the private information state variables to form beliefs about their opponents’ decisions.

Player \( i \)'s strategy is a function, denoted \( \rho_i (\cdot) \), that maps the information set to the choice set, i.e. \( \rho_i (\cdot) : \mathcal{I}_i \to \mathcal{Y} \). The conditional choice probability (CCP) of player \( i \) choosing \( y_i \in \mathcal{Y} \) at a given realization of the common-knowledge state variables \( \mathbf{s} \) is therefore equal to:

\[
P_i(y_i|\mathbf{s}) = \int 1 \{\rho_i(\mathcal{I}_i) = y_i\} \, dG_{\mathcal{E}_i}(\varepsilon_i)
\]

which can be interpreted as the beliefs formed by player \( i \)'s opponents regarding player \( i \)'s decision given strategy \( \rho_i(\cdot) \). Let \( \mathbf{P}_i(\mathbf{s}) \) be a \((J + 1)\)-dimensional vector collecting player \( i \)'s CCPs with each element

\footnote{In terms of notation, random variables are denoted by capital calligraphic letters and their realizations are written in lower case letters. Boldface letters refer to vectors and matrices. Finally, sets are denoted with capital script letters.}
corresponding to a different action, i.e. \( P_i(s) \equiv [P_1(0|s), \ldots, P_1(J|s)]' \), and collect these vectors of CCPs in a \( N(J+1) \)-dimensional vector \( P(s) \equiv [P_1(s)', \ldots, P_N(s)']' \). Similarly as before, \( P_{-i}(s) \) would be the \((N-1)(J+1)\)-dimensional vector collecting the vectors of CCPs for all players except player \( i \).

Since player \( i \) does not know \( y_{-i} \) at the time of decision, its strategy is based on its expected payoffs. \( P_{-i}(s) \) is used to define the expected payoff of player \( i \) choosing \( y_i \) given \( s \) and \( \varepsilon_i \) as:

\[
\pi_i^P(y_i, s, \varepsilon_i(y_i)) = \sum_{y_{-i} \in Y^{N-1}} P(y_{-i}|s) \pi_i(y_i, y_{-i}, s, \varepsilon_i(y_i)). \tag{1.3}
\]

Player \( i \)'s strategy is therefore a function of \( P_{-i}(s) \), such that (1.2) can be written as:

\[
P_i(y_i|s) = \varphi_i(y_i, s, P_{-i}(s)) \tag{1.4}
\]

where the right-hand side of equation (1.4) is interpreted as player \( i \)'s best response to its beliefs regarding its opponents' actions, i.e. \( P_{-i}(s) \). Let \( \Psi(s, P(s)) \) be the \( N(J+1) \)-dimensional vector collecting these best response functions for all actions and all players organized in the same order as in \( P(s) \), i.e.:

\[
\Psi(s, P(s)) = [\varphi_1(0, s, P_{-1}(s)), \ldots, \varphi_N(J, s, P_{-N}(s))]'. \tag{1.5}
\]

It follows that:

\[
P(s) = \Psi(s, P(s)) \tag{1.6}
\]

and \( \Psi(s, \cdot) : [0, 1]^{N(J+1)} \to [0, 1]^{N(J+1)} \) is henceforth referred to as the best response mapping.

Defining a Bayesian-Nash Equilibrium (BNE) in the probability space is convenient to analyze equilibrium existence and potential multiplicity (Milgrom and Weber, 1985). Definition 1 is the definition of equilibrium that is used throughout the chapter.

**Definition 1** (Bayesian-Nash Equilibrium in pure strategies). For a given realization of common-knowledge state variables \( s \), a pure strategy BNE in the probability space is a vector of CCPs, \( P^*(s) \), such that \( P^*(y_i|s) = \varphi_i(y_i, s, P^*_{-i}(s)) \) for any \( y_i \in \mathcal{Y} \) and any \( i \in \mathcal{N} \).

In other words, a BNE in pure strategies is a fixed point of the best response mapping, i.e. \( P^*(s) = \Psi(s, P^*(s)) \), such that each player’s beliefs are consistent with their opponents’. Since \( \Psi(s, \cdot) \) maps a compact set to itself and since it is continuous in \( P(s) \), equilibrium existence follows from Brouwer’s fixed point theorem. However, equilibrium uniqueness is not guaranteed.

Let \( \mathcal{T} \) be a random variable with realization \( \tau \in \mathcal{F}(s) \) indicating which equilibrium is played at a specific realization \( s \).\(^8\) The distribution of \( \mathcal{T} \) conditional on \( s \), which will be denoted \( \lambda(\tau|s) \), should be interpreted as an equilibrium selection mechanism. The following assumption is maintained for the rest of the chapter.

**Assumption 2** (Equilibrium selection mechanism). \( \mathcal{T}, \mathcal{E}_1, \ldots, \mathcal{E}_N \) are mutually independent. Moreover, \( \mathcal{T}(s) \) is finite and discrete, such that \( \lambda(\tau|s) \) is a probability mass function.

Assumption 2 is also maintained by Grieco (2014) and Aguirregabiria and Mira (2016), among others. It has two important consequences. First, it ensures that players do not learn about their opponents’ private information through equilibrium selection. Second, it rules out a continuum of equilibria.

\(^8\)Rigorous notation would use \( \mathcal{T}(s) \) and \( \tau(s) \), but the argument \( s \) is dropped when referring to the random variable and its realization for simplicity.
Before turning to the econometric representation, it is worth introducing a simple stylized example related to the current chapter’s empirical application.

**Example 1** (Simple game of transceivers’ location decisions). Consider $N$ network operators simultaneously deciding how many transceivers to install in a given market. Each firm can install at most $J$ transceivers in that market, i.e. $\mathcal{Y} = \{0,1,\ldots,J\}$. In this case, strategic interactions arise due to relative network quality and therefore each player’s payoff depends on its competitors’ decisions. All firms observe some information about the market of interest, such as population density, existing stocks of transceivers, spectrum licenses ownership and different agreements (e.g., network sharing, tower sharing and domestic roaming). This information is included in $\mathcal{S}$. On top of this common-knowledge information, each firm has some private information about its costs of installing a given number of transceivers. For instance, a firm may have an agreement with the property management service of a high-rise building that makes it relatively cheap to install a transceiver at this location. While such information included in $\mathcal{E}_i$ is not known to competitors, they form beliefs about its realization.

Because of the choice set’s natural ordering, players’ decisions can be defined through an ordered-response model. Some restrictions are imposed on player $i$’s payoff function to ensure that its expected payoffs are globally concave in $y_i$ for any $s$ and $\varepsilon_i(y_i)$. First, let $\varepsilon_i(y_i) = \varepsilon_i$, $\forall y_i \in \mathcal{Y}$ and $\forall i \in \mathcal{N}$. Then, let firm $i$’s payoff function be given by:

$$\pi_i(y_i, y_{-i}, s, \varepsilon_i) = y_i \left[ \zeta_i(y_{-i}, s) + \varepsilon_i \right] - \bar{\kappa}_i(y_i)$$  \hspace{1cm} (1.7)

where strategic interactions are captured by the index function $\zeta_i(\cdot)$. Defining the expected index as $\zeta_i^P(s) = \sum_{y_{-i} \in \mathcal{Y}_{N-i}} P(y_{-i}|s) \zeta_i(y_{-i}, s)$, the corresponding expected payoff is:

$$\pi_i^P(y_i, s, \varepsilon_i) = y_i \left[ \zeta_i^P(s) + \varepsilon_i \right] - \bar{\kappa}_i(y_i) .$$  \hspace{1cm} (1.8)

Let $\kappa_i(j) = \bar{\kappa}_i(j) - \bar{\kappa}_i(j-1)$. In this case, player $i$ chooses:

$$y_i = \begin{cases} 
0 & \iff \zeta_i^P(s) + \varepsilon_i \leq \kappa_i(1) \\
0 < j < J & \iff \kappa_i(j) < \zeta_i^P(s) + \varepsilon_i \leq \kappa_i(j+1) \\
J & \iff \zeta_i^P(s) + \varepsilon_i > \kappa_i(J) .
\end{cases}$$  \hspace{1cm} (1.9)

For $\varepsilon_i$ following the logistic distribution, the CCPs have the familiar ordered logit expression:

$$P_i(j|s) = \begin{cases} 
\Lambda (\kappa_i(1) - \zeta_i^P(s)) , & \text{for } j = 0 \\
\Lambda (\kappa_i(j+1) - \zeta_i^P(s)) - \Lambda (\kappa_i(j) - \zeta_i^P(s)) , & \text{for } 0 < j < J \\
1 - \Lambda (\kappa_i(J) - \zeta_i^P(s)) , & \text{for } j = J
\end{cases}$$  \hspace{1cm} (1.10)

where $\Lambda (c) = \exp\{c\} / [1 + \exp\{c\}]$. Finally, the right-hand side of (1.10) corresponds to $\varphi_i(j, s, P_{-i}(s))$, which is used to define a BNE of the game.

**1.3.2 The econometric model**

Typically, when estimating a game of simultaneous decisions, the econometrician observes several independent repetitions of the game. For instance, one could observe players’ decisions in a large number of isolated markets.
For each of these repetitions, the researcher only observes some of the state variables that are known to all players. More formally, divide the vector \( \mathcal{S} \) in two subvectors, i.e. \( \mathcal{S} = [\mathcal{X}', \mathcal{V}]' \). Let \( \mathcal{X} \) with realizations \( \mathbf{x} \in \mathcal{X} \) be the common-knowledge state variables observable to the econometrician. \( \mathcal{V} \) with realizations \( \mathbf{v} \in \mathcal{V}(\mathbf{x}) \) is left as part of the unobservables. Notice that \( \mathcal{X} \) and \( \mathcal{V} \) do not have to be independent of each other.\(^9\) By including this vector of unobservables, one introduces some asymmetry between the researcher’s and the players’ information regarding common-knowledge payoff-relevant state variables. Of course, assumptions on \( \mathcal{S} \) stated in Assumption 1 are also imposed on \( \mathcal{V} \). In addition, the following assumption is maintained on the nature of the unobservables included in \( \mathcal{V} \).

**Assumption 3** (Common-knowledge unobservable heterogeneity). Let \( \mathcal{V} \) be an \( N \times 1 \) vector \( [\nu_1, \ldots, \nu_N]' \) with realizations \( \mathbf{v} = [\nu_1, \ldots, \nu_N]' \in \times_{i \in \mathcal{N}} \mathcal{V}_i(\mathbf{x}) = \mathcal{V}(\mathbf{x}) \), such that \( \forall i \in \mathcal{N}, \pi_i(\mathbf{y}, \mathbf{x}, \mathbf{v}, \varepsilon_i(y_i)) = \pi_i(\mathbf{y}, \mathbf{x}, \nu_i, \varepsilon_i(y_i)) \).

In other words, \( \mathcal{V} \) is a vector of common-knowledge payoff-relevant unobservables. The variable \( \mathcal{V}_i \) is therefore interpreted as common-knowledge unobservable heterogeneity specific to player \( i \). The exclusion of competitors’ unobserved heterogeneity from player \( i \)’s payoff has a natural explanation: all the relevant effect of \( \nu_{-i} \) on \( i \)'s payoff is summarized in \( y_{-i} \). Despite this exclusion, player \( i \) still considers the full vector \( \mathbf{v} \) when forming beliefs about other players’ decisions. In particular, \( \nu_{-i} \) satisfies the exclusion restriction required to identify player \( i \)’s payoff function.

Common-knowledge unobservable heterogeneity should be included in the model whenever one believes that the game of interest is not a game of pure incomplete information, i.e. when there exists some payoff-relevant information that is known to all players, but that is unobservable to the econometrician. For instance, in an industry where firms produce a fairly homogenous good, there may be some component of the cost function that is common knowledge to all players, but not to the researcher. Furthermore, specially when one is studying a case where the same few firms are involved in each repetition of the game, it is reasonable to assume that they know more about each other’s payoff function than is observable from the data. If players take this information into account when they make their decisions, omitting to control for these unobservables may bias estimation.

The identification argument proposed in the current chapter suggests using predetermined outcomes, denoted \( \mathcal{Z} = [\mathcal{Z}_1', \ldots, \mathcal{Z}_N']' \) to recover common-knowledge unobservable heterogeneity, separately from the variable indexing which equilibrium is realized in the data. The following assumption is key to achieve this goal. It states two conditions that must be satisfied by these predetermined outcome variables. Different specifications of \( \mathcal{Z} \) satisfying this assumption are suggested in Section 1.4.

**Assumption 4** (Predetermined outcomes). (i) \( \mathcal{Z} \) depends on \( \mathcal{X} \) and \( \mathcal{V} \), i.e. \( \mathcal{Z} \perp \mathcal{X}, \mathcal{V} \). (ii) \( \mathcal{Z} \) is conditionally independent of \( \mathcal{T} \) given \( \mathbf{x} \) and \( \mathbf{v} \), i.e. \( \mathcal{Z} \perp \mathcal{T} | \mathbf{x}, \mathbf{v} \).

When choosing which predetermined outcome variables to use, one must argue that these two conditions are satisfied. While the validity of the chosen outcome variables is inherently application-specific, there are some natural candidates likely to satisfy these requirements. In particular, it is easier to argue that the necessary conditional independence holds for predetermined outcomes realized before the game is played. For instance, one may take advantage of the existence of some pre-game endogenous sunk costs, such as costs of licenses required to operate in a market, franchises, market research, exploratory investments, etc. If they vary across repetitions of the game, these sunk costs are typically correlated.

\(^9\)Again, one could write \( \mathcal{V}(\mathbf{x}) \) and \( \nu(\mathbf{x}) \), but the arguments are dropped to make notation less burdensome.
with common-knowledge payoff-relevant state variables $\mathbf{X}$ and $\mathbf{V}$. Moreover, these costs being incurred before the game is played makes them more likely to be conditionally independent of $T$.

The data available to the econometrician are summarized as:

$$\{y_m, x_m, z_m : m = 1, \ldots, M\}$$

where $m$ indexes each repetition of the game and $M$ is large. In this econometric model, there are three types of unobservables with a structural interpretation.

1. The realizations of players’ private information, i.e. $\varepsilon_1, \ldots, \varepsilon_N$, which are unobservable to both the econometrician and each player’s opponents.

2. The realizations of common-knowledge unobservable heterogeneity $\nu_1, \ldots, \nu_N$, which are known to all players, but not to the econometrician. Since they enter payoff functions directly, these variables are payoff-relevant unobservables.

3. The index identifying which equilibrium is selected, i.e. $\tau$. Again, while this information is unobservable to the econometrician, it is known to all players. Notice that $\tau$ is excluded from players’ payoffs and is therefore payoff-irrelevant.

Before addressing the identification of the primitives of the model, it is worth revisiting Example 1 to fix ideas regarding the type of unobservables that correspond to $\mathbf{V}$ and the outcome variables $\mathbf{Z}$ to be used when studying transceivers’ location decisions.

**Example 1** (Simple game of transceivers’ location decisions, continued). The researcher may observe population density, firms’ initial stocks of transceivers and spectrum licenses ownership, but may fail to observe different types of agreements prevailing between network operators. Moreover, because few telecommunications companies are competing in a large number of markets and since mobile services are fairly homogenous across firms, each firm may know more about its competitors’ payoffs than the econometrician does. Let the effect of such unobservables on player $i$’s payoff be summarized by $V_i$, such that:

$$\pi_i (y_i, y_{-i}, x, \nu_i, \varepsilon_i) = y_i \left[ \zeta_i (y_{-i}, x, \nu_i) + \varepsilon_i \right] - \bar{\kappa}_i (y_i).$$

(1.12)

The corresponding expected payoff is:

$$\pi^P_i (y_i, x, \nu, \varepsilon_i) = y_i \left[ \zeta^P_i (x, \nu) + \varepsilon_i \right] - \bar{\kappa}_i (y_i)$$

(1.13)

where $\zeta^P_i (x, \nu) \equiv \sum_{y_{-i} \in Y ^{n-1} \setminus y_i} P (y_{-i}|x, \nu) \zeta_i (y_{-i}, x, \nu_i)$. Player $i$ chooses $y_i$ such that:

$$y_i = \begin{cases} 0 & \iff \zeta^P_i (x, \nu) + \varepsilon_i \leq \kappa_i (1) \\ 0 < j < J & \iff \kappa_i (j) < \zeta^P_i (x, \nu) + \varepsilon_i \leq \kappa_i (j + 1) \\ J & \iff \zeta^P_i (x, \nu) + \varepsilon_i > \kappa_i (J). \end{cases}$$

(1.14)

The resulting CCPs are given by:
An important sunk cost paid prior to installing transceivers is the cost of spectrum licenses needed to operate a network in a given geographic area. In Canada, these licenses are allocated through auctions. As it will be argued in Section 1.5, bids for spectrum licenses satisfy the two conditions defining valid predetermined outcome variables $Z$.

### 1.4 Identification of payoff functions

This section presents the main methodological result of the chapter. By recovering the realized $\nu$’s, one can explicitly control for these unobservables, which if omitted would bias the estimation. Moreover, in addition to the variation generated by different equilibria realized in the data (i.e., variation in $\tau$), the recovered $\nu$’s provide an extra source of exogenous variation required to identify the primitives of the model. Predetermined outcome variables $Z$ play a key role in recovering these unobservables.

It is worth noting that the idea of using predetermined outcomes to recover unobservables has been proposed in other settings as well. For example, Olley and Pakes (1996) use investment in capital to identify unobserved productivity in production function estimation. In education and in labour economics, test scores are often used to recover students’ unobserved ability (e.g., Carneiro, Hansen, and Heckman, 2003). Furthermore, leveraging the identifying power of exogenous variation in unobservable variables shares similarities with Matzkin (2004, 2016)’s unobservable instruments approach. For instance, consider a model with two simultaneous equations where an instrument is available for only one of these equations. Matzkin shows how to recover the unobservable random term from the equation for which a valid instrument is available and uses this unobservable as an instrument for the other equation. As in the argument I propose, this approach takes advantage of some unobservable variables’ useful exclusion properties to identify the primitives of the model.

In the current setting, the objective is to identify players’ payoff functions or, more precisely, their normalized versions. Let $\tilde{\pi}_i (y_i, y_{-i}, x, \nu_i) = [\tilde{\pi}_i (y_i, y_{-i}, x, \nu_i) - \tilde{\pi}_i (0, y_{-i}, x, \nu_i)] / \sigma (y_i)$ be player $i$’s payoff normalized with respect to the baseline action $y_i = 0$ and scaled by the standard deviation of the relative private information shocks, i.e. $\sigma^2 (y_i) = \text{Var} [\varepsilon_i (y_i) - \varepsilon_i (0)], \forall y_i \in Y \setminus \{0\}$.

A considerable difficulty associated with the identification and the estimation of empirical games is that it may be computationally very costly to find all the equilibria of a game. This problem is especially challenging when the game of interest involves several players choosing from rich choice sets. Identification and estimation arguments that avoid having to solve for different equilibria of the model are therefore very appealing in practice.

A key result that is commonly used in structural econometrics to avoid having to solve the underlying economic model is Hotz and Miller (1993)’s CCPs inversion. While it was first introduced to estimate dynamic discrete choice models, it has also been used to estimate games. Under some assumptions, this inversion allows the researcher to point-identify (normalized) expected payoff functions from the CCPs. Therefore, once one obtains an estimate of the CCPs, one can get an estimate of the (normalized)
expected payoffs without having to solve the model.

However, to apply the CCPs inversion approach one must first identify the choice probabilities conditional on \( x, \nu \) and \( \tau \). The main challenge in doing so is to be able to recover \( \nu \) separately from \( \tau \). To this end, the conditional independence between predetermined outcomes and the variable indexing which equilibrium is realized is an important identifying assumption. The identification argument I propose is based on a constructive approach which is detailed below.

### 1.4.1 Controlling for common-knowledge unobservable heterogeneity

A natural approach to identify choice probabilities conditional on both \( x \) and \( \nu \) is to first recover \( \nu \) from the observable data. The method I present consists in recovering \( \nu \) from the realized outcome variables \( z \). I propose two alternative specifications for the relationship among \( X \), \( V \) and \( Z \) that may suit different types of applications. The first one is based on a control function approach and the second one uses a factor model representation of the predetermined outcomes. Both specifications are in line with Assumption 4.

**Control function**

The control function approach to recover \( \nu \) from \( z \) is based on the following assumption.

**Assumption 5** (Control function). Let \( Z_m \equiv [Z_{1m}, \ldots, Z_{Nm}]' \) be an \( N \times 1 \) vector of observable random variables with realizations \( z_m \equiv [z_{1m}, \ldots, z_{Nm}] \in \times_{i \in N} Z_i(x) \) such that, for \( i \in N \):

\[
Z_{im} = b_i(X_m, V_{im})
\]

where \( b_i(\cdot) : X \times V_i(x) \mapsto Z_i(x) \) is a bijection in \( V_{im} \).

The main advantage of Assumption 5 is that it is straightforward to implement. In particular, \( V_m \) does not have to be explicitly recovered. One can simply replace \( V_{im} \) by a flexible function of \( X_m \) and \( Z_{im} \). However, one important drawback of the control function is that it implies a deterministic relationship between \( X \), \( V \) and \( Z \). Nonetheless, this type of approach has often been used, for instance, in production function estimation (e.g. Olley and Pakes, 1996) and in triangular systems (e.g. Newey, Powell, and Vella, 1999; Kasy, 2011; Hoderlein, Holzmann, Kasy, and Meister, 2015).

An important feature of Assumption 5 is that \( V_{-im} \) is excluded from \( b_i(\cdot) \). While it is restrictive in some cases, timing assumptions on the realization of the common-knowledge information may make it plausible in other settings. For instance, player \( i \) may know \( V_{im} \), but ignore \( V_{-im} \), when \( Z_{im} \) is determined. \( V_{-im} \) may be revealed to player \( i \) at a later stage, though still before the game is played. Such a specification \( b_i(\cdot) \) is needed in order to rewrite \( V_{im} \) as a function that does not depend on \( Z_{-im} \), such that \( Z_{-im} \) can be used instead of \( V_{-im} \) as variables satisfying the exclusion restriction needed to identify player \( i \)'s payoffs. More information on the role of the variables satisfying the exclusion restriction is provided in Section 1.4.4.

**Factor model**

In some settings, it is preferable to allow player \( i \)'s predetermined outcomes to depend on its competitors' common-knowledge payoff-relevant unobservables as well as its own. Moreover, a deterministic relationship between \( X \), \( V \) and \( Z \) may be rejected by the data if one observes several predetermined outcome
variables for each player. Using a factor model to allow for multi-dimensional unobservables addresses these two issues. Goldberger (1972) is one of the first papers to introduce factor models to economics. Since then, factor models have been used, among other things, for forecasting purposes (e.g. Stock and Watson, 2002) and to control for multidimensional unobservables when estimating treatment effects (e.g. Carneiro, Hansen, and Heckman, 2003). The required structure is summarized in Assumption 6.

**Assumption 6 (Factor model).** Let $Z_{im} = [Z_{im1}, \ldots, Z_{imK}]'$ be a $K \times 1$ vector such that, for $i \in \mathbb{N}$:

$$Z_{imk} = X'_m \alpha_{ik} + V'_m \gamma_{ik} + U_{imk}$$

where $V_m$ are interpreted as unobservable factors; $\gamma_{ik}$ are the corresponding loadings; $U_{imk}$ is an idiosyncratic shock independent across $i$’s, $m$’s and $k$’s and independent of $X_m, V_m$ and $\gamma_{ik}$.

Under Assumption 6, the common-knowledge payoff-relevant unobservable heterogeneity is interpreted as the part of the predetermined outcome variables that is not explained by $X_m$ and that is constant across $k$. Notice that the key for the separate identification of $V_m$ and $\gamma_{ik}$ is that while the former varies across $k$’s, the latter does not. Finally, the introduction of the idiosyncratic shock $U_{imk}$ makes the relationship between $X$, $V$ and $Z$ stochastic.

For a given player, the coefficients associated with $X_m$, the factors, and the factor loadings are all treated as parameters to be estimated simultaneously. An estimation method that borrows heavily from the least-squares estimator studied in Bai (2009) (related to an estimator introduced by Kiefer, 1980) is described in Appendix 1.7.1. It is worth noting that consistent estimation of the factors and their corresponding loadings requires both $M$ and $K$ to be large.

### 1.4.2 Controlling for multiplicity of equilibria

Given that $\nu$ can either be expressed as a function of observables from the control function approach (Assumption 5) or consistently estimated from the factor model (Assumption 6), one can construct a consistent estimate of $P(x, \nu)$. However, since the variable indicating which equilibrium is selected in a given repetition of the game is unobservable to the econometrician, $P(x, \nu)$ may still be mixed over multiple equilibria. In other words, the joint distribution of players’ decisions given $x$ and $\nu$, i.e. $P(y|x, \nu)$, is a finite mixture of different equilibrium-specific joint CCPs denoted by $P(y|x, \nu, \tau)$ for $\tau \in T(x, \nu)$. More formally:

$$P(y|x, \nu) = \sum_{\tau \in T(x, \nu)} P(y|x, \nu, \tau) \lambda(\tau|x, \nu)$$  \hspace{1cm} (1.16)

where the private information shocks being independent across players (Assumption 1(i)) implies that $P(y|x, \nu, \tau) = \prod_{i \in \mathbb{N}} P_i(y_i|x, \nu, \tau)$. The nonparametric identification of such finite mixtures, where each element of a random vector (in this case, $y$ conditional on $x$ and $\nu$) are independent after conditioning on the latent variable (i.e., $\tau$), has been studied by Hall and Zhou (2003), Hall, Neeman, Pakyari, and Elmore (2005), Allman, Matias, and Rhodes (2009), Kasahara and Shimotsu (2014) and Bonhomme, Jochmans, and Robin (2016), among others. Such identification results are also closely related to the measurement errors literature. See Hu (2017) for a recent survey of the literature and its applications to empirical industrial organization and labour economics. Lemma 1 below states how results from Kasahara and Shimotsu (2014) and Bonhomme, Jochmans, and Robin (2016) apply to the current setting. Some notation must first be introduced.
Let \( \{N_1, N_2, N_3\} \) be a partition of \( N \) in 3 subsets, such that \( |N_q| = N_q \) for \( q = 1, 2, 3 \) where \(||\) refers to the number of elements in the set. Moreover, let \( Y_{N_q} \) be an \( N_q \times 1 \) random vector with realizations \( y_{N_q} \in Y_{N_q} \equiv \{ y_{N_q}^1, y_{N_q}^2, \ldots, y_{N_q}^{|N_{N_q}|} \} \). For say \( q = 1, 2 \), one can construct a \( \{ |Y_{N_1}| \times |Y_{N_2}| \} \) bivariate probability matrix \( P(y_{N_q}, y_{N_2}) \) with (\( i, j \))-th entry given by \( P \left( y_{N_1}^i, y_{N_2}^j \mid x, \nu \right) \). Moreover, let \( P(y_{N_i}, y_{N_j}^q \mid x, \nu, y_{N_q}) \) be another bivariate probability matrix with (\( i, j \))-th entry given by \( P \left( y_{N_1}^i, y_{N_2}^j \mid x, \nu, y_{N_q} \right) \). Notice that for a given partition, there are therefore \( |Y_{N_1}| \) such matrices. Finally, let \( P(y_{N_q} \mid x, \nu, \tau) \) be a \( |Y_{N_q}| \times |T(x, \nu)| \) matrix with (\( i, j \))-th entry given by \( P \left( y_{N_q}^i \mid x, \nu, \tau = j \right) \).

**Lemma 1** (Identification of the equilibrium conditional choice probabilities). Fix \( x \) and \( \nu \). Suppose that \( N \geq 3 \). Consider a partition of \( N \) in 3 subsets of players. Suppose that \( |T(x, \nu)| \leq \min \{ |Y_{N_1}|, |Y_{N_2}|, |Y_{N_3}| \} \) and that \( P(y_{N_q} \mid x, \nu, \tau) \) is full column rank for \( q = 1, 2, 3 \). Then,

(i) \( |T(x, \nu)| = \text{rank} \{ P(y_{N_1}, y_{N_2} \mid x, \nu) \} \); and

(ii) \( P(y_{N_q} \mid x, \nu, \tau) \) for \( q = 1, 2, 3 \) and \( \lambda(\tau \mid x, \nu) \) for \( \tau \in \{ 1, \ldots, |T(x, \nu)| \} \) are all identified up to permutation.

**Proof.** Part (i) directly follows from Kasahara and Shimotsu (2014, Proposition 4, p. 102) and part (ii) follows from Bonhomme, Jochmans, and Robin (2016, Theorems 1-3, pp. 549-551). Appendix 1.7.2 shows how these results apply to the current setting.

### 1.4.3 Inverting the conditional choice probabilities

Given equilibrium-specific choice probabilities, one can now use the CCPs inversion approach to recover expected payoffs. Assumption 7 states extra restrictions needed to proceed. Let \( \tilde{E}_i(y) = E_i(y) - E_i(0) \) \( \forall y \in \mathcal{Y} \setminus \{0\} \).

**Assumption 7** (Known distribution of private information). \( G_{\tilde{E}_i} \left( \tilde{E}_i \right) \), i.e. the joint distribution of \( \tilde{E}_i = \left[ \tilde{E}_i(1), \ldots, \tilde{E}_i(J) \right] \), is known to the econometrician. Moreover, \( \sigma(y_i) = \sigma_i \forall y_i \in \mathcal{Y} \setminus \{0\} \).

Without loss of generality, \( \sigma_i = 1, \forall i \in N \). Denote player \( i \)'s normalized expected payoff as:

\[
\tilde{\pi}_i^P(y_i, x, \nu, \tau) = \sum_{y_{-i} \in \mathcal{Y}^{N-1}} P \left( y_{-i} \mid x, \nu, \tau \right) \tilde{\pi}_i(y_i, y_{-i}, x, \nu_i) 
\]

where \( \tau \) is included as an argument of \( \tilde{\pi}_i^P(\cdot) \) to make explicit that the CCPs used to compute the expectation are specific to equilibrium \( \tau \). The following Lemma formally states the CCPs inversion result due to Hotz and Miller (1993) as it applies to the current setting.

**Lemma 2** (Conditional choice probabilities inversion). Fix \( x \) and \( \nu \). Under Assumptions 1 to 7, there exists a one-to-one mapping between player \( i \)'s \( J \) expected normalized payoffs and its corresponding \( J \) best-response conditional choice probabilities. For \( y_i \in \mathcal{Y} \setminus \{0\} \) and \( i \in N \), this mapping is:

\[
\varphi_i(y_i, x, \nu, P_{-i}(x, \nu, \tau)) = G_{\tilde{E}_i(y_i)}(\tilde{\pi}_i^P(y_i, x, \nu, \tau))
\]

where \( G_{\tilde{E}_i(y_i)}(\cdot) \) is the marginal distribution of \( \tilde{E}_i(y_i) \). This mapping is invertible.

Chapter 1. Games with unobservable heterogeneity and multiple equilibria

Lemma 2 is very powerful. It is used to identify (normalized) expected payoffs without having to solve for the equilibria of the model, which may be computationally intensive. The identified expected payoffs can then be used to recover the corresponding payoffs, as described in the following subsection. In the simple case where \( E_i (y) \) follows the extreme value type one distribution (i.e., in the multinomial logit model), the inversion implies that:

\[
\hat{\pi}_i^P (y_i, x, \nu, \tau) = \ln \left[ \frac{P_i (y_i | x, \nu, \tau)}{P_i (0 | x, \nu, \tau)} \right]
\]  

(1.18)

The inversion corresponding to the ordered logit case is described in Example 1.

**Example 1** (Simple game of transceivers’ location decisions, continued). Similarly as before, but slightly abusing notation, let \( \hat{\pi}_i^P (y_i, x, \nu, \tau) = y_i \hat{\zeta}_i^P (x, \nu, \tau) - \kappa_i (y_i) \). In this case, for given \( x \) and \( \nu \), one wants to identify:

\[
\hat{\pi}_i^P (y_i, x, \nu, \tau) = \hat{\pi}_i^P (y_i, x, \nu, \tau) - \hat{\pi}_i^P (y_i - 1, x, \nu, \tau) = \zeta_i^P (x, \nu, \tau) - \kappa_i (y_i)
\]

\( \forall y_i \in \mathcal{Y} \setminus \{0\} \) and \( \forall \in \mathbb{N} \). For \( P_i (j | x, \nu, \tau) \) defined according to (1.15) (again including \( \tau \) as an argument to refer to equilibrium-specific CCPs), it is then easy to see that, for \( y = 1, \ldots, J \):

\[
\hat{\pi}_i^P (y, x, \nu, \tau) = \ln \left[ \frac{1 - \sum_{j=1}^{y-1} P_i (j | x, \nu, \tau)}{\sum_{j=1}^{y-1} P_i (j | x, \nu, \tau)} \right]
\]  

(1.19)

**1.4.4 Identification result**

In a simplified setting where there is no \( \mathbf{Y} \) and where the equilibrium selection mechanism is assumed to be degenerate, the only commonKnowledge payoff-relevant state variables are all included in the observable vector \( \mathbf{X} \). Given the choice probabilities conditional on \( x \), which are consistently estimated from the data, one can apply Lemma 2 to identify players’ (normalized) expected payoffs, i.e. \( \hat{\pi}_i^P (y_i, x) \). For a given \( i \in \mathbb{N} \), conditional on a specific realization \( x \):

\[
\hat{\pi}_i^P (y_i, x) = \sum_{y_{-i} \in \mathcal{Y}^{N-1}} P (y_{-i} | x) \hat{\pi}_i (y_i, y_{-i}, x)
\]  

(1.20)

where \( \hat{\pi}_i^P (y_i, x) \) \( \forall y_i \in \mathcal{Y} \setminus \{0\} \) and \( P (y_{-i} | x) \forall y_{-i} \in \mathcal{Y}^{N-1} \) are known, but \( \hat{\pi}_i (y_i, y_{-i}, x) \) \( \forall y \in \mathcal{Y}^{N-1} \times (\mathcal{Y} \setminus \{0\}) \) are unknown. Consequently, for each player, one is left with an under-identified system of \( J \) linear equations in \( J (J + 1)^{N-1} \) unknowns.

If player-specific regressors satisfying the typical exclusion restriction used to identify empirical games were available, one could use exogenous variation in the competitors’ regressors to deal with this under-identification (Pesendorfer and Schmidt-Dengler, 2003; Bajari, Hong, Krainer, and Nekipelov, 2010). In fact, exogenous variation in the competitors’ regressors shifts observable CCPs, without varying player \( i \)’s payoffs. In other words, exogenous variation in competitors’ excluded regressors increases the number of equations without increasing the number of unknowns, which eventually lead to just- or over-identification.
In the worst-case scenario, player-specific regressors satisfying the usual exclusion restriction may not be observable. The main identification result proposed here is to exploit exogenous variation in competitors’ common-knowledge unobservable heterogeneity separately from the variation generated by multiple equilibria being realized in the data to overcome the under-identification of the system of linear equations constructed from the expected payoffs. The argument is similar to the one based on the usual exclusion restriction: multiple equilibria and common-knowledge unobservable heterogeneity generate exogenous variation that shifts choice probabilities, but leaves payoffs unaffected, therefore increasing the number of equations without increasing the number of unknowns.

Let \( \mathcal{V}_{-i} (x) \equiv \times_{l \neq i \in \mathcal{N}} V_l (x) \). Proposition 1 states that each player’s normalized payoffs are identified, provided that there is enough variation in the realizations of \( T \) and/or \( \mathcal{V}_{-i} \forall i \in \mathcal{N} \).

**Proposition 1** (Identification of the normalized payoffs). Fix \( x \) and \( \nu_i \). Under Assumptions 1 to 7, \( \pi_i (y_i, y_{-i}, x, \nu_i) \) is identified \( \forall y_i \in Y \setminus \{0\} \), \( \forall y_{-i} \in Y^{N-1} \) and \( \forall i \in \mathcal{N} \) provided that there is sufficient variation in \( P (y_{-i} | x, \nu, \tau) \) across \( \tau \) and \( \mathcal{V}_{-i} \forall i \in \mathcal{N} \) and \( \sum_{\nu_{-i} \in \mathcal{V}_{-i} (x)} | T (x, \nu_i, \nu_{-i}) | \geq (J + 1)^{N-1} \).

Proof. Consider the normalized expected payoff in (1.17). In this equation, the left-hand side and the CCPs on the right-hand side are treated as known since they can be identified from the data using Lemmas 1 and 2. The only unknowns are the normalized payoffs on the right-hand side. Therefore, for fixed \( x \) and \( \nu_i \), and for each \( i \in \mathcal{N} \), (1.17) can be used to construct a system of \( J \sum_{\nu_{-i} \in \mathcal{V}_{-i} (x)} | T (x, \nu_i, \nu_{-i}) | \) linear equations (i.e., the normalized expected payoffs for player \( i \)'s \( J \) non-normalized decisions, at a total of \( J \sum_{\nu_{-i} \in \mathcal{V}_{-i} (x)} | T (x, \nu_i, \nu_{-i}) | \) different equilibria) in \( J (J + 1)^{N-1} \) unknowns (i.e., player \( i \)'s normalized payoffs for \( J (J + 1)^{N-1} \) different realizations of \( y \)). For each player, identification therefore requires \( \sum_{\nu_{-i} \in \mathcal{V}_{-i} (x)} | T (x, \nu_i, \nu_{-i}) | \geq (J + 1)^{N-1} \) (with over-identification if the inequality is not binding). Sufficient variation in the CCPs is needed to make sure the system of equations satisfies the rank condition. This condition can be checked since the CCPs on the right-hand side of (1.17) are identified.

The identifying power of exogenous variation in \( \nu_{-i} \) and \( \tau \) is made obvious by the inequality condition stated in Proposition 1. As the number of players increases and the choice set becomes richer, more variation in the number of equilibria and/or common-knowledge unobservable heterogeneity is needed to achieve identification. Notice that if there was no \( \mathcal{V} \) included in the model (as in Sweeting, 2009), the condition would simplify to \( | T (x) | \geq (J + 1)^{N-1} \) and, therefore, many equilibria would be required for identification of games with a large number of players choosing from a rich choice set. The gains from introducing \( \mathcal{V} \) can therefore be substantial. In particular, even if the equilibrium selection mechanism is degenerate, variation in \( \nu_{-i} \) may still be enough to identify the primitives of the model.

It is also worth noting that the inequality in Proposition 1 would be slightly different in the ordered-response model used in Example 1. While the number of equations is the same as in the multinomial setting, the number of unknowns to identify for fixed \( x \) and \( \nu_i \) for each \( i \in \mathcal{N} \) is \( (J + 1)^{N-1} + J \) (i.e., player \( i \)'s \( (J + 1)^{N-1} \) index functions \( \zeta_i (y_{-i}, x, \nu_i) \) and \( J \) threshold parameters \( \kappa_i (y) \)). The inequality therefore becomes \( \sum_{\nu_{-i} \in \mathcal{V}_{-i} (x)} | T (x, \nu_i, \nu_{-i}) | \geq \frac{(J+1)^{N-1}}{J} + 1 \).

### 1.5 Application: transceivers’ locations in Canada

I now turn to the application of the method presented above to study transceivers’ location decisions by mobile service providers in Canada. The primary focus is on the three players that offer nationwide
coverage (Rogers, Bell and Telus, henceforth referred to as the incumbents) and two new entrants (Vidéotron and Wind).

1.5.1 Industry background

Before laying out the empirical analysis, it is worth discussing some background information. This section is far from being a comprehensive description of the Canadian mobile telecommunications industry. Instead, I report relevant facts to justify the model used.

The main reason for focusing on transceivers’ location decisions is to assess the properties of the Canadian mobile telecommunications industry at the wholesale level. The wholesale market in mobile telecommunications relates to transactions through which a mobile service provider allows another firm to use parts of its network and/or its installations to provide network services. For instance, such transactions include domestic roaming, tower-sharing and network-sharing agreements which, as mentioned in Section 1.2, have raised concerns about incumbents’ market power.

One of the main policies used by the federal government to encourage entry in the mobile telecommunications industry has been setting aside some blocks of spectrum for new and regional players. In fact, prior to operating a network, mobile service providers must first obtain region-specific licenses to use spectrum. Since 2001, such licenses have been allocated through auctions (see Industry Canada, 2004, 2011). Due to spectrum set-asides, incumbents have only been allowed to bid on a subset of the available frequencies in some of these auctions.

On one hand, spectrum set-asides have allowed new and regional service providers to acquire parts of the spectrum and to operate a network. For instance, among new entrants associated with the 2008 auction set-asides are Vidéotron, Wind, Eastlink, Mobilicity and Public Mobile. On the other hand, two drawbacks of spectrum set-asides are worth mentioning. First, some of the new entrants do not use all the spectrum they are assigned and eventually sell unused frequencies to other firms, including incumbents. Second, some of these firms go out of business and are acquired by incumbents. In fact, upon Industry Canada’s and the Competition Bureau’s approval, Telus bought Public Mobile and Rogers acquired Mobilicity.\(^\text{10}\)

The main objective of the current empirical application is to provide an explanation for these drawbacks. The identification argument presented above is well suited to study the problem at hand. The need to include common-knowledge unobservable heterogeneity is namely justified by agreements between firms. The available information regarding these agreements is typically quite sparse. It is therefore natural to treat this payoff-relevant information as being common-knowledge to players, but unobservable to the econometrician. Since the number of firms competing in this industry is quite small, it is more realistic to assume that these firms know more about each other’s costs and payoffs than the econometrician does. Furthermore, given the relatively rich choice set firms can choose from and the potentially large number of equilibria that may arise in the economic model, assuming that the same equilibrium is played in every repetition of the game may be hard to defend \textit{a priori}.

\(^{10}\)See the statements of approval given in Industry Canada (2013b, 2015) and Competition Bureau (2014).
1.5.2 Empirical model

Sequence of simultaneous decisions

Modelling current stocks of transceivers owned by each provider as a one-shot game of simultaneous decisions would be unrealistic. In Canada, the first cellular network came into service in the mid-1980’s. Since then, there have been many important changes in terms of technology, regulation and market structure. It would be simplistic to assume that any firm would commit to their decisions over such a long period of time despite all these changes.

A more realistic view is to consider transceivers’ locations as resulting from a sequence of games of simultaneous decisions. Different spectrum auctions define the timing of each game. More precisely, I model providers’ decisions as an investment problem. After an auction is completed, firms that have been granted licenses simultaneously choose how many transceivers to add to their accumulated stocks of transceivers. Each player’s payoff is affected by its own and its competitors’ current stocks of transceivers, as well as the simultaneously chosen additions.

Dynamics are therefore included in the model by allowing current payoffs to be affected by previous decisions through stock accumulation. In other words, existing stocks of transceivers are part of the state variables, such that past investments affect current decisions even if transceivers’ installation is not explicitly modelled as a dynamic game. The resulting payoff functions should be interpreted as a semi-reduced-form of the sum of the current and discounted expected future valuations.

There are two important reasons for not explicitly modelling forward-looking behaviour through a dynamic game. First, I only observe a single auction with relevant strategic interactions between incumbents and new entrants. At this stage, as will be described in subsection 1.5.3, subsequent auctions are arguably too recent to interpret transceivers’ locations as equilibrium investment decisions. Second, it is not clear to what extent firms are able to anticipate future regulatory and technological changes. In some sense, interpreting payoff functions as semi-reduced-form value functions amounts to being fairly agnostic about firms’ anticipation of the future.

Auctions for spectrum licences

Auctioned licenses allow service providers to use a block of frequencies for a given geographic area. The size of the geographic area covered by a license is determined by which tier the license belongs to. Industry Canada uses three different types of tier when auctioning mobile spectrum licenses. Each type of tier partitions the whole country in regions of different sizes. There are 14 tier-2 regions, 59 tier-3 regions and 172 tier-4 regions. Tier-2 regions are collections of tier-3’s, which are collections of tier-4’s. There is more than one block of spectrum in each of these size-varying regions, such that competitors simultaneously operate in the same geographic area, but on different frequencies.

Using spectrum auctions when defining the timing of the game has two advantages. First, since providers must bid for licenses prior to setting up a network, they must plan their transceivers’ locations, at least to some extent, before knowing their competitors’ decisions. In other words, the need to simultaneously plan transceivers’ locations before the spectrum is auctioned argues in favour of modelling the investment problem as one involving simultaneous decisions.

The second advantage of using auction-specific games of simultaneous decisions is that providers’ bids can be used as predetermined outcome variables to control for common-knowledge payoff-relevant unobserved heterogeneity. Even if firms’ bidding behaviour is not interpreted through a structural
model, bids should reflect players’ valuations of operating a network including common-knowledge payoff-relevant unobservables. As a result, players’ bids depend on both $\mathbf{X}$ and $\mathbf{V}$. Moreover, these outcome variables are determined before transceivers’ location decisions. In particular, at the bidding stage, firms do not know who will successfully acquire which frequency blocks and, therefore, do not know which equilibrium will be played. For this reason, bids are considered to be conditionally independent of the equilibrium selection variable.

**Isolated markets**

As stated in Section 1.3, the identification result requires the econometrician to observe a large number ($M$) of independent repetitions of the game. I consider location decisions in different isolated markets as being different repetitions of the transceivers’ location game. Population centres, i.e. relatively more densely populated areas, are used as isolated markets.

A nice feature of population centres is that they are physically isolated and surrounded by rural areas. Of course, it is likely that there is some dependence between transceivers’ locations within and around a given population centre. For instance, providers may want to ensure that their subscribers can use their mobile devices when they travel in rural areas around where they live or work. By focusing on population centres, I avoid having to model this potential correlation between population centres and their surrounding rural areas.

**Ordered-response model**

Transceivers’ location decisions are studied through an ordered-response model similar to the model introduced in Example 1. Consider $N = 5$ firms (Rogers, Bell, Telus, Vidéotron and Wind). Following a spectrum auction, they simultaneously decide how many transceivers to install in a given market, i.e. $\mathbf{Y}_i$ is the addition to player $i$’s stocks of transceivers. This quantity is discretized such that $\mathbf{Y} = \{0, y^1, \ldots, y^J\}$, where $y^1, \ldots, y^J$ defines an arbitrary partition of the positive numbers of transceivers installed by players.

Each player’s stock of transceivers accumulated prior to the auction of interest are included in $\mathbf{x}$. Notice that stocks and additions of transceivers are not treated as substitutes. In particular, there can be substantial heterogeneity in spectrum quality across auctions. Moreover, to the extent that relative network quality is an important determinant of players’ payoffs and since both stocks and additions affect relative network quality, all players’ stocks of transceivers are included in player $i$’s payoffs. As a result, even if stocks are player-specific regressors, they are not assumed to satisfy the usual exclusion restriction needed to identify normalized payoffs.

Other payoff-relevant market characteristics may also be included in $\mathbf{x}$. Then, in line with the ordered-response model in Example 1, the equilibrium-specific CCPs are defined as in (1.15). The objective is to estimate: (i) the index functions $\zeta_i (\mathbf{y}_{-i}, \mathbf{x}, \nu_i) \forall \mathbf{y}_{-i} \in \mathbf{Y}_{i-1}, \forall \mathbf{x} \in \mathbf{X}, \forall \nu_i \in \mathbf{V}_i (\mathbf{x}), \forall i \in N$; and (ii) the thresholds $\kappa_i (j) \forall i \in N$ and $\forall j \in \mathbf{Y} \setminus \{0\}$.

1.5.3 Data

**Transceivers**

Data about transceivers’ locations are constructed using the Technical and Administrative Frequency Lists maintained by Industry Canada. I observe a 2016 snapshot of all transceivers associated with
mobile services in Canada. For each transceiver, I observe who owns it, its geographic location and the frequencies on which it operates, among other characteristics.

Frequencies are used to link transceivers to different auctions. All transceivers operating on frequencies allocated prior to a given auction are considered to be stocks at this auction; transceivers operating on frequencies allocated during this auction are additions. Implicitly, since I do not observe the timing of transceivers’ installation, I am assuming that all transceivers operating on frequencies allocated during a given auction are installed simultaneously. Therefore, I am ignoring some of the dynamics that could be associated with delays in towers’ construction and/or spectrum transfers between providers.

Population centres as isolated markets

I use Statistics Canada’s population centres as isolated markets. Population centres are defined as areas with a minimum population of 1,000 people and a minimum population density of 400 persons per square kilometre. Any territory outside population centres is treated as a rural area. I restrict the analysis to population centres located in Alberta, British Columbia, Ontario and Quebec since these are the provinces where all 5 players own spectrum licenses. Overall, there are \( M = 722 \) population centres in these four provinces (out of 942 population centres for the whole country).

Geographic coordinates are used to match transceivers with the market they belong to. For each market, one can count how many transceivers are associated with each provider. Additions are binned in four different categories (i.e., \( J = 3 \)): 0, 1-10, 11-50 and more than 50 transceivers. A non-trivial share of markets, specially the smaller ones, are such that there is no transceiver within the boundary of the population centre, but some right outside of it. I therefore consider all transceivers within a 5 kilometers buffer zone around a population centre as being part of that population centre.

As mentioned earlier, an important factor determining the need for more transceivers in a given area is the population density in this area. This variable is also observed for each population centre.

The Advanced Wireless Service 1 auction

The focus of the current application is the 2008 Advanced Wireless Service 1 (AWS1-2008) auction, which was the first auction during which the two new entrants considered in the empirical application (Vidéotron and Wind) obtained licenses required to set up a network. Figure 1.1 in Section 1.2 shows the location of each player’s transceivers operating on frequencies allocated up to the AWS1-2008 auction. In other words, these maps include both stocks (for incumbents) and additions of transceivers associated with this auction.

Between the AWS1-2008 auction and the construction of the dataset used for estimation, there have been three other spectrum auctions relevant for mobile services: the 2014 Mobile Broadband Service (MBS-2014), the 2015 Advanced Wireless Service 3 (AWS3-2015) and the 2015 Broadband Radio Service (BRS-2015) auctions. However, these auctions are excluded from the analysis since they may be too recent for network services providers to have had sufficient time to install new transceivers. In fact, for each of these auctions, there is at least one player (incumbent and/or new entrant) that has not installed a single transceiver operating on its allocated frequencies. In particular, none of the mobile service providers has installed transceivers operating on AWS3-2015 frequencies according to the 2016 dataset

\[11\] Notice that spectrum owned by Vidéotron in Alberta and British Columbia has been allocated during the 2014 Mobile Broadband Service auction, i.e. after the auction of interest. As mentioned below, the analysis controls for license ownership when estimating the model.
I use. Notice that the AWS3-2015 frequencies were sold in 2015 even if no mobile device available in Canada were operating on these frequencies at that time.

Frequencies from 8 different blocks (denoted A, B, C, D, E, F, G and I) were allocated during the AWS1-2008 auction. Blocks B, C and D were set aside for new entrants and regional players only. Two different sizes of tiers were used: tier 2 for blocks B, C, G and I; tier 3 for blocks A, D, E and F. A license is auctioned for a given block, in a given tier. A total of 282 licenses were allocated during this auction, with winning bids summing to approximately CAD$ 4.3 billion.

Population centres are matched with their corresponding tier-2 and -3 regions. There are 8 tier-2 regions and 46 tier-3’s in the four provinces of interest. Because tiers are usually larger than population centres, there are several markets in each tier region. By combining license allocations resulting from the auction with public information regarding licenses transfers, I construct variables indicating whether or not each provider owns some spectrum license in each market. License ownership is used to determine potential entrants.

Finally, I obtained each players’ bids in each round of the AWS1-2008 auction from Industry Canada. Players’ highest bids for each license are used to construct predetermined outcomes \( Z \). Two features of the bids are worth mentioning. First, I observe both winning and non-winning bids. In particular, observing non-winning bids is helpful to recover information about unobservable heterogeneity for all players, i.e. not only winners. Second, firms typically bid on multiple blocks of spectrum in a given tier. Therefore, several predetermined outcomes are available for each player.

### 1.5.4 Estimation

Given the relatively small sample available (718 markets), the estimation method used here differs from the semi-parametric identification argument presented above. In particular, some parametric assumptions are made when estimating reduced-form CCPs and payoff functions. However, as suggested in Section 1.4, these assumptions could be relaxed if one observed sufficient variation in the (recovered) unobservable heterogeneity and/or multiple equilibria, together with a large enough number of markets to consistently estimate CCPs non-parametrically.

Estimation proceeds in three stages. First, a factor model is estimated for each player in order to recover player-specific common-knowledge unobservable heterogeneity from firms’ bids. Second, equilibrium-specific reduced-form choice probabilities are obtained via a relatively flexible specification, taking the unobserved heterogeneity recovered from the first stage as given. Third, expected payoffs are constructed to estimate semi-reduced-form payoffs and threshold parameters defining players’ choices. Standard errors of the third stage’s estimates are computed from bootstrapped samples to take into account sampling variation from the first two stages’ estimates.

#### Estimation of the factor model

I use bids on frequency blocks A, D, E and F to recover player-specific unobservable heterogeneity. More precisely, for each license, \( Z_{ik} \) is constructed as player \( i \)’s bid divided by the maximum bid for this

---

12 For detailed information about the AWS1-2008 auction, see Industry Canada (2007) and Industry Canada’s webpage on this specific auction (http://www.ic.gc.ca/eic/site/smt-gst.nsf/eng/h_sf08891.html).

13 Blocks B, C, G and I are dropped since their allocation used tier 2 regions, which implies at most 8 observations per block for each player. Block D is also dropped for the incumbents since it is part of the blocks set aside in the AWS1-2008 auction.
license.\textsuperscript{14} Notice that bids vary at the tier level, not at the market level. It follows that the recovered common-knowledge payoff-relevant unobserved heterogeneity should be interpreted as being constant across markets belonging to the same tier. Each player’s factor model is therefore estimated using 46 tiers with $K = 3$ (incumbents\textsuperscript{15}) or $K = 4$ (new entrants) blocks of spectrum to estimate 2 factors ($\mathcal{V}_i$ and $\tilde{\mathcal{V}}_{-i}$, where $\mathcal{V}_i$ is assumed to correspond to the factor explaining most of the variance in player $i$’s bids).

Table 1.1 is the correlation matrix for the estimated $\nu_i$’s obtained separately from each player’s factor model. Correlations are reported in absolute value since factors are identified up to a column-wise change of sign. In particular, it is interesting to see that the largest correlation is between Bell and Telus. In fact, their network-sharing agreement is arguably the most substantial agreement currently prevailing in the industry. Neither are the fairly high correlations between new entrants and some of the incumbents surprising since Vidéotron and Wind typically have domestic roaming agreements with some of the larger firms.

<table>
<thead>
<tr>
<th></th>
<th>Rogers</th>
<th>Bell</th>
<th>Telus</th>
<th>Vidéo</th>
<th>Wind</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rogers</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bell</td>
<td>0.07</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Telus</td>
<td>0.20</td>
<td>0.52</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vidéo</td>
<td>0.44</td>
<td>0.30</td>
<td>0.42</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Wind</td>
<td>0.38</td>
<td>0.01</td>
<td>0.14</td>
<td>0.11</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Notes: Coefficients of correlation between estimated player-specific common-knowledge unobservable heterogeneity. Unobservable heterogeneity is estimated separately for each player using a model with two factors. Correlations are in absolute value since factors are identified up to a column-wise change of sign.

### Estimation of the reduced-form conditional choice probabilities

Given the estimated unobservable heterogeneity, one can construct choice probabilities conditioned on both $\mathbf{x}$ and $\hat{\nu}$. These probabilities are still potentially mixed over multiple equilibria. In order to apply the non-parametric estimator based on the joint diagonalization of a sequence of matrices proposed by Bonhomme, Jochmans, and Robin (2016), one would need non-parametric estimates of the mixed marginal and joint conditional choice probabilities. In very large samples, where one observes many markets with similar realizations $\mathbf{x}$ and $\hat{\nu}$, such non-parametric estimates are easily obtained. However, in small samples, nonparametric estimators are heavily biased.

Alternatively, I use the Expectation-Maximization (EM) algorithm (Dempster, Laird, and Rubin, 1977) together with a flexible ordered logit to estimate the equilibrium-specific reduced-form CCPs. More precisely, consider $T$ types of equilibrium realized in the data with probability $\omega_\tau$, $\tau = 1, \ldots, T$. Let $h (\mathbf{x}, \hat{\nu}; \phi_{i\tau})$ be a parametric function of $\mathbf{x}$, $\hat{\nu}$ with a vector of parameters $\phi_{i\tau}$ specific to each player and each equilibrium. In the current application, $h (\mathbf{x}, \hat{\nu}; \phi_{i\tau})$ is a simple first-order polynomial, but higher orders could be accommodated with a larger number of observations. Letting $\xi_{i\tau} = [\xi_{i\tau} (1), \ldots, \xi_{i\tau} (J)]'$

\textsuperscript{14}In cases where the maximum bid is 0, $Z_{ik} = 0$.
\textsuperscript{15}The only exception is Rogers, for which $K = 2$. In fact, block A is dropped for Rogers since it is the highest bidder on all licenses.
be threshold parameters (also specific to each player and each equilibrium), the reduced-form equilibrium choice probability of player $i$ choosing $y_i = j$, which is denoted as $P_i \left( j \mid x, \hat{\nu}, \tau; \phi_{i\tau}, \xi_{i\tau} \right)$, is given by:

$$
\begin{align*}
\begin{cases}
\Lambda \left( \xi_{i\tau} (1) - h \left( x, \hat{\nu}; \phi_{i\tau} \right) \right), & \text{for } j = 0 \\
\Lambda \left( \xi_{i\tau} (j + 1) - h \left( x, \hat{\nu}; \phi_{i\tau} \right) \right) - \Lambda \left( \xi_{i\tau} (j) - h \left( x, \hat{\nu}; \phi_{i\tau} \right) \right), & \text{for } 0 < j < J \\
1 - \Lambda \left( \xi_{i\tau} (J) - h \left( x, \hat{\nu}; \phi_{i\tau} \right) \right), & \text{for } j = J.
\end{cases}
\end{align*}
$$

(1.21)

Let $\phi = [\phi_{11}', \ldots, \phi_{NT}']'$, $\xi = [\xi_{11}', \ldots, \xi_{NT}']'$ and $\omega = [\omega_1, \ldots, \omega_T]'$. The contribution of market $m$ to the likelihood function is given by:

$$
L_m \left( y_{m} \mid x, \hat{\nu}; \phi, \xi, \omega \right) = \sum_{\tau=1}^{T} \prod_{i \in \tilde{N}_m} \prod_{j=0}^{J} \omega_{\tau} P_i \left( y_{im} = j \mid x_{m}, \hat{\nu}_m, \tau; \phi_{i\tau}, \xi_{i\tau} \right) \mathbf{1} \left\{ y_{im} = j \right\}
$$

(1.22)

where $\tilde{N}_m$ is the set of providers owning spectrum licenses in market $m$. Estimates of $\phi$, $\xi$ and $\omega$ are computed by using the EM algorithm, allowing for $T = 2$ types of equilibrium. The estimated weights associated with each type of equilibrium are $\hat{\omega}_1 = 0.45$ and $\hat{\omega}_2 = 0.55$, providing some evidence that multiple equilibria are realized in the data.

**Estimation of the payoff parameters**

Given the relatively small sample size at hand, a parametric specification is proposed for the index functions $\zeta_i(\cdot)$:

$$
\zeta_i \left( y_{-i}, x, \hat{\nu}; \theta_i \right) = x' \beta_i + \sum_{n \in \tilde{N} \setminus \{i\}} \delta_{in} y_n + \eta_i \hat{\nu}_i
$$

(1.23)

where $\theta_i = [\beta_i, \{\delta_{in}\}_{n \neq i}, \eta_i]'$ is the vector of payoff parameters for firm $i$. Given the estimated equilibrium-specific CCPs, which are denoted by $\hat{P}_i \left( y_i \mid x, \hat{\nu}, \tau \right) = P_i \left( y_i \mid x, \hat{\nu}, \tau; \phi_{i\tau}, \xi_{i\tau} \right)$, one can construct the equilibrium-specific expected index:

$$
\zeta_i^{P} \left( x, \hat{\nu}, \tau; \theta_i \right) = x' \beta_i + \sum_{n \in \tilde{N} \setminus \{i\}} \sum_{j=0}^{J} \delta_{in} j \hat{P}_n \left( j \mid x, \hat{\nu}, \tau \right) + \eta_i \hat{\nu}_i.
$$

(1.24)

Let $\kappa_i = [\kappa_i (1), \ldots, \kappa_i (J)]'$. The structural equilibrium-specific probability of player $i$ choosing $y_i = j$, denoted $P_i \left( j \mid x, \hat{\nu}, \tau; \theta_i, \kappa_i \right)$, is therefore:

$$
\begin{align*}
\begin{cases}
\Lambda \left( \kappa_i (1) - \zeta_i^{P} \left( x, \hat{\nu}, \tau; \theta_i \right) \right), & \text{for } j = 0 \\
\Lambda \left( \kappa_i (j + 1) - \zeta_i^{P} \left( x, \hat{\nu}, \tau; \theta_i \right) \right) - \Lambda \left( \kappa_i (j) - \zeta_i^{P} \left( x, \hat{\nu}, \tau; \theta_i \right) \right), & \text{for } 0 < j < J \\
1 - \Lambda \left( \kappa_i (J) - \zeta_i^{P} \left( x, \hat{\nu}, \tau; \theta_i \right) \right), & \text{for } j = J.
\end{cases}
\end{align*}
$$

(1.25)

Let $\theta = [\theta_1', \ldots, \theta_N']'$ and $\kappa = [\kappa_1', \ldots, \kappa_N']'$. Using the equilibrium weights obtained from the estimation of the reduced-form CCPs, $\hat{\omega}$, market $m$’s contribution to the likelihood is therefore:

$$
L_m \left( y_{m} \mid x, \hat{\nu}; \theta, \kappa \right) = \sum_{\tau=1}^{T} \prod_{i \in \tilde{N}_m} \prod_{j=0}^{J} \hat{\omega}_{\tau} P_i \left( y_{im} = j \mid x_{m}, \hat{\nu}_m, \tau; \theta_i, \kappa_i \right) \mathbf{1} \left\{ y_{im} = j \right\}.
$$

(1.26)
Maximum likelihood estimates of $\theta$ and $\kappa$ are reported in Tables 1.2 and 1.3.

Table 1.2: Ordered-response estimates for control variables and thresholds

<table>
<thead>
<tr>
<th></th>
<th>Rogers</th>
<th>Bell</th>
<th>Telus</th>
<th>Vidéo</th>
<th>Wind</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population density</td>
<td>0.057</td>
<td>-0.014</td>
<td>0.846***</td>
<td>0.050</td>
<td>0.488</td>
</tr>
<tr>
<td>(0.382)</td>
<td>(0.355)</td>
<td>(0.308)</td>
<td>(0.358)</td>
<td>(0.532)</td>
<td></td>
</tr>
<tr>
<td>Stocks of transceivers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rogers</td>
<td>2.048***</td>
<td>0.562</td>
<td>-0.176</td>
<td>0.256</td>
<td>1.537</td>
</tr>
<tr>
<td>(0.225)</td>
<td>(0.511)</td>
<td>(0.351)</td>
<td>(0.919)</td>
<td>(1.169)</td>
<td></td>
</tr>
<tr>
<td>Bell</td>
<td>0.158</td>
<td>1.717***</td>
<td>0.023</td>
<td>0.403</td>
<td>0.921**</td>
</tr>
<tr>
<td>(0.178)</td>
<td>(0.317)</td>
<td>(0.220)</td>
<td>(0.335)</td>
<td>(0.454)</td>
<td></td>
</tr>
<tr>
<td>Telus</td>
<td>-0.636*</td>
<td>-1.063</td>
<td>2.227***</td>
<td>0.725</td>
<td>0.127</td>
</tr>
<tr>
<td>(0.366)</td>
<td>(0.675)</td>
<td>(0.226)</td>
<td>(0.476)</td>
<td>(0.772)</td>
<td></td>
</tr>
<tr>
<td>Unobserved heterogeneity</td>
<td>-0.432***</td>
<td>-0.405**</td>
<td>-0.318</td>
<td>-1.472</td>
<td>0.963***</td>
</tr>
<tr>
<td>(0.141)</td>
<td>(0.174)</td>
<td>(0.347)</td>
<td>(1.325)</td>
<td>(0.402)</td>
<td></td>
</tr>
<tr>
<td>Threshold parameters</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-10</td>
<td>2.609***</td>
<td>8.056***</td>
<td>6.789***</td>
<td>5.373***</td>
<td>9.350***</td>
</tr>
<tr>
<td>(0.526)</td>
<td>(0.941)</td>
<td>(0.711)</td>
<td>(1.168)</td>
<td>(1.824)</td>
<td></td>
</tr>
<tr>
<td>(0.632)</td>
<td>(0.977)</td>
<td>(0.824)</td>
<td>(1.189)</td>
<td>(1.970)</td>
<td></td>
</tr>
<tr>
<td>(1.375)</td>
<td>(1.228)</td>
<td>(1.220)</td>
<td>(1.466)</td>
<td>(2.397)</td>
<td></td>
</tr>
</tbody>
</table>

Notes: Estimates of the coefficients corresponding to control variables and threshold parameters for the ordered-response model where players simultaneously choose between installing 0, 1-10, 11-50 or 51+ transceivers operating on frequencies allocated in 2008. Parameters are player-specific and each column refers to a different player. Population density is in $10^3$ persons per square kilometre. Stocks of transceivers, in natural logarithm (after adding 1 to avoid 0’s), are incumbents’ transceivers operating on frequencies allocated before 2008. Variables corresponding to player-specific unobserved heterogeneity estimated using factor models are included as regressors. Values in brackets are standard errors computed using 99 bootstrapped samples. Significance levels: * = 0.10, ** = 0.05 and *** = 0.01.

An important result that is worth mentioning is the relative contribution of each incumbent’s own stock of transceivers to their payoffs. The corresponding coefficients (respectively 2.048, 1.717 and 2.227 for Rogers, Bell and Telus) suggests that increasing a provider’s own accumulated stock of transceivers has a positive and significant effect on his payoffs. In other words, transceivers’ location decisions may be associated with important economies of density: incumbents are more likely to add new transceivers where they already own some. Intuitively, it may be less costly for providers to improve an existing network rather than building a brand-new one. In that sense, adding transceivers is more profitable for an incumbent than for a new entrant, keeping everything else constant. This observation potentially explains why new entrants fail to use some of their spectrum or go out of business. In fact, new entrants cannot benefit from such economies of density, since they are building a brand-new network without having accumulated a stock of transceivers. This asymmetry between incumbents and new entrants motivates the counterfactual experiments proposed in Section 1.5.5.

Threshold parameters for all players are also significantly different from 0 at the 1% significance level in all cases. Even if estimated payoff functions have the semi-reduced-form interpretation of capturing both the current and the discounted expected future valuation of adding transceivers, threshold parameters still have a structural interpretation. They provide information about each firm’s effective costs of installing a given number of transceivers, with higher values of thresholds suggesting higher costs.

In this vein, there are interesting similarities and differences across players. First of all, Rogers’ effective costs seem to be quite different from the other providers. Its relatively low threshold for installing
Table 1.3: Ordered-response estimates for strategic interactions

<table>
<thead>
<tr>
<th></th>
<th>Rogers</th>
<th>Bell</th>
<th>Telus</th>
<th>Vidéo</th>
<th>Wind</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rogers</td>
<td></td>
<td>0.930</td>
<td>0.915</td>
<td>1.400</td>
<td>-0.560</td>
</tr>
<tr>
<td></td>
<td>(–)</td>
<td>(1.547)</td>
<td>(1.028)</td>
<td>(2.206)</td>
<td>(2.586)</td>
</tr>
<tr>
<td>Bell</td>
<td>0.816**</td>
<td>–</td>
<td>0.798</td>
<td>0.285</td>
<td>–0.326</td>
</tr>
<tr>
<td></td>
<td>(0.382)</td>
<td>(–)</td>
<td>(0.513)</td>
<td>(0.617)</td>
<td>(0.721)</td>
</tr>
<tr>
<td>Telus</td>
<td>1.976**</td>
<td>1.555</td>
<td>–</td>
<td>-0.479</td>
<td>0.537</td>
</tr>
<tr>
<td></td>
<td>(0.834)</td>
<td>(1.476)</td>
<td>(–)</td>
<td>(1.007)</td>
<td>(1.784)</td>
</tr>
<tr>
<td>Vidéo</td>
<td>1.322**</td>
<td>-0.775</td>
<td>0.230</td>
<td>–</td>
<td>1.050</td>
</tr>
<tr>
<td></td>
<td>(0.653)</td>
<td>(0.628)</td>
<td>(0.699)</td>
<td>(–)</td>
<td>(2.031)</td>
</tr>
<tr>
<td>Wind</td>
<td>1.023</td>
<td>-0.080</td>
<td>-0.633</td>
<td>-4.002</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>(0.740)</td>
<td>(0.614)</td>
<td>(0.785)</td>
<td>(3.597)</td>
<td>(–)</td>
</tr>
</tbody>
</table>

Notes: Estimates of the parameters corresponding to strategic interactions for the ordered-response model where players simultaneously choose between installing 0, 1-10, 11-50 or 51+ transceivers operating on frequencies allocated in 2008. The estimate in row $i$ and column $j$ corresponds to the effect of player $i$’s decision on player $j$’s payoffs. Values in brackets are standard errors computed using 99 bootstrapped samples. Significance levels: $*$ = 0.10, $**$ = 0.05 and $***$ = 0.01.

1 to 10 transceivers (2.609, compared with estimates between 5.373 and 9.350 for other providers) and relatively high threshold for more than 50 transceivers (15.802 compared with estimated thresholds varying between 11.096 and 13.535) suggests that it installs transceivers in more markets, but typically installs fewer transceivers than its competitors in each market. Second, the effective costs of transceivers’ additions are very similar for Bell and Telus, specially for 11-50 transceivers (9.613 vs 9.214) and 51+ transceivers (13.535 vs 13.468). In some sense, this is not surprising, given that these two providers are known to have reached some network-sharing agreements. Interestingly, the estimated threshold parameters suggest that these agreements share the costs of installing new transceivers roughly equally between the two providers. Finally, there are important differences across new entrants’ effective costs, Vidéotron’s being relatively lower than Wind’s. In particular, Vidéotron’s estimated threshold associated with 1-10 transceivers is much lower than Wind’s (5.373 vs 9.350).

1.5.5 Counterfactual experiments: subsidizing transceivers

As already mentioned, new entrants do not benefit from economies of density generated by accumulated stocks of transceivers since they must build brand-new networks. Taking as given the government’s objective to increase competition in the telecommunications industry and the drawbacks associated with spectrum set-asides, this asymmetry between incumbents and new entrants may justify subsidizing the first few transceivers installed by the latter.

I use the estimated structure of the model to provide valuable insights regarding the effects of such a policy. Intuitively, subsidizing new entrants’ transceivers should increase the probability that they install some transceivers in a given market. This change in probability of installing transceivers will also have an effect on the incumbents’ investment probabilities since, in equilibrium, these are defined as best responses to their competitors’ expected behaviour. This indirect effect of the subsidy would not be captured by a model ignoring strategic interactions between network operators.

Counterfactual experiments are performed separately for three average markets, respectively corre-
sponding to a small (1,000 to 29,999 people), a medium (30,000 to 99,999 people) and a large (100,000 and more people) population centre. More precisely, each market is classified in one of these three categories. Average markets are constructed by averaging the values of $\mathbf{x}$ and $\hat{\nu}$ in each category. Table 1.4 describes each of these markets.

The subsidy is modelled as a reduction of the threshold associated with installing 1 to 10 transceivers for Vidéotron and Wind only. I consider the impact of reducing the estimated thresholds from 0% to 100% simultaneously for both new entrants. Using the estimated model, one can analyze the effect of the subsidy, distinguishing between the probability of installing at least one transceiver (Figures 1.2 to 1.4) and the probabilities associated with different quantities of transceivers (Figures 1.5 to 1.7).

Table 1.4: Average markets

<table>
<thead>
<tr>
<th>Markets</th>
<th>Obs.</th>
<th>Pop. dens.</th>
<th>Stocks of transceivers</th>
<th>Unobserved heterogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Rogers</td>
<td>Bell</td>
</tr>
<tr>
<td>Small</td>
<td>646</td>
<td>0.71</td>
<td>1.75</td>
<td>2.01</td>
</tr>
<tr>
<td>Medium</td>
<td>47</td>
<td>1.02</td>
<td>3.81</td>
<td>3.38</td>
</tr>
<tr>
<td>Large</td>
<td>25</td>
<td>1.35</td>
<td>5.95</td>
<td>5.58</td>
</tr>
</tbody>
</table>

Notes: Markets in the dataset are divided in three categories according to their population: small (1,000 to 29,999), medium (30,000 to 99,999) and large (100,000+). The number of markets in each category is in column “Obs.”. Average values of population density (in $10^3$ persons per square kilometre), stocks of transceivers (in natural logarithms, after adding 1 to avoid 0’s) and unobserved heterogeneity (estimated using factor models) are computed for each category.

At the extensive margin, i.e. whether or not firms install at least one transceiver, it is worth noting that the effect of the policy differs across markets of different sizes. In particular, while Wind is much more responsive than Vidéotron in small markets, the reverse is true in large ones. This contrast can be explained by the sizeable difference between the estimated new entrants’ 1-10 transceivers threshold parameters.
Chapter 1. Games with unobservable heterogeneity and multiple equilibria

Notes: Equilibrium probabilities of adding at least one transceiver for different levels of subsidy. Probabilities are plotted for a 0% to 100% reduction of Vidéotron’s and Wind’s costs of installing 1-10 transceivers.

Figure 1.3: Effect of subsidy on probability of installing at least one transceiver – medium market

Notes: Equilibrium probabilities of adding at least one transceiver for different levels of subsidy. Probabilities are plotted for a 0% to 100% reduction of Vidéotron’s and Wind’s costs of installing 1-10 transceivers.

Figure 1.4: Effect of subsidy on probability of installing at least one transceiver – large market
Chapter 1. Games with unobservable heterogeneity and multiple equilibria

Notes: Equilibrium probabilities of adding 0, 1-10, 11-50 or 51+ transceivers for different levels of subsidy. Probabilities are plotted for a 0% to 100% reduction of Vidéotron’s and Wind’s costs of installing 1-10 transceivers.

Figure 1.5: Effect of subsidy on probabilities of installing many transceivers – small market

Notes: Equilibrium probabilities of adding 0, 1-10, 11-50 or 51+ transceivers for different levels of subsidy. Probabilities are plotted for a 0% to 100% reduction of Vidéotron’s and Wind’s costs of installing 1-10 transceivers.

Figure 1.6: Effect of subsidy on probabilities of installing many transceivers – medium market
Notes: Equilibrium probabilities of adding 0, 1-10, 11-50 or 51+ transceivers for different levels of subsidy. Probabilities are plotted for a 0% to 100% reduction of Vidéotron’s and Wind’s costs of installing 1-10 transceivers.

Figure 1.7: Effect of subsidy on probabilities of installing many transceivers – large market

Moreover, in small markets, the slight reduction in Telus’ probability of installing transceivers is more than compensated by Rogers’ increase. However, for incumbents, subsidizing Vidéotron and Wind has a very modest effect on the probability of installing some transceivers in medium and large markets. In particular, Rogers, Bell and Telus almost certainly install transceivers in large markets, regardless of the level of the subsidy. In general, since most markets in the sample are small (646/718), the counterfactual results suggest that subsidizing new entrants’ first few transceivers may have an overall positive effect on both new entrants’ and incumbents’ investment probabilities.

At the intensive margin, i.e. focusing on the number of transceivers installed, probabilities of installing a larger number of transceivers become positive only in medium and large ones. In some sense, this prediction is reassuring: the subsidy keeps investments proportional to market size. While most of the action in small markets comes from shifting Vidéotron’s, Wind’s and Rogers’ probabilities from installing 0 to 1-10 transceivers, more nuanced variations are observable in larger markets. In particular, in medium markets, Rogers responds to Vidéotron’s and Wind’s increased probability of installing 1-10 transceivers by increasing its probability of installing 11-50 transceivers. In this case, subsidizing new entrants’ first few transceivers has a positive effect on one of the incumbents’ network qualities.

Moreover, the difference between new entrants’ intensive margin response to the subsidy is striking when looking at large markets. On one hand, while the subsidy increases the probability of Vidéotron installing up to ten transceivers, its probability of installing more than 10 transceivers is basically 0. On the other hand, similarly to the incumbents, Wind’s probabilities of installing transceivers remain mostly unaffected by this subsidy. The relatively high costs of installing a large number of transceivers for Vidéotron may explain this difference.
To summarize, the counterfactual experiments suggest two important interpretations:

1. Subsidizing the first 10 transceivers can increase new entrants’ and incumbents’ overall network investments in small markets, which represent the vast majority of the markets studied.

2. The subsidy can also improve some of the incumbents’ network quality by increasing their probability of installing a larger number of transceivers in response to new entries.

Two important caveats are worth mentioning at this stage. First, the constructed average markets took the common-knowledge unobservable heterogeneity as exogenously given. Presumably, since part of this heterogeneity may include the effects of agreements on each provider’s payoff, it would be preferable to allow these agreements to vary with the counterfactual experiments. However, it is not easy to model how these unobservables would change in a hypothetical world. Second, the experiments treated all providers as being potential entrants. Implicitly, this amounts to assuming that all five players own some spectrum licenses in all markets. While this is true in some areas of the country (thanks to spectrum set-aside), there are other areas where the three incumbents and only one of the two new entrants own licenses.

1.6 Conclusion

The identification argument I propose introduces predetermined outcomes to recover two types of unobservables separately when estimating empirical games: player-specific common-knowledge unobservable heterogeneity and a variable indicating which equilibrium is realized in the data. Recovering such unobservables brings two important advantages. First, it allows the researcher to explicitly control for unobservables that, if omitted, would bias the estimation. Second, it provides an extra source of exogenous variation that can then be leveraged to identify the primitives of the model, which is especially helpful in cases where observable regressors satisfying the exclusion restriction typically used for identification are not available to the researcher. Even in cases where observable exclusion restrictions are available, such an extra source of exogenous variation should improve the precision of the estimates.

I apply the method to estimate a game of transceivers’ location decisions between three national incumbents and two new entrants using data from the Canadian mobile telecommunications industry. Bids that network operators make to obtain spectrum licenses are used as predetermined outcomes to recover the effects of unobservable agreements between firms (domestic roaming, tower sharing, network sharing, etc.) and other common-knowledge unobservables on transceivers’ locations. Estimation results suggest that economies of density may explain why new entrants do not always use their allocated spectrum or even go out of business. In counterfactual experiments, I study the impact of subsidizing new entrants’ first ten transceivers to compensate for the absence of economies of density generated by accumulated stocks. I find that, especially in smaller markets, such a subsidy can increase the probability of both new entrants and incumbents installing new transceivers. Furthermore, incumbents may improve the quality of their network (by adding more transceivers) as a response to new entries.

One extension of the empirical application, which I leave for future projects, is to include a full-blown structural model of network operators’ bids for spectrum licenses. In the current chapter, I combine a reduced-form representation of firms’ bids with a structural model of market entry. While the reduced-form approach allows me to control for common-knowledge unobservable heterogeneity without relying on structural assumptions, there would also be some advantages in explicitly modelling firms’ bidding
Chapter 1. Games with unobservable heterogeneity and multiple equilibria

behaviour. One could then interpret transceivers’ location decisions as a two-stage game in which firms first bid on spectrum licenses and then choose how many transceivers to install where they are allowed to. Both structural models (i.e., the auction and the entry models) would be driven by the same primitives (i.e., the payoffs of providing coverage in a given market). The full-blown auction model could be used to answer other questions related to the effect of spectrum set-asides on transceivers’ locations instead of taking this policy as given, as is done with the reduced-form approach in the current chapter. In particular, the structural auction model would help to assess the optimality of this policy.

1.7 Appendix

1.7.1 Factor model estimation

This appendix provides some information about the estimation of the factor model introduced in Assumption 6. The method borrows heavily from Bai (2009). Let the realization of $Z_{imk}$, i.e. the $k$-th predetermined outcome variable for player $i$ in market $m$, be such that:

\[ z_{imk} = x_m'r_i + \nu_m'\gamma_{ik} + u_{imk} \]  

(1.27)

where $z_{imk}$ and $x_m$ are observed; $\nu_m = [\nu_{1m}, \ldots, \nu_{N_m}]'$ are interpreted as realized unobservable factors; $\gamma_{ik} = [\gamma_{i1k}, \ldots, \gamma_{ikN}]'$ are the loadings; $u_{imk}$ is idiosyncratic shock independent across $i$'s, $m$'s and $k$'s and independent of $x_m$, $\nu_m$ and $\gamma_{ik}$.

For a given $i$, $\nu_m$ and $\gamma_{ik}$ are treated as fixed effects to be estimated, which requires both $M$ and $K$ to be large. Some notation must be introduced to write the problem in matrix notation. For player $i$, let $z_{ik} = [z_{i1k}, \ldots, z_{iMk}]'$, $A_i = [\alpha_{i1}, \ldots, \alpha_{iK}]'$ and $\Gamma_i = [\gamma_{i1}, \ldots, \gamma_{iK}]'$. Similarly, define $X = [x_1, \ldots, x_M]'$ and $V = [\nu_1, \ldots, \nu_M]'$.

It is well known that some normalizations are needed to uniquely identify $\Gamma_i$ and $V$. In fact, notice that for any invertible matrix $C$, $V'C' = VCC^{-1} = (VC)(C'C^{-1})' = V'C''$. Therefore, in order to uniquely identify $\Gamma_i$ and $V$ (up to a rotation and a change of columns’ signs), the following normalizations are made: (i) $V'V/M = I_N$, where $I_N$ is the $N \times N$ identity matrix; and (ii) $\Gamma_i'\Gamma_i$ is diagonal.

For player $i$, the least-squares estimates of $A_i$, $V$ and $\Gamma_i$ are defined as:

\[
\begin{bmatrix} \hat{A}_i, \hat{V}, \hat{\Gamma}_i \end{bmatrix} = \arg\min_{A_i, V, \Gamma_i} \sum_{k=1}^{K} (z_{ik} - X\alpha_{ik} - V\gamma_{ik})'(z_{ik} - X\alpha_{ik} - V\gamma_{ik}) \\
\text{s.t. } V'V/M = I_N, \\
\Gamma_i'\Gamma_i \text{ is diagonal.}
\]  

(1.28)

First, for a fixed $V$, notice that $\hat{A}_i$ can be obtained by applying Frisch-Waugh theorem. Let $M_V = I_M - V(V'V)^{-1}V' = I_M - VV'/M$. Then,

\[
\hat{\alpha}_{ik}(V) = (X'M_VX)^{-1}X'M_Vz_{ik}.
\]  

(1.29)

Moreover, given $A_i$, let $w_{ik}(\alpha_{ik}) = z_{ik} - X\alpha_{ik} = V\gamma_{ik} + u_{ik}$, where $u_{ik} = [u_{ik1}, \ldots, u_{ikM}]'$. In this case, $\hat{V}(A_i)$ and $\hat{\Gamma}(A_i)$ solve the following minimization problem:
\[
\left[ \hat{V}(A_i), \hat{\Gamma}_i(A_i) \right] = \arg \min_{V, \Gamma_i} \sum_{k=1}^{K} (w_{ik}(\alpha_{ik}) - V\gamma_{ik})' (w_{ik}(\alpha_{ik}) - V\gamma_{ik}) \\
\text{s.t. } V'V/M = I_N, \\
\Gamma_i'\Gamma_i \text{ is diagonal.}
\]

This last objective function can be written as \( \text{Tr} \left\{ \left( W_i(A_i) - V\Gamma_i \right)' \left( W_i(A_i) - V\Gamma_i \right) \right\} \), where \( W_i(A_i) = [w_{i1}(\alpha_{i1}), \ldots, w_{iK}(\alpha_{iK})] \). Notice that, at this stage, the estimation problem becomes a principal components problem. Concentrating out \( \Gamma_i \), one gets:

\[
\Gamma_i = W_i(A_i)'V(V'V)^{-1} = \frac{W_i(A_i)'V}{M}
\]

The concentrated objective function is equivalent to:

\[
\text{Tr} \left\{ W_i(A_i)'W_i(A_i) \right\} - \frac{\text{Tr} \left\{ V'W_i(A_i)W_i(A_i)'V \right\}}{M}
\]

which is minimized with respect to \( V \) when \( V \) maximizes \( \text{Tr} \left\{ V'W_i(A_i)W_i(A_i)'V/M \right\} \). The estimator of \( V(A_i) \) therefore corresponds to the \( N \) eigenvectors (multiplied by \( \sqrt{M} \)) associated with the \( N \) largest eigenvalues of \( W_i(A_i)W_i(A_i)' \). Finally, the estimator of \( \Gamma_i(A_i) \) is:

\[
\hat{\Gamma}_i(A_i) = M^{-1}W_i(A_i)'\hat{V}.
\]

Notice that in order to estimate \( V \), which really is what is needed in the identification argument, one does not need to estimate \( \Gamma_i \). That being said, in order to estimate \( V \), one must estimate \( A_i \) and \textit{vice versa}, which is achieved through an iterative procedure.

### 1.7.2 Proof of Lemma 1

Given the point-identified realizations of \( \mathcal{Y} \), one can construct choice probabilities conditional on \( x \) and \( \nu \). However, these conditional probabilities are still potentially mixed over multiple equilibria, such that:

\[
P(y|x, \nu) = \sum_{\tau \in \mathcal{T}(x, \nu)} P(y|x, \nu, \tau) \lambda(\tau|x, \nu)
\]

The key for identification is that the different elements in \( y \) are independent after conditioning on \( x, \nu \) and \( \tau \), i.e. \( P(y|x, \nu, \tau) = \prod_{i \in N} P(y_i|x, \nu, \tau) \).

The first part of Lemma 1 identifies the number of equilibria realized in the data given \( x, \nu \), which is denoted \( |\mathcal{T}(x, \nu)| \). This result is a simple application of Kasahara and Shimotsu (2014, Proposition 4, p. 102). Let \( A(x, \nu) \) be the \( |\mathcal{T}(x, \nu)| \times |\mathcal{T}(x, \nu)| \) diagonal matrix with weights \( \lambda(\tau|x, \nu) \) for \( \tau \in \{1, \ldots, |\mathcal{T}(x, \nu)|\} \) on its diagonal. Given the notation introduced above, one can write:

\[
P(y_{N1}, y_{N2}|x, \nu) = P(y_{N1}|x, \nu, \tau) A(x, \nu) P(y_{N2}|x, \nu, \tau)'.
\]

As noted by Kasahara and Shimotsu (2014), the number of components in the finite mixture representation of \( P(y_{N1}, y_{N2}|x, \nu) \) is closely related to the nonnegative rank (Cohen and Rothblum, 1993) of \( P(y_{N1}, y_{N2}|x, \nu) \), denoted \( \text{rank}^+_\{P(y_{N1}, y_{N2}|x, \nu)\} \). The nonnegative rank of \( P(y_{N1}, y_{N2}|x, \nu) \) is de-
fined as the smallest number of nonnegative rank 1 matrices such that \( P (y_{N_1}, y_{N_2}|x, \nu) \) equals their sum. In particular, given that \(|\mathcal{T}(x, \nu)| \leq \min \{|y_{N_1}|, |y_{N_2}|, |y_{N_3}|\} \) and that \( P (y_{N_q}|x, \nu, \tau) \) is full column rank for \( q = 1, 2, 3 \) (which holds if the distribution of \( Y_{N_q} \) differs across equilibria), we have that:

\[
\text{rank}_+ \{ P (y_{N_1}, y_{N_2}|x, \nu) \} = |\mathcal{T}(x, \nu)|. \tag{1.36}
\]

Since it is often computationally difficult to obtain the nonnegative rank of a matrix, it is preferable to write \(|\mathcal{T}(x, \nu)|\) in terms of \( \text{rank} \{ P (y_{N_1}, y_{N_2}|x, \nu) \} \). From Cohen and Rothblum (1993, Lemma 2.3, p. 152):

\[
\text{rank} \{ P (y_{N_1}, y_{N_2}|x, \nu) \} \leq \text{rank}_+ \{ P (y_{N_1}, y_{N_2}|x, \nu) \}. \tag{1.37}
\]

However, in the current setting, this inequality is always binding. In fact, by applying Frobenius inequality on (1.35), one gets:

\[
\text{rank} \{ P (y_{N_1}, y_{N_2}|x, \nu) \} \geq |\mathcal{T}(x, \nu)| \tag{1.38}
\]

since \( \text{rank} \{ \Lambda (x, \nu) \}, \text{rank} \{ P (y_{N_1}|x, \tau) \Lambda (x, \nu) \} \) and \( \text{rank} \{ \Lambda (x, \nu) P (y_{N_2}|x, \nu, \tau) \} \) are all equal to \(|\mathcal{T}(x, \nu)|\). Therefore, since \( \text{rank} \{ P (y_{N_1}, y_{N_2}|x, \nu) \} > |\mathcal{T}(x, \nu)| = \text{rank}_+ \{ P (y_{N_1}, y_{N_2}|x, \nu) \} \) would contradict (1.37), it must be the case that \( \text{rank} \{ P (y_{N_1}, y_{N_2}|x, \nu) \} = |\mathcal{T}(x, \nu)| \), which completes the proof of the first part of the Lemma.

The second part of Lemma 1 identifies the equilibrium-specific conditional choice probabilities and the equilibrium selection mechanism. This result is an application of Bonhomme, Jochmans, and Robin (2016, Theorems 1-3, pp. 549-551). From the first part of the Lemma, \( \text{rank} \{ P (y_{N_1}, y_{N_2}|x, \nu) \} = |\mathcal{T}(x, \nu)| \) implies that \( P (y_{N_1}, y_{N_2}|x, \nu) \) has a singular value decomposition:

\[
P (y_{N_1}, y_{N_2}|x, \nu) = R_1 (x, \nu) S (x, \nu) R_2 (x, \nu)' \tag{1.39}
\]

where \( R_q (x, \nu) \) is a \(|y_{N_q}| \times |\mathcal{T}(x, \nu)|\) unitary matrix for \( q = 1, 2 \) and \( S (x, \nu) \) is a nonsingular \(|\mathcal{T}(x, \nu)| \times |\mathcal{T}(x, \nu)|\) diagonal matrix.

Let \( L_q (x, \nu) = S (x, \nu)^{-1/2} R_q (x, \nu)' \). Consider transforming \( P (y_{N_1}, y_{N_2}|x, \nu) \) as:

\[
L_1 (x, \nu) P (y_{N_1}, y_{N_2}|x, \nu) L_2 (x, \nu)' = I_{|\mathcal{T}(x, \nu)|} = Q (x, \nu) Q (x, \nu)^{-1} \tag{1.40}
\]

where \( I_{|\mathcal{T}(x, \nu)|} \) is the \(|\mathcal{T}(x, \nu)| \times |\mathcal{T}(x, \nu)|\) identity matrix. In other words the transformed matrix \( L_1 (x, \nu) P (y_{N_1}, y_{N_2}|x, \nu) L_2 (x, \nu)' \) admits an eigenvalue decomposition such that:

\[
L_1 (x, \nu) P (y_{N_1}, y_{N_2}|x, \nu) L_2 (x, \nu)' = Q (x, \nu) D_0 (x, \nu) Q (x, \nu)^{-1} \tag{1.41}
\]

where \( D_0 (x, \nu) = I_{|\mathcal{T}(x, \nu)|} \). Furthermore, consider the matrix \( P (y_{N_1}, y_{N_2}|x, \nu, y_{N_3}^k) \) for a given \( y_{N_3}^k \in Y_{N_3} \) and let \( D_k (x, \nu) \) be the diagonal matrix whose diagonal is the \( k\)-th row of \( P (y_{N_3}|x, \nu, \tau) \). Then, one can write:

\[
P (y_{N_1}, y_{N_2}|x, \nu, y_{N_3}^k) = P (y_{N_1}|x, \nu, \tau) \Lambda (x, \nu) D_k (x, \nu) P (y_{N_2}|x, \nu, \tau)' \tag{1.42}
\]

By applying the same transformation as above, one gets:
\[ \mathbf{L}_1 (x, \nu) \mathbf{P} (y_{N_1}, y_{N_2} | x, \nu, y_{N_3}^k) \mathbf{L}_2 (x, \nu)' = \mathbf{Q} (x, \nu) \mathbf{D}_k (x, \nu) \mathbf{Q} (x, \nu)^{-1}. \]  

(1.43)

Therefore, the matrices \( \mathbf{L}_1 (x, \nu) \mathbf{P} (y_{N_1}, y_{N_2} | x, \nu, y_{N_3}^k) \mathbf{L}_2 (x, \nu)' \) for \( k = 1, \ldots, |y_{N_3}| \) can be diagonalized in the same basis as \( \mathbf{L}_1 (x, \nu) \mathbf{P} (y_{N_1}, y_{N_2} | x, \nu, y_{N_3}^k) \mathbf{L}_2 (x, \nu)' \), i.e. the columns of \( \mathbf{Q} (x, \nu) \). In other words, \( \mathbf{Q} (x, \nu) \) is the joint diagonalizer of a sequence of matrices. In particular, the equilibrium-specific probability of observing \( y_{N_3} \) are identified (up to permutation) by the eigenvalues of the matrix \( \mathbf{L}_1 (x, \nu) \mathbf{P} (y_{N_1}, y_{N_2} | x, \nu, y_{N_3}^k) \mathbf{L}_2 (x, \nu)' \). This procedure can therefore be used to identify \( \mathbf{P} (y_{N_q} | x, \nu, \tau) \) for \( q = 1, 2, 3 \).

Finally, since the equilibrium selection mechanism is constant across all players, one may use the identified \( \mathbf{P} (y_{N_q} | x, \nu, \tau) \) and the corresponding \( \mathbf{P} (y_{N_q} | x, \nu) \) for any \( q = 1, 2, 3 \) to recover the weights associated with each equilibrium. Let \( \lambda (x, \nu) \) be the \( |\mathcal{F} (x, \nu)| \times 1 \) vector corresponding to the diagonal of \( \mathbf{A} (x, \nu) \). Then,

\[ \mathbf{P} (y_{N_q} | x, \nu) = \mathbf{P} (y_{N_q} | x, \nu, \tau) \lambda (x, \nu) \]  

(1.44)

and since \( \mathbf{P} (y_{N_q} | x, \nu, \tau) \) is full column rank:

\[ \lambda (x, \nu) = \left[ \mathbf{P} (y_{N_q} | x, \nu, \tau)' \mathbf{P} (y_{N_q} | x, \nu, \tau) \right]^{-1} \mathbf{P} (y_{N_q} | x, \nu, \tau)' \mathbf{P} (y_{N_q} | x, \nu). \]  

(1.45)

In other words, the equilibrium selection mechanism corresponds to the least-squares coefficients obtained by regressing \( \mathbf{P} (y_{N_q} | x, \nu) \) on \( \mathbf{P} (y_{N_q} | x, \nu, \tau) \) for any \( q = 1, 2, 3 \).
Chapter 2

Sharp test for equilibrium uniqueness in discrete games with a flexible information structure

2.1 Introduction

Economic models of strategic interactions among agents often admit multiple equilibria. Equilibria multiplicity in the model may be seen as an economic problem and some equilibrium refinement can be used to determine which equilibrium or which equilibria should be considered. In empirical games, equilibria multiplicity in the data generating process is an econometric issue that must be taken into account when trying to recover the primitives of the model from the data observable to the econometrician. In fact, identification arguments available in the literature differ according to the assumptions maintained on the number of equilibria realized in the data. Testing for equilibrium uniqueness is therefore desirable to guide empiricists towards an appropriate estimation approach. Furthermore, such assumptions have different implications depending on the information structure of the game. It follows that tests allowing for a relatively flexible information structure may be preferable in practice.

The current chapter provides a test of equilibrium uniqueness in the data generating process under a flexible information structure. It allows for private as well as common-knowledge payoff-relevant information to be unobservable to the econometrician. The identification argument I propose is nonparametric: no parametric assumptions are needed for the payoff functions, the distribution of the private information shocks, the distribution of the common-knowledge unobservables, nor the equilibrium selection mechanism. More precisely, the observable joint distribution of players’ decisions can be written as a finite mixture and I use partial identification results from Henry, Kitamura, and Salanié (2014) to derive sharp bounds for the distributions defining this finite mixture. Equilibrium uniqueness in the data imposes further restrictions on these distributions. Testing the null hypothesis of equilibrium uniqueness in the data generating process simply amounts to testing whether or not the identified set constructed from all these restrictions is empty. The identification result relies on the existence of an observable variable that can be interpreted as a proxy for the common-knowledge payoff-relevant unobservables. It must (i) have sufficient variation; (ii) be correlated with these common-knowledge unobservables; and
(iii) provide only redundant information about players’ decisions and the equilibrium selection if such unobservables were actually observed. The test is implemented through a simple two-stage approach and simulation results suggest that it performs well.

How one treats multiple equilibria typically depends on how one is willing to interpret the information unobservable to the econometrician. In games where one assumes that all unobservables are known to all players, i.e. games of complete information, set-identified estimators have been proposed to recover the set of the model’s primitives that can rationalize the data for any possible equilibrium selection mechanism (e.g. Tamer, 2003; Ciliberto and Tamer, 2009; Beresteau, Molchanov, and Molinari, 2011; Galichon and Henry, 2011; etc.). In that sense, such estimation methods are robust to equilibria multiplicity. In contrast, many estimation methods ask the econometrician to take a stance on whether or not there are multiple equilibria in the data when estimating games of incomplete information, i.e. games assuming that unobservables are players’ private information. On one hand, many estimation methods assume that the data have been generated by a single equilibrium (e.g., Aguirregabiria and Mira, 2007; Bajari, Benkard, and Levin, 2007; Pakes, Ostrovsky, and Berry, 2007; Pesendorfer and Schmidt-Dengler, 2008; Bajari, Hong, Krainer, and Nekipelov, 2010; Aradillas-Lopez, 2012; etc.). This assumption is often labeled the “single-equilibrium in the data” or the “degenerate equilibrium selection mechanism” assumption. On the other hand, multiple equilibria realized in the data generate an extra source of variation that helps to identify the primitives of the model (e.g., Sweeting, 2009; De Paula and Tang, 2012; Aradillas-Lopez and Gandhi, 2016). In those cases, one should therefore explicitly identify such multiple equilibria instead of assuming them away.

According to the single-equilibrium in the data assumption, every time the same players play the same game, the same equilibrium is realized. For example, consider a game of market entry between two players, firm A and firm B. Suppose that this game has the following two equilibria: either A is more likely to enter the market than B, or vice versa. Such equilibria may arise in markets that are typically too small to justify simultaneous entry. In an econometric study of the entry behaviour of A and B, one would typically observe firms’ entry decisions in several markets. The single equilibrium in the data assumption states that if A is more likely to enter than B in one specific market, then it also has to be more likely to enter than B whenever the same game is realized in another market. This assumption is maintained even if B being more likely to enter than A is also sustainable in equilibrium.

Of course, the single-equilibrium in the data assumption substantially simplifies the estimation by avoiding the need of solving for the different equilibria of the model: the only relevant equilibrium is the one realized in the data, which can be estimated. However, if the assumption is falsely maintained, the resulting estimates are associated with a mixture of equilibria, which is typically not an equilibrium in itself.

Some tests of equilibrium uniqueness in the data have been proposed in the literature.\footnote{I am focusing on static games. Otsu, Pesendorfer, and Takahashi (2014) have proposed some tests for dynamic games which also suffer from the limitations discussed here.} Two different approaches can be distinguished. The first one, which includes the tests proposed by De Paula and Tang (2012) and Xiao (2015), requires players’ decisions to be mutually independent after controlling for observable common-knowledge information and the selected equilibrium. This requirement implies that any unobservables are players’ private information or, in other words, the players do not know more about each other than the econometrician does. Therefore, common-knowledge payoff-relevant information known to all players, but unobservable to the econometrician is ruled out by assumption.
More importantly, such unobservables may lead to the false rejection of the null hypothesis of a single equilibrium in the data.

The second approach, which includes the papers by Aguirregabiria and Mira (2016), has the advantage of allowing for common-knowledge payoff-relevant unobservable information. A test of equilibrium uniqueness in the data generating process can be obtained as a by-product of their identification results. However, their identification argument generally does not apply to games between two players and can restrict the number of equilibria that is identifiable from the data. The test I propose has the advantage of dealing with these potential shortcomings at the cost of losing point-identification. Another important advantage of the approach I propose compared to theirs is that I do not put parametric restrictions on the distribution of players’ private information.

The chapter is organized as follows. Related literature is summarized in Section 2.2, with a special attention being paid to some useful results on the nonparametric identification of finite mixtures. A static discrete game with simultaneous decisions is introduced in Section 2.3. The nonparametric identification result and the statistical test that it suggests are respectively presented in Sections 2.4 and 2.5. Section 2.6 concludes.

2.2 Related literature

As mentioned above, a few papers from the static games literature propose tests of equilibrium uniqueness in the data generating process. These tests are usually obtained as by-products of some identification results. While these identification results provide great insights and are very promising on their own, some caveats about the corresponding tests for equilibrium uniqueness are worth pointing out.

An important feature of some existing tests is that they require players’ equilibrium-specific decisions to be independent given the observable common-knowledge information. In a static game of pure incomplete information this simply follows from the conditional independence of unobservable private shocks. This independence is the basis of De Paula and Tang (2012)’s test: it amounts to testing whether or not players’ decisions are conditionally independent.

The same conditional independence is also key to use recent nonparametric identification results from the literature on finite mixtures and measurement errors (e.g., Hall and Zhou, 2003; Hu, 2008; Kasahara and Shimotsu, 2009, 2014; Hu and Shum, 2012; Bonhomme, Jochmans, and Robin, 2016. For a recent survey of the applications of measurement errors models to empirical industrial organization and labour economics, see Hu, 2017). In games of pure incomplete information, one can use results from this literature to identify a lower bound on the number of equilibria occurring in the sample, which corresponds to the number of components in the finite mixture representation of the joint distribution of players’ decisions. This is the approach proposed by Xiao (2015).

Unfortunately, even if there is a single equilibrium in the data, conditional independence breaks down if players also take into account payoff-relevant information that is known to all of them, but unobservable to the econometrician. In such cases, De Paula and Tang (2012)’s and Xiao (2015)’s tests cannot be applied. The main issue is that if one finds the correlation between players’ decisions to be non-zero or if one finds more than one components in the finite mixture representing the observable choice probabilities, it could either be because there are multiple equilibria in the data and/or because there are some common-knowledge payoff-relevant unobservables. In other words, such unobservables may lead to the false rejection of the null hypothesis of a single equilibrium in the data.
Grieco (2014) proposes a framework that allows for multiple equilibria and a flexible information structure in a parametric setting. His results suggest that both private and common-knowledge unobservable information may be relevant in empirical applications. While this insight justifies the need to extend tests of equilibrium uniqueness to a flexible information structure, it seems desirable to relax the parametric assumptions maintained therein.

To my knowledge, the only semi-parametric identification results that allow for multiple equilibria and common-knowledge payoff-relevant unobservables are due to Aguirregabiria and Mira (2016) and Xiao (2017). Their main proposition follows from a sequential identification argument which combines results from the literature about nonparametric identification of finite mixtures. In a first step, they identify the nonparametric distribution of a discrete random variable with finite support that summarizes the information of the common-knowledge payoff-relevant unobservables and the unobservable variable that indicates which equilibrium is realized. Using the property that the equilibrium selection variable is not payoff-relevant, the distribution of the common-knowledge unobservables and the equilibrium selection can be separately identified. In their setting, the number of equilibria corresponds to the cardinality of the support of the unobservable variable that selects the equilibrium realized in the data.

Unfortunately, the sequential approach in their main result can be problematic when testing for equilibrium uniqueness. The potential limitation is that the finite mixture framework may restrict the number of components identifiable from the data. This restriction is not innocuous. Notably, the largest number of components that one can identify in the first step of their sequential argument is given by the number of alternatives available in the players’ choice set, raised to the power \( \lfloor \frac{N-1}{2} \rfloor \), where \( N \) is the number of players and \( \lfloor \cdot \rfloor \) is the floor function. As a consequence, no mixture would be identifiable in a game of market entry between two players. Aguirregabiria and Mira (2016) also propose some non-sequential results which require an exclusion restriction commonly used in empirical games (e.g. Bajari, Hong, Krainer, and Nekipelov, 2010) to be sufficiently over-identifying. While their non-sequential identification results allow for the identification of more components than in the sequential case, it still puts an upper bound on the identifiable number of equilibria. Again, this upper bound has some important implications when testing for equilibrium uniqueness. For instance, in this non-sequential approach, they show that one could test for equilibrium uniqueness in a game between two players, as long as there are at most two realizations of the common-knowledge unobservable and that one of these two realizations is known to be associated with a single equilibrium prior to testing.

Of course, similar restrictions on the identifiable number of components are likely to arise in other identification arguments that are based on the finite mixture framework. The approach that I propose is no exception, but the conditions imposed are less restrictive. A considerable advantage of this method is that the number of equilibria is not restricted prior to the test. The restriction may only affect the support of the common-knowledge payoff-relevant unobservables, which is still allowed to be larger than in Aguirregabiria and Mira (2016).

At this point, it is worth distinguishing between two different approaches that have been proposed in the literature about nonparametric identification of finite mixtures: (i) the conditional independence, and (ii) the exclusion restriction approaches. In both approaches, the main objective is to identify the number of components, the conditional component distributions and the mixing probability weights.

In the conditional independence approach, the joint distribution conditional on the latent mixing

\[ \text{Notice that, while Xiao (2017) studies the problem of unobservable heterogeneity and multiple equilibria in dynamic games, the current chapter focuses on static games.} \]
variable can be factored as the product of its marginals. A system of equations is constructed by considering different sub-vectors of the vector of mixed variables. For instance, in the context of a game with $N$ players, one would consider the joint distribution of any subset of players’ decisions. Point identification is reached if one can construct enough equations to identify all the corresponding marginal conditional component distributions and mixing probability weights. The conditional independence approach has been used by, among others, Hall and Zhou (2003), Kasahara and Shimotsu (2009), Hu and Shum (2012) and Bonhomme, Jochmans, and Robin (2016). This conditional independence is also needed to identify the number of mixtures when using Kasahara and Shimotsu (2014)’s results.

Alternatively, the exclusion restriction approach assumes that there exists an observable variable with sufficient variation that affects the mixing probability weights, but not the conditional component distributions. With this exclusion restriction, one can write the conditional component distributions and the mixing probability weights in terms of some set-identified parameters. To my knowledge, Henry, Kitamura, and Salanié (2014) are the first to propose this approach and they also provide an extensive discussion suggesting that the required exclusion restriction often arises naturally in applied work. A specially relevant feature of this alternative approach is that the joint distribution conditional on the mixing variable does not have to be factorable in the product of its marginals. This is key in the context of the current chapter: because the observable conditional choice probabilities are potentially mixed over multiple equilibria and common-knowledge payoff-relevant unobservables, players’ decisions may fail to be independent after controlling for only one of these two types of unobservables.

Finally, it should be emphasized that the main object of interest in the current chapter is the number of equilibria in the data generating process, which form a subset of the equilibria in the model. In that sense, the objective is very different from other works, such as Kasy (2015), focusing on the number of equilibria in the model. In fact, when applying his inference method to a game of incomplete information, Kasy (2015) first estimates the model using a two-step approach. Such two-step estimation relies on the single equilibrium in the data assumption, which can be tested using the method I propose.

### 2.3 A static discrete game with simultaneous decisions

While the model could be described in a more general way, I focus on a simple static discrete game with simultaneous decisions: a binary game between two players with two realizations of the common-knowledge payoff-relevant unobservables. This simple $2 \times 2 \times 2$ case helps build the intuition behind the identification result. In particular, the simplification makes it fairly easy to represent the corresponding identified set graphically and to show some interesting results.

Even though the model introduced here is very similar to the model in Chapter 1, it is worth repeating its basic features for completeness. Once again, I first describe the economic model, i.e. the game as it is played by the players. I then turn to its econometric counterpart observed by the econometrician.

#### 2.3.1 Economic model

Consider a game where $N = 2$ players, indexed by $i \in \{1, 2\}$, simultaneously choose from a binary choice set. Players’ decisions are stored in a vector of random variables $\mathbf{y} = [y_1, y_2]'$ with realizations.
Players’ decisions are contingent on some state variables, which are separated into two categories, depending on whether they are observed by all players. Let \( \mathcal{S} = [\mathcal{S}_1', \mathcal{S}_2'] \) with realizations \( s = [s_1', s_2']' \in \mathcal{S}^2 \) be some information that is common knowledge to both players. Furthermore, let \( \mathcal{E} = [\mathcal{E}_1, \mathcal{E}_2]' \) with realizations \( \varepsilon = [\varepsilon_1, \varepsilon_2]' \in \mathbb{R}^2 \) be some private information. Let \( G_{\mathcal{E}_i} (\cdot) \) denote the cumulative density function of \( \mathcal{E}_i \). Because player \( i \)'s opponent does not observe \( \varepsilon_i \), this is a game of incomplete information.

Let \( \pi_i (\cdot) : \mathcal{Y} \times \mathcal{S} \times \mathbb{R} \mapsto \mathbb{R} \) be player \( i \)'s payoff function. While the payoff of player \( i \) choosing \( y_i = 0 \) is normalized to 0, the payoff when choosing \( y_i = 1 \) is denoted by \( \pi_i (y_{-i}, s, \varepsilon_i) \), where \( -i \) denotes player \( i \)'s opponent. The following assumption is maintained on the payoff functions and the distributions of \( \mathcal{S}, \mathcal{E} \).

**Assumption 8** (State variables and payoffs). (i) \( \mathcal{S}, \mathcal{E}_1 \) and \( \mathcal{E}_2 \) are mutually independent. (ii) \( G_{\mathcal{E}_1} (\cdot), G_{\mathcal{E}_2} (\cdot) \) are common-knowledge to both players and they are absolutely continuous with respect to the Lebesgue measure on \( \mathbb{R} \). (iii) \( \pi_1 (\cdot) \) and \( \pi_2 (\cdot) \) are common-knowledge to both players.

The timing of the decision process is as follows. First, \( s \) and \( \varepsilon \) are realized. Even if players do not observe the realization of their opponent’s private information, they can still form beliefs about their opponent’s decision under Assumption 8. Then, both players simultaneously decide, i.e. \( y \) is realized and commonly observed. To sum up, at the time of the simultaneous decisions, player \( i \)'s information set is:

\[
J_i = \{ s, \varepsilon_i, \pi_1 (\cdot), \pi_2 (\cdot), G_{\mathcal{E}_1} (\cdot), G_{\mathcal{E}_2} (\cdot) \}. \tag{2.1}
\]

Player \( i \)'s strategy is a function that maps the information set, \( J_i \), to the choice set, i.e. \( \sigma_i (\cdot) : J_i \mapsto \mathcal{Y} \). For a given strategy, the conditional choice probability of player \( i \) choosing \( y_i = 1 \) at a given \( s \in \mathcal{S} \) is:

\[
p_i (s) \equiv p (y_i = 1 | s) = \int 1 \{ \sigma_i (J_i) = 1 \} \, dG_{\mathcal{E}_i} (\varepsilon_i) \tag{2.2}
\]

which can be interpreted as player \( i \)'s opponent’ beliefs regarding player \( i \)'s decision, when player \( i \) behaves according to strategy \( \sigma_i (J_i) \). Collect those probabilities in \( p (s) \equiv [p_1 (s), p_2 (s)]' \). Using these choice probabilities, one can write the expected payoff of player \( i \) choosing \( y_i = 1 \) as:

\[
\pi^p_i (s, \varepsilon_i) \equiv p_{-i} (s) \pi_i (1, s, \varepsilon_{-i}) + [1 - p_{-i} (s)] \pi_i (0, s, \varepsilon_{-i}) \tag{2.3}
\]

If each player’s strategy is to maximize expected payoffs, (2.2) can be written as:

\[
p_i (s) = \int 1 \{ \pi^p_i (s, \varepsilon_i) \geq 0 \} \, dG_{\mathcal{E}_i} (\varepsilon_i) \tag{2.4}
\]

The right hand side of equations (2.4) is the best response mapping of player \( i \) for given beliefs regarding the opponent’s decision. Let \( \psi_i (\cdot) \) denote this mapping and define \( \Psi (\cdot) \equiv [\psi_1 (\cdot), \psi_2 (\cdot)]' \). It follows that \( p (s) \) can be written as:

\[
p (s) = \Psi (s, p (s)) \tag{2.5}
\]
where $\Psi(s, \cdot) : [0, 1]^2 \mapsto [0, 1]^2$ is a mapping in the probability space. Given this best response mapping, one can define a Bayesian Nash Equilibrium (BNE) in pure strategies. Defining a BNE in the probability space is very convenient to analyze equilibrium existence and multiplicity (Milgrom and Weber, 1985).

**Definition 2** (BNE in probability space). A pure strategy BNE in the probability space is a set of conditional choice probabilities $p^*(s)$ such that $p^*(s) = \Psi(s, p^*(s))$.

Definition 2 simply states that, in equilibrium, players’ beliefs are consistent with their opponent’s. In fact, a BNE in the probability space is a fixed point of the best response mapping. Since $\Psi(s, \cdot)$ maps a compact set to itself and since it is continuous in $p(s)$, the existence of an equilibrium follows from Brouwer’s fixed point theorem for any $s \in S$. However, uniqueness is not guaranteed.

Let $T$ be a random variable labelling which equilibrium is played. More precisely, each equilibrium is indexed by a realization $\tau \in \mathcal{T}(s)$ with conditional probability mass function $\lambda(\tau|s)$ which can be interpreted as the equilibrium selection mechanism given the information observable to both players.

**Assumption 9** (Equilibrium index). (i) $T$ is independent from $\mathcal{E}_1$ and $\mathcal{E}_2$. (ii) $T$ is discrete with support $\mathcal{T}(s) \equiv \{1, \ldots, |\mathcal{T}(s)|\}$.

In order to fix ideas, I now introduce a running example that will be used as a data generating process to illustrate several concepts and results throughout the chapter. This example is a very simple static game of market entry between two firms.

**Example 2** (Simple game of market entry). Consider two firms deciding whether or not they want to operate in a given market, such that $y_i = 1$ if firm $i$ enters the market and $y_i = 0$ otherwise. In this case, $S$ could be some common-knowledge information about the size of each firm, whether or not they operate in surrounding markets and consumers’ preferences. Moreover, $E$ could refer to some private information cost shifters such as managerial ability. Let players’ payoffs when entering the market be:

$$
\pi_1(y_2, s, \varepsilon_1) = s_1 - 4y_2 - \varepsilon_1 
$$

$$
\pi_2(y_1, s, \varepsilon_2) = s_2 - 3y_1 - \varepsilon_2.
$$

Furthermore, let $E \sim \text{Normal}(0, I_2)$, where $I_2$ is the $2 \times 2$ identity matrix. Using $\Phi(\cdot)$ to denote the standard normal cumulative density function, the best response mapping is:

$$
p(s) = \begin{bmatrix} p_1(s) \\ p_2(s) \end{bmatrix} = \begin{bmatrix} \Phi(s_1 - 4p_2(s)) \\ \Phi(s_2 - 3p_1(s)) \end{bmatrix} \equiv \Psi(s, p(s)).
$$

Figure 2.1, is the graphical representation of the best response mapping in (2.8). The BNE(s) are given by the intersection(s) of the two best response functions. This figure clearly illustrates that, for a given set of primitives, different realizations of $S$ are associated with different BNE’s and in particular with different numbers of BNE’s.

---

5A more rigorous notation would be to write $T(s)$ and $\tau(s)$ instead of $T$ and $\tau$, but the latter is preferred to simplify notation.
2.3.2 Econometric model

The game described so far is the economic game as it is played by the players. We now turn to the econometric game, i.e., the game as it is observed by the econometrician. An important difference between the two is that the researcher only observes some of the common-knowledge payoff-relevant state variables in $S$. The following assumption is maintained on $S$.

Assumption 10 (Common-knowledge payoff-relevant state variables). (i) $S = [X', V']'$, where $X$ is observable to the econometrician, but $V$ is not. (ii) $X$ and $V(x)$ have finite and discrete supports $X = \{x^0, x^1, \ldots, x^{|X|-1}\}$ and $V(x) = \{\nu^0, \nu^1, \ldots, \nu^{V(x)-1}\}$. (iii) Realized $\nu$'s are drawn from the conditional distribution with probability mass function $\Gamma (\cdot | x)$. (iv) Conditional on $X$, $V$ is independent from $E$.

Typically, $V$ can be thought of as a vector of common-knowledge payoff-relevant state variables that are unobservable to the researcher. In particular, $V$ could be a vector of player-specific unobservables. In the simple $2 \times 2 \times 2$ case, $|V(x)| = 2$ and $V(x) = \{\nu^0, \nu^1\}$.

While $X$ being a finite and discrete set is not essential in theory, it is very convenient in practice since the identification result holds for fixed values of $x$. The applied researcher would find it helpful to discretize any continuous variables in $X$. Assumption 10 allows for a fairly flexible information structure. Notice that $X$ and $V$ are not assumed to be independent across markets. This will be useful when justifying the choice of exclusion restriction proposed below.

Example 2 (Simple game of market entry, continued). In the simple game of market entry between two firms, $V$ could refer to consumers’ preferences in a given market. While both firms may have gathered information about these preferences through market research, such information is typically unobservable.

---

Footnote 6: Since the support of $V$ depends on $x$, it may be preferable to write $V(x)$ and $\nu(x)$. However, similarly as for $T$ and $\tau$, the argument is dropped to alleviate notation.
to the econometrician. Since consumers’ preferences may affect firms’ payoffs differently, the vector $\mathbf{V}$ is allowed to have firm-specific components such that $\mathbf{V} = [\nu_1, \nu_2]^\prime$. The discrete support of $\mathbf{V}$ can account for one firm being typically preferred to the other. For instance, one could have $\mathbf{V}(\mathbf{x}) = \{\nu^0, \nu^1\}$ such that $\nu^0_1 > \nu^0_2$ and $\nu^1_1 < \nu^1_2$. Moreover, $\mathbf{V}$ varying with $\mathbf{x}$ captures that consumer’s preferences may vary with observed market characteristics.

Another important difference between the game played by the players and the game observed by the econometrician is that not all the equilibria of the model need to be realized in the data. Let $\mathcal{T}^* (\mathbf{x}, \mathbf{V}) \subseteq \mathcal{T} (\mathbf{x}, \mathbf{V})$ be the subset of the model’s equilibria that are realized in the data according to probabilities given by $\lambda^* (\cdot|\mathbf{x}, \mathbf{V})$. Notice that $1 \leq |\mathcal{T}^* (\mathbf{x}, \mathbf{V})| \leq |\mathcal{T} (\mathbf{x}, \mathbf{V})|$ and the special case $|\mathcal{T}^* (\mathbf{x}, \mathbf{V})| = 1$ for all $\mathbf{V} \in \mathcal{V}(\mathbf{x})$ corresponds to the single equilibrium in the data assumption, conditional on the observable $\mathbf{x}$.

When estimating an empirical game, the econometrician typically observes $M$ independent realizations of the game. For each of these realizations, the data consist in:

$$\{y_m, x_m : m = 1, \ldots, M\}.$$  \hfill (2.9)

There are therefore three random variables that are unobservable from the point of view of the researcher: (i) the private information shocks $\mathcal{E}$; (ii) the common-knowledge variables $\mathcal{V}$; and (iii) the variable indicating which equilibrium is realized in the data $\mathcal{T}$.

Let $p(y|x)$ be the observable conditional joint distribution of players’ decisions. Given Assumptions 8 to 10, such distributions are double finite mixtures of the equilibrium conditional choices realized in the data, denoted $p^* (y|x, \mathbf{V}, \tau)$. To alleviate notation a bit, let $p(y|x, \mathbf{V}) \equiv p(y|x, \mathbf{V})$ and $\Gamma (\mathbf{V}|x) \equiv \Gamma (\mathbf{V}|x)$ for $\mathbf{V} \in \mathcal{V}(\mathbf{x})$. Similarly, let $p^* (y|x, \mathbf{V}, \tau) \equiv p^* (y|x, \mathbf{V}, \tau)$ and $\lambda^* (\tau|x, \mathbf{V}) \equiv \lambda^* (\tau|x, \mathbf{V})$ for $\mathbf{V} \in \mathcal{V}(\mathbf{x})$ and $\tau \in \mathcal{T}^* (\mathbf{x}, \mathbf{V})$. Then, for a given $(y, x) \in \mathcal{Y} \times X$:

$$p(y|x) = \sum_{\mathbf{V} \in \mathcal{V}(\mathbf{x})} p(y|x, \mathbf{V}) \Gamma (\mathbf{V}|x)$$  \hfill (2.10)

where

$$p(y|x, \mathbf{V}) = \sum_{\tau \in \mathcal{T}^* (\mathbf{x}, \mathbf{V})} p^* (y|x, \mathbf{V}, \tau) \lambda^* (\tau|x, \mathbf{V}).$$  \hfill (2.11)

From equations (2.10) and (2.11), one can see that $p(y|x)$ consists in a double finite mixture with a total of $\sum_{j=0}^{\mathcal{V}(\mathbf{x})} |\mathcal{T}^* (\mathbf{x}, \mathbf{V})|$ components. In the current setting, the main object of interest is the number of equilibria associated with each $\mathbf{V} \in \mathcal{V}(\mathbf{x})$ conditional on $\mathbf{x}$. In the simple $2 \times 2 \times 2$ case, this object of interest corresponds to each element of the following set:

$$\left\{|\mathcal{T}^* (\mathbf{x}, \mathbf{V})|\right\}_{j=0}^{\mathcal{V}(\mathbf{x})} \equiv \left\{|\mathcal{T}^* (\mathbf{x}, \mathbf{V}^0)|, |\mathcal{T}^* (\mathbf{x}, \mathbf{V}^1)|\right\}.$$  \hfill (2.12)

---

7While this distinction was not made explicit in Chapter 1, it is very important to understand that the number of equilibria realized in the data can differ from the number of equilibria in the model when testing for the single equilibrium in the data assumption. In Chapter 1, one should understand $\mathcal{T} (\mathbf{x}, \mathbf{V})$ as the set of equilibria in the model that arise with positive probability in the data, i.e. the current $\mathcal{T}^* (\mathbf{x}, \mathbf{V})$. 
2.4 Nonparametric identification

2.4.1 The need for more structure

Notice that, conditional on $x$, $\nu$ and $\tau$, players’ decisions are independent. This is a consequence of Assumption 8(i). Under some conditions on $\|Y\|$, one can use recent results in the literature about the nonparametric identification of finite mixtures (Kasahara and Shimotsu, 2014), to identify the total number of components in the double finite mixture in (2.10) and (2.11), i.e. $\sum_{j=0}^{V(x)|-1} |\mathcal{T}^*(x, \nu')|$. However, in the current setting, we want to test whether or not $|\mathcal{T}^*(x, \nu')| = 1 \forall \nu \in \mathcal{V}(x)$. We are therefore not necessarily interested in this total number of components.

Unfortunately, any combinations of $|V(x)|$ and $\{|\mathcal{T}^*(x, \nu')|\}_{j=0}^{|V(x)|-1}$ that generate the same total number of components in the double finite mixture are observationally equivalent. Therefore, in order to separate $|V(x)|$ from $\{|\mathcal{T}^*(x, \nu')|\}_{j=0}^{|V(x)|-1}$, one needs to put more structure on the double finite mixture of interest.

The additional structure that I impose is based on the exclusion restriction approach of Henry, Kitamura, and Salanié (2014). While details about the exclusion restriction are presented below, one can interpret the main requirement of this approach as the availability of a proxy variable for the common-knowledge payoff-relevant unobservables. In that sense, it is fairly natural. Such proxy variable allows us to separate the mixture over $\nu$ from the mixture over $\tau$.

However, one important drawback of their approach is that it leads to a partial identification result, which may be seen as quite unfortunate since other results currently available in the literature about nonparametric identification of finite mixtures deliver point identification. The reason why point identification fails in the current setting is that, conditional on $x$ and $\nu$, players’ decisions are not independent if there are multiple equilibria realized in the data. Therefore, when identifying the finite mixture over $\nu$, one cannot use the point identification results typically associated with the conditional independence approach.

In order to recover point identification, one could interpret the double finite mixture as a single finite mixture over a discrete variable that summarizes the information contained in both $V$ and $T$. This approach is used by Aguirregabiria and Mira (2016) and Xiao (2017). Unfortunately, the finite mixture representation of the problem puts restrictions on the total number of elements in the support of such a combined variable. In other words, it potentially restricts $\sum_{j=0}^{V(x)|-1} |\mathcal{T}^*(x, \nu')|$. As a result, a test for equilibrium uniqueness based on such an identification approach would have the important drawback of restricting the number of equilibria that are identifiable from the model prior to testing. By treating the identification of the mixture over $\nu$ separately from the mixture over $\tau$, the partial identification approach proposed here does not suffer from this limitation.

2.4.2 Exclusion restriction

Let the vector of observable state variables, $\mathcal{X}$, be divided into a sub-vector of variables that do not satisfy the exclusion restriction, $\mathcal{X}_{NE}$, and a subvector of variables that do satisfy it, $\mathcal{X}_E$, such that $\mathcal{X} = [\mathcal{X}_{NE}, \mathcal{X}_E']'$, with realizations $x = [x_{NE}, x_E]' \in \mathcal{X}_{NE} \times \mathcal{X}_E \equiv \{x_{NE}^0, x_{NE}^1, \ldots, x_{NE}^{[X_{NE}]|1} \} \times \{x_E^0, x_E^1, \ldots, x_E^{[X_E]|1} \}$. Assumption 11 states conditions that must be satisfied by the variables satisfying the exclusion restriction needed for identification.
**Assumption 11** (Exclusion restriction). For any \( y \in \mathbb{Y}^2, x_{\text{NE}} \in X_{\text{NE}}, x_e \in X_e \) and \( \nu \in \mathcal{V}(x_{\text{NE}}, x_e) \): (i) (Support independence) \( \mathcal{V}(x_{\text{NE}}, x_e) \) does not depend on \( x_e \); (ii) (Cardinality of the support) \( |\mathcal{V}(x_{\text{NE}}, x_e)| \leq \min \{ |\mathbb{Y}|^2, |X_e| \} \); (iii) (Relevance) \( \Gamma(\nu|x_{\text{NE}}, x_e) \) depends on \( x_e \); and (iv) (Redundancy) \( p(y|x_{\text{NE}}, x_e, \nu) \) does not depend on \( x_e \).

The support independence condition implies that the set of values of the mixing variables realized with a positive probability does not vary with \( x_e \). This condition is important: in order to use variation in \( x_e \) to identify the finite mixture over \( \nu \in \mathcal{V} \), such variation should not generate changes in the set of possible realizations of the mixing variable. Slightly abusing notation, \( \mathcal{V}(x_{\text{NE}}) \) will be used for the rest of the chapter to make this condition explicit.

The condition on the cardinality of the support is included to make sure that there is enough variation in \( \mathcal{V} \) and in \( X_{\text{NE}} \) for the exclusion restriction approach to be able to capture all the relevant realizations of \( \mathcal{V} \). The rationale behind \( \min \{ |\mathbb{Y}|^2, |X_e| \} \) will become clear with Lemma 3.

The relevance condition requires the distribution of the unobservable \( \mathcal{V} \) to depend on both \( X_{\text{NE}} \) and \( X_e \). More precisely, one needs different values of \( x_e \) to be associated with different weights corresponding to each possible realization of \( \mathcal{V} \). Notice that this condition does not contradict the support independence one as long as realizations of \( \mathcal{V} \) do not become zero probability events for some values of \( x_e \).

Finally, by the redundancy condition, the conditional choice probabilities, the equilibrium selection mechanism and the set of equilibria in the data generating process have to be independent of \( X_e \) after conditioning on \( \mathcal{V} \).\(^8\) Once again, slightly abusing notation, such independence is made obvious by using \( p(y|x_{\text{NE}}, \nu), p^*(y|x_{\text{NE}}, \tau), \lambda^*(\tau|x_{\text{NE}}, \nu) \) and \( \Gamma^*(x_{\text{NE}}, \nu) \). In other words, \( X_e \) provides some information about the distribution of the unobservable \( \mathcal{V} \), but would not provide any information about the players’ decisions nor the equilibrium selection if \( \mathcal{V} \) would be observable.

Of course, a natural question to ask at this point is whether or not an exclusion restriction satisfying Assumption 11 is easy to find. Remember that \( \mathcal{V} \) is observed by both players, but not by the econometrician. In some sense, one simply needs an observable variable that plays the role of a proxy for the unobservable common-knowledge payoff-relevant variables. A suggestion of variables that can be used in the context of the simple market entry example is presented in Example 2 below.

**Example 2** (Simple game of market entry, continued). Given this exclusion restriction, let firm’s payoffs be written as:

\[
\pi_1(y_2, x_{\text{NE}}, \nu_1, \varepsilon_1) = x_{\text{NE},1} + \nu_1 - 4y_2 - \varepsilon_1 \tag{2.13}
\]

\[
\pi_2(y_1, x_{\text{NE}}, \nu_2, \varepsilon_2) = x_{\text{NE},2} + \nu_2 - 3y_1 - \varepsilon_2. \tag{2.14}
\]

Provided that there is some level of correlation between the realizations of \( \mathcal{V} \) across markets, the realizations of \( X_{\text{NE}} \) in other markets can be used as a proxy for \( \mathcal{V} \). In other words, one can use some of the observable common-knowledge state variables in surrounding markets to control for common-knowledge payoff-relevant unobservables in a given market. Here is the argument. Consider a market \( m \) and let

\(^8\) Notice that the redundancy condition implies independence between \( \mathcal{Y} \) and \( X_e \) after conditioning on \( X_{\text{NE}} \) and \( \mathcal{V} \). In that sense, one may find the exclusion restriction approach for the identification of finite mixtures to be somewhat similar to the alternative approach relying on the independence of some observable variables after conditioning on the latent variable. However, if one was to leverage such independence between \( \mathcal{Y} \) and \( X_e \) to identify the mixture over \( \mathcal{V} \) as in the conditional independence approach, one would not be allowed to condition the distribution of the latent variable \( \mathcal{V} \) on \( X_e \).
denote some surrounding market(s). In each market, $V$ is correlated with $X_{\text{NE}}$. For instance, the unobservable information that both firms may have regarding consumers’ preferences in a given market is typically correlated with this market’s observable demographics. Provided that $\nu_{m'}$ is correlated with $\nu_m$, then $x_{\text{NE},m'}$ contains some information about $\nu_m$ through its correlation with $\nu_{m'}$. $x_{\text{NE},m'}$ therefore satisfies the relevance condition. Furthermore, by the definition of the BNE, only market $m$’s state variables affect firms’ decisions and equilibrium selection in this market. Therefore, $x_{\text{NE},m'}$ is only informative about market $m$ when one fails to observe $\nu_m$. In other words, $x_{\text{NE},m'}$ satisfies the redundancy requirement. Finally, a necessary condition for the support constraint to hold is that the support of $x_{\text{NE},m'}$ must be large enough to capture the different possible realizations $\nu_m$, which can be checked for a given $|V(x_{\text{NE}})|$.

There is one potential caveat that should be pointed out. Even if being able to treat the mixture over $\nu$ separately from the mixture over $\tau$ avoids restricting the number of equilibria in the data generating process, the exclusion restriction approach still puts an upper bound on $|V(x_{\text{NE}})|$ identifiable from the data.$^9$ The cardinality of the support condition in Assumption 11 assumes away the potential issue caused by the restriction on $|V(x_{\text{NE}})|$ implied by the finite mixture framework. Of course, one could simply impose a similar restriction in order to use Aguirregabiria and Mira (2016)’s identification results to test for equilibrium uniqueness in the data generating process. However, as mentioned above, the needed assumption would restrict both $|V(x)|$ and $|T^*(x,\nu)| \forall \nu \in V(x)$. Also, the support condition in Assumption 11 allows for a larger $|V(x_{\text{NE}})|$ than in Aguirregabiria and Mira (2016).

To wrap up the discussion about the exclusion restriction needed in the current identification argument, it is important to note that this exclusion restriction is different from the one commonly used to identify payoffs in static games of incomplete information (e.g., Bajari, Hong, Krainer, and Nekipelov, 2010). The latter requires the existence of player-specific variables that do not affect opponents’ payoffs. The main reason why this conventional exclusion restriction is not used here is because the current method relies on the joint distribution of players’ decisions given all common-knowledge payoff-relevant information, but still mixed over the equilibria realized in the data. However, player specific observables provide identifying power only if one is trying to identify the marginal distributions of the players’ decisions. As a result, the conventional exclusion restriction requires the probabilities to be conditioned on the equilibrium realized in the data which, in the finite mixture framework, puts an upper bound on the number of equilibria that can be identified.

Even though we are now considering the simple case of $|V(x_{\text{NE}})| = 2$, it is worth noting that $|V(x_{\text{NE}})|$ is identifiable given the exclusion restriction. Let $\Delta P(x_{\text{NE}})$ be the $(|X_E| - 1) \times |Y|^2$ matrix with element $(i, j)$ given by $p(x^1|\nu, x_E^1) - p(x^2|\nu, x_E^2)$, for $i = 1, \ldots, |X_E| - 1$ and $j = 1, \ldots, |Y|^2$. As it is stated in Lemma 3, for each $x_{\text{NE}} \in X_{\text{NE}}$, $|V(x_{\text{NE}})|$ is identified through the rank of the matrix $\Delta P(x_{\text{NE}})$.

**Lemma 3** (Identification of $|V|$). Under Assumptions 8 to 11, for each $x_{\text{NE}} \in X_{\text{NE}}$, $|V(x_{\text{NE}})| = \text{rank} \{\Delta P(x_{\text{NE}})\} + 1$.

**Proof.** This result follows from Henry, Kitamura, and Salanié (2014, Lemma 2, p.138). Notice that they only consider mixtures of marginal distributions. However, their result also applies to mixtures of joint distributions, provided that one is interested in the joint conditional component distributions themselves, not the corresponding marginal distributions.

$^9$More details on the number of components that can be identified are given below.
Since the elements in matrix \( \Delta P(x_{NE}) \) are probabilities, its column rank is at most \( |Y|^2 - 1 \). Therefore, since the rank of a matrix is bounded by the minimum number of its rows and columns, the finite mixture representation restricts \( |V(x_{NE})| \) to be at most \( \min\{|Y|^2, |X_E|\} \), which corresponds to the cardinality of the support condition stated in Assumption 11.

### 2.4.3 Constructing the identified set

The identification result presented in this chapter is conditional on \( x_{NE} \), such that all functions and statistics presented below depend on \( x_{NE} \). In order to alleviate notation, I omit \( x_{NE} \) as an argument, but the reader should keep in mind that all the derivations that follow hold for a given realization \( x_{NE} \).

Since we are considering the simple \( 2 \times 2 \times 2 \) case, let \( \Gamma(x_E) \equiv \Gamma(\nu^1|x_E) \) such that \( 1 - \Gamma(x_E) = \Gamma(\nu^0|x_E) \). Under Assumptions 8 to 11, (2.10) and (2.11) become:

\[
p(y|x_E) = p(y|\nu^0)[1 - \Gamma(x_E)] + p(y|\nu^1)\Gamma(x_E)
\]

where

\[
p(y|\nu^j) = \sum_{\tau \in \mathcal{T}(\nu^j)} p^*(y|\nu^j, \tau) \lambda^*(\tau|\nu^j).
\]

Let \( \theta \equiv [\phi, \Upsilon]' \) where \( \phi \equiv \Gamma(x_E^0) \) and \( \Upsilon \equiv \Gamma(x_E^1) - \Gamma(x_E^0) \). The main intuition behind the identification argument is as follows. Given the exclusion restriction introduced above, one can construct the identified set \( \Theta_1 \) such that each \( \theta \) in that set can rationalize the data through a finite mixture over \( \nu \in \mathcal{V} \) and a single equilibrium being realized for each \( \nu \in \mathcal{V} \). For a given \( x_{NE} \), testing for the single equilibrium in the data assumption, i.e. \( |\Upsilon^*(\nu)| = 1 \) \( \forall \nu \in \mathcal{V} \), boils down to testing whether or not \( \Theta_1 \) is empty.

To construct \( \Theta_1 \), I follow very closely the partial identification approach proposed by Henry, Kitamura, and Salanié (2014). More precisely, I rewrite the unknown probabilities \( p(y|\nu) \) and \( \Gamma(x_E) \) as functions of observable probabilities \( p(y|x_E) \)'s and \( \theta \). There are two types of restrictions that must be satisfied by \( \theta \) to belong to the identified set. In particular, \( \theta \in \Theta_1 \) if:

1. the resulting \( p(y|\nu) \)'s and \( \Gamma(x_E) \)'s are proper probabilities, i.e. they belong to the unit interval; and

2. \( p(y|\nu) \)'s satisfy the conditions implied by a single equilibrium being realized in the data.

The first type of restrictions are satisfied if the data can be rationalized through a finite mixture over the common-knowledge payoff-relevant unobservables. In that sense, they correspond to the restrictions provided in Henry, Kitamura, and Salanié (2014). The second ones can be interpreted as extra restrictions that are satisfied if the single equilibrium in the data assumption holds for each \( \nu \in \mathcal{V} \) at a given \( x_{NE} \).

Similarly as in Henry, Kitamura, and Salanié (2014)'s, notice that (2.15) can be written as:

\[
p(y|x_E) = p(y|\nu^0) + [p(y|\nu^1) - p(y|\nu^0)] \Gamma(x_E).
\]

By evaluating (2.17) at \( x_E^1 \) and \( x_E^0 \), it follows that:

\[
p(y|x_E^1) - p(y|x_E^0) = [p(y|\nu^1) - p(y|\nu^0)] \Upsilon
\]
and

\[ p(y|x_E) - p(y|x_E^0) = [p(y|\nu^1) - p(y|\nu^0)] [\Gamma(x_E) - \phi]. \]  

(2.19)

By combining (2.18) and (2.19), one gets:

\[ \Gamma(x_E) = \phi + \frac{\nu \{p(y|x_E) - p(y|x_E^0)\}}{p(y|x_E^1) - p(y|x_E^0)}. \]  

(2.20)

By rearranging (2.17) evaluated at \( x_E^0 \) and using (2.18), one gets:

\[ p(y|\nu^0) = p(y|x_E^0) - \frac{\phi}{T} \left[ p(y|x_E^1) - p(y|x_E^0) \right]. \]  

(2.21)

Finally, noting that \( p(y|\nu^1) = p(y|\nu^0) + p(y|\nu^1) - p(y|\nu^0) \) and using (2.18) together with (2.21), it follows that:

\[ p(y|\nu^1) = p(y|x_E^0) + \frac{1 - \phi}{T} \left[ p(y|x_E^1) - p(y|x_E^0) \right]. \]  

(2.22)

The unobservable probabilities in (2.20), (2.21) and (2.22) are therefore functions of \( \Theta \) and observable probabilities. Let’s now turn to the conditions defining \( \Theta_1 \). Consider the following functions:

\[ L_0(y) \equiv \frac{-p(y|x_E^0)}{p(y|x_E^1) - p(y|x_E^0)}; \]  

(2.23)

\[ L_1(y) \equiv \frac{1 - p(y|x_E^0)}{p(y|x_E^1) - p(y|x_E^0)}; \]  

(2.24)

\[ Q(x_E) \equiv \frac{p(y|x_E) - p(y|x_E^0)}{p(y|x_E^1) - p(y|x_E^0)} \]  

(2.25)

Moreover, let:

\[ a_0 \equiv p(0,0|x_E^0) p(1,1|x_E^0) - p(1,0|x_E^0) p(0,1|x_E^0); \]  

(2.26)

\[ a_1 \equiv p(0,0|x_E^0) \left[ p(1,1|x_E^0) - p(1,1|x_E^0) \right] + p(1,1|x_E^0) \left[ p(0,0|x_E^0) - p(0,0|x_E^0) \right] 
- p(1,0|x_E^0) \left[ p(0,1|x_E^0) - p(0,0|x_E^0) \right] - p(0,1|x_E^0) \left[ p(1,0|x_E^0) - p(1,0|x_E^0) \right]; \]  

(2.27)

\[ a_2 \equiv \left[ p(0,0|x_E^0) - p(0,0|x_E^0) \right] \left[ p(1,1|x_E^0) - p(1,1|x_E^0) \right] 
- \left[ p(0,1|x_E^0) - p(0,1|x_E^0) \right] \left[ p(1,0|x_E^0) - p(1,0|x_E^0) \right]. \]  

(2.28)

Without loss of generality, let \( T > 0 \). Then, the identified set is defined in Proposition 2.

**Proposition 2** (Identified set). **Under Assumptions 8 to 11, \( \Theta \in \Theta_1 \) provided that:**

(i) \( \max_{y \in Y^2} \{ \min \{ L_0(y), L_1(y) \} \} \leq \frac{-\phi}{T} < \min_{x_E \in x_E} \{ Q(x_E) \}; \)

(ii) \( \max_{x_E \in x_E} \{ Q(x_E) \} \leq \frac{1 - \phi}{T} \leq \min_{y \in Y^2} \{ \max \{ L_0(y), L_1(y) \} \}; \)
(iii) \( a_2 \left[ -\frac{\phi}{Y} \right]^2 - a_1 \frac{\phi}{Y} + a_0 = 0; \)

(iv) \( a_2 \left[ 1 - \frac{\phi}{Y} \right]^2 + a_1 \frac{1 - \phi}{Y} + a_0 = 0. \)

Proof. See Appendix 2.7.1.

### 2.4.4 Properties of the identified set

I now turn to some interesting properties of \( \Theta_I \) collected in two corollaries. Corollary 1 states that, in the simple \( 2 \times 2 \times 2 \) case, if \( \Theta_I \) is not empty, it is a singleton.

**Corollary 1** (Singleton nonempty \( \Theta_I \)). If \( \Theta_I \neq \emptyset \), then \( \Theta_I \) is a singleton.

**Proof.** See Appendix 2.7.2.

From Proposition 2 and Corollary 1, it is easy to represent the identified set graphically for the reparameterization \(-\phi/Y \) and \((1 - \phi)/Y \). In Figure 2.2, \( \Theta_I \) is the singleton \((-\phi/Y)^*, (-\phi/Y)^{**}\). The shaded area corresponds to \( \theta \in \Theta_I \) that satisfy conditions (i) and (ii) from Proposition 2. The couple \((-\phi/Y)^*, (-\phi/Y)^{**}\), if it exists, is the only one that satisfies conditions (iii) and (iv). As a result, \( \Theta_I \neq \emptyset \) if and only if this point falls in the shaded area, i.e.

\[
\max_{y \in Y^2} \{ \min \{ L_0(y), L_1(y) \} \} \leq \left( -\frac{\phi}{Y} \right)^* < \min_{x_E \in X_E} \{ Q(x_E) \}; \tag{2.29}
\]

\[
\max_{x_E \in X_E} \{ Q(x_E) \} < \left( \frac{1 - \phi}{Y} \right)^{**} \leq \min_{y \in Y^2} \{ \max \{ L_0(y), L_1(y) \} \}. \tag{2.30}
\]

![Figure 2.2: Graphical representation of the identified set for the 2 × 2 × 2 case](image)

The identified set \( \Theta_I \) is sharp in the following sense. Given \( x_{NE} \), if the true data generating process corresponds to a single equilibrium being realized at \( \nu^0 \) and \( \nu^1 \), then \( \Theta_I \) is not empty. Conversely, if \( \Theta_I \)
is not empty, then the joint distribution of the players' decisions conditional on \( x_{\text{NE}} \) can be rationalized through a single equilibrium at \( \nu^0 \) and \( \nu^1 \). One drawback of the partial identification approach is that the latter could hold even if the true data generating process does not correspond to a single equilibrium for \( \nu^0 \) and/or \( \nu^1 \). Nonetheless, a test based on \( \Theta_1 \) is still informative since \( \Theta_1 \) being empty implies that the data cannot be rationalized by the single equilibrium in the data assumption and a finite mixture over the common-knowledge payoff-relevant unobservables given \( x_{\text{NE}} \). In other words, \( \Theta_1 \neq \emptyset \) is necessary, but not sufficient for equilibrium uniqueness. This limitation of the test can easily be understood from Figure 2.2: it is due to the fact that we do not observe the true \( p(y|\nu) \)'s, but we know that they belong to a set. Some of the probabilities in this set may be rationalized through a single equilibrium even if the data is generated by multiple ones. Intuitively, the more informative the bounds coming from conditions (i) and (ii) in Proposition 2 are, the less likely this problem is to arise.

Of course, given that we are representing the joint distribution of players’ decisions as a double finite mixture, a natural question to ask is to which extent are we able to separate the two levels of mixing, i.e. the mixture over multiple equilibria and the mixture over the common-knowledge payoff-relevant unobservables? Here, the exclusion restriction is key. In some sense, it allows us to identify \( p(y|\nu) \)'s, regardless of the number of equilibria realized at each \( \nu \in V \), using conditions (i) and (ii) in Proposition 2. Then, whether one can factor \( p(y|\nu) \) as the product of \( p(y_1|\nu) \) and \( p(y_2|\nu) \), i.e. conditions (iii) and (iv), determines whether one can rationalize the data through the single equilibrium assumption.

One important feature of \( \Theta_1 \), as stated in Corollary 2 below, is that the emptiness of \( \Theta_1 \) does not depend on the weights associated with each realization of the common-knowledge payoff-relevant unobservables. More precisely, consider two different data generating processes that are associated with the same \( p(y|\nu) \)'s, but different \( \Gamma(x_E)'s \). Corollary 2 states that the identified sets corresponding to each data generating processes are either both empty or both nonempty.

**Corollary 2** (Emptiness of \( \Theta_1 \) independent of \( \Gamma(x_E) \)). Let \( \theta^0 \neq \emptyset \) correspond to two different data generating processes with choice probabilities such that \( \bar{p}^0(y|\nu) = \bar{p}^0(y|\nu) \forall y \in Y^2 \) and \( \forall \nu \in V \). Then, the corresponding identified sets are such that \( \Theta_1 \neq \emptyset \) if and only if \( \bar{\Theta}_1 \neq \emptyset \).

**Proof.** See Appendix 2.7.3.

Corollary 2 highlights an important property of the identified set. For a given vector of equilibrium-specific choice probabilities, whether or not \( \Theta_1 \) is empty only depends on the equilibrium selection mechanism, not on the weights associated with each realization of the common-knowledge payoff-relevant unobservables. In that sense, a test based on \( \Theta_1 \neq \emptyset \) effectively separates the two level of mixing.

### 2.4.5 Identifying power of \( \Theta_1 \neq \emptyset \)

As already mentioned above, a potential drawback of a test based on the testable implication \( \Theta_1 \neq \emptyset \) is that it tests a necessary, but insufficient condition for equilibrium uniqueness. If there actually is a single equilibrium in the data generating process (and the joint distribution of players’ decisions can be rationalized through a finite mixture over \( \nu^0 \) and \( \nu^1 \)), then \( \Theta_1 \) is not empty. However, it could be the case that \( \Theta_1 \) is not empty even if the data has been generated by multiple equilibria. In that case, despite equilibria multiplicity in the data generating process, the data could be rationalized by a single equilibrium. Therefore, for a test based on \( \Theta_1 \neq \emptyset \) to be informative, it must be the case that \( \Theta_1 \) is empty at least for some data generating processes. In theory, there are two reasons why \( \Theta_1 \) may be empty:
Chapter 2. Sharp test for equilibrium uniqueness in discrete games

1) The data cannot be rationalized by a finite mixture over \( \nu \in \mathcal{V} \). This would be the case if (i) and/or (ii) from Proposition 2 do not hold for any \( \theta \in \Theta \). More precisely, the data cannot be rationalized by such a finite mixture if either:

\[
\min_{x_E \in \mathcal{X}_E} \{ Q(x_E) \} \leq \max_{y \in \mathcal{Y}} \{ \min \{ L_0(y), L_1(y) \} \}; \tag{2.31}
\]

\[
\min_{y \in \mathcal{Y}} \{ \max \{ L_0(y), L_1(y) \} \} \leq \max_{x_E \in \mathcal{X}_E} \{ Q(x_E) \}. \tag{2.32}
\]

2) The joint distribution of players’ decisions conditional on \( \nu \in \mathcal{V} \) cannot be factored as the product of their marginal distributions. In other words, either (iii) and/or (iv) from Proposition 2 do not hold. This could either be due to the quadratic equations not admitting a real solution or admitting solutions that fall outside of the bounds in (i) and (ii).

I now provide simulation evidence to show that \( \Theta_1 \) may indeed be empty in practice. I use the simple data generating process introduced in Example 2. Similarly as before, let \( p_i(\nu, \tau) \equiv p(y_i = 1|\nu, \tau) \). Given \( x_{NE} \) and \( \nu \), a BNE in pure strategies is such that:

\[
p^*(\nu, \tau) \equiv \begin{bmatrix} p^*_1(\nu, \tau) \\ p^*_2(\nu, \tau) \end{bmatrix} = \begin{bmatrix} \Phi(x_{NE,1} + \nu_1 - 4p^*_2(\nu, \tau)) \\ \Phi(x_{NE,2} + \nu_2 - 3p^*_1(\nu, \tau)) \end{bmatrix}. \tag{2.33}
\]

I consider two different data generating processes that take the same values of \( x_{NE} \equiv \{x_{NE,1}, x_{NE,2}\}' = [1, 1]' \), but vary according to the values of \( \nu^0 \) and \( \nu^1 \). In the first one, let \( \nu^0 = [1, 0.5]' \) and \( \nu^1 = [1.25, 1]' \). This case admits three solutions for each \( \nu \in \mathcal{V}(x_{NE}) \), with two of these three equilibria being stable. Even though stability is not required for the identification result to hold, I focus on stable equilibria since they have a more natural economic interpretation. These equilibria are such that \( p^*(\nu^0, \tau = 1) = [0.0499, 0.9116]' \), \( p^*(\nu^0, \tau = 2) = [0.9501, 0.0884]' \) for \( \nu^0 \); \( p^*(\nu^1, \tau = 1) = [0.9003, 0.2417]' \), \( p^*(\nu^1, \tau = 2) = [0.0527, 0.9672]' \) for \( \nu^1 \).

Figure 2.3 summarizes whether or not \( \Theta_1 \neq \emptyset \) for \( \lambda^*(\tau|\nu) \in \{0, 0.05, \ldots, 0.95, 1\} \) with the vertical axis corresponding to \( \lambda^*(\tau = 1|\nu^0) \) and the horizontal axis corresponding to \( \lambda^*(\tau = 1|\nu^1) \). While \( \Theta_1 \) is empty in the white area, it is nonempty in the black one. For \( \{\lambda^*(\tau = 1|\nu^0), \lambda^*(\tau = 1|\nu^1)\} \in \{(0, 0), (1, 0), (0, 1), (1, 1)\} \), i.e. the four “corners”, we have \( \Theta_1 \neq \emptyset \). This observation is as expected: these four equilibrium selection mechanisms satisfy the single equilibrium in the data assumption. If the whole area was black, a test based on \( \Theta_1 \neq \emptyset \) would not be informative. One could then always rationalize these mixtures over two equilibria by a single equilibrium. The presence of the white area implies that for some equilibrium selection mechanisms, the resulting joint distributions of the players’ decisions cannot be rationalized by a single equilibrium.

The second data generating process is such that \( \nu^0 = [1.25, 1]' \) and \( \nu^1 = [1.30, 1]' \). The equilibria kept in the data generating process are \( p^*(\nu^0, \tau = 1) = [0.0590, 0.9659]' \), \( p^*(\nu^0, \tau = 2) = [0.9201, 0.2235]' \) for \( \nu^0 \); \( p^*(\nu^1, \tau = 1) = [0.9003, 0.2417]' \), \( p^*(\nu^1, \tau = 2) = [0.0527, 0.9672]' \) for \( \nu^1 \). Figure 2.4 suggests that \( \Theta_1 \) is nonempty for most equilibrium selection mechanisms. Notice that an important difference between this data generating process and the previous one is that equilibria do not vary much across \( \nu \).
Chapter 2. Sharp test for equilibrium uniqueness in discrete games

Note: $\Theta_I$ is empty in white area.

Figure 2.3: Emptiness of $\Theta_I$ – First data generating process

Note: $\Theta_I$ is empty in white area.

Figure 2.4: Emptiness of $\Theta_I$ – Second data generating process
2.5 Statistical test

2.5.1 A two-stage minimum distance approach

The identification result derived above suggests that $\Theta_1$ being nonempty is an informative necessary condition of the single equilibrium assumption in the presence of common-knowledge payoff-relevant unobservables. More formally, this necessary condition can be tested as:

$$H_0 : \Theta_1 \neq \emptyset ; \quad H_1 : \Theta_1 = \emptyset.$$  \hspace{1cm} (2.34)

In other words, the proposed test of the single equilibrium in the data assumption boils down to a specification test in partial identification.

There are three important features of the problem at hand that should be taken into account when thinking about the inference methodology to be used to test this null hypothesis. First, the conditions defining $\Theta_1$ are functions of some estimable probabilities, which are expectations of discrete variables. However, since these expectations enter the conditions nonlinearly, the usual moment inequalities framework does not readily apply. Second, these probabilities are unknown to the researcher and must therefore be consistently estimated prior to the construction of the identified set. Third, some of the conditions defining $\Theta_1$ are equality constraints.

The two-stage minimum distance estimator proposed by Shi and Shum (2015) allows us to take advantage of these three features of the inference problem. Their approach can be interpreted as an extension of classical minimum distance estimation procedures to include inequality constraints and partial identification. In order to apply their method, one must be able to write the conditions defining $\Theta_1$ as:

1. Some equality constraints that depend on first-stage probabilities, the parameters of interest and potentially some nuisance parameters.
2. Some inequality constraints that depend on the parameters of interest and potentially some nuisance parameters, but not the first-stage probabilities.

Let $p$ denote the vector of joint probabilities of players’ decisions $\forall y \in Y^2$ and $\forall x_E \in X_E$. Let $\hat{p}$ be the first-stage estimate of the population probabilities $p^0$ such that:

$$M^{1/2} (\hat{p} - p^0) \overset{d}{\rightarrow} N (0, \Sigma).$$ \hspace{1cm} (2.35)

Moreover, let $\eta \in H$ be a vector of nuisance parameters to be defined below. Essentially, these nuisance parameters are used to transform inequalities that depend on $p$ into equalities.

There are two types of equality conditions in this context. First, let the vector $g^e (p, \theta)$ be some equality conditions such that $g^e (p, \theta) = 0$. These are conditions defining the identified set that do not need the introduction of nuisance parameters to be written as equalities, which is the reason why they do not depend on $\eta$. Second, let $g^{ie,1} (p, \theta) - \eta$ be a vector of equality conditions such that $g^{ie,1} (p, \theta) - \eta = 0$. These conditions correspond to inequalities that have been transformed into equalities by introducing nuisance parameters. Typically, one must rewrite all equalities such that the matrix of their joint variance (defined below) is full rank. Finally, one must introduce the remaining inequality conditions. Let $g^{ie,2} (\theta, \eta)$, which do not depend on $p$, be such that $g^{ie,2} (\theta, \eta) \geq 0$. In the simple $2 \times 2 \times 2$ case, $g^e (p, \theta)$ consists in:
The inequality conditions transformed into equality ones in $g^{ie,1}(p, \theta) - \eta$ are:

\[
p(y|x_0^E) - \frac{\phi}{\hat{\Gamma}} [p(y|x_1^E) - p(y|x_0^E)] - \eta_0(y), \ y \in \{(0,0), (1,0)\};
\]

\[
p(y|x_0^E) + \frac{1-\phi}{\hat{\Gamma}} [p(y|x_1^E) - p(y|x_0^E)] - \eta_1(y), \ y \in \{(0,0), (1,0)\};
\]

\[
\phi + \Gamma \left[\frac{p(0,0|x_E) - p(0,0|x_E^0)}{p(0,0|x_E^0)}\right] - \eta_0(x_E), \ \forall x_E \in \mathcal{X}_E \setminus \{x_0^0, x_1^0\}.
\]

The inequality conditions in $g^{ie,2}(\theta, \eta)$ are:

\[
\eta_j(y), \ y \in \{(0,0), (1,0)\}, \ j \in \{0,1\};
\]

\[
1 - \eta_j(y), \ y \in \{(0,0), (1,0)\}, \ j \in \{0,1\};
\]

\[
1 - \eta_j(0,0) - \eta_j(1,0), \ j \in \{0,1\};
\]

\[
\eta_0(x_E), \ \forall x_E \in \mathcal{X}_E \setminus \{x_0^0, x_1^0\};
\]

\[
1 - \eta_0(x_E), \ \forall x_E \in \mathcal{X}_E \setminus \{x_0^1, x_1^1\};
\]

\[
\phi; \ 1 - \phi; \ \phi + \gamma; \ \text{and} \ 1 - \phi - \gamma.
\]

At a given $\theta$ the full vector of equality conditions evaluated at the first stage estimates $\hat{p}$, i.e. $g(\hat{p}, \theta, \eta) \equiv [g^e(\hat{p}, \theta), g^{ie,1}(\hat{p}, \theta) - \eta]'$, inherits the asymptotic normality of $\hat{p}$:

\[
M^{1/2} \left(g(\hat{p}, \theta, \eta) - g(p^0, \theta, \eta)\right) \xrightarrow{d} N(0, W(\theta))
\]

where $W(\theta) \equiv [G(p, \theta, \eta) \Sigma G(p, \theta, \eta)']$ and $G(p, \theta, \eta)$ is the Jacobian of $g(p, \theta, \eta)$ with respect to $p$.

Consider the following profiled criterion function: \(^{10}\)

\[
a_2 \left[\frac{-\phi}{\hat{\Gamma}}\right]^2 - a_1 \frac{\phi}{\hat{\Gamma}} + a_0;
\]

\[
a_2 \left[\frac{1-\phi}{\hat{\Gamma}}\right]^2 + a_1 \frac{1-\phi}{\hat{\Gamma}} + a_0.
\]

\(^{10}\)This criterion is slightly different from the one in Shi and Shum (2015) since it does not weight the equality conditions by the inverse of $W(\theta)$. While the test statistic derived from (2.48) is not pivotal, it avoids having to invert a matrix that may be close to singular due to sampling error even if it is invertible in the population.
\[ Q(p, \theta) = \min_{\eta \in \mathcal{H}} \left[ g^c(p, \theta) \begin{bmatrix} g^{ie,1}(p, \theta) - \eta \\ g^{ie,1}(p, \theta) - \eta \end{bmatrix} \right]' \left[ g^{ie,2}(p, \theta) \right] \]

\[ \text{s.t. } g^{ie,2}(\theta, \eta) \geq 0. \]  

(2.48)

Given the quadratic form of this criterion function, it reaches its minimum (zero) at \( p^0 \) and \( \theta \in \Theta_1 \). Therefore, one can define the identified set as:

\[ \Theta_1 = \{ \theta \in \Theta : Q(p^0, \theta) = 0 \}. \]  

(2.49)

A natural estimate of \( \Theta_1 \) is obtained by minimizing the sample version of this criterion function:

\[ \hat{\Theta}_1 = \arg \min_{\theta \in \Theta} Q_M(p, \theta). \]  

(2.50)

The corresponding confidence set is defined as:

\[ \text{CS}_M(\alpha) = \{ \theta \in \Theta : MQ_M(p, \theta) \leq c_M(\theta, 1 - \alpha) \}. \]  

(2.51)

One can take advantage of the asymptotic normality of \( g(p, \theta, \eta) \) when computing critical values \( c_M(\theta, 1 - \alpha) \) used to construct this confidence set. In the presence of nuisance parameters, Shi and Shum (2015) propose to use the corresponding quantile of the variable \( J_M(\theta) \) defined as:\(^{11}\)

\[ J_M(\theta) = \min_{h \in \mathcal{H}(\theta) - \hat{\eta}(\theta)} \left[ Z^c_M - \kappa^{-1}_M M^{1/2} h \right]' \left[ Z^c_M - \kappa^{-1}_M M^{1/2} h \right] \]  

(2.52)

where \( \mathcal{H}(\theta) \equiv \{ \eta \in \mathcal{K} : g^{ie,2}(\theta, \eta) \geq 0 \} ; \{ \kappa_M \} \) is a sequence of tuning parameters that diverges to \( \infty \); \( Z^c_M, Z^c_M^{-1} \)' \sim N(0, W_0(\theta)) \); and

\[ \hat{\eta}(\theta) = \min_{\eta \in \mathcal{K}} \left[ g^{ie,1}(p, \theta) - \eta \right]' \left[ g^{ie,1}(p, \theta) - \eta \right] \]  

\[ \text{s.t. } g^{ie,2}(\theta, \eta) \geq 0. \]  

(2.53)

The ability of computing the critical values used to construct \( \text{CS}_M(\alpha) \) by drawing from a normal distribution is an important computational advantage of Shi and Shum (2015)’s approach over other estimators that rely on resampling methods. Another important advantage of the current setting that is worth mentioning is that one does not need to construct the full identified set to test the hypothesis of interest, provided that \( g(p, \theta, \eta) \) is smooth in \( \theta \). In fact, even if \( \theta \) minimizing \( Q_M(p, \theta) \) is not unique, the minimum value of \( Q_M(p, \theta) \) is unique.\(^{12}\)

For smooth \( g(p, \theta, \eta) \), one can easily find one of the \( \theta \)'s that minimize \( Q_M(p, \theta) \) using gradient methods and check if \( MQ_M(p, \theta) \leq c_M(\theta, 1 - \alpha) \). In other words, since one does not need to estimate the identified set in itself, but simply needs to check if it is

\(^{11}\)Again, while Shi and Shum (2015) use the inverse of \( W(\theta) \) as a weighting matrix when constructing \( J_M(\theta) \), the current definition of \( J_M(\theta) \) does not. This is because even if \( W(\theta) \) is invertible, its sample counterpart \( W(\theta) \) may be close to not being invertible, therefore creating some numerical issues that make the constructed quantiles of \( J_M(\theta) \) unreliable.

\(^{12}\)In the simple \( 2 \times 2 \times 2 \) case, \( \Theta_1 \) is a singleton under \( H_0 : \Theta_1 \neq \emptyset \), but \( Q_M(p, \theta) \) may still admit several minimizers when \( \Theta_1 = \emptyset \).
Chapter 2. Sharp test for equilibrium uniqueness in discrete games

empty, the optimization problem to solve is actually as simple as it would be in a point-identification setting.

\[ MQ_M \left( \hat{p}, \hat{\theta} \right) \] is therefore the statistic of interest. The asymptotic level of a test based on this statistic is discussed in the next section.

### 2.5.2 Asymptotic level

Consider the following non-randomized decision rule:

\[
\xi_M = \begin{cases} 
1, & \text{if } MQ_M \left( \hat{p}, \hat{\theta} \right) \geq c_M \left( \hat{\theta}, 1 - \alpha \right) \\
0, & \text{otherwise}
\end{cases}
\] (2.54)

Let \( E_{\theta} \left[ \cdot \right] \) denote the expectation under the data generating process corresponding to \( \theta \). In particular, for \( \theta \in \Theta_I \), \( E_{\theta} \left[ \xi_M \right] \) is the probability of rejecting \( H_0 : \Theta_I \neq \emptyset \) when it's true. Proposition 3 states that the statistical test of this null hypothesis based on the decision rule \( \xi_M \) is asymptotically level \( \alpha \).

**Proposition 3** (Asymptotic level of the test). Suppose that: (i) \( G(p, \theta, \eta) \) is well defined and continuous in \( \theta \) and \( p \); (ii) \( W(\theta) \overset{\Delta}{=} W(\theta) \); (iii) \( \Theta \) is compact; (iv) \( g(p, \theta, \eta) \) is continuous in \( \theta \) for all \( p \); (v) \( g^{\pi,2}(\theta, \eta) \) is continuous in \( \theta \) and \( \eta \); (vi) \( \mathcal{H}(\theta) \) is convex \( \forall \theta \in \Theta_I \); and (vii) \( \kappa_M \to \infty \). Then:

\[
\limsup_{M \to \infty} \sup_{\theta \in \Theta_I} E_{\theta} \left[ \xi_M \right] \leq \alpha.
\] (2.55)

**Proof.** Shi and Shum (2015, Theorem 4.1, p. 503) show that:

\[
\liminf_{M \to \infty} \inf_{\theta \in \Theta_I} \Pr \left( \theta \in CS_M (\alpha) \right) \geq 1 - \alpha.
\] (2.56)

Since \( \hat{\theta} \) minimizes \( Q_M(\hat{p}, \hat{\theta}) \), \( E_{\theta} \left[ \xi_M \right] = \Pr \left( \theta \in CS_M (\alpha) \right) = \emptyset \). Moreover, for any \( \theta \in \Theta_I \), since \( CS_M (\alpha) = \emptyset \Rightarrow \theta \notin CS_M (\alpha) \), it follows that \( \Pr \left( \theta \notin CS_M (\alpha) \right) \geq \Pr \left( \Theta \neq \emptyset \right) \). Therefore:

\[
\limsup_{M \to \infty} \sup_{\theta \in \Theta_I} E_{\theta} \left[ \xi_M \right] \leq \limsup_{M \to \infty} \sup_{\theta \in \Theta_I} \Pr \left( \theta \notin CS_M (\alpha) \right) \leq 1 - \liminf_{M \to \infty} \inf_{\theta \in \Theta_I} \Pr \left( \theta \in CS_M (\alpha) \right) \leq \alpha.
\] (2.57)

which completes the proof. \( \square \)

The properties of the test are studied through Monte-Carlo experiments below.

### 2.5.3 Monte-Carlo experiments

I now provide simulation evidence in order to investigate the statistical size and power of a test based on the statistic \( MQ_M \left( \hat{p}, \hat{\theta} \right) \) and the decision rule \( \xi_M \). More precisely, I investigate the probability of rejecting the null hypothesis \( \Theta_I \neq \emptyset \) for equilibrium selection mechanisms that correspond to the single-equilibrium in the data assumption and some equilibrium selection mechanisms that correspond to multiple equilibria.
Once again, I consider the same data generating process as in the simple game of market entry introduced in Example 2. Let the values of $x_{\text{NE}}$, $\nu^0$ and $\nu^1$ be as in the first data generating process used when investigating the identifying power of the testable restriction in Section 2.4.5. The weights associated with $\nu^1$ are $\Gamma(x_E^0) = 0.2$ and $\Gamma(x_E^1) = 0.9$. Given these probabilities, it follows that $\phi = 0.2$ and $\gamma = 0.7$.

Different data generating processes are created by varying equilibrium selection mechanisms. Consider $\lambda^\ast (\tau = 1|\nu^j) \in \{0, 0.2, 0.4, 0.6, 0.8, 1\}$ for $j = 0, 1$. The data is generated by a single equilibrium when both $\lambda^\ast (\tau = 1|\nu^0)$ and $\lambda^\ast (\tau = 1|\nu^1)$ are equal to 0 or 1; it is generated by two equilibria in all other cases. For each equilibrium selection mechanism, samples of $M = 1,000$, $M = 2,000$ and $M = 20,000$ with $M/2$ markets for each $x_E$ are generated. The joint conditional choice probabilities in $p$ are estimated using a simple frequency count estimator. One of the potentially multiple $\hat{\theta}$’s that minimize $Q_M (\hat{p}, \hat{\theta})$ is estimated using simple gradient methods and the test statistic $MQ_M (\hat{p}, \hat{\theta})$ is computed. The critical value used to determine whether or not one rejects the null hypothesis corresponds to the 95-th empirical percentile obtained from 100 draws of $J_M (\theta)$, i.e. $\alpha = 0.05$. Following Shi and Shum (2015)’s advice, I use $\{\kappa_M\} = \sqrt{\ln[M]}$. The probabilities of rejecting the null hypothesis that $\Theta_1 \neq \emptyset$ are reported in Table 2.1.

The equilibrium selection mechanisms associated with a single equilibrium are indicated by a check mark under the column denoted “Single”. The probabilities of rejecting $H_0 : \Theta_1 \neq \emptyset$ that are reported in these rows therefore correspond the probability of rejecting the necessary condition for a single equilibrium in the data generating process when the data is indeed generated by a single equilibrium. In other words, these rows denote the size of the test based on the necessary condition of equilibrium uniqueness that the identified set should be nonempty. As it can be seen from Table 2.1, the size of the test is smaller than 0.05 for large samples, which is in line with Proposition 3.

The remaining rows of the table, i.e. the ones without a check mark, correspond to data generating processes for which the single equilibrium in the data assumption does not hold. These probabilities therefore correspond to the statistical power of the test for equilibrium uniqueness based on the necessary condition that $\Theta_1 \neq \emptyset$. In several case, one observes that the statistical power of the test tends to 1 as $M$ increases. Unfortunately, this is not the case for all equilibrium selection mechanisms where the data is generated from multiple equilibria. Of course, this is expected from the analysis of the identifying power of the test in Section 2.4.5: data generated from multiple equilibria may still be rationalized by a single equilibrium. This is a consequence of testing the single equilibrium in the data assumption from a testable necessary (as opposed to a necessary and sufficient) condition. It explains why the probability of rejecting $H_0 : \Theta_1 \neq \emptyset$ for some equilibrium selection mechanisms associated with multiple equilibria may be asymptotically smaller than $\alpha$. Nonetheless, the fact that the power of the test tends to 1 for some alternatives implies that this test is still statistically informative.

### 2.6 Concluding remarks

To sum up, the test for equilibrium uniqueness that I present addresses two important issues associated with the procedures previously proposed in the literature. First, I allow for common-knowledge payoff-relevant unobservables, therefore making the information structure of the model relatively more flexible. Second, the test that I propose does not restrict the number of equilibria identifiable from the data prior to testing for equilibrium uniqueness. Moreover, no parametric assumption is needed for the payoff
Table 2.1: Probability of rejecting $H_0 : \Theta_1 \neq \emptyset$

| $\lambda^* (\tau = 1|\nu^0)$ | $\lambda^* (\tau = 1|\nu^1)$ | Markets $(M)$ | Single |
|-------------------------------|-------------------------------|--------------|-------|
| 0                             | 0                             | 0            | ✓     |
| 0                             | 0.2                           | 0            | 0     |
| 0                             | 0.4                           | 0            | 0     |
| 0                             | 0.6                           | 0.21, 0.03   | 0     |
| 0                             | 0.8                           | 0.95, 1      | 1     |
| 0                             | 1                             | 0.04, 0.07   | 0     |
| 0.2                           | 0                             | 0            | 0     |
| 0.2                           | 0.2                           | 0            | 0     |
| 0.2                           | 0.4                           | 0            | 0     |
| 0.2                           | 0.6                           | 0.12, 0.03   | 0     |
| 0.2                           | 0.8                           | 0.50, 0.65   | 1     |
| 0.2                           | 1                             | 1            | 1     |
| 0.4                           | 0                             | 0            | 0     |
| 0.4                           | 0.2                           | 0.07         | 0     |
| 0.4                           | 0.4                           | 0.34, 0.33   | 0.11  |
| 0.4                           | 0.6                           | 0.40, 0.57   | 1     |
| 0.4                           | 0.8                           | 0.97, 1      | 1     |
| 0.4                           | 1                             | 1            | 1     |
| 0.6                           | 0                             | 0.97, 0.98   | 1     |
| 0.6                           | 0.2                           | 0.97         | 1     |
| 0.6                           | 0.4                           | 0.22, 0.36   | 1     |
| 0.6                           | 0.6                           | 0.44, 0.77   | 0.99  |
| 0.6                           | 0.8                           | 0.91, 0.97   | 1     |
| 0.6                           | 1                             | 0.62, 0.78   | 0.97  |
| 0.8                           | 0                             | 1            | 1     |
| 0.8                           | 0.2                           | 0.18, 0.32   | 0.85  |
| 0.8                           | 0.4                           | 0.23, 0.58   | 1     |
| 0.8                           | 0.6                           | 0.05, 0.02   | 0     |
| 0.8                           | 0.8                           | 0            | 0     |
| 0.8                           | 1                             | 0            | 0     |
| 0.8                           | 1                             | 0.02         | 0.03  |
| 1                             | 0                             | 0            | ✓     |
| 1                             | 0.2                           | 1            | 1     |
| 1                             | 0.4                           | 1            | 1     |
| 1                             | 0.6                           | 0.05         | 0     |
| 1                             | 0.8                           | 0            | 0     |
| 1                             | 1                             | 0            | 0     |

Notes: $\lambda^* (\tau = 1|\nu^j)$ is the weight associated with equilibrium 1 at $x_{NE}$ and $\nu^j$. The data is generated as in Example 2 using $p^* (\nu^0, \tau = 1) = [0.0499, 0.9116]'$, $p^* (\nu^0, \tau = 2) = [0.9501, 0.0884]'$, $p^* (\nu^1, \tau = 1) = [0.9003, 0.2417]'$, $p^* (\nu^1, \tau = 2) = [0.0527, 0.9672]'$. Different columns correspond to different numbers of realizations of the game $(M)$ and there are $M/2$ realizations for $x_{NE}$ and $x_{E}$ each. The column “Single” indicates which data generating processes correspond to the single equilibrium assumption. Rejection probabilities are computed over 100 Monte-Carlo samples. The critical values used correspond to the 95-th empirical percentile obtained from 100 draws of $J_M (\theta)$. 
functions nor the distributions of the unobservables. The main identifying assumption is the existence of an observable variable satisfying some exclusion restriction requirements, which can be interpreted as a proxy variable for the payoff-relevant information that is known to all players, but unobservable to the econometrician. The test boils down to a specification test and it can be implemented through a simple two-stage minimum distance procedure that has nice properties according to the provided simulation evidence.

There are at least two natural extensions of the current test that I leave for future research. First, it may be interesting to allow for continuous common-knowledge unobservables. In this vein, it may be possible to use a deconvolution argument along the lines of Khan and Nekipelov (2012). Unfortunately, these results impose much more structure on the payoffs, the unobservables and the equilibrium selection mechanism compared to the test proposed above.

Second, the simulation results presented in Sections 2.4.5 and 2.5.3 clearly suggest that the test has higher power under some alternatives than others. For now, I do not have a clear characterization of the directions for which the test I propose has more power. This is a feature of the test that is worth investigating.

2.7 Appendix

2.7.1 Proof of Proposition 2

Conditions (i) and (ii) come from \( p(y|\nu)'s \) and \( \Gamma(x_E)'s \) being in the unit interval. First, consider the case where \( p(y|x_1^E) - p(y|x_0^E) > 0 \). From equations (2.20), (2.21) and (2.22), one gets:

\[
0 \leq p(y|\nu^0) \leq 1 \Rightarrow L_0(y) \leq -\frac{\phi}{\Upsilon} \leq L_1(y); \tag{2.59}
\]

\[
0 \leq p(y|\nu^1) \leq 1 \Rightarrow L_0(y) \leq \frac{1 - \phi}{\Upsilon} \leq L_1(y); \tag{2.60}
\]

\[
\Gamma(x_E) > 0 \Rightarrow -\frac{\phi}{\Upsilon} < Q(x_E); \tag{2.61}
\]

\[
\Gamma(x_E) < 1 \Rightarrow \frac{1 - \phi}{\Upsilon} > Q(x_E). \tag{2.62}
\]

Notice that \( p(y|x_1^E) - p(y|x_0^E) > 0 \) implies \( L_0(y) \leq Q(x_E) \leq L_1(y) \). Therefore the upper bound on \(-\phi/\Upsilon\) in (2.59) is satisfied whenever (2.61) holds. Similarly, the lower bound on \((1 - \phi)/\Upsilon\) in (2.60) is satisfied whenever (2.62) holds. Inequalities (2.59) to (2.62) can therefore be summarized as:

\[
L_0(y) \leq -\frac{\phi}{\Upsilon} < Q(x_E); \tag{2.63}
\]

\[
Q(x_E) < \frac{1 - \phi}{\Upsilon} \leq L_1(y). \tag{2.64}
\]

A similar argument for the case where \( p(y|x_1^E) - p(y|x_0^E) < 0 \), which in turn implies that \( L_1(y) \leq Q(x_E) \leq L_0(y) \), gives:
\[ L_1(y) \leq -\frac{\phi}{1} < Q(x_E); \]  
\[ Q(x_E) < \frac{1 - \phi}{1} \leq L_0(y). \]  
(2.65)

(2.66)

Since inequalities (2.63) to (2.66) must hold for all \( y \in Y^2 \) and all \( x_E \in X_E \), the restrictions on \( \theta \) coming from \( p(y|\nu) \)'s and \( \Gamma(x_E) \)'s being in the unit interval boil down to:

\[
\max_{y \in Y^2} \{ \min \{ L_0(y), L_1(y) \} \} \leq -\frac{\phi}{1} < \min_{x_E \in X_E} \{ Q(x_E) \};
\]
(2.67)

\[
\max_{x_E \in X_E} \{ Q(x_E) \} < \frac{1 - \phi}{1} \leq \min_{y \in Y^2} \{ \max \{ L_0(y), L_1(y) \} \};
\]
(2.68)

which corresponds to conditions (i) and (ii). Conditions (iii) and (iv) hold if the single equilibrium in the data assumption holds, i.e. \( |\mathcal{T}^*(\nu^0)| = |\mathcal{T}^*(\nu^1)| = 1 \). More precisely, these are restrictions on \( \theta \) that hold if players' decisions are independent. Following Kasahara and Shimotsu (2014), one way to characterize this independence is by evaluating the rank of the following matrix \( \forall \nu \in \mathcal{V} \):

\[
P(\nu) = \begin{bmatrix} p(0,0|\nu) & p(0,1|\nu) \\ p(1,0|\nu) & p(1,1|\nu) \end{bmatrix}.
\]
(2.69)

Intuitively, if \( |\mathcal{T}^*(\nu)| = 1 \), players’ decisions are independent conditional on \( x_{NE} \) and \( \nu \). This independence is a consequence of Assumption 8(i), more precisely the independence of private information shocks across players. Then, \( p(y|\nu) = p(y_1|\nu)p(y_2|\nu) \) and \( \text{rank}(P(\nu)) = 1 \), which can be written as:

\[
p(0,0|\nu)p(1,1|\nu) - p(1,0|\nu)p(0,1|\nu) = 0.
\]
(2.70)

Evaluating (2.70) at each \( \nu \in \mathcal{V} \) using (2.21) and (2.22) generates quadratic equations in \(-\phi/\Upsilon\) and \((1-\phi)/\Upsilon\) that correspond to conditions (iii) and (iv) defining \( \Theta_1 \).

### 2.7.2 Proof of Corollary 1

This statement follows from the quadratic equations associated with conditions (iii) and (iv) in Proposition 2 having at most one solution in \(-\phi/\Upsilon\) and \((1-\phi)/\Upsilon\) after having fixed the sign of \( \Upsilon \). To see this, notice that these quadratic equations each have at most two solutions. Denote these solutions \((-\phi/\Upsilon)^*\) and \((-\phi/\Upsilon)^{**}\) for condition (iii); \((1-\phi)/\Upsilon)^*\) and \((1-\phi)/\Upsilon)^{**}\) for condition (iv). There are therefore four possible combinations of solutions. Notice that, since the quadratic equations in (iii) and (iv) have the same coefficients \(a_0, a_1\) and \(a_2\), the solutions are the same for both equations, i.e. \((-\phi/\Upsilon)^* = ((1-\phi)/\Upsilon)^*\) and \((-\phi/\Upsilon)^{**} = ((1-\phi)/\Upsilon)^{**}\). However, since \(-\phi/\Upsilon \neq (1-\phi)/\Upsilon\), there are two potentially valid solution couples: \((-\phi/\Upsilon)^*, ((1-\phi)/\Upsilon)^{**}\) and \((-\phi/\Upsilon)^{**}, ((1-\phi)/\Upsilon)^*\). Finally, since by definition \( \phi \in (0,1) \), \(-\phi/\Upsilon\) and \((1-\phi)/\Upsilon\) must be of opposite signs. Once one fixes \( \Upsilon > 0 \), which is without loss of generality, there is only one solution couple such that \(-\phi/\Upsilon < 0\) and \((1-\phi)/\Upsilon > 0\). If it exists, this solution couple defines a system of two equations in two unknowns which corresponds to at most one \( \theta \in \Theta_1 \).
2.7.3 Proof of Corollary 2

Let \( \tilde{p}(y|\nu) \) be the choice probability constructed from (2.21) and (2.22) using an arbitrary \( \tilde{\theta} \in \tilde{\Theta}_1 \). Similarly for \( \tilde{p}(y|\nu) \). In order to show the result stated in the Corollary, it is useful to show an intermediary one. I first show that, if \( \tilde{\Theta}_1 \neq \emptyset \) and \( \tilde{\Theta}_1 \neq \emptyset \), then \( \tilde{p}(y|\nu) = \tilde{p}(y|\nu) \) \( \forall \theta \in \tilde{\Theta}_1 \), \( \forall \theta \in \tilde{\Theta}_1 \), \( \forall y \in \mathcal{Y}^2 \) and \( \forall \nu \in \mathcal{V} \).

This intermediary result is straightforward to show if \( \tilde{\Theta} = \tilde{\Theta}^0 = \tilde{\Theta}_1 \) and \( \tilde{\Theta} = \tilde{\Theta}^0 = \tilde{\Theta}_1 \), i.e. when the true data generating processes are included in the identified sets. The result then holds by construction since \( \tilde{p}^0(y|\nu) = \tilde{p}^0(y|\nu) \) and it holds for all elements of the identified sets since \( \tilde{\Theta}_1 \) and \( \tilde{\Theta}_1 \) are at most singletons.

However, the intermediary result requires more work when \( \tilde{\Theta} \in \tilde{\Theta}_1 \) and \( \tilde{\Theta} \in \tilde{\Theta}_1 \), but \( \tilde{\Theta} \neq \tilde{\Theta}^0 \), \( \tilde{\Theta} \neq \tilde{\Theta}^0 \). This case would arise when the data can be rationalized by the single equilibrium assumption despite that there are multiple equilibria in the data generating processes. For both \( \tilde{\Theta}^0 \) and \( \tilde{\Theta}^0 \), notice that one can write the observable choice probabilities \( p(y|x_E) \) as functions of \( \theta^0 = [\phi^0, \Phi^0] \), i.e. the parameters associated with the true data generating process, and the corresponding unobservable-specific choice probabilities \( p^0(y) \equiv p^0(y|\nu) = \tilde{p}^0(y|\nu) \). In fact:

\[
p(y|x_E^0) = p^0(y|\nu^0) + \phi^0 [p^0(y|\nu^1) - p^0(y|\nu^0)]
\]

(2.71)

\[
p(y|x_E^0) = p^0(y|\nu^0) + \left[ \Phi^0 \phi^0 + \phi^0 \Phi^0 \right] [p^0(y|\nu^1) - p^0(y|\nu^0)].
\]

(2.72)

By using (2.71), (2.72) one can rewrite (2.21), (2.22) for an arbitrary \( \theta \) as:

\[
p(y|\nu^0) = p^0(y|\nu^0) + \left[ \phi^0 - \Phi^0 \phi^0 \right] [p^0(y|\nu^1) - p^0(y|\nu^0)]
\]

(2.73)

\[
p(y|\nu^1) = p^0(y|\nu^0) + \left[ \phi^0 + \Phi^0 \phi^0 \right] [p^0(y|\nu^1) - p^0(y|\nu^0)].
\]

(2.74)

Define the following functions:

\[
b_0 \equiv p^0(0,0|\nu^0) p^0(1,1|\nu^0) - p^0(1,0|\nu^0) p^0(0,1|\nu^0);
\]

(2.75)

\[
b_1 \equiv p^0(0,0|\nu^0) [p^0(1,1|\nu^1) - p^0(1,1|\nu^0)] + p^0(1,1|\nu^0) [p^0(0,0|\nu^1) - p^0(0,0|\nu^0)]
\]

\[ - p^0(1,0|\nu^0) [p^0(0,0|\nu^1) - p^0(0,1|\nu^0)] - p^0(0,1|\nu^0) [p^0(1,0|\nu^1) - p^0(1,0|\nu^1)];
\]

(2.76)

\[
b_2 \equiv \left[ p^0(0,0|\nu^1) - p^0(0,0|\nu^0) \right] \left[ p^0(1,1|\nu^1) - p^0(1,1|\nu^0) \right]
\]

\[ - \left[ p^0(0,1|\nu^1) - p^0(0,1|\nu^0) \right] \left[ p^0(1,0|\nu^1) - p^0(1,0|\nu^0) \right].
\]

(2.77)

By evaluating condition (2.70) at \( p(y|\nu^0) \) given in (2.73), one gets a quadratic equation in \( h_0(\theta) \equiv \phi^0 - \Phi^0 \phi^0 / \Phi^0 \) such that:

\[
b_2 h_0(\theta)^2 + b_1 h_0(\theta) + b_0 = 0.
\]

(2.78)
Similarly, evaluating (2.70) at \( p(y|\nu^1) \) given in (2.74), one obtains the following quadratic equation in \( h_1(\theta) \equiv \phi^0 + \Upsilon^0 (1 - \phi) / \Upsilon \):

\[
b_2 h_1(\theta)^2 + b_1 h_1(\theta) + b_0 = 0. \tag{2.79}
\]

Since \( \theta \in \Theta_1 \) and \( \theta \in \hat{\Theta}_1 \), (2.78) and (2.79) are both satisfied by \( \theta \) and \( \hat{\theta} \). Notice that the solutions to these equations will be functions of \( b_0, b_1 \) and \( b_2 \), which are the same for both \( \theta \) and \( \hat{\theta} \). There are at most two solutions for each quadratic equation. After fixing the sign of \( \Upsilon \), using an argument similar to what has been done above, one can show that there is at most one couple, say \( h_0(\theta)^* \) and \( h_1(\theta)^{**} \), that solve (2.78) and (2.79) simultaneously for a given \( \theta \). Therefore, as what is needed to show:

\[
\bar{p}(y|\nu^0) = p^0(y|\nu^0) + h_0(\theta)^* \left[p^0(y|\nu^1) - p^0(y|\nu^0)\right] = \bar{p}(y|\nu^0); \tag{2.80}
\]

\[
\bar{p}(y|\nu^1) = p^0(y|\nu^0) + h_1(\theta)^* \left[p^0(y|\nu^1) - p^0(y|\nu^0)\right] = \bar{p}(y|\nu^1) \tag{2.81}
\]

which holds \( \forall y \in \mathbb{Y}^2 \).

I now turn to the proof of Corollary 2 per se. In particular, we want to show that, under the stated conditions, \( \hat{\Theta}_1 \neq \emptyset \) implies \( \hat{\Theta}_I \neq \emptyset \). The converse is symmetric. Suppose that \( \theta \in \hat{\Theta}_I \) such that \( \hat{\Theta}_1 \neq \emptyset \). To show that \( \hat{\Theta}_1 \neq \emptyset \), one must find a \( \hat{\theta} \) that satisfies the conditions defining \( \hat{\Theta}_1 \). The converse of the intermediary result implies that the only candidate \( \hat{\theta} \) must satisfy \( \bar{p}(y|\nu) = \bar{p}(y|\nu) \forall y \in \mathbb{Y}^2 \) and \( \forall \nu \in \mathbb{V} \).

Such a \( \hat{\theta} \) exists and is unique because: (a) (2.21) defines a one-to-one mapping between \( \bar{p}(y|\nu^0) \) and \( \hat{\phi}/\bar{\Upsilon} \); and (b) (2.22) defines a one-to-one mapping between \( \bar{p}(y|\nu^1) \) and \( \left(1 - \hat{\phi}\right)/\bar{\Upsilon} \). Notice that since \( \theta \in \hat{\Theta}_I \), \( \bar{p}(y|\nu) = \bar{p}(y|\nu) \forall y \in \mathbb{Y}^2 \) and \( \forall \nu \in \mathbb{V} \) implies that \( 0 \leq \bar{p}(y|\nu) \leq 1 \) and \( \text{rank} \left\{ \bar{\mathbf{P}}(\nu) \right\} = 1 \forall y \in \mathbb{Y}^2 \) and \( \forall \nu \in \mathbb{V} \). The only other restrictions that must be satisfied by \( \hat{\theta} \) to belong to \( \hat{\Theta}_I \) are \( 0 \leq \bar{\Gamma}(x_E) \leq 1 \forall x_E \in X_E \). Again, without loss of generality, consider \( \bar{\Upsilon} > 0 \). Given \( x_E \), these conditions can be written as:

\[
\frac{-\ddot{\phi}}{\bar{\Upsilon}} < \frac{\bar{p}(y|x_E) - \bar{p}(y|x_E^0)}{\bar{p}(y|x_E^1) - \bar{p}(y|x_E^0)}; \tag{2.82}
\]

\[
\frac{\bar{p}(y|x_E) - \bar{p}(y|x_E^0)}{\bar{p}(y|x_E^1) - \bar{p}(y|x_E^0)} < \frac{1 - \ddot{\phi}}{\bar{\Upsilon}}. \tag{2.83}
\]

From (2.21), one gets:

\[
\frac{-\ddot{\phi}}{\bar{\Upsilon}} = \frac{\bar{p}(y|\nu^0) - \bar{p}(y|x_E^0)}{\bar{p}(y|x_E^1) - \bar{p}(y|x_E^0)}. \tag{2.84}
\]

Similarly, from (2.22):

\[
\frac{1 - \ddot{\phi}}{\bar{\Upsilon}} = \frac{\bar{p}(y|\nu^1) - \bar{p}(y|x_E^0)}{\bar{p}(y|x_E^1) - \bar{p}(y|x_E^0)}. \tag{2.85}
\]

\[\text{Actually, since } \bar{p}(y|\nu) = \bar{p}(y|\nu), \text{ equation (2.91) below implies that:}
\]

\[
\frac{\bar{p}(y|x_E^1) - \bar{p}(y|x_E^0)}{\bar{\Upsilon}} = \frac{\bar{p}(y|x_E^1) - \bar{p}(y|x_E^0)}{\bar{\Upsilon}}.
\]

Using both \( \bar{\Upsilon} > 0 \) and \( \bar{\Upsilon} > 0 \) ensures that \( \bar{p}(y|x_E^1) - \bar{p}(y|x_E^0) \) and \( \bar{p}(y|x_E^1) - \bar{p}(y|x_E^0) \) have the same signs.
It follows that (2.21) and (2.22) hold if and only if:

\[
\frac{\hat{p}(y|\nu^0) - \hat{p}(y|x_E)}{\bar{p}(y|x_E^0) - \bar{p}(y|x_E^1)} < 0; \tag{2.86}
\]

\[
\frac{\hat{p}(y|\nu^1) - \hat{p}(y|x_E)}{\bar{p}(y|x_E^0) - \bar{p}(y|x_E^1)} > 0. \tag{2.87}
\]

Notice that, by rearranging

\[
\hat{p}(y|x_E) = \hat{p}(y|\nu^0) \left[1 - \hat{\Gamma}(x_E)\right] + \hat{p}(y|\nu^1) \hat{\Gamma}(x_E), \tag{2.88}
\]

one can write:

\[
\hat{p}(y|\nu^0) - \hat{p}(y|x_E) = - \left[\hat{p}(y|\nu^1) - \hat{p}(y|\nu^0)\right] \hat{\Gamma}(x_E); \tag{2.89}
\]

\[
\hat{p}(y|\nu^1) - \hat{p}(y|x_E) = \left[\hat{p}(y|\nu^1) - \hat{p}(y|\nu^0)\right] \left[1 - \hat{\Gamma}(x_E)\right]. \tag{2.90}
\]

Moreover, by evaluating (2.88) at \(x_E^0, x_E^1\) and rearranging:

\[
\hat{p}(y|\nu^1) - \hat{p}(y|\nu^0) = \frac{\hat{p}(y|x_E^1) - \hat{p}(y|x_E^0)}{\hat{\Gamma}}. \tag{2.91}
\]

Therefore, \(\hat{p}(y|\nu^1) - \hat{p}(y|\nu^0)\) and \(\hat{p}(y|x_E^1) - \hat{p}(y|x_E^0)\) have the same sign. Suppose that \(\hat{p}(y|x_E^1) - \hat{p}(y|x_E^0) > 0\). From (2.91), \(\hat{p}(y|\nu^1) - \hat{p}(y|\nu^0) > 0\), (2.89) implies that \(\hat{p}(y|\nu^0) - \hat{p}(y|x_E) < 0\) and (2.90) implies that \(\hat{p}(y|\nu^1) - \hat{p}(y|x_E) > 0\). As a result, inequalities (2.86) and (2.87) hold. A similar argument can be used to show that these inequalities also hold when \(\hat{p}(y|x_E^1) - \hat{p}(y|x_E^0) < 0\). It follows that \(\hat{\theta} \in \hat{\Theta}_I\) and \(\hat{\Theta}_I \neq \emptyset\), which completes the proof.
Chapter 3

Two-step semiparametric estimation of dynamic discrete games: A comparison of bias reduction techniques

3.1 Introduction

The focus of this chapter is the estimation of a fairly general class of economic models in which the parametric distribution of the agents’ decisions is implicitly defined as a fixed point. Empirical games are a special case of this class of models. There are typically two objects to be estimated: (1) the conditional choice probabilities (CCPs), which can be interpreted as agents’ beliefs about their own and the other players’ actions; and (2) the structural parameters or the primitives of the model. One could estimate all these objects simultaneously by using a full-solution method like Rust (1987)’s nested fixed point algorithm. However, that approach is computationally very intensive and can be impractical for large state spaces. While it is tempting to reduce the size of the state space to make estimation tractable, it may also affect the credibility of the results.

To alleviate the computational burden in large state spaces, one may separate the estimation of the beliefs, treated as nuisance parameters, from the estimation of the structural parameters, as in Hotz and Miller (1993)’s CCP estimator. This approach consists in estimating the beliefs in a first step, typically using nonparametric techniques, and then estimating the structural parameters given the first-step estimates. Such a practice implies substantial computational savings, but those benefits also come at a cost: Monte-Carlo evidence suggests that the two-step semiparametric estimator may be heavily biased in small samples. The usual poor performance of nonparametric estimators in small samples and the first-step estimates entering nonlinearly in the objective function of the second-step estimator are well-accepted explanations for this bias.

One method that has been used to deal with the finite sample bias of the structural parameters’

\footnote{See Manski (1991) and Ahn and Manski (1993) for a similar approach applied to another class of models.}

\footnote{See, for example, Aguirregabiria and Mira (2002) and Kasahara and Shimotsu (2008) for single-agent dynamic discrete choice models; Aguirregabiria and Mira (2007), Pakes, Ostrovsky, and Berry (2007), Pesendorfer and Schmidt-Dengler (2008) and Egesdal, Lai, and Su (2013) for dynamic games. Furthermore, notice that the same problem has also been pointed out by Hotz, Miller, Sanders, and Smith (1994) for a simulated version of Hotz and Miller (1993)’s estimator.}
Chapter 3. Bias reduction in dynamic discrete games

estimates is the iterative $K$-step estimator introduced by Aguirregabiria and Mira (2002, 2007). The Monte-Carlo evidence they provide indicates that the $K$-step estimator can be very effective at reducing finite sample bias. However, subsequent papers (Pesendorfer and Schmidt-Dengler, 2008, 2010; Egesdal, Lai, and Su, 2013) have pointed out that in dynamic games, the generated sequence of estimates may often fail to converge and the performance of this estimator relies on the stability of the equilibrium generating the data.\footnote{Stability holds if, after having slightly disturbed the equilibrium CCPs, iterating over the fixed point mapping generates a sequence of CCPs that converge to the same equilibrium (e.g., Aguirregabiria and Nevo, 2013).} An even more concerning issue pointed out by these criticisms is that, when the $K$-step estimator converges, it may converge to a point far from the true value of the parameters, therefore making the $K$-step estimator less effective at reducing the finite sample bias than previously suggested. Notice that this is true in finite samples only. The $K$-step estimator is consistent regardless of the stability of the equilibrium in the data generating process.

Compared with recent criticisms of the $K$-step estimator, a very important difference in the approach that I follow is that, instead of iterating the $K$-step method until it converges, I consider a fixed number of iterations, whether or not convergence is reached. In other words, instead of focusing only on the properties of the estimator upon convergence, I study the finite sample properties of the $K$-step estimator without getting rid of the sequences of estimates that fail to converge.

Using a dynamic game of market entry and exit, under both stable and unstable equilibria, I compare the finite sample properties of the $K$-step approach (for a fixed $K$) with alternative estimators that do not require the stability condition: two modified $K$-step estimators; bootstrap and jackknife bias-reduced estimators; and a single Newton full maximum likelihood iteration initialized at the two-step estimates. Even the best of all the two-step estimators considered here, which vary according to the first and the second stage estimator used, has disappointing finite sample properties. As expected, the results suggest that the $K$-step estimator does not reduce finite sample bias as much when the equilibrium generating the data is unstable compared to the stable case. However, the $K$-step estimator works better than any of the alternative methods considered here, even when the equilibrium is not stable. None of the alternatives considered in this chapter would be an obvious candidate estimator robust to equilibrium instability. In fact, when compared to the $K$-step estimator, none of these alternatives achieves lower root-mean-square errors (RMSE) for all parameters in both stable and unstable equilibria. Even when the $K$-step estimator does not converge to a single point, it may still lead to an appreciable bias correction for a fixed number of iterations.

More precisely, when the equilibrium generating the data is stable, the two-step estimator is associated with RMSE that are on average (over all parameters) more than twice larger than an unfeasible two-step approach in which the true beliefs would be known. For the unstable case, the corresponding RMSE are almost three times larger, suggesting that the poor small sample properties of the two-step approach are even more problematic when stability fails. The $K$-step estimator RMSE is 16% and 56% greater than the unfeasible two-step estimator for the stable and the unstable equilibria respectively. Aguirregabiria and Mira (2011)’s modified $K$-step does not perform particularly well (35% and 125% greater than the unfeasible two-step RMSE on average), except that it generally leads to a lower RMSE for player-specific parameters compared to the original $K$-step. Despite attractive results in the unstable case (on average, RMSE 49% larger than the unfeasible two-step estimator), the feasible version of Kasahara and Shimotsu (2012)’s proposed modification is less convincing when stability holds (85% larger). Finally, the bootstrap, the jackknife and the single Newton full maximum likelihood iteration
all perform worse than the usual $K$-step approach, both in the stable and in the unstable equilibria.

The rest of the chapter proceeds as follows. Next section briefly reviews the related literature and emphasizes the contributions of the chapter. In Section 3.3, I describe a dynamic game of market entry and exit following Aguirregabiria and Mira (2007) and I introduce the two-step semiparametric estimator of this model. The different bias reduction techniques and alternative estimators are presented in Section 3.4. Monte-Carlo experiments are reported in Section 3.5 and the last section concludes.

## 3.2 Literature review

This chapter contributes to the still emerging literature of bias reduction techniques for two-step estimators applied to dynamic discrete games by shedding light on the convergence and stability issues of the $K$-step estimator.

On one hand, for single-agent dynamic discrete choice models, Kasahara and Shimotsu (2008) provide theoretical arguments justifying the good finite sample properties of the $K$-step estimator suggested by Monte-Carlo evidence from Aguirregabiria and Mira (2002). These theoretical results imply that one can expect a monotone improvement of the finite sample bias as the number of iterations increases. On the other hand, such theoretical results do not directly extend to dynamic discrete games and, therefore, despite encouraging simulation results, Aguirregabiria and Mira (2007) recognize that convergence of the $K$-step estimator could be a concern when strategic interactions are part of the model.

Among the early evidence of the potential problems associated with the $K$-step estimator applied to dynamic discrete games, it is worth mentioning the paper by Pesendorfer and Schmidt-Dengler (2008) in which the authors propose a different technique of estimation applicable to dynamic discrete games and compare it to Aguirregabiria and Mira (2007)’s estimator (among others). Their results depict a less conclusive picture of the $K$-step’s properties.

Pesendorfer and Schmidt-Dengler (2010) provide a possible explanation for the poor performance of the $K$-step method in some dynamic games. Simulation evidence shows that if the equilibrium generating the data is unstable, the sequence of estimates generated by the $K$-step approach rarely converges and, when it does, the resulting estimates exhibit poor finite sample properties. Some details about stability requirements are available in Aguirregabiria and Nevo (2013). In order to deal with this equilibrium stability issue, Aguirregabiria and Mira (2011) and Kasahara and Shimotsu (2012) propose modified $K$-step estimators that do not rely on the stability of the equilibrium generating the data.

However, equilibrium stability does not seem to be the only cause of the rare convergence of the $K$-step method in dynamic games. In fact, some of the simulations in Egesdal, Lai, and Su (2013) suggest that even when the equilibrium generating the data is stable, the iterative process may still often fail to converge. In other words, there seems to be some properties of the $K$-step estimator that are not yet fully understood in the literature.

While most papers on the subject have been focusing on the properties of the estimates upon convergence, I ignore convergence and I iterate the updating process a fixed number of times. This difference is key to the nuances in my results. Another contribution can also be seen in the fact that the Monte-Carlo results that I present are the first attempt to assess the finite sample properties for most alternative estimators included here. To my knowledge, the finite sample properties of Aguirregabiria and Mira (2011)’s modified $K$-step estimator and the single Newton full maximum likelihood iteration (as pro-
posed by Aguirregabiria and Mira, 2007) have not been studied previously. Moreover, while Kasahara and Shimotsu (2012) provide Monte-Carlo evidence of the good properties of their modified $K$-step estimator, they test its unfeasible version by using the true value of the parameters and the true beliefs when determining the value of a key parameter used to guarantee convergence of the updating process. One of the methods that they propose in practice is to set this parameter close to zero. This is the feasible version that I consider.

Notice that the bootstrap did get some attention in the context of two-step semiparametric estimation for the class of models studied here and in some related settings, but not for its finite sample bias reduction properties. For instance, Kasahara and Shimotsu (2008) use the bootstrap to improve the size of hypothesis tests and the coverage probability of confidence intervals in single-agent models. To my knowledge, this current chapter is the first attempt at using resampling methods to improve the finite sample properties of the two-step semiparametric estimator in games.

### 3.3 Model and two-step estimation

#### 3.3.1 General framework

Let $\mathcal{Y}$ be a discrete and finite set and let $P^0(y|x)$ be the true probability distribution of the realization of a $N \times 1$ vector $y \in \mathcal{Y}^N$, for $N \geq 1$, conditional on some vector of realized exogenous variables $x \in \mathcal{X}$. Also, denote $P^0 = \{P^0(y|x) : y \in \mathcal{Y}^N, x \in \mathcal{X}\}$. One wishes to rationalize this distribution through a structural model leading to a parametrized vector of conditional probabilities. Consider a $q$-dimensional vector ($q < \infty$) of structural parameters $\theta \in \Theta$, for a compact $\Theta$ and denote the vector of parametrized probabilities as $P(\theta) = \{P(y|\theta, x) : y \in \mathcal{Y}^N, x \in \mathcal{X}\}$.

There is no closed-form analytical expression for $P(y|\theta, x)$. The parametric distribution of $y$ is implicitly defined as a fixed point of the following mapping in the probability space:

$$P(\theta) = \Psi(\theta, P(\theta)) \quad (3.1)$$

where $\Psi(\cdot) : \mathcal{X} \times \Theta \times \mathcal{F} \rightarrow \mathcal{F}$ for $\mathcal{F}$ being the space of the probability distribution of $y$ and $\Psi(\theta, P) = \{\Psi(y|x, \theta, P) : y \in \mathcal{Y}^N, x \in \mathcal{X}\}$.

Typically, the researcher has a sample of $M$ observations indexed by $m \in \{1, 2, \ldots, M\}$. For each observation, the data set includes the realizations $y_m \in \mathcal{Y}^N$ and $x_m \in \mathcal{X}$.

The index of the data could vary over more than one dimensions, allowing for individual, time and geographical variations for instance. The key element here is that all the asymptotic results depend on one of these dimensions being large. For what follows, this dimension is the one indexed by $m$.

#### 3.3.2 Dynamic discrete game of market entry and exit

In this section, I introduce the dynamic game of market entry and exit that will be used for Monte-Carlo experiments in Section 3.5. This model comes from Aguirregabiria and Mira (2007) and it is a special case of the general framework presented above.

---

4 Nonetheless, Aguirregabiria and Mira (2011)’s modified $K$-step estimator is one of the estimators used by Sweeting (2013) in a study of the commercial radio industry.

5 Egesdal, Lai, and Su (2013) also reports Monte-Carlo results for a feasible version of this modified $K$-step estimator by fixing this key parameter to 0.5. I use a value much closer to 0, i.e. 0.05.
Each firm (indexed by \( i \in \{1,2,\ldots, N \} \)) decides whether or not to operate in a market \( (m \in \{1,2,\ldots, M \}) \) at a given period \( (t \in \{1,2,\ldots, T \}) \). In this context, \( y = \{0,1\} \) is each firm’s choice set, such that \( y_{imt} = 1 \) if firm \( i \) operates in market \( m \) at time \( t \).\(^6\) Suppose that a firm’s decision to operate depends on the market size \( (S_{mt}) \), the number of competitors that are operating \( \left( \sum_{j \neq i} y_{jmt} \right) \), an entry cost, a firm-specific fixed cost and some \( \text{dim (} y \text{)} \)-dimensional vector of unobservables \( (\varepsilon_{imt}) \), which can be interpreted as some private information distributed according to the cumulative density function \( G(\cdot) \). The dynamic dimension to the firm’s decision comes from the firm paying the entry cost only if it has not been operating in the same market during the previous period.

Let \( y_{mt} = \{y_{imt} : i = 1, \ldots, N \} \) and let \( \varepsilon(y) \) denote vector \( \varepsilon \)'s element associated with choice \( y \). An observation is a firm-market-period-tuple. Besides the operating decision, for each observation, the econometrician knows the observable state variables, i.e. the market size \( (S_{mt}) \) and the incumbency status of the firms \( (y_{mt-1}) \). To be consistent with the notation introduced above, denote the observable state variables in market \( m \) at period \( t \) as \( x_{mt} = [S_{mt}, y'_{mt-1}]' \). The following assumptions are made by Aguirregabiria and Mira (2007) and are fairly standard in the literature.

**Assumption 12** (Additive separability). The contemporaneous utility function is additively separable in its observable and its unobservable components, such that \( U_i (y_{mt}, x_{mt}, \varepsilon_{imt} (y_{imt})) = u_i (y_{mt}, x_{mt}) + \varepsilon_{imt} (y_{imt}) \).

**Assumption 13** (Conditional independence). The joint transition probability of the observable state variables and the unobservables conditional on choices is written as \( \Pr (x_{mt+1}, \varepsilon_{imt+1} | x_{mt}, \varepsilon_{imt}, y_{mt}) = \Gamma (x_{mt+1} | x_{mt}, y_{mt}) G (\varepsilon_{imt+1}) \), where \( \varepsilon \) is iid across players, markets and time periods.

**Assumption 14** (Discrete support for observable state variables). The observable state variables \( x_{mt} \) have a discrete and finite support.

Under these assumptions, the contemporaneous utility of firm \( i \) in market \( m \) at period \( t \), i.e. \( U_i (y_{mt}, x_{mt}, \varepsilon_{imt} (y_{imt})) \) is:

\[
\begin{aligned}
\left\{ \begin{array}{l}
\theta_{RS} S_{mt} - \theta_{RN} \left[ 1 + \sum_{j \neq i} y_{jmt} \right] - \theta_{EC} (1 - y_{imt-1}) - \theta_{FC,i} + \varepsilon_{imt} (1) ,
\varepsilon_{imt} (0) \\
\end{array} \right. ,
\end{aligned}
\]

\( y_{imt} = 1 \)

\( y_{imt} = 0 \)

\( \text{(3.2)} \)

where \( \theta_{RS} \) measures the effect of the market size on the firm’s payoff; \( \theta_{RN} \) is referred to as the strategic interaction parameter; \( \theta_{EC} \) corresponds to the entry cost; and \( \theta_{FC,i} \) is the firm-specific fixed cost.

Denoting the discount factor as \( \beta \), the firm’s optimization problem can be written recursively using the following stochastic value function where \( -i \) denotes all players except player \( i \):

\[
V_i (x_{mt}, \varepsilon_{imt}) = \max_{y \in \{0,1\}} \left\{ E_{\varepsilon_{-imt}} \left[ U_i (y, x_{-imt}, x_{mt}, \varepsilon_{imt} (y)) + \beta \int \sum_{x_{mt+1} \in \mathcal{X}} V_i (x_{mt+1}, \varepsilon_{imt+1}) \right. \\
\times d\Gamma (x_{mt+1}, [x_{mt}, y, x_{-imt}] G (\varepsilon_{imt+1}) \left. \right] \right\}
\]

\( \equiv \max_{y \in \{0,1\}} \left\{ E_{\varepsilon_{-imt}} \left[ v_i (y, x_{-imt}, x_{mt}) + \varepsilon_{imt} (y) \right] \right. \)

\( \text{(3.3)} \)

\( \text{(3.4)} \)

\(^6\)In this chapter, I focus on binary choice models, but multinomial models could easily be accommodated.
From player \(i\)'s point of view, given \(x_{mt}, v_i(y, y_{-imt}, x_{mt})\) is the sum of its current and discounted future expected utility of choosing \(y\) when player \(i\)'s competitors choose \(y_{-imt}\) (minus \(\varepsilon_{imt}(y)\)). Its expectation with respect to \(\varepsilon_{-imt}\) can be written as:

\[
v_P^i(y, x_{mt}) \equiv E_{\varepsilon_{-imt}}[v_i(y, y_{-imt}, x_{mt})] = \sum_{y_{-i} \in Y_{N-1}} \prod_{j \neq i} P(y_j | \theta, x_{mt}) v_i(y, y_{-i}, x_{mt})
\]

where the superscript \(P\) in \(v_P^i(y, x_{mt})\) is used to make explicit the dependence on the distribution of \(y\).

Provided that each firm enters if and only if the expected payoff of entering is greater than the expected payoff of not entering, the conditional choice probability (CCP) of firm \(i\) entering market \(m\) at time \(t\), denoted \(P_i(1|\theta, x_{mt}) \equiv P(y_i = 1|\theta, x_{mt})\), is:

\[
P_i(1|\theta, x_{mt}) = \int 1 \{v^P_i(1, x_{mt}) - v^P_i(0, x_{mt}) \geq \varepsilon_{imt}(0) - \varepsilon_{imt}(1)\} dG(\varepsilon_{imt})
\]

which can be used to construct the fixed point mapping in (3.1).

### 3.3.3 Two-step semiparametric estimator

Suppose that a researcher wants to estimate the true CCPs (\(P_0 \in \mathcal{F}\)) and the true structural parameters (\(\theta^0 \in \Theta\)) from a parametric discrete choice model in which the probability distribution of \(y\) is implicitly defined by the fixed point mapping in (3.1). One natural option could be the maximum likelihood (ML) estimator defined as:

\[
(\hat{P}^\text{ML}, \hat{\theta}^\text{ML}) = \arg \max_{P \in \mathcal{F}, \theta \in \Theta} \sum_{m=1}^{M} \ln \left[ \Psi(y_m | x_m, \theta, P) \right]
\]

i.e., the maximum likelihood estimates are the parameters and the CCPs that maximize the likelihood function and satisfy the fixed point mapping. Such estimates share all the nice properties associated with maximum likelihood estimation. Unfortunately, especially for large state spaces, the high number of objects to be estimated makes this full solution approach less attractive. Egesdal, Lai, and Su (2013) uses the Mathematical Program with Equilibrium Constraints (MPEC), a constrained optimization algorithm proposed by Su and Judd (2012) to estimate dynamic games by maximum likelihood. While their results indicates that the MPEC has nice finite sample properties, it remains computationally more intensive than the \(K\)-step approach.

Alternatively, one could consider a two-step semiparametric estimator that separates the estimation of \(P^0\) from the estimation of \(\theta^0\). In a first step, nonparametric estimates of \(P^0(x) = \{P^0(y|x) : y \in Y^N\}\) (treated as nuisance parameters) can be obtained from the data. These estimates are then used in a (parametric) second-step estimation in order to estimate \(\theta^0\). The next Assumption is typically maintained when using such a two-step procedure to estimate games that admit multiple equilibria. It is necessary for the first step estimator to be a consistent estimator of \(P^0\).

**Assumption 15** (Single equilibrium in the data generating process). If \(P(\theta) = \Psi(\theta, P(\theta))\) admits multiple fixed points, all observations associated with the same realization of \(x\) are drawn from the same
\( P(y|\theta, x) \) with probability one.

For a binary choice model, different possible estimators are available in the first step: frequency count, Nadaraya-Watson, local linear regression, splines, etc. For sake of generality, I write \( \hat{P}(y|x) \) as a weighted sum of the \( y_m \)'s with different weighting functions corresponding to different estimators:

\[
\hat{P}(y|x) = \sum_{m=1}^{M} y_m w(x_m, x).
\] (3.8)

For instance, \( w(x_m, x) = \frac{\mathbb{1}\{x_m=x\}}{\sum_{m=1}^{M} \mathbb{1}\{x_m=x\}} \) and \( w(x_m, x) = \frac{\mathbb{1}\{x_m=x\}}{\sum_{m=1}^{M} \mathbb{1}\{x_m=x\}} \) respectively correspond to the frequency count and to the Nadaraya-Watson (with vector of bandwidths \( h \)) estimators. The pseudo-maximum likelihood (PML) estimator in the second step is then given by:

\[
\hat{\theta}_{\text{PML}} = \arg \max_{\theta \in \Theta} M^{-1} \sum_{m=1}^{M} \ln \left[ \Psi(y_m|x_m, \theta, \hat{P}) \right].
\] (3.9)

Some general results such as consistency, asymptotic normality and conditions for efficiency of the two-step semiparametric estimator can be found in Manski (1991), Ahn and Manski (1993) and Newey and McFadden (1994) among others. In this vein, it is worth reporting similar results that have already been stated for the PML estimator applied to the class of models studied here. A general characterization of such asymptotic properties can be obtained as a special case of Aguirregabiria (2004, Proposition 1, p. 338).

**Lemma 4 (Asymptotic properties of the PML estimator)**. Let Assumptions 12-15 and the following conditions hold:

(a) \( \Psi(\cdot) \) is twice continuously differentiable in \( \mathbf{P} \) and in \( \theta \);

(b) \( \Psi(y|x, \theta, \mathbf{P}) > 0 \) for any \( y \in Y^N, x \in X, \theta \in \Theta, \mathbf{P} \in \mathcal{F} \); and

(c) \( \exists (\mathbf{P}^0, \theta^0) \in \mathcal{F} \times \Theta \) such that \( \mathbf{P}^0 = \Psi(\theta^0, \mathbf{P}^0) \) and \( \theta^0 = \arg \max_{\theta \in \Theta} E \left[ \ln \left[ \Psi(y|x, \theta, \mathbf{P}^0) \right] \right] \) is unique.

Then,

(i) \( \hat{\theta} \xrightarrow{P} \theta^0 \); and

(ii) \( M^{1/2} \left( \hat{\theta} - \theta^0 \right) \xrightarrow{d} N(0, \mathbf{G} \Sigma [\mathbf{I} - \Psi_\theta] \mathbf{G}') \), where \( \Sigma \) is the asymptotic variance of \( M^{1/2} (\hat{\mathbf{P}} - \mathbf{P}^0) \); \( \mathbf{G} \equiv \left( \Psi_\theta \mathbf{diag} \{ \mathbf{P}^0 \}^{-1} \Psi_{\theta} \right)^{-1} \); \( \Psi_\theta \) and \( \Psi_{\mathbf{P}} \) are the Jacobian matrices evaluated at the true values \( \mathbf{P}^0 \) and \( \theta^0 \).

**Proof.** This Lemma can be seen as a special case of Aguirregabiria (2004, Proposition 1). For completeness, the derivation is included in Appendix 3.7.1.

### 3.4 Bias reduction techniques and alternative estimators

The nice properties of the PML estimator stated at the end of the previous section are asymptotic. A drawback associated with this two-step approach is that it is heavily biased in small samples, which justifies the interest of studying bias reduction techniques.
The reasons explaining the important finite sample bias of the PML estimator are well known. The most obvious explanation is that nonparametric estimators generally perform poorly in small samples and both the bias and the variance of the first-step estimates affect the bias in the second step. Another problem related to nonparametric estimation is the curse of dimensionality: the greater the dimension of the vector of conditioning variables in the nonparametric estimation, the slower the rate of convergence. In fact, the curse of dimensionality is usually more problematic when estimating models with strategic interactions. In games, the utility of a given player depends on the other players’ decisions, which may be contingent on different observable variables. The beliefs are, therefore, typically estimated by conditioning on a larger number of exogenous variables than in single-agent models. In practice, it is common to be left with very few or even no observations for some states.

Another obvious explanation is that, because the expectation of a nonlinear function is not equal to this nonlinear function evaluated at its expected argument, the fact that the estimated beliefs enter nonlinearly in the second-step objective function also contributes to the finite sample bias problem.

Finally, errors in the estimation of the beliefs from the first step can be seen as measurement errors that generate a finite sample correlation between regressors and error terms in the moment conditions defining the second step estimator. This correlation introduces a finite sample bias in the estimation (Pakes, Ostrovsky, and Berry, 2007).

Of course, different choices of first-step and second-step estimators should generate different finite sample properties. Pakes, Ostrovsky, and Berry (2007)’s Monte-Carlo simulations suggest that using a smooth estimator in the first step and a general method of moments (GMM) estimator in the second step can reduce the finite sample bias. Even if I have omitted the GMM from the discussion so far, I will get back to this alternative second step when I compare the performance of different two-step estimators in the Monte-Carlo experiments below.

### 3.4.1 $K$-step estimator

A key feature of the PML estimator is that $\Psi(\cdot)$ maps $X \times \Theta \times \mathcal{F}$ into $\mathcal{F}$, i.e. the mapping generates a new set of CCPs given a vector of independent variables, a vector of parameters and a set of CCPs. The two-step semiparametric estimator uses a nonparametric estimate of the CCPs to construct the pseudo-likelihood function. One can iterate this procedure and use, together with the estimated parameters, the CCPs generated by $\Psi(\cdot)$ as a new set of CCPs to construct a new pseudo-likelihood function, and so on. When iterated until convergence, this approach is referred to as the nested pseudo-likelihood (NPL) estimator.

**Definition 3 ($K$-step estimator).** The $k$-th iteration of the NPL estimator is defined as:

$$\hat{\theta}_k = \arg \max_{\theta \in \Theta} M^{-1} \sum_{m=1}^{M} \ln \left[ \Psi \left( y_m | x_m, \theta, \hat{P}_{k-1} \right) \right]$$

and the $k$-th iteration CCPs are given by:

$$\hat{P}_k = \Psi \left( \hat{P}_{k-1}, \hat{\theta}_k \right)$$

where $\hat{P}_0$ is a nonparametric estimate of the CCPs. The $K$-step estimator iterates until $k = K$.

This $K$-step estimator has been introduced by Aguirregabiria and Mira (2002, 2007). A general
description of its asymptotic properties when applied to the class of models studied here is given in
Aguirregabiria (2004, Proposition 1, p. 338). Again, for completeness, they are stated in Lemma 5.

**Lemma 5** (Asymptotic properties of the $K$-step estimator). Under Assumptions 12-15 and conditions
(a)-(c) stated in Lemma 4, \{\hat{\theta}_K, \hat{P}_K : K \geq 1\} are such that:

(i) $\hat{P}_K \overset{P}{\to} P^0$;

(ii) $\hat{\theta}_K \overset{P}{\to} \theta^0$;

(iii) $M^{1/2} \left( \hat{P}_K - P^0 \right) \overset{d}{\to} N(0, A_K \Sigma A_K')$, where $A_K \equiv [I - \Psi_\theta G] \Psi_P A_{K-1} + \Psi_\theta G; A_0 = I; \Psi_\theta, \Psi_P, \Sigma$ and $G$ have already been defined above; and

(iv) $M^{1/2} \left( \hat{\theta}_K - \theta^0 \right) \overset{d}{\to} N(0, B_K \Sigma B_K')$, where $B_K \equiv G [I - \Psi_P A_{K-1}]$.

**Proof.** See Appendix 3.7.2. \hfill \Box

### 3.4.2 Modified $K$-step estimators

In finite samples, as argued by Pesendorfer and Schmidt-Dengler (2010), the $K$-step estimator may fail
to converge to a single point if the equilibrium generating the data is unstable, even if $K$ is large and
\( \hat{P}_0 \) is a consistent estimator of $P^0$. Let $\varphi(P) \equiv \Psi(P, \hat{\theta}(P))$ where:

$$
\hat{\theta}(P) = \arg \max_{\theta \in \Theta} M^{-1} \sum_{m=1}^{M} \ln \left[ \Psi(y_m|x_m, \theta, P) \right]
$$

(3.10)

is the NPL mapping. A necessary condition for the convergence of the $K$-step estimator initiated at
$P \neq P^0$ is the Lyapunov stability of $\varphi(\cdot)$ around $P^0$. A sufficient condition for the equilibrium to be
Lyapunov stable is the spectral radius (i.e., the maximum absolute eigenvalue) of the NPL mapping’s
Jacobian matrix evaluated at the equilibrium CCPs and the true parameters being smaller than one.
When stability holds, successively iterating $\varphi(\cdot)$ starting at $P$ belonging to a neighbourhood of the
equilibrium $P^0$ should converge to $P^0$. Therefore, in finite samples\footnote{Notice that this observation does not prevent the $K$-step estimator from being consistent. As pointed out in the proof of Lemma 5, the $K$-step estimator is consistent if the first-stage estimator is consistent. In other words, even if $P^0$ is associated with an unstable equilibrium, $\hat{\theta}_K$ is a consistent estimator as long as $\hat{P}$ consistently estimates $P^0$.}, initiating the iterative process with
a consistent estimate of $P^0$ will converge to $P^0$, only if the equilibrium generating the data is stable.
More details about the stability requirements for convergence are stated in Kasahara and Shimotsu
(2012) and Aguirregabiria and Nevo (2013). The idea behind the modified $K$-step estimators currently
available in the literature is to achieve convergence even if the equilibrium is unstable.

At this point, it is worth pointing that this convergence issue is only problematic if one hopes to
iterate the fixed point mapping until convergence. In other words, while the NPL estimator is simply
not well defined if convergence fails – therefore making it impossible to properly study its finite sample
properties in this case – one can still iterate the fixed point mapping for a finite number of times and
evaluate the resulting estimates. The finite sample properties of the $K$-step estimator with a fixed $K$ then
depend on the type of non-convergence observed on average.
Aguirregabiria and Mira (2011)’s modified $K$-step

For dynamic games, Aguirregabiria and Mira (2011) propose to update only player $i$’s CCPs when computing player $i$’s contribution to the likelihood function, leaving other players’ CCPs at their first-step estimates. The reason behind this modification is that, by keeping other players’ CCPs fixed, each player’s contribution to the likelihood function is updated as it would be in a single-agent model, for which stability is not an issue.

If $\hat{P}_{k,i}$ denotes player $i$’s CCPs at the $k$-th iteration and $\hat{P}_{k,-i}$ refers to the CCPs of all players except player $i$, Aguirregabiria and Mira (2011)’s modified $K$-step estimator is defined as follows:

**Definition 4** (Aguirregabiria and Mira (2011)’s modified $K$-step estimator). The $k$-th iteration of Aguirregabiria and Mira (2011)’s modified $K$-step estimator is defined as:

$$\hat{\theta}^{AM}_k = \arg \max_{\theta \in \Theta} (NM)^{-1} \sum_{i=1}^{N} \sum_{m=1}^{M} \ln \left[ \Psi \left( y_{im} | x_{im}, \theta, \hat{P}^{AM}_{k-1,i}, \hat{P}_{0,-i} \right) \right]$$

and the $k$-th iteration CCPs are given by:

$$\hat{P}^{AM}_{k,i} = \Psi \left( \hat{P}^{AM}_{k-1,i}, \hat{P}_{0,-i}, \hat{\theta}^{AM}_k \right), \forall i \in \{1, 2, \ldots, N\}$$

where $\hat{P}_0$ is a nonparametric estimate of the CCPs. This modified $K$-step estimator is obtained by iterating until $k = K$.

Since the finite sample properties of this alternative estimator have not been studied yet, it deserves to be included in the Monte-Carlo experiments reported in Section 3.5.

Kasahara and Shimotsu (2012)’s modified $K$-step

Among the modified $K$-step estimators that they propose, Kasahara and Shimotsu (2012) suggest to alter the way the CCPs are updated, while leaving the computation of the structural parameters the same as in the usual $K$-step. Their estimator is described more formally in the next definition. Some explanations follow.

**Definition 5** (Kasahara and Shimotsu (2012)’s modified $K$-step estimator). The $k$-th iteration of Kasahara and Shimotsu (2012)’s modified $K$-step estimator of the structural parameters is defined as:

$$\hat{\theta}^{KS}_k = \arg \max_{\theta \in \Theta} M^{-1} \sum_{m=1}^{M} \ln \left[ \Psi \left( y_m | x_m, \theta, \hat{P}^{KS}_{k-1} \right) \right]$$

and the $k$-th iteration CCPs are given by:

$$\hat{P}^{KS}_{k,i} = \left[ \Psi \left( \hat{P}^{KS}_{k-1,i}, \hat{\theta}^{KS}_k \right) \right]^\alpha \left[ \hat{P}^{KS}_{k-1} \right]^{1-\alpha}$$

where $\hat{P}_0$ is a nonparametric estimate of the CCPs and $\ast$ refers to the Hadamard product. This modified $K$-step estimator is obtained by iterating until $k = K$.

This approach borrows from the relaxation method in numerical analysis. Its properties depend on a key parameter: $\alpha$. Notice that the usual $K$-step estimator is a special case of Definition 5 with $\alpha = 1$. For any $\alpha$, this modified $K$-step approach is consistent as long as $\hat{P}_0$ is consistent. However, one may
Chapter 3. Bias reduction in dynamic discrete games

want to choose a value of \( \alpha \) in order to guarantee bias reduction and to minimize the number of iterations needed to obtain appreciable finite sample improvements. Any \( \alpha \) such that the spectral radius of the Jacobian of \( [\Psi(P, \theta)]^\alpha \ast [P]^{1-\alpha} \) with respect to \( P \) evaluated at \( (P^0, \theta^0) \) is less than 1 will reduce bias. In order to make this reduction as fast as possible, one should aim for a small spectral radius. Therefore, the optimal \( \alpha \), say \( \alpha^* \), would be the one that minimizes the spectral radius of this Jacobian matrix.

At least two reasons make the computation of \( \alpha^* \) challenging. First, minimizing the spectral radius of a matrix with respect to a parameter cannot be done by using typical gradient methods. A costly alternative could be to perform a grid search, keeping in mind that \( \alpha^* \) can take both positive and negative values. Second, computing the Jacobian of the mapping used to update the CCPs can be relatively hard, especially for large state spaces. In order to deal with these challenges, Kasahara and Shimotsu (2012) propose to use \( \alpha \approx 0 \) in practice. For \( \alpha \) small enough, the spectral radius of interest will always be smaller than 1 (c.f. the proof of their Proposition 5).

Kasahara and Shimotsu (2012) provide Monte-Carlo evidence suggesting that their estimator has very good finite sample properties. Nonetheless, it should be noted that they use \( \alpha^* \), which is usually unfeasible when the data generating process is unknown. Moreover, their application is a very simple model, i.e. a simplified version of the model I use in Section 3.5. Egesdal, Lai, and Su (2013) study the finite sample properties of this estimator applied to the same model as in my Monte-Carlo experiments, but they use \( \alpha = 0.5 \), which may seem far from the \( \alpha \approx 0 \) suggested by Kasahara and Shimotsu (2012). In the Monte-Carlo simulations below, I set \( \alpha = 0.05 \).

3.4.3 Bootstrap

One of the refinements available from the bootstrap is the bias reduction of an estimator (see for instance Hall, 1992 and Horowitz, 1999). The general idea consists of using several bootstrap samples that are drawn with replacement from the original sample in order to estimate the bias and then applying the appropriate correction.

More formally, let \( \theta^0 \) be the vector of the true parameters in the population and let \( \hat{\theta}^{PML} \) be its PML estimate computed from the original sample of \( M \) observations. The bias of the estimator is defined as:

\[
\text{Bias} \left[ \hat{\theta}^{PML} \right] = E \left[ \hat{\theta}^{PML} \right] - \theta^0.
\] (3.11)

The nonparametric bootstrap amounts to considering the original sample as being a population in itself and drawing bootstrap samples from this population. Construct \( B \) bootstrap samples, each time drawing with replacement \( M \) observations from the original sample, and compute \( \hat{\theta}^{PML}_b \) for each \( b \) bootstrap sample. An estimate of the PML estimator’s bias is therefore given by:

\[
\text{Bias} \left[ \hat{\theta}^{PML}_b \right] = B^{-1} \sum_{b=1}^{B} \hat{\theta}^{PML}_b - \hat{\theta}^{PML}.
\] (3.12)

The estimated bias is then subtracted from the PML estimates to construct the bootstrap bias correction in Definition 6.

---

8 The objective function to be minimized is the largest absolute eigenvalue and may therefore not be differentiable everywhere.

9 This comment also applies to the single Newton full maximum likelihood iteration method presented below.

10 See their cases 3 and 4. The only difference with my simulations is the value of \( \theta_{RS} \) in the unstable equilibrium.
Chapter 3. Bias reduction in dynamic discrete games

Definition 6 (Bootstrap bias correction). The bias-corrected estimator using a bootstrap correction is defined as:

$$\hat{\theta}_{BS} = 2\hat{\theta}_{PML} - B^{-1} \sum_{b=1}^{B} \hat{\theta}_{b}^{PML}.$$  

While there is no “rule” in choosing $B$, it should be large enough so that further increasing the number of bootstrap samples used to compute the bias-corrected estimator should have no important effect on this estimator’s properties. Formal arguments regarding why one should expect a bootstrap bias-corrected estimator to perform better in small samples than an uncorrected estimator have already been provided in the literature. For instance, see Horowitz (1999). Some evidence of how the bootstrap effectively reduces the finite sample bias can be found in Hsu, Lau, Fung, and Ulveling (1986) for two-stage least squares estimation of structural parameters in systems of equations; in Horowitz (1998) for the optimal minimum distance estimator of covariance structures; and in Pesaran and Zhao (1998) for the mean group estimator of long run coefficients in dynamic heterogenous panels.

Besides this nonparametric approach, it is also interesting to consider a parametric version of the bootstrap. Drawing with replacement from the original sample amounts to drawing from the distribution defined by $\hat{P}_0$. Alternatively, one could draw observations from the distribution defined by $\Psi(\hat{P}_0, \hat{\theta}_1)$, i.e. the probabilities obtained from a single iteration of the $\Psi$ mapping. Unless $\hat{P}_0$ corresponds to the equilibrium vector of CCPs, these distributions are different. The finite sample properties of the parametric bootstrap should therefore differ from the properties of its nonparametric version.

3.4.4 Jackknife

To some extent, the finite sample bias problem studied here shares similarities with at least two other bodies of the econometric literature: the bias in the estimation of fixed effects in nonlinear panel data models and the bias resulting from the simulation or the discretization of an objective function. The resampling bias corrections recently suggested to deal with these problems often are jackknife rather than bootstrap corrections. Examples are Hahn and Newey (2004), Dhaene and Jochmans (2015) and Kristensen and Salanié (2017).

When Efron (1979) introduced the bootstrap, he showed that the jackknife proposed by Quenouille (1956) and Tukey (1958) was a linear approximation of his method. It is not clear, at this point, if one of these two approaches should perform better than the other when applied to dynamic discrete games.

Typically, one obtains different estimates of a vector of parameters when different subsamples of the initial sample of $M$ observations are used. The jackknife reduces the bias of an initial estimator by exploiting the variation in the estimates from $M$ distinct subsamples of $M - 1$ observations each. More precisely, if $\hat{\theta}_{(m)}$ is a PML estimate computed using all observations except observation $m$, the (first-order) jackknife estimator is defined as follows.

Definition 7 (Jackknife bias correction). The bias-corrected estimator using a jackknife correction is:

$$\hat{\theta}_{JK} = M\hat{\theta}_{PML} - (M - 1)M^{-1} \sum_{m=1}^{M} \hat{\theta}_{(m)}^{PML}.$$  

Notice that, while $B$ is fixed by the econometrician for the bootstrap, the number of samples needed to compute the jackknife correction from the last definition is determined by the sample size $M$. 

3.4.5 Single Newton full maximum likelihood iteration

As proposed by Aguirregabiria and Mira (2007, section 3.6), the applied econometrician could combine the NPL estimator with a single Newton iteration in the full likelihood. This single iteration has the advantage of generating an estimator that is asymptotically efficient, which is an attractive feature given the inefficiency of the NPL estimator when applied to games. However, this single Newton iteration is not feasible if the NPL algorithm fails to converge. Alternatively, building on a similar idea, I propose to initiate a single Newton iteration at the PML estimator, after having updated the CCPs only once. While the resulting estimator is asymptotically efficient, not much can be said about its finite sample properties without performing a Monte-Carlo simulation.

In order to implement this alternative version of the single Newton iteration, one first computes the CCPs associated with \( \hat{\theta}_{\text{PML}} \) from \( \hat{P}_1 = \Phi(\hat{P}_0, \hat{\theta}_{\text{PML}}) \). Then, given \( \hat{P}_1 \) and \( \hat{\theta}_{\text{PML}} \), appropriate estimates of the expected score of the log-likelihood \( \bar{g}_M(\hat{P}_1, \hat{\theta}_{\text{PML}}) \) and of the information matrix \( \hat{I}(\hat{P}_1, \hat{\theta}_{\text{PML}}) \) are constructed to update the vector of structural parameters. The expressions provided in Aguirregabiria and Mira (2007, p. 29) lead to the following estimator:

**Definition 8** (Single Newton full maximum likelihood iteration). The single Newton full maximum likelihood iteration initialized at the PML estimates is:

\[
\hat{\theta}_{\text{NI}} = \hat{\theta}_{\text{PML}} + \hat{I}(\hat{P}_1, \hat{\theta}_{\text{PML}})^{-1} \bar{g}_M(\hat{P}_1, \hat{\theta}_{\text{PML}})
\]

where

\[
\hat{I}(\hat{P}_1, \hat{\theta}_{\text{PML}}) = M^{-1} \sum_{m=1}^{M} \hat{g}_m(\hat{P}_1, \hat{\theta}_{\text{PML}}) \hat{g}_m(\hat{P}_1, \hat{\theta}_{\text{PML}})';
\]

\[
\hat{g}_m(\hat{P}_1, \hat{\theta}_{\text{PML}}) = \frac{\partial \ln \left[ \Psi(y_m|x_m, \hat{\theta}_{\text{PML}}, \hat{P}_1) \right]}{\partial \theta} + \frac{\partial \hat{P}_1(\hat{\theta}_{\text{PML}})'}{\partial \theta} \frac{\partial \ln \left[ \Psi(y_m|x_m, \hat{\theta}_{\text{PML}}, \hat{P}_1) \right]}{\partial P};
\]

\[
\frac{\partial \hat{P}_1(\hat{\theta}_{\text{PML}})'}{\partial \theta} = \frac{\partial \Psi(\hat{\theta}_{\text{PML}}, \hat{P}_1)'}{\partial \theta} \left[ I - \frac{\partial \Psi(\hat{\theta}_{\text{PML}}, \hat{P}_1)}{\partial P'} \right]^{-1};
\]

\[
\bar{g}_M(\hat{P}_1, \hat{\theta}_{\text{PML}}) = M^{-1} \sum_{m=1}^{M} \hat{g}_m(\hat{P}_1, \hat{\theta}_{\text{PML}}).
\]

Of course, this method requires the computation of the Jacobian matrix of \( \Psi(\cdot) \), which can be computationally very intensive in some applications.

3.5 Monte-Carlo experiments

3.5.1 Data generating processes

The data generating processes considered here borrow heavily from Aguirregabiria and Mira (2007)’s experiment 3. The model is presented in subsection 3.3.2 above. Suppose that \( N = 5 \) firms must decide
whether or not to operate in \(M\) independent markets. There are \(N + 1 = 6\) observable state variables: the market size \((S_{mt})\) and the incumbency status of each player \((y_{mt-1})\). \(y_{mt-1}\) is obviously a vector of discrete variables and Aguirregabiria and Mira (2007) consider the support of \(S_{mt}\) to be \(S = \{1, 2, 3, 4, 5\}\). The observable state space is denoted by \(X \equiv S \times Y^N\) and the total number of possible different states is given by \(\dim(X) = 5 \times 2^5 = 160\).

A transition probability distribution has to be specified for the vector of observable state variables in market \(m\) at period \(t\) denoted by \(x_{mt} = [S_{mt}, y'_{mt-1}]'\). At the steady state, the transition probability of \(y_{mt-1}\) trivially corresponds to the firms’ decisions CCPs. Let the transition matrix of \(S_{mt}\) be given by:

\[
\begin{pmatrix}
0.8 & 0.2 & 0 & 0 & 0 \\
0.2 & 0.6 & 0.2 & 0 & 0 \\
0 & 0.2 & 0.6 & 0.2 & 0 \\
0 & 0 & 0.2 & 0.6 & 0.2 \\
0 & 0 & 0 & 0.2 & 0.8
\end{pmatrix}
\]

which means that the market size is most likely to stay the same from one period to another. If the market size does change, it either increases or decreases by one “unit” with equal probability (except for \(S = 1\) and \(S = 5\) where it can only increase and decrease respectively).

Since the contemporaneous utility is linear in parameters, let:

\[
u_i(y_{mt}, x_{mt}) + \varepsilon_{imt}(y_{imt}) = \alpha_i(y_{mt}, x_{mt}) \theta + \varepsilon_{imt}(y_{imt})
\]

where

\[
\alpha_i(1, y_{-imt}, x_{mt}) = \left[ S_{mt}, - \left(1 + \sum_{j \neq i} y_{jmt}\right), -(1 - y_{imt-1}), 0, \ldots, -1, \ldots, 0 \right],
\]

\[
\alpha_i(0, y_{-imt}, x_{mt}) = [0, 0, 0, 0, \ldots, 0, \ldots, 0]
\]

and

\[
\theta = [\theta_{RS}, \theta_{RN}, \theta_{EC}, \theta_{FC,1}, \ldots, \theta_{FC,N}]'.
\]

Let \(\theta_{RS} = 1, \theta_{EC} = 1, \theta_{FC,1} = 1.9, \theta_{FC,2} = 1.8, \theta_{FC,3} = 1.7, \theta_{FC,4} = 1.6\) and \(\theta_{FC,5} = 1.5\). The discount rate is assumed to be known and equal to \(\beta = 0.95\). I consider two different values for the strategic interaction parameter: \(\theta_{RN} = 2\) and \(\theta_{RN} = 4\). Notice that only the former case is considered in Aguirregabiria and Mira (2007)’s experiment 3.

The equilibrium CCPs needed to generate the data solve the fixed point mapping in (3.1). From (3.6) and extreme value type 1 distributed unobservables, the fixed point mapping is:

\[
\left\{ \frac{\exp \{v_i^P(y, x)\}}{\exp \{v_i^P(1-y, x)\} + \exp \{v_i^P(y, x)\}} : i = 1, \ldots, N; \ y \in \{0, 1\} \ \text{and} \ x \in X. \right\}
\]

Since the model considered here is a binary choice model, there is a total of \(N \times \dim(X) = 5 \times 160 = 800\)
Chapter 3. Bias reduction in dynamic discrete games

The equilibrium CCPs can be found by using a built-in solver for a system of nonlinear equations. For $\theta_{RN} = 2$, the spectral radius of $\Psi_p$ is $0.9237 < 1$ and the equilibrium generating the data is therefore Lyapunov stable under this data generating process. For $\theta_{RN} = 4$, it is $1.6748 > 1$, which implies instability.

Given the equilibrium CCPs, the steady state distribution of the observable states must be computed. This is done by iterating the following fixed point mapping implied by the ergodic distribution of the observable states:

$$ p = F(P)p $$ (3.19)

where $p$ is a $\dim(X)$-dimensional vector of the steady state probabilities associated with each $x \in X$; $F(P)$ is the $\dim(X) \times \dim(X)$ matrix of transition probabilities induced by the vector of CCPs, i.e. the probabilities of observing $x'$ in the next period given $x$ being observed today and $P$.

I consider an infinite horizon model. The econometrician observes $M$ independent markets over $T = 1$ period, i.e.

$$ \{y_{1m0}, \ldots, y_{Nm0}, s_{m1}, y_{1m1}, \ldots, y_{Nm1} : m = 1, \ldots, M\} $$

where the observable state variables $\{y_{1m0}, \ldots, y_{Nm0}, s_{m1} : m = 1, \ldots, M\}$ are randomly drawn from the ergodic distribution of the observable states and $\{y_{1m1}, \ldots, y_{Nm1} : m = 1, \ldots, M\}$ are drawn according to the equilibrium CCPs. When implementing the bootstrap and the jackknife, the market is the unit of resampling. In other words, $N$ observations are drawn each time.

### 3.5.2 Comparison of alternative estimators

Several Monte-Carlo experiments are used to compare the performance of the estimators presented above. The first question worth asking is the following: how problematic is the two-step estimator in small samples? Tables 3.1 and 3.2 report the average estimates and RMSE computed over 1,000 samples associated with different possible two-step estimators for the stable and the unstable equilibria respectively. I explore all the possible combinations of the frequency count/Nadaraya-Watson kernel estimators in the first step and the PML/GMM estimators in the second step. Details about the two-step estimation are provided in Appendix 3.7.3. When comparing the different two-step estimators, I use two sample sizes: 400 and 5,000 markets.

Besides the frequency count and the kernel first-step estimators, I report the estimates of an unfeasible two-step estimator in which the CCPs are known to the econometrician, i.e. not estimated in a first step. This unfeasible estimator can be seen as a benchmark since it is not affected by the estimation error in the first step. When 5,000 markets are used in the samples, both the unfeasible PML and the unfeasible GMM estimators generate estimates that are close to their true values. One gets a different pictures when the sample size drops to 400 markets: while the unfeasible PML estimator remains unbiased, the unfeasible GMM second-step estimator becomes heavily biased and very imprecise.

---

11 Only the CCPs associated with one decision (either operating or not operating) are needed since $P_i(1|\theta, x) = 1 - P_i(0|\theta, x)$.
12 I used OxMetrics’ SolveNLE function with its default settings. Similar solvers can be found in other softwares. For instance, one could use Matlab’s fsolve.
13 However, after having chosen a preferred two-step estimator, I only use samples of 400 markets to compare bias reduction techniques.
Table 3.1: Stable equilibrium: Monte-Carlo comparison of two-step estimators

<table>
<thead>
<tr>
<th></th>
<th>( \theta_{RS} )</th>
<th>( \theta_{RN} )</th>
<th>( \theta_{EC} )</th>
<th>( \theta_{FC,1} )</th>
<th>( \theta_{FC,2} )</th>
<th>( \theta_{FC,3} )</th>
<th>( \theta_{FC,4} )</th>
<th>( \theta_{FC,5} )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Step 1</strong></td>
<td><strong>Step 2</strong></td>
<td>1.00</td>
<td>2.00</td>
<td>1.00</td>
<td>1.90</td>
<td>1.80</td>
<td>1.70</td>
<td>1.60</td>
</tr>
<tr>
<td><strong>M = 400</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>True</strong> PML</td>
<td></td>
<td>1.008</td>
<td>2.023</td>
<td>1.002</td>
<td>1.903</td>
<td>1.806</td>
<td>1.706</td>
<td>1.607</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.214]</td>
<td>[0.805]</td>
<td>[0.110]</td>
<td>[0.185]</td>
<td>[0.174]</td>
<td>[0.160]</td>
<td>[0.152]</td>
</tr>
<tr>
<td><strong>Freq</strong> PML</td>
<td></td>
<td>0.389</td>
<td>0.175</td>
<td>1.159</td>
<td>1.386</td>
<td>1.322</td>
<td>1.256</td>
<td>1.187</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.620]</td>
<td>[1.858]</td>
<td>[0.200]</td>
<td>[0.561]</td>
<td>[0.523]</td>
<td>[0.488]</td>
<td>[0.457]</td>
</tr>
<tr>
<td><strong>Kernel</strong> PML</td>
<td></td>
<td>0.490</td>
<td>0.359</td>
<td>1.139</td>
<td>1.606</td>
<td>1.517</td>
<td>1.429</td>
<td>1.330</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.521]</td>
<td>[1.690]</td>
<td>[0.184]</td>
<td>[0.368]</td>
<td>[0.355]</td>
<td>[0.334]</td>
<td>[0.327]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[5.627]</td>
<td>[20.13]</td>
<td>[4.456]</td>
<td>[9.330]</td>
<td>[9.583]</td>
<td>[10.61]</td>
<td>[9.867]</td>
</tr>
<tr>
<td><strong>Freq</strong> GMM</td>
<td></td>
<td>0.240</td>
<td>-6.654</td>
<td>0.480</td>
<td>6.974</td>
<td>7.304</td>
<td>7.160</td>
<td>7.487</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[1.056]</td>
<td>[8.992]</td>
<td>[1.023]</td>
<td>[10.74]</td>
<td>[10.39]</td>
<td>[12.02]</td>
<td>[10.46]</td>
</tr>
<tr>
<td><strong>Kernel</strong> GMM</td>
<td></td>
<td>1.555</td>
<td>-6.798</td>
<td>0.887</td>
<td>11.84</td>
<td>11.74</td>
<td>11.96</td>
<td>12.84</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[2.286]</td>
<td>[9.806]</td>
<td>[2.156]</td>
<td>[18.84]</td>
<td>[18.46]</td>
<td>[19.52]</td>
<td>[20.90]</td>
</tr>
<tr>
<td><strong>M = 5000</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>True</strong> PML</td>
<td></td>
<td>1.001</td>
<td>2.005</td>
<td>1.001</td>
<td>1.897</td>
<td>1.798</td>
<td>1.699</td>
<td>1.598</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.059]</td>
<td>[0.221]</td>
<td>[0.032]</td>
<td>[0.051]</td>
<td>[0.050]</td>
<td>[0.047]</td>
<td>[0.043]</td>
</tr>
<tr>
<td><strong>Freq</strong> PML</td>
<td></td>
<td>0.704</td>
<td>0.919</td>
<td>1.042</td>
<td>1.890</td>
<td>1.781</td>
<td>1.670</td>
<td>1.555</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.303]</td>
<td>[1.107]</td>
<td>[0.054]</td>
<td>[0.061]</td>
<td>[0.062]</td>
<td>[0.064]</td>
<td>[0.071]</td>
</tr>
<tr>
<td><strong>Kernel</strong> PML</td>
<td></td>
<td>0.724</td>
<td>1.049</td>
<td>1.074</td>
<td>1.827</td>
<td>1.717</td>
<td>1.607</td>
<td>1.492</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.280]</td>
<td>[0.969]</td>
<td>[0.080]</td>
<td>[0.095]</td>
<td>[0.102]</td>
<td>[0.109]</td>
<td>[0.120]</td>
</tr>
<tr>
<td><strong>True</strong> GMM</td>
<td></td>
<td>1.010</td>
<td>2.015</td>
<td>1.016</td>
<td>1.914</td>
<td>1.821</td>
<td>1.711</td>
<td>1.629</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.036]</td>
<td>[0.066]</td>
<td>[0.066]</td>
<td>[0.240]</td>
<td>[0.238]</td>
<td>[0.238]</td>
<td>[0.233]</td>
</tr>
<tr>
<td><strong>Freq</strong> GMM</td>
<td></td>
<td>0.755</td>
<td>0.095</td>
<td>0.848</td>
<td>2.906</td>
<td>2.765</td>
<td>2.734</td>
<td>2.571</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.399]</td>
<td>[3.247]</td>
<td>[0.313]</td>
<td>[2.334]</td>
<td>[2.451]</td>
<td>[2.555]</td>
<td>[2.645]</td>
</tr>
<tr>
<td><strong>Kernel</strong> GMM</td>
<td></td>
<td>0.904</td>
<td>0.624</td>
<td>1.084</td>
<td>2.749</td>
<td>2.908</td>
<td>2.793</td>
<td>2.470</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.770]</td>
<td>[3.347]</td>
<td>[0.479]</td>
<td>[4.680]</td>
<td>[7.136]</td>
<td>[6.392]</td>
<td>[4.195]</td>
</tr>
</tbody>
</table>

Notes: Averages and root-mean-square errors (in square brackets) are computed over 1,000 Monte-Carlo simulations. The estimated model corresponds to experiment 3 from Aguirregabiria and Mira (2007). The instruments used in the GMM estimation are the incumbency status dummies, the market size, the square of the market size and the interaction of the market size with the incumbency status dummies. In some cases (typically with \( M = 400 \)), the GMM estimation does not converge. Such cases are dropped.
Table 3.2: Unstable equilibrium: Monte-Carlo comparison of two-step estimators

<table>
<thead>
<tr>
<th>Step 1</th>
<th>Step 2</th>
<th>$\theta_{RS}$</th>
<th>$\theta_{RN}$</th>
<th>$\theta_{EC}$</th>
<th>$\theta_{FC,1}$</th>
<th>$\theta_{FC,2}$</th>
<th>$\theta_{FC,3}$</th>
<th>$\theta_{FC,4}$</th>
<th>$\theta_{FC,5}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>PML</td>
<td>1.018</td>
<td>4.093</td>
<td>0.998</td>
<td>1.905</td>
<td>1.810</td>
<td>1.711</td>
<td>1.610</td>
<td>1.520</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.141]</td>
<td>[0.775]</td>
<td>[0.155]</td>
<td>[0.229]</td>
<td>[0.215]</td>
<td>[0.199]</td>
<td>[0.188]</td>
<td>[0.196]</td>
</tr>
<tr>
<td>Freq</td>
<td>PML</td>
<td>0.341</td>
<td>0.313</td>
<td>1.475</td>
<td>1.693</td>
<td>1.581</td>
<td>1.455</td>
<td>1.276</td>
<td>0.906</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.666]</td>
<td>[3.711]</td>
<td>[0.502]</td>
<td>[0.376]</td>
<td>[0.359]</td>
<td>[0.363]</td>
<td>[0.406]</td>
<td>[0.630]</td>
</tr>
<tr>
<td>Kernel</td>
<td>PML</td>
<td>0.365</td>
<td>0.519</td>
<td>1.506</td>
<td>1.682</td>
<td>1.596</td>
<td>1.470</td>
<td>1.288</td>
<td>0.906</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.648]</td>
<td>[3.546]</td>
<td>[0.551]</td>
<td>[0.400]</td>
<td>[0.372]</td>
<td>[0.363]</td>
<td>[0.399]</td>
<td>[0.625]</td>
</tr>
<tr>
<td>True</td>
<td>GMM</td>
<td>1.713</td>
<td>6.624</td>
<td>1.626</td>
<td>4.348</td>
<td>3.847</td>
<td>3.125</td>
<td>3.147</td>
<td>2.691</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[9.691]</td>
<td>[55.63]</td>
<td>[4.984]</td>
<td>[11.31]</td>
<td>[10.21]</td>
<td>[8.063]</td>
<td>[8.331]</td>
<td>[8.991]</td>
</tr>
<tr>
<td>Freq</td>
<td>GMM</td>
<td>0.465</td>
<td>-4.705</td>
<td>0.904</td>
<td>9.365</td>
<td>9.535</td>
<td>8.658</td>
<td>7.590</td>
<td>4.415</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[1.311]</td>
<td>[11.66]</td>
<td>[2.304]</td>
<td>[16.41]</td>
<td>[16.95]</td>
<td>[13.85]</td>
<td>[13.30]</td>
<td>[8.298]</td>
</tr>
<tr>
<td>Kernel</td>
<td>GMM</td>
<td>0.906</td>
<td>-4.087</td>
<td>1.100</td>
<td>11.99</td>
<td>11.97</td>
<td>11.11</td>
<td>9.596</td>
<td>5.911</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[1.710]</td>
<td>[10.82]</td>
<td>[2.506]</td>
<td>[21.42]</td>
<td>[23.32]</td>
<td>[21.96]</td>
<td>[18.80]</td>
<td>[10.64]</td>
</tr>
</tbody>
</table>

$M = 400$

| True   | PML    | 1.001        | 4.007        | 0.999        | 1.899          | 1.799          | 1.701          | 1.599          | 1.500          |
|        |        | [0.039]      | [0.219]      | [0.044]      | [0.062]        | [0.057]        | [0.055]        | [0.050]        | [0.055]        |
| Freq   | PML    | 0.664        | 2.075        | 1.202        | 1.939          | 1.824          | 1.693          | 1.515          | 1.200          |
|        |        | [0.340]      | [1.946]      | [0.209]      | [0.093]        | [0.084]        | [0.078]        | [0.113]        | [0.311]        |
| Kernel | PML    | 0.705        | 2.463        | 1.202        | 1.835          | 1.722          | 1.595          | 1.429          | 1.155          |
|        |        | [0.299]      | [1.564]      | [0.210]      | [0.106]        | [0.113]        | [0.130]        | [0.186]        | [0.353]        |
| True   | GMM    | 1.003        | 4.007        | 1.011        | 1.903          | 1.813          | 1.718          | 1.609          | 1.509          |
|        |        | [0.027]      | [0.076]      | [0.074]      | [0.234]        | [0.222]        | [0.210]        | [0.194]        | [0.131]        |
| Freq   | GMM    | 0.815        | 1.263        | 0.831        | 4.555          | 4.129          | 3.904          | 3.058          | 1.956          |
|        |        | [0.569]      | [4.343]      | [0.379]      | [6.569]        | [5.935]        | [8.082]        | [4.206]        | [2.118]        |
| Kernel | GMM    | 0.792        | 1.933        | 0.980        | 3.402          | 3.052          | 3.121          | 2.649          | 1.901          |
|        |        | [0.371]      | [3.436]      | [0.405]      | [5.034]        | [3.789]        | [5.864]        | [6.252]        | [3.110]        |

Notes: Averages and root-mean-square errors (in square brackets) are computed over 1,000 Monte-Carlo simulations. The estimated model corresponds to experiment 3 from Aguirregabiria and Mira (2007), with $\theta_{RN} = 4$. The instruments used in the GMM estimation are the incumbency status dummies, the market size, the square of the market size and the interaction of the market size with the incumbency status dummies. In some cases (typically with $M = 400$), the GMM estimation does not converge. Such cases are dropped.
When looking at the feasible versions of the two-step estimators, one can see that, for both sample sizes, it is typically the case that the PML second-step estimator performs better than the GMM estimator for a given first-step estimator. The GMM estimator actually performs very badly when estimating player-specific parameters. Unfortunately, determining the best first-step estimator is not as obvious. Keeping the second-step estimator fixed, there are some cases where the Nadaraya-Watson kernel first-step estimator performs better than the frequency count estimator and other cases where the frequency count estimator is preferable. However, since the kernel first-step together with the PML second-step generate the lowest RMSE for all parameters in the stable case and for most parameters in the unstable case when \( M = 400 \), I choose this combination to be my preferred two-step estimator.

Two more remarks are worth pointing out about this comparison of different two-step estimators. First, even the preferred two-step estimator exhibits disappointing finite sample properties. In other words, there does not seem to be a specific combination of a first and a second step estimators that allows one to get rid of the poor finite sample performance of the two-step approach. Second, this preferred two-step estimator is not exactly in line with Pakes, Ostrovsky, and Berry (2007)'s best estimator. On one hand, their experiments suggest that a smooth first step can improve the finite sample properties, which is also what I get in some cases. On the other hand, they note that a GMM second-step usually outer performs the PML approach, which is not confirmed by Tables 3.1 and 3.2. Perhaps, the difference in the sample sizes, the sizes of the state spaces and the presence of some player-specific parameters could explain the discrepancy between my preferred two-step estimator and theirs.

I can now discuss the performance of the different bias reduction methods. Average estimates and RMSE computed over 1,000 samples of 400 markets are reported in Tables 3.3 and 3.4 for the stable and the unstable cases respectively. The first sets of results reported in these tables correspond to the unfeasible preferred two-step estimator. In Table 3.5, I report the ratios of the RMSE of each estimator divided by the RMSE of this unfeasible two-step estimator. I use the average of these RMSE ratios computed over all parameters to compare the different estimators. The full solution maximum likelihood could also have been used as a benchmark, but the two-step approach with known CCPs is easier to implement in Monte-Carlo simulations.

Replacing the true values of the CCPs with their nonparametric first-step estimates (using the Nadaraya-Watson kernel estimator) generates an important bias for all parameters in both equilibria. As expected, this suggests that the first-step estimation error is driving most of the finite sample bias in the second step. It is worth noting that the RMSE of the two-step estimator is larger for the unstable equilibrium compared to the stable one: on average, almost three times larger than the RMSE of its unfeasible version versus a little bit more than two. The two-step approach itself is therefore more problematic when stability fails.

\[^{14}\text{It should also be noted that when } M = 400, \text{ the GMM estimator often failed to converge, in which case I simply dropped the samples from the Monte-Carlo.}\]
Table 3.3: Stable equilibrium: Monte-Carlo comparison of the estimators

<table>
<thead>
<tr>
<th>Estimator</th>
<th>$\theta_{RS}$ 1.00</th>
<th>$\theta_{RN}$ 2.00</th>
<th>$\theta_{EC}$ 1.00</th>
<th>$\theta_{FC,1}$ 1.90</th>
<th>$\theta_{FC,2}$ 1.80</th>
<th>$\theta_{FC,3}$ 1.70</th>
<th>$\theta_{FC,4}$ 1.60</th>
<th>$\theta_{FC,5}$ 1.50</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-step, true CCPs</td>
<td>1.0075</td>
<td>2.0226</td>
<td>1.0023</td>
<td>1.9025</td>
<td>1.8062</td>
<td>1.7059</td>
<td>1.6066</td>
<td>1.5070</td>
</tr>
<tr>
<td></td>
<td>[0.2137]</td>
<td>[0.8054]</td>
<td>[0.1095]</td>
<td>[0.1852]</td>
<td>[0.1743]</td>
<td>[0.1599]</td>
<td>[0.1517]</td>
<td>[0.1500]</td>
</tr>
<tr>
<td>2-step</td>
<td>0.4903</td>
<td>0.3592</td>
<td>1.1388</td>
<td>1.6061</td>
<td>1.5174</td>
<td>1.4286</td>
<td>1.3302</td>
<td>1.2301</td>
</tr>
<tr>
<td></td>
<td>[0.5211]</td>
<td>[1.6903]</td>
<td>[0.1837]</td>
<td>[0.3683]</td>
<td>[0.3552]</td>
<td>[0.3343]</td>
<td>[0.3267]</td>
<td>[0.3227]</td>
</tr>
<tr>
<td>K-step, $K=100$</td>
<td>0.9485</td>
<td>1.7979</td>
<td>1.0120</td>
<td>1.9122</td>
<td>1.8120</td>
<td>1.7082</td>
<td>1.6052</td>
<td>1.5018</td>
</tr>
<tr>
<td></td>
<td>[0.1952]</td>
<td>[0.7018]</td>
<td>[0.1123]</td>
<td>[0.2259]</td>
<td>[0.2185]</td>
<td>[0.2091]</td>
<td>[0.2045]</td>
<td>[0.2054]</td>
</tr>
<tr>
<td>AM 2011, $K=100$</td>
<td>0.5787</td>
<td>0.4229</td>
<td>1.0729</td>
<td>1.9413</td>
<td>1.8204</td>
<td>1.6932</td>
<td>1.5645</td>
<td>1.4306</td>
</tr>
<tr>
<td></td>
<td>[0.4335]</td>
<td>[1.6248]</td>
<td>[0.1300]</td>
<td>[0.2021]</td>
<td>[0.1881]</td>
<td>[0.1742]</td>
<td>[0.1719]</td>
<td>[0.1773]</td>
</tr>
<tr>
<td>KS 2012, $K=100$</td>
<td>0.5285</td>
<td>0.4618</td>
<td>1.1257</td>
<td>1.6653</td>
<td>1.5726</td>
<td>1.4777</td>
<td>1.3758</td>
<td>1.2721</td>
</tr>
<tr>
<td></td>
<td>[0.4845]</td>
<td>[1.5965]</td>
<td>[0.1702]</td>
<td>[0.3127]</td>
<td>[0.3030]</td>
<td>[0.2888]</td>
<td>[0.2855]</td>
<td>[0.2843]</td>
</tr>
<tr>
<td>Bootstrap, $B=99$</td>
<td>0.5611</td>
<td>0.4937</td>
<td>1.0613</td>
<td>1.7625</td>
<td>1.6615</td>
<td>1.5582</td>
<td>1.4449</td>
<td>1.3328</td>
</tr>
<tr>
<td></td>
<td>[0.4635]</td>
<td>[1.6129]</td>
<td>[0.1333]</td>
<td>[0.2892]</td>
<td>[0.2825]</td>
<td>[0.2637]</td>
<td>[0.2599]</td>
<td>[0.2592]</td>
</tr>
<tr>
<td>Param bootstrap, $B=99$</td>
<td>0.5751</td>
<td>0.6362</td>
<td>1.0506</td>
<td>1.6468</td>
<td>1.5545</td>
<td>1.4683</td>
<td>1.3670</td>
<td>1.2680</td>
</tr>
<tr>
<td></td>
<td>[0.4619]</td>
<td>[1.5433]</td>
<td>[0.1265]</td>
<td>[0.3724]</td>
<td>[0.3626]</td>
<td>[0.3338]</td>
<td>[0.3255]</td>
<td>[0.3179]</td>
</tr>
<tr>
<td>Jackknife</td>
<td>0.6255</td>
<td>0.6646</td>
<td>1.0579</td>
<td>1.8560</td>
<td>1.7379</td>
<td>1.6218</td>
<td>1.4949</td>
<td>1.3733</td>
</tr>
<tr>
<td></td>
<td>[0.4268]</td>
<td>[1.6212]</td>
<td>[0.1948]</td>
<td>[0.4802]</td>
<td>[0.4609]</td>
<td>[0.4312]</td>
<td>[0.4171]</td>
<td>[0.4020]</td>
</tr>
<tr>
<td>Single NI</td>
<td>1.1290</td>
<td>2.5152</td>
<td>1.0374</td>
<td>1.8389</td>
<td>1.7513</td>
<td>1.6617</td>
<td>1.5729</td>
<td>1.4852</td>
</tr>
<tr>
<td></td>
<td>[0.4665]</td>
<td>[1.7256]</td>
<td>[0.1458]</td>
<td>[0.2795]</td>
<td>[0.2708]</td>
<td>[0.2620]</td>
<td>[0.2583]</td>
<td>[0.2660]</td>
</tr>
</tbody>
</table>

Notes: Averages and root-mean-square errors (in square brackets) are computed over 1,000 Monte-Carlo simulations. The estimated model corresponds to experiment 3 from Aguirregabiria and Mira (2007). In KS 2012, $\alpha = 0.05$. 

CHAPTER 3. BIAS REDUCTION IN DYNAMIC DISCRETE GAMES
All other estimators studied here have lower finite sample bias compared to the preferred two-step estimator for all parameters and for both equilibria. Now, the main question of interest is whether or not any of these estimators does better than the $K$-step estimator. With a RMSE only 16\% larger than the unfeasible estimator on average, the $K$-step approach is associated with good finite sample properties for the stable equilibrium. Actually, for the logarithm of the population ($\theta_{RS}$) and for the strategic interaction ($\theta_{RN}$) parameters, the $K$-step estimator is associated with RMSEs that are smaller than the unfeasible two-step approach. Even if these two parameters estimates are more biased than their unfeasible counterparts, they exhibit a much lower variance. However, in line with criticisms made towards the $K$-step estimator, its finite sample properties are less attractive when the equilibrium generating the data is unstable: the RMSE becomes 56\% larger than the unfeasible two-step on average. Notice that this is still a substantial improvement compared to the two-step semiparametric estimator.

Aguirregabiria and Mira (2011)’s modified $K$-step estimator has, overall, worse finite sample properties than the $K$-step: its RMSE are respectively 35\% and 126\% larger than the unfeasible benchmark on average. Interestingly, this poor performance is mainly driven by the estimation of the parameters common to all players. In fact, the player-specific fixed cost estimates are often associated with lower RMSE than the $K$-step estimator, both in the stable and the unstable equilibria. This observation suggests that keeping fixed the CCPs of the other players when updating the beliefs of a given firm is not a “perfect substitute” to a process that updates each player’s CCPs simultaneously. If it were, one would expect both the player-specific and the common parameters estimates to exhibit similar finite sample properties as the usual $K$-step estimator for $K$ large enough.

With a RMSE 49\% larger than the unfeasible two-step on average, Kasahara and Shimotsu (2012)’s modified $K$-step estimator does perform better than the $K$-step estimator for the unstable equilibrium, but not for the stable one (85\% larger). These results contrast with the Monte-Carlo evidence provided by Kasahara and Shimotsu (2012), at least for the stable case. When the equilibrium generating the data is stable, in a simplified version of the model considered here and with the optimal $\alpha^*$, they obtain the same finite sample properties as the $K$-step estimator for $k$ as low as 10.\footnote{They do not study the $K$-step \textit{per se}, but rather the NPL estimator. However, since the NPL estimator seems to have converged for small $k$’s in the stable equilibrium for their simplified version, the NPL and the $K$-step estimators should lead to the same results.} The main conclusion one should draw from these results is that the finite sample properties of this version of the $K$-step estimator with $\alpha \approx 0$, suggested to be used in practice, may be fairly different from the results associated with the optimal $\alpha^*$, even after iterating $K = 100$ times.

The resampling methods perform very poorly. Even if they reduce the finite sample bias compared to the two-step estimator, their RMSE ratios remain large. On average, the nonparametric bootstrap RMSE is 71\% larger than the unfeasible benchmark in the stable case and 176\% larger in the unstable case. The parametric bootstrap gives somewhat similar results with a RMSE 96\% and 162\% greater than the benchmark, on average, for the stable and the unstable equilibria respectively. For the jackknife, the corresponding values are 139\% and 220\% larger, which indicates that this resampling method performs actually worse than the feasible two-step estimator in finite samples.

The single Newton full maximum likelihood iteration seems to generally perform worse than the $K$-step estimator. On average, its RMSE ratios are 73\% and 123\% larger than the unfeasible two-step estimator in the stable and the unstable equilibria respectively. However, a closer look at the estimates suggest that, in the unstable case, this method can generate a lower bias than the $K$-step estimator (but with a higher variance) for some parameters. On one hand, it is reasonable to think that increasing...
Table 3.4: Unstable equilibrium: Monte-Carlo comparison of the estimators

<table>
<thead>
<tr>
<th>Estimator</th>
<th>$\theta_{\text{RS}}$</th>
<th>$\theta_{\text{RN}}$</th>
<th>$\theta_{\text{EC}}$</th>
<th>$\theta_{\text{FC,1}}$</th>
<th>$\theta_{\text{FC,2}}$</th>
<th>$\theta_{\text{FC,3}}$</th>
<th>$\theta_{\text{FC,4}}$</th>
<th>$\theta_{\text{FC,5}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-step, true CCPs</td>
<td>1.0180</td>
<td>4.0928</td>
<td>0.9980</td>
<td>1.9051</td>
<td>1.8104</td>
<td>1.7113</td>
<td>1.6104</td>
<td>1.5199</td>
</tr>
<tr>
<td></td>
<td>[0.1409]</td>
<td>[0.7754]</td>
<td>[0.1553]</td>
<td>[0.2288]</td>
<td>[0.2147]</td>
<td>[0.1994]</td>
<td>[0.1879]</td>
<td>[0.1964]</td>
</tr>
<tr>
<td>2-step</td>
<td>0.3646</td>
<td>0.5192</td>
<td>1.5060</td>
<td>1.6824</td>
<td>1.5960</td>
<td>1.4698</td>
<td>1.2884</td>
<td>0.9065</td>
</tr>
<tr>
<td></td>
<td>[0.6475]</td>
<td>[3.5455]</td>
<td>[0.5514]</td>
<td>[0.4000]</td>
<td>[0.3724]</td>
<td>[0.3630]</td>
<td>[0.3993]</td>
<td>[0.6249]</td>
</tr>
<tr>
<td>K-step, $K=100$</td>
<td>0.7512</td>
<td>2.5293</td>
<td>1.2210</td>
<td>1.9613</td>
<td>1.8455</td>
<td>1.7107</td>
<td>1.5384</td>
<td>1.2531</td>
</tr>
<tr>
<td></td>
<td>[0.2642]</td>
<td>[1.5027]</td>
<td>[0.2537]</td>
<td>[0.2911]</td>
<td>[0.2806]</td>
<td>[0.2655]</td>
<td>[0.2608]</td>
<td>[0.3378]</td>
</tr>
<tr>
<td>AM 2011, $K=100$</td>
<td>0.4575</td>
<td>0.8763</td>
<td>1.3403</td>
<td>2.0830</td>
<td>1.9404</td>
<td>1.7600</td>
<td>1.5054</td>
<td>1.0045</td>
</tr>
<tr>
<td></td>
<td>[0.5500]</td>
<td>[3.1641]</td>
<td>[0.3661]</td>
<td>[0.3125]</td>
<td>[0.2794]</td>
<td>[0.2330]</td>
<td>[0.2275]</td>
<td>[0.5247]</td>
</tr>
<tr>
<td>KS 2012, $K=100$</td>
<td>1.0975</td>
<td>4.5597</td>
<td>0.8988</td>
<td>1.9046</td>
<td>1.8145</td>
<td>1.7312</td>
<td>1.6738</td>
<td>1.6630</td>
</tr>
<tr>
<td></td>
<td>[0.1855]</td>
<td>[0.9175]</td>
<td>[0.2041]</td>
<td>[0.3297]</td>
<td>[0.3244]</td>
<td>[0.3182]</td>
<td>[0.3204]</td>
<td>[0.3682]</td>
</tr>
<tr>
<td>Bootstrap, $B=99$</td>
<td>0.4172</td>
<td>0.7494</td>
<td>1.3982</td>
<td>1.8036</td>
<td>1.7182</td>
<td>1.5780</td>
<td>1.3790</td>
<td>0.9645</td>
</tr>
<tr>
<td></td>
<td>[0.6088]</td>
<td>[3.4024]</td>
<td>[0.4794]</td>
<td>[0.4115]</td>
<td>[0.3791]</td>
<td>[0.3530]</td>
<td>[0.3673]</td>
<td>[0.5832]</td>
</tr>
<tr>
<td>Param bootstrap, $B=99$</td>
<td>0.4882</td>
<td>1.1711</td>
<td>1.3068</td>
<td>1.7852</td>
<td>1.6854</td>
<td>1.5453</td>
<td>1.3649</td>
<td>1.0244</td>
</tr>
<tr>
<td></td>
<td>[0.5539]</td>
<td>[3.1029]</td>
<td>[0.3973]</td>
<td>[0.4309]</td>
<td>[0.4031]</td>
<td>[0.3801]</td>
<td>[0.3902]</td>
<td>[0.5403]</td>
</tr>
<tr>
<td>Jackknife</td>
<td>0.4845</td>
<td>1.0819</td>
<td>1.3049</td>
<td>1.9235</td>
<td>1.8305</td>
<td>1.6728</td>
<td>1.4533</td>
<td>1.0225</td>
</tr>
<tr>
<td></td>
<td>[0.5850]</td>
<td>[3.3439]</td>
<td>[0.4980]</td>
<td>[0.6441]</td>
<td>[0.6001]</td>
<td>[0.5338]</td>
<td>[0.4899]</td>
<td>[0.6021]</td>
</tr>
<tr>
<td>Single NI</td>
<td>0.7865</td>
<td>2.8580</td>
<td>1.3804</td>
<td>1.7231</td>
<td>1.6295</td>
<td>1.5112</td>
<td>1.3639</td>
<td>1.1277</td>
</tr>
<tr>
<td></td>
<td>[0.3491]</td>
<td>[1.9197]</td>
<td>[0.4416]</td>
<td>[0.4008]</td>
<td>[0.3863]</td>
<td>[0.3749]</td>
<td>[0.3925]</td>
<td>[0.5010]</td>
</tr>
</tbody>
</table>

Notes: Averages and root-mean-square errors (in square brackets) are computed over 1,000 Monte-Carlo simulations. The estimated model corresponds to experiment 3 from Aguirregabiria and Mira (2007), with $\theta_{\text{RN}} = 4$. In KS 2012, $\alpha = 0.05$. The ML estimation did not converge for a very small proportion of bootstrap samples. Such cases were simply dropped.
Table 3.5: RMSE ratios (compared to the unfeasible two-step estimator)

<table>
<thead>
<tr>
<th>Estimator</th>
<th>$\theta_{RS}$</th>
<th>$\theta_{RN}$</th>
<th>$\theta_{EC}$</th>
<th>$\theta_{FC,1}$</th>
<th>$\theta_{FC,2}$</th>
<th>$\theta_{FC,3}$</th>
<th>$\theta_{FC,4}$</th>
<th>$\theta_{FC,5}$</th>
<th>average</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Stable equilibrium</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-step</td>
<td>2.44</td>
<td>2.10</td>
<td>1.68</td>
<td>1.99</td>
<td>2.04</td>
<td>2.09</td>
<td>2.15</td>
<td>2.15</td>
<td>2.08</td>
</tr>
<tr>
<td>K-step, K=100</td>
<td>0.91</td>
<td>0.87</td>
<td>1.03</td>
<td>1.22</td>
<td>1.25</td>
<td>1.31</td>
<td>1.35</td>
<td>1.37</td>
<td>1.16</td>
</tr>
<tr>
<td>AM 2011, K=100</td>
<td>2.03</td>
<td>2.02</td>
<td>1.19</td>
<td>1.09</td>
<td>1.08</td>
<td>1.09</td>
<td>1.13</td>
<td>1.18</td>
<td>1.35</td>
</tr>
<tr>
<td>KS 2012, K=100</td>
<td>2.27</td>
<td>1.98</td>
<td>1.55</td>
<td>1.69</td>
<td>1.74</td>
<td>1.81</td>
<td>1.88</td>
<td>1.90</td>
<td>1.85</td>
</tr>
<tr>
<td>Bootstrap, B=99</td>
<td>2.17</td>
<td>2.00</td>
<td>1.22</td>
<td>1.56</td>
<td>1.62</td>
<td>1.65</td>
<td>1.71</td>
<td>1.73</td>
<td>1.71</td>
</tr>
<tr>
<td>Param bootstrap, B=99</td>
<td>2.16</td>
<td>1.92</td>
<td>1.15</td>
<td>2.01</td>
<td>2.08</td>
<td>2.09</td>
<td>2.14</td>
<td>2.12</td>
<td>1.96</td>
</tr>
<tr>
<td>Jackknife</td>
<td>2.00</td>
<td>2.01</td>
<td>1.78</td>
<td>2.59</td>
<td>2.64</td>
<td>2.70</td>
<td>2.75</td>
<td>2.68</td>
<td>2.39</td>
</tr>
<tr>
<td>Single NI</td>
<td>2.18</td>
<td>2.14</td>
<td>1.33</td>
<td>1.51</td>
<td>1.55</td>
<td>1.64</td>
<td>1.70</td>
<td>1.77</td>
<td>1.73</td>
</tr>
<tr>
<td><strong>Unstable equilibrium</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-step</td>
<td>4.60</td>
<td>4.57</td>
<td>3.55</td>
<td>1.75</td>
<td>1.73</td>
<td>1.82</td>
<td>2.13</td>
<td>3.18</td>
<td>2.97</td>
</tr>
<tr>
<td>K-step, K=100</td>
<td>1.88</td>
<td>1.94</td>
<td>1.63</td>
<td>1.27</td>
<td>1.31</td>
<td>1.33</td>
<td>1.39</td>
<td>1.72</td>
<td>1.56</td>
</tr>
<tr>
<td>AM 2011, K=100</td>
<td>3.90</td>
<td>4.08</td>
<td>2.36</td>
<td>1.37</td>
<td>1.30</td>
<td>1.17</td>
<td>1.21</td>
<td>2.67</td>
<td>2.26</td>
</tr>
<tr>
<td>KS 2012, K=100</td>
<td>1.32</td>
<td>1.18</td>
<td>1.31</td>
<td>1.44</td>
<td>1.51</td>
<td>1.60</td>
<td>1.71</td>
<td>1.87</td>
<td>1.49</td>
</tr>
<tr>
<td>Bootstrap, B=99</td>
<td>4.32</td>
<td>4.39</td>
<td>3.09</td>
<td>1.80</td>
<td>1.77</td>
<td>1.77</td>
<td>1.95</td>
<td>2.97</td>
<td>2.76</td>
</tr>
<tr>
<td>Param bootstrap, B=99</td>
<td>3.93</td>
<td>4.00</td>
<td>2.56</td>
<td>1.88</td>
<td>1.88</td>
<td>1.91</td>
<td>2.08</td>
<td>2.75</td>
<td>2.62</td>
</tr>
<tr>
<td>Jackknife</td>
<td>4.15</td>
<td>4.31</td>
<td>3.21</td>
<td>2.82</td>
<td>2.80</td>
<td>2.68</td>
<td>2.61</td>
<td>3.07</td>
<td>3.20</td>
</tr>
<tr>
<td>Single NI</td>
<td>2.48</td>
<td>2.48</td>
<td>2.84</td>
<td>1.75</td>
<td>1.80</td>
<td>1.88</td>
<td>2.09</td>
<td>2.55</td>
<td>2.23</td>
</tr>
</tbody>
</table>

Notes: The numbers correspond to the RMSE of the parameters estimates divided by the corresponding RMSE of the unfeasible two-step estimator with known CCPs. The averages are the average ratios computed over all parameters.
the number of iterations should further improve the finite sample properties as each iteration brings the estimates closer to their MLE counterparts. On the other hand, each iteration is computationally very costly.

To summarize, for a fixed number of iterations ($K = 100$), none of the estimators compared here does better than the $K$-step estimator for all parameters in both the stable and the unstable equilibria. It follows that none of these estimators stands as an obvious instability-robust alternative to the $K$-step estimator.

### 3.5.3 Convergence issues

As already mentioned above, previous studies have pointed out that the NPL estimator may fail to converge. Pesendorfer and Schmidt-Dengler (2008, 2010) provide examples for which the NPL estimator does not converge or converge to wrong estimates. While instability of the equilibrium generating the data is the explanation for their results, Monte-Carlo evidence from Egesdal, Lai, and Su (2013) suggests that the NPL estimator may fail to converge even when stability holds. An interesting question to ask is whether this lack of convergence prevents any gains in terms of finite sample properties.

Even if I updated the vectors of the CCPs and the structural parameters for a fixed $K$ number of times without considering convergence, looking at the sequences of estimates generated by the three different versions of the $K$-step estimator allows me to assess whether or not convergence has been reached. The percentage of sequences of CCPs and structural parameters which converge are reported in Table 3.6. Even in the stable case, the $K$-step estimator converges in less than 60% of the samples, which is in line with Egesdal, Lai, and Su (2013)’s results. As expected, the $K$-step estimator almost never converges in the unstable equilibrium. As for the modified $K$-step estimators, Aguirregabiria and Mira (2011)’s version almost always converges and Kasahara and Shimotsu (2012)’s rarely converges (for $\alpha = 0.05$ and $K = 100$).

At first, the relative convergence of the different $K$-step estimators is puzzling: even if Aguirregabiria and Mira (2011)’s modified $K$-step estimator almost always converges, its finite sample properties are worse than the original $K$-step estimator in the stable case and worse than both the original $K$-step estimator and Kasahara and Shimotsu (2012)’s modified version in the unstable equilibrium. Figures 3.1 and 3.2 shed some light on these results. In these figures, I plot the evolution of the absolute value of the bias and the RMSE for $k \in \{1, 2, \ldots, K\}$ associated with the strategic interaction parameter for the stable and the unstable equilibria respectively. In the interest of space, I do not include the plots associated with the other parameters. They would lead to results that are qualitatively similar. The

---

**Table 3.6: Proportions (in %) of Monte-Carlo samples that converged**

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Stable CCP</th>
<th>Stable $\theta$</th>
<th>Unstable CCP</th>
<th>Unstable $\theta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K$-step, $K = 100$</td>
<td>59.6</td>
<td>59.7</td>
<td>3.4</td>
<td>4.0</td>
</tr>
<tr>
<td>AM 2011, $K = 100$</td>
<td>99.9</td>
<td>99.7</td>
<td>99.9</td>
<td>99.8</td>
</tr>
<tr>
<td>KS 2012, $K = 100$</td>
<td>0.8</td>
<td>12.5</td>
<td>0.0</td>
<td>0.3</td>
</tr>
</tbody>
</table>

Notes: Convergence is reached if the supremum norm of the difference between two successive vectors of estimates in the sequence of $K = 100$ estimates is smaller than $10^{-3}$ at least once. $10^{-3}$ corresponds to 0.1% of the smallest absolute value of the structural parameters to be estimated.
values used to draw these plots are averages over 1,000 Monte-Carlo samples. Since $k = 1$ corresponds to the two-step semiparametric estimator, all three methods have the same absolute value of the bias and the same RMSE at this specific point.

Consider the modified $K$-step estimators first. From Figures 3.1 and 3.2, one can clearly see that the finite sample properties of Aguirregabiria and Mira (2011)’s modified $K$-step estimator converge very quickly, but generate a relatively small improvement upon the two-step semiparametric estimator. The evolution of the finite sample properties of Kasahara and Shimotsu (2012)’s modified $K$-step estimator follows a very interesting path, where two sequences of improvement are separated by a sequence of degradation of the bias and the RMSE. Of course, different paths would be obtained with different $\alpha$’s.

The plots associated with the $K$-step estimator suggest that, for both equilibria, its bias and RMSE fail to converge to a single point, because the average estimates of $\theta_{RN}$ oscillate between sets of values that are far enough from each others to prevent convergence prior to $K$.\footnote{The oscillation may not be obvious on the plot associated with the stable equilibrium because the difference between two subsequent estimates is very small (about $10^{-2}$ at $k = 100$), but it still remains.} Of course, such cycles are problematic if one is trying to assess the finite sample properties of the NPL estimator: one cannot properly assess the properties of an estimator defined upon convergence if convergence to a single point is not reached. However, if one omits the convergence and iterates the updating process $K$ times, one can still aggregate all the sequences of estimates to assess the finite sample properties of the estimator. In particular, even if the average over all sequences does not converge to a single point, but converges to a stable cycle, all that matters to assess an improvement in the finite sample properties at a fixed $k$ compared to the two-step estimator is where this stable cycle lays at this $k$ compared to $k = 1$.\footnote{Convergence to a stable cycle has already been noted by Kasahara and Shimotsu (2012), but to my knowledge, the potential improvement of the finite sample properties associated with this stable cycle has been, so far, overlooked.}

In other words, as long as the “upper bound” of the cycle of the RMSE is lower than the RMSE at $k = 1$, better finite sample properties can be expected from the $K$-step estimator for $k$ large enough to have reached the stable cycle.

Table 3.7 and Figures 3.3 and 3.4 provide further evidence about the cyclical patterns in the estimates of the structural parameters. For each type of equilibrium, Table 3.7 breaks down the 1,000 Monte-Carlo samples in three different subsamples according to the convergence of the generated sequences of estimates: convergence to a single point, convergence to a cycle of two points and other samples. Notice that 21.4% and 52.7% of the sequences of estimates converge to a two-points cycle in the stable and the unstable equilibrium respectively. The properties of the estimated structural parameters in each subsample are also reported in this table.

The most striking observation is that subsamples in which the sequence fails to converge to a single point can be associated with better finite sample properties compared to converging sequences, specially for $\theta_{RS}$ and $\theta_{RN}$. Actually, this remark is probably the key reason explaining why the $K$-step estimator is performing better than other studies have suggested. By focussing only on converging sequences, one throws away estimates with finite sample properties that are, on average, better than the estimates actually kept. However, it is surprising to see that even the samples that fall under the other category can be associated with relatively better finite sample properties compared to the sequences of estimates converging to a single point. One possible explanation would be that the sequences of estimates in this category are not necessarily “ill behaved”. For instance, some of them could actually exhibit another form of convergence that is different from convergence to a single point or to a pair of points. In fact, a small amount of sequences did converge to cycles of three or four points in the stable equilibrium. Another
Figure 3.1: Stable equilibrium: estimation of $\theta_{RN}$
Figure 3.2: Unstable equilibrium: estimation of $\theta_{RN}$
Table 3.7: Convergence of the $K$-step estimator

<table>
<thead>
<tr>
<th>Convergence</th>
<th>$\theta_{RS}$</th>
<th>$\theta_{RN}$</th>
<th>$\theta_{EC}$</th>
<th>$\theta_{FC,1}$</th>
<th>$\theta_{FC,2}$</th>
<th>$\theta_{FC,3}$</th>
<th>$\theta_{FC,4}$</th>
<th>$\theta_{FC,5}$</th>
<th>% of samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stable equilibrium</td>
<td>1.00</td>
<td>2.00</td>
<td>1.00</td>
<td>1.90</td>
<td>1.80</td>
<td>1.70</td>
<td>1.60</td>
<td>1.50</td>
<td></td>
</tr>
<tr>
<td>Single point</td>
<td>0.8442</td>
<td>1.4064</td>
<td>1.0456</td>
<td>1.9183</td>
<td>1.8144</td>
<td>1.7034</td>
<td>1.5918</td>
<td>1.4809</td>
<td>59.7</td>
</tr>
<tr>
<td></td>
<td>[0.2272]</td>
<td>[0.8303]</td>
<td>[0.1152]</td>
<td>[0.2152]</td>
<td>[0.2060]</td>
<td>[0.1969]</td>
<td>[0.1884]</td>
<td>[0.1942]</td>
<td></td>
</tr>
<tr>
<td>Two points</td>
<td>1.0867</td>
<td>2.3248</td>
<td>0.9655</td>
<td>1.8926</td>
<td>1.8081</td>
<td>1.7085</td>
<td>1.6107</td>
<td>1.5248</td>
<td>21.4</td>
</tr>
<tr>
<td></td>
<td>[0.1274]</td>
<td>[0.4140]</td>
<td>[0.1035]</td>
<td>[0.2364]</td>
<td>[0.2376]</td>
<td>[0.2280]</td>
<td>[0.2253]</td>
<td>[0.2249]</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>1.1099</td>
<td>2.3931</td>
<td>0.9546</td>
<td>1.9244</td>
<td>1.8176</td>
<td>1.7303</td>
<td>1.647</td>
<td>1.5425</td>
<td>18.9</td>
</tr>
<tr>
<td></td>
<td>[0.1377]</td>
<td>[0.4559]</td>
<td>[0.1135]</td>
<td>[0.2424]</td>
<td>[0.2327]</td>
<td>[0.2215]</td>
<td>[0.2252]</td>
<td>[0.2154]</td>
<td></td>
</tr>
<tr>
<td>Unstable equilibrium</td>
<td>1.00</td>
<td>4.00</td>
<td>1.00</td>
<td>1.90</td>
<td>1.80</td>
<td>1.70</td>
<td>1.60</td>
<td>1.50</td>
<td></td>
</tr>
<tr>
<td>Single point</td>
<td>0.6333</td>
<td>1.8721</td>
<td>1.3617</td>
<td>1.978</td>
<td>1.8207</td>
<td>1.6494</td>
<td>1.4902</td>
<td>1.1280</td>
<td>4.0</td>
</tr>
<tr>
<td></td>
<td>[0.3805]</td>
<td>[2.1715]</td>
<td>[0.3812]</td>
<td>[0.2608]</td>
<td>[0.2560]</td>
<td>[0.2672]</td>
<td>[0.2761]</td>
<td>[0.4396]</td>
<td></td>
</tr>
<tr>
<td>Two points</td>
<td>0.7569</td>
<td>2.5671</td>
<td>1.2276</td>
<td>1.9511</td>
<td>1.8447</td>
<td>1.7085</td>
<td>1.5372</td>
<td>1.2575</td>
<td>52.7</td>
</tr>
<tr>
<td></td>
<td>[0.2595]</td>
<td>[1.4639]</td>
<td>[0.2556]</td>
<td>[0.2882]</td>
<td>[0.2829]</td>
<td>[0.2600]</td>
<td>[0.2545]</td>
<td>[0.3334]</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>0.7634</td>
<td>2.586</td>
<td>1.2041</td>
<td>1.9696</td>
<td>1.8466</td>
<td>1.7172</td>
<td>1.5436</td>
<td>1.2598</td>
<td>43.3</td>
</tr>
<tr>
<td></td>
<td>[0.2509]</td>
<td>[1.4344]</td>
<td>[0.2357]</td>
<td>[0.2869]</td>
<td>[0.2732]</td>
<td>[0.2627]</td>
<td>[0.2617]</td>
<td>[0.3356]</td>
<td></td>
</tr>
</tbody>
</table>

Notes: Averages and root-mean-square errors (in square brackets) are computed over 1,000 Monte-Carlo simulations. Other indicates that the sequence of estimates does not converge to a single point, nor to a two-points cycle. The tolerance level to determine convergence is $10^{-3}$.
Figure 3.3: Stable equilibrium: estimation of $\theta_{RN}$, odd and even iterations
Figure 3.4: Unstable equilibrium: estimation of $\theta_{RN}$, odd and even iterations.
potential explanation could be that the criterion used to assess convergence may be too restrictive: even if the supremum norm of the difference between two vectors of estimates may fail to converge, some elements in these vectors could still be close to the true values of the corresponding parameters.

In order to better understand the behaviour of the estimates that converge to two-points cycles, I compare the evolution of the even and the odd iterations of the K-step estimator for the strategic interaction parameter. The plots for this specific subsample are available in Figures 3.3 and 3.4 for the stable and the unstable equilibrium respectively. Of course, the difference between the properties of the even and odd estimates is more obvious in the unstable case. Interestingly, this difference seems to be decreasing (very) slowly as the number of iterations increases.

To sum up, even if the K-step estimator may fail to converge to a single point for many or even most samples, it seems that, on average, it can still converge to a stable cycle with nice finite sample properties. In particular, this stable cycle may lead to better properties compared to the modified K-step estimators studied here.

3.6 Concluding remarks

Probably one of the most striking results from the Monte-Carlo experiments is that, even if the K-step estimator rarely converges and even if it performs worse when stability of the equilibrium fails than when it holds, it is hard to find a better estimator among the alternatives considered here, i.e. one that would work better both in the stable and in the unstable equilibria.

This conclusion contrasts with what one may understand from Egesdal, Lai, and Su (2013)’s negative results about the K-step estimator. The main reason behind this disparity is that I consider a fixed number of iterations, whether or not convergence is reached. This specific way to approach the problem makes it obvious that even if most sequences of estimates do not converge on their own, they can still converge to a stable cycle on average.

This being said, at least one interesting issue associated with the K-step estimator deserves to be addressed in future work. In fact, even if, for K large enough, any estimates from the generated sequence have, on average, better finite sample properties than the two-step estimator, some of these estimates are preferable than others. In other words, some estimates in the cycle should be closer to the truth. At this point, no selection mechanism has been proposed regarding which estimates should be picked when the sequence oscillates.

3.7 Appendix

3.7.1 Proof of Lemma 4

A standard first-order expansion of the first-order conditions around \( \theta^0 \) gives us:

\[
M^{1/2} (\hat{\theta} - \theta^0) = -E \left[ \frac{\partial^2 \ln [\Psi (y|x, \theta^0, P^0)]}{\partial \theta \partial \theta'} \right]^{-1} M^{-1/2} \sum_{m=1}^{M} \frac{\partial \ln [\Psi (y|x, \theta^0, \hat{P})]}{\partial \theta} + o_p (1). \tag{3.20}
\]

Furthermore, a first-order expansion of \( \partial \ln \left[ \Psi (y|x, \theta^0, \hat{P}) \right] / \partial \theta \) around \( P^0 \) allows us to write:
\[ M^{1/2} (\hat{\theta}\text{PML} - \theta^0) = -E \left[ \frac{\partial^2 \ln \left[ \Psi \left( y | x, \theta^0, P^0 \right) \right]}{\partial \theta \partial \theta'} \right]^{-1} \left\{ M^{-1/2} \sum_{m=1}^{M} \frac{\partial \ln \left[ \Psi \left( y | x, \theta^0, \hat{P} \right) \right]}{\partial \theta} \right\} + E \left[ \frac{\partial^2 \ln \left[ \Psi \left( y | x, \theta^0, P^0 \right) \right]}{\partial \theta \partial \theta'} \right] M^{-1/2} \left( \hat{P} - P^0 \right) + o_p(1). \] (3.21)

Notice that:

\[
M^{-1} \sum_{m=1}^{M} \frac{\partial \ln \left[ \Psi \left( y_m | x_m, \theta^0, \hat{P} \right) \right]}{\partial \theta} = \sum_{y \in Y} \sum_{x \in X} \frac{\partial \ln \left[ \Psi \left( y | x, \theta^0, \hat{P} \right) \right]}{\partial \theta} \hat{P} (y|x) \] (3.22)

and

\[
E \left[ \frac{\partial \ln \left[ \Psi \left( y | x, \theta^0, \hat{P} \right) \right]}{\partial \theta} \right] = \sum_{y \in Y} \sum_{x \in X} \frac{\partial \ln \left[ \Psi \left( y | x, \theta^0, \hat{P} \right) \right]}{\partial \theta} P^0 (y|x) = 0. \] (3.23)

It follows that:

\[
M^{-1/2} \sum_{m=1}^{M} \frac{\partial \ln \left[ \Psi \left( y_m | x_m, \theta^0, \hat{P} \right) \right]}{\partial \theta} = \sum_{y \in Y} \sum_{x \in X} \frac{\partial \ln \left[ \Psi \left( y | x, \theta^0, \hat{P} \right) \right]}{\partial \theta} M^{1/2} \left( \hat{P} (y|x) - P^0 (y|x) \right)
= \Psi'_\theta \text{diag} \{P^0\}^{-1} M^{1/2} \left( \hat{P} - P^0 \right). \] (3.24)

Moreover, using Bartlett’s equalities and writing expectations as weighted sums with weights from \( P^0 \), one gets:

\[-E \left[ \frac{\partial^2 \ln \left[ \Psi \left( y | x, \theta^0, P^0 \right) \right]}{\partial \theta \partial \theta'} \right] = \Psi'_\theta \text{diag} \{P^0\}^{-1} \Psi_\theta \] (3.25)

and

\[
E \left[ \frac{\partial^2 \ln \left[ \Psi \left( y | x, \theta^0, P^0 \right) \right]}{\partial \theta \partial \theta'} \right] = -\Psi'_\theta \text{diag} \{P^0\}^{-1} \Psi_P. \] (3.26)

Equation (3.21) can therefore be written as:

\[
M^{1/2} (\hat{\theta}\text{PML} - \theta^0) = \left[ \Psi'_\theta \text{diag} \{P^0\}^{-1} \Psi_\theta \right]^{-1} \left\{ \Psi'_\theta \text{diag} \{P^0\}^{-1} M^{1/2} \left( \hat{P} - P^0 \right) \right\} - \Psi'_\theta \text{diag} \{P^0\}^{-1} \Psi_P M^{1/2} \left( \hat{P} - P^0 \right) + o_p(1). \] (3.27)

Finally, using the definition of \( G \) given in the Lemma, (3.27) becomes:

\[
M^{1/2} (\hat{\theta}\text{PML} - \theta^0) = G \left[I - \Psi_P \right] M^{1/2} \left( \hat{P} - P^0 \right). \] (3.28)
The consistency of \( \hat{\theta}^{PML} \) follows from \( \hat{P} \stackrel{p}{\rightarrow} P^0 \). The asymptotic normality comes from \( M^{1/2} \left( \hat{P} - P^0 \right) \overset{d}{\rightarrow} N \left( 0, \Sigma \right) \).

### 3.7.2 Proof of Lemma 5

The consistency of \( \left\{ \hat{\theta}_K, \hat{P}_K : K \geq 1 \right\} \) directly follows from the consistency of the first-stage estimator, i.e. \( \hat{P}_0 \overset{p}{\rightarrow} P^0 \), and the consistency of \( \hat{\theta}^{PML} \) stated in Lemma 4. In fact, by definition, \( \hat{\theta}_1 = \hat{\theta}^{PML} \) and:

\[
\hat{\theta}_1 = \arg\max_{\theta} M^{-1} \sum_{m=1}^M \ln \left[ \Psi \left( y_m | x_m, \theta, \hat{P}_0 \right) \right] \overset{p}{\rightarrow} \arg\max_{\theta} E \left[ \ln \left[ \Psi \left( y | x, \theta, P^0 \right) \right] \right] = \theta^0. \tag{3.29}
\]

Applying the argument recursively for \( K \geq 1 \) proves parts (i) and (ii).

The asymptotic distributions of \( \left\{ \hat{\theta}_K, \hat{P}_K : K \geq 1 \right\} \) are derived in Aguirregabiria (2004, Proposition 1, p. 338). A first-order expansion of the first-order conditions defining \( \hat{\theta}_K \) around \( \theta^0 \) gives:

\[
M^{1/2} \left( \hat{\theta}_K - \theta^0 \right) = -E \left[ \frac{\partial^2 \ln \left[ \Psi \left( y | x, \theta^0, P^0 \right) \right]}{\partial \theta \partial \theta'} \right]^{-1} \left( M^{-1/2} \sum_{m=1}^M \frac{\partial \ln \left[ \Psi \left( y | x, \theta^0, \hat{P}_{K-1} \right) \right]}{\partial \theta} \right) + M^{1/2} \left( \hat{P}_{K-1} - P^0 \right) + o_p(1) \tag{3.30}
\]

which, using the notation introduced in Lemma 4, can be written as:

\[
M^{1/2} \left( \hat{\theta}_K - \theta^0 \right) = GM^{1/2} \left( \hat{P}_0 - P^0 \right) - G \Psi_P M^{1/2} \left( \hat{P}_{K-1} - P^0 \right) + o_p(1). \tag{3.31}
\]

Moreover, a standard first-order expansion of \( \hat{P}_K = \Psi \left( \hat{P}_{K-1}, \hat{\theta}_K \right) \) around \( P^0 \) and \( \theta^0 \) gives:

\[
M^{1/2} \left( \hat{P}_K - P^0 \right) = \Psi_P M^{1/2} \left( \hat{P}_{K-1} - P^0 \right) + \Psi_\theta M^{1/2} \left( \hat{\theta}_K - \theta^0 \right) + o_p(1). \tag{3.32}
\]

By substituting (3.31) in (3.32), we get:

\[
M^{1/2} \left( \hat{P}_K - P^0 \right) = \left[ I - \Psi_\theta G \right] \Psi_P M^{1/2} \left( \hat{P}_{K-1} - P^0 \right) + \Psi_\theta G M^{1/2} \left( \hat{P}_0 - P^0 \right) + o_p(1) \tag{3.33}
\]

which can be solved backwards to get \( M^{1/2} \left( \hat{P}_K - P^0 \right) = A_K M^{1/2} \left( \hat{P}_0 - P^0 \right) + o_p(1) \) after using the definition of \( A_K \). Furthermore, evaluating (3.33) at \( K - 1 \), substituting in (3.31) and using the definition of \( B_K \), we get \( M^{1/2} \left( \hat{\theta}_K - \theta^0 \right) = B_K M^{1/2} \left( \hat{P}_0 - P^0 \right) + o_p(1) \). Parts (iii) and (iv) in the Lemma simply follow from \( M^{1/2} \left( \hat{P}_0 - P^0 \right) \overset{d}{\rightarrow} N \left( 0, \Sigma \right) \).

### 3.7.3 Two-step semiparametric estimation of the model

In this Appendix, I describe how to apply the two-step semiparametric estimator to the dynamic game of market entry and exit used in the Monte-Carlo experiments. In order to focus on the estimation of
Chapter 3. Bias reduction in dynamic discrete games

the CCPs and the structural parameters, I assume that the researcher knows the distributions of the observable and the unobservable state variables.

The first step consists of a nonparametric estimation of the CCPs. For the Nadaraya-Watson kernel estimator and for \( x_1 = [S_1, y_0] \) the estimated CCP of firm \( i \) being active in period 1 is:

\[
P_i(1|x_1) = \frac{\sum_{m=1}^{M} y_{im1} K\left( \frac{x_{m1} - x_1}{h} \right)}{\sum_{m=1}^{M} K\left( \frac{x_{m1} - x_1}{h} \right)},
\]

(3.34)
The kernel I use is a product of Gaussian kernels:

\[
K\left( \frac{x_{m1} - x_1}{h} \right) = \frac{1}{(2\pi)^{d}} \exp \left\{ -\frac{1}{2} \left[ \left( \frac{S_{m1} - S_1}{h_S} \right)^2 + \left( \frac{y_{1m0} - y_{10}}{h_1} \right)^2 + \left( \frac{y_{2m0} - y_{20}}{h_2} \right)^2 + \left( \frac{y_{3m0} - y_{30}}{h_3} \right)^2 + \left( \frac{y_{4m0} - y_{40}}{h_4} \right)^2 + \left( \frac{y_{5m0} - y_{50}}{h_5} \right)^2 \right] \right\}.
\]

(3.35)

I select the bandwidths (one for each conditioning variable) by using Silverman (1986)'s rule of thumb. In this model, \( x_1 \) is a \( d = 6 \)-dimensional vector of conditioning variables. Assuming that the marginal density of each of these variable is twice continuously differentiable, the \( l \)-th bandwidth corresponding to the \( l \)-th element of \( x_1 \) is given by \( h_l = 0.9330 \hat{\sigma}_{x_{1l}} M^{-1/10} \) where \( \hat{\sigma}_{x_{1l}} \) is the estimated standard deviation of the corresponding \( l \)-th variable in \( x_1 \). For the frequency count estimator, the kernel is simply replaced by the indicator function \( \mathbb{1}\{x_{m1} = x_1\} \).

The second-step estimators considered in the Monte-Carlo are the PML and the GMM respectively defined as:

\[
\hat{\theta}^{PML} = \arg \max_{\theta \in \Theta} (N M)^{-1} \sum_{i=1}^{N} \sum_{m=1}^{M} \ln \left[ \Psi \left( y_{im1}|x_{m1}, \theta, \hat{P} \right) \right]
\]

(3.36)

and

\[
\hat{\theta}^{GMM} = \arg \max_{\theta \in \Theta} (N M)^{-1} \left( Z' \xi (\theta) \right)' \left( W(\theta)^{-1} (Z' \xi (\theta)) \right)
\]

(3.37)

where \( W(\theta) = Z' \xi (\theta) \xi (\theta)' Z \) and, for \( z_{im} = [x_{im1}, S_{im} \times x_{im1}]' \),

\[
Z = \begin{bmatrix} z_{11}' \\ \vdots \\ z_{NM}' \end{bmatrix} \quad \text{and} \quad \xi (\theta) = \begin{bmatrix} y_{111} - \Psi \left( y_{111}|x_{111}, \theta, \hat{P} \right) \\ \vdots \\ y_{NM1} - \Psi \left( y_{NM1}|x_{M1}, \theta, \hat{P} \right) \end{bmatrix}.
\]

In order to evaluate these objective functions, one has to compute \( u^p_i (y, x_{mt}) \). Let \( \alpha^p_i (y, x_{mt}) \) be such that \( u^p_i (y, x_{mt}) = \alpha^p_i (y, x_{mt}) \theta \) for \( u^p_i (\cdot) \) defined as:

\[
u^p_i (y, x_{mt}) = \sum_{y_{-i} \in y^{N-1}} \prod_{j \neq i} P(y_j|x_{mt}) u_i(y, y_{-i}, x_{mt}).
\]

(3.38)

Furthermore, let \( \alpha^p_i (x_{mt+1}|x_{mt}, y) \) be defined by:
\[ \gamma_i^P (x_{mt+1} | x_{mt}, y) = \sum_{y_{-i} \in Y^{x_{mt}}} \prod_{j \neq i} P (y_j | x_{mt}) \gamma_i (x_{mt+1} | x_{mt}, y_{-i}). \] (3.39)

Then, for \( y \in Y \),

\[ v_i^P (y, x_{mt}) = \alpha_i^P (y, x_{mt}) \theta + \beta \sum_{x_{mt+1} \in \mathcal{X}} \gamma_i^P (x_{mt+1} | x_{mt}, y) \bar{V}_i^P (x_{mt+1}) \] (3.40)

which requires the evaluation of \( \bar{V}_i^P (x_{mt+1}) \), the expectation of the integrated value function \( \bar{V}_i (x_{mt+1}) = \int V_i (x_{mt+1}, \epsilon_{imt+1}) dG (\epsilon_{imt+1}) \). Let \( e_i^P (y, x_{mt}) \) be the expected unobservable given the state and the optimality of the decision. In vector notation, write \( \bar{V}_i^P \) (a \( \dim (\mathcal{X}) \)-dimensional vector with each element corresponding to a different state) as:

\[
\bar{V}_i^P = \sum_{y \in Y} p_i (y) \ast \left[ \alpha_i^P (y) \theta + e_i^P (y) + \beta \Gamma_i^P (y) \bar{V}_i^P \right] \\
= \left[ I - \beta \sum_{y \in Y} p_i (y) \ast \Gamma_i^P (y) \right]^{-1} \sum_{y \in Y} p_i (y) \ast \left[ \alpha_i^P (y) \theta + e_i^P (y) \right] \tag{3.41}
\]

where \( I \) is the identity matrix. \( p_i (y) \), \( \alpha_i^P (y) \) and \( e_i^P (y) \) are all \( \dim (\mathcal{X}) \)-dimensional vectors with each element corresponding to a different state. Similarly, \( \Gamma_i^P (y) \) is a \( \dim (\mathcal{X}) \times \dim (\mathcal{X}) \) matrix collecting the \( \gamma_i^P \)'s associated with each state.

Notice that the extreme value type 1 assumption comes with a closed-form expression for \( e_i^P (y, x_{mt}) \) in binary choice models: \( e_i^P (y, x_{mt}) = \zeta - \ln [p_i (y | x_{mt})] \), where \( \zeta \) denotes Euler’s constant.

By considering the element of \( \bar{V}_i^P \) corresponding to the observable state \( x_{mt} \), one computes \( v_i^P (y, x_{mt}) \) for \( y \in Y \) and evaluates the \( \Psi (\cdot) \) mapping needed in the pseudo-likelihood function.
Bibliography


BIBLIOGRAPHY


