Essays in Applied Econometrics

by

Jacob Schwartz

BSc, University of Victoria, 2009

MA, University of Victoria, 2011

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

Doctor of Philosophy

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Economics)

The University of British Columbia

(Vancouver)

September 2018

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the dissertation entitled: Essays in Applied Econometrics, submitted by Jacob Schwartz in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Economics.

Examining Committee:

David Green, Economics	Co-Supervisor
Kevin Song, Economics	Co-Supervisor
Vadim Marmer, Economics	Supervisory Committee Member
Florian Hoffmann, Economics	Supervisory Committee Member
Thomas Davidoff, Sauder School of Business	University Examiner
Kevin Milligan, Economics	University Examiner

Abstract

Chapter 1 develops an empirical two-sided matching model with endogenous pre-investment. The model can be used to measure the impact of frictions in labour markets using a single cross-section of matched employer-employee data. The observed matching of workers to firms is the outcome of a discrete, two-sided matching process where firms with heterogeneous preferences over education sequentially choose workers according to an index correlated with worker preferences over firms. The distribution of education arises in equilibrium from a Bayesian game: workers, knowing the distribution of worker and firm types, invest in education prior to the matching process. I propose an inference procedure combining discrete choice methods with simulation. Counterfactual analysis using Canadian data shows that changes in matching frictions can lead to economically significant equilibrium changes in both inequality and the probability of investing in higher education. These effects are more pronounced when worker and firm attributes are complements in the match surplus function.

In many economic settings, agents behave similarly because they share information with one another. Information-sharing relations among agents can be modeled as a network, and the strategic interactions among them as a game on a network. Chapter 2, coauthored with Kyungchul Song and Nathan Canen, develops a tractable empirical model of social interactions where each agent - without seeing the full information network - shares information with their neighbors and best responds to the other players based on simple beliefs about their strategies. We provide conditions on the information networks and beliefs of agents such that their best responses exhibit economically intuitive features and desirable external validity relative to equilibrium models of social interaction. Moreover, the setup

admits asymptotic inference without requiring that the researcher observes all the players in the game, nor that the they know precisely the sampling process.

Chapter 3 discusses how discrete distributions of unobserved heterogeneity can be identified using information on sample attrition. Although attrition is often seen as a source of selection problems, we argue that it can also be used to solve selection problems - even in the absence of covariates or panel data.

Lay Summary

The quality of information that people have can affect the decisions they make prior to entering a market. For example, education may be less valuable to workers in labour markets with poor information. Chapter 1 develops tools for measuring these frictions in markets where agents' decisions affect one anothers' hiring outcomes. I apply the methodology to study labour markets in Canada.

Chapter 2 develops an approach for studying how information-sharing agents affect one another in social networks. The methods exhibit good external validity and do not require strong assumptions about how the data were sampled. We apply the methods to study public goods provision in Colombia.

Chapter 3 argues that information on whether or not agents are observed to leave a researcher's data set can provide the researcher with useful information with which to learn about the agents' unobserved attributes.

Preface

Chapter 1 of this thesis "*Schooling Choice, Labour Market Matching, and Wages*" is my original work. The empirical section of this chapter uses data from Statistics Canada's Workplace Employee Survey (WES).

The second chapter, "*Estimating Local Interactions Among Many Agents Who Observe Their Neighbors*", is an unpublished working paper that I co-authored with Nathan Canen and Kyungchul Song. The authors contributed equally to the project overall.

In Chapter 3, "*Identification Using Attrition*, Section 3.2 up to Assumption 3.2.1 of Section 3.2.2 is drawn from a working paper I co-authored with Hugo Jales entitled "*Type-Targeted Treatment Effects and Type Revelation*", of which I am the lead author and principal contributor. The remainder of Chapter 3 is my original work that is new for the thesis.

Any views expressed in this thesis are mine alone and do not reflect the views of Statistics Canada or the Government of Canada.

Table of Contents

Abstract . . . . **iii**

Lay Summary . . . . **v**

Preface . . . . **vi**

Table of Contents . . . . **vii**

List of Tables . . . . **x**

List of Figures . . . . **xii**

Acknowledgments . . . . **xiii**

Introduction **1**

1 Schooling Choice, Labour Market Matching, and Wages **3**

1.1 Introduction 3

1.1.1 Background and Related Literature 7

1.2 The Labour Market As a Two-Sided Matching Market 10

1.2.1 Baseline model 10

1.2.2 Frictional Matching Model with Worker Investments . . . 14

1.2.3 Some Implications of Frictional Matching Model 17

1.3 Econometric Inference 19

1.3.1 Two-Stage Inference Accommodating Cross-Sectional Dependence of Observed Matching 22

1.3.2 First-Stage Estimation of θ 27
1.3.3 Matching Probabilities 32
1.4 Analysis of a Labour Matching Market in Canada 33
1.4.1 Background 33
1.4.2 Data 33
1.4.3 Model Estimates and Counterfactuals 35
1.5 Conclusion 37

2 Estimating Local Interactions Among Many Agents Who Observe Their Neighbors 39
2.1 Introduction 39
2.2 Strategic Interactions with Information Sharing 45
2.2.1 A Model of Interactions with Information Sharing 45
2.2.2 Predictions from Rationality 48
2.2.3 Belief Projection and Best Linear Responses 52
2.2.4 The External Validity of Network Externality 58
2.3 Econometric Inference 62
2.3.1 General Overview 62
2.3.2 Asymptotic Theory 77
2.4 A Monte Carlo Simulation Study 79
2.5 Empirical Application: State Presence across Municipalities . . . 81
2.5.1 Motivation and Background 81
2.5.2 Empirical Set-up 85
2.5.3 Model Specification 86
2.5.4 Results 87
2.6 Conclusion 89

3 Identification Using Attrition 91
3.1 Introduction 91
3.2 Identification of the Type-Targeted Treatment Effect 94
3.2.1 General Setup 96
3.2.2 Identification Under One Period of Attrition 96
3.2.3 Identification Using Multiple Periods of Attrition 101

3.3 Estimation and Inference 105
3.3.1 Small Simulation Exercise 106
3.4 Using Attrition to Correct for Selection on Unobservables 106
3.5 Conclusion 108

Conclusions 109

Bibliography 111

A Supporting Materials 121
A.1 Appendix to Chapter 1 121
A.1.1 Equilibrium of Investment and Matching Game 121
A.1.2 Additional Mathematical Results 125
A.1.3 A Monte Carlo Simulation Study 133
A.1.4 Empirical Section 137
A.1.5 Estimation 137
A.1.6 Estimation Results 138
A.1.7 Model Counterfactuals 140
A.2 Appendix to Chapter 2 149

List of Tables

Table 2.1 The Characteristics of the Payoff Graphs 60

Table 2.2 The Degree Characteristics of the Graphs Used in the Simulation Study 80

Table 2.3 The Empirical Coverage Probability of the Confidence Intervals for β_0 at 95% Nominal Level. 81

Table 2.4 The Average Length of Confidence Intervals for β_0 at 95% Nominal Level. 82

Table 2.5 The Empirical Coverage Probability of Confidence Intervals for $a'\rho_0$ at 95% Nominal Level. 83

Table 2.6 The Average Lengths of Confidence Intervals for $a'\rho_0$ at 95% Nominal Level. 84

Table 2.7 State Presence and Networks Effects across Colombian Municipalities 88

Table 3.1 The Empirical Coverage Probability and Average Length of Confidence Intervals for Attrition Process Parameters at 95% Nominal Level 107

Table A.1 The Empirical Coverage Probability of Asymptotic Confidence Intervals for $a'\theta_0$ at 95% Nominal Level When β_0 is Known. . . 135

Table A.2 Average Length of Confidence Intervals for $a'\theta_0$ at 95% Nominal Level When β_0 is Known. 136

Table A.3 Matching Technology In Canadian Manufacturing and Finance Industries, 1995-2005 140

Table A.4 Estimation of Worker and Firm Preferences in Canadian Manufacturing Industry, 1999-2005 141
Table A.5 Estimation of Worker and Firm Preferences in Canadian Finance Industry, 1999-2005 142
Table A.6 Counterfactual Estimated Probabilities of Investing in High Education, Manufacturing Industry 143
Table A.7 Counterfactual Estimated Probabilities of Investing in High Education, Finance Industry 144
Table A.8 Counterfactual Estimated Probabilities of Investing in High Education, Manufacturing Industry 145
Table A.9 Counterfactual Estimated Probabilities of Investing in High Education, Finance Industry 146
Table A.10 Matching Technology Counterfactuals, Simulated Gini Coefficient, Manufacturing Industry 147
Table A.11 Matching Technology Counterfactuals, Gini Coefficient, Finance Industry 148

List of Figures

Figure 1.1 Education and Wage Inequality Under Specification 1 20
Figure 1.2 Education and Wage Inequality Under Specification 2 20
Figure 1.3 Supply of Highly Educated Workers and Education Wage Premia 21

Figure 2.1 Network Externality Comparison Between Equilibrium and Behavioral Models: Erdös-Rényi Graphs 61
Figure 2.2 Network Externality Comparison Between Equilibrium and Behavioral Models: Barabási-Albert Graphs 61
Figure 2.3 Degree Distribution of G_P 85
Figure 2.4 Average Network Externality from being a Department Capital 89

Figure A.1 Wage Inequality in Canada's Workplace-Employee Survey: 99-50 Difference in Quantile of Log Hourly Wages 139
Figure A.2 Income Inequality in Canada: Gini coefficients, 1990-2015 . . 139

Acknowledgments

This thesis could not have been written without the expert advice, patience, and encouragement of each of my main committee members: Kevin Song, David Green, and Vadim Marmer. My work also benefited greatly from conversations with Florian Hoffmann. I could not have asked for a better committee. Special thanks are also due to Hiro Kasahara, Paul Schrimpf, Michael Peters, and Joris Pinkse.

I would also like to thank my family, friends, and colleagues: in particular Anujit Chakraborty, Hugo Jales, Jon Graves, Pierluca Pannella, Oscar Becerra, Ruoying Wang, Tom Cornwall, Nathan Canen, Denis Kojevnikov, Tom Thorn, and Alison Watt. Lastly, I thank David Giles and Judy Clarke for their support during my studies in Victoria, and for introducing me to the field of econometrics.

Introduction

This thesis is mainly concerned with developing methodology for inference in models with many simultaneously interacting agents. The first two chapters develop tools to help researchers better exploit the rich information contained in matched employer-employee data and network data: *"Schooling Choice, Labour Market Matching, and Wages"* and *"Estimating Local Interactions Among Many Agents Who Observe Their Neighbors"*. The third chapter, *"Identification Using Attrition"*, provides a general case for the usefulness of data on sample attrition for learning forms of unobserved heterogeneity.

Understanding how and why workers and firms match with one another is important for understanding the sources of wage inequality. Studying matching in labour markets presents a particular challenge in settings where the decisions of individual job seekers affect one another's hiring outcomes, thereby creating strong cross-sectional dependence. The first chapter of this thesis develops a methodology to address this challenge. The method can be used for studying the role that matching frictions plays in shaping education and wage patterns using cross-sections of matched employer-employee data. In this paper, the observed matching is the outcome of a two-sided matching process where firms with hetereogeneous preference rankings over skill sequentially choose workers according to an index correlated with worker preference rankings over firms - the index being a simple and flexible way of modeling matching frictions. The distribution of worker skill in my model arises in equilibrium from a Bayesian pre-match investment game: workers, knowing the distribution of worker and firm types, invest in education prior to the matching process. In simulations of my model, I demonstrate how improvements in the matching technology can lead to an increase in college wage premia without leading to a significant increase in the supply of highly educated workers. A dynamic extension to the static structural framework that I develop in this chapter may thus offer a novel explanation to a longstanding puzzle in the empirical literature. I estimate my model using data from Canada's Workplace Employee Survey for the years 1999-2005. My study of counterfactual distributions of education and wages reveals that changes in the matching technology can lead to economically significant changes in both wage inequality and the probability of investing in higher education, and that these effects are more pronounced when worker and firm attributes are complements in the match surplus function.

The increased availability of large network datasets has provided economists with detailed information to study social interactions. In practice, inference using such network data can present a serious challenge to the researcher, since the sampling scheme used to collect such data is often both non-random and not known to the researcher with precision. In work with Nathan Canen and Kyungchul Song, *"Estimating Local Interactions Among Many Agents Who Observe Their Neighbors"*, we develop a linear model of interactions convenient for inference when the econometrician sees many (possibly non-random) samples of local interactions between agents. The inference procedure we propose exhibits asymptotic validity regardless of the sampling scheme used, provided it satisfy certain weak conditions. A key feature of our model is that correlation between the actions of neighbors can emerge partly due to information sharing between them when such information is unobserved by the econometrician. In our setup, we show how the presence of such information sharing can be detected using a test based on the cross-sectional correlation of the residuals.

Sample attrition is typically considered a nuisance that prevents inference on the underlying population of interest. Chapter 3 takes an alternative view, arguing that when the attrition decision depends on unobserved heterogeneity, attrition patterns over time can be used to identify the distribution of unobserved heterogeneity - even in the absence of covariates or panel data on the outcome of interest.

Chapter 1

Schooling Choice, Labour Market Matching, and Wages

1.1 Introduction

Since the 1980s, economists have attributed rising wage inequality to a number of sources. One possible source of such inequality is positive assortative matching between workers and firms - the tendency for the quality of workers and firms who match with one another to be positively correlated.[1] Understanding how and why certain workers and firms match with one another is thus a key to understanding this assortativity channel and, ultimately, a source of wage inequality. Unfortunately, studying matching in labour markets presents a serious challenge when the decisions of individual job seekers affect each other's hiring outcomes. This chapter develops a methodology to address this challenge. In particular, I show how cross-sections of matched employer-employee data can be used to study the role that a labour market matching technology plays in shaping the equilibrium distributions of education and wages. A key result is that - even in the absence of complementarities between worker and firm types in the match production function - the model

[1] Recent empirical papers examining the role of sorting on wage inequality include Card et al. 35, Barth et al. [13], and Kantenga and Law [75].

can capture assortative matching between workers and firms.[2]

A general overview of the labour market in my model is as follows. Agents from one side of the market sequentially choose agents from the other side according to their preferences. Preference rankings of the choosers depend on a preference parameter, along with the capital of both types of agents. The order in which the choosers pick depends on the chooser's capital and a matching technology parameter. Before matching, the agents who will be chosen are allowed to simultaneously decide their capital given the distribution of the chooser's capital and the underlying parameters of the economy (including the frictions).

The structural approach I develop allows me to examine counterfactual distributions of education and wages under different matching technologies and preferences. Although the model of the labour market I develop in this chapter is fundamentally a static one, an extension to my approach may help explain a long-standing empirical puzzle – namely, how college wage premia can increase without an associated increase in the supply of highly educated workers. I show that this pattern is captured by simulating a special case of my model in which matching frictions become less severe over time.

This chapter contributes to the econometric literature concerned with inference in two-sided matching models. I propose a two-stage approach for inference on the agents' preferences and the matching technology. In the first stage, I fix the matching technology and construct confidence regions for the preference parameter by estimating the Bayesian game associated with the workers' pre-match investment in education decision. I show that this problem can be cast in a discrete choice framework that is tractably estimable using maximum likelihood when the workers' educational decision takes one of two values (college, or no college). In the second stage, I construct confidence intervals for the matching technology using a simulation-based inference approach. In the first stage, the presence of the matching function in workers' expected utility function makes estimating workers' equilibrium expectations highly non-trivial. Nevertheless, under reasonable assumptions, I show that workers' equilibrium expectations can be written in a

[2] The value of a match between any type of worker, h, and any type of firm, k, can be represented using a positive, increasing function, $f(h,k)$. We say that the types are complements in f when the marginal product of an h type is higher when matched with a higher k type (and vice versa).

closed form suitable for estimation. The second-stage inference on the matching technology uses the following insight: once the matching process is specified, the finite sample distribution of the observed matching is known up to a parameter.[3] I construct a test statistic that measures the distance between the observed joint distribution of worker education and matched firm capital to simulated counterparts. A confidence interval for the matching technology can then be constructed by inverting the test.

A unique feature of such a model is that worker and firm types need not be complements in the match production function for sorting between workers and firms to occur. This chapter builds on the fundamental insights of Becker [15] and Gale and Shapley [56] to highlight the fact that, under certain circumstances, a meaningful notion of sorting can be captured in an empirical model with additive worker and firm effects. In his seminal 1973 paper on the marriage market, Becker argued the following: when the match production function is supermodular and utility is transferable between matched agents, high types can outbid low types for the best partners, leading to an equilibrium with positive assortative matching.[4] The reason why positive assortative matching can occur in my model even when such complementarities are absent is answered in Becker's same 1973 paper. Becker notes that sorting can arise in an non-transferable utility (NTU) framework when the payoffs of the agents on both sides of the market are monotonic in the other agent's type. To explain why this is so, Becker invokes the logic of pairwise stability (Gale and Shapley, 1962).[5] For example, consider an economy with four agents where a low-type firm is paired with a high-type worker and vice-versa. Such a matching is unstable when high types are preferred, because both high-types will agree to abandon their low-type partners for one another. In a special case of my model where preferences are indeed monotonic, sorting - and some inequality - may emerge. In this case, complementarities are not necessary for sorting but merely amplify the

[3] This idea of using a structural model to characterize the joint distribution of a discrete matching model that can then be used for inference on the model parameters builds from work-in-progress I am pursuing with Taehoon Kim, Kyungchul Song, and Yoon-Jae Whang. Although computationally intractable when the dimension of the parameter is large, this approach is attractive for inference on the matching technology parameter in the second stage of my approach.

[4] When f is differentiable, (strict) supermodularity is equivalent to $\partial^2 f(h,k)/\partial h \partial k > 0$.

[5] This insight comes to me by way of Chade et al. [37]'s excellent review of the search and matching literature.

effects of sorting, since interactions between worker and firm types in the wage function lead to more wage dispersion than when such interactions are absent.

I estimate my model using matched employer-employee data on the finance and manufacturing industries using Canada's Workplace-Employee Survey (WES). I find that frictions in the matching technology rose in the middle of the sample period, a time corresponding with relatively stable wage inequality. My model counterfactuals imply that the matching technology frictions matter: for example, in the 2001 finance industry, the effect of eliminating frictions causes a roughly 8% increase in the equilibrium probability of high education in the complementarities case and only a 3% increase in the specification without.

The empirical findings highlight the importance of production complementarities to wage inequality. The model-predicted level of wage inequality is much higher (and more reasonable for Canada) in the case that complementarities are present - for example, the predicted Gini coefficient for the 2002 finance industry is 0.2542 for the case of complementarities, while it is only 0.1655 in the additive case. The effect of information frictions on wage inequality are complicated by the presence of two competing effects: when frictions are lowered, the number of workers investing in education rises (a supply effect), but assortative matching between worker and firms also increases (a sorting effect). The sorting effect tends to increase inequality while the supply effect tends to decrease it. I argue that the role of the latter effect is relevant in the subgroups I study, where the equilibrium probability of investment in education is typically quite high. For example, among high skilled workers in the manufacturing industry in 2005, the Gini is 0.2220 and the investment in education is 77%. This rises to 0.2507 (education investment 76%) when information frictions are highest and 0.2452 (education investment 85%) when frictions are lowest. Overall, however, the result results suggest that changes in wage inequality over time are mostly driven by changes in exogenous worker and firm characteristics and preferences (including shifts in the underlying production technology) rather than changes in the matching technology.

The key features of the model are explored in Section 1.2.3. In addition to illustrating the model's main implications concerning the relationship between complementarities, frictions, and sorting, the section also suggests that my model may be able to shed some light on other empirical puzzles. In particular, we see how the model predicts that a fall in information frictions can lead to a dramatic rise in the education wage premium through sorting while at the same time, the same fall in information frictions leads to a much more modest increase in the supply of highly educated workers. Thus, changes in informational frictions may be a useful way to explain a puzzling empirical findings concerning the relationship between wage premia and educational attainment.[6]

1.1.1 Background and Related Literature

Since Abowd et al. [2] (AKM), the availability of matched employer-employee data has allowed researchers to study the role that unobserved worker and firm attributes play in driving wage variation over time. In AKM, the correlation of worker and firm fixed effects from wage regressions is taken to capture a notion of sorting. Although popular for investigating the wage structure, a burgeoning literature has criticised the viability of AKM for detecting sorting on unobservables. In particular, the additive structure of AKM implies that wages are monotone in firm type - an implication that is difficult to reconcile with equilibrium models of sorting with and without frictions (Eeckhout and Kircher [52], Lopes de Melo [87]).[7] For example, in Eeckhout and Kircher [52], a low-type worker can receive a lower wage at a high-type firm since the worker must implicitly compensate the high-type firm in equilibrium for forgoing the oppotunity to fill a vacant job with a higher-type worker.[8]

Search and matching models have emerged as the leading alternative to the AKM framework for studying sorting in labour markets.[9] In this literature, the standard matching technology is one that converts aggregates of vacancies and unemployed workers into matches. Although treating matching at the aggregate level simplifies the analysis considerably, any strategic interdependence that may be present in the matching process is assumed away (Chade et al. [37]). In many

[6] See Card and Lemieux [33].

[7] Gautier and Teulings [57] was an early empirical study that detected a concave relationship between wages and firm type.

[8] There are other reasons wages may be non-monotonic in firm type. In Postel-Vinay and Robin [99], workers may be willing willing to accept lower wages at higher type firms when they expect to receive higher wages in the future.

[9] Hagedorn et al. [65] Bagger and Lentz [11] , Lise et al. [85] and Lopes de Melo [87] all find evidence of positive sorting when an AKM approach finds negligible sorting.

labour markets, it is unrealistic to suppose that the decisions of individual workers do not affect the outcomes of other workers.[10] One contribution of this chapter is to develop and estimate a model that takes such strategic interdependence in the matching process seriously. In the equilibrium of my model, the probability that a worker matches to a given firm typically depends on the decisions of all the other agents in the economy.

Another key facet of search models is that workers direct their job search based on the wages that employers set for them. However, many recent studies of online job markets have found that it is relatively uncommon for positions to explicitly post wages.[11][12] This chapter differs from the traditional search literature by supposing that workers do not observe posted wages directly. Instead, workers know the underlying distributions of job characteristics and the matching process prior to simultaneously investing in education. In this sense, the worker's decision to invest in education is the channel by which workers are able to direct their search.

This chapter is related to a literature studying the role that match function complementarities play in driving sorting patterns. Since Becker [15]'s seminal study of the marriage market, it has been known that in two-sided matching markets with non-transferable utility (NTU), a sufficient condition for a stable matching to exhibit positive assortativity is that the agents' payoff be monotonic in their partners' type.[13] NTU arises naturally in my model from the assumption that wages for any matched pair are determined exogenously (in fact, by imposing Nash bargaining). In my approach, sorting can arise in the absence of direct interactions between worker and firm types in the (post-match) surplus function in the special case that workers and firms have monotone preferences over one another.[14]

[10]For example, a worker's decision to get a master's degree in finance will not only affect the likelihood that he gets a job at an investment bank, but also the likelihood that his competitors get the job.

[11]For example, Marinescu and Wolthoff [91] study the role of job titles in directing the search of workers report that only 20% of job the advertisements CareerBuilder.com report a wage.

[12]More generally, Cardoso et al. [36] is another recent paper that studies the impact of information frictions in search and matching models.

[13]Becker remarks this without explicitly invoke a notion of pairwise stability, such as that in Gale and Shapley [56] (Chade et al. [37]).

[14] A desire to accommodate complementarities does not necessarily require us to abandon the AKM framework entirely. Bonhomme et al. [23] incorporate worker-firm complementarities into a framework resembling AKM, while relaxing the exogenous mobility assumption. While providing

This chapter contributes to the literature concerned with inference in two-sided matching models.[15] A key feature of my setup is that the characteristics of agents on one side of the market are endogenous - in particular, arising in equilibrium from a pre-matching investment game. I show how, rather than making the empirical analysis intractable, accommodating such pre-matching investments provides the researcher useful information for inference.[16]

My framework supposes that each equilibrium gives rise to a single large matching between workers and firms.[17] Under familiar assumptions (e.g., iid and separable private information), I follow similar arguments to Aguirregabiria and Mira [6] to prove that an equilibrium exists. My setup, however, also allows me to provide sufficient conditions for equilibrium uniqueness.

This chapter is also part of the literature concerned with estimating cross-sectionally dependent observations. In my setup, the observed matching of workers to firms exhibits cross-sectional dependence of an unknown form due to the matching process. This means that asymptotic inference approaches that appeal to the the law of large numbers and central limit theorems will not work. The approach I pursue builds on an idea from work-in-progress that demonstrates how inference in structural matching models are possible when knowledge of the matching process can be used to characterize the joint distribution of the observed matching. This chapter shows how such a simulation-based inference approach, cumbersome when the dimension of the parameter space is high dimensional or complex, is useful for estimating a subset of the parameters in structural models with cross-sectional dependence.[18]

---

empirical support for the existence of complementarities, they also note that the additive specification does not appear to be a bad approximation in practice.

[15]See Chiappori and Salanié [38] for a review of this literature. A seminal paper in this literature is Choo and Siow [39], which considers inference in a transferable utility setup with a continuum of agents.

[16]A popular approach for estimating two-sided matching models builds on the notion that the observed matching is pairwise stable. For example, see Fox and Bajari [55], Echenique et al. [51], and Menzel [93]. Requiring that the observed matching be pairwise stable may be unrealistic in the context of a frictional labour matching market of the sort that is the focus of this chapter.

[17]This contrasts with cases in which the researcher sees many independent copies of games involving few players, such as those studied by Bresnahan and Reiss [26], Ciliberto and Tamer [40], Berry [17] and many others. See Xu [114], Song [107], Menzel [94] for more papers discussing the estimation of large Bayesian games.

[18]The simulation-based approach used in the second-stage of the inference procedure is known

Section 1.2 introduces the model of two-sided labour market matching with frictions. In the baseline model of Section 1.2.1, workers and firms with exogenous characteristics match with one another and split the match surplus according to a Nash bargaining rule. The rest of the chapter is organized as follows. Section 1.2.2 extends the baseline model to allow for endogenous worker characteristics - after observing their type, workers simultaneously invest in education prior to entering the labour market. Section 1.3 outlines an approach for inference on the parameters of the structural model of Section 1.2.2. Section 1.4 applies the structural inference methodology to matched employer-employee data from Statistics Canada's Workplace Employee Survey to study labour markets in Canada. Section 1.5 concludes. Mathematical proofs, additional details of the empirical application, and a simulation study illustrating reasonable performance of the estimators, are confined to the Appendix of this chapter.

1.2 The Labour Market As a Two-Sided Matching Market

Our goal is to study the distribution of education and wages using separate cross-sections of matched employer-employee data. The first subsection introduces the core elements of the model that will serve as the basis for the structural model in the second subsection.

1.2.1 Baseline model

Let $N_h = \{1,...,n_h\}$ be the set of workers and $N_f = \{1,...,n_f\}$ be the set of firms, where n_h and n_f are used to denote the total number of workers and firms, respectively. Each worker seeks one job and each firm seeks to hire one worker.

The matching of workers to firms will be determined by the preference rankings of workers and firms. Workers value the capital of firms, $K = (K_j)_{j\in N_f}$, and firms value the human capital of workers $H = (H_i)_{i\in N_h}$, where K_j and H_i are scalars. Any worker i who is matched with firm j receives wage $w_{ij} \geq 0$ while firm j re-

as a Monte Carlo test. Monte Carlo tests have a history in econometrics dating back at least to the 1950s, as discussed by Dufour and Khalaf [48] in their overview of the technique.

ceives profit $\rho_{ji} \geq 0$, where both wages and profits may also depend on a parameter, $\theta \in \mathbf{R}^d$.[19] Since our framework supposes that wages and profits are always non-negative for any worker and firm that could match, I will assume throughout the chapter that no agent will ever unilaterally dissolve a match to become unmatched. This requirement that any matching satisfy an individual rationality constraint is embodied in the following condition:[20]

Condition IR (Individual rationality of matches): *For each $i \in N_h$, and $j \in N_f$ $w_{ij} \geq 0$ and $\rho_{ji} \geq 0$.*

Based on the values of $\rho_j = (\rho_{ji})_{i \in N_h}$ each firm j can construct preference rankings over the workers. We suppose that if the firm is ever indifferent between one or more workers, then the firm picks preference rankings over these workers at random. Next, we introduce a condition on the worker's wage function that will be useful for interpreting the matching process (along with our notion of information frictions).

Condition H (Homogeneous worker preferences): *For each $i \in N_h$, the wage of worker i is increasing in the capital of their matched firm.*

The assumption is tantamount to a notion of worker preference homogeneity, implying that all workers prefer higher capital firms. Supposing that workers accurately observe the capital of firms, this assumption implies that a matching algorithm in which the highest capital firm, j_1, choose his preferred worker, $i_1 \in N_h$, the second highest capital firm, j_2, choose his preferred worker $i_2 \in N_h \backslash \{i_1\}$ and so on is an example of the *serial dictatorship* mechanism (Abdulkadiroğlu and Sönmez [1]) and would produce a stable matching.[21]

In order to build a model that accounts for the possibility of mismatches between workers and firms, we suppose there are information frictions in the market.

[19] In this setup, θ represents the preferences of both workers and firms. As we will see, w_{ij} and ρ_{ji} depend on the output of worker i at firm j, and the production function that gives rise to this output will depend on a part of θ.

[20] The current setup is tailored to settings where the researcher has at least one cross-section of matched employer-employee data and the agents who are unmatched are not of primary interest in the analysis. An interesting (and challenging) extension of the current framework would accommodate the possibility of unmatched agents, and hence unemployment.

[21] See Section 2.2. of Roth and Sotomayor [102].

Specifically, workers do not directly observe realizations of the firm's capital. Instead, each worker sees $v = (v_j)_{j \in N_f}$, where v_j is a 'noisy' measure of firm j's capital. In particular, suppose that workers see

$$v_j = \beta K_j + \eta_j, \tag{1.1}$$

for each j, where $\beta \in B$, $B \subset \mathbf{R}$ is the parameter space of β, and η_j is a random variable that is independent across j.[22] The size of the variance of η_j relative to the magnitude β represents the magnitude of information frictions in the matching process. It is clear that when β is zero and the variance of η_j is positive, then this setup yields random matching from firm to worker the characteristics, since variation across firm capital plays no role in determining the realizations of v. Furthermore, when $\beta \neq 0$ and $\mathrm{Var}(\eta_j) = 0$, it will be as if firm capital is observed by the worker, since v_j is determined entirely by the firm's capital. In the latter case, when $\beta > 0$ workers would favour firms with the largest realizations of v, while in the case that $\beta < 0$, workers would favour firms with the smallest realizations of v. However, even in the case that $\mathrm{Var}(\eta_j) > 0$, v_j still conveys some useful information to the worker under certain circumstances. When $\beta > 0$, η_j's are iid and workers see $v_{j_1} > v_{j_2}$, then workers would still prefer matching with Firm j_1 over Firm j_2 since the distribution of K_{j_1} stochastically dominates distribution of K_{j_2}. The following condition specifies the matching process we will use throughout the chapter.

Condition SD (Matching process): *The matching of workers to firms in the economy arises as follows. The highest v firm, j_1, chooses his preferred worker, $i_1 \in N_h$, the second highest v_2 firm, j_2, chooses his preferred worker $i_2 \in N_h \backslash \{i_1\}$, and so on, until the lowest v firm, j_{n_f}, chooses his preferred worker among those not chosen by any higher ranked firms.*

One way of understanding this matching algorithm in economic terms is to

[22] We make all independence assumptions explicit in Section 1.2.2 when the structural model is introduced (in particular, see Assumption 1.2.1). Sections 1.2.2 and 1.3.2 also include further assumptions regarding the distribution of the model variables (Assumption 1.3.2) and the information that agents have (Assumption 1.3.1). In addition, although we do not discuss inference until Section 1.3, the reader may presume that the econometrician observes firm and human capital, wages, and the matching of workers to firms, but observes neither η nor v.

consider the following thought experiment. Imagine a situation in which a group of job-seekers have assembled in a large room on the day of a job fair. Workers do not observe the true quality of any of the firms, (represented by K), but they do see each firm's value of v. When $\beta > 0$ and η_j's are iid, each worker is happiest to match with the highest v firm, since the distribution of capital associated with the highest v firm stochastically dominates the distribution of capital associated with any of the lower v firms. A procedure in which the highest v firm, j_1, chooses his preferred worker, $i_1 \in N_h$, the second highest capital firm, j_2, chooses his preferred worker $i_2 \in N_h \backslash \{i_1\}$ and so on, will have no complaints from any of the participants at the job fair – that is, until uncertainty associated with K is revealed. In this world, agents will typically have more regret (and hence a greater desire to rematch) when the frictions in v are large. However, rematching is outside the scope of the model.

Next, we add some further structure to wages and profits. In particular, we will assume that the payoffs for any two matched agents follow a Nash bargaining structure. Let $\tau \in (0,1)$ be the bargaining weight. A worker i who matches with a firm j receives

$$w_{ij} = \tau f(H_i, K_j) + (1-\tau) g(H_i), \tag{1.2}$$

while firm j receives

$$\rho_{ji} = (1-\tau)\left(f(H_j, K_j) - g(H_i)\right), \tag{1.3}$$

where f is the worker-firm output function and $g(H_i)$ is an outside option function, both of which may depend on elements of θ. In a subsequent section, we will allow worker covariates, X_i, to effect wages through the outside option function, g.[23] The following condition requires f to satisfy some intuitive properties with respect to the worker and firm capital variables.

Condition F (Production function): *f is increasing in human capital and firm capital.*

Condition F merely requires that more capital leads to more output - it does not impose that the worker and firm attributes be complements in f. Section 1.2.2 goes

[23] X_i's have support $\mathscr{X} \subset \mathbf{R}^d$, where d is an integer greater than or equal to one.

into further detail about the role of f in this model.

1.2.2 Frictional Matching Model with Worker Investments

We now introduce a structural model where workers simultaneously invest in education prior to the serial dictatorship matching process as outlined in the previous section. A general overview of the matching process is as follows: i) workers, observing only their type, simultaneously choose a level of education, ii) v is realized, iii) firms, seeing only the education of workers, match according to Condition SD.

Although firms select their preferred workers in the serial dictatorship phase after constructing preference rankings over the workers, firms are not considered strategic agents within the context of the investment game itself.

There are n_h players indexed by $i \in N_h$. Each player chooses an education level, h_i, from the discrete set $\mathscr{H} \equiv \{1, ..., J\}$ to maximize their expected payoff. Let $\lambda = (\theta', \beta)$, where β is the matching frictions parameter and $\theta \in \mathbf{R}^d$ is a preference parameter. The payoff function of player i comprises the wage less a cost of education,

$$u(h_i, h_{-i}, x_i, k, \eta, \varepsilon_i; \lambda) = \omega(h_i, h_{-i}, x_i, k, \eta; \lambda) - c(h_i, x_i, \varepsilon_i; \lambda), \tag{1.4}$$

where $h_{-i} \in \mathscr{H}_{-i}$ are the choices of the other agents,[24] $x_i \in \mathscr{X}$ and $\varepsilon_i \in \mathbf{R}^J$ are the private information of worker i, and $k \in \mathbf{R}^{n_f}$ and $\eta \in \mathbf{R}^{n_f}$ are vectors of exogenous firm variables that are unobserved by the workers. Although ε_i and x_i are private information of the worker, we will assume x_i is observed by the econometrician in a subsequent section. The variable ε_i represents the worker's private cost associated with each of the J education levels. In Section 1.3.2 we will supply explicit assumptions on worker and firm information that illustrates why, given the matching process, the components of the payoff function depend on model's underlying variables in the way stipulated by equation (1.4).

We now provide additional conditions that establish the existence of a Bayesian Nash equilibrium for our game (which we prove in Appendix A.1.1).

[24] Since the set of pure strategies for each agent is $\mathscr{H}$, it follows that $\mathscr{H}_{-i} = \mathscr{H}^{n_h - 1}$ for each i, where $\mathscr{H}^{n_h-1}$ denotes the $(n_h - 1)$-ary Cartesian power of $\mathscr{H}$.

Assumption 1.2.1. *(a) K_j's, η_j's are independent across j. X_i's, ε_i's are independent across i. X, K, ε, and η are independent. (b) ε_i's are continuously distributed.*

Assumption 1.2.2. *The cost function is separable in private information*:

$$c(h_i, x_i, \varepsilon_i; \lambda) = c_0(h_i, x_i; \lambda) + \varepsilon_i' d(h_i),$$

where $d(h_i)$ is a J dimensional vector with one in the h_i-th row and zero otherwise.

The assumptions of separability and independence are common in the structural literature.[25] In Appendix A.1.1, we show that Assumptions 1.2.1 and 1.2.2 are sufficient for establishing the existence of the Bayesian Nash equilibrium for the game of this section. For now, we will provide some intuition into the worker's education decision problem. First, we define the set of pure strategies as $\sigma = \{\sigma_i(x_i, \varepsilon_i) : i \in N_h\}$ where σ_i is a function that maps from $\mathscr{X} \times \mathbf{R}^{J-1}$ into $\mathscr{H}$. Assumption 1.2.2 says that we can write the expected utility of agent i with covariates x_i, who chooses h_i under beliefs σ as

$$U_i(h_i, x_i, \sigma, \varepsilon_i) \quad = \quad \tilde{U}_i(h_i, x_i, \sigma) + \varepsilon_i' d(h_i), \tag{1.5}$$

where the first term in the expected utility is

$$\tilde{U}_i(h_i, x_i, \sigma) \quad = \quad \sum_{h_{-i} \in \mathscr{H}_{-i}} \tilde{u}_i(h_i, h_{-i}, x_i) P_{-i}(h_{-i}|\sigma), \tag{1.6}$$

and

$$\tilde{u}_i(h_i, h_{-i}, x_i) \equiv \tilde{\omega}_i(h_i, h_{-i}, x_i) - c_0(h_i, x_i), \tag{1.7}$$

where $\tilde{\omega}_i(h_i, h_{-i}, x_i)$ is given by

$$\tilde{\omega}_i(h_i, h_{-i}, x_i) = \mathbf{E}[\omega(H_i, H_{-i}, X_i, K, \eta; \lambda) | H_i = h_i, H_{-i} = h_{-i}, X_i = x_i],$$

and expectation is taken with respect to the distributions of K and η. By Lemma

[25] For example, see the discussions in Kasahara and Shimotsu [76] and Xu [113].

(A.1.1) (on page 122) we can rewrite equation 1.6 as

$$\tilde{U}_i(h_i, x_i, \sigma) = \sum_{h_{-i} \in \mathscr{H}_{-i}} \tilde{u}_i(h_i, h_{-i}, x_i) \prod_{j \in N_h \setminus \{i\}} P_j(h_j | \sigma_j).$$

Throughout this chapter, we will consider the case in which the wages of workers are determined by Nash bargaining. As in equation 1.2, we will suppose that firm capital only enters the worker's payoff through the production function. Denote $\mathscr{M}(i)$ as the identity of the firm that worker i matches to as a result of the matching process, and $K_{\mathscr{M}(i)}$ as the level of capital associated with firm $\mathscr{M}(i)$. Under Nash Bargaining wages where firm capital only enters the production function, we may write $\tilde{\omega}_i(h_i, h_{-i}, x_i)$ as

$$\tilde{\omega}_i(h_i, h_{-i}, x_i) = \tau \tilde{f}_i(h_i, h_{-i}) + (1 - \tau) g(h_i, x_i), \tag{1.8}$$

where

$$\tilde{f}_i(h_i, h_{-i}) = \mathbf{E}[f(H_i, K_{\mathscr{M}(i)}) | H_i = h_i, H_{-i} = h_{-i}, X_i = x_i],$$

the expectation is taken with respect to the distributions of K and η,[26] and we have allowed the worker's characteristics to enter the payoff function through the outside option function, g.[27]

Education affects the worker's expected utility in a number of ways. The first two are obvious: since f is increasing in h_i by Condition F, the worker who invests in a higher level of education obtains a higher wage at any firm he matches to. The worker's choice of education also affects his payoff through the outside option function, g. The novel channel in this setup is that h_i also determines the expected quality of the firm that i matches to. Even though (as mentioned before) firms in this model are non-strategic agents, the functional form of the production function, f, plays a key role in determining whether or not firms with different levels of

[26] In Section 1.3.2, we will demonstrate the precise form that this expectation takes under particular lower-level assumptions.

[27] Here, $\tau_i = \tau$ for each i. My framework can be extended to incorporate heterogeneity in worker bargaining positions. In the empirical results, however, I ignore this channel. Bagger and Lentz [11] emphasize the role that disentangling variation such as endogenous search intensity from matching variation (e.g., Postel-Vinay and Robin [99]) plays in understanding the causes of wage inequality.

capital exhibit different preferences for workers of differing levels of education. To see how f determines whether or not firms' preferences are heterogeneous, consider the Nash bargaining preferences of a firm for any worker who chooses education level h:

$$\rho(k,h;\theta) = (1-\tau)(f(h,k;\theta) - \tilde{g}(h;\theta)), \tag{1.9}$$

where $\tilde{g} = \mathbf{E}g(h,X_i)$ and the expectation is taken with respect to the distribution of X_i.[28] Suppose that X_i is iid, K takes two values k_1, k_2 and there are two levels of education, h_1, h_2 with $h_2 > h_1$. Let us denote the set of firms that prefer high education (h_2) as

$$M_2^+(\theta) = \{m \in \{1,2\} : \rho(k_m,h_2;\theta) \geq \rho(k_m,h_1;\theta)\}.$$

If $f(h,k)$ is of the form $a(h)+b(k)$, where a and b are two functions that map the capital variables to the real numbers, then $M_2^+(\theta)$ will be either $\{1,2\}$ or $\emptyset$. In this case we say that firms have homogeneous preferences, since both types of firms in the economy prefer the higher educated workers. Alternatively, if $f(h,k)$ is of the form $a(h)b(k)$ then $M_2^+(\theta)$ will be either $\{1,2\}$, $\emptyset$, or $\{2\}$. This is the case of heterogeneous firm preferences. In this latter case where f exhibits complementarities in worker and firm types, the set of firms types that prefer high to low education is more finely partitioned. Moreover, the presence or absence of complementarities will play a key role in determining the severity of wage inequality, as we will see in Section 1.4. More general than all these points, however, is the following fact about the model: as long as k appears somewhere in f, k does not have to interact directly with h in f for the information frictions represented by β to matter in worker's investment decision.

1.2.3 Some Implications of Frictional Matching Model

In this section, we explore some key features of the model. We will suppose that the functional forms, underlying distributions, and firm preferences are such that

[28] Here, we implicitly assume that firms do not observe workers' covariates and rank workers only in terms of their education. We make these assumptions concerning firm information explicit in a subsequent section.

firms always strictly prefer higher educated workers. In the following subsection, we will illustrate sorting without any direct interactions between worker and firm types in the production function.

Sorting Without Complementarities

In Figure 1.1 and Figure 1.2, we compare the equilibrium probability of investing in education and the equilibrium Gini coefficient for a range of the friction parameters under two specifications of the production function: Specification 1 allows direct interaction between worker and firm types, $f = \theta_1 hk$, while such interactions are absent in Specification 2, $f = \theta_1(h+k)$.[29] Each point on the plot is the average of 500 simulations of endogenous variable from the equilibrium of the model. The outside option parameter is set to $\theta_2 = (-.75, .25, .5)$. There are 100 workers and firms. In Specification 1, the high value of θ_1 is 3.5, and the low value of θ_1 is 1.5. In Specification 2, the high value of θ_1 is 1.8, and the low value of θ_1 is 0.8. There are two levels of of firm capital: $K = 1/2$ and $K = 1$. The fraction of each type of firm is .5 in the economy.

A number of implications are straightforward: the equilibrium probability of investing in high education is higher when θ_1 is higher and frictions are lower. When θ_1 is higher, workers will be compensated more for higher levels of education. When β is higher, the probability of matching to a higher type firm when they choose high education is higher.

The effect of increasing β (lowering matching frictions) on both the education and wage inequality is typically much more dramatic in Specification 1. A rise in β (a lessening in matching frictions) increases sorting in both specifications, though the effect is more dramatic in the complementarities case: in Figure 1.1, the correlation between worker and firm types rises from 60% to 68% when θ_1 is high, but from 64% to 81% when θ_1 is lower; in Figure 1.2, the correlation between worker and firm types rises from 84% to 90% in the high theta case whereas it rises from 77% to 87% in the low theta case. The overall level of inequality in Specification 1 is also higher since whatever sorting is present is amplified to a greater extent when the types interact in the wage equation than when they do not.

[29]The precise functional forms are the same as the ones used in Section A.1.3.

The high θ_1 case in the right hand panel of Figure 1.1 also illustrates the role that two competing effects of changes in β play on the level of wage inequality. When β rises from 0 to 1, the level of inequality increases through the sorting channel. However, as β continues rises, the equilibrium probability of investing in education also continues to rise. As the fraction of highly educated surpasses 80%, the level of inequality begins to level off (at $\beta = 2$) and then begins to fall. This phenomenon is also illustrated to a lesser degree in the high θ_1 case of the right hand side panel of Figure 1.2.

Supply of Highly Educated Workers and Education Premia

In this section, we show how simulation of our static model can capture a puzzling phenomenon discussed in Card and Lemieux [33]. How can dramatic increases in the education wage premium lead to only modest increases in the supply of highly educated workers? The authors note that, over a roughly 30 year period beginning in the early 1970s, the college-high school wage gap rose considerably in the United States, Canada, and the United Kingdom, and that this rise occurred mostly for younger workers. They argue that an important source of this trend is a stagnation in the rate of educational attainment among workers born in the 1950s and thereafter.

In Figure 1.3, we show how this pattern can be driven entirely by changes in the matching technology over time. The wage premium is measured as the difference between the average wages of the workers with high education and the average wages of workers with low education. Each point on the plot represents the average of 500 simulations of the model. We use Specification 1, $f = \theta_1 hk$, under the same setup as before with only one difference; we choose the low value of θ_1 to be 0.6 and the high θ_1 to be 3.5. In the case that θ_1 is very low, the effect of raising β is to dramatically increase sorting without a large benefit to any particular worker.

1.3 Econometric Inference

In this section, we outline the general empirical strategy for performing inference on the underlying model parameters. In Section 1.3, we describe how the main model can be used to characterize the observed distribution of the matching of

Figure 1.1: Education and Wage Inequality Under Specification 1

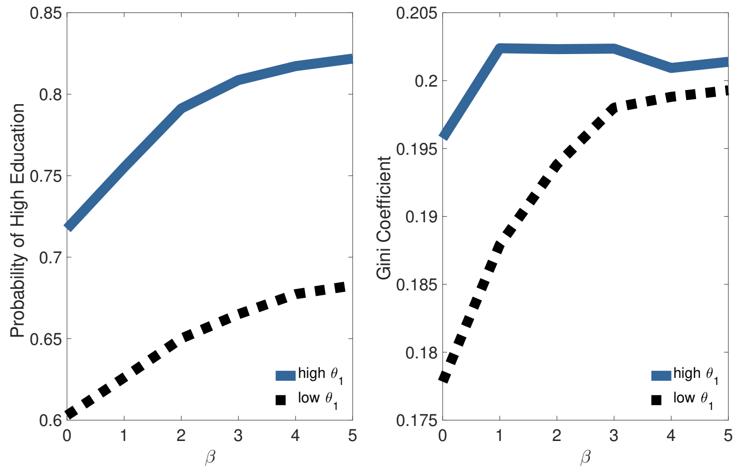
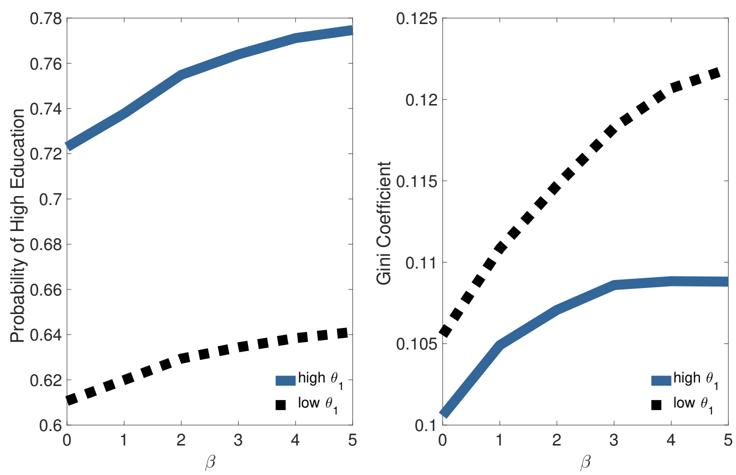
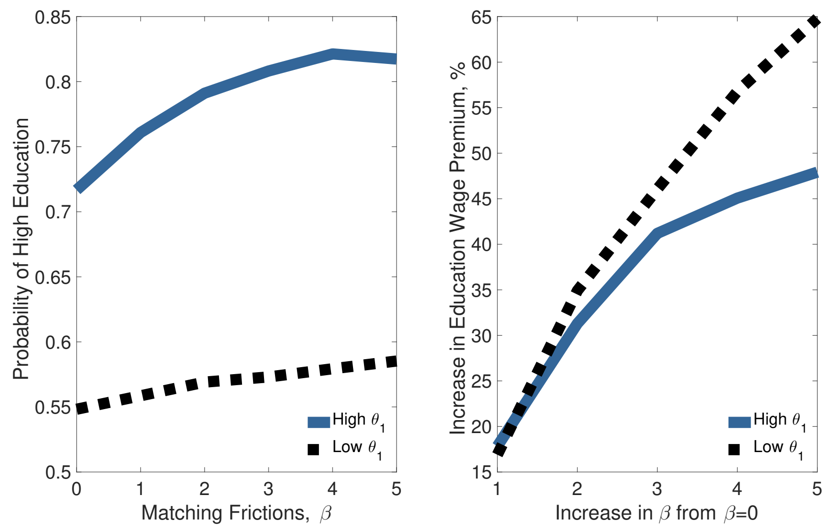


Figure 1.2: Education and Wage Inequality Under Specification 2



Figures 1.1 and 1.2 plot the equilibrium probability of high education investment and the Gini coefficient for a range of values of the matching frictions parameter, β. We consider two specifications for the production function: Specification 1 includes interactions between worker and firm types while Specification 2 does not. Lowering matching frictions (increasing β) increases the equilibrium level of education across specifications. A rise in β impacts inequality through two competing effects: a sorting effect that increases inequality and an a supply effect that lowers inequality. This can be seen most dramatically in Figure 1.1: as β rises past a value of three, the fraction of highly educated rises more and more and inequality falls, dominating the effects of sorting on inequality.

Figure 1.3: Supply of Highly Educated Workers and Education Wage Premia



Figures 1.3 offers an explanation to an empirical puzzle discussed in Card and Lemieux [33]: why are increases in wage premia not associated with large increases in the supply of highly educated workers? We plot the equilibrium probability of high education investment and the returns to education for a range of values of the matching frictions parameter, β. We consider Specification 1. In the case that θ_1 is very low, the effect of increasing β is to dramatically increase sorting while keeping the returns to education for any particular worker reasonably low.

workers to firm and hence the wages of all the workers in the economy. The goal is to then use these representations to construct confidence regions for the preference and matching technology parameters.

However, if the model is high dimensional, the Monte Carlo inference approach may be cumbersome to apply in practice. For this reason, we propose a two-stage inference approach that relies on the construction of a first-stage confidence interval for a subset of the model parameters. We demonstrate this approach in practice in Section 1.3.2 by estimating the Bayesian game from 1.2.2 for fixed values of β.

1.3.1 Two-Stage Inference Accommodating Cross-Sectional Dependence of Observed Matching

The econometrician observes a matching of workers to firms, $\mathbf{M} = (\mathbf{M}(i))_{i \in N_h}$, where for each $i \in N_h$, $\mathbf{M}(i)$ takes values in the set of firms.[30] The main challenge associated with inference is the fact that the distribution of $\mathbf{M}$ exhibits cross-sectional dependence of a complicated form. The matching of workers to firms can be thought of as discrete choice problem on the part of the firm where the choice sets of firms are endogenously constrained by the choices of firms with higher v-indices, which depends on β, η and k. Hence, the event that worker i matches to firm j cannot be considered independent from the event that a worker $i' \neq i$ matches to firm j. Also, the fact that firm preferences are heterogeneous means we cannot condition on the v-index and firm preferences in a way to remove the cross-sectional dependence as was done by Agarwal and Diamond [5].

The econometrician observes the vector $\mathbf{M} \in \mathbf{R}^{n_h}$, which represents a matching of workers to firms. Given the serial dictatorship matching process, the joint distribution of $\mathbf{M}$ is known up to a parameter. Let $\mathbf{K} = (\mathbf{K}(i))_{i \in N_h}$, where $\mathbf{K}(i) = K_{\mathbf{M}(i)}$; i.e., the capital of the firm matched to by worker i.

Our model also implies that the finite sample distribution of wages, $(\mathbf{W}(i))_{i \in N_h}$, is known up to a parameter. Under Nash bargaining (and a specification of the post-match wage function based off an equation such as 1.2), we have for each $i \in N_h$

$$\mathbf{W}(i) = w(H_i, \mathbf{K}(i)).$$

We denote all the match-related observables as $\mathbf{Y} = (\mathbf{K}, \mathbf{M})$. $\mathbf{M}$ is observed whenever the researcher has matched employer-employee data. $\mathbf{K}$ is observed when the researcher can use the matching data, $\mathbf{M}$, and the firm capital data, K, to find the capital of the firm each worker in the sample is employed at. Using $\mathbf{Y}$ and worker observables H and X, the econometrician wishes to infer λ_0.

[30] Throughout this chapter, we will suppose that the matching is one-to-one between workers and firms. In practice, "firms" in this context can be viewed as positions at particular firms.

Inference on Parameters

Next, we consider a test statistic that matches the moments of the distribution of the matched-related observables with their simulated counterparts. To simplify the exposition, we discuss the construction of a confidence interval for β_0 alone, i.e., supposing that we knew the true values of θ_0. Denote $R+1$ as the total number of simulations in the Monte Carlo inference procedure. Drawing η_r from some continuous parametric distribution F_θ,[31] we simulate a version of the matching for each $\beta \in B$ and each $r = 1, \ldots, R+1$, which we write as $\mathbf{M}_r(\beta) = \{\mathbf{M}_r(i;\beta) : i \in N_h\}$. The simulated wages are then

$$\mathbf{W}_r(i;\beta) = w(H_i, K_{\mathbf{M}_r(i;\beta)})).$$

It is convenient to define

$$\begin{aligned} \mathbf{Y}_r(\beta) &= \{\mathbf{Y}_r(i;\beta) : i \in N_h\}, \\ \mathbf{Y}^*_{R+1}(\beta) &= \{\mathbf{Y}_r(i;\beta) : i \in N_h, r = 1, \ldots, R+1\}, \text{ and} \\ \mathbf{Y}_{-r}(\beta) &= \mathbf{Y}^*_{R+1}(\beta) \backslash \mathbf{Y}_r(\beta). \end{aligned}$$

Next, we will propose a test statistic that depends on both the observed matching data, $\mathbf{Y}$, and the simulated matching data, (along with simulated versions of this test statistic). That is,

$$T(\beta) = \phi_n(\mathbf{Y}, \mathbf{Y}^*_R(\beta)), \quad \text{and}$$
$$T_r(\beta) = \phi_n(\mathbf{Y}_r(\beta), \mathbf{Y}_{-r}(\beta)).$$

An example of such a test statistic is one that compares the observed joint distribution of worker human capital and matched firm capital with simulated counterparts. For example, we may consider the test statistic

$$T(\beta) = \frac{1}{R} \sum_{r=1}^{R} \left\| \hat{P} - \hat{P}_r(\beta) \right\|,$$

[31] We will specify a particular parametric family that F_θ belongs to, along with additional assumptions, in Section 1.3.2.

where $\hat{P}$ is an $M \times J$ matrix[32] whose (m, j) element is the estimated probability that a worker of education level h_j matches to a firm of capital level m, $\hat{P}_r(\beta)$ is defined similarly to $\hat{P}$, except we replace the observed matching with the rth simulated matching, $\mathbf{M}_r(\beta)$, and $\|\cdot\|$ is the matrix norm.[33]

Using our test statistic, we may compute a confidence region for β as

$$C^{\beta}_{\alpha,R} = \{\beta \in B : T(\beta) \leq c_{\alpha,R}(\beta)\},$$

where the critical value is computed as the $(1-\alpha)$ -quantile of the empirical distribution of $\{T_r(\beta) : r = 1,...,R\}$:

$$c_{\alpha,R}(\beta) \quad = \quad \inf\left\{c \in \mathbf{R} : \frac{1}{R}\sum_{r=1}^{R} 1\{T_r(\beta) \leq c\} \geq 1-\alpha\right\}.$$

Under Assumption 1.3.2, it can easily be shown that finite sample inference on β_0 satisfies $P\{\beta_0 \in C^{\beta}_{\alpha,R}\} \geq 1-\alpha$ when the procedure outlined above involves the true parameter, θ_0.

In practice, we do not know the true value of θ_0. In situations in which the full parameter vector λ_0 is not very large, it may be feasible to construct a $(1-\alpha)100\%$ confidence region for this parameter that exhibits finite sample validity. That is, we construct

$$C^{\lambda}_{\alpha,R} = \{\lambda \in \Lambda : T(\lambda) \leq c_{\alpha,R}(\lambda)\}, \tag{1.10}$$

where $T(\lambda)$ and $c_{\alpha,R}(\lambda)$ are defined analogously to $T(\beta)$ and $c_{\alpha,R}(\beta)$. In the case that Λ is high-dimensional, the finite sample procedure outlined above may not be practical due to the unreasonable computational cost. In the following subsection, we explore a two-stage inference approach that admits inference on β_0 when the researcher is able to construct a first-stage confidence region for a subset of the parameters, θ_0.

Plugging in a consistent estimator of θ_0, $\hat{\theta}_n$, for the true value in inference pro-

[32]In this example, we are implicitly assuming that the distribution of K is discrete and has M support points. We will make this assumption explicit in a subsequent section.

[33]See equation A.12 in the empirical section for more on constructing $\hat{P}$ and $\hat{P}_r(\beta)$. In this section, we also choose the matrix norm to be the Frobenius norm. That is, for an $m \times n$ matrix A, $||A|| = \left(\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2\right)^{1/2}$, where a_{ij} denotes the (i, j)-element of A.

cedure outlined above will (in general) not lead to valid inference on β_0. This is because there is no reason to suspect that plugging in $\hat{\theta}_n$ for θ_0 will make the distribution of the simulated matching, $\mathbf{M}_r$, equal to the distribution of the observed matching, $\mathbf{M}$. The fact that $\mathbf{M}_r$ is not equal in distribution to $\mathbf{M}$, in turn implies that $\mathbf{K}_r$ does not follow the same distribution as $\mathbf{K}$. The severe consequences of estimation error in $\hat{\theta}_n$ occur because the firm preferences are typically misspecified at all values of θ other than the true value, θ_0. Moreover, this problem is not alleviated by conditioning on H, K, or exogenous variables. In the following section, we discuss a general two-stage inference approach when the econometrician can construct an (asymptotically) valid confidence first-stage confidence interval for θ_0. In Section 1.2.2, we extend our baseline economic model of Section 1.2 in a manner that admits the application of this two-stage inference approach to our setup.

Two-Stage Inference on β using Test-Inversion Confidence Interval

Suppose that we wish a $(1-\alpha)$-level asymptotic confidence interval for β_0, and can construct a confidence interval for θ_0. Let us denote the test statistic and its simulated counterpart from the previous section, where the θ arguments make explicit the test statistic's dependence upon a given value of $\theta \in \Theta$:

$$
\begin{aligned}
T(\beta;\theta_0,\theta_1) &= \phi_n(\mathbf{Y}(\beta_0,\theta_0),\mathbf{Y}_R^*(\beta,\theta_1)), \quad \text{and} \\
T_r(\beta;\tilde{\theta},\theta_1) &= \phi_n(\mathbf{Y}_r(\beta,\tilde{\theta}),\mathbf{Y}_{-r}(\beta,\theta_1)).
\end{aligned}
$$

Note that according to the notation we used in the last section we have $T(\beta;\theta_0,\theta_0) = T(\beta)$. Our inference on β proceeds in two steps:

Step 1. Using the first stage estimates of $\hat{\theta}(\beta)$, we construct a confidence region for θ_0, $\hat{C}_{\alpha/2}(\beta)$, with $(1-(\alpha/2))$ asymptotic coverage.
Step 2. Next, we construct a test statistic that doesn't involve θ. Define

$$
\begin{aligned}
S(\beta) &= \inf_{\theta_1 \in \hat{C}_{\alpha/2}(\beta)} T(\beta;\theta_0,\theta_1), \text{ and} \\
S_r^*(\beta) &= \sup_{\tilde{\theta} \in \hat{C}_{\alpha/2}(\beta)} \inf_{\theta_1 \in \hat{C}_{\alpha_{/2}}(\beta)} T_r(\beta;\tilde{\theta},\theta_1).
\end{aligned}
$$

We now construct a confidence set for β as

$$\hat{C}_{\alpha,R} = \{\beta \in B : S(\beta) \leq c^*_{1-(\alpha/2),R}(\beta)\}, \tag{1.11}$$

where the critical value $c^*_{1-(\alpha/2),R}(\beta)$ is computed as the $(1-(\alpha/2))$ -quantile of the empirical distribution of $\{S^*_r(\beta) : r = 1, ..., R\}$; that is,

$$c^*_{1-(\alpha/2),R}(\beta) \quad = \quad \inf\left\{c \in \mathbf{R} : \frac{1}{R}\sum_{r=1}^{R} 1\{S^*_r(\beta) \leq c\} \geq 1-(\alpha/2)\right\}.$$

The following lemma establishes the asymptotic validity of the two-stage inference procedure.

Lemma 1.3.1. *Suppose that the econometrician can construct $\hat{C}_{\alpha/2}(\beta_0)$ such that*

$$\lim_{n\to\infty} P\left(\theta_0 \in \hat{C}_{\alpha/2}(\beta_0)\right) \geq 1-(\alpha/2).$$

Then

$$\lim_{n\to\infty} P\left(\beta_0 \in \hat{C}_{\alpha,R}\right) \geq 1-\alpha. \tag{1.12}$$

Proof. By the definition of $\hat{C}_{\alpha,R}$, $P\left(\beta_0 \in \hat{C}_{\alpha,R}\right)$ is equal to

$$\begin{aligned} & P\left(S(\beta_0) \leq c^*_{1-(\alpha/2),R}(\beta_0)\right) \\ = \quad & P\left(\inf_{\theta_1 \in \hat{\mathscr{C}}_{\alpha/2}(\beta)} T(\beta_0;\theta_0,\theta_1) \leq c^*_{1-(\alpha/2),R}(\beta_0)\right) \\ \geq \quad & P\left[\left\{\inf_{\theta_1 \in \hat{\mathscr{C}}_{\alpha/2}(\beta)} T(\beta_0;\theta_0,\theta_1) \leq c^*_{1-(\alpha/2),R}(\beta_0)\right\} \cap A_1\right], \end{aligned} \tag{1.13}$$

where $A_1 \equiv \left\{\theta_0 \in \hat{C}_{\alpha/2}(\beta_0)\right\}$. Then, the right hand side of the right hand side of

(1.13) is greater than or equal to

$$
\begin{aligned}
& P\left[\left\{\sup_{\tilde{\theta}\in\hat{C}_{\alpha/2}(\beta)}\ \inf_{\theta_1\in\hat{\mathscr{C}}_{\alpha/2}(\beta)} T_r(\beta;\tilde{\theta},\theta_1)\leq c^*_{1-(\alpha/2),R}(\beta_0)\right\}\cap A_1\right], \\
\geq\ & P\left(\sup_{\tilde{\theta}\in\hat{C}_{\alpha/2}(\beta)}\ \inf_{\theta_1\in\hat{\mathscr{C}}_{\alpha/2}(\beta)} T_r(\beta;\tilde{\theta},\theta_1)\leq c^*_{1-(\alpha/2),R}(\beta_0)\right)-P\left(A_1^c\right).
\end{aligned}
$$

Now since

$$
\lim_{n\to\infty} P\left(\theta_0\notin\hat{C}_{\alpha/2}(\beta_0)\right) \quad \leq \quad \alpha/2,
$$

we have

$$
\lim_{n\to\infty} P\left(\beta_0\in\hat{C}_{\alpha,R}\right) \quad \geq \quad 1-\alpha.
$$

□

In the following section, we describe how to construct $\hat{C}_{\alpha/2}(\beta)$ for each $\beta\in B$.[34]

1.3.2 First-Stage Estimation of θ

In this section, we show how θ can be estimated for a particular fixed value of β. We will write an estimator of such an object as $\hat{\theta}(\beta)$. The main challenge associated with this problem is that of estimating the worker's expected utility from equation 1.7. The problem is difficult because the workers must somehow resolve uncertainty associated with the serial dictatorship matching process in order to compute the expected output under the equilibrium education choices. In spite of these complications, it turns out that, under reasonable assumptions, the parameters are tractably estimable using discrete choice methods with a fixed point constraint when there are only two education choices. We now provide and discuss these assumptions.

[34] In the empirical application of this chapter, we estimate θ_0 by maximum likelihood and construct confidence regions using numerical derivatives of the likelihood function. A Monte Carlo simulation (reported in the Appendix) - illustrates acceptable finite sample performance of this inference approach.

Assumption 1.3.1. *(a) Firms observe (i) workers' education decisions, H, and (ii) the distribution of characteristics, X. (b) Workers observe (i) the distribution of firm capital, (ii) the distribution of η, (iii) the distribution of X, and (iv) the distribution of the number of firms preferring each education level $h_j \in \mathscr{H}$.*

Under part (a) of Assumption 1.3.1, firms do not take workers' covariates into account when forming their preference rankings over workers. Thus, workers with the same education level are equally desirable to any given firm. When worker i considers the desirability of choosing education h_j, he need only consider the capital a generic agent who chooses level h_j expects to receive in the matching process. In many contexts, (a) will be reasonable for a host of variables that affects the worker's education decision (e.g., marital status, number of dependent children).[35] Part (b) says that workers know only the distribution of firm capital without knowing the precise realizations of capital. Assumption 1.3.1 (b) also stresses that the worker's knowledge of the distribution of capital is not sufficient for knowledge of the distribution of the number of firms that prefers each education class, which will turn out to be crucial for our results of this section.

Assumption 1.3.2. *(a) K is discrete with probability mass $q = (q_m)_{m=1}^{M}$ where $q_m = P(K = k_m)$ for $m = 1, ..., M$.*
(b) η_j's are iid $N(0, \sigma^2)$ (c) ε_i's follow the Type I extreme value distribution.

Part (a) says the distribution of firm capital has discrete support. In practice, we can let M be as large as our application requires. In concert with (b) and the parametric structure for v stipulated by equation 1.1, (a) allows us to express the unconditional distribution of v_j as a mixture of normals, $G \equiv \sum_{m=1}^{M} q_m F_m$, where F_m is $N(\beta k_m, \sigma^2)$.[36] Part (c) is an assumption on the worker's unobserved costs that allows us to estimate the model parameters using conventional discrete choice methods.

We wish to obtain a convenient representation of each worker's conditional expectation of the production function, for each education level that the worker

[35] In some cases in which employers do see these worker characteristics, they are prohibited from discriminating based on them due to state or federal anti-discrimination laws.

[36] In the simulations and empirical sections of the chapter we normalize $\sigma^2 = 1$ when we perform inference on the model parameters.

can choose. Under the model of Section 1.2.2 the identity of the firm that worker i matches with, $\mathscr{M}(i)$, depends on K, H, β, and, θ. Therefore, for each $i \in N_h$ and $h_j \in \mathscr{H}$, we wish to estimate

$$\tilde{f}_{ij} \equiv \mathbf{E}[f(H_i, K_{\mathscr{M}(i)})|H_i = h_j, X_i = x_i],$$

where the expectation is taken with respect to the distribution of K, H_{-i} and η. Under Assumption 1.3.2 (a), we can express the expectation on the preceding line as

$$\tilde{f}_{ij} = f_j' \pi_j^{(i)}, \tag{1.14}$$

where $f_j = (f_{j1}, ..., f_{jm})'$ is an $M \times 1$ vector with the m-th element of f_j given as $f_{jm} = f(h_j, k_m)$ and $\pi_j^{(i)} = (\pi_{1j}^{(i)}, ..., \pi_{Mj}^{(i)})'$ is an $M \times 1$ vector with the m-th element of $\pi_j^{(i)}$ given as

$$\pi_{mj}^{(i)} = \sum_{h_{-i} \in \mathscr{H}_{-i}} P(\mathscr{M}(i) = m | H_i = h_j, H_{-i} = h_{-i}, X_i = x_i) P(h_{-i}|x_i). \tag{1.15}$$

This is the probability that worker i matches to a firm of capital level k_m when he has chosen education level h_j.[37] Given that there are M education levels, J choices, and n_h workers, the dimensionality of the problem appears daunting. However, under our assumptions the problem is simplified considerably, and we can show that for each j and m, $\pi_{mj}^{(i)} = \pi_{mj}$, and hence, $\tilde{f}_{ij} = \tilde{f}_j$.[38]

Although it is unclear how to represent π_{mj}'s analytically when the worker faces a choice between a large number of education levels, the problem becomes tractable when there are only two (i.e., $J = 2$). Proposition A.1.1 shows that under our informational assumptions, firms (and workers) cannot distinguish between workers with the same education level during the matching process. As a consequence, we find that a worker is only concerned with the number of other workers who picked one of the two education levels (and not which particular workers chose what). Independence and identical distributions assumptions imply that the probability that n_j workers picked education level h_j can be represented using the

[37] Note that although these terms depend on θ and β, we will occasionally omit these from our notation for convenience.

[38] The argument for why this is the case is given in the proof of Proposition A.1.1.

binomial probability mass function. However, the number of workers choosing education level h_j is unknown to workers, so they must take expectations. Thus, instead of having to sum over $n_h - 1$ indices associated with actions of each of the other workers to compute the worker's expectation, we need only sum over one: the number of workers choosing a particular education level.

We will also allow θ to enter π_{mj}'s through the distribution of the number of firms that prefer high (or low) education. The following assumption is a natural way to specify this distribution. We use the notation $M_j^+(\theta)$ to denote the set of firm types that prefer education level h_j.[39]

Assumption 1.3.3. *In the model with $J = 2$, the probability that exactly $n^{(j)}$ firms prefer workers with education level h_j follows the binomial distribution with probability $\sum_{m \in M_j^+(\theta)} q_m$.*

The explicit representation of the matching probabilities are given in Propositions A.1.2, A.1.3, and Lemma A.1.3. These results can be used to construct estimates of the π_{mj}'s - and hence the $\tilde{f}_j$'s - for fixed values of θ and β. Using a given functional form for the production function, we denote an estimate of the expected production function when the worker chooses education level h_j as

$$\hat{f}_j(\theta, \beta) = f_j' \hat{\pi}_j(\theta, \beta),$$

where our notation emphasizes the dependence of the objects upon the parameter values. To construct $\hat{\pi}_{mj}$'s we must estimate the terms of equation A.9. $\hat{P}(n_j)$ is constructed as $B(n_j; n_h - 1, \hat{p}_j)$ where the latter denotes the binomial probability mass function with $\hat{p}_j = P(H_i = h_j)$.[40] Similarly, $\hat{P}(n^{(j)}; \theta)$ is constructed as $B(n_j; n_h - 1, \hat{q}_j(\theta))$, where $\hat{q}_j(\theta) = \sum_{m \in \hat{M}_j^+(\theta)} \hat{q}_m$, with $\hat{q}_m = \hat{P}(K_j = m)$,

$$\hat{M}_j^+(\hat{\theta}) = \left\{ m \in \{1, ..., M\} : \hat{\rho}(k_m, h_j; \theta) \geq \hat{\rho}(k_m, h_{j'}; \theta), j \neq j' \right\},$$

and $\hat{\rho}(k_m, h_j; \theta)$ is as in equation (1.9), except we use $\hat{g}_j = \frac{1}{n} \sum_{i=1}^n g(h_j, X_i)$ in place of $\tilde{g}$.

[39] That is, $M_j^+(\theta) = \{m \in \{1, ..., M\} : \rho(k_m, h_j; \theta) \geq \rho(k_m, h_{j'}; \theta), j \neq j'\}$. See also the discussion before Proposition A.1.2.

[40] In so doing, we pursue a two-step approach for estimating the choice probabilities, such as Bajari et al. [12]. See for example Kasahara and Shimotsu [77] for an alternative approach.

Lastly, the $P_{h_j,n_j,n^{(j)}}(m)$'s, from equation A.9 - that is, the probability that a worker matches to a firm of type m when they choose education level h_j, n_j other workers choose h_j, and $n^{(j)}$ firms prefer h_j - can be simulated for fixed values of θ and β. Proposition A.1.2 and Proposition A.1.3 show how these can be represented using probabilities involving order statistics. Under Assumption 1.3.2 (b), we can construct $\hat{P}_{h_j,n_j,n^{(j)}}(m)$'s by averaging functions of simulated draws of beta-distributed random variables (in particular, see Corollaries A.1.1 and A.1.2, which follow the order statistic result in Lemma A.1.3).

Once we have estimated $\hat{f}_j(\theta,\beta)$ for each education level, we may use the specification of the wage from equation 1.8 to write the expected wage as

$$\hat{\omega}_{ji}(\theta,\beta) = \tau\hat{f}_j(\theta,\beta) + (1-\tau)g(H_i,X_i;\theta). \tag{1.16}$$

When there are two choices ($J=2$), the worker chooses high education ($h_j=1$) if and only if

$$U^*_{1i} - U^*_{i0} > 0.$$

Under the assumption that ε_i's follow the extreme value distribution (Assumption 1.3.2), the probability that worker i chooses high education can be written as

$$\hat{p}_i(\theta,\beta) = \frac{\exp(\hat{\omega}_{1i}(\theta,\beta) - \hat{\omega}_{0i}(\theta,\beta))}{1+\exp(\hat{\omega}_{1i}(\theta,\beta) - \hat{\omega}_{0i}(\theta,\beta))}.$$

Since the covariates $\{X_i\}_{i=1}^n$ are iid we can write the joint likelihood as the product of the marginal likelihoods. We can then define the estimator of θ (for a fixed value of β) as the minimizer of the standard logit likelihood function:[41]

$$\ln L_n(\theta,\beta) = -\sum_{i=1}^{n}\left(h_i \ln \hat{p}_i(\theta,\beta) + (1-h_i)\ln(1-\hat{p}_i(\theta,\beta))\right).$$

[41] When β is fixed, maximizing the likelihood by computing the $\hat{f}_j(\theta,\beta)$'s for each candidate value of θ can be slow. The following strategy can be used to estimate θ for fixed β more quickly provided that the support of K is not too large. First, note that θ enters $\hat{f}_j(\theta,\beta)$ only through the set of firm types that prefer education level h_j, $\hat{M}^+_j(\theta)$. Given our assumptions on the production function and firm preferences, $\hat{M}^+_j(\theta)$ must take one of $M+1$ possible values. Therefore, for fixed β , we can avoid simulating $\hat{f}_j(\theta,\beta)$ for each candidate value of θ by pre-allocating the $\hat{q}_j(\theta)$'s and $P_{h_j,n_j,n^{(j)}}(m)$'s for each of the $M+1$ cases for $\hat{M}^+_j(\theta)$. It then suffices to evaluate $\hat{M}^+_j(\theta)$, select the appropriate dimension of the array of terms, then assemble the terms according to equation A.9.

1.3.3 Matching Probabilities

In this section, we consider the role of frictions, or the magnitude of β relative to the variance of η, in shaping matching patterns between workers and firms. Note that these frictions play no role in determing firm preferences, or which firm types prefer high education.[42] Nevertheless, because the frictions do affect sorting patterns, they are of considerable importance to workers when they decide how much to invest in education.

In the following example, we will suppose that that the set of firms that prefer education level h_j, M_j^+, contains at least two types of firms, m and $\tilde{m}$ with $k_m \neq k_{\tilde{m}}$. Suppose we fix N_j, the number of workers who chose education level h_j, at some n_j and we fix $N^{(j)}$, the number of firms who prefer highly-educated workers at some $n^{(j)}$ such that $n_j + 1 < n^{(j)}$. In this situation, there are strictly more firms who prefer type h_j workers than there are workers of this type. Let $\kappa = n^{(j)} - n_j + 1$, and denote $p_{m\kappa} \equiv P(v_m > v_{(\kappa)})$ for each m in M_j^+. Proposition A.1.2 says that the difference in the probability of matching to a type $\tilde{m}$ versus a type m firm at these values of n_j and $n^{(j)}$ in such a situation is given by

$$\left(p_{\tilde{m}\kappa} - p_{m\kappa}\right) q_{\tilde{m}}^+ / c_\kappa + p_{m\kappa} \left(q_{\tilde{m}}^+ - q_m^+\right) / c_\kappa, \tag{1.17}$$

with

$$c_\kappa \equiv \sum_{m \in M_j^+} p_{m\kappa} q_m^+,$$

where $q_m^+ = q_m / \sum_{m \in M_j^+} q_m$. Under Assumption 1.3.2, the case of $\beta = 0$ gives us that $p_{mk} = p_{\tilde{m}k}$, implying that the first term in the parentheses of equation 1.17 is zero. This means that when matching frictions are highest (i.e., when $\beta = 0$), the difference in the probability of matching to one type of firm that prefers h_j over another is captured by the relative prevalence of those types of firms in the economy.

In the case that $\beta > 0$, under Assumption 1.3.2, $p_{\tilde{m}\kappa} - p_{m\kappa}$ becomes larger as $k_{\tilde{m}} - k_m$ becomes larger. This means that higher capital firms have a better chance of matching with the high education workers when $\beta > 0$. On the other hand, in

[42] We discuss the role of firm preferences on matching patterns at the end of Section 1.2.2.

the case that $n_j + 1 > n^{(j)}$ (i.e., h_j is demanded by fewer firms than there are in the economy), then the above probabilities are independent of firm capital and β once again depend solely on the relative prevalence of the each type of firm.

1.4 Analysis of a Labour Matching Market in Canada

1.4.1 Background

In this section, we investigate the role of a labour matching technology in shaping education and wage patterns for the Canadian economy. Since the 1980's, many scholars studying the Canadian economy have focused on the rise of wage inequality, contrasting it to similar trends in the United States and Western Europe (Fortin et al. [54], Lemieux [83], Saez and Veall [104]). To explain this recent growth in wage inequality, some have emphasized forces on the demand side of the labour market, such as increasing wage premia for highly-skilled workers (Boudarbat et al. [24]), and a declining demand for jobs in the middle of the skill distribution (Green and Sand [63]). Another set of explanations emphasizes the role of institutional changes and government policy; namely, the effect of minimum wages and falling unionization rates (Fortin et al. [54], Lemieux [83]). One less-explored cause of this inequality may lie in the underlying process by which firms find workers to hire. This chapter emphasizes such a channel, focusing on the role of the labour market matching mechanism in recent wage patterns in Canada. As in the study of the German labour market in Card et al. [34], this channel considers the role that sorting patterns play in wage patterns, with a particular focus on the process by which firms find workers to hire.

1.4.2 Data

The matched employer-employee data we consider come from the Workplace Employee Survey (WES) of Statistics Canada (Statscan). WES is a longitundinal survey of Canadian firms and the workers they employ. WES allows researchers to study how the characteristics and outcomes of workers and firms are related. Thus, WES goes beyond other surveys that track only one side of the market, such Labour Force Survey (LFS) in the case of workers, or the Longitudinal Employment Analysis Program (LEAP) in the case of firms. Furthermore, the WES allows, in more detail than in previous surveys, to understand how firms adopted new technology and what the impacts of this was (Statscan). The WES is especially rich in terms of information concerning worker-firm bargaining, outside options, and technology use. As the 2006 release only contains employer data, I only consider WES panels for the years 1999-2005.

WES only collects data on firms and workers in Canadian provinces whose information was obtainable from Statcan's Business Register. The target population of the study was all non-governmental firms aside from agricultural and religious organizations. Furthermore, the focus was only on firms that hired more than one worker (who was not the owner or the employer). A firm employee is defined as a person associated with that firm who is working or on paid leave in March of the survey year who receives a T-4 slip from Canada Revenue Agency.

The workplace component of WES was conducted from 1999-2006. The firms were followed throughout the course of the study. Every two years, a sample of firms which are new to the Business Register are added to the base sample. The employee component of WES was conducted from 1999-2005. In each workplace survey firm that employs more than four workers, up to 24 workers are randomly sampled. All firms with fewer than four workers are included in the sample. Workers are only followed for two years in the workplace survey. For this reason, every second year, workers are resampled from the firms.

WES data has been used by other researchers. Dionne and Dostie [46] use WES data from 1999-2002 to study the impact of work arrangements on employee absenteeism. Dostie and Jayaraman [47] investigate the role of computer use on firm productivity gains. Pendakur and Woodcock [98] study the extent to which immigrant and minority access to high-paying jobs is determined by barriers to becoming hired at high-paying firms.

The WES (1999-2006) time frame coincides with a period of somewhat modest wage inequality (See A.1.7). In Figure A.1.7, we use the WES data to plot the difference between the 99th and 50th quantiles of total hourly wages for all workers in the WES sample. A comparable pattern can be seen in the shares of market income accruing to the top 1% of recipients (Veall [110]).

1.4.3 Model Estimates and Counterfactuals

In this section, we explore the evolution of the matching technology and preferences in two industries from the WES sample: Secondary Products Manufacturing (WES industry 4), and the Finance and Insurance (WES industry). Table A.3 reports results for the matching technology for a subgroup of higher skilled workers: namely, managers (WES occupation category 1) and professionals (WES occupation category 2), while tables A.4 and A.5 report preference estimates for the same subgroup of workers for the manufacturing and finance industries respectively. The estimates of preferences are reported at the minimum distance estimate of β. We consider the two specifications for the expected wage equation 1.8 that are found to behave reasonably in the simulation studies of Section A.1.3. Specification 1 uses the production function where worker and firm types are multiplicative while the production function in Specification 2 is additive. The results in this section provide similar insights on parameter inference. However, the results from A.1.7 illustrate the importance consequences of production function interactions in wage inequality.

The results for the matching technology suggest a period of low frictions in 1999-2000, followed by an increase in frictions. Whereas the frictions remain high towards the end of the sample in the finance industry, the frictions fall in the manufacturing industry towards the end of the sample (2004-2005). Note that the standard errors are so small in general that we report confidence intervals for β using the plug-in values of $\hat{\theta}$ (that is, we are not strictly taking into account estimation error of $\hat{\theta}$ into account).

In the preference estimates for both industries, we typically estimate θ_2 - the coefficient on female$_i$ in the worker's outside option function - to be negative. The estimated coefficients on θ_3 (marital status), and θ_4 (number of dependent children) are less conclusive. In both industries, θ_1 is found to be largest towards the end of the sample - 2004 in the case of the manufacturing industry, 2005 in the finance industry. The production technology appears more stable in the finance industry than it does in the manufacturing industry over time.

In section A.1.7, we use the structural model developed in this chapter along with the structural estimates of section to A.1.6 to generate statistics from two key counterfactual distributions of interest: wages and education. We consider the counterfactual implications of different (in-sample) estimated levels of matching frictions. For example, we can see what level of inequality would have prevailed in 1999 if the matching frictions had been as low as they were in 2005. Counterfactuals education levels in the two specifications are generated in a similar way as the outcome variables generated in the Monte Carlo study from in Section A.1.3, and the wage is generated using these values along with the simulated matchings (that involve iid draws of η). Here, of course, we use the relevant covariates, firm capital data, and parameter estimates for each of the cell in the tables. Overall, the results highlight the important role that matching technology and the production complementarities play in educational decisions and wage patterns.

Tables A.6 and A.7 shows results for counterfactual education levels. Production complementarities lead to higher investment in education in both cases, but appear to matter more in the finance industry. For instance, in 1999, the effect of switching to a multiplicative production function from an additive one increases the equilibrium investment in education by about 2% in the manufacturing industry but by about 5% in finance industry. We also see the substantial role that both preferences and the matching technology play in the decision to obtain higher education. In the manufacturing sector in 1999 (a high β year), the effect of switching to the matching technology from 2001 causes a fall in the equilibrium probability of attending college by roughly 8%. Overall, however, there is evidence that - taken together - changes to preferences (including the parameter in the production function) - matter much more to the worker's college decision than changes in the matching frictions. For example, in the year 2001, the effect of switching to 1999's preference parameter is a fall in the probability of investing in education of almost 20%.

Tables A.10 and A.11 report counterfactual (weighted) Gini coefficients for the WES sample years along with two counterfactual levels: maximal frictions ($\beta = 0$) and very low frictions ($\beta = 5$). The Gini coefficients in the row $\hat{\beta}_{year}$ were simulated from the equilibrium of the model taking the exogenous variables and preference estimates from that year.

In Tables A.10 and A.11 we see that in both industries, inequality is typically much higher in the case with production complementarities (Specification 1). In the finance industry in Specification 2, the effect of lowering matching frictions raises wage inequality in every year. In this case, the sorting effect raises inequality and dominates the inequality-lowering effects of a greater supply of highly educated workers. In other cases, however, the effect is ambiguous. In Specification 1 in the manufacturing industry, the level of inequality at the estimated value of the frictions is lower than at the counterfactual levels for most years (except 2002). For example, in 2005 the simulated Gini is 0.222 and the investment in education is 77%. This rises to 0.2507 (education investment 76%) when information frictions are highest and 0.2452 (education investment 85%) when frictions are lowest. The opposite is the case in Specification 1 in the finance industry, where the level of inequality at the estimated value of the frictions is higher than at the counterfactual levels for each year (except 1999).

1.5 Conclusion

This chapter presents an empirical strategy for studying wages and education in a labour market where the decisions of workers matter in the matching process. In particular, I perform inference on a labour market matching technology using matched employer-employee data. I demonstrate the feasibility of my approach in the case that the worker faces a choice between two education levels.

The methodology I develop in this chapter can be extended in a number of useful directions. Although the data used in this chapter did not include information on firms' profit, Bartolucci and Devicienti [14] have shown that such data is useful for investigating sorting. Another natural extension of the current setup is to consider the role that heterogeneity in the worker's bargaining strength plays in driving wage variation.

A number of inriguing extensions to the model would prove much more challenging. One limitation of the current approach is its reliance on cross-sectional variation alone for inference. In effect, useful information concerning unemployment and job-to-job transitions by workers is unused in my framework.

This chapter has also demonstrated how the decision to invest in education - and wage inequality - is sensitive to the presence of a particular source of matching frictions in the economy. Although firm capital is exogenous in this chapter, the role of information frictions on capital accumulation in an extended framework could be a fruitful way to study not only wage inequality, but also economic growth.

Chapter 2

Estimating Local Interactions Among Many Agents Who Observe Their Neighbors

2.1 Introduction

Interactions between agents - for example, through personal or business relations - generally lead to their actions being correlated. In fact, such correlated behaviors form the basis of identifying and estimating peer effects, neighborhood effects, or more generally social interactions in the literature. (See Blume et al. [20] and Durlauf and Ioannides [49] for a review of this literature.)

Empirical modeling becomes nontrivial when one takes seriously the fact that people are often connected directly or indirectly on a large complex network, observing some others' types, and that the econometrician observes only a small fraction of those on the network. Furthermore, strategic environments are highly heterogeneous across agents as each agent occupies a nearly "unique" position in the network. Information sharing potentially creates a complex form of cross-sectional dependence among the observed actions and yet the econometrician rarely has precise information about the actual network on which people observe other people.

The main contribution of this chapter is to develop a tractable empirical model of linear interactions among agents with the following two major features. First, assuming a large game on a complex exogenous network, the empirical model allows the agents not to observe the full network, but to observe only part of the types of their neighbors.[1]

Second, our model explains strategic interdependence among agents through correlated observed behaviors. In this model, the locality of cross-sectional dependence among the observed actions reflects the locality of strategic interdependence among the agents. Most importantly, unlike most incomplete information game models in the literature, our set-up allows for *information sharing on unobservables*, i.e., each agent is allowed to observe his neighbors' payoff relevant signals that are not observed by the econometrician.

Third, the econometrician does not need to observe the whole set of players in the game for inference. It suffices that he observes many (potentially) non-random samples of local interactions. The inference procedure that this chapter proposes is asymptotically valid *independently of* the actual sampling process, as long as the sampling process satisfies certain weak conditions. Accommodating a wide range of sampling processes is useful because random sampling is rarely used for the collection of network data, and a precise formulation of the actual sampling process is often difficult in practice.

A standard approach to model interactions among agents is to model them as a game and use equilibrium strategies from the game to obtain predictions and testable implications. Such an approach is cumbersome in our set-up. Since a particular realization of any agent's type affects all the other agents' actions in equilibrium through a chain of information sharing, each agent needs to form a "correct" belief about the entire information graph. Apart from such an assumption being highly unrealistic, it also implies that predictions from an equilibrium that the econometrician uses to form testable implications generally involve all the players in the game, when it is often the case that only part of the players are observed in practice. Thus an empirical analysis which regards the players in the sample as coincident with the actual set of players in the game will suffer from lack of

[1]For example, a recent paper by Breza et al. [27] documents that people in a social network has a substantial lack of knowledge on the network, and that the violation of this assumption may have significant implications in the predictions of the model.

external validity when his target "population" is the original large game involving much more players than those in the sample.

Instead, this chapter adopts an approach of behavioral modeling, where it is assumed that each agent, not knowing fully the information sharing relations, optimizes according to his simple beliefs about other players' strategies. The crucial part of our behavioral assumption is a primitive form of *belief projection* which says that each agent, not knowing who his payoff neighbors observe, projects his own beliefs about other players onto his payoff neighbors. More specifically, if agent i gives more weight to agent j than to agent k, agent i believes that each of his payoff neighbor s does the same in comparing agents j and k.

Belief projection in this chapter is a variant of inter-personal projection studied in behavioral economics. A related behavioral concept is projection bias of Loewenstein et al. [86] which refers to the tendency of a person projecting his own current taste to his future taste. See also Van Boven et al. [109] who reported experiment results on the interpersonal projection of tastes onto other agents. Since formation of belief is often tied to the information set the agent has, belief projection is closely related to information projection in Madarász [88] who focuses on the tendency of a person projecting his information to other agents' information. The main difference here is that our focus is to formulate the assumption in a way that is useful for inference using observational data on actions on a network.

We show that our primitive form of belief projection yields an explicit form of the best linear response which has intuitive features. For example, the best linear response is such that each agent i gives more weights to those agents with a higher *local centrality* to him, where the local centrality of agent j to agent i is defined to be high if and only if a high fraction of agents from those whose actions affect agent i's payoff have their payoffs affected by agent j's action. Also, each agent's action responds to a change in his own type more sensitively when there are stronger strategic interactions, due to what we call the *reflection effect*. The reflection effect of player i captures the way player i's type affects his own action through his payoff neighbors whose payoffs are affected by player i's types and actions.

Furthermore, the best linear response from the belief assumption provides a testable implication for information sharing on unobservables in data. The main idea is as follows. When the agents are strategically interdependent, the best linear response gives a linear reduced form for observed actions where the cross-sectional correlation of residuals indicates information sharing on unobservables. Hence as this chapter shows, using cross-sectional correlation of residuals, one can test for the role of information sharing on unobservables.

One might wonder how close the predictions from our behavioral model is to the predictions from an equilibrium model. For this we consider a simple linear interactions model as a complete information game where one can compute the equilibrium explicitly. The equilibrium strategies are given in a primitive form of a spatial autoregressive model. We compare the network externality from our behavioral model and that from the complete information game model using simulated graphs, one from Erdös-Rényi random graphs and the other from a scale-free random graph generation of Barabási-Albert. In both cases, it is shown that both models have similar predictions when the payoff externality parameter is less than or equal to 0.5. However, when it is close to one, the network externality becomes much higher in the equilibrium model than in the behavioral model. This is because while strong local interactions induce global cross-sectional dependence in the equilibrium model due to extensive information transmission, it does not in our behavioral model. Also, as the network size increases, the network externality from our behavioral model changes more stably than that from the equilibrium strategies from a complete information game.

We investigate the finite sample properties of asymptotic inference through Monte Carlo simulations using various payoff graphs. The results show reasonable performance of the inference procedures. In particular, the size and the power of the test for the strategic interaction parameter work well in finite samples. We also apply our method to an empirical application of decisions of municipalities on state presence revisiting the study by Acemoglu et al. [4]. We consider an incomplete information game model which permits information sharing. The fact that our best linear responses explicitly reveal the local dependence structure means that it is unnecessary to separately correct for spatial correlation following, for example, the procedure of Conley [41].

The literature of social interactions often look for evidence of interactions through correlated behaviors. For example, linear interactions models investigate correlation between Y_i and the average of outcomes over agent i's neighbors. See for example Manski [90], De Giorgi et al. [45], Bramoullé et al. [25] and Blume et al. [21] for identification analysis in linear interactions models, and see Calvó-Armengol et al. [29] for an application in the study of peer effects. Goldsmith-Pinkham and Imbens [61] considers nonlinear interactions on a social network and discusses endogenous network formation. These models often assume that we observe many independent samples of such interactions, where each independent sample constitutes a game which contains the entire set of the players in the game.

In the context of a complete information game, linear interactions models on a large social network can generally be estimated without assuming independent samples. The outcome equations frequently take the form of spatial autoregressive models which have been actively studied in the literature of spatial econometrics. (Anselin [9]) A recent study by Johnsson and Moon [74] consider a model of linear interactions on a large social network which allows for endogenous network formation. Developing inference on a large game model with nonlinear interactions is more challenging. See Menzel [94], Xu [114], Song [107], Xu and Lee [115], and Yang and Lee [116] for a large game model of nonlinear interactions. This large game approach is suitable when the data set does not have many independent samples of interactions. One of the major issues in the large game approach is that the econometrician often observes only part of the agents in the original game.[2]

Our approach of empirical modeling is also based on a large game model which is closer to the tradition of linear interactions models in the sense that our approach attempts to explain strategic interactions through correlated behaviors among neighbors. In our set-up, the cross-sectional dependence of the observed actions is not merely a nuisance that complicates asymptotic inference; it provides the very piece of information that reveals the strategic interdependence among agents. The correlated behaviors also arise in equilibrium in models of complete information games or games with types that are either privately or commonly observable.

[2] Song [107], Xu [114], Johnsson and Moon [74], Xu and Lee [115] and Yang and Lee [116] assume observing all the players in the large game. In contrast, Menzel [94] allows for observing i.i.d. samples from the many players, but assumes that each agent's payoff involves all the other agents' actions exchangeably.

(See Bramoullé et al. [25] and Blume et al. [21].) However, as emphasized before, such an approach can be cumbersome in our context of a large game primarily because the testable implications from the model typically involve the entire set of players, when in many applications the econometrician observes only a small subset of the players in the large game. After finishing the first draft of the working paper that forms the basis of this chapter, we learned of a recent paper by Eraslan and Tang [53] who model the interactions as a Bayesian game on a large network with private link information. Like this chapter, they permit the agents not to observe the full network, and show identification of the model primitives adopting a Bayesian Nash equilibrium as a solution concept. One of the major differences of this chapter from their paper is that this chapter permits information sharing on unobservables, so that the actions of neighboring agents are potentially correlated even after controlling for observables.

A departure from the equilibrium approach in econometrics is not new in the literature. Aradillas-Lopez and Tamer [10] studied the implications of various rationality assumptions for identification of the parameters in a game. Unlike their approach, our focus is on a large game where many agents interact with each other on a single complex network, and, instead of considering all the beliefs which rationalize observed choices, we consider a particular set of beliefs that satisfy a simple rule and yield an explicit form of best linear responses. (See also Goldfarb and Xiao [60] and Hwang [73] for empirical research adopting behavioral modeling for interacting agents.)

This chapter is organized as follows. In Section 2, we introduce an incomplete information game of interactions with information sharing. This section derives the crucial result of best linear responses under simple belief rules. In this section, we discuss the issue of external validity of network externality comparing two simple interactions models: a complete information game with equilibrium strategies and our behavioral model. Section 3 focuses on econometric inference. This section presents inference procedures, explains a situation where we can measure the role of information sharing on unobservables and compares our approach with a standard linear-in-means model. Section 4 investigates the finite sample properties of our inference procedure through a study of Monte Carlo simulations. Section 5 presents an empirical application on state capacity among municipalities. Section

6 concludes.

2.2 Strategic Interactions with Information Sharing

2.2.1 A Model of Interactions with Information Sharing

Strategic interactions among a large number of information-sharing agents can be modeled as an incomplete information game. Let N be the set of a finite yet large number of players. Each player $i \in N$ is endowed with his type vector (T_i, η_i), where η_i is a private type and T_i a sharable type. As we will elaborate later, information η_i is kept private to player i whereas T_i is observed by his "neighbors" which we define later. Throughout this chapter, we set $T_i = (X_i', \varepsilon_i)'$, where X_i is the vector of characteristics of player i that are observed by the econometrician, and ε_i the unobserved characteristic of player i. Thus *the model permits information sharing on unobservables* ε_i. This feature in fact makes a significant departure from many existing incomplete information interactions models which assume that variables that the econometrician observes are public among the agents whereas the variables that the econometrician does not observe are kept private among themselves. (e.g. Blume et al. [21])

To capture the strategic interactions among players, let us introduce an *undirected* graph $G_P = (N, E_P)$, where E_P denotes the set of edges ij, $i, j \in N$ with $i \neq j$ and each edge $ij \in E_P$ represents that the action of player i affects player j's payoff.[3] We denote $N_P(j)$ to be the G_P-neighborhood of player j, i.e., the collection of players whose actions affect the payoff of player j:

$$N_P(j) = \{i \in N : ij \in E_P\},$$

and let $n_P(j) = |N_P(j)|$. We define $\overline{N}_P(i) = N_P(i) \cup \{i\}$ and let $\overline{n}_P(i) = |\overline{N}_P(i)|$.

Player i choosing action $y_i \in \mathscr{Y}$ with the other players choosing $y_{-i} = (y_j)_{j \neq i}$ obtains payoff:

$$u_i(y_i, y_{-i}, T, \eta_i) \quad = \quad y_i \left(X_{i,1}' \gamma_0 + \tilde{X}_{i,2}' \delta_0 + \beta_0 \tilde{y}_i + \varepsilon_i + \eta_i \right) - \frac{1}{2} y_i^2,$$

[3] A graph $G = (N, E)$ is undirected if $ij \in E$ whenever $ji \in E$ for all $i, j \in N$.

where $T = (T_i)_{i \in N}$, $X_{i,1}$ and $X_{i,2}$ are subvectors of X_i,

$$\tilde{X}_{i,2} = \frac{1}{n_P(i)} \sum_{k \in N_P(i)} r_{ik} X_{k,2}, \text{ and } \tilde{y}_i = \frac{1}{n_P(i)} \sum_{k \in N_P(i)} r_{ik} y_k,$$

if $N_P(i) \neq \varnothing$, and $\tilde{X}_{i,2} = 0$ and $\tilde{y}_i = 0$ otherwise. The factor r_{ik} measures the "relative weight" of individual k in the network from the viewpoint of i. In this chapter, we consider two specifications.

$$\begin{aligned} \text{Specification A} \quad &: \quad r_{ik} = 1, \text{ for all } i,k \in N. \qquad (2.1) \\ \text{Specification B} \quad &: \quad r_{ik} = \overline{n}_P(k)/\overline{n}_P(i), \text{ for all } i,k \in N. \end{aligned}$$

The simple choice $r_{ik} = 1$ gives equal weight to every other agent, but the choice of $r_{ik} = \overline{n}_P(k)/\overline{n}_P(i)$ give more weights to those who have more edges with others relative to agent i. Thus the payoff depends on other players' actions and types only through those of his G_P-neighbors. We call G_P the *payoff graph*.

The parameter β_0 measures the payoff externality among agents. In the terminology of Manski (1993), δ_0 captures the exogenous effect and β_0 the endogenous effect of social interactions. As for β_0, we make the following assumption:

Assumption 2.2.1. $0 \leq |\beta_0| < 1$.

This assumption is often used to derive a characterization of a unique pure strategy equilibrium in the literature. (See e.g. Bramoullé et al. [25] and Blume et al. [21] for its use.) When $\beta_0 > 0$, the game is called a game of strategic complements and, when $\beta_0 < 0$, it is called a game of strategic substitutes.

Let us introduce information sharing relations in the form of a *directed* graph (or a network) $G_I = (N, E_I)$ on N so that each ij in E_I represents the edge *from* player i *to* player j, where the presence of edge ij joining players i and j indicates that T_i is observed by player j. Hence the presence of an edge ij between agents i and j represents information flow from i to j. This chapter calls graph G_I the *information graph*. For each $j \in N$, define

$$N_I(j) = \{i \in N : ij \in E_I\},$$

that is, the set of G_I-neighbors observed by player j.[4] Also write

$$\overline{N}_I(i) = N_I(i) \cup \{i\},$$

i.e., the G_I-neighborhood of i including i himself. We define $\overline{n}_I(i) = |\overline{N}_I(i)|$.

In this chapter, we do not assume that each agent knows the whole information graph G_I and the payoff graph G_P. To be precise about each agent's information set, let us introduce some notation. For each $i \in N$, we set $\overline{N}_{P,1}(i) = \overline{N}_P(i)$ and $\overline{N}_{I,1}(i) = \overline{N}_I(i)$, and for $k \geq 2$, define recursively

$$\overline{N}_{P,k}(i) = \bigcup_{j \in \overline{N}_P(i)} \overline{N}_{P,k-1}(j), \text{ and } \overline{N}_{I,k}(i) = \bigcup_{j \in \overline{N}_I(i)} \overline{N}_{I,k-1}(j).$$

Thus $\overline{N}_{P,k}(i)$ denotes the set of players which consist of player i and those players who are connected to player i through at most k edges in G_P, and similarly with $\overline{N}_{I,k}(i)$. Also, define $N_{P,k}(i) = \overline{N}_{P,k}(i) \setminus \{i\}$ and $N_{I,k}(i) = \overline{N}_{I,k}(i) \setminus \{i\}$. For each $k \geq 1$, let $\mathscr{N}_{i,k-1}$ be the σ-field generated by $\overline{N}_{P,k+1}(i)$, $\overline{N}_I(i)$ and some additional information $\mathscr{C}_i$ which potentially causes correlation between types across different players. (We will explain $\mathscr{C}_i$ later.) That is, for $k \geq 1$,

$$\mathscr{N}_{i,k-1} = \sigma(\overline{N}_{P,k+1}(i), \overline{N}_{P,k}(i), ..., \overline{N}_{P,2}(i), \overline{N}_I(i)) \vee \mathscr{C}_i,$$

where $\vee$ between two σ-fields is the smallest σ-field among those which contain the two σ-fields. Define for each $k \geq 0$,

$$\mathscr{I}_{i,k} = \sigma(T_{\overline{N}_I(i)}, \eta_i) \vee \mathscr{N}_{i,k},$$

where $T_{\overline{N}_I(i)} = (T_j)_{j \in \overline{N}_I(i)}$. We use $\mathscr{I}_{i,k}$ to represent the information set of agent i. For example, when agent i has $\mathscr{I}_{i,1}$ as his information set, it means that agent i knows the set of agents whose types he observes (i.e., $N_I(i)$), the set of agents j whose actions affect his payoff (i.e., $N_{P,1}(i)$) and the set of agents whose actions affect the payoff of his G_P-neighbors j (i.e., $N_{P,2}(i)$), and the sharable types of his G_I-neighbors (i.e., $T_{N_I(i)}$) and his own private signal η_i.

[4] More precisely, the neighbors in $N_I(j)$ are called *in-neighbors* and $n_I(i) = |N_I(j)|$ *in-degree*. Throughout this chapter, we simply use the term neighbors and degrees, unless specified otherwise.

Throughout the chapter, *it is not assumed that any agent i knows $N_I(k)$ for any of his G_P-neighbors k.* In other words, there might be some G_P-neighbor k who may observe other agents that agent i does not observe, and agent i does not know who such G_P-neighbor k is or who those other agents player k observes are.

Regarding the joint distribution of the profile of sharable types T, we make the following assumption:

Assumption 2.2.2. For each $i \in N$, $T_{N \setminus \overline{N}_I(i)}$ and $T_{\overline{N}_I(i)}$ are conditionally independent given $(G_P, \overline{N}_I(i))$ and $\mathscr{C}$, where

$$\mathscr{C} = \vee_{i \in N} \mathscr{C}_i.$$

This assumption allows the individual types to be correlated unconditionally. Each player i has information $\mathscr{C}_i$ which can cause correlation between his type and other agents' types. For example, any two types T_i and T_j may contain a common signal which comes from a common observation by the two agents i and j.[5] Assumption 2.2.2 says that the sharable types between two non-neighbors in G_I are independent conditional on all such pieces of information $\mathscr{C}_i$.

The assumption permits the situation where the payoff network G_P is exogenously formed, for example, as a dyadic regression model degree heterogeity, a_i, with errors u_{ij}'s that are independent of ε_i's, η_j's, X_i's and a_i's. (See e.g. Graham [62].) In this case, if we set $\mathscr{C}_i = \sigma(X_i, a_i)$, Assumption 2.2.2 is reduced to that for each $i \in N$, $\varepsilon_{N \setminus \overline{N}_I(i)}$ and $\varepsilon_{\overline{N}_I(i)}$ are conditionally independent given $(G_P, \overline{N}_I(i), X, a)$, where $X = (X_i)_{i \in N}$ and $a = (a_i)_{i \in N}$.

2.2.2 Predictions from Rationality

Each player chooses a strategy that maximizes his expected payoff according to his beliefs. This provides predictions for their actions given their beliefs. For the sake of analytical facility, we assume throughout the chapter that each agent having

[5]The signal $\mathscr{C}_i$ may contain information accumulated from the past information obtained when the information sharing takes place over time, such as information used at the stage of forming information and payoff graphs G_I and G_P. The supplemental note contains details about an extended model where people share information over time and shows how this fits the current set-up in the chapter.

information set $\mathscr{I}_i = \mathscr{I}_{i,k}$ for some $k \geq 0$ chooses from a class of linear strategies:

$$s_i(\mathscr{I}_i) = \sum_{j \in \overline{N}_I(i)} w_{ij}' T_j + \eta_i,$$

where $T_i = [X_i', \varepsilon_i]'$ and w_{ij} denotes the nonstochastic vector of nonnegative numbers. We call w_{ij} *the weight given to player* j *by player* i. The weight vector w_{ij} summarizes the influence of player j on player i's decision making.

To characterize predictions from rationality, we introduce some notation. For $i, j, k \in N$, let w_{kj}^i denote the weight that player i believes that player k gives to player j. Then the strategy of player k as believed by player i is given as follows:

$$s_k^i(\mathscr{I}_k) = \sum_{j \in \overline{N}_I^i(k)} T_j' w_{kj}^i + \eta_k,$$

where $\overline{N}_I^i(k)$ denotes the set of players (including player k) who player i believes that player k observes. Given player i's strategy and his expected strategy of other players $s_{-i}^i = (s_k^i)_{k \in N \setminus \{i\}}$, the (interim) expected payoff of player i is defined as

$$U_i(s_i, s_{-i}^i; \mathscr{I}_i) = \mathbf{E}[u_i(s_i(\mathscr{I}_i), s_{-i}^i(\mathscr{I}_{-i}), T, \eta_i) | \mathscr{I}_i],$$

where $s_{-i}^i(\mathscr{I}) = (s_k^i(\mathscr{I}_k))_{k \in N \setminus \{i\}}$, $\mathscr{I}_{-i} = \vee_{k \neq i} \mathscr{I}_k$ and $T = (T_i)_{i \in N}$. A *best linear response* s_i^{BR} of player i corresponding to the strategies s_{-i}^i of the other players as expected by player i is a linear strategy such that for any linear strategy s_i,

$$U_i(s_i^{\mathsf{BR}}, s_{-i}^i; \mathscr{I}_i) \geq U_i(s_i, s_{-i}^i; \mathscr{I}_i), \text{ a.e.}$$

Under the assumptions of the model, the best linear responses can be shown to produce a map from beliefs to actions. To see this, first let

$$w^B = (w^1, ..., w^n)$$

be the belief profile of all the agents, where $w^i = (w_{kj}^i)_{k,j \in N}$. Then the rationality of agents (i.e., their choosing a best linear response given their beliefs) gives the following relation:

$$w = \mathscr{M} w^B,$$

where $w = (w_{ij})_{i,j \in N}$ corresponds to best responses and $\mathscr{M}$ is the best response operator which assigns a strategy profile (in terms of weights w_{ij}) to a given belief profile w^B.

Given our set-up of quadratic payoffs and linear strategies, we can make explicit the best response operator $\mathscr{M}$. To see this, given $w_j^i = (w_{kj}^i)_{k \in N}$, let us define

$$\mathscr{M}_i w_j^i = \frac{1}{n_P(i)} \sum_{k \in N_P(i)} r_{ik} w_{kj}^i 1\{j \in \overline{N}_I^i(k)\}.$$

Recall that player i's payoff is affected by his G_P-neighbors' actions. Hence player i perceives player j as important to him even if player j's action does not directly influence the payoff of player i, if player j's type is observed by and influences many of player i's G_P-neighbors. The expression $\mathscr{M}_i w_j^i$ captures this perceived importance of player j to player i that comes through player j's influence (as perceived by player i) on his G_P-neighbors.

Suppose that each agent i has information set $\mathscr{I}_{i,k}$ for some $k \geq 0$. Then the best response operator $\mathscr{M}$ is given by the following relations:

$$\begin{aligned} w_{ii,1} &= \gamma_0 + \beta_0 \mathscr{M}_i w_{i,1}^i, \\ w_{ii,\varepsilon} &= 1 + \beta_0 \mathscr{M}_i w_{i,\varepsilon}^i, \\ w_{ii,2} &= \beta_0 \mathscr{M}_i w_{i,2}^i, \end{aligned} \tag{2.2}$$

and for all $j \in N_I(i)$,

$$\begin{aligned} w_{ij,1} &= \beta_0 \mathscr{M}_i w_{j,1}^i, \\ w_{ij,\varepsilon} &= \beta_0 \mathscr{M}_i w_{j,\varepsilon}^i, \text{ and} \\ w_{ij,2} &= \begin{cases} \delta_0 r_{ij}/n_P(i) + \beta_0 \mathscr{M}_i w_{j,2}^i, & \text{if } j \in N_P(i), \\ \beta_0 \mathscr{M}_i w_{j,2}^i, & \text{if } j \in N_I(i) \setminus N_P(i), \end{cases} \end{aligned} \tag{2.3}$$

where we write $w_{ij} = (w'_{ij,1}, w'_{ij,2}, \varepsilon_{ij})'$, $w_{ij,1}, w_{ij,2}$ and $w_{ij,\varepsilon}$ being weights given by player i to player j's type components, $X_{i,1}, X_{i,2}$ and ε_{ij}.

In order to generate predictions, one needs to deal with the beliefs w^B. There are three approaches to model these beliefs. The first approach is an equilibrium approach where the beliefs w^B coincide with the actual weights implemented by the agents in equilibrium. The second approach uses rationalizability where all the linear strategies that are rationalizable given some belief w^B are in consideration. The third approach is a behavioral approach where one considers a set of simple behavioral assumptions on the beliefs w^B and focuses on the best linear responses to corresponding to these beliefs.

There are pros and cons among the three approaches. One of the main differences between the equilibrium approach and the behavioral approach is that the former approach requires the beliefs w^i_{-i} to be "correct" for all players i in equilibrium. However, since each player i generally does not know who each of his G_P-neighbors observes, a Bayesian player in a standard model with rational expectations would need to know the distribution of the entire information graph G_I (or at least have a common prior on the information graph commonly agreed upon by all the players) to form a "correct" belief given his information. Given a potentially complex form of G_P (partially observed in data) and that the econometrician rarely observes G_I with precision, producing a testable implication from this equilibrium model appears far from a trivial task.

The rationalizability approach can be used to relax this rational expectations assumption by eliminating the requirement that the beliefs be correct. The approach considers all the predictions that are rationalizable given *some* beliefs. However, in our context, the best response operator $\mathscr{M}$ depends on unknown parameters in general, and hence the set of predictions from rationalizability can potentially be large and may fail to produce sharp predictions that would be useful in practice.

As we explain later in detail, this chapter takes the third approach. We adopt a set of simple behavioral assumptions on players' beliefs which can be incorrect from the viewpoint of a person with full knowedge on the distribution of the information graph, yet useful as a rule-of-thumb guidance for an agent in a complex decision-making environment such as one in our model. As we shall see later, this approach can give a sharp prediction that is intuitive and analytically tractable.

2.2.3 Belief Projection and Best Linear Responses

In this chapter, we consider the following set of behavioral assumptions on the beliefs.

Condition BP (Belief Projection): (i) For each $i \in N$ and $k \in N_P(i)$,
(a) $w^i_{kk} = w_{ii}$,
(b) $w^i_{kj} = \tau^i_{kj} w_{ij}$ for all $j \in \overline{N}_I(i) \cap N^i_I(k)$ for some positive number τ^i_{kj}, where

$$\begin{aligned} \tau^i_{ki} &= 1/(r_{ik}\overline{n}_P(k)), \text{ and} \qquad (2.4) \\ \tau^i_{kj} &= 1/r_{ik}, \text{ for all } j \in N_I(i) \cap N^i_I(k), \text{ and} \end{aligned}$$

(ii) $w^i_{kj} = 0$ for all $j \notin \bar{N}_P(k)$.

As mentioned before, each player i does not know who his G_P-neighbors observe, and Condition BP describes a simple rule of belief formation in this environment. The main premise of Condition BP is that each agent *projects his own beliefs about himself and other players onto* his G_P neighbors. Condition BP (i)(a) says that each player i believes that the self-weight his G_P-neighbor k gives to himself is the same as the self-weight of player i himself. Condition BP (i)(b) says that player i's belief on his G_P neighbor k's weight to player j is formed in reference to his own weight to player j. This assumption says that each agent believes that his G_P neighbors follow the same ranking of other agents as he does. The belief projection is taken as a rule of thumb for each agent i who needs to form an expectation about his G_P-neighbors' actions when he does not know who his G_P-neighbors observe.

The specification of τ^i_{ki} in (2.4) reflects that player i believes that player k does not care much about player i's type in choosing an action if the player k has many G_P-neighbors. The specification of τ^j_{ki} in (2.4) says that each player i believes that the weight of each of his G_P-neighbors given to a G_P-neighbor j is $(1/r_{ik})w_{ij}$. For example, if $r_{ik} = \bar{n}_P(k)/\bar{n}_P(i)$, we have

$$w^i_{kj} = \frac{\bar{n}_P(i)}{\bar{n}_P(k)} w_{ij}.$$

Therefore, player i believes that when player k has more G_P-neighbors than he does, player k gives less weight to player j than he does. Not knowing who player

k observes, player i employs this rule-of-thumb belief regarding player k's weights given to other players.

Condition BP(ii) is concerned with player i's belief about the players that his G_P-neighbors observe. A standard approach in an incomplete information game with Bayesian players assumes that the players agree on a common prior on the entire information graph G_I. From this, each agent i derives his posterior on the G_I-neighbors of each of his G_P-neighbors. Instead, Condition BP(ii) states that player i simply considers only those players in $N_P(k)$ when he deliberates on those players whose action affects the payoff of player k. This is because while player i knows player k's G_P-neighborhood, he does not know player k's G_I-neighborhood.

Let us distinguish between different environments with different information structures of the game.

Definition 2.2.1. *(i) Each agent $i \in N$ having beliefs about the other players' strategies as in Condition BP and having information set $\mathscr{I}_i = \mathscr{I}_{i,0}$ with $\overline{N}_{P,2}(i) \subset \overline{N}_I(i)$ is said to be of simple type.*

(ii) Each agent $i \in N$ who believes that the other players are of simple type and has information set $\mathscr{I}_i = \mathscr{I}_{i,1}$ with $\overline{N}_{P,3}(i) \subset \overline{N}_I(i)$ is said to be of first order sophisticated type.[6]

The difference between the simple type and a sophisticated type lies not only in the difference in the rationality type but also in the information set. A first order sophisticated type agent knows who the neighbors of the neighbors of their neighbors in G_P (i.e., $\overline{N}_{P,3}(i)$) are, whereas a simple type agent knows only who the neighbors of their neighbors in G_P (i.e., $\overline{N}_{P,2}(i)$) are.

Regarding the sophistication of agents, we make explicit the following basic assumption which we assume throughout the chapter.

Assumption 2.2.3. The game is populated by agents with the same order of sophistication.

[6]One can also define a higher order sophisticated type, although this chapter does not fully elaborate on such a case. More specifically, for $k \geq 2$, each agent $i \in N$ who believes that the other players are of the $(k-1)$-th order sophisticated type and has information set $\mathscr{I}_i = \mathscr{I}_{i,k}$ with $\overline{N}_{P,k+2}(i) \subset \overline{N}_I(i)$ is said to be *of the k-th order sophisticated type*.

Different levels of reasoning for agents of the game are assumed in level k models which have received a great deal of attention as a behavioral model in the experiment literature. (See Chapter 5 of Camerer [30] for a review.) In these experiments, a simple type agent is often much simpler than those in our set-up, where the agent chooses an action without considering any strategic interdependence. In contrast, our simple type agent already considers strategic interdependence and forms a best linear response. On the other hand, the experiment literature of level-k models allows the agents to be of different rationality type in the same game. In our set-up which focuses on observational data, identification of the unknown proportion of each rationality type appears far from trivial. Hence in this chapter, we consider a game where all the agents have the same order of sophistication.

Our focus on linear strategies in combination with other assumptions gives an explicit form of best linear responses. For the expression, let us introduce some notation: for each $i \in N$ and $j \in \overline{N}_I(i)$,

$$\begin{aligned} c_{ij} &\equiv \frac{1}{n_P(i)} \sum_{k \in N_P(i)} 1\{j \in N_P(k)\}, \text{ if } i \neq j, \text{ and} \\ c_{ii} &\equiv \frac{1}{n_P(i)} \sum_{k \in N_P(i)} \frac{1\{i \in N_P(k)\}}{\bar{n}_P(k)} = \frac{1}{n_P(i)} \sum_{k \in N_P(i)} \frac{1}{\bar{n}_P(k)}, \end{aligned} \tag{2.5}$$

where the last equality follows due to G_P being undirected. Note that c_{ij} is the proportion of player i's G_P-neighbors whose payoffs are influenced by the type and action of player j. Hence c_{ij} represents the *local centrality* of player j to player i in terms of player j's influence on player i's G_P-neighbors. On the other hand, c_{ii} is the average of $1/\bar{n}_P(k)$ among player i's G_P-neighbors k whose payoffs are affected by player i's sharable type and action.

Using the explicit form of the best response operator $\mathscr{M}$ and Condition BP, we can derive the explicit form of best linear responses. The following theorem gives the form in the case where all the players are of simple type.

Theorem 2.2.1. *Suppose that Assumptions 2.2.1 - 2.2.3 hold and all the players are of simple type. Suppose further that for each $i \in N$ and $k \neq i$, $\mathbf{E}[\eta_k|\mathscr{I}_i] = 0$.*

Then each player i's best linear response s_i^{BR} *takes the following form:*

$$\begin{aligned} s_i^{\mathsf{BR}}(\mathscr{I}_i) &= \lambda_{ii}\left(\gamma_0' X_{i,1} + \varepsilon_i + \frac{\beta_0}{n_P(i)} \sum_{j \in N_P(i)} \lambda_{ij}(\gamma_0' X_{j,1} + \varepsilon_j)\right) \\ &+ \frac{1}{n_P(i)} \sum_{j \in N_P(i)} \lambda_{ij} \delta_0' X_{j,2} + \eta_i, \end{aligned}$$

where $\lambda_{ij} \equiv r_{ij}/(1-\beta_0 c_{ij})$.

The result in Theorem 2.2.1 shows multiple intuitive features. First, it shows that each player i's best linear response does not depend on the types of payoff-irrelevant agents whose types player i observes but whose actions do not affect player i's payoff. Note that agents indirectly connected to agent i in G_P can still shape the player's strategies through the local centralities c_{ij}. (Later, we also consider the case of sophisticated type, where the types of indirectly connected agents are permitted to influence the agent i's actions.)[7] Furthermore, observe that for $j \in N_P(i)$,

$$\begin{aligned} \frac{\partial s_i^{\mathsf{BR}}(\mathscr{I}_i)}{\partial x_{j,1}} &= \frac{\beta_0 r_{ij}}{n_P(i)(1-\beta_0 c_{ii})(1-\beta_0 c_{ij})} \gamma_0 \text{ and} \\ \frac{\partial s_i^{\mathsf{BR}}(\mathscr{I}_i)}{\partial x_{j,2}} &= \frac{\delta_0 r_{ij}}{n_P(i)(1-\beta_0 c_{ij})}, \end{aligned} \tag{2.6}$$

both of which measure the response of actions of agent i to a change in the observed type change of his G_P-neighbors. Hence, these quantities capture the *network externality* in the strategic interactions.

First, note that the network externality for agent i from a particular agent j decreases in the neighborhood size $n_P(i)$ of agent i. More importantly, the network externality for each agent i is different across i's and across their G_P neighbors j depending on their "importance" to agent i in the payoff graph. This is seen from the network externality (2.6) being an increasing function of agent j's local centrality to agent i, i.e., c_{ij}, when the game is that of strategic complements (i.e.,

[7]The local dependence of actions from best linear responses regardless of what values β_0 take in $(-1,1)$ is in contrast with the complete information version of the game, where a high value of β_0 makes the dependence close to be global.

$\beta_0 > 0$). In other words, the larger the fraction of agent i's G_P-neighbors whose payoff is affected by agent j's action, the higher the network externality of agent i from agent j's type change becomes. Therefore, in our model network externality is heterogeneous across agents, depending on the local feature of the payoff graph around each agent.

It is interesting to note that the network externality for agent i with respect to his own type $X_{i,1}$ has a factor $\lambda_{ii} \equiv r_{ii}/(1-\beta_0 c_{ii}) = 1/(1-\beta_0 c_{ii})$ which is increasing in c_{ii} when $\beta_0 > 0$. We call

$$\frac{1}{1-\beta_0 c_{ii}} - 1$$

the *reflection effect* which captures the way player i's type affects his own action through his G_P neighbors whose payoffs are affected by player i's types and actions. The reflection effect arises because each agent, in decision making, considers the fact that his type affects other G_P-neighbors' decision making. When there is no payoff externality (i.e., $\beta_0 = 0$), the reflection effect is zero. However, when there is a strong strategic interactions or when a majority of player i's G_P-neighbors have a small G_P-neighborhood (i.e., small $\bar{n}(k)$ in the definition of c_{ii} in (2.5)), the reflection effect is large. Note that for those agents whose c_{ii} the econometrician observes, the reflection effect is easily recovered once one estimates the payoff externality β_0.

Now let us turn to the case where the game is played among the first-order sophisticated players.[8]

Theorem 2.2.2. *Suppose that Assumptions 2.2.1 - 2.2.3 hold and that all the players are of first-order sophisticated type. Suppose further that for each $i \in N$ and $k \neq i$, $\mathbf{E}[\eta_k|\mathscr{I}_i] = 0$. Then each player i's best linear response $s_i^{\mathsf{BR.FS}}$ takes the following form:*

[8]For this result, we focus on Specification A (i.e., r_{ik}'s are set to 1).

$$
\begin{aligned}
s_i^{BR.FS}(\mathscr{I}_{i1}) = {} & \gamma_0 X_{i,1} + \varepsilon_i + \frac{\beta_0}{n_{P(i)}} \sum_{j \in N_P(i)} \lambda_{jj}(\gamma_0' X_{j,1} + \varepsilon_j) + \beta_0^2 \sum_{j \in \bar{N}_{P,2}(i)} \tilde{\lambda}_{ij}(\gamma_0' X_{j,1} + \varepsilon_j) \\
& + \delta_0 \tilde{X}_{i,2} + \delta_0' \beta_0 \sum_{j \in \bar{N}_{P,2}(i)} \bar{\lambda}_{ij} X_{j,2} + \eta_i.
\end{aligned}
$$

where,

$$
\begin{aligned}
\bar{\lambda}_{ij} &= \frac{1}{n_P(i)} \sum_{k \in N_P(i)} \frac{\lambda_{kj} 1\{j \in N_P(k)\}}{n_P(k)}, \\
\textit{and} \quad \tilde{\lambda}_{ij} &= \frac{1}{n_P(i)} \sum_{k \in N_P(i)} \frac{\lambda_{kj} 1\{j \in N_P(k)\}}{n_P(k)(1-\beta_0 c_{kk})}.
\end{aligned}
$$

Note that as compared to the case of the game with agents of simple type, the game with agents of the first order sophisticated type predicts outcomes with broader network externality. For example, in contrast to the case of simple type agents, the types of neighbors whose actions do not affect player i's payoff can affect his best response. More specifically, note that for $j \in N_{P,2}(i) \setminus N_P(i)$,

$$
\begin{aligned}
\frac{\partial s_i^{\mathsf{BR.FS}}(\mathscr{I}_i)}{\partial x_{j,1}} &= \beta_0^2 \gamma_0 \tilde{\lambda}_{ij} \text{ and} \\
\frac{\partial s_i^{\mathsf{BR.FS}}(\mathscr{I}_i)}{\partial x_{j,2}} &= \beta_0 \delta_0 \overline{\lambda}_{ij}.
\end{aligned}
$$

This externality from player j on player i is strong when c_{kj}'s are large for many $k \in N_P(i)$, i.e., when player j has a high local centrality to a large fraction of player i's G_P-neighbors.[9]

[9] Using the explicit form of the best response operator $\mathscr{M}$ and Condition BP, we can derive best linear responses in a game populated by agents of a higher order sophisticated type. As the sophistication of agents becomes of higher order, the network externality of each agent broadens to a wider set of agents. The derivation is easy but tedious algebraically, giving a more complex form of best linear responses. Hence, details are omitted in the chapter.

2.2.4 The External Validity of Network Externality

Through a simple model of linear interactions, we explore two issues of external validity. The first issue is about generalizing the results that come from a model with a smaller graph to the population with a larger graph. We see how sensitively the network externality changes as the network grows. If the sensitivity is not high, this supports the external validity of a model toward a larger graph. The second issue is about misspecification of behavioral assumptions. Here we set the benchmark (true) model to be a complete information model with equilibrium strategies, but assume that the econometrician adopts our behavioral model to make the analysis tractable. Then we explore how close the network externality from the behavioral model is to the true model of complete information game. Both models assume the same payoff function and the same payoff graph. For simplicity, we remove X_i's and η_i's. The main focus here is on the stability of the prediction of the network externalities as we progressively move from a small payoff graph to a large payoff graph. Let Y_i be the observed outcome of player i as predicted from either of the two game models.

The complete information game model assumes that every agent observes all the types ε_i's of other agents. This model yields the following equilibrium equation:

$$Y_i = \frac{\beta_0}{n_P(i)} \sum_{j \in N_P(i)} Y_j + \varepsilon_i.$$

Then the reduced form for Y_i's can be written as

$$y = (I - \beta_0 A)^{-1} \varepsilon,$$

where $y = (Y_1, ..., Y_n)'$, $\varepsilon = (\varepsilon_1, ..., \varepsilon_n)'$, and A is a row-normalized adjacency matrix of the payoff graph G_P, i.e., the (i, j)-th entry of A is $1/n_P(i)$ if $j \in N_P(i)$ and zero otherwise. Thus in the complete information equilibrium model, each Y_i is a linear combination of *all* ε_i's. The model implies that when β_0 is close to one (i.e., the local interaction becomes strong), the equilibrium outcome can exhibit extensive cross-sectional dependence.

On the other hand, our behavioral model (with specification A: $r_{ik} = 1$ in (2.1) and with the assumption that all the players are of simple type) predicts the outcomes in the following simple reduced form:

$$Y_i = \lambda_{ii}\left(\varepsilon_i + \frac{\beta_0}{n_P(i)} \sum_{j \in N_P(i)} \lambda_{ij}\varepsilon_j\right),$$

which comes from Theorem 2.2.1 without X_i's and η_i's. It is important to note that the two models have the same payoff with the same payoff externality parameter β_0. The only difference is the information set assumptions and the solution concepts of the game.

The parameter of interest is the average network externality:

$$\begin{aligned}
&\frac{1}{n}\sum_{i\in N}\frac{1}{n_P(i)}\sum_{j\in N_P(i)}\frac{\partial s_i^{\mathsf{BR}}(\mathscr{I}_i)}{\partial \varepsilon_j} \\
= &\begin{cases} \dfrac{1}{n}\displaystyle\sum_{i\in N}\frac{1}{n_P(i)}\sum_{j\in N_P(i)}[(I-\beta_0 A)^{-1}]_{ij}, & \text{from the equilibrium model} \\[2ex] \dfrac{\beta_0\lambda_{ii}}{n}\displaystyle\sum_{i\in N}\frac{1}{n_P(i)}\sum_{j\in N_P(i)}\lambda_{ij}, & \text{from the behavioral model,} \end{cases}
\end{aligned}$$

where $[(I-\beta_0 A)^{-1}]_{ij}$ denotes the (i,j)-th entry of the matrix $(I-\beta_0 A)^{-1}$.

Note that the network externalities depend only on β_0 and the payoff graph G_P. For the payoff graph G_P, we considered two different models for random graph generation. The first kind of random graphs are Erdös-Rényi (ER) random graph with the probability equal to $5/n$ and the second kind of random graphs are Barabási-Albert (BA) random graph such that beginning with an Erdös-Rényi random graph of size 20 with each link forming with equal probability 1/19 and grows by including each new node with two links formed with the existing nodes with probability proportional to the degree of the nodes.

For each random graph, we first generate a random graph of size 10,000, and then construct three subgraphs A,B,C such that network A is a subgraph of network B and the network B is a subgraph of network C. We generate these subgraphs as follows. First, we take a subgraph A to be one that consists of agents within distance k from agent $i=1$. Then network B is constructed to be one that consists of the

Table 2.1: The Characteristics of the Payoff Graphs

	Erdös-Rényi			Barabási-Albert		
	Network A	Network B	Network C	Network A	Network B	Network C
n	162.0	766.4	3067.4	432.2	2080.1	4663.5
d_{mx}	10.72	12.50	14.14	76.62	98.82	113.66
d_{av}	2.043	2.296	3.186	1.437	1.902	2.233

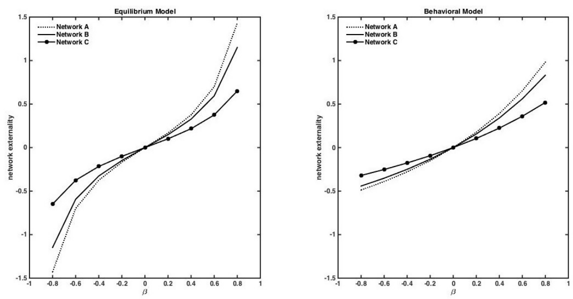
Notes: This table gives average characteristics of the payoff graphs, G_P, used in the simulation study, where the average was over 50 simulations. d_{av} and d_{mx} denote the average and maximum degrees of the payoff graphs.

neighbors of the agents in network A and network C is constructed to be one that consists of the neighbors of the agents in network B. For an ER random graph, we took $k = 3$ and for a BA random graph, we took $k = 2$. We repeated the process 50 times to construct an average behavior of network externality as we increase the network. Table 1 shows the average network sizes and degree characteristics as we move from Networks A, B to C.

First, we would like to see how sensitive the predicted average network externality becomes as we move across three networks of increasing sizes. The results are in Figures 2.1 and 2.2. Figure 2.1 captures the relation between β_0 and the average network externality for the case of ER graphs and Figure 2.2 captures that for the case of BA graphs. The left panel depicts the relation from the complete information equilibrium model and the right panel depicts the relation from the behavioral model.

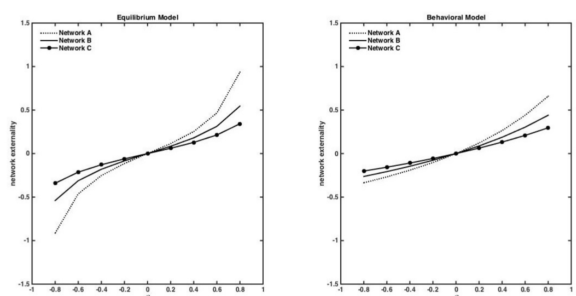
As shown in Figures 2.1-2.2, the predicted network externality from the behavioral model is less sensitive to the change of the networks than that from the equilibrium model. In particular, this contrast is stark when β_0 is close to 1. The main reason behind this contrast is that in the case of the equilibrium model, stronger local strategic interactions induce extensive cross-sectional dependence. This extensiveness will sensitively depend on the size and the shape of the network. On the other hand, the behavioral model limits the extent of the cross-sectional dependence even when β_0 is high. Hence the predicted network externality does not vary as much as the equilibrium model as we change the network. The result illustrates the point that our behavioral model translates local strategic interactions to

Figure 2.1: Network Externality Comparison Between Equilibrium and Behavioral Models: Erdös-Rényi Graphs



Notes: Each line gives the average network externality as a function of β_0, where the network is generated through an ER graph. The complete information game shows how the relationship between the network externality and β_0 changes as we expand the graph from a subgraph of agents within distance k from the agent 1. (Networks A, B, and C correspond to networks with $k = 3, 4, 5$ from a small graph to a large one.) The figures show that the average network externality from the behavioral model behaves more stably across different networks than that from the equilbrium model in particular when β_0 (local interaction parameter) is high.

Figure 2.2: Network Externality Comparison Between Equilibrium and Behavioral Models: Barabási-Albert Graphs



Notes: The figure is similar to the previous one except that the graph is now BA. The complete information game shows the relation changes as we expand the graph from a subgraph of agents within distance k from the agent 1. Again, the behavioral model gives a prediction of the relation that tends to be more stable than the complete information game in this network generation.

local stochastic dependence of observed actions gives a better property of external validity than the complete information equilibrium model.

Suppose that the econometrician believes the true model is an equilibrium model, but uses our behavioral model as a proxy for the equilibrium model. If these two models generate “similar” predictions, using our behavioral model as a proxy will not be a bad idea. The results in Figures 2.1 and 2.2 again show that the answer depends on the payoff externality β_0. Unless the parameter β_0 is very high (say larger than or equal to 0.5), both the equilibrium approach and the behavioral approach give similar network externality. However, the discrepancy widens when β_0 is high. Hence in this set-up, using our behavioral approach as a proxy for an equilibrium approach makes sense only when strategic interdependence is not too high.

The comparison here uses a set-up where the econometrician observes all the players in the game. However, it should be kept in mind that as we shall see later when we propose inference, the behavioral model naturally accommodates the case where one observes only part of the players whereas the complete information game model does not in general. Hence when the local strategic interactions are not very high, the behavioral model can be a good proxy for a complete information game model with predictions from an equilibrium when only part of the players are observed in the sample.

2.3 Econometric Inference

2.3.1 General Overview

Partial Observation of Interactions

A large network data set is often obtained through a non-random sampling process. (See e.g. Kolaczyk [80].) The main difficulty in practice is that the actual sampling process by which the network data are gathered is hard to formulate formally with accuracy. Our approach of empirical modeling can be useful in such a situation where interactions are observed only partially through a certain non-random sampling scheme that is not precisely known. In this section, we make explicit the data requirements for the econometrician and propose inference procedures. We mainly focus on the game where all the players in the game are of simple type. Later, we discuss the situation with agents of first order sophisticated type.

Suppose that the original game of interactions consists of a large number of agents whose set we denote by N. Let the set of players N be on a payoff graph G_P and an information graph G_I, facing the strategic environment as described in the preceding section. Denote the best response as an observed dependent variable Y_i: for $i \in N$,

$$Y_i = s_i^{\mathsf{BR}}(\mathscr{I}_i).$$

Let us make the following additional assumption on this original large game. Let us first define

$$\mathscr{F} = \sigma(X, G_P, G_I) \vee \mathscr{C},$$

i.e., the σ-field generated by $X = (X_i)_{i \in N}$, G_P, G_I and $\mathscr{C}$.

Assumption 2.3.1. (i) ε_i's and η_i's are conditionally i.i.d. across i's given $\mathscr{F}$.
(ii) $\{\varepsilon_i\}_{i=1}^n$ and $\{\eta_i\}_{i=1}^n$ are conditionally independent given $\mathscr{F}$.
(iii) For each $i \in N$, $\mathbf{E}[\varepsilon_i|\mathscr{F}] = 0$ and $\mathbf{E}[\eta_i|\mathscr{F}] = 0$.

The last condition (iii) excludes endogenous formation of G_P or G_I, because the condition requires that the unobserved type components ε_i and η_i be conditionally mean independent of these graphs, given $X = (X_i)_{i \in N}$ and $\mathscr{C}$. However, the condition does not exclude the possibility that G_P and G_I are formed based on $(X, \mathscr{C})$. Hence the formation of networks by agents using information in X or $\mathscr{C}$ is permitted in the chapter.

The econometrician observes only a subset $N^* \subset N$ of agents and part of G_P through a potentially stochastic sampling process of unknown form. We assume for simplicity that $n^* \equiv |N^*|$ is nonstochastic. This assumption is satisfied, for example, if one collects the data for agents with predetermined sample size n^*. We assume that though being a small fraction of N, the set N^* is still a large set justifying our asymptotic framework that sends n^* to infinity. Most importantly, constituting only a small fraction of N, the observed sample N^* of agents induces a payoff subgraph which one has no reason to view as "approximating" or "similar to" the original payoff graph G_P. Let us make precise the data requirements.

Condition A: The stochastic elements of the sampling process are conditionally independent of $\{(T_i', \eta_i)'\}_{i \in N}$ given $\mathscr{F}$.

Condition B: For each $i \in N^*$, the econometrician observes $N_P(i)$ and (Y_i, X_i), and for each $j \in N_P(i)$, the econometrician observes $|N_P(i) \cap N_P(j)|$, $n_P(j)$ and X_j.

Condition C: Either of the following two conditions is satisfied:

(a) For $i, j \in N^*$ such that $i \neq j$, $N_P(i) \cap N_P(j) = \varnothing$.

(b) For each agent $i \in N^*$, and for any agent $j \in N^*$ such that $N_P(i) \cap N_P(j) \neq \varnothing$, the econometrician observes Y_j, $|N_P(j) \cap N_P(k)|$, $n_P(k)$ and X_k for all $k \in N_P(j)$.

Before we discuss the conditions, it is worth noting that these conditions are trivially satisfied when we observe the full payoff graph G_P and $N^* = N$. Condition A is satisfied, for example, if the sampling process is based on observed characteristics X and some characteristics of the strategic environment that is commonly observed by all the players. This condition is violated if the sampling is based on the outcomes Y_i's or unobserved payoff-relevant signals such as ε_i or η_i. Condition B essentially requires that in the data set, we observe (Y_i, X_i) of many agents i, and for each G_P-neighbor j of agent i, observe the number of the agents who are common G_P-neighbors of i and j and the size of G_P-neighborhood of j along with the observed characteristics X_j.[10] As for a G_P-neighbor j of agent $i \in N^*$, this condition does not require that the agent j's action Y_i or the full set of his G_P-neighbors are observed. Condition C(a) is typically satisfied when the sample of agents N^* is randomly selected from a much larger set of agents so that no two agents have overlapping G_P-neighbors in the sample.[11] In practice for use in inference, one can take the set N^* to include only those agents that satisfy Conditions A-C as long as N^* thereof is still large and the selection is based only on (X, G_P). One can simply use only those agents whose G_P-neighborhoods are not overlapping, as long as there are many such agents in the data.

[10] Note that this condition is violated when the neighborhoods are top-coded in practice. For example, the maximum number of friends in the survey for a peer effects study can be set to be lower than the actual number of friends for many students. The impact of this top-coding upon the inference procedure is an interesting question on its own which deserves exploration in a separate paper.

[11] This random selection does not need to be a random sampling from the population of agents. Note that the random sampling is extremely hard to implement in practice in this situation, because one needs to use the equal probability for selecting each agent into the collection N^*, but this equal probability will be feasible only when one has at least the catalog of the entire population N.

Estimating Payoff Parameters and the Average Network Externality

In order to introduce inference procedures for β_0 and other payoff parameters, let us define for $i \in N$,

$$\begin{aligned} Z_{i,1} &= \lambda_{ii} X_{i,1} + \frac{\beta_0 \lambda_{ii}}{n_P(i)} \sum_{j \in N_P(i)} \lambda_{ij} X_{j,1}, \text{ and} \\ Z_{i,2} &= \frac{1}{n_P(i)} \sum_{j \in N_P(i)} \lambda_{ij} X_{j,2}. \end{aligned} \tag{2.7}$$

(Note that $Z_{i,1}$ and $Z_{i,2}$ rely on β_0 although it is suppressed from notation for simplicity as we do frequently below for other quantities.) By Theorem 2.2.1, we can write

$$Y_i = Z_{i,1}' \gamma_0 + Z_{i,2}' \delta_0 + v_i,$$

where

$$v_i = \lambda_{ii} \varepsilon_i + \frac{\beta_0 \lambda_{ii}}{n_P(i)} \sum_{j \in N_P(i)} \lambda_{ij} \varepsilon_j + \eta_i.$$

Note that the observed actions Y_i are cross-sectionally dependent (conditional on X_i's) due to information sharing on unobservables ε_i. However, since only the types of G_P-neighbors turn out to be relevant in the best linear response, the correlation between Y_i and Y_j is non-zero only when agents i and j are G_P-neighbors.

We define $Z_i = [Z_{i,1}', Z_{i,2}']' \in \mathbf{R}^{d_{x_1} + d_{x_2}}$ and $\rho_0 = [\gamma_0', \delta_0']' \in \mathbf{R}^{d_{x_1} + d_{x_2}}$, where $X_{i,1} \in \mathbf{R}^{d_{x_1}}$ and $X_{i,2} \in \mathbf{R}^{d_{x_2}}$, so that we can rewrite the linear model as

$$Y_i = Z_i' \rho_0 + v_i.$$

Suppose that φ_i is $M \times 1$ vector of instrumental variables (which potentially depend on β_0) with $M > d \equiv d_{x_1} + d_{x_2}$ such that for all $i \in N$,

$$\mathbf{E}[v_i \varphi_i] = 0.$$

Note that the orthogonality condition above holds for any φ_i as long as for each

$i \in N$, φ_i is $\mathscr{F}$-measurable, i.e., once $\mathscr{F}$ is realized, there is no extra randomness in φ_i. This is the case, for example, when φ_i is a function of $X = (X_i)_{i \in N}$. We also allow that each φ_i depends on β_0.

While the asymptotic validity of our inference procedure admits a wide range of choices for φ_i's, one needs to choose them with care to obtain sharp inference on the payoff parameters. Especially, it is important to consider instrumental variables which involve the characteristics of G_P-neighbors to obtain a sharp inference on payoff externality parameter β_0. This is because the cross-sectional dependence of observations carries substantial information for estimating strategic interdependence among agents.

The moment function is nonlinear in the payoff externality β_0 and it is not easy to ensure that these moment conditions uniquely determine the true parameter vector even in the limit as n^* goes to infinity.[12] In this chapter, we adopt a Bonferroni procedure in which we first obtain a confidence interval for β_0 and, using this, we perform inference on ρ_0. This approach works well even when β_0 is not consistently estimable.

We proceed first to estimate ρ_0 assuming knowledge of β_0. Define

$$S_{\varphi\varphi} = \varphi'\varphi / n^*,$$

and let

$$\tilde{\varphi} = \varphi S_{\varphi\varphi}^{-1/2},$$

where φ is an $n^* \times M$ matrix whose i-th row is given by φ_i', $i \in N^*$. Define

$$\Lambda = \frac{1}{n^*} \sum_{i \in N^*} \sum_{j \in N^*} \mathbf{E}[v_i v_j | \mathscr{F}] \tilde{\varphi}_i \tilde{\varphi}_j', \tag{2.8}$$

and let $\hat{\Lambda}$ be a consistent estimator of Λ. (We will explain how we construct this

[12] One might consider following the nonlinear iterated least squares approach of Blundell and Robin [22]. However, it is not clear in our context whether the parameter β_0 is consistently estimable across various payoff graph configurations as n^* diverges to infinity. Thus, this chapter takes a Bonferroni approach.

estimator later.) Define

$$S_{Z\tilde{\varphi}} = Z'\tilde{\varphi}/n^*, \text{ and } S_{\tilde{\varphi}y} = \tilde{\varphi}'y/n^*,$$

where Z is an $n^* \times d$ matrix whose i-th row is given by Z_i' and y is an $n^* \times 1$ vector whose i-th entry is given by Y_i, $i \in N^*$. Since (from the fact that G_P is undirected)

$$c_{ij} = \frac{|N_P(i) \cap N_P(j)|}{n_P(i)},$$

we can construct Z_i for each $i \in N^*$ from the data satisfying Conditions A-C. Then we estimate

$$\hat{\rho} = \left[S_{Z\tilde{\varphi}}\hat{\Lambda}^{-1}S_{Z\tilde{\varphi}}'\right]^{-1} S_{Z\tilde{\varphi}}\hat{\Lambda}^{-1}S_{\tilde{\varphi}y}. \tag{2.9}$$

Using this estimator, we construct a vector of residuals $\hat{v} = [\hat{v}_i]_{i \in N^*}$, where

$$\hat{v}_i = Y_i - Z_i'\hat{\rho}. \tag{2.10}$$

Finally, we form a test statistic as follows:

$$T(\beta_0) = \frac{\hat{v}'\tilde{\varphi}\hat{\Lambda}^{-1}\tilde{\varphi}'\hat{v}}{n^*}, \tag{2.11}$$

making it explicit that the test statistic depends on β_0. Later we show that

$$T(\beta_0) \to_d \chi^2_{M-d}, \text{ as } n^* \to \infty,$$

where χ^2_{M-d} denotes the χ^2 distribution with degree of freedom $M-d$. Let $C^{\beta}_{1-(\alpha/2)}$ be the $(1-(\alpha/2))100\%$ confidence set for β_0 defined as

$$C^{\beta}_{1-(\alpha/2)} \equiv \{\beta \in (-1,1) : T(\beta) \leq c_{1-(\alpha/2)}\},$$

where $T(\beta)$ is computed as $T(\beta_0)$ with β_0 replaced by β and the critical value $c_{1-(\alpha/2)}$ is the $(1-(\alpha/2))$-quantile of χ^2_{M-d}.

Then we establish[13] that under regularity conditions,

$$\sqrt{n^*}\hat{V}^{-1/2}(\hat{\rho}-\rho_0) \to_d N(0,I),$$

as $n^* \to \infty$, where

$$\hat{V} = \left[S_{Z\tilde{\varphi}}\hat{\Lambda}^{-1}S'_{Z\tilde{\varphi}}\right]^{-1}.$$

Using this estimator $\hat{\rho}$, we can construct a $(1-\alpha)100\%$ confidence interval for $a'\rho_0$ for any non-zero vector a. For this define

$$\hat{\sigma}^2(a) = a'\hat{V}a.$$

Let $c^a_{1-(\alpha/4)}$ be the $(1-(\alpha/4))$-percentile of $N(0,1)$. Define for a vector a with the same dimension as ρ,

$$C^{\rho}_{1-(\alpha/2)}(\beta_0,a) = \left[a'\hat{\rho} - \frac{c^a_{1-(\alpha/4)}\hat{\sigma}(a)}{\sqrt{n}}, a'\hat{\rho} + \frac{c^a_{1-(\alpha/4)}\hat{\sigma}(a)}{\sqrt{n}}\right].$$

Then the confidence set for $a'\rho$ is given by

$$C^{\rho}_{1-\alpha}(a) = \bigcup_{\beta \in C^{\beta}_{1-(\alpha/2)}} C^{\rho}_{1-(\alpha/2)}(\beta,a).$$

Notice that since β runs in $(-1,1)$ and the estimator $\hat{\rho}$ has an explicit form, the confidence interval is not computationally costly to construct in general.

Often the eventual parameter of interest is one that captures how strongly the agents's decisions are inter-dependent through the network. Here let us introduce parameters representing the sensitivity. Let $s_i^{\mathsf{BR}}(\mathscr{I}_i)$ be the best linear response of agent i having information set $\mathscr{I}_i$. Let us define *the average network externality*

[13] The asymptotic theory proofs can be found in the working paper version of this chapter, (Canen et al. [32]).

with respect to variable $X_{i,1,r}$ (where $X_{i,1,r}$ represents the r-th entry of $X_{i,1}$) to be

$$\begin{aligned}\theta_1(\beta_0,\gamma_{0,r}) &= \frac{1}{n^*}\sum_{i\in N^*}\frac{1}{n_P(i)}\sum_{j\in N_P(i)}\frac{\partial s_i^{\mathsf{BR}}(\mathscr{I}_i)}{\partial x_{j,1,r}}\\ &= \frac{1}{n^*}\sum_{i\in N^*}\frac{1}{n_P(i)}\sum_{j\in N_P(i)}\frac{\beta_0 r_{ij}}{n_P(i)(1-\beta_0 c_{ii})(1-\beta_0 c_{ij})}\gamma_{0,r},\end{aligned}$$

where $\gamma_{0,r}$ denotes the r-th entry of γ_0. See (2.6). Thus the confidence interval for $\theta_1(\beta_0,\gamma_0)$ can be constructed from the confidence interval for β_0 and γ_0 as follows:

$$C_{1-\alpha}^{\theta_1} = \left\{\theta_1(\beta,\gamma_r) : \beta \in C_{1-\alpha}^{\beta}, \text{ and } \gamma_r \in C_{1-\alpha}^{\gamma_r}\right\},$$

where $C_{1-\alpha}^{\beta}$ denotes the confidence interval for $\gamma_{0,r}$. We can define similarly the average network externality with respect to an entry of $X_{i,2}$ and construct a confidence interval for it. Details are omitted.

Downweighting Players with High Degree Centrality

When there are players who are linked to many other players in G_P, the graph G_P tends to be denser, making it difficult to obtain good variance estimators that perform stably in finite samples. To remedy this situation, this chapter proposes a downweighting of those players with high degree centrality in G_P. More specifically, in choosing an instrument vector φ_i, we may consider the following:

$$\varphi_i(X) = \frac{1}{\sqrt{\bar{n}_P(i)}} g_i(X), \tag{2.12}$$

where $g_i(X)$ is a function of X. This choice of φ_i downweights players i who have a large G_P-neighborhood. Thus we rely less on the variations of the characteristics of those players who have many neighbors in G_P.

Taking downweighting agents too heavily may hurt the power of the inference because the actions of agents with high centrality contain information about the parameter of interest through the moment restrictions. On the other hand downweighting them too lightly will hurt the finite sample stability of the inference due to strong cross-sectional dependence they cause to the observations. Since a model with agents of higher order sophisticated type results in observations with more extensive cross-sectional dependence, the role of downweighting can be prominent in securing finite sample stability in such a model.

Comparison with Linear-in-Means Models

One of the most frequently used interaction models in the econometrics literature is a linear-in-means model specified as follows:

$$Y_i = X_{i,1}'\gamma_0 + \overline{X}_{i,2}'\delta_0 + \beta_0\mu_i^e(\overline{Y}_i) + v_i, \tag{2.13}$$

where $\mu_i^e(\overline{Y}_i)$ denotes the player i's expectation of $\overline{Y}_i$, and

$$\overline{Y}_i = \frac{1}{n_P(i)}\sum_{i \in N_P(i)} Y_i \text{ and } \overline{X}_i = \frac{1}{n_P(i)}\sum_{i \in N_P(i)} X_i.$$

The literature assumes rational expectations by equating $\mu_i^e(\overline{Y}_i)$ to $\mathbf{E}[\overline{Y}_i|\mathscr{I}_i]$, and then proceeds to identification analysis of parameters γ, δ_0 and β_0. For actual inference, one needs to use an estimated version of $\mathbf{E}[\overline{Y}_i|\mathscr{I}_i]$. One standard way in the literature is to replace it by $\overline{Y}_i$ so that we have

$$Y_i = X_{i,1}'\gamma_0 + \overline{X}_{i,2}'\delta_0 + \beta_0\overline{Y}_i + \tilde{v}_i,$$

where $\tilde{v}_i$ is an error term defined as $\tilde{v}_i = \beta_0(\mathbf{E}[\overline{Y}_i|\mathscr{I}_i] - \overline{Y}_i) + v_i$. The complexity arises due to the presence of $\overline{Y}_i$ which is an endogneous variable that is involved in the error term $\tilde{v}_i$.[14]

One of the frequently used approaches is to use instrumental variables. There are two types of instrumental variables. The first kind is a peers-of-peers type instrumental variable which is based on the observed characteristics of the neighbors of the neighbors. This strategy was proposed by Kelejian and Robinson [78], Bramoullé, Djebbari, and Fortin [25] and De Giorgi, Pellizzari, and Redaelli [45]. The second kind of an instrumental variable is based on observed characteristics

[14] A similar observation applies in the case of a complete information version of the model, where one directly uses $\overline{Y}_i$ in place of $\mu_i^e(\overline{Y}_i)$ in (2.13). Still due to simultaneity of the equations, $\overline{Y}_i$ necessarily involve error terms v_i not only of agent i's own but other agents' as well.

excluded from the group characteristics as instrumental variables. (See Brock and Durlauf [28] and Durlauf and Tanaka [50].) However, finding such an instrumental variable in practice is not always a straightforward task in empirical research.

Our approach of empirical modeling is different in several aspects. Our modeling uses behavioral assumptions instead of rational expectations, and produces a reduced form for observed actions Y_i from using best linear responses. This reduced form gives a rich set of testable implications and makes explicit the source of cross-sectional dependence in relation to the payoff graph. Our approach permits any nontrivial functions of $\mathscr{F}$ as instrumental variables at least for the validity of the inference. Furthermore, one does not need to observe many independent interactions for inference.

Estimation of Asymptotic Covariance Matrix

The inference requires an estimator of $\hat{V}$. First, let us find the population version of $\hat{V}$. After some algebra, it is not hard to see that the population version (conditional on $\mathscr{F}$) of $\hat{V}$ is given by

$$V = \left[S_{Z\tilde{\varphi}}\Lambda^{-1}S'_{Z\tilde{\varphi}}\right]^{-1}. \tag{2.14}$$

For estimation, it suffices to estimate Λ defined in (2.8). For this, we need to incorporate the cross-sectional dependence of the residuals v_i properly. From the definition of v_i, it turns out that v_i and v_j can be correlated if i and j are connected indirectly through two edges in G_P. However, constructing an estimator of Λ simply by imposing this dependence structure and replacing v_i by $\hat{v}_i$ can result in a conservative estimator with unstable finite sample properties, especially when each player has many players connected through two edges. Instead, this chapter proposes an alternative estimator of Λ as follows. This estimator is found to work well in our simulation studies.

We first explain our proposal to estimate Λ consistently for the case of $\beta_0 \neq 0$. Then we later show how the estimator works even for the case of $\beta_0 = 0$. We first write

$$v_i = R_i(\varepsilon) + \eta_i, \tag{2.15}$$

where

$$R_i(\varepsilon) = \lambda_{ii}\varepsilon_i + \frac{\beta_0\lambda_{ii}}{n_P(i)} \sum_{j \in N_P(i)} \lambda_{ij}\varepsilon_j.$$

Define for $i, j \in N$,

$$e_{ij} \quad = \quad \mathbf{E}\left[R_i(\varepsilon)R_j(\varepsilon)|\mathscr{F}\right]/\sigma_\varepsilon^2,$$

where σ_ε^2 denotes the variance of ε_i. It is not hard to see that for all $i \in N$,

$$e_{ii} = \lambda_{ii}^2 + \frac{\beta_0^2\lambda_{ii}^2}{n_P^2(i)} \sum_{j \in N_P(i)} \lambda_{ij}^2,$$

and for $i \neq j$ such that $N_P(i) \cap N_P(j) \neq \varnothing$, $e_{ij} = \beta_0 q_{\varepsilon,ij}$, where

$$q_{\varepsilon,ij} = \frac{\lambda_{ji}\lambda_{ii}\lambda_{jj}}{n_P(j)} + \frac{\lambda_{ij}\lambda_{ii}\lambda_{jj}}{n_P(i)} + \frac{\beta_0\lambda_{ii}\lambda_{jj}}{n_P(i)n_P(j)} \sum_{k \in N_P(i) \cap N_P(j)} \lambda_{ik}\lambda_{jk}.$$

Thus, we write

$$\begin{aligned} \frac{1}{n^*} \sum_{i \in N^*} \mathbf{E}[v_i^2|\mathscr{F}] &= a_\varepsilon\sigma_\varepsilon^2 + \sigma_\eta^2, \text{ and} \\ \frac{1}{n^*} \sum_{i \in N^*} \sum_{j \in N_P(i) \cap N^*} \mathbf{E}[v_i v_j|\mathscr{F}] &= \beta_0 b_\varepsilon \sigma_\varepsilon^2, \end{aligned} \tag{2.16}$$

where σ_η^2 denotes the variance of η_i,

$$a_\varepsilon = \frac{1}{n^*} \sum_{i \in N^*} e_{ii}, \text{ and } b_\varepsilon = \frac{1}{n^*} \sum_{i \in N^*} \sum_{j \in N_P(i) \cap N^*} q_{\varepsilon,ij}.$$

(Note that since not all agents in $N_P(i)$ are in N^* for all $i \in N^*$, the set $N_P(i) \cap N^*$ does not necessarily coincide with $N_P(i)$.) When $\beta_0 \neq 0$, the solution takes the following form:

$$\sigma_\varepsilon^2 = \frac{1}{n^*\beta_0 b_\varepsilon}\sum_{i\in N^*}\sum_{j\in N_P(i)\cap N^*}\mathbf{E}[v_i v_j|\mathscr{F}] \text{ and} \tag{2.17}$$

$$\sigma_\eta^2 = \frac{1}{n^*}\sum_{i\in N^*}\mathbf{E}[v_i^2|\mathscr{F}] - \frac{a_\varepsilon}{n^*\beta_0 b_\varepsilon}\sum_{i\in N^*}\sum_{j\in N_P(i)\cap N^*}\mathbf{E}[v_i v_j|\mathscr{F}].$$

In other words, when $\beta_0 \neq 0$, i.e., when there is strategic interaction among the players, we can "identify" σ_ε^2 and σ_η^2 by using the variances and covariances of residuals v_i's. The intuition is as follows. Since the source of cross-sectional dependence of v_i's is due to the presence of ε_i's, we can identify first σ_ε^2 using covariance between v_i and v_j for linked pairs i, j, and then identify σ_η^2 by subtracting from the variance of v_i the contribution from ε_i.

In order to obtain a consistent estimator of Λ which does not require that $\beta_0 \neq 0$, we derive its alternative expression. Let us first write

$$\Lambda = \Lambda_1 + \Lambda_2, \tag{2.18}$$

where

$$\Lambda_1 = \frac{1}{n^*}\sum_{i\in N^*}\mathbf{E}[v_i^2|\mathscr{F}]\tilde{\varphi}_i\tilde{\varphi}_i', \text{ and}$$

$$\Lambda_2 = \frac{1}{n^*}\sum_{i\in N^*}\sum_{j\in N_{-i}^*}\mathbf{E}[v_i v_j|\mathscr{F}]\tilde{\varphi}_i\tilde{\varphi}_j',$$

where $N_{-i}^* = N^*\backslash\{i\}$. Using (2.15) and (2.17), we can rewrite

$$\begin{aligned}\Lambda_2 &= \frac{1}{n^*}\sum_{i\in N^*}\sum_{j\in N_{-i}^*:N_P(i)\cap N_P(j)\neq\varnothing} e_{ij}\sigma_\varepsilon^2\tilde{\varphi}_i\tilde{\varphi}_j' \\ &= \frac{\beta_0}{n^*}\sum_{i\in N^*}\sum_{j\in N_{-i}^*:N_P(i)\cap N_P(j)\neq\varnothing} q_{\varepsilon,ij}\sigma_\varepsilon^2\tilde{\varphi}_i\tilde{\varphi}_j' \\ &= \frac{s_\varepsilon}{n^*}\sum_{i\in N^*}\sum_{j\in N_{-i}^*:N_P(i)\cap N_P(j)\neq\varnothing} q_{\varepsilon,ij}\tilde{\varphi}_i\tilde{\varphi}_j',\end{aligned}$$

where

$$s_\varepsilon = \frac{\sum_{i\in N^*}\sum_{j\in N_P(i)\cap N^*}\mathbf{E}[v_i v_j|\mathscr{F}]}{\sum_{i\in N^*}\sum_{j\in N_P(i)\cap N^*} q_{\varepsilon,ij}}.$$

Now, it is clear that with this expression for Λ_2, the definition of Λ is well defined regardless of whether $\beta_0 = 0$ or $\beta_0 \neq 0$.

Thus, to obtain an estimator $\hat{\Lambda}$ of Λ (up to β_0), we first obtain a first-step estimator of ρ as follows:

$$\tilde{\rho} = \left[S_{Z\tilde{\varphi}}S'_{Z\tilde{\varphi}}\right]^{-1} S_{Z\tilde{\varphi}}S_{\tilde{\varphi}y}. \tag{2.19}$$

Using this estimator, we construct a vector of residuals $\tilde{v} = [\tilde{v}_i]_{i\in N^*}$, where

$$\tilde{v}_i = Y_i - Z_i'\tilde{\rho}. \tag{2.20}$$

Then we estimate[15]

$$\hat{\Lambda} = \hat{\Lambda}_1 + \hat{\Lambda}_2,$$

where

$$\begin{aligned}\hat{\Lambda}_1 &= \frac{1}{n^*}\sum_{i\in N^*}\tilde{v}_i^2\tilde{\varphi}_i\tilde{\varphi}_i', \text{ and}\\ \hat{\Lambda}_2 &= \frac{\hat{s}_\varepsilon}{n^*}\sum_{i\in N^*}\sum_{j\in N^*_{-i}:N_P(i)\cap N_P(j)\neq\varnothing} q_{\varepsilon,ij}\tilde{\varphi}_i\tilde{\varphi}_j',\end{aligned}$$

and

$$\hat{s}_\varepsilon = \frac{\sum_{i\in N^*}\sum_{j\in N_P(i)\cap N^*}\tilde{v}_i\tilde{v}_j}{\sum_{i\in N^*}\sum_{j\in N_P(i)\cap N^*} q_{\varepsilon,ij}}. \tag{2.21}$$

(Note that the quantity $q_{\varepsilon,ij}$ can be evaluated once β_0 is fixed.) Using $\hat{\Lambda}$, we con-

[15]Under Condition C(a) for sample N^*, we have $\Lambda_2 = 0$ because the second sum in the expression for λ_2 is empty. Hence in this case, we can simply set $\hat{\Lambda}_2 = 0$.

struct the estimator for the covariance matrix V, i.e.,

$$\hat{V} = \left[S_{Z\tilde{\varphi}}\hat{\Lambda}^{-1}S'_{Z\tilde{\varphi}}\right]^{-1}. \tag{2.22}$$

Later we provide conditions for the estimator to be consistent for V.[16]

Testing for Information Sharing on Unobservables

One may want to see how much empirical relevance there is for incorporating information sharing on unobservables. Observe that when $\beta_0 = 0$, presence of information sharing on unobservables is not testable. When $\beta_0 = 0$, it follows that

$$s_i^{\mathsf{BR}}(\mathscr{I}_i) = X'_{i,1}\gamma_0 + \tilde{X}'_{i,2}\delta_0 + v_i,$$

where $v_i = \varepsilon_i + \eta_i$. In this case, it is not possible to distinguish between contributions from ε_i and η_i.

Consider the following hypotheses:

$$\begin{aligned} H_0 &: \sigma_\varepsilon^2 = 0, \text{ and} \\ H_1 &: \sigma_\varepsilon^2 > 0. \end{aligned}$$

The null hypothesis tells us that there is no information sharing on unobservables. Let $\hat{v}_i(\beta)$, $a_\varepsilon(\beta)$ and $b_\varepsilon(\beta)$ be the same as previously defined $\hat{v}_i$, a_ε and b_ε only with β_0 replaced by generic β. From here on we assume that $\beta_0 \neq 0$.

The main idea for testing the hypothesis is that when $\sigma_\varepsilon^2 > 0$, this implies cross-sectional dependence of residuals v_i. We need to compute the sample version of the covariance between v_i and v_j for G_P-neighbors i and j. However, Condition C alone does not guarantee that for each $i \in N^*$, we will be able to compute $\hat{v}_j$ for some $j \in N_P(i)$, because there may not exist such j for some $i \in N^*$ at all. Thus let us introduce an additional data requirement as follows:

[16] In finite samples, $\hat{V}$ is not guaranteed to be positive definite. We can modify the estimator by using spectral decomposition similarly as in Cameron et al. [31]. More specifically, we first take a spectral decomposition $\hat{V} = \hat{B}\hat{A}\hat{B}'$, where $\hat{A}$ is a diagonal matrix of eigen values $\hat{a}_j$ of $\hat{V}$. We replace each $\hat{a}_j$ by the maximum between $\hat{a}_j$ and some small number $c > 0$ in $\hat{A}$ to construct A_*. Then the modified version $\tilde{V} \equiv \hat{B}A_*\hat{B}'$ is positive definite. For $c > 0$, taking $c = 0.005$ seems to work well in the simulation studies.

Condition D: For each $i \in N^*$, the econometrician observes a nonempty subset $\widetilde{N}(i) \subset N_P(i)$ (possibly a singleton) of agents where for each $j \in \widetilde{N}(i)$, the econometrician observes Y_j, $|N_P(j) \cap N_P(k)|$, $n_P(k)$ and X_k for all $k \in N_P(j)$.

Condition D is satisfied if there are many agents in the data set where each agent has at least one G_P-neighbor j for which the econometrician observes the outcome Y_j, the number of their G_P-neighbors, the observed characteristics of their G_P-neighbors, and the number of the agents who are both their G_P-neighbors and the neighbors of their G_P-neighbors. While this data requirement can be restrictive in some cases where one obtains a partial observation of G_P, it is still weaker than the usual assumption that the econometrician observes G_P fully together with $(Y_i, X_i')_{i \in N}$.

Now let us reformulate the null and the alternative hypothesis as follows:

$$H_0 \;:\; \frac{1}{n^*} \sum_{i \in N^*} \sum_{j \in \widetilde{N}(i)} \mathbf{E}[v_i v_j | \mathscr{F}] = 0, \text{ and} \tag{2.23}$$

$$H_1 \;:\; \frac{1}{n^*} \sum_{i \in N^*} \sum_{j \in \widetilde{N}(i)} \mathbf{E}[v_i v_j | \mathscr{F}] \neq 0. \tag{2.24}$$

For testing, we propose the following method. Let $C^{\beta}_{1-(\alpha/2)}$ be the $(1-(\alpha/2))$-level confidence interval for β. We consider the following test statistics:

$$\widehat{IU} = \inf_{\beta \in C^{\beta}_{1-(\alpha/2)}} \frac{1}{2\hat{S}^4(\beta) n^*} \left(\sum_{i \in N^*} \sum_{j \in \widetilde{N}(i)} \hat{v}_i(\beta) \hat{v}_j(\beta) \right)^2,$$

where

$$\hat{S}^2(\beta) = \frac{\widetilde{d}_{av}^{1/2}}{n^*} \sum_{i \in N^*} \hat{v}_i^2(\beta), \text{ and } \widetilde{d}_{av} = \frac{1}{n^*} \sum_{i \in N^*} |\widetilde{N}(i)|.$$

When the confidence set includes zero, the power of the test becomes asymptotically trivial, as expected from the previous remark that information sharing on unobservables is not testable when $\beta_0 = 0$.

As for the critical value, we take the $(1-(\alpha/2))$-percentile from the χ^2 distribution with degree of freedom 1, which we denote by $c_{1-(\alpha/2)}$. Then the level

α-test based on the test statistic $\widehat{IU}$ rejects the null hypothesis if and only if $\widehat{IU} > c_{1-(\alpha/2)}$. We investigate the finite sample properties of this test in the supplemental note to this chapter.

2.3.2 Asymptotic Theory

In this section, we present the assumptions and formal results of asymptotic inference. We introduce some technical conditions.

Assumption 2.3.2. There exists $c > 0$ such that for all $n^* \geq 1$, $\lambda_{\min}(S_{\varphi\varphi}) \geq c$, $\lambda_{\min}(S_{Z\tilde{\varphi}}S'_{Z\tilde{\varphi}}) \geq c$, $\lambda_{\min}(S_{Z\tilde{\varphi}}\Lambda^{-1}S'_{Z\tilde{\varphi}}) \geq c$, $\lambda_{\min}(\Lambda) \geq c$, and

$$\frac{1}{n^*}\sum_{i\in N^*}\frac{\lambda_{ii}}{n_P(i)}\sum_{j\in N_P(i)\cap N^*}\lambda_{ij} > c,$$

where $\lambda_{\min}(A)$ for a symmetric matrix A denotes the minimum eigenvalue of A.

Assumption 2.3.3. There exists a constant $C > 0$ such that for all $n^* \geq 1$,

$$\max_{i\in N^\circ}||X_i|| + \max_{i\in N^\circ}||\tilde{\varphi}_i|| \leq C$$

and $\mathbf{E}[\varepsilon_i^4|\mathscr{F}] + \mathbf{E}[\eta_i^4|\mathscr{F}] < C$, where $n^\circ = |N^\circ|$ and

$$N^\circ = \bigcup_{i\in N^*}\bar{N}_P(i).$$

Assumption 2.3.2 is used to ensure that the asymptotic distribution is nondegenerate. This regularity condition is reasonable, because an asymptotic scheme that gives a degenerate distribution would not be adequate to derive a finite sample, nondegenerate distribution of an estimator. Assumption 2.3.3 can be weakened at the expense of complexity in the conditions and the proofs.

We introduce an assumption which requires the payoff graph to have a bounded degree over i in the observed sample N^*.

Assumption 2.3.4. There exists $C > 0$ such that for all $n^* \geq 1$,

$$\max_{i\in N^*}|N_P(i)| \leq C.$$

We may relax the assumption to a weaker, yet more complex condition at the expense of longer proofs, but in our view, this relaxation does not give additional insights. When N^* is large, one can remove very high-degree nodes to obtain a stable inference. As such removal is solely based on the payoff graph G_P, the removal does not lead to any violation of the conditions in the chapter.

The following theorem establishes the asymptotic validity of the inference based on the best linear responses in Theorem 2.2.1. The proof is found in the supplemental note to this chapter.

Theorem 2.3.1. *Suppose that the conditions of Theorem 2.2.1 and Assumptions 2.3.1 - 2.3.4 hold. Then,*

$$T(\beta_0) \to_d \chi^2_{M-d}, \text{ and } \hat{V}^{-1/2}\sqrt{n^*}(\hat{\rho}-\rho_0) \to_d N(0,I),$$

as $n^* \to \infty$. Furthermore, under the null hypothesis in (2.23),

$$\lim_{n^*\to\infty} P\left\{\widehat{IU} > c_{1-\alpha/2}\right\} \leq \alpha.$$

The asymptotic validity of inference is not affected if the researcher chooses a nonempty subset $\widetilde{N}(i)$ in Condition D as a singleton subset, say, $j(i) \subset N_P(i)$, $j(i) \in N$, such that we observe $Y_{j(i)}$, $|N_P(j(i)) \cap N_P(k)|$, $n_P(k)$ and X_k for all $k \in N_P(j(i))$ are available in the data, so far as the choice is not based on Y_i's but on X only.

Testing for Information Sharing on Unobservables

When $\beta_0 = 0$, it follows that

$$s_i^{\mathsf{BR.FS}}(\mathscr{I}_i) = X'_{i,1}\gamma_0 + \tilde{X}'_{i,2}\delta_0 + v_i^{\mathsf{FS}},$$

where $v_i^{\mathsf{FS}} = \varepsilon_i + \eta_i$. Therefore, we have $s_i^{\mathsf{BR}}(\mathscr{I}_{i,0}) = s_i^{\mathsf{BR.FS}}(\mathscr{I}_{i,1})$ and just as in the case of a simple type model, it is not possible to distinguish between contributions from ε_i and η_i. Thus let us assume that $\beta_0 \neq 0$. The presence of cross-sectional correlation of residuals v_i^{FS} serves as a testable implications from information sharing on unobservables. As in the case of a model with agents of simple type, we need to strengthen Condition D as follows:

Condition D1: For each $i \in N^*$, the econometrician observes a nonempty subset $\widetilde{N}(i) \subset N_P(i)$ (possibly a singleton) of agents where for each $j \in \tilde{N}(i)$, the econometrician observes Y_j, $|N_P(j) \cap N_P(k)|$, $n_P(k)$ and X_k for all $k \in N_{P,2}(j)$.

Similarly as before, we consider the following test statistics:

$$\widehat{IU}^{\mathsf{FS}} = \inf_{\beta \in C^{\beta}_{1-(\alpha/2)}} \frac{1}{2(\hat{S}^{\mathsf{FS}}(\beta))^4 n^*} \left(\sum_{i \in N^*} \sum_{j \in \widetilde{N}(i)} \hat{v}_i^{\mathsf{FS}}(\beta) \hat{v}_j^{\mathsf{FS}}(\beta) \right)^2,$$

where

$$(\hat{S}^{\mathsf{FS}}(\beta))^2 = \frac{\widehat{d}_{av}^{1/2}}{n^*} \sum_{i \in N^*} \hat{v}_i^2(\beta).$$

As before, we reject the null hypothesis of no information sharing on unobservables if and only if $\widehat{IU}^{\mathsf{FS}} > c_{1-(\alpha/2)}$, where $c_{1-(\alpha/2)}$ is the $(1-(\alpha/2))$-percentile of χ^2_1.

2.4 A Monte Carlo Simulation Study

In this section, we investigate the finite sample properties of the asymptotic inference across various configurations of the payoff graph, G_P. The payoff graphs are generated according to two models of random graph formation, which we call Specifications 1 and 2. Specification 1 uses the Barabási-Albert model of preferential attachment, with m representing the number of edges each new node forms with existing nodes. The number m is chosen from $\{1,2,3\}$. Specification 2 is the Erdös-Rényi random graph with probability $p = \lambda/n$, where λ is also chosen from $\{1,2,3\}$.[17] In the first table, we report degree characteristics of the payoff graphs used in the simulation study.

For the simulations, we also set the following: $\rho_0 = (\gamma_0', \delta_0')'$, with $\gamma_0 = (2,4,1)'$, and $\delta_0 = (3,4)'$. We choose a to be a vector of ones so that $a'\rho_0 = 14$. The variables ε and η are drawn i.i.d. from $N(0,1)$. The first column of $X_{i,1}$ is a column of

[17]Note that in Specification 1, the Barabási-Albert graph is generated with an Erdös-Rényi seed graph, where the number of nodes in the seed is set to equal the smallest integer above $5\sqrt{n}$. All graphs in the simulation study are undirected.

Table 2.2: The Degree Characteristics of the Graphs Used in the Simulation Study

n		Specification 1 $m=1$	$m=2$	$m=3$	Specification 2 $\lambda=1$	$\lambda=2$	$\lambda=3$
500	d_{mx}	17	21	30	5	8	11
	d_{av}	1.7600	3.2980	4.8340	0.9520	1.9360	2.9600
1000	d_{mx}	18	29	34	6	7	9
	d_{av}	1.8460	3.5240	5.2050	0.9960	1.9620	3.0020
5000	d_{mx}	32	78	70	7	10	11
	d_{av}	1.9308	3.7884	5.6466	0.9904	2.0032	3.0228

Notes: This table gives characteristics of the payoff graphs, G_P, used in the simulation study. d_{av} and d_{mx} denote the average and maximum degrees of the payoff graphs.

ones, while the remaining columns of $X_{i,1}$ are drawn independently from $N(1,1)$. The columns of $X_{i,2}$ are drawn independently from $N(3,1)$.

For instruments, we consider the following nonlinear transformations of X_1 and X_2:

$$\varphi_i = [\tilde{Z}_{i,1}, X_{i,1}^2, X_{i,2}^2, X_{i,2}^3]',$$

where we define

$$\tilde{Z}_{i,1} \equiv \frac{1}{n_P(i)} \sum_{j \in N_P(i)} \lambda_{ij} X_{j,1}.$$

We generate Y_i from the best response function in Theorem 2.1. While the instruments $X_{i,1}^2, X_{i,2}^2, X_{i,2}^3$ capture the nonlinear impact of X_i's, the instrument $\tilde{Z}_{i,1}$ captures the cross-sectional dependence along the payoff graph. The use of this instrumental variable in crucial in obtaining a sharp inference for β_0. Note that since we have already concentrated out ρ in forming the moment conditions, we cannot use linear combinations of $X_{i,1}$ and $X_{i,2}$ as our instrumental variables. The nominal size in all the experiments is set at $\alpha = 0.05$.

Overall, the simulation results illustrate the good power and size properties for the asymptotic inference on β_0 and $a'\rho_0$. As expected, the average length of the confidence intervals for both β_0 and $a'\rho_0$ become shorter as the sample size

Table 2.3: The Empirical Coverage Probability of the Confidence Intervals for β_0 at 95% Nominal Level.

		Specification 1			Specification 2		
β_0		$m=1$	$m=2$	$m=3$	$\lambda=1$	$\lambda=2$	$\lambda=3$
-0.5	$n=500$	0.9760	0.9715	0.9710	0.9740	0.9685	0.9700
	$n=1000$	0.9665	0.9720	0.9660	0.9785	0.9705	0.9755
	$n=5000$	0.9745	0.9705	0.9680	0.9715	0.9765	0.9690
-0.3	$n=500$	0.9730	0.9690	0.9690	0.9740	0.9690	0.9655
	$n=1000$	0.9665	0.9710	0.9660	0.9750	0.9700	0.9710
	$n=5000$	0.9700	0.9670	0.9645	0.9710	0.9755	0.9645
0	$n=500$	0.9610	0.9660	0.9715	0.9735	0.9670	0.9625
	$n=1000$	0.9640	0.9690	0.9690	0.9745	0.9665	0.9670
	$n=5000$	0.9685	0.9670	0.9655	0.9710	0.9745	0.9660
0.3	$n=500$	0.9640	0.9665	0.9725	0.9780	0.9720	0.9675
	$n=1000$	0.9670	0.9695	0.9655	0.9770	0.9700	0.9730
	$n=5000$	0.9705	0.9660	0.9635	0.9725	0.9755	0.9675
0.5	$n=500$	0.9725	0.9735	0.9735	0.9800	0.9715	0.9700
	$n=1000$	0.9690	0.9705	0.9660	0.9770	0.9770	0.9785
	$n=5000$	0.9720	0.9695	0.9650	0.9770	0.9755	0.9730

Notes: The table reports the empirical coverage probability of the asymptotic confidence interval for β_0. The simulated rejection probability at the true parameter is close to the nominal size of $\alpha = 0.05$. The simulation number is $R = 2000$.

increases. We find that the confidence interval for β_0 exhibits empirical coverage close to the 95% nominal level, while the confidence interval for $a'\rho_0$ is somewhat conservative. This conservativeness is expected, given the fact that the interval is constructed using a Bonferroni approach.

2.5 Empirical Application: State Presence across Municipalities

2.5.1 Motivation and Background

State capacity (i.e., the capability of a country to provide public goods, basic services, and the rule of law) can be limited for various reasons. (See e.g. Besley

Table 2.4: The Average Length of Confidence Intervals for β_0 at 95% Nominal Level.

β_0		Specification 1			Specification 2		
		$m=1$	$m=2$	$m=3$	$\lambda=1$	$\lambda=2$	$\lambda=3$
-0.5	$n=500$	0.2735	0.3069	0.3334	0.3318	0.2933	0.2832
	$n=1000$	0.2264	0.2513	0.2749	0.2874	0.2489	0.2491
	$n=5000$	0.1771	0.1783	0.1971	0.2368	0.1917	0.1818
-0.3	$n=500$	0.2541	0.2971	0.3245	0.3189	0.2818	0.2703
	$n=1000$	0.2132	0.2445	0.2680	0.2810	0.2398	0.2412
	$n=5000$	0.1710	0.1739	0.1952	0.2387	0.1901	0.1791
0	$n=500$	0.2434	0.2881	0.3175	0.3043	0.2692	0.2606
	$n=1000$	0.2051	0.2377	0.2627	0.2730	0.2316	0.2332
	$n=5000$	0.1639	0.1715	0.1925	0.2373	0.1877	0.1746
0.3	$n=500$	0.2481	0.2891	0.3149	0.2982	0.2652	0.2629
	$n=1000$	0.2114	0.2377	0.2639	0.2673	0.2260	0.2314
	$n=5000$	0.1657	0.1729	0.1913	0.2325	0.1884	0.1710
0.5	$n=500$	0.2613	0.2974	0.3186	0.2977	0.2693	0.2691
	$n=1000$	0.2193	0.2429	0.2688	0.2620	0.2267	0.2350
	$n=5000$	0.1703	0.1770	0.1912	0.2253	0.1851	0.1687

Notes: The table reports average length of the asymptotic confidence interval for β_0. The average lengths of the confidence intervals decrease with n. The simulation number is $R = 2000$.

and Persson [19] and Gennaioli and Voth [59]). A "weak state" may arise due to political corruption and clientelism, and result in spending inadequately on public goods (Acemoglu [3]), accommodating armed opponents of the government (Powell [100]), and war (McBride et al. [92]). Empirical evidence has shown how these weak states can persist from precolonial times. Higher state capacities seem related to the current level prosperity at the ethnic and national levels (Gennaioli and Rainer [58] and Michalopoulos and Papaioannou [95]).

Our empirical application is based on a recent study by Acemoglu et al. [4] who investigate the local choices of state capacity in Colombia, using a model of a complete information game on an exogenously formed network. In their setup, municipalities choose a level of spending on public goods and state presence (as measured by either the number of state employees or state agencies). There is network externality in a municipality's choice because municipalities that are

Table 2.5: The Empirical Coverage Probability of Confidence Intervals for $a'\rho_0$ at 95% Nominal Level.

β_0		Specification 1			Specification 2		
		$m=1$	$m=2$	$m=3$	$\lambda=1$	$\lambda=2$	$\lambda=3$
-0.5	$n=500$	0.9945	0.9955	0.9925	0.9955	0.9950	0.9910
	$n=1000$	0.9925	0.9930	0.9955	0.9955	0.9970	0.9910
	$n=5000$	0.9850	0.9875	0.9900	0.9915	0.9925	0.9930
-0.3	$n=500$	0.9935	0.9960	0.9940	0.9960	0.9950	0.9905
	$n=1000$	0.9905	0.9935	0.9955	0.9955	0.9960	0.9910
	$n=5000$	0.9795	0.9880	0.9895	0.9885	0.9940	0.9895
0	$n=500$	0.9935	0.9965	0.9945	0.9955	0.9925	0.9925
	$n=1000$	0.9925	0.9935	0.9945	0.9960	0.9970	0.9920
	$n=5000$	0.9785	0.9875	0.9895	0.9840	0.9910	0.9860
0.3	$n=500$	0.9950	0.9940	0.9935	0.9955	0.9915	0.9925
	$n=1000$	0.9940	0.9940	0.9940	0.9920	0.9965	0.9925
	$n=5000$	0.9790	0.9895	0.9925	0.9790	0.9860	0.9850
0.5	$n=500$	0.9940	0.9925	0.9935	0.9940	0.9915	0.9920
	$n=1000$	0.9945	0.9940	0.9945	0.9890	0.9930	0.9930
	$n=5000$	0.9835	0.9900	0.9910	0.9645	0.9850	0.9860

Notes: The true $a'\rho_0$ is equal to 14.The table reports the empirical coverage probability of the asymptotic confidence interval. The coverage probability for $a'\rho_0$ is generally conservative, which is expected from the use of the Bonferroni approach. Nevertheless, the length of the confidence interval is reasonably small. The simulation number, R, is 2000.

adjacent to each other can benefit from their neighbors' choices of public goods provisions, such as increased security, infrastructure and bureaucratic connections. Thus, a municipality's choice of state capacity can be thought of as a strategic decision on a geographic network.

It is not obvious that public good provision in one municipality leads to higher spending on public goods in neighboring municipalities. Some neighbors may free-ride and under-invest in state presence if they anticipate others will invest highly. Rent-seeking by municipal politicians would also limit the provision of public goods. On the other hand, economies of scale could yield complementarities in state presence across neighboring municipalities.[18]

[18]Note that our analysis excludes the possibility that the provision of of public goods by municipal-

Table 2.6: The Average Lengths of Confidence Intervals for $a'\rho_0$ at 95% Nominal Level.

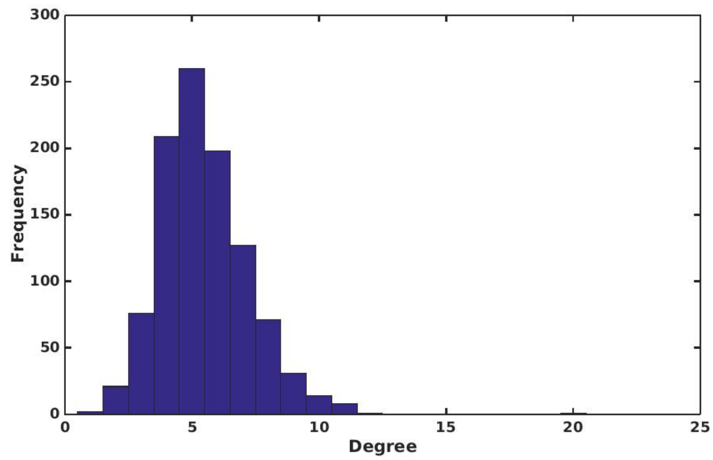
β_0		Specification 1			Specification 2		
		$m=1$	$m=2$	$m=3$	$\lambda=1$	$\lambda=2$	$\lambda=3$
-0.5	$n=500$	1.5840	1.6752	1.7643	1.3326	1.4123	1.8075
	$n=1000$	1.3626	1.4522	1.4819	1.1218	1.1474	1.4152
	$n=5000$	0.9995	0.9979	1.0333	0.8658	0.7786	0.8503
-0.3	$n=500$	1.5337	1.6361	1.7225	1.4037	1.4196	1.7567
	$n=1000$	1.3263	1.4140	1.4466	1.2028	1.1714	1.3966
	$n=5000$	0.9896	0.9749	1.0145	0.9656	0.8244	0.8650
0	$n=500$	1.5068	1.6007	1.6761	1.5646	1.4632	1.6976
	$n=1000$	1.3060	1.3607	1.4031	1.3721	1.2259	1.3798
	$n=5000$	0.9840	0.9486	0.9873	1.1527	0.9154	0.8922
0.3	$n=500$	1.5516	1.6019	1.6501	1.8290	1.5653	1.6869
	$n=1000$	1.3416	1.3257	1.3754	1.6133	1.3204	1.3959
	$n=5000$	1.0066	0.9412	0.9690	1.3938	1.0199	0.9213
0.5	$n=500$	1.6553	1.6353	1.6567	2.1022	1.6772	1.7101
	$n=1000$	1.4069	1.3146	1.3731	1.8362	1.4272	1.4376
	$n=5000$	1.0552	0.9420	0.9542	1.5865	1.1029	0.9593

Notes: The true $a'\rho_0$ is equal to 14. The table reports the average lengths of the asymptotic confidence interval for $a'\rho_0$. The length of the confidence interval is reasonably small. The simulation number, R, is 2000.

In our study, we extend the model in Acemoglu et al. [4] to an incomplete information game where information may be shared across municipalities. In particular, we do not assume that all municipalities know and observe all characteristics and decisions of the others. It seems reasonable that the decisions made across the country may not be observed or well known by those municipalities that are geographically remote.

ities is related to the preferences of agents who may sort into regions according to their preferences for such public goods (as may be the case in models based on the insights of Tiebout [108]).

Figure 2.3: Degree Distribution of G_P



Notes: The figure presents the degree distribution of the graph G_P used in the empirical specification. The average degree is 5.48, the maximum degree is 20, and the minimum degree is 1.

2.5.2 Empirical Set-up

Let y_i denote the state capacity in municipality i (as measured by the (log) number of public employees in municipality i) and G_P denote the geographic network, where an edge is defined on two municipalities that are geographically adjacent.[19] We assume that G_P is exogenously formed. The degree distribution of G_P is shown in Figure 2.3. We study the optimal choice of y_i, where y_i leads to a larger prosperity p_i. Prosperity in municipality i is modeled as:

$$p_i = \left(\beta \bar{y} + x_{1,i}\gamma + \eta_i + \varepsilon_i + \varsigma_i^D\right) y_i, \tag{2.25}$$

where ς_i^D is a district specific dummy variable, ε_i and η_i are our sharable and non-sharable private information, and $\bar{y} = \frac{1}{n_P(i)} \sum_{j \in N_P(i)} y_j$. The term $x_{1,i}$ represents municipality characteristics. These include geographic characteristics, such as land quality, altitude, latitude, rainfall; and municipal characteristics, such as distance to highways, distance to royal roads and Colonial State Presence.[20]

[19] This corresponds to the case in of $\delta_1 = \delta_2 = 0$ in Acemoglu et al. [4].

[20] Note that, from our notation in Section 3, here we take $x_{2,i} = 0$. This is done for a closer correspondence to the specification in Acemoglu et al. [4]. Finally, note that p_i is only a function of

The welfare of a municipality is given by

$$u_i(y_i, y_{-i}, T, \eta_i) = p_i(y_i, \bar{y}, T, \eta_i) - \frac{1}{2}y_i^2, \tag{2.26}$$

where the second term refers to the cost of higher state presence, and the first term is the prosperity p_i.

We can rewrite the welfare of the municipality by substituting (2.25) into (2.26):

$$u_i(y_i, y_{-i}, T, \eta_i) = \left(\beta\bar{y} + x_{1,i}\gamma + \eta_i + \varepsilon_i + \varsigma_i^D\right) y_i - \frac{1}{2}y_i^2, \tag{2.27}$$

which is our model from Section 3. We assume that municipalities (or the mayor in charge), wishes to maximize welfare by choosing state presence, given their beliefs about the types of the other municipalities.

In our specification, we allow for incomplete information. This is reflected in the terms ε_i, η_i, which will be present in the best response function. The municipality, when choosing state presence y_i, will be able to observe ε_i of its neighbors and will use its beliefs over the types of the others to generate its best response. The best response will follow the results from Theorem (2.2.1).

2.5.3 Model Specification

We follow closely Table 3 in Acemoglu et al. [4] for the choice of specifications and variables. Throughout the specifications, we include longitude, latitude, surface area, elevation, annual rainfall, department fixed effects and a department capital dummy (all in X_1). We further consider the effect of variables distance to current highways, land quality and presence of rivers in the municipality.

For the choice of instruments, we consider two separate types of instruments. The first is the sum of neighbor values (across G_P) of the historical variables (denoted as C_i).[21] The historical variables used are Total Crown Employees (also

terms are multiplied by y_i. This is a simplification from their specification. We do so because we will focus on the best response equation. The best response equation, derived from the first order condition to this problem, would not include any term that is not a function of y_i itself.

[21] For this, we assume the exclusion restriction in Acemoglu et al. [4], namely that historical variables only affect prosperity in the same municipality. This means that although one's historical variables (Total Crown Employees, Distance to Royal Roads, Colonial State Agencies and Historical Population, as well as functions thereof) can affect the same municipality's prosperity, it can only

called Colonial State Officials), Distance to Royal Roads, Colonial State Agencies and Historical Population, as well as Colonial State Presence Index squared and Distance to Royal Roads squared. The later two add additional power to the inference. We also use the variable $\tilde{Z}_i = n_P(i)^{-1} \sum_{j \in N_P(i)} \lambda_{ij} X_{j,1}$ as part of the instrumental variable, which was shown to perform well in the Monte Carlo Simulations in Section 5. This variable captures cross sectional dependence as a crucial source of variation for inference on the strategic interactions. We use downweighting of our instruments as explained in a preceding section.

2.5.4 Results

The results across a range of specifications are presented in Table 2.7. In these results, we see that the effect is statistically different than 0 and stable across specifications. It indicates that there is complementarity in the provision of public goods and state presence ($\beta > 0$).

Let us compare our results to those in Acemoglu et al. [4]. There, the authors report the average marginal effects over their weighted graph. The (weighted) average degree is 0.0329, so our results can be compared in an approximation, by considering 0.0329 $\hat{\beta}$.

In general, our estimates have the same sign and significance as those of Acemoglu et al. [4]. Our estimates are in the range of [0.002, 0.013], after reweighting as mentioned before, somewhat comparable to theirs of [0.016, 0.022] (in the case of the outcome of the number of public employees, in Table 3). Hence, we find similar qualitative effects, although a smaller magnitude. Recall that our confidence set is built without assuming that β_0 is consistently estimable.

In Figure 2.4, we show the results of our estimated network externalities for the estimates from Table 2.7, for the importance of being a department capital. The average network externality is computed from

$$\frac{1}{N} \sum_{i \in N} \frac{1}{n_P(i)} \sum_{j \in N_P(i)} \frac{\beta_0 \hat{\gamma}_{dc}}{n_P(i)(1-\beta_0 c_{ii})(1-\beta_0 c_{ij})},$$

affect those of the neighbors by impacting the choice of state capacity in the first, which then impacts the choice of the state capacity in the neighbors.

Table 2.7: State Presence and Networks Effects across Colombian Municipalities

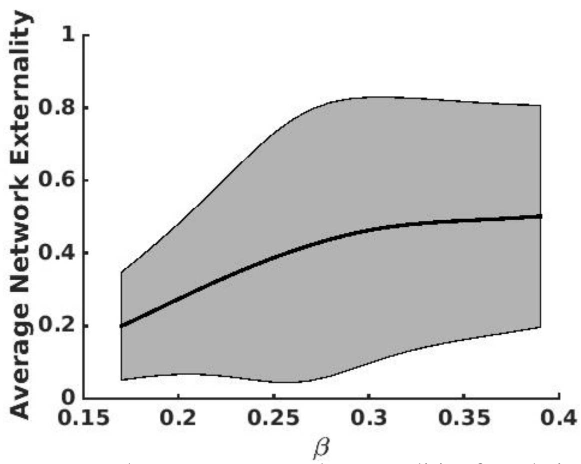
	Outcome: The Number of State Employees			
	Baseline	Distance to Highway	Land Quality	Rivers
	(1)	(2)	(3)	(4)
β_0	$[0.16, 0.31]$	$[0.15, 0.32]$	$[0.17, 0.39]$	$[0.09, 0.38]$
dy_i/d(colonial state officials)	$[-0.051, 0.001]$	$[-0.045, -0.001]$	$[-0.043, 0.000]$	$[-0.024, 0.003]$
Average dy_i/d(colonial state agencies)	$[-1.138, 3.760]$	$[-1.335, 2.742]$	$[-0.609, 3.388]$	$[-1.775, 1.987]$
Average dy_i/d(distance to Royal Roads)	$[-0.010, 0.009]$	$[-0.008, 0.010]$	$[-0.007, 0.015]$	$[-0.005, 0.012]$
n	1018	1018	1003	1003

Notes: Confidence sets for β are presented in the table, obtained from inverting the test statistic $T(\beta)$ from Section 3, with confidence level of 95%. The critical values in the first row come from the asymptotic statistic. Downweighting is used. The average marginal effects for historical variables upon state capacity are also shown. This is computed from finding the confidence set for the appropriate γ estimate. For Colonial State Agencies and Distance to Royal Roads, since they enter in quadratic form in X_1, we show the average marginal effect. All specifications include controls of latitude, longitude, surface area, elevation, rainfall, as well as Department and Department capital dummies. Instruments are constructed from payoff neighbors' sum of the G_P neighbors values of the historical variables Total Crown Employees, Colonial State Agencies, Colonial State Agencies squared, population in 1843, distance to Royal Roads, distance to Royal Roads squared, together with the non-linear function $\tilde{Z}_i = n_P(i)^{-1} \sum_{j \in N_P(i)} \lambda_{ij} X_{j,1}$. Column (2) includes distance to current highway in X_1, Column (3) expands the specification of Column (2) by also including controls for land quality (share in each quality level). Column (4) controls for rivers in the municipality and land quality, in addition to those controls from Column (1). One can see that the results are very stable across specifications.

where $\hat{\gamma}_{dc}$ is the estimated parameter of the X_1 variable department capital, and we vary β_0 within its confidence set. The parameter is defined in Section 3.1.2, and captures the average effect of a neighbour being a department capital.

The figure shows that there is a strong and increasing network externality from being a department capital over the range of the confidence set of β. This indicates that the effect of being a capital has spillovers on other municipalities: since $\beta > 0$, and one expects that department capitals have more state presence and resources,

Figure 2.4: Average Network Externality from being a Department Capital



Notes: The figure presents the average network externalities from being a department capital. We use the estimated results from Column (3) in Table 2.7. This captures the externality for a municipality from being a department capital, which involves higher state presence and centralization of resources. This effect is not only the direct effect, but it also quantifies a reflection effect: neighbors of department capitals also benefit from it. The grey shaded area represents the 95% confidence interval for β_0.

being a department capital yields increasing returns the stronger the complementarity.

2.6 Conclusion

This chapter proposes a new approach of empirical modeling for interactions among many agents when the agents observe the types of their neighbors. The main challenge arises from the fact that the information sharing relations are typically connected among a large number of players whereas the econometrician observes only a fraction of those agents. Using a behavioral model of belief formation, this chapter produces an explicit form of best linear responses from which an asymptotic inference procedure for the payoff parameters is developed. As we showed in our chapter, this explicit form gives a reduced form for the observed actions, and exhibits various intuitive features. For example, the best linear responses show that network externality is heterogeneous across agents depending on the relations of their payoff neighbors.

The advantage of our chapter's approach is two-fold. First, the empirical modeling according to our approach accommodates a wide range of sampling processes. Such a feature is crucial because the econometrician rarely has precise knowledge about the actual sampling process through which data are generated. Second, the model can be used when only part of the players are observed from a large connected network of agents.

Chapter 3

Identification Using Attrition

3.1 Introduction

[Attrition is] perhaps the most potentially damaging and frequently-mentioned threat to the value of panel data.
— Moffit, Fitzgerald, and Gottschalk (1999)

Shortly after Heckman's seminal 1976 work on sample selection bias (Heckman [69]), the author diagnosed a related problem - that of selection bias due to non-random sample attrition.[1] Since then, a large literature has emerged proposing solutions to the problem (e.g., Hausman and Wise [66], Moffit et al. [96]).

This chapter discusses the value associated with the information contained in attrition patterns. The key insight is as follows: provided that agents attrit from a sample differently according to their unobserved attributes, the researcher can use this information to back out the distribution of unobserved heterogeneity. Overall, information on attrition can be used to address a variety of interesting identification problems - both in economics and beyond.

A key related work that recognizes the value of attrition for identification is Bellemare [16]. The author uses sample attrition to identify the probability that a migrant leaves his adopted country (termed "outmigration") when this decision is unobserved by the researcher. Using identification arguments from Lewbel [84],

[1] See discussion in Moffit et al. [96]

the author shows that the outmigration probability can be identified from sample attrition when the researcher observes a continuous variable that affects attrition solely through outmigration via a single index. The author stresses that, in contrast with the literature on mismeasured binary variables (e.g., Hausman et al. [68], Lewbel [84]), the key properties of attrition imply that missclassification is only possible when attrition occurs. One purpose of this chapter is to stress that attrition may be informative for unboserved heterogeneity in a more general sense.

This chapter is related to the literature on identifying nonlinear models with measurement error. One way to view the role of attrition in this chapter is as a type of (possibly repeated) measurement of the unobserved quantity. However, there are at least two important differences between attrition and repeated measurements as they are typically considered in the nonlinear measurement error literature. The first difference recognizes the strong possibility that an agent's attrition decisions will be correlated across time - even conditional on the unobserved heterogeneity itself. In Cunha et al. [43], the distribution of unobservables are identified when repeated measurements of the unseen quantity satisfy an assumption of mutual conditional dependence.[2,3] A second key difference between attrition and traditional measurements is the tendency for attrition to be irrevocable, which imparts a special structure to the data. A common data structure when the researcher sees whether or not an agent of has attrited in a T-period sample is as follows: if the agent has attrited at any period t with $t \leq T$, the researcher sees a $t \times 1$ vector of ones with a zero in the the last entry, and a $T \times 1$ vector of ones otherwise. Throughought this chapter, we will maintain the assumption that attrition is an absorbing state - in other words, once an agent has attrited from the researcher's sample, they are gone from that sample for good.[4]

Far from being a nuisance - dependence in the attrition process over time may be precisely what is informative for the distribution of unobserved heterogeneity under certain functional form restrictions - even when there are no covariates.

[2] This is Assumption 2 of Theorem 2 of Cunha et al. [43]. This is also equivalent to Assumption 2 of Hu and Schennach [72] (see Cunha et al. [44]).

[3] In fact, in the absence of additional information, Lemma 3.2.1 shows how such conditional independence can lead to non-identification. Examples are then provided where parametric assumptions and temporal dependence may restore identification.

[4] This is what Robins et al. [101] call a monotone missing data pattern.

In Section 3.2.3, we show how an analogue of the hazard function familiar from the econometrics of duration analysis arises as a natural byproduct of a particular specification of the attrition process. Thus, this chapter is related to a literature that shows how duration models can be used to identify unobserved heterogeneity.[5]

Another relevant related literature is that concerned with identification in non-linear panel models with mismeasurement. Like Hausman et al. [67], Schennach [106], we consider repeated measurements of the mismeasured variable. Although we use repeated measurements of attrition as information on heterogeneity, the relationship is clearly not linear. Moreover, the models we discuss are not strictly of a panel nature, since the researcher does not need to observe outcome variable for more than one period.

As the main goal of this chapter is to highlight the usefulness of attrition data for identification in as simple an exposition as possible, we restrict ourselves to discrete forms of unobserved heterogeneity and focus on the problem of identification in parametric models.[6] Moreover, to stress the usefulness of focusing on discrete unboserved heterogeneity, we center much of our discussion of identification on a new parameter known as the Type-Targeted Treatment Effect (TTTE).[7] In our focus of identifying the TTTE, we will suppose that the main outcome variables satisfy a non-differential error assumption.[8] In other words, attrition doesn't affect the outcome of interest conditional on the unobserved heterogeneity.

The remainder of this chapter is structured as follows. In Section 3.2 we introduce the notion of the TTTE and discusses identification. In Section 3.3 we propose a minimum distance inference procedure appropriate for our setup. A short Monte Carlo simulation shows the procedure performs acceptably in finite samples. In Section 3.4 we show how - in conjunction with the preceding discussions on identification - attrition data can be used to correct for selection on unobservables.

[5]For example, e.g., see pages 88-93 in Heckman and Singer [70]) for a discussion of non-parametric identification of duration models

[6]Williams [111] show how continuous forms of heterogeneity may not be identified in a models with binary proxies.

[7]This parameter was first proposed in work-in-progress the author is pursuing with Hugo Borges Jales, titled "Type-Targeted Treatment Effects".

[8]This assumption has been employed widely in the measurement error literature (e.g., see Hu [71] Assumption 1 and associated references).

3.2 Identification of the Type-Targeted Treatment Effect

The Average Treatment Effect (ATE) is a parameter informative for the mean change in outcome induced by assigning someone to one treatment (or policy) state of interest compared to another. The ATE is widely used for evaluating treatments in many fields for its simplicity of implementation and ease of interpretation. Despite these advantages, a naïvely deployed ATE may lead the researcher astray. For instance, a researcher interested in the mean impact of a new drug versus the industry standard would ideally select a sample of individuals who actually have the disease the drug is designed to treat. Unfortunately, a researcher will likely have to face an imperfectly composed sample, due to the difficulty of accurately observing the health status of each patient when the decision to admit them into the study is made.

We suppose that an agent can have one of two possible types - either the agent is a 'target' type, θ_1, or a 'mistarget', θ_0. Let $\tilde{Y}_i = Y_{i1} - Y_{i0}$, where Y_{i1} represents the person's potential outcome when they receive the treatment and Y_{i0} represents their potential outcome without the treatment. For example, in a study of the impact of financial assistance on graduate student outcomes, the type might capture whether the student is an 'academic type' (e.g., has a talent and passion for research), and $\tilde{Y}_i$ may represent a student's potential gain from receiving financial assistance in terms of the student's course grades. We denote the probability of targeted agents in the population as $p = P(\theta_i = \theta_1)$. By the law of iterated expectations, we can write the standard average treatment effect (ATE) [9] as

$$a = a_1 p + a_0(1-p), \tag{3.1}$$

where the type-targeted treatment and (mis-)targeted treatment effect are defined

[9]Conditioning on covariates is trivial, so we omit them for notational convenience. At this stage, the reader can implicitly assume that treatment assignment is independent of potential outcomes, or, as in a model with covariates, conditionally independent of potential outcomes. This simplification is done to highlight the main issues of unobserved heterogeneity and attrition.

as

$$a_1 \equiv \mathbf{E}[\tilde{Y}_i|\theta_1], \text{ and}$$
$$a_0 \equiv \mathbf{E}[\tilde{Y}_i|\theta_0],$$

where we express $\mathbf{E}[\tilde{Y}_i|\theta] = \mathbf{E}[\tilde{Y}_i|\theta_i = \theta]$ for each type for notational convenience. The average treatment effect is a weighted average of the treatment effect on the targets a_1, and the mistargets, a_0, where weights are given by the proportion of individuals from each type in the sample. Since the effect of the policy on the targets is the true object of interest, we can think of a_1 as a 'Type-Targed' Treatment Effect (TTTE). Provided that p is strictly less than one, then there will necessarily be some mistargets in the population. Depending on how potential outcomes are affected by type, the ATE and TTTE may differ substantially. For instance, if p is low then a may be very small even when a_1 is not. In such a case, a researcher who considers the ATE and ignores the possibility of mistargeting may erroneously conclude that financial assistance doesn't improve student outcomes, or - in the medical example - a drug is not an improvement over the status quo for the subjects of interest.

Type mistargeting is also a relevant issue in economics more broadly. When examining how receiving disability insurance affects the labour supply decision (e.g., Maestas et al. [89]), it is natural to question what impact such policies would have on the subpopulation of agents who actually suffer from chronic disability. The question is not obvious, since disability status is typically only imperfectly observed. In welfare reform studies, policy changes such as workfare legislation may make it less desirable for noncompliers to receive welfare in both the treatment and control groups of the study.[10]

[10] Economic theory can help us understand how people of different types respond to such policy variations. For example, in Besley and Coate [18] work requirement legislation can induce full separation by type; i.e., only the needy find it incentive compatible to continue collecting welfare payments after the policy change is introduced. Here the policy acts as a screening mechanism.

3.2.1 General Setup

Agents have one of two types, θ_1 or θ_0, where θ_1 is the target type of individuals in the study and θ_0 is the mistargeted type. Our object of identification is the type targeted treatment effect on the targets, a_1, untreated, a_0, and the marginal type probability p. In this section, we will write the parameters of interest as functions of differences in the potential outcomes, to avoid discussing separate identification that have been well explored in the literature. In all that follows, we assume that the econometrician observes some outcomes, Y_i, a treatment, D_i, and a variable $\{L_{ij}\}_{j=0}^{J_i}$ for all agents, where $J_i \geq 1$ for all i. L_{ij} is a variable that takes value l_0 if agent i fails the jth test and l_1 if the agent passes a test. We can think of "tests" as institutional changes or policy interventions, the "score" being a binary behavioral response to the policies, such as attrition from the econometrician's sample. We also suppose that once an agent attrits from the sample they are gone for good - that is; $L_{ij} = l_0$ implies that $J_i = j$.

The broad intuition of our approach is as follows: once the outcomes of agents are observed, behavior of participants that is both type-specific and observable may, under certain assumptions, allow the econometrician to compute causal effects in a way that is targeted to the subpopulation of interest.

3.2.2 Identification Under One Period of Attrition

In this section, we consider what can be learned when the econometrician observes at most one test for each agent. That is, $J_i = 1$ for all i. By the Law of Iterated expectations, we know that the ATE and ATE conditional on conditional on passing one test can be written as:

$$\mathbf{E}[\tilde{Y}_i|L_i = l_1] = \mathbf{E}[\tilde{Y}_i|\theta_1, l_1]P(\theta_1|l_1) + \mathbf{E}[\tilde{Y}_i|\theta_0, l_1]P(\theta_0|l_1).$$

We now introduce an assumption that reduces the dimensionality of the identification problem.

Assumption 3.2.1. For each i, $\tilde{Y}_i$ is independent of L_i conditional on θ_i.

Assumption 3.2.1 implies that the difference in the agent's potential outcomes are conditionally independent of the previous test outcomes once the agent's true type is accounted for. Effectively, the only way the treatment effect varies once conditioning on test passing is through changes in the composition of each type.[11] The reasonability of Assumption 3.2.1 should be carefully justified for each application. This condition fails if the probabilty that an individual of a particular type passes a test is conditionally correlated with her gain from treatment. For example, suppose that L_i represents whether or not a student decides drop out of a graduate program during the first year. Then Assumption 3.2.1 is reasonable if the researcher believes that - conditional on the student's type - the difference in the student's first term grade-point average with and without financial assistance, $\tilde{Y}_i$, does not influence their decision to drop out of the program. Under Assumption 3.2.1 , the previous equation simplifies to:

$$\mathbf{E}[\tilde{Y}_i|l_1] = a_1\lambda p/\ell + (1-\lambda p/\ell)a_0,$$

where $\lambda = P(L_i = l_1|\theta_i = \theta_1)$, and $\ell = P(L_i = l_1)$. Therefore, provided that $p > 0$, the difference in TTTE, $\Delta \equiv a_1 - a_0$, can be written as:

$$\Delta = \frac{\left(\mathbf{E}[\tilde{Y}_i|l_1] - \mathbf{E}[\tilde{Y}_i]\right)\ell}{(\lambda-\ell)p}. \tag{3.2}$$

Since p and λ are unknown, Δ is unidentified in the absence of further information. However, the sign of Δ may be identified if the researcher knows that targeted types are strictly or more or less likely to pass than the mistargeted agents; e.g., $\lambda - \ell > 0$.

In the education example, the parameter a_1 captures the expected difference in the student's first year grades with and without financial assistance for those students of innately high quality, while a_0 captures the same for the students of low quality.

Identification Using Covariates

Since λ and p are unknown, we now explore how these can be learned when attrition varies according to both the agent's unknown type and some observable characteristics. Consider the joint distribution of attrition and a variable X_i with

[11] This is an exclusion restriction of the kind that is used in Section 4.1. of Bellemare [16].

support $\mathscr{X} \subset \mathbf{R}^m$:

$$P(L_i = l_1, X_i = x) = \sum_{\theta} P(l_1|x,\theta)P(x|\theta)P(\theta).$$

We will consider a parametric approach to identification. In particular, we will suppose that the researcher is willing to take a stand on attrition process.

Assumption 3.2.2. a) $L_i = 1\{\gamma'\psi(X_i,\theta_i) + \eta_i > 0\}$, where η_i is continuous with known cdf, Ψ, and where $\psi : \mathbf{R}^{m+1} \to \mathbf{R}^d$ is a known, non-stochastic function. b) X_i is independent of θ_i for each i.

Denote $g(\tau,x) \equiv P_\tau(l_1|x)$ for each $\tau \in \mathscr{T}$. Our model is identified if for all $x \in \mathscr{X}$, there exists no $\tau \neq \tau_0$ such that $P_\tau(l_1|x) = P(l_1|x)$. Under Assumption 3.2.2, we can express the conditional probability of passing a test as

$$g(\tau,x) = \Psi(\gamma'\psi(x,\theta_1))p + \Psi(\gamma'\psi(x,\theta_0))(1-p),$$

where $\tau = (\gamma', p)$. We can approach identification in our context by considering solutions to $k \geq d+1$ non-linear equations of the following form

$$g(\tau_0) - \pi_0 = 0, \tag{3.3}$$

where $g(\tau_0) = (g(\tau_0,x_1),...,g(\tau_0,x_k))'$ and $\pi_0 = (P(l_1|x_1),...,P(l_1|x_k))'$. Sufficient conditions for identification of τ_0 are that the parameter space, $\mathscr{T}$, is a compact subset of $\mathbf{R}^k$, $g(\tau)$ is continuous on $\mathscr{T}$, and the solution to equation 3.3 is unique. Demonstrating that equation 3.3 has a unique solution is not straightforward. Instead, we may argue that local identification should hold for many models of interest satisfying our functional form assumptions.[12] In our case, τ_0 is locally identified if g is continuously differentiable in τ over $\mathscr{T}$, and the Jacobian of $g(\tau_0)$ has rank k.[13] If p and γ are identified, then we may also establish the identification of the

[12] See Rothenberg [103], Sargan [105], and pages 2127 of Newey and McFadden [97].

[13] This is the standard condition for local identification in minimum distance models (see page 2128 of Newey and McFadden [97]).

difference in type targeted effects, Δ, from equation 3.2, since

$$\lambda = \sum_{x \in \mathscr{X}} g(\tau_0, x) P(x).$$

In order to separately identify a_1 and a_0, we may extend Assumption 3.2.2 so that X_i also satisfies an exclusion restriction.

Assumption 3.2.3. (i) $\tilde{Y}_i$ is independent of L_i conditional on θ_i and (ii) $\tilde{Y}_i$ is independent of X_i conditional on θ_i.

Assumption 3.2.3 says that the change in potential outcomes are not influenced by X_i once the type is taken into account. Returning to the education example from before, suppose X_i measures the student's experience in the private sector. Then, conditional on the student's ability, it may be reasonable to imagine that the student's private sector experience may influence their likelihood of dropping out of graduate school in the first year without affecting their first term grades (with and without financial assistance). The usefulness of Assumption 3.2.3 for identifying the model can be seen by considering equations of the form

$$\begin{aligned} \mathbf{E}[\tilde{Y}_i | L_i = l_1, X_i = x] = \mathbf{E}[\tilde{Y}_i | l_1, x, \theta_1] P(\theta_1 | l_1, x) \\ + \mathbf{E}[\tilde{Y}_i | l_1, x, \theta_0](1 - P(\theta_1 | l_1, x)). \end{aligned}$$

Under Assumption 3.2.3 we have:

$$\begin{aligned} \mathbf{E}[\tilde{Y}_i | l_1, x, \theta_1] =& a_1, \text{ and} \\ \mathbf{E}[\tilde{Y}_i | l_1, x, \theta_0] =& a_0. \end{aligned}$$

Therefore, using Bayes' Rule and Assumption 3.2.2 we have[14]

$$\begin{aligned} \mathbf{E}[\tilde{Y}_i | L_i = l_1, X_i = x] =& a_1 P(l_1 | x, \theta_1) P(l_1 | x) p \\ &+ a_0 \left(1 - P(l_1 | x, \theta_1) P(l_1 | x) p\right). \end{aligned}$$

[14]That is

$$P(\theta_1 | l_1, x) = \frac{P(l_1 | x, \theta_1) P(\theta_1, x)}{P(l_1, x)} = \frac{P(l_1 | x, \theta_1) P(\theta_1) P(x)}{P(l_1, x)} = P(l_1 | x, \theta_1) P(\theta_1) P(l_1 | x).$$

We wish to identify the following $d+3$ parameters: a_1, a_0, p and γ. For the purposes of this exposition, we use the identification of γ and p from the previous section. Denote

$$y(x) \equiv \mathbf{E}[\tilde{Y}_i | l_1, x], \text{ and}$$
$$c(x) \equiv P(l_1 | x, \theta_1) P(l_1 | x) P(\theta_1).$$

Then, provided $C^{-1}(x)$ exists, the TTTE, $\alpha = (a_1, a_0)$, are identified as

$$\alpha = C^{-1}(x) y,$$

where $y = (y(x_1), y(x_2))'$, and $C(x)$ is a 2x2 matrix with row r equal to $(c(x_r), 1 - c(x_r))'$.

Example Local Identification Using Covariates

In this section, we provide some simple examples of specifications for the attrition process that admit local identification. Take $L_i = 1\{\tau_1 \theta_i X_i + \eta_i > 0\}$, where $\tau_1 \in \mathbf{R}$, η_i is standard normal, θ_i takes values in $\{0, 1\}$, and X_i takes at least two distinct values. Then we can express the conditional probability of attrition given $X_i = x$ as

$$g(\tau, x) = \Psi(\tau_1 x) \tau_2 + \Psi(0)(1 - \tau_2),$$

where $\Psi(\cdot)$ is the standard normal cdf and $\tau_2 \in (0, 1)$. In this example, we may compute the Jacobian matrix, $J = \partial g(\tau)/\partial \tau'$, explicitly. The rank condition is satisfied when the determinant of J is nonzero. Therefore, τ is locally identified for the current example provided that

$$\frac{\Psi'(\tau_1)}{\Psi'(2\tau_1)2} \neq \frac{(\Psi(\tau_1) - \Psi(0))}{(\Psi(2\tau_1) - \Psi(0))}, \tag{3.4}$$

noting that we have assumed that X_i takes values in the set $\{1, 2\}$, without any loss of generality. Similar arguments can be used to establish the identification of τ_1, τ_2, and p when the attrition process is given by $L_i = 1\{\tau_1 X_i + \tau_2 \theta_i + \eta_i > 0\}$ (as in 3.8), provided that X_i has additional support points.

3.2.3 Identification Using Multiple Periods of Attrition

In this section, we investigate what can be learned from outcomes and attrition alone. That is, the researcher observes the outcome Y_i for all agents, two rounds of attrition L_{i1}, L_{i2} - but no covariates X_i. We will show that - even in the absence of additional information - the researcher can back out heterogeneity using attrition alone, provided they are willing to make parametric assumptions about the attrition process.

We are interested in identifying a_1, a_0, and p. Consider the average treatment effect conditional on subsequent rounds of the attrition variable:

$$\begin{aligned}\mathbf{E}[\tilde{Y}_i|L_{i1}=l_1] = &\mathbf{E}[\tilde{Y}_i|\theta_1, L_{i1}=l_1]\frac{P(L_{i1}=l_1|\theta)p}{P(L_{i1}=l_1)}\\ &-\mathbf{E}[\tilde{Y}_i|\theta_0, L_{i1}=l_1]\frac{P(L_{i1}=l_1|\theta)p}{P(L_{i1}=l_1)}\\ &+\mathbf{E}[\tilde{Y}_i|\theta_0, L_{i1}=l_1],\end{aligned}$$

and

$$\begin{aligned}\mathbf{E}[\tilde{Y}_i|L_{i1}=l_1, L_{i2}=l_1] = &\mathbf{E}[\tilde{Y}_i|\theta_1, L_{i1}=l_1, L_{i2}=l_1]\frac{P(L_{i1}=l_1, L_{i2}=l_1|\theta_1)p}{P(L_{i1}=l_1, L_{i2}=l_1)}\\ &-\mathbf{E}[\tilde{Y}_i|\theta_0, L_{i1}=l_1, L_{i2}=l_1]\frac{P(L_{i1}=l_1, L_{i2}=l_1|\theta_1)p}{P(L_{i1}=l_1, L_{i2}=l_1)}\\ &+\mathbf{E}[\tilde{Y}_i|\theta_0, L_{i1}=l_1, L_{i2}=l_1].\end{aligned}$$

To identify the model, we extend the conditional independence of tests on outcomes from Assumption 3.2.1 to accommodate multiple periods of attrition.

Assumption 3.2.4. $\tilde{Y}_i$ is independent of $\{L_{ij}\}_{j=1}^{J}$ conditional on θ_i, for all $J \geq 1$.

Under Assumption 3.2.4 we can rewrite the preceding equations as

$$\begin{aligned}\mathbf{E}[\tilde{Y}_i|L_{i1}=l_1] &= (a_1-a_0)\frac{P(L_{i1}=l_1|\theta)p}{P(L_{i1}=l_1)}+a_0, \text{ and}\\ \mathbf{E}[\tilde{Y}_i|L_{i1}=l_1, L_{i2}=l_1] &= (a_1-a_0)\frac{P(L_{i1}=l_1, L_{i2}=l_1|\theta_1)p}{P(L_{i1}=l_1, L_{i2}=l_1)}+a_0.\end{aligned}$$

Assumption 3.2.4 can also be illustrated using the education example we have been considering. The assumption holds provided some outcome of interest (such as first-year course grades) do not induce attrition at any subsequent phase of the program once we have controlled for the student's underlying type.

Non-Identification Under Conditionally Independent Attrition

When the attrition process is unknown, we have too many parameters to identify. To reduce the dimensionality of the problem, one possibility is to assume that L_{ij} is independent across j given θ_i. Regrettably, the following result shows that the model cannot be identified under such a strong assumption.

Lemma 3.2.1. *Suppose that Assumption 3.2.4 holds. Then (a_1, a_0, p, λ) are not identified when L_{ij} are independent across j given θ_i.*

Proof. Define $\ell_1 \equiv P(L_{i1} = l_1)$, $\ell_2 \equiv P(L_{i1} = l_1, L_{i2} = l_1)$, and $\ell_3 \equiv P(L_{i1} = l_1, L_{i2} = l_1, L_{i3} = l_1)$. Under the assumption that L_{ij} is independent across j given θ_i the system of equations that we consider to identify the model are

$$\begin{aligned}
\mathbf{E}[\tilde{Y}_i] &= a_1 p + (1-p)a_0, \\
\mathbf{E}[\tilde{Y}_i | L_{i1} = l_1] &= a_1 \lambda p/\ell_1 + a_0(1 - \lambda p/\ell_1), \\
\mathbf{E}[\tilde{Y}_i | L_{i1} = l_1, L_{i2} = l_1] &= a_1 \lambda^2 p/\ell_2 + a_0(1 - \lambda^2 p/\ell_2), \text{ and} \\
\mathbf{E}[\tilde{Y}_i | L_{i1} = l_1, L_{i2} = l_1, L_{i3} = l_1] &= a_1 \lambda^3 p/\ell_3 + a_0(1 - \lambda^3 p/\ell_3).
\end{aligned}$$

The third equality uses the assumption that $P(L_{i1} = l_1, L_{i2} = l_1 | \theta_i = \theta_1) = P(L_{i1} = l_1 | \theta_i = \theta_1)^2 \equiv \lambda^2$ (and similarly for the fourth equality). Recall that the Jacobian of this nonlinear system is defined as $J = [j_1, j_2, j_3, j_4]$, where for $k = 1, .., 4$,

$$j_k \equiv \left(\frac{\partial g_1}{\partial \theta_k}, \ldots, \frac{\partial g_4}{\partial \theta_k} \right)'.$$

Denote $\Delta = a_1 - a_0$. In our case, since

$$
\begin{aligned}
j_1 &= \left(p, \frac{\lambda p}{\ell_1}, \frac{\lambda^2 p}{\ell_2}, \frac{\lambda^3 p}{\ell_3}\right)' \\
j_2 &= \left(1-p, 1-\frac{p\lambda}{\ell_1}, 1-\frac{\lambda^2 p}{\ell_2}, 1-\frac{\lambda^3 p}{\ell_3}\right)', \\
j_3 &= \left(\Delta, \Delta\frac{\lambda}{\ell_1}, \Delta\frac{\lambda^2}{\ell_2}, \Delta\frac{\lambda^3}{\ell_3}\right)', \\
j_4 &= \left(0, \Delta\frac{p}{\ell_1}, \Delta\frac{2\lambda p}{\ell_2}, \Delta\frac{3\lambda^2 p}{\ell_3}\right)',
\end{aligned}
$$

we have that $\det(J) = 0$. This follows from the fact that, for the choice of $c_1 = -1$, $c_2 = 0$, $c_3 = p/\Delta$, $c_4 = 0$, we have

$$c_1 j_1 + c_2 j_2 + c_3 j_3 + c_4 j_4 = 0,$$

Hence the model is not identified. □

The following section illustrates the value of dependence in the attrition process for identification.

Example of Local Identification Under Two Periods of Conditionally Dependent, Monotone Attrition

In this section, we illustrate local identification of the unobserved heterogeneity when the researcher only has information on attrition, but is willing to make parametric assumptions on the attrition process. Define $\mathscr{L}_{j-1} \equiv \{L_{ik}\}_{k=0}^{j-1}$. Suppose, for example that $L_{i0} = 1$ and the attrition process is given recursively for $j > 0$ as

$$L_{ij} = 1\{\gamma_1 \theta_i h(\mathscr{L}_{j-1}) + \eta_j > 0\}, \tag{3.5}$$

where $h(\cdot)$ is a known, non-stochastic function.

Define $h_{j-1} \equiv h(\mathscr{L}_{j-1})$. Then the system of nonlinear equations that we must solve to identify p and γ_1 are:

$$P(L_{i1}=1)=\Psi(\gamma_1 h_0)p+\Psi(0)(1-p), \text{ and}$$
$$P(L_{i2}=1|L_{i1}=1)=\Psi(\gamma_1 h_1)p+\Psi(0)(1-p).$$

The following additional assumption on h allows for convenient discussion of identification of the model with conditionally dependent attrition.

Assumption 3.2.5. h is monotone in j.

An example of an h that satisfies Assumption 3.2.5 is one where $h_{j-1}=\sum_{k=0}^{j-1}L_{ik}$ for each j, implying that $h_0=1$ and $h_1=2$ in the preceding equations. Then, γ_1 and p are locally identified provided the Jacobian is nonsingular, as it will be under a condition that is identical to equation 3.4. On the other hand, one additional period of attrition is required to identify a specification such as:

$$L_{ij}=1\{\gamma_1\theta_i+\gamma_2 h(\mathscr{L}_{j-1})+\eta_i>0\}. \tag{3.6}$$

We now take a moment to highlight the source of identification of the model parameters using this type of approach. Continuing with the example from equation 3.5, we can write the probability that an agent attrits in the jth period given that the agent has never previously attrited as:

$$\nu_j\equiv 1-\left(p\Psi(\gamma_1 h_{j-1})+(1-p)\Psi(0)\right). \tag{3.7}$$

In our setup, the object ν_j fulfills a similar conceptual purpose to the hazard function that is familiar from the econometric literature on duration analysis that began in the 1970s (e.g., Cox [42], Lancaster [82]). For example, under a positive monotonic h, it is clear that the pseudo-hazard of 3.7, ν_j, is increasing in j when $\gamma_1<0$, and that ν_j is decreasing in j when $\gamma_1>0$. Thus, the sign of γ_1 determines whether the attrition process exhibits positive or negative duration dependence. The parameters of the model also governs how the likelihood of attrition changes with j. Here, a key difference from the usual duration literature is that we did not model the duration process - we instead modeled the underlying economic phenomenon of interest (attrition). The pseudo-hazard in 3.7 simply arises as a byproduct of our attempt to identify the model using properties of an attrition process we specify in a reduced-form such as equation 3.5.[15]

3.3 Estimation and Inference

In this section, we propose a general inference strategy for the models we have considered. In all cases, the parameter of interest, $\tau_0 \in \mathbf{R}^k$, can be expressed as the solution to a system of non-linear equations of the form

$$g(\tau_0) = \pi_0,$$

where $g(\cdot)$ is a known, non-stochastic function.

For the problem of identifying the TTTE, we are interested in the parameters $\tau_0 = (a_1, a_0, \gamma_0', p_0)$, where a_1 and a_0 are the TTTE, p_0 represents the distribution of discrete, unobserved heterogeneity, and γ_0 represents the parameters of the attrition process. As for π_0, example 3.2.3 from the section concerned with identification without covariatess[16] considers the moments $\pi_0 = (\pi_{01}, \pi_{02})'$, where

$$\begin{aligned}
\pi_{01} &= (\mathbf{E}[\tilde{Y}_i], \mathbf{E}[\tilde{Y}_i | L_{i1} = l_1], \mathbf{E}[\tilde{Y}_i | L_{i1} = l_1, L_{i2} = l_1]), \text{ and} \\
\pi_{02} &= (P(L_{i1} = l_1), P(L_{i2} = l_1 | L_{i1} = l_1)).
\end{aligned}$$

Letting $\hat{\pi}_n$ be the sample analogue estimator of π_0 and A_n be a random $k \times k$ weight matrix, we can define a minimum distance estimator, $\hat{\tau}_n$, as the minimzer of the following criterion function over $\tau_0 \in \mathscr{T}$, where $\mathscr{T} \subset \mathbf{R}^k$:

$$Q_n(\tau) = \|A_n(\hat{\pi}_n - g(\tau))\|^2 / 2.$$

Under the usual assumptions, the asymptotic covariance matrix of $\hat{\tau}_n$ has the familiar sandwich form

$$\Sigma_0 = B_0^{-1} \Omega_0 B_0^{-1},$$

[15]Although this chapter has focused on monotone attrition patterns, one can consider relaxing Assumption 3.2.5. For example, we may take an attrition process such as one where $L_{ij} = 1\{\gamma_1 \theta_i (h_{j-1} - a)^2 + \eta_j > 0\}$ with a unknown. Although the local identification argument in such a case is similar to before, the rank becomes harder to verify.

[16]In the case where covariates are used to identiy τ_0, π_0 follows a very similar structure, except we use the moments of X_i and the conditional moments of $\tilde{Y}_i$ given X_i.

where $B_0 = \Gamma_0' A' A \Gamma_0$, $\Omega_0 = \Gamma_0' A' A V_0 A' A \Gamma_0$, and $\Gamma_0 \equiv \frac{\partial}{\partial \tau'} g(\tau_0)$. In the following section, I consider inference on τ_0 when Σ_0 is estimated using $\hat{\Gamma}_n = \frac{\partial}{\partial \tau'} g(\hat{\tau}_n)$ and V_0 is estimated using a non-parametric bootstrap procedure.

3.3.1 Small Simulation Exercise

In this section, we assess the finite sample performance of the minimum distance inference approach outlined in the previous section. We consider the following model for the attrition equation:

$$L_i = 1\{\gamma_1 X_{i1} + \gamma_2 X_{i2} \theta_i + \eta_i > 0\},$$

where η_i is standard normal, X_{i1} and X_{i2} are both discrete uniform on $\mathscr{X} = \{1,2,3,4\}$. θ_i takes the value of 1.5 with probability p and 0.5 with probability $1-p$. X_{i1}, X_{i2}, θ_i are all independent of one another and drawn independently across i in each iteration of the simulations. Table 3.1 considers the size and power properties of the minimum distance estimation. We set the parameter values $\gamma_1 = -0.65$, $\gamma_2 = 0.8$, and $p = 0.5$. Although the procedure suffers from size distortion in the very small sample sizes, the performance is drastically improved as the sample size increases. Moreover, the procedure does not appear particularly sensitive to the choice of starting value.

3.4 Using Attrition to Correct for Selection on Unobservables

The following discussion draws from the arguments of Section 3.2 to illustrate that - although attrition is often considered a type of selection problem - it can also be seen as solution to selection problems under special circumstances.

We consider the the incidental truncation model of Gronau [64], where a variable W_i is only observed if another variable, L_i, takes on the value 1.[17] We wish to study the distribution of W_i. In the spirit of Gronau's 1974 example, suppose that W_i represents the worker's wage offer and L_i represents whether or not the worker

[17] See also Wooldridge [112].

Table 3.1: The Empirical Coverage Probability and Average Length of Confidence Intervals for Attrition Process Parameters at 95% Nominal Level

	Estimation 1			Estimation 2		
	γ_1	γ_2	p	γ_1	γ_2	p
$n = 1000$	0.9100	0.8655	0.9195	0.9210	0.8575	0.8810
$n = 2500$	0.9420	0.9175	0.9385	0.9270	0.9025	0.9155
$n = 5000$	0.9425	0.9405	0.9440	0.9575	0.9350	0.9370
$n = 10000$	0.9470	0.9395	0.9510	0.9505	0.9460	0.9460
$n = 15000$	0.9520	0.9385	0.9380	0.9520	0.9430	0.9480
	γ_1	γ_2	p	γ_1	γ_2	p
$n = 1000$	0.3302	0.5745	0.3822	0.3357	0.5687	0.3697
$n = 2500$	0.2051	0.3733	0.2230	0.2059	0.3708	0.2235
$n = 5000$	0.1442	0.2665	0.1540	0.1443	0.2659	0.1537
$n = 10000$	0.1014	0.1891	0.1078	0.1012	0.1890	0.1078
$n = 15000$	0.0828	0.1546	0.0875	0.0829	0.1545	0.0873

Notes: The first half of the table reports the empirical coverage probability of the confidence interval for the attrition process parameters, $\tau = (\gamma_1, \gamma_2, p)'$, and the second half reports its average length. We choose the value $\tau_0 = (-0.65, 0.8, 0.5)'$. The simulation number is $S = 2000$. We use $B = 1000$ in the bootstrap procedure. Estimation 1 uses a starting value of $(0, 0, 0.5)'$ for the minimum distance procedure while Estimation 2 uses a starting value $(-0.3, 0.4, 0.3)'$. As n increases, we find that the empirical rejection probabilities approach the nominal size and that confidence intervals shrink (as expected).

took the job. Here, ε_i may represent the worker's abilities, θ_i may capture whether or not the worker is discouraged, and η_i captures some additional income shocks. Suppose further, for concreteness, that the reduced form for these variables are given as follows:

$$\begin{aligned} W_i &= X_{i1}'\beta_1 + \beta_2\theta_i + \varepsilon_i, \text{ and} \\ L_i &= 1\{Z_i'\gamma_1 + \gamma_2\theta_i + \eta_i > 0\}, \end{aligned} \tag{3.8}$$

where $Z_i = (X_{i1}, X_{i2})'$, ε_i and η_i are possibly correlated, but independent of Z_i, and θ_i is a discrete, unobserved variable, possibly correlated with Z_i. If $\beta_2 = \gamma_2 = 0$, then the model is the familiar Type II Tobit model described in Amemiya [8].

When β_2 and γ_2 are not zero, the usual Heckman correction approach may not lead to consistent estimation of the model.

In the case that L_i depends on covariates as in 3.8, we can identify γ_1 and the distribution of θ_i using using an assumption such as 3.2.2 and the arguments of Section 3.2.2, provided Z_i has enough support points. Then it is a simple matter to show that:

$$\mathbf{E}[W_i|L_i = 1, Z_i] = X_{i1}'\beta_1 + \beta_2 + \rho\left(p\lambda(Z_i'\gamma_1) + (1-p)\lambda(Z_i'\gamma_1 + \gamma_2)\right),$$

where ρ is the correlation between η_i and ε_i and $\lambda(\cdot)$ is the inverse Mills ratio.

We can solve the problem of selection when the attrition process takes the form of 3.2.3 in a similar way when the econometrician observes L_{i2} in addition to L_{i1}. Note that none of the arguments in this section require panel data on W_i.

The following example demonstrates an economic application of the pure attrition identification strategy. Let W_i represent a student's midterm exam grade, which is observed if the student remained enrolled in class and wrote the midterm exam, i.e., $L_{i1} = 1$. Here, ε_i may represent the student's underlying intellectual abilities, θ_i represents their love of the subject matter, and η_i captures a general measure of their academic ambition. Let $L_{i2} = 1$ be the event that the student remained enrolled in the class and wrote the final exam. In this situation, if L_{i2} and L_{i1} are both influenced by θ_i and are not themselves conditionally independent given θ_i, we may use functional form restrictions to back out the distribution of θ_i. Then we follow arguments similar from the preceding discussion to make inference on the distribution of W_i.

3.5 Conclusion

This chapter makes a general case for the use of information on attrition to uncover unobserved heterogeneity and other parameters of interest. We show that attrition is special from an economic point of view in that modeling the attrition process can yield identification strategies that exploit duration dependence without any actual need to model the duration itself. The usefulness of information on sample attrition for identifying for identifying more complicated forms of heterogeneity - such as that varying over time - is left for future work.

Conclusions

In this thesis, I devise novel econometric approaches to extract information from two-sided matching data, network data, and data on sample attrition.

In the Chapter 1, I develop an empirical model of matching with endogenous pre-investments. Although the chapter focuses on investment in education by workers prior to entering a labour market, the model can be applied to many other situations of interest. For example, the model can be used to investigate the role that frictions and preferences play in a worker's choice of job sector. The model can be applied to the economics of education, to see how pre-investments on the part of students impact their college placements. The approach can also be used study marriage markets in which agents make observable pre-investments prior to looking for a partner.

One special feature of the model of Chapter 1 is that it can capture assortative matching between workers and firms when the match production function does not exhibit complementarities between their types. In this chapter, we investigate the role of these complementarities by performing counterfactuals under different specifications of the production technology. One avenue for future research would consider how the insights from such counterfactuals change in a model that uses richer information on the agents, such as firm profits.

One obvious extension to the framework of Chapter 1 would consider whether a tractable model of matching with endogeneous pre-investment is possible when we allow workers to choose between many different options.

Chapter 1 illustrates that preferences and frictions in two-sided matching markets can be studied when the researcher only observes a single cross-section of matched employer-employee data. One very challenging extension to the static setup considered here would use information on unemployment and job-to-job transitions to learn more about frictions in markets along with the preferences of the agents who participate in them.

One key insight of Chapter 1 is that capital and wages are affected in subtle ways by the presence of market frictions when the decisions of workers are endogenous. A useful extension that may yield further insights on wage inequality would consider the role of endogenous firm capital. Such an extended framework may also provide insights on economic growth.

In Chapter 2 we develop a model of social interactions with many economically attractive features that is also highly practical from the point of view of empirical modeling. In our approach, we suppose that agents optimize by projecting their own beliefs onto those of their neighbors. This condition is well-motivated from a literature in behavioural economics on inter-personal projection. The methods we develop illustrate the usefulness of behavioural approaches to economic modeling. An open question is whether a tractable model of interactions is possible under other assumptions about belief formation. A related avenue for future research is whether researchers can develop empirical approaches to test alternative models of belief formation.

In Chapter 3, we find that when an agent's attrition decision depends on their unobserved heterogeneity, we may consider using attrition patterns to learn the distribution of heterogeneity, even in the absence of covariates and panel data on the main outcome of interest. These approaches can be usefully applied to correct for selection on unobservables when the researcher believes that the attrition process satisfies a certain parametric structure. A useful extension of the approaches discussed in this chapter would allow for heterogeneity that may be time varying or continuous, and consider identification under minimal assumptions on the relationship between attrition and the unobserved heterogeneity.

Bibliography

[1] A. Abdulkadiroğlu and T. Sönmez. Random serial dictatorship and the core from random endowments in house allocation problems. *Econometrica*, 66 (3):689–701, 1998. → page 11

[2] J. M. Abowd, F. Kramarz, and D. N. Margolis. High wage workers and high wage firms. *Econometrica*, 67(2):251–333, 1999. → page 7

[3] D. Acemoglu. Politics and economics in weak and strong states. *Journal of Monetary Economics*, 52(7):1199–1226, 2005. → page 82

[4] D. Acemoglu, C. García-Jimeno, and J. A. Robinson. State capacity and economic development: A network approach. *American Economic Review*, 105:2364–2409, 2015. → pages 42, 82, 84, 85, 86, 87

[5] N. Agarwal and W. Diamond. Identification and estimation in two-sided matching markets. 2014. → page 22

[6] V. Aguirregabiria and P. Mira. Identification of games of incomplete information with multiple equilibria and unobserved heterogeneity. 2016. → pages 9, 123

[7] M. Ahsanullah, V. B. Nevzorov, and M. Shakil. *An introduction to order statistics*. Springer, 2013. → page 131

[8] T. Amemiya. *Advanced econometrics*. Harvard university press, 1985. → page 107

[9] L. Anselin. *Spatial Econometrics: Methods and Models*. Kluwer Academic Publishers, The Netherlands, 1988. → page 43

[10] A. Aradillas-Lopez and E. Tamer. The identification power of equilibrium in simple games. *Journal of Business and Economic Statistics*, 26: 261–283, 2008. → page 44

[11] J. Bagger and R. Lentz. An empirical model of wage dispersion with sorting. Technical report, National Bureau of Economic Research, 2014. → pages 7, 16

[12] P. Bajari, H. Hong, and D. Nekipelov. Game theory and econometrics: A survey of some recent research. In *Advances in Economics and Econometrics, 10th World Congress*, volume 3, pages 3–52, 2013. → page 30

[13] E. Barth, A. Bryson, J. C. Davis, and R. Freeman. It?s where you work: Increases in the dispersion of earnings across establishments and individuals in the united states. *Journal of Labor Economics*, 34(S2): S67–S97, 2016. → page 3

[14] C. Bartolucci and F. Devicienti. Better workers move to better firms: a simple test to identify sorting. 2012. → page 37

[15] G. S. Becker. A theory of marriage: Part i. *The Journal of Political Economy*, pages 813–846, 1973. → pages 5, 8

[16] C. Bellemare. A life-cycle model of outmigration and economic assimilation of immigrants in germany. *European Economic Review*, 51(3): 553–576, 2007. → pages 91, 97

[17] S. T. Berry. Estimation of a model of entry in the airline industry. *Econometrica: Journal of the Econometric Society*, pages 889–917, 1992. → page 9

[18] T. Besley and S. Coate. Workfare versus welfare: Incentive arguments for work requirements in poverty-alleviation programs. *The American Economic Review*, 82(1):249–261, 1992. → page 95

[19] T. Besley and T. Persson. The origins of state capacity: property rights, taxation and politics. *American Economic Review*, 99(4):1218–1244, 2009. → page 82

[20] L. E. Blume, W. A. Brock, S. N. Durlauf, and Y. M. Ioannides. Identification of social interactions. In A. B. Benhabib, J. and M. Jackson, editors, *Handbook of Social Economics*, volume 1B, pages 853–964. Elsevier, 2010. → page 39

[21] L. E. Blume, W. A. Brock, S. N. Durlauf, and R. Jayaraman. Linear social interactions models. *Journal of Political Economy*, 123:444–496, 2015. → pages 43, 44, 45, 46

[22] R. Blundell and J. M. Robin. Estimation in large and disaggregated demand systems: An estimator for conditionally linear systems. *Journal of Applied Econometrics*, 14:209–232, 1999. → page 66

[23] S. Bonhomme, T. Lamadon, and E. Manresa. A distributional framework for matched employer employee data. *Unpublished manuscript, University of Chicago*, 2015. → page 8

[24] B. Boudarbat, T. Lemieux, and W. C. Riddell. The evolution of the returns to human capital in canada, 1980-2005. *Canadian Public Policy*, 36(1): 63–89, 2010. → page 33

[25] Y. Bramoullé, H. Djebbari, and B. Fortin. Identification of peer effects through social networks. *Journal of Econometrics*, 150:41–55, 2009. → pages 43, 44, 46, 70

[26] T. F. Bresnahan and P. C. Reiss. Entry and competition in concentrated markets. *Journal of Political Economy*, 99(5):977–1009, 1991. → page 9

[27] E. Breza, A. G. Chandrasekhar, and A. Tahbaz-Salehi. Seeing the forest for the trees? an investigation of network knowledge. *Working Paper*, 2018. → page 40

[28] W. A. Brock and S. N. Durlauf. Discrete choice with social interaction. *Review of Economic Studies*, 68:235–260, 2001. → page 71

[29] A. Calvó-Armengol, E. Pattacchini, and Y. Zenou. Peer effects and social networks in education. *Review of Economic Studies*, 76:1239–1267, 2009. → page 43

[30] C. Camerer. *Behavioral Game Theory*. Princeton University Press, New York, 2003. → page 54

[31] A. C. Cameron, J. B. Gelbach, and D. L. Miller. Robust inference with multiway clustering. *Journal of Business and Economic Statistics*, 29: 238–249, 2011. → page 75

[32] N. Canen, J. Schwartz, and K. Song. Estimating local interactions among many agents who observe their neighbors. *arXiv preprint arXiv:1704.02999*, 2017. → page 68

[33] D. Card and T. Lemieux. Can falling supply explain the rising return to college for younger men? a cohort-based analysis. *The Quarterly Journal of Economics*, 116(2):705–746, 2001. → pages 7, 19, 21

[34] D. Card, J. Heining, and P. Kline. Workplace heterogeneity and the rise of west german wage inequality. 2012. → page 33

[35] D. Card, J. Heining, and P. Kline. Workplace heterogeneity and the rise of west german wage inequality. *The Quarterly journal of economics*, 128(3): 967–1015, 2013. → page 3

[36] A. R. Cardoso, A. Loviglio, and L. Piemontese. Information frictions and labor market outcomes. 2015. → page 8

[37] H. Chade, J. Eeckhout, and L. Smith. Sorting through search and matching models in economics. *Journal of Economic Literature*, 55(2):493–544, 2017. → pages 5, 7, 8

[38] P.-A. Chiappori and B. Salanié. The econometrics of matching models. *Journal of Economic Literature*, 54(3):832–861, 2016. → page 9

[39] E. Choo and A. Siow. Who marries whom and why. *Journal of political Economy*, 114(1):175–201, 2006. → page 9

[40] F. Ciliberto and E. Tamer. Market structure and multiple equilibria in airline markets. *Econometrica*, 77(6):1791–1828, 2009. → page 9

[41] T. G. Conley. GMM estimation with cross sectional dependence. *Journal of Econometrics*, 92:1–45, 1999. → page 42

[42] D. R. Cox. Regression models and life-tables. 34(2):187–220, 1972. → page 104

[43] F. Cunha, J. J. Heckman, and S. M. Schennach. Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78(3): 883–931, 2010. → page 92

[44] F. Cunha, J. J. Heckman, and S. M. Schennach. Supplement to ”estimating the technology of cognitive and noncognitive skill formation”. *Econometrica*, 78(3):883–931, 2010. → page 92

[45] G. De Giorgi, M. Pellizzari, and S. Redaelli. Identification of social interactions through partially overlapping peer groups. *American Economic Journal: Applied Economics*, 2:241–275, 2010. → pages 43, 70

[46] G. Dionne and B. Dostie. New evidence on the determinants of absenteeism using linked employer-employee data. *Industrial & Labor Relations Review*, 61(1):108–120, 2007. → page 34

[47] B. Dostie and R. Jayaraman. Organizational redesign, information technologies and workplace productivity. 2008. → page 34

[48] J.-M. Dufour and L. Khalaf. Monte carlo test methods in econometrics. *Companion to Theoretical Econometrics, Blackwell Companions to Contemporary Economics, Basil Blackwell, Oxford, UK*, pages 494–519, 2001. → page 10

[49] S. N. Durlauf and Y. M. Ioannides. Social interactions. *Annual Review of Economics*, 2, 2010. → page 39

[50] S. N. Durlauf and H. Tanaka. Understanding regression versus variance tests for social interactions. *Economic Inquiry*, 46, 2008. → page 71

[51] F. Echenique, S. Lee, and M. Shum. Partial identification in two-sided matching models. In *Structural Econometric Models*, pages 117–139. Emerald Group Publishing Limited, 2013. → page 9

[52] J. Eeckhout and P. Kircher. Identifying sorting - in theory. *The Review of Economic Studies*, page rdq034, 2011. → page 7

[53] H. Eraslan and X. Tang. Identification and estimation of large network games with private link information. *Working Paper*, 2017. → page 44

[54] N. Fortin, D. A. Green, T. Lemieux, K. Milligan, and W. C. Riddell. Canadian inequality: Recent developments and policy options. *Canadian Public Policy*, 38(2):121–145, 2012. → page 33

[55] J. T. Fox and P. Bajari. Measuring the efficiency of an fcc spectrum auction. *American Economic Journal: Microeconomics*, 5(1):100–146, 2013. → page 9

[56] D. Gale and L. S. Shapley. College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1):9–15, 1962. → pages 5, 8

[57] P. A. Gautier and C. N. Teulings. How large are search frictions? *Journal of the European Economic Association*, 4(6):1193–1225, 2006. → page 7

[58] N. Gennaioli and I. Rainer. The modern impact of precolonial centralization in africa. *Journal of Economic Growth*, 12(3):185–234, 2007. → page 82

[59] N. Gennaioli and H.-J. Voth. State capacity and military conflict. *Review of Economic Studies*, 82(4):1409–1448, 2015. → page 82

[60] A. Goldfarb and M. Xiao. Who thinks about the competition? managerial ability and strategic entry in us local telephone markets. *American Economic Review*, 101:3130–3161, 2011. → page 44

[61] P. Goldsmith-Pinkham and G. W. Imbens. Social networks and the identification of peer effects. *Journal of Business and Economic Statistics*, 31(3):253–264, 2013. → page 43

[62] B. S. Graham. An econometric model of link formation with degree heterogeneity. *Econometrica*, 85:1033–1063, 2017. → page 48

[63] D. A. Green and B. M. Sand. Has the canadian labour market polarized? *Canadian Journal of Economics/Revue canadienne d'économique*, 48(2): 612–646, 2015. → page 33

[64] R. Gronau. Wage comparisons–a selectivity bias. *Journal of political Economy*, 82(6):1119–1143, 1974. → page 106

[65] M. Hagedorn, T. H. Law, and I. Manovskii. Identifying equilibrium models of labor market sorting. *Econometrica*, 85(1):29–65, 2017. → page 7

[66] J. A. Hausman and D. A. Wise. Attrition bias in experimental and panel data: the gary income maintenance experiment. *Econometrica: Journal of the Econometric Society*, pages 455–473, 1979. → page 91

[67] J. A. Hausman, W. K. Newey, H. Ichimura, and J. L. Powell. Identification and estimation of polynomial errors-in-variables models. *Journal of Econometrics*, 50(3):273–295, 1991. → page 93

[68] J. A. Hausman, J. Abrevaya, and F. M. Scott-Morton. Misclassification of the dependent variable in a discrete-response setting. *Journal of econometrics*, 87(2):239–269, 1998. → page 92

[69] J. J. Heckman. The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. In *Annals of Economic and Social Measurement, Volume 5, number 4*, pages 475–492. NBER, 1976. → page 91

[70] J. J. Heckman and B. Singer. Econometric duration analysis. *Journal of Econometrics*, 24(1-2):63–132, 1984. → page 93

[71] Y. Hu. Identification and estimation of nonlinear models with misclassification error using instrumental variables: A general solution. *Journal of Econometrics*, 144(1):27–61, 2008. → page 93

[72] Y. Hu and S. M. Schennach. Instrumental variable treatment of nonclassical measurement error models. *Econometrica*, 76(1):195–216, 2008. → page 92

[73] S. I. M. Hwang. A robust redesign of high school match. *Working Paper*, 2016. → page 44

[74] I. Johnsson and R. H. Moon. Estimation of peer effects in endogenous social networks: control function approach. *Working Paper*, 2016. → page 43

[75] K. Kantenga and T. H. Law. Sorting and wage inequality. In *2016 Meeting Papers*, number 660. Society for Economic Dynamics, 2016. → page 3

[76] H. Kasahara and K. Shimotsu. Pseudo-likelihood estimation and bootstrap inference for structural discrete markov decision models. *Journal of Econometrics*, 146(1):92–106, 2008. → page 15

[77] H. Kasahara and K. Shimotsu. Sequential estimation of structural models with a fixed point constraint. *Econometrica*, 80(5):2303–2319, 2012. → page 30

[78] H. H. Kelejian and D. Robinson. A suggested method of estimation for spatial interdependent models with autocorrelated errors, and an application to a county expenditure model. *Papers in Regional Science*, 72: 297–312, 1993. → page 70

[79] R. Kellogg. Uniqueness in the schauder fixed point theorem. *Proceedings of the American Mathematical Society*, 60(1):207–210, 1976. → page 124

[80] E. D. Kolaczyk. *Statistcal Analysis of Network Data*. Springer Verlag, New York, 2009. → page 62

[81] A. Konovalov and Z. Sándor. On price equilibrium with multi-product firms. *Economic Theory*, 44(2):271–292, 2010. → page 124

[82] T. Lancaster. Econometric methods for the duration of unemployment. *Econometrica: Journal of the Econometric Society*, pages 939–956, 1979. → page 104

[83] T. Lemieux. The changing nature of wage inequality. *Journal of Population Economics*, 21(1):21–48, 2008. → page 33

[84] A. Lewbel. Identification of the binary choice model with misclassification. *Econometric Theory*, 16(4):603–609, 2000. → pages 91, 92

[85] J. Lise, C. Meghir, and J.-M. Robin. Matching, sorting and wages. *Review of Economic Dynamics*, 19:63–87, 2016. → page 7

[86] G. Loewenstein, T. O'Donohue, and M. Rabin. Projection bias in predicting future utility. *Quarterly Journal of Economics*, 118:1209–1248, 2003. → page 41

[87] R. Lopes de Melo. Firm wage differentials and labor market sorting: Reconciling theory and evidence. *Unpublished Manuscript*, 2013. → page 7

[88] K. Madarász. Information projection: Model and applications. *Review of Economic Studies*, 79, 2012. → page 41

[89] N. Maestas, K. J. Mullen, and A. Strand. Does disability insurance receipt discourage work? using examiner assignment to estimate causal effects of ssdi receipt. *American Economic Review*, 103(5):1797–1829, 2013. → page 95

[90] C. F. Manski. Identification of endogenous social effects: The reflection problem. *Review of Economic Studies*, 60(3):531–542, 1993. → page 43

[91] I. Marinescu and R. Wolthoff. Opening the black box of the matching function: The power of words. 2016. → page 8

[92] M. McBride, G. Milante, and S. Skaperdas. Peace and war with endogenous state capacity. *Journal of Conflict Resolution*, 55(3):446–468, 2011. → page 82

[93] K. Menzel. Large matching markets as two-sided demand systems. *Econometrica*, 83(3):897–941, 2015. → page 9

[94] K. Menzel. Inference for games with many players. *Review of Economic Studies*, 83:306–337, 2016. → pages 9, 43

[95] S. Michalopoulos and E. Papaioannou. Pre-colonial ethnic institutions and contemporary african development. *Econometrica*, 81(1):113–152, 2013. → page 82

[96] R. Moffit, J. Fitzgerald, and P. Gottschalk. Sample attrition in panel data: The role of selection on observables. *Annales d'Economie et de Statistique*, pages 129–152, 1999. → page 91

[97] W. K. Newey and D. McFadden. Large sample estimation and hypothesis testing. *Handbook of econometrics*, 4:2111–2245, 1994. → page 98

[98] K. Pendakur and S. Woodcock. Glass ceilings or glass doors? wage disparity within and between firms. *Journal of Business & Economic Statistics*, 28(1):181–189, 2010. → page 34

[99] F. Postel-Vinay and J.-M. Robin. Equilibrium wage dispersion with worker and employer heterogeneity. *Econometrica*, 70(6):2295–2350, 2002. → pages 7, 16

[100] R. Powell. Monopolizing violence and consolidating power. *Quarterly Journal of Economics*, 128(2):807–859, 2013. → page 82

[101] J. M. Robins, A. Rotnitzky, and L. P. Zhao. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the american statistical association*, 90(429):106–121, 1995. → page 92

[102] A. E. Roth and M. A. O. Sotomayor. *Two-sided matching: A study in game-theoretic modeling and analysis*. Number 18. Cambridge University Press, 1992. → page 11

[103] T. J. Rothenberg. Identification in parametric models. *Econometrica: Journal of the Econometric Society*, pages 577–591, 1971. → page 98

[104] E. Saez and M. R. Veall. The evolution of high incomes in northern america: Lessons from canadian evidence. *American Economic Review*, 95 (3):831–849, 2005. doi:10.1257/0002828054201404. URL http://www.aeaweb.org/articles.php?doi=10.1257/0002828054201404. → page 33

[105] J. D. Sargan. Identification and lack of identification. *Econometrica: Journal of the Econometric Society*, pages 1605–1633, 1983. → page 98

[106] S. M. Schennach. Estimation of nonlinear models with measurement error. *Econometrica*, 72(1):33–75, 2004. → page 93

[107] K. Song. Econometric inference on large bayesian games with heterogeneous beliefs. *arXiv:1404.2015 [stat.AP]*, 2014. → pages 9, 43

[108] C. M. Tiebout. A pure theory of local expenditures. *Journal of political economy*, 64(5):416–424, 1956. → page 84

[109] L. Van Boven, G. Loewenstein, and D. Dunning. Mispredicting the endowment effect: Underestimation of owners' selling prices by buyer's agents. *Journal of Economic Behavior and Organization*, 51:351–365, 2003. → page 41

[110] M. R. Veall. Top income shares in canada: recent trends and policy implications. *Canadian Journal of Economics/Revue canadienne d'économique*, 45(4):1247–1272, 2012. → page 34

[111] B. Williams. Identification of a nonseparable model under endogeneity using binary proxies for unobserved heterogeneity. 2017. → page 93

[112] J. M. Wooldridge. *Econometric analysis of cross section and panel data*. MIT press, 2010. → page 106

[113] H. Xu. Estimation of discrete games with correlated types. *The Econometrics Journal*, 17(3):241–270, 2014. → page 15

[114] H. Xu. Social interactions in large networks: A game theoretic approach. *Working Paper*, 2015. → pages 9, 43

[115] X. Xu and L. Lee. Estimation of a binary choice game model with network links. *Working Paper*, 2015. → page 43

[116] C. Yang and L.-F. Lee. Social interactions under incomplete information with heterogeneous expectations. *Journal of Econometrics, Forthcoming*, 2016. → page 43

Appendix A

Supporting Materials

A.1 Appendix to Chapter 1

A.1.1 Equilibrium of Investment and Matching Game

In this section, we characterize the equilibrium of the incomplete information game of Section 1.2.2. First, we introduce a representation of the worker's expected utility function that proves useful for establishing the existence of the Bayesian Nash equilibrium of the game as a fixed point of a best probability response operator. We begin by defining relevant terms. A profile of *strategy functions* (or decision rules) is

$$\sigma = \{\sigma_i(x_i, \varepsilon_i) : i \in N_h\},$$

where the functions $\sigma_i : \mathscr{X} \times \mathbf{R}^{J-1} \to \mathscr{H}$. The conditional probability that a worker with covariates x_i chooses action h_i can be written

$$P_i(h_i|x_i, \sigma_i) \equiv \int 1\{\sigma_i(x_i, \varepsilon_i) = h_i\} dF(\varepsilon_i).$$

Since X_i's are private information in this model, each agent i must take expectations with respect to the distribution of X_{-i}. The following result shows that under the independence assumptions embodied by Assumption 1.2.1, the agent's expected utility has a very convenient form - it is only affected by the behaviour of the other agents through the choice probabilities.

Lemma A.1.1. *In the model of Section (1.2.2) and Assumptions 1.2.1 and 1.2.2, we can represent the first term in the expected utility of agent i from equation 1.5 as*

$$\tilde{U}_i(h_i, x_i, \sigma) = \sum_{h_{-i} \in \mathscr{H}_{-i}} \tilde{u}_i(h_i, h_{-i}, x_i) \prod_{j \neq i} P_j(h_j | \sigma_j).$$

Proof. First, we write equation 1.5 as

$$\tilde{U}_i(h_i, x_i, \sigma) \quad = \quad \sum_{x_{-i} \in \mathscr{X}_{-i}} \sum_{h_{-i} \in \mathscr{H}_{-i}} \tilde{u}_i(h_i, h_{-i}, x_i) P_{-i}(h_{-i} | x_{-i}, \sigma) P(x_{-i}), \quad \text{(A.1)}$$

where $x_{-i} = (x_j)_{j \in N_h \setminus \{i\}}$ and we use the shorthand $P(x_{-i}) \equiv P(X_{-i} = x_{-i})$. Without loss of generality, let $i = 1$. Then we write $\tilde{U}_1(h_1, x_1, \sigma)$ as

$$\sum_{h_{-1} \in \mathscr{H}_{-1}} \sum_{x_2 \in \mathscr{X}} \cdots \sum_{x_n \in \mathscr{X}} \tilde{u}_1(h_1, h_{-1}, x_1) P_{-1}(h_{-1} | x_2, ..., x_n, \sigma) \prod_{j=2}^{n_h} P_j(x_j), \qquad \text{(A.2)}$$

where we used the independence of X_i's from Assumption 1.2.1. Next, since Assumption 1.2.1 says that X_i's and ε_i's are independent, we know that the actions of each of the agents are independent and depend only on their personal value of X_i and ε_i. Therefore,

$$P(h_{-1} | x_2, ..., x_n, \sigma) \quad = \quad \prod_{j=2}^{n_h} P_j(h_j | x_2, ... x_n, \sigma_j) = \prod_{j=2}^{n_h} P_j(h_j | x_j, \sigma_j). \quad \text{(A.3)}$$

Plugging (A.3) back into (A.2) yields that $\tilde{U}_1(h_1, x_1, \sigma)$ is equal to

$$\sum_{h_{-i} \in \mathscr{H}_{-i}} \tilde{u}_i(h_i, h_{-i}, x_i) \sum_{x_2 \in \mathscr{X}} \cdots \sum_{x_n \in \mathscr{X}} \prod_{j=2}^{n_h} P_j(h_j | x_j, \sigma_j) \prod_{j=2}^{n_h} P_j(x_j). \qquad \text{(A.4)}$$

Grouping the sums in (A.4) and restoring the generic i index gives

$$\sum_{h_{-i} \in \mathscr{H}_{-i}} \tilde{u}_i(h_i, h_{-i}, x_i) \prod_{j \neq i} \sum_{x_j \in \mathscr{X}} P_j(h_j | X_j = x_j, \sigma_j) P_j(x_j)$$

and hence we have the desired result. □

We will show the existence of the equilibrium for our model. The solution concept for the game described in Section 1.2.2 is Bayesian Nash Equilibrium (BNE), which we now define.

Definition A.1.1. *A Bayesian Nash Equilibrium (BNE) of the game described in Section (1.2.2) is a profile of decision rules* σ^* *such that for any player i and for any* (x_i, ε_i)*:*

$$\sigma_i^*(x_i, \varepsilon_i) \quad = \quad argmax_{\ h_i \in \mathscr{H}} \left\{ U_i(h_i, x_i, \varepsilon_i, \sigma^*) \right\}. \tag{A.5}$$

The notation and arguments in this section follow Aguirregabiria and Mira [6], but we include them here for completeness. Under Assumption 1.2.2, we write the expected utility of i as

$$U_i(h_i, x_i, \sigma, \varepsilon_i) \quad = \quad \tilde{U}_i(h_i, x_i, \sigma) + \varepsilon_i' d(h_i).$$

By Lemma A.1.1 we can express the first term on the right hand side of the the preceding equation as

$$\tilde{U}_i(h_i, x_i, \sigma) \quad = \quad \sum_{h_{-i} \in \mathscr{H}_{-i}} \tilde{u}_i(h_i, h_{-i}, x_i) \prod_{j \in N_h \setminus \{i\}} P_j(h_j | \sigma_j).$$

Note that $\tilde{U}_i(h_i, x_i, \sigma)$ only depends on the choices of other agents through the choice probabilities of the other players that are induced by σ. We write the choice probabilities of the people other than i as

$$\mathbf{P}_{-i} \equiv \{P_j(h_j) : (j, h_j) \in N \setminus \{i\} \times \mathscr{H} \setminus \{1\}\}.$$

For any $\mathbf{P}_{-i}$, we can define a *best response probability function* as:

$$\tilde{\Psi}_i(h_i | x_i, \mathbf{P}_{-i}) \quad \equiv \quad \int 1\{\text{argmax}_{\ h_i \in \mathscr{H}} \tilde{U}_i(h_i, x_i, \sigma) + \varepsilon_i' d(h_i) = h_i\} dF(\varepsilon_i).$$

$\tilde{\Psi}_i$ tells us the probability that a particular action is optimal for i with covariates x_i

when others choose according to probabilities $\mathbf{P}_{-i}$.[1] Let

$$\Psi_i(h_i|\mathbf{P}_{-i}) \quad = \quad \sum_{x_i \in \mathscr{X}} \tilde{\Psi}_i(h_i|x_i, \mathbf{P}_{-i}) P(x_i).$$

An equivalent to Definition 1 on the preceding page is that the equilibrium probabilities, $\mathbf{P}^* \equiv \mathbf{P}(\sigma^*)$, satisfy the fixed point constraint, $\mathbf{P}^* = \Psi(\mathbf{P}^*)$, where Ψ is the best response probability mapping:

$$\Psi(\mathbf{P}) = \{\Psi_i(h_i|\mathbf{P}_{-i}) : (i, h_i) \in N \times \mathscr{H} \backslash \{1\}\}. \tag{A.6}$$

Lemma A.1.2. *Under Assumption 1.2.1 and Assumption 1.2.2 the game described in Section (1.2.2) has a Bayesian Nash Equilibrium.*

Proof. Since $\Psi(\cdot)$ maps from a compact convex set, $[0,1]^{n \times (J-1)}$, to itself and is continuously differentiable (by the continuity of ε_i's (Assumption 1.2.1)), $\Psi(\cdot)$ has a fixed point by Brouwer's fixed point theorem. $\square$

We can appeal to a result of Kellogg [79] to show that the Bayesian equilibrium is unique under a mild condition on the derivatives of the best response probability mapping. Let I_n be an identity matrix of size n. Kellogg's result - as stated in Konovalov and Sándor [81] - stipulates that the equilibrium is unique if the determinant of $\tilde{J}_n \equiv \partial \Psi(\mathbf{P}) / \partial \mathbf{P}' - I_n$ is nonzero and the mapping Ψ has no fixed points on the boundary of $[0,1]^{n \times (J-1)}$. In our setup, the former condition holds provided that the best response of any agent is not excessively sensitive to a change in the probability of any other agent.[2]

[1] Note that when ε_i's have the extreme value distribution (as in Assumption 1.3.2) then we have

$$\tilde{\Psi}_i(h_i|x_i, \mathbf{P}_{-i}) = \frac{\exp(\tilde{U}_i(h_i, x_i))}{\sum_{j=1}^{J} \exp(\tilde{U}_i(h_j, x_i))}.$$

[2] Under our assumptions, $\tilde{J}_n$ is a matrix with -1's on the diagonal and $\tilde{p} \equiv \partial \Psi_i / \partial p_j$ for all $i \neq j$ on the off diagonals. Using the fact that $\tilde{J}_n$ is a circulant matrix, we can express its determinant explicitly as $(\tilde{p}(n-1) - 1)\,(-(1+\tilde{p}))^{n-1}$. A sufficient condition for $\det(\tilde{J}_n) \neq 0$ is thus $\tilde{p} < 1/(n-1)$.

A.1.2 Additional Mathematical Results

The remaining results of this section allow us to represent the matching probabilities from equation 1.15, hence workers' expectations, in a convenient way. These representations can then be used to estimate $\theta(\beta)$ using maximum likelihood.

Proposition A.1.1. *Suppose that $J = 2$, that Assumptions 1.2.1, 1.2.2, 1.3.1, and 1.3.2 hold. Then for each $m = 1,\dots,M$ and each $h_j \in \mathscr{H}$ the conditional probability of an arbitrary worker i matching to firm with capital class m after choosing education level h_j is*

$$\pi_{mj} = \sum_{n_j=0}^{n_h-1} P(\mathscr{M}(i) = m | H_i = h_j, N_j = n_j) B(n_j; n_h - 1, p_j),$$

where N_j is the number of workers other than i who picked education level h_j, $B(n_j; n_h - 1, p_j)$ is the binomial p.m.f. and $p_j = P(H_i = h_j)$.

Proof. Part (a) of Assumption 1.3.1 that says firms do not consider the workers' covariates when ranking them in the matching process. This means that for each $m = 1,\dots,M$ we have that

$$P(\mathscr{M}(i) = m | H_i = h_j, H_{-i} = h_{-i}, X_i = x_i) = P(\mathscr{M}(i) = m | H_i = h_j, H_{-i} = h_{-i}).$$

Combining this with equation 1.15, we can write

$$\pi_{mj}^{(i)} = \sum_{h_{-i} \in \mathscr{H}_{-i}} P(\mathscr{M}(i) = m | H_i = h_j, H_{-i} = h_{-i}) P(H_{-i} = h_{-i} | X_i = x_i).$$

Next, it is straightfoward to see that[3]

$$P(H_{-i} = h_{-i}|X_i = x_i) = \prod_{j\neq i}^{n_h} P_j(h_j). \tag{A.8}$$

Since ε_i are identically distributed by Assumption 1.2.1, for each j and m, have $\pi_{mj}^{(i)} = \pi_{mj}$.

When there are only two education levels, any $h_{-i} \in \mathscr{H}_{-i}$ can be represented as a total number of workers other than i who picked education level h_j, n_j. From worker i's point of view, n_j is a particular realization of the random variable N_j that takes values in the set $\{0,...,n_h - 1\}$. Since there are $n_h - 1$ agents other than i in the economy, the sum over $h_{-i} \in \mathscr{H}_{-i}$ amounts to a sum over the support of N_j. Now consider any n_j in the support of N_j. The assumption that ε_i's are iid implies that the probability that exactly n_j out of $n_h - 1$ workers pick h_j can be represented as

$$\frac{(n_h-1)!}{n_j!(n_h-1-n_j)!} p_j^{n_j} (1-p_j)^{n_h-1-n_j},$$

which is the binomial probability mass function, $B(n_j; n_h - 1, p_j)$. □

[3] This can be shown using the same arguments as those in Lemma A.1.1. The private information and independence of X_i's (Assumption 1.2.1) implies that the left hand size of A.8 equals

$$\sum_{x_{-i}\in\mathscr{X}_{-i}} P(h_{-i}|x_{-i},x_i)P(x_{-i}|x_i) \quad = \quad \sum_{x_{-i}\in\mathscr{X}_{-i}} P(h_{-i}|x_{-i})P(x_{-i}). \tag{A.7}$$

Suppose without loss of generality that $i = 1$. It is convenient to rewrite the above as follows (using independence):

$$\sum_{x_2\in\mathscr{X}} \cdots \sum_{x_n\in\mathscr{X}} P(h_{-1}|x_2,...,x_n) \prod_{j=2}^{n_h} P_j(x_j).$$

Next, since X_i's and ε_i's are independent across i and each i's strategy function is only a function of X_i and ε_i we have

$$P(h_{-1}|x_{-1}) = \prod_{j\neq 1}^{n_h} P_j(h_j|x_{-1}) \quad = \quad \prod_{j\neq 1}^{n_h} P_j(h_j|x_j).$$

Combining these two results we write A.7 as

$$\sum_{x_2\in\mathscr{X}} P_j(h_2|x_2)P_j(x_2)\ldots \sum_{x_n\in\mathscr{X}} P_n(h_n|x_n)P_n(x_n) = \prod_{j\neq 1}^{n_h} P_j(h_j).$$

Characterization of Matching Probabilities when J equals 2

When $J = 2$ we can partition the types of firms, $m = 1, ..., M$ into two sets: those who prefer $h_j \in \mathscr{H}$ and those who prefer $h_{j'}$ with $j' \neq j$. It is convenient to introduce the following notation:

$M_j^+(\theta) = \{m \in \{1, ..., M\} : \rho(k_m, h_j; \theta) \geq \rho(k_m, h_{j'}; \theta), j \neq j'\}$, and
$M_j^-(\theta) = \{1, ..., M\} \backslash M_j^+$,

recalling that firm preferences are given in 1.9. The firm classes that prefer h_j are pinned down by the functional form for firm preferences, ρ, along with the preference parameter, θ, and the distribution of X_i. Furthermore let us denote

$$\pi_{mj} \quad \equiv \quad \sum_{n^{(j)}=0}^{n_f} \sum_{n_j=0}^{n_h-1} P_{h_j, n_j, n^{(j)}}(m) P(n_j) P(n^{(j)}; \theta), \tag{A.9}$$

where

$$P_{h_j, n_j, n^{(j)}}(m) = P(\mathscr{M}(i) = m | h_i = h_j, N_j = n_j, N^{(j)} = n^{(j)}). \tag{A.10}$$

Note that this object depends on both β and θ through the matching function. For each firm type $m = 1, ...M$ let $F_m \equiv N(\beta k_m, \sigma^2)$, and define the following for each education choice h_j:

$$G_{j+} \equiv \sum_{m \in M_j^+} q_m F_m \text{ and } G_{j-} \equiv \sum_{m \in M_j^+} q_m F_m.$$

Furthermore, define the posterior firm types as follows:

$$q_m^+ \equiv q_m / \sum_{m \in M_j^+} q_m \text{ and } q_m^- \equiv q_m / \sum_{m \in M_j^+} q_m.$$

We also define $v_{(b_1, b_2; F)}$ as the b_1-order statistic of b_2 random variables independently distributed according to cdf F. Propositions A.1.2 and A.1.3 are characterizations of $P_{h_j, n_j, n^{(j)}}(m)$'s of the model in the case that $J = 2$ and $n_h = n_f = n$.

When considering these results, it is important to recall one core feature of the matching model as we outline it in Section 1.2: that there is no unemployment. Therefore, when reading the arguments, the reader should take for granted the fact that the probability that each worker matches to some firm occurs with probability one.

Proposition A.1.2. *(Heterogeneous firm preferences). Denote* $\bar{n}_j \equiv n_j + 1$ *and suppose that* $n_h = n_f = n$*. Then under the assumptions of Proposition A.1.1 we have the following for any* n_j *such that* $1 \leq n_j \leq n$ *and* $n^{(j)}$ *such that* $0 < n^{(j)} < n$*:*

i) For each $m \in M_j^+$*,*

$$P_{h_j,n_j,n^{(j)}}(m) = \begin{cases} q_m^+ n^{(j)}/\bar{n}_j & \text{if } \bar{n}_j \geq n^{(j)} \\ \frac{P(v_m > \hat{v})q_m^+}{\sum_{m \in M_j^+} P(v_m > \hat{v})q_m^+} & \text{if } \bar{n}_j < n^{(j)} \end{cases},$$

where $\hat{v} \equiv v_{(a,b;F)}$ *with* $a = n^{(j)} - \bar{n}_j$*,* $b = n^{(j)}$*, and* $F = G_{j+}$*.*

ii) For each $m \in M_j^-$*,*

$$P_{h_j,n_j,n^{(j)}}(m) = \begin{cases} \frac{P(v_m < \hat{v})q_m^- \left((\bar{n}_j - n^{(j)})/\bar{n}_j\right)}{\sum_{m \in M_j^-} P(v_m < \hat{v})q_m^-} & \text{if } \bar{n}_j > n^{(j)} \\ 0 & \text{if } \bar{n}_j \leq n^{(j)} \end{cases},$$

where $\hat{v} \equiv v_{(a,b;F)}$ *with* $a = \bar{n}_j - n^{(j)} + 1$*,* $b = n - n^{(j)}$*, and* $F = G_{j-}$*.*

Proof. We begin by introducing some notation. We denote the event that a worker who chose education level h_j matches to any firm of type $m \in M_j^+$ or $m \in M_j^-$ as M_{ij}^+ and M_{ij}^- respectively.[4]

First, we consider the probability that a worker who chose h_j matches to any firm in the class $m \in M_j^+$. Consider the case that $\bar{n}_j \geq n^{(j)}$. In this case, there are at least as many workers who chose h_j as firms who prefer h_j. Given that Condition IR implies that no worker or firm will never unilaterally dissolve a match to become unmatched, the case of $\bar{n}_j \geq n^{(j)}$ implies that every firm in class m who wants a worker with h_j will hire one in the matching process. For each class of firm $m \in M_j^+$, the probability that a worker who chose h_j matches to a firm in the set of firms that prefers h_j and to the particular class $m \in M_j^+$ is given as follows

[4] That is, $M_{ij}^+ \equiv \{\mathscr{M}_i \in M_j^+\}$ and similarly for $M_{ij}^- \equiv \{\mathscr{M}_i \in M_j^+\}$.

when $\bar{n}_j \geq n^{(j)}$:

$$\begin{aligned} P_j(\mathscr{M}_i = m, M_{ij}^+) &= P_j(\mathscr{M}_i = m | M_{ij}^+) P_j(M_{ij}^+). \\ &= q_m^+ n^{(j)} / \bar{n}_j, \end{aligned}$$

where the j-subscript on the probabilities denote a probability conditional on the event $H_i = h_j$. $P_j(M_{ij}^+)$ is equal to $n^{(j)}/\bar{n}_j$ because workers with the same h_j are indistinguishable to the firms that prefer them, so firms choose among these workers at random. The probability of matching to a firm of type $m \in M_j^+$ given that the worker has already matched to some firm in M_j^+ is equal to the relative proportion of type m firms in this category, q_m^+.

Next, we consider the case that $\bar{n}_j < n^{(j)}$. Since there are strictly more firms that prefer h_j than workers who chose h_j, the probability that a worker who chose h_j matches to a firm that prefers workers with h_j occurs with probability one; that is $P_j(M_{ij}^+) = 1$.[5]

Although $P_j(M_{ij}^+) = 1$, only the firms with the $\bar{n}_j$ largest v-indices will be able to match with a worker who chose h_j. Thus, a firm in M_j^+ matches to a worker with h_j if and only if its v statistic exceeded the $\kappa = n^{(j)} - \bar{n}_j$ order statistic among all $n^{(j)}$ firms in M_j^+. Thus, by Assumptions 1.3.1 and 1.3.2, the probability that a worker who chose h_j matches with a firm from class $m \in M_j^+$ conditional on matching to some firm in M_j^+ is

$$P(v(K) = v(k_m) | v(K) > \hat{v}, m \in M_j^+),$$

which by Bayes' rule equals

$$\frac{P(v(K) > \hat{v} | v(K) = v(k_m), m \in M_j^+) P(v(K) = v(k_m) | m \in M_j^+)}{\sum_{m \in M_j^+} P(v(K) > \hat{v} | v(K) = v(k_m), m \in M_j^+) P(v(K) = v(k_m) | m \in M_j^+)}, \quad \text{(A.11)}$$

where $\hat{v} \equiv v_{(\kappa, n^{(j)}; G_{j+})}$. Equation A.11 represents the relative proportion of type m firms represented among threshhold crossers among all firms that prefer h_j. We

[5]This follows from Condition IR and the following two facts: i) h_j workers are scarce relative to the firms that prefer them ii) firms that prefer $h_{j'}$ will never choose a h_j worker in the matching process since the condition $n_h = n_f = n$ and $J = 2$ implies that $h_{j'}$ workers are always available (i.e., when $n_h = n_f = n$, $n^{(j)} > \bar{n}_j$ implies that $n_{j'} > n^{(j')}$, since $n_{j'} = n - \bar{n}_j$ and $n^{(j')} = n - n^{(j)}$).

next consider the probability of matching to each firm with $m \in M_j^-$. We consider first the case that $\bar{n}_j > n^{(j)}$. The relevant probability is

$$\begin{aligned} P_j(\mathscr{M}_i = m, M_{ij}^-) &= P_j(\mathscr{M}_i = m | M_{ij}^-) P_j(M_{ij}^-) \\ &= P_j(\mathscr{M}_i = m | M_{ij}^-)(1 - n^{(j)}/\bar{n}_j). \end{aligned}$$

As stated above, the of case $\bar{n}_j > n^{(j)}$ combined with our assumption that $n_h = n_f = n$ implies that $n^{(j')} > n_{j'}$, since $n_{j'} = n - \bar{n}_j$ and $n^{(j')} = n - n^{(j)}$. Therefore by similar logic to before, firms who prefer $h_{j'}$ match to workers with h_j if their v-index is lower than the $n^{(j')} - n_{j'} + 1 = \bar{n}_j - n^{(j)} + 1$ order statistic among those firms in M_j^-. Letting $\kappa \equiv \bar{n}_j - n^{(j)} + 1$, the probability of a worker who chose h_j matching to a type $m \in M_j^-$ firm conditional on matching to some firm in M_{ij}^- is given as the proportion of type m firms whose v index falls below this threshhold:

$$P_j(\mathscr{M}_i = m | M_{ij}^-) = \frac{P(v_m < \hat{v}) q_m^-}{\sum_{m \in M_j^-} P(v_m < \hat{v}) q_m^-},$$

where $\hat{v} \equiv v_{(\kappa, n^{(j')}; G_{j-})}$. Lastly, in the case that $\bar{n}_j \leq n^{(j)}$, $P(M_{ij}^-) = 0$. This completes the proof. $\square$

Next we define $G \equiv \sum_{m=1}^{M} F_m q_m$. Proposition A.1.3 characterizes the matching probabilities in the case that all firms types prefer one level of education; that is, in the case that firm preferences are homogeneous over worker education types. The arguments are abridged, since they are very similar to those used in the proof of Proposition A.1.2.

Proposition A.1.3. *(Homogeneous firm preferences). Suppose that $n_h = n_f = n$. Then under the assumptions of Proposition A.1.1 we have the following for the cases that $n^{(j)} = n$ and $n^{(j)} = 0$.*

1. *if $n^{(j)} = n$, then $M_j^- = \emptyset$ and for each $m \in M_j^+ = M$ we have*

$$P_{h_j, n_j, n^{(j)}}(m) = \begin{cases} q_m & \text{if } \bar{n}_j = n \\ \frac{P(v_m > \hat{v}) q_m}{\sum_{m \in M} P(v_m > \hat{v}) q_m} & \text{if } \bar{n}_j < n \end{cases},$$

   where $\hat{v} \equiv v(a_1, a_2; G)$, with $a = n - \bar{n}_j$ and $b = n$.

2. *If $n^{(j)} = 0$, then $M_j^+ = \emptyset$ and for each $m \in M_j^- = M$ we have*

$$P_{h_j,n_j,n^{(j)}}(m) = \begin{cases} q_m & \text{if } \bar{n}_j = n \\ \frac{P(v_m < \hat{v})q_m}{\sum_{m \in M} P(v_m < \hat{v})q_m} & \text{if } \bar{n}_j < n \end{cases},$$

where $\hat{v} \equiv v(a_1, a_2; G)$, with $a = \bar{n}_j + 1$ and $b = n$.

Proof. When $n^{(j)} = n$ and $\bar{n}_j = n$ the probability of matching to firm m is simply equal to the marginal probability of that firm type in the economy, q_m. When $n^{(j)} = n$ and $\bar{n}_j < n$, using logic identical to that employed in the proof of Proposition A.1.2, we conclude that the probability of matching to a firm from class m is equal to the proportion of type m firms above the $n - \bar{n}_j$ order statistic of the v's.

When $n^{(j)} = 0$, we must have $\bar{n}_j > n^{(j)} = 0$ (since at least one person is assumed to choose h_j). Since the top $n_{j'} = n - \bar{n}_j$ ranked firms in terms of v receive a worker with their preferred education, $h_{j'}$, the probability of matching to a firm in class m is equal to the proportion of type m below the $\bar{n}_j + 1$ order statistic of the v's. □

The following result takes for granted a well-known fact that uniform order statistics follow the Beta distribution.[6]

Lemma A.1.3. *Let: i) $\{X_i\}_{i=1}^n$ be iid random variables from continuous distribution function G; ii) Z be normally distributed with mean μ and variance σ^2; iii) $X_{(i)}$ be the i-th order statistic of $\{X_i\}_{i=1}^n$; iv) $U_{(i)}$ be the i-th order statistic of iid uniform random variables $\{U_i\}_{i=1}^n$. Then,*

$$P(Z \geq X_{(i)}) = 1 - \mathbf{E}\Phi((G^{-1}(U_{(i)}) - \mu)/\sigma),$$

where $\Phi(\cdot)$ is the standard normal cdf, and $\mathbf{E}(\cdot)$ is taken over the distribution of $U_{(i)}$, which follows the Beta distribution with parameters i and $n+1-i$.

Proof. Note that since X_i's are continuously distributed according to G it follows from the probability integral transformation result that for each i

$$X_i =_d G^{-1}(U_i).$$

[6] For example, see Chapter 2 Ahsanullah et al. [7].

Also, since G is monotone we have that for each i

$$X_{(i)} =_d G^{-1}(U_{(i)}).$$

The previous line implies that

$$\begin{aligned}
P(Z \geq X_{(i)}) &= P(Z \geq G^{-1}(U_{(i)})) \\
&= 1 - P(Z \leq G^{-1}(U_{(i)})) \\
&= 1 - \mathbf{E}\Phi((G^{-1}(U_{(i)}) - \mu)/\sigma),
\end{aligned}$$

where $\mathbf{E}(\cdot)$ is taken over the distribution of $U_{(i)}$. The last equality used the fact that Z is normal with mean μ and variance σ^2. □

The following results are direct application of the previous results. They are useful for constructing the π_{mj}'s that are used in the structural estimation of this paper. Recall the definitions of G, G_{j+}, G_{j-}, and $v(b_1, b_2, F)$ from before. We introduce introduce the following notation:

$$a(\kappa, n, m; G) \equiv \mathbf{E}\Phi\left((G^{-1}(U_{(\kappa;n)}) - \beta k_m)/\sigma_m\right),$$

where $U_{(\kappa;n)}$ is the κ-order statistic of n uniform random variables and $\mathbf{E}(\cdot)$ is taken over the distribution of $U_{(\kappa;n)}$.

Corollary A.1.1. *Suppose the conditions of Proposition A.1.2 hold and let v_m be distributed according to F_m. Then, in the heterogeneous preferences case with $\bar{n}_j < n^{(j)}$,*

1. *For each $m \in M_j^+$, $P(v_m > v_{(\kappa, n^{(j)}; G_{j+})}) = 1 - a(\kappa, n^{(j)}, m; G_{j+})$, where $\kappa = n^{(j)} - \bar{n}_j$.*
2. *For each $m \in M_j^-$, $P(v_m < v_{(\kappa; n^{(j')}; G_{j-})}) = a(\kappa, n^{(j')}, m; G_{j-})$, where $\kappa = \bar{n}_j - n^{(j)} + 1$.*

Corollary A.1.2. *Suppose the conditions of Proposition A.1.3 hold and let v_m be distributed according to F_m. Then, in the homogeneous preferences case with $\bar{n}_j < n$,*

1. *If* $n^{(j)} = 0$*,* $P(v_m < v_{(\kappa;n,G)}) = a(\kappa, n, m; G)$ *for each* $m \in M$*, where* $\kappa = \bar{n}_j + 1$.

2. *If* $n^{(j)} = n$*,* $P(v_m > v_{(\kappa;n,G)}) = 1 - a(\kappa, n, m; G)$ *for each* $m \in M$*, where* $\kappa = n - \bar{n}_j$.

Proof. The proofs of Corollaries 1 and 2 follows directly from Lemma A.1.3. □

A.1.3 A Monte Carlo Simulation Study

In this section, we investigate the finite sample size and power properties of the estimator of preferences, $\hat{\theta}_n(\beta)$, under a variety of parameters and functional form assumptions. The results in this section are for the case the matching technology, β, is known to the econometrician.

In this study, we choose the following general structure for the worker's expected utility function:

$$\tilde{U}_i = (f_i + g_i)/2 + d(H_i)\varepsilon_i,$$

where $\theta = (\theta_1, \theta_2')' \in \mathbf{R}^4$, $X_i \in \mathbf{R}^3$. $d(H_i)$ be 2×1 vector with one in the H_i-th row where $H_i \in \{1,2\}$.We suppose that $\varepsilon_i \in \mathbf{R}^2$ follows the extreme value distribution so that the best response probability function of each worker has the logit structure. We consider two functional forms for the production function f_i, which we call Specification 1 (f_{i1}) and Specification 2 (f_{i2}):

$$f_{i1} = \theta_1 H_i \cdot \pi_i(\theta, \beta)'k, \text{ and}$$
$$f_{i2} = \theta_1 (H_i + \pi_i(\theta, \beta)'k).$$

f_{i1} implies direct production complementarities between the worker and firm variables whereas any complementarities in f_{i2} are forced through the worker's expectation of firm capital $\pi_i'k$. We also choose the following g_i that ensures that the worker's outside option is positive[7]:

$$g_i = \exp(H_i \cdot X_i' \theta_2).$$

$X_i = (X_{1i}, X_{2i}, X_{3i})'$ are drawn independently across i and one another from $U[0,1]$. For each simulation sample, the H_i's are generated as follows. First, we solve for fixed point in the best response operator to obtain $\mathbf{P}^*$.[8] Then we compute the best response at the simulated covariates

$$\Psi_i(H_i|X_i, \mathbf{P}^*_{-i}) = \frac{\exp(\tilde{U}_i^*(H_i, X_i))}{\sum_{j=1}^{2} \exp(\tilde{U}_i^*(H_j, X_i))}.$$

Letting $\Psi_i^*(X_i) \equiv \Psi_i(H_i|X_i, \mathbf{P}^*_{-i})$ we generate the simulated actions as,

$$H_i = 1\{\Psi_{i1}^* > \omega_i\}$$

where ω_i's are drawn iid from the uniform distribution on $[0,1]$.

[7]Note that when $g_i = H_i \exp(X_i' \theta_2)$ - that is, a functional form guaranteeing that g_i is increasing in H_i - also yielded comparable performacne in the the simulation studies.

[8]In experiments with different starting values, iterating the best response operator yielded the same fixed point each time.

Table A.1: The Empirical Coverage Probability of Asymptotic Confidence Intervals for $a'\theta_0$ at 95% Nominal Level When β_0 is Known.

		Specification 1			Specification 2		
β_0		$M=2$	$M=3$	$M=5$	$M=2$	$M=3$	$M=5$
-1	$n=500$	0.9480	0.9430	0.9470	0.9510	0.9410	0.9430
	$n=1000$	0.9480	0.9560	0.9490	0.9450	0.9450	0.9270
	$n=2000$	0.9380	0.9300	0.9250	0.9590	0.9300	0.9090
-0.5	$n=500$	0.9470	0.9490	0.9490	0.9530	0.9500	0.9310
	$n=1000$	0.9480	0.9370	0.9550	0.9480	0.9310	0.9300
	$n=2000$	0.9430	0.9430	0.9450	0.9170	0.9180	0.8860
0	$n=500$	0.9460	0.9470	0.9530	0.9380	0.9350	0.9380
	$n=1000$	0.9490	0.9490	0.9530	0.9370	0.9480	0.9440
	$n=2000$	0.9360	0.9510	0.9400	0.9580	0.9530	0.9490
0.5	$n=500$	0.9520	0.9530	0.9600	0.9420	0.9310	0.9330
	$n=1000$	0.9410	0.9500	0.9360	0.9460	0.9450	0.9440
	$n=2000$	0.9430	0.9260	0.9230	0.9560	0.9500	0.9280
1	$n=500$	0.9440	0.9430	0.9380	0.9330	0.9360	0.9270
	$n=1000$	0.9260	0.9180	0.8930	0.9470	0.9340	0.9220
	$n=2000$	0.9170	0.9050	0.8850	0.9200	0.9240	0.9190

Notes: The table reports the empirical coverage probability of the asymptotic confidence interval for θ_0. The simulated rejection probability at the true parameter is close to the nominal size of $\alpha = 0.05$. The simulation number is $R = 1000$.

Table A.2: Average Length of Confidence Intervals for $a'\theta_0$ at 95% Nominal Level When β_0 is Known.

β_0		$M=2$	$M=3$	$M=5$	$M=2$	$M=3$	$M=5$
-1	$n=500$	15.5881	1.8333	1.8353	1.6223	1.4490	1.3557
	$n=1000$	1.3573	1.2290	1.2920	1.1102	0.9449	0.9240
	$n=2000$	0.9093	0.8762	0.8643	0.6812	0.6858	0.6385
-0.5	$n=500$	2.3023	2.0498	2.0104	2.4145	1.9593	1.6319
	$n=1000$	3.4657	1.508	1.3757	1.6552	1.2743	1.0797
	$n=2000$	1.0376	1.0841	1.0300	1.0803	0.8720	0.7451
0	$n=500$	2.2851	2.2847	2.2773	1.1456	1.1542	1.2197
	$n=1000$	1.5994	1.5993	1.6397	0.7371	0.8457	0.7375
	$n=2000$	3.1524	1.1593	1.1208	0.5205	0.5196	0.5206
0.5	$n=500$	3.0442	2.9003	2.9723	2.4889	3.0612	3.1814
	$n=1000$	3.0954	1.9949	2.1289	2.0466	2.0044	2.3200
	$n=2000$	1.5633	1.4310	1.5487	0.9286	1.3152	1.2469
1	$n=500$	6.0386	11.0447	4.7156	4.8434	4.7347	4.3571
	$n=1000$	2.7670	3.1329	4.6450	2.8590	3.1888	3.2592
	$n=2000$	1.6956	2.0135	2.2905	2.1166	2.5055	2.2695

Notes: This table reports the average length of the asymptotic confidence interval for θ_0. The lengths of the of the confidence intervals decrease with n. The simulation number is $R = 1000$.

A.1.4 Empirical Section

Variables

We use variables EDC_1 through EDC_12 to construct the indicator variable for high educated. A number of definitions are possible. The results record the value of $H_i = 1$ if the individual has completed both high school and a college degree. We rely on firm size as our measure of firm productivity. There is specific evidence that firm size is a useful proxy for firm productivity in both manufacturing and non-manufacturing sectors in the Canadian context. Leung et al. (2008), using Canadian administrative data for the period 1984-1997, argue that firm size is positively correlated with measures of labour and total factor productivity, particularly within the manufacturing sector.

For the outside option function, we use the worker's marital status, the number of dependent children the worker has, and the worker's gender. We drop the few individuals who reported having six dependent children (the maximum allowable in the sample). The form of the outside option is as reported in the simulations study.

A.1.5 Estimation

The estimation proceeds in two steps, broadly as outlined in 1.3. We begin by discussing the estimation of θ_0, then we discuss the Monte-Carlo inference approach used to construct confidence intervals for β in oru samples of interest.

We set the parameter space for β_0 to be $B = [-1 : 0.075 : 3]'$. We normalize $\sigma = 1$ throughout. For each value of $\beta \in B$ we obtain $\hat{\theta}(\beta)$ using the employee-final weights provided for WES by Statistics Canada. Standard errors are constructed as the square roots of the diagonal elements of inverse of the sample Hessian, constructed using numerical differentiation of the log-likelihood function via a five-point stencil approach.

Much of computational difficulty associated with the estimation of θ involves the construction of the worker's expectations, i.e., the $\hat{\pi}_j$'s. For the number of support points for the capital variable, we set a value of $M = 5$, but the results are not very sensitive to similar values of M (i.e., $M = 3$, $M = 7$). The empirical distribution of firm size, $\hat{q}$, is constructed using the WES workplace weights after grouping the firms into categories based on log-firm size. We also use a value of fifty draws of random variables from the beta distribution for the construction of the thresh-hold crossing-probabilities. We use the empirical distribution of high education as the equilibrium choice probabilities. For the distribution of, $N^{(j)}$, the number of firms that prefer workers who chose education level j, we use the binomial probability mass function with probability equal to the expected fraction of firms who prefer education level j. In practice, we construct the matching probabilities using an interpolation of the support of $N^{(j)}$. A choice of forty support points is found to work well.

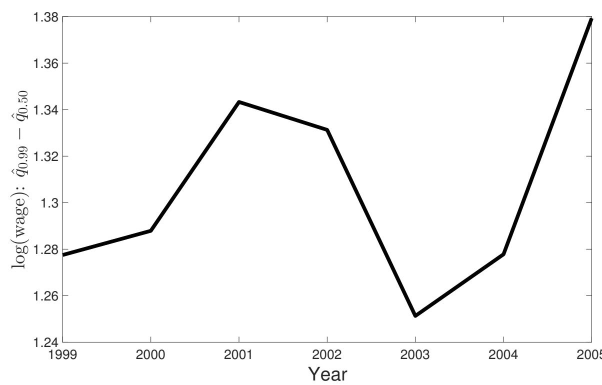
The test statistic of interest is based on matched observed characteristics in the data. In particular, we consider a test statistic that compares the observed joint distribution of worker human capital and the matched firm capital to the simulated counterpart. That is,

$$T(\beta) \quad = \quad \frac{1}{R}\sum_{r=1}^{R}\left\|\hat{P}-\hat{P}_r(\beta,\hat{\theta}_n(\beta))\right\|, \tag{A.12}$$

where $\hat{P}$ is an $M \times J$ matrix whose (m,j) element is the probability that a worker of education level j matches to a firm of capital level m (i.e., $\hat{P}(\mathbf{M}(i)=m,h_i=j)$), $\hat{P}_r$ is similarly defined except we use the the simulated matching, $\mathbf{M}_{ir}(i;\beta,\hat{\theta}(\beta))$, in place of the observed matching, $\mathbf{M}(i)$, and $\|\cdot\|$ is the Frobenius norm. Note that when we construct $\hat{P}_r$ we use the $\hat{\theta}$ that were estimated using the employee-weights. However, no set of weights provided by Statistics Canada are appropriate at the level of the match itself. We also define an estimator of β as the minimizer of the $T(\beta)$ as a heuristic measure of β.

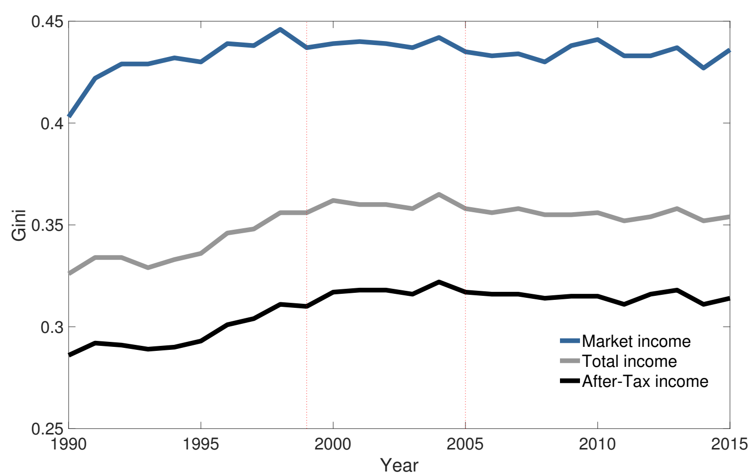
A.1.6 Estimation Results

Figure A.1: Wage Inequality in Canada's Workplace-Employee Survey: 99-50 Difference in Quantile of Log Hourly Wages



The plot shows the difference in the weighted sample quantiles of log hourly wages (HR_WAGET) in the WES sample years of 1999-2005.

Figure A.2: Income Inequality in Canada: Gini coefficients, 1990-2015



The figure plots Gini coefficients for total, adjusted market, and after-tax income for Canada. The period of 1999-2005, coinciding with WES, is a period of relatively stable income inequality in Canada. The source of the data is CANSIM Table 206-0033 from Income Statistics Division, Statistics Canada.

Table A.3: Matching Technology In Canadian Manufacturing and Finance Industries, 1995-2005

	Manufacturing		Finance	
Year	Specification 1	Specification 2	Specification 1	Specification 2
1999	1.475	1.775	0.950	0.875
	[0.5000,2.450]	[0.575, 2.450]	[0.200,1.625]	[0.200, 1.625]
2000	1.475	1.475	1.625	1.550
	[0.2000,2.525]	[0.275 2.600]	[0.950,1.850]	[1.100, 1.925]
2001	-1.000	-0.175	0.575	0.575
	[-1.000,1.325]	[-1.000, 1.400]	[-0.775,1.700]	[-0.700, 1.700]
2002	-0.175	-0.100	0.725	0.650
	[-1.000,1.775]	[-1.000, 1.625]	[-.550,2.150]	[-0.550, 2.225]
2003	0.725	0.875	0.875	0.800
	[-1.000,2.375]	[-1.000 2.600]	[0.0250,1.750]	[-0.025, 1.700]
2004	1.325	1.100	0.650	0.575
	[-0.325,2.450]	[-0.325, 2.525]	[-0.550,1.775]	[-0.550, 1.850]
2005	2.300	2.375	0.425	0.425
	[0.950,2.975]	[0.950, 2.975]	[-0.550,1.550]	[-0.550, 1.550]

This table reports minimum distances estimates of β using the test statistics considered A.12 along with 95% confidence intervals for β for the years 1999-2005 using WES data for managers and professionals in the Secondary Products Manufacturing sector and the Finance and Insurance Industry. Specification 1 and Specification 2 refer to the cases in which worker and firm capital are multiplicative and additive, respectively. The results for β are similar across specifications. The matching technology in the manufacturing industry exhibited the least frictions at the start of the sample, rising in the middle, then falling again towards the end of the sample. In the finance industry, we find an increase in matching frictions from 1999 onwards. Weighted sample sizes for relevant years and industries are reported in table A.4 and A.5.

A.1.7 Model Counterfactuals

Table A.4: Estimation of Worker and Firm Preferences in Canadian Manufacturing Industry, 1999-2005

Specification 1

	1999	2000	2001	2002	2003	2004	2005
θ_1	3.6061	1.9910	3.2229	0.0596	4.4090	5.4205	3.2859
	(0.0023)	(0.0012)	(0.0006)	(0.0016)	(0.0016)	(0.0010)	(0.0015)
θ_2	-0.9760	0.1460	0.0240	-0.3201	-0.4391	0.7434	-29.7608
	(0.0269)	(0.0011)	(0.0002)	(0.0005)	(0.0012)	(0.0003)	$(6.785 \cdot 10^9)\dagger$
θ_3	0.2742	-30.7611	0.2259	0.6203	0.5536	-1.3002	0.0877
	(0.0009)	$(3.132 \cdot 10^9)\dagger$	(0.0005)	(0.0003)	(0.0006)	(0.0058)	(0.0007)
θ_4	-0.0889	0.5008	0.2017	0.1410	-0.0299	-0.3937	0.2015
	(0.0006)	(0.0004)	(0.0002)	(0.0001)	(0.0003)	(0.0014)	(0.0002)
b_i	71361	72464	80738	76214	87026	153520	98879

Specification 2

	1999	2000	2001	2002	2003	2004	2005
θ_1	1.7022	0.9582	0.2891	0.0257	2.1104	2.5843	1.8401
	(0.0004)	(0.0005)	(0.0004)	(0.0005)	(0.0005)	(0.0003)	(0.0006)
θ_2	-0.9760	-5.5360	-0.1571	-0.3201	-0.4391	0.5897	-30.3271
	(0.0002)	(0.0110)	(0.0003)	(0.0072)	(0.0046)	(0.0015)	(0.0071)
θ_3	0.2742	0.6741	0.6012	0.6203	0.5536	-0.1873	0.0877
	(0.0009)	(0.0004)	(0.0077)	(0.0003)	(0.0007)	(0.0007)	(0.0002)
θ_4	-0.0889	-0.2500	0.1432	0.1410	-0.0299	0.1068	0.2015
	(0.0039)	(0.0002)	$(2.423 \cdot 10^9)\dagger$	(0.0004)	(0.0003)	(0.0003)	(0.0001)
b_i	71361	72464	80738	76214	87026	153520	98879

This table reports preference parameter estimates and standard errors (in parantheses) for professionals and managers in the Canadian manufacturing industry for two specifications for the years 1995-2005 (WES sample frame). b_i denotes the weighted sample size in year t. For each year, we report the value of $\hat{\theta}_n(\hat{\beta})$, where $\hat{\beta}$ is the value of the minimum distance estimate for that year. Specification 1 and Specification 2 refer to the cases in which worker and firm capital are multiplicative and additive respectively. θ_1 is the coefficient on worker and firm attributes in the production function. The remainder are coefficients on the outside option function θ_2: female$_i$, θ_3: marital status, θ_4: number of dependent children. All coefficients other than ones with † are statistically significant at $\alpha = 0.01$. Both specifications suggest modest increases in the production technology parameter over time, θ_1 . Both specifications, particularly the additive, suggest a negative coefficient on female$_i$ in the worker's wage equation.

Table A.5: Estimation of Worker and Firm Preferences in Canadian Finance Industry, 1999-2005

Specification 1

	1999	2000	2001	2002	2003	2004	2005
θ_1	2.8284	2.5676	3.2546	2.9339	2.4986	2.5479	4.5527
	(0.0014)	(0.0012)	(0.0009	(0.0014)	(0.0013)	(0.0007)	(0.0016)
θ_2	-1.3669	-1.2849	0.4988	-1.2993	-1.7819	-0.9356	-1.3559
	(0.0055)	(0.011)	(0.0003)	(0.0072)	(0.0046)	(0.0015)	(0.0071)
θ_3	0.0842	0.4970	-1.1876	0.4788	-0.3668	-0.3690	0.6391
	(0.0007)	(0.0004)	(0.0053)	(0.0003)	(0.0007)	(0.0007)	(0.0002)
θ_4	0.1541	0.0397	-0.3510	-0.2085	0.4309	0.4245	-0.0591
	(0.0003)	(0.0002)	(0.0014)	(0.0004)	(0.0003)	(0.0003)	(0.0001)
b_i	193420	159320	191340	180890	188320	190750	185670

Specification 2

	1999	2000	2001	2002	2003	2004	2005
θ_1	1.0408	1.0157	1.1637	1.1011	0.9652	0.9886	1.7575
	(0.0004)	(0.0005)	(0.0003)	(0.0005)	(0.0005)	(0.0003)	(0.0006)
θ_2	0.5115	-1.2849	0.4988	-1.2993	-1.7819	-0.9356	-1.3559
	(0.0002)	(0.0110)	(0.0003)	(0.0072)	(0.0046)	(0.0015)	(0.0071)
θ_3	-0.4881	0.4970	-1.1876	0.4788	-0.3668	-0.3690	0.6391
	(0.0009)	(0.0004)	(0.0053)	(0.0003)	(0.0007)	(0.0007)	(0.0002)
θ_4	-1.0452	0.0397	-0.3510	-0.2085	0.4309	0.4245	-0.0591
	(0.0039)	(0.0002)	(0.0014)	(0.0004)	(0.0003)	(0.0003)	(0.0001)
b_i	193420	159320	191340	180890	188320	190750	185670

This table reports preference parameter estimates and standard errors (in parantheses) for professionals and managers in the Canadian finance and insurance industry for two specifications for the years 1995-2005 (WES sample frame). b_i denotes the weighted sample size in year t. For each year, we report the value of $\hat{\theta}_n(\hat{\beta})$, where $\hat{\beta}$ is the value of the minimum distance estimate for that year. Specification 1 and Specification 2 refer to the cases in which worker and firm capital are multiplicative and additive respectively. θ_1 is the coefficient on worker and firm attributes in the production function. The remainder are coefficients on the outside option function θ_2: female$_i$, θ_3: marital status, θ_4: number of dependent children. All results are statistically significant at $\alpha = 0.01$. Both specifications suggest a relatively stable production technology, θ_1, over time with increases in 2005 . Both specifications suggest a negative coefficient on female$_i$ in the worker's wage equation.

Table A.6: Counterfactual Estimated Probabilities of Investing in High Education, Manufacturing Industry

	Specification 1						
Year	1999	2000	2001	2002	2003	2004	2005
$\hat{\beta}^{CF}_{1999}$	**0.7443**	0.7310	0.8605	0.6723	0.8205	0.8276	0.7622
$\hat{\beta}^{CF}_{2000}$	0.7442	**0.7310**	0.8605	0.6723	0.8205	0.8275	0.7622
$\hat{\beta}^{CF}_{2001}$	0.6656	0.7263	**0.8596**	0.6736	0.7494	0.7436	0.7049
$\hat{\beta}^{CF}_{2002}$	0.6947	0.7281	0.8600	**0.6732**	0.7763	0.7758	0.7259
$\hat{\beta}^{CF}_{2003}$	0.7237	0.7298	0.8603	0.6727	**0.8026**	0.8066	0.7470
$\hat{\beta}^{CF}_{2004}$	0.7404	0.7308	0.8604	0.6724	0.8172	0.**8235**	0.7594
$\hat{\beta}^{CF}_{2005}$	0.7624	0.7321	0.8607	0.6719	0.8358	0.8400	**0.7762**

	Specification 2						
Year	1999	2000	2001	2002	2003	2004	2005
$\hat{\beta}^{CF}_{1999}$	**0.7253**	0.6761	0.7036	0.6724	0.8066	0.8283	0.7762
$\hat{\beta}^{CF}_{2000}$	0.7224	**0.6744**	0.7042	0.6725	0.8040	0.8254	0.7738
$\hat{\beta}^{CF}_{2001}$	0.7036	0.6635	**0.7074**	0.6728	0.7860	0.8055	0.7582
$\hat{\beta}^{CF}_{2002}$	0.7045	0.6641	0.7072	**0.6728**	0.7868	0.8065	0.7590
$\hat{\beta}^{CF}_{2003}$	0.7160	0.6707	0.7053	0.6726	**0.7980**	0.8187	0.7685
$\hat{\beta}^{CF}_{2004}$	0.7185	0.6721	0.7049	0.6725	0.8003	**0.8213**	0.7705
$\hat{\beta}^{CF}_{2005}$	0.7306	0.6793	0.7027	0.6723	0.8116	0.8336	**0.7806**

This table reports the model-simulated probability of investing in high education at the estimated parameter values for the secondary products manufacturing sector (WES industry 4) for two-specifications. Specification 1 and 2 are defined in the Section A.1.6. The results demonstrate the importance of the matching technology and complementarities on education decisions. The equilibrium probability of education is typically higher in the complementarities case (Specification 1). In the manufacturing sector in 1999 (a high β year), the effect of switching to the matching technology from 2001 causes a fall in the equilibrium probability of attending college by roughly 8% in Specification 1, and 2.5% in Specification 2. The variation of preferences and worker distribution of characteristics over time is also significant. In 2001, if preferences and characteristics and the were as they were in 2002, the probability of investing in higher education would plummet from 86% to 67%. The effect in greatly attenuated in Specification 2, which assumes no production complementarities between worker and firm types are present in production.

Table A.7: Counterfactual Estimated Probabilities of Investing in High Education, Finance Industry

	Specification 1						
Year	1999	2000	2001	2002	2003	2004	2005
$\hat{\beta}^{CF}_{1999}$	**0.6903**	0.6961	0.7163	0.6979	0.6833	0.6799	0.7889
$\hat{\beta}^{CF}_{2000}$	0.7059	**0.7097**	0.7341	0.7146	0.6965	0.6932	0.8050
$\hat{\beta}^{CF}_{2001}$	0.6807	0.6877	**0.7052**	0.6877	0.6751	0.6719	0.7787
$\hat{\beta}^{CF}_{2002}$	0.6847	0.6911	0.7097	**0.6918**	0.6784	0.6752	0.7829
$\hat{\beta}^{CF}_{2003}$	0.6885	0.6944	0.7142	0.6960	**0.6817**	0.6784	0.7870
$\hat{\beta}^{CF}_{2004}$	0.6827	0.6894	0.7075	0.6898	0.6768	**0.6735**	0.7808
$\hat{\beta}^{CF}_{2005}$	0.6767	0.6842	0.7005	0.6833	0.6717	0.6686	**0.7744**

	Specification 2						
Year	1999	2000	2001	2002	2003	2004	2005
$\hat{\beta}^{CF}_{1999}$	**0.6399**	0.6619	0.6507	0.6415	0.6472	0.6431	0.7380
$\hat{\beta}^{CF}_{2000}$	0.6449	**0.6665**	0.6565	0.6470	0.6514	0.6474	0.7442
$\hat{\beta}^{CF}_{2001}$	0.6376	0.6598	**0.6480**	0.6390	0.6452	0.6411	0.7351
$\hat{\beta}^{CF}_{2002}$	0.6382	0.6603	0.6487	**0.6396**	0.6457	0.6416	0.7358
$\hat{\beta}^{CF}_{2003}$	0.6394	0.6614	0.6500	0.6409	**0.6467**	0.6426	0.7373
$\hat{\beta}^{CF}_{2004}$	0.6376	0.6598	0.6480	0.6390	0.6452	**0.6411**	0.7351
$\hat{\beta}^{CF}_{2005}$	0.6364	0.6587	0.6466	0.6376	0.6441	0.6401	**0.7336**

Specification 1 and 2 are as defined in Seciton A.1.6. As in the manufacturing sector, the equilibrium probability of investing in education is higher in the case of complemenarities (Specification 1). Changes in preferences and technology and characteristics matter to education patterns: for example, in 1999, the effect of switching to 2005's preferences and exogenous characteristics leads to an increase of almost 10% in both specifications.

Table A.8: Counterfactual Estimated Probabilities of Investing in High Education, Manufacturing Industry

	Specification 1						
Year	1999	2000	2001	2002	2003	2004	2005
$\hat{\beta}_{year}$	0.7443	0.7310	0.8596	0.6732	0.8026	0.8235	0.7762
$\beta = 0$	0.7007	0.7285	0.8600	0.6731	0.7817	0.6822	0.7302
$\beta = 5$	0.7954	0.7342	0.8610	0.6710	0.8629	0.8758	0.8036

	Specification 2						
Year	1999	2000	2001	2002	2003	2004	2005
$\hat{\beta}_{year}$	0.7253	0.6744	0.7074	0.6728	0.7980	0.8213	0.7806
$\beta = 0$	0.7058	0.6647	0.7070	0.6727	0.7880	0.8078	0.7600
$\beta = 5$	0.7456	0.6888	0.6998	0.6721	0.8250	0.8483	0.7936

In this table we consider the effects of very low frictions ($\beta = 5$) and maximal frictions ($\beta = 0$) in the case that the production function exhibits interactions between worker and firm characteristics (Specification 1) and when they do not (Specification 2). The table illustrates the importance of both production complementarities and matching frictions to educational attainment. In 2004, the effect of lowering β to 0 causes a fall in the probability of high education by 14% in Specification 1, but only by 2% in Specificaiton 2. The overall level of investment in education is typically much lower in the case with high matching frictions (low β). β has a greater effect on the outcome in the complementarities case (Specification 1).

Table A.9: Counterfactual Estimated Probabilities of Investing in High Education, Finance Industry

	Specification 1						
Year	1999	2000	2001	2002	2003	2004	2005
$\hat{\beta}_{year}$	0.6903	0.7097	0.7052	0.6918	0.6817	0.6735	0.7744
$\beta = 0$	0.6648	0.6739	0.6867	0.6707	0.6618	0.6588	0.7615
$\beta = 5$	0.7509	0.7500	0.7830	0.7618	0.7366	0.7335	0.8492

	Specification 2						
Year	1999	2000	2001	2002	2003	2004	2005
$\hat{\beta}_{year}$	0.6399	0.6665	0.6480	0.6396	0.6467	0.6411	0.7336
$\beta = 0$	0.6330	0.6555	0.6425	0.6339	0.6412	0.6372	0.7292
$\beta = 5$	0.6614	0.6817	0.6755	0.6650	0.6658	0.6618	0.7645

In this table we consider the effects of very low frictions ($\beta = 5$) and maximal frictions ($\beta = 0$) for Specifications 1 and 2 in the finance industry. The implications are similar to those for the manufacturing industry in the previous table. A rise in β causes a much greater increase in the probability of high education in Specification 1: In 2001, the effect of a rise in the estimated β to $\beta = 5$ leads to a 8% increase in the equilibrium probability of high education under Specification 1, but only a 3% increase in Specification 2. The level of investment in education is lower in the case without complementarities (Specification 2).

Table A.10: Matching Technology Counterfactuals, Simulated Gini Coefficient, Manufacturing Industry

	Specification 1						
Year	1999	2000	2001	2002	2003	2004	2005
$\hat{\beta}_{year}$	0.1916	0.4207	0.2112	0.4216	0.1646	0.2121	0.2220
$\beta = 0$	0.2138	0.4299	0.2149	0.4218	0.2013	0.2534	0.2507
$\beta = 5$	0.2112	0.4330	0.2280	0.4208	0.1951	0.2360	0.2452

	Specification 2						
Year	1999	2000	2001	2002	2003	2004	2005
$\hat{\beta}_{year}$	0.1129	0.2203	0.3481	0.4206	0.1183	0.1268	0.1653
$\beta = 0$	0.1181	0.2238	0.3499	0.4207	0.1301	0.1285	0.1710
$\beta = 5$	0.1255	0.2311	0.3478	0.4201	0.1294	0.1251	0.1728

This table reports the model-simulated Gini coefficients under two-specifications. Specification 1 and 2 are as defined in Seciton A.1.6. The predicted level of wage inequality is typically much lower in Specification 2, where there are no production complementarities. In Specification 1, the level of inequality at the estimated value of the frictions is lower than at the counterfactual levels for most years (except 2002). For example, in 2005 the simulated Gini is 0.222 and the investment in education is 77%. This rises to 0.2507 (education investment 76%) when information frictions are highest and 0.2452 (education investment 85%) when frictions are lowest. Similar patterns can be seen in Specification 2.

Table A.11: Matching Technology Counterfactuals, Gini Coefficient, Finance Industry

	Specification 1						
Year	1999	2000	2001	2002	2003	2004	2005
$\hat{\beta}_{year}$	0.2602	0.2929	0.2769	0.2542	0.3594	0.2976	0.2607
$\beta = 0$	0.2516	0.2814	0.2502	0.2332	0.3408	0.2813	0.2495
$\beta = 5$	0.2647	0.2899	0.2523	0.2380	0.3421	0.2856	0.2432

	Specification 2						
Year	1999	2000	2001	2002	2003	2004	2005
$\hat{\beta}_{year}$	0.1998	0.2245	0.1756	0.1655	0.2801	0.2200	0.1815
$\beta = 0$	0.1984	0.2237	0.1714	0.1623	0.2759	0.2185	0.1793
$\beta = 5$	0.2075	0.2315	0.1803	0.1697	0.2827	0.2266	0.1837

This table reports the model-simulated Gini coefficients under two-specifications. Specification 1 and 2 are as defined in Section A.1.6. As in the manufacturing industry, the levels of inequality are typically much lower when there are no production complementarities (Specification 2). In Specification 1, the level of inequality at the estimated value of the frictions is higher than at the counterfactual levels for each year (except 1999). In Specification 2, the effect of lowering matching frictions raises wage inequality in every year. In this case, when frictions are lowered, the effect of increased sorting is stronger than the inequality-lowering effect of the greater supply of highly educated workers .

A.2 Appendix to Chapter 2

Proof of Theorem 2.2.1: From the optimization of agent i, we have

$$s_i^{\mathsf{BR}}(\mathscr{I}_i) = X_{i,1}'\gamma_0 + \tilde{X}_{i,2}'\delta_0 + \beta_0 \mathbf{E}[\tilde{Y}_i^B|\mathscr{I}_i] + \varepsilon_i + \eta_i, \tag{A.13}$$

where $\tilde{Y}_i^B$ denotes the weighted average (over $N_P(i)$) of $s_k(\mathscr{I}_k)$ where the weights are given by beliefs of player i and r_{ik}. Thus we write

$$\mathbf{E}[\tilde{Y}_i^B|\mathscr{I}_i] = \frac{1}{n_P(i)} \sum_{k \in N_P(i)} r_{ik} \sum_{j \in \overline{N}_I(i)} w_{kj}^{i\prime} T_j 1\{j \in \overline{N}_I^i(k)\}.$$

Plugging in this in (A.13), we have

$$\begin{aligned} s_i^{\mathsf{BR}}(\mathscr{I}_i) &= \left(\gamma_0 + \frac{\beta_0}{n_P(i)} \sum_{k \in N_P(i)} r_{ik} w_{ki,1}^i 1\{i \in \overline{N}_I^i(k)\}\right)' X_{i,1} \\ &\quad + \left(1 + \frac{\beta_0}{n_P(i)} \sum_{k \in N_P(i)} r_{ik} w_{ki,\varepsilon}^i 1\{i \in \overline{N}_I^i(k)\}\right) \varepsilon_i + A_n + B_n, \end{aligned}$$

where

$$\begin{aligned} A_n &= \frac{\beta_0}{n_P(i)} \sum_{k \in N_P(i)} r_{ik} w_{ki,2}^{i\prime} X_{i,2} 1\{i \in \overline{N}_I^i(k)\}, \text{ and} \\ B_n &= \beta_0 \sum_{j \in N_I(i)} \frac{1}{n_P(i)} \sum_{k \in N_P(i)} r_{ik} (w_{kj,1}^{i\prime} X_{j,1} + w_{kj,\varepsilon}^i \varepsilon_j) 1\{j \in \overline{N}_I^i(k)\} \\ &\quad + \beta_0 \sum_{j \in N_I(i)} \frac{1}{n_P(i)} \sum_{k \in N_P(i)} r_{ik} w_{kj,2}^{i\prime} X_{j,2} 1\{j \in \overline{N}_I^i(k)\} \\ &\quad + \frac{1}{n_P(i)} \sum_{k \in N_P(i)} r_{ik} \delta_0' X_{k,2}. \end{aligned}$$

By setting the coefficients of $X_{j,1}$, ε_j and $X_{j,2}$ to be $w^i_{j,1}$, $w^i_{j,\varepsilon}$ and $w^i_{j,2}$, we obtain that

$$\begin{aligned} w_{ii,1} &= \gamma_0 + \beta_0 \mathscr{M}_i w^i_{i,1}, & \text{(A.14)} \\ w_{ii,\varepsilon} &= 1 + \beta_0 \mathscr{M}_i w^i_{i,\varepsilon}, \\ w_{ii,2} &= \beta_0 \mathscr{M}_i w^i_{i,2}, \end{aligned}$$

and for all $j \in N_I(i)$,

$$\begin{aligned} w_{ij,1} &= \beta_0 \mathscr{M}_i w^i_{j,1}, & \text{(A.15)} \\ w_{ij,\varepsilon} &= \beta_0 \mathscr{M}_i w^i_{j,\varepsilon}, \text{ and} & \text{(A.16)} \\ w_{ij,2} &= \begin{cases} \delta_0 r_{ij}/n_P(i) + \beta_0 \mathscr{M}_i w^i_{j,2}, & \text{if } j \in N_P(i), \\ \beta_0 \mathscr{M}_i w^i_{j,2}, & \text{if } j \in N_I(i) \backslash N_P(i), \end{cases} & \text{(A.17)} \end{aligned}$$

where

$$\mathscr{M}_i w^i_j = \frac{1}{n_P(i)} \sum_{k \in N_P(i)} r_{ik} w^i_{kj} 1\{j \in \bar{N}^i_I(k)\}.$$

Now, we apply the behavioral assumptions to this operator to obtain the following:

$$\begin{aligned} \mathscr{M}_i w^i_i &= \frac{1}{n_P(i)} \sum_{k \in N_P(i)} r_{ik} \tau^i_{ki} w_{ii} 1\{i \in \overline{N}_P(k)\} \\ &= \frac{1}{n_P(i)} \sum_{k \in N_P(i)} r_{ik} \tau^i_{ki} w_{ii} \\ &= \frac{1}{n_P(i)} \sum_{k \in N_P(i)} \frac{1}{\bar{n}_P(k)} w_{ii} = c_{ii} w_{ii}. \end{aligned}$$

By plugging this into (A.14), we have

$$\begin{aligned} w_{ii,1} &= \gamma_0 + \beta_0 w_{ii,1} c_{ii}, & \text{(A.18)} \\ w_{ii,\varepsilon} &= 1 + \beta_0 w_{ii,\varepsilon} c_{ii}, \end{aligned}$$

and

$$\begin{aligned} w_{ii,2} &= w_{ii,2} \cdot \frac{\beta_0}{n_P(i)} \sum_{k \in N_P(i)} \frac{1\{i \in \bar{N}_I^i(k)\}}{\bar{n}_P(k)} \\ &= w_{ii,2} \cdot \frac{\beta_0}{n_P(i)} \sum_{k \in N_P(i)} \frac{1\{i \in N_P(k)\}}{\bar{n}_P(k)} = \beta_0 w_{ii,2} c_{ii}. \end{aligned}$$

The last equation gives $w_{ii,2} = 0$, because $|\beta_0 c_{ii}| < 1$, and the first two equations give

$$w_{ii,1} = \frac{\gamma_0}{1 - \beta_0 c_{ii}} \tag{A.19}$$

and

$$w_{ii,\varepsilon} = \frac{1}{1 - \beta_0 c_{ii}}. \tag{A.20}$$

Also, we turn to $\mathscr{M} w_j^i$:

$$\mathscr{M}_i w_j^i = \frac{1}{n_P(i)} \sum_{k \in N_P(i)} w_{ij} 1\{j \in N_P(k)\} + \frac{r_{ij} w_{jj}^i 1\{j \in N_P(i)\}}{n_P(i)}, \tag{A.21}$$

where the last term corresponds to the case $j = k \in N_P(i)$. Using the definition c_{ij}, we rewrite

$$\mathscr{M}_i w_{j,1}^i = c_{ij} w_{ij,1} + \frac{w_{ii,1} r_{ij} 1\{j \in N_P(i)\}}{n_P(i)} = c_{ij} w_{ij,1} + \frac{\gamma_0}{1 - \beta_0 c_{ii}} \frac{1}{n_P(i)} r_{ij} 1\{j \in N_P(i)\}$$

and

$$\mathscr{M}_i w_{j,\varepsilon}^i = c_{ij} w_{ij,\varepsilon} + \frac{w_{ii,\varepsilon} r_{ij} 1\{j \in N_P(i)\}}{n_P(i)} = c_{ij} w_{ij,\varepsilon} + \frac{1}{1 - \beta_0 c_{ii}} \frac{1}{n_P(i)} r_{ij} 1\{j \in N_P(i)\}.$$

We plug this into (A.15) to obtain

$$w_{ij,1} = \frac{\beta_0 \gamma_0 r_{ij} 1\{j \in N_P(i)\}}{n_P(i)(1 - \beta_0 c_{ij})(1 - \beta_0 c_{ii})},$$

and

$$w_{ij,\varepsilon} = \frac{\beta_0 r_{ij} 1\{j \in N_P(i)\}}{n_P(i)(1-\beta_0 c_{ij})(1-\beta_0 c_{ii})}.$$

Finally, let us consider $w_{ij,2}$. Note that from (A.21),

$$\mathscr{M}_i w^i_{j,2} = c_{ij} w_{ij,2},$$

because $w_{ii,2} = 0$. By plugging this into (A.15), we obtain that

$$\begin{aligned} w_{ij,2} &= \begin{cases} \delta_0 r_{ij}/n_P(i) + \beta_0 c_{ij} w_{ij,2}, & \text{if } j \in N_P(i), \\ \beta_0 c_{ij} w_{ij,2}, & \text{if } j \in N_P^2(i) \backslash N_P(i), \end{cases} \\ &= \begin{cases} \delta_0 r_{ij}/n_P(i) + \beta_0 c_{ij} w_{ij,2}, & \text{if } j \in N_P(i), \\ 0, & \text{if } j \in N_P^2(i) \backslash N_P(i), \end{cases} \end{aligned}$$

where the last zero follows from the equality $w_{ij,2} = \beta_0 c_{ij} w_{ij,2}$ with $|\beta_0 c_{ij}| < 1$. Therefore, we have

$$w_{ij,2} = \frac{\delta_0 r_{ij} 1\{j \in N_P(i)\}}{n_P(i)(1-\beta_0 c_{ij})}.$$

From the form of a linear strategy for $s_i^{\mathsf{BR}}(\mathscr{I}_i)$ with the weights as solved thus far, we obtain the desired result.■

Proof of Theorem 2.2.2:

Suppose each agent is first-order sophisticated (FS) type; i.e., each $i \in N$ believes that each $k \neq i$ is simple type and chooses strategies according to:

$$s^i_k = \sum_{j \in \overline{N}_P(k)} T'_j w^i_{kj} + \eta_k.$$

The best responses of FS types are linear because the utility is quadratic in the player's own actions, and they believe simple types play according to linear strategies, with i's best response taking the form

$$s_i^{\mathsf{BR.FS}}(\mathscr{I}_{i,1}) = X'_{i,1}\gamma_0 + \overline{X}'_{i,2}\delta_0 + \beta_0\left(\frac{1}{n_P(i)}\sum_{k\in N_P(i)} s_k^i\right) + \eta_i + \varepsilon_i.$$

Since an agent i with FS-type believes that all other agents are of simple type, we have that w_{kj}^i equal the weights given by player k to player j according to the best response strategies in Theorem 2.2.1. Using $\lambda_{ij} = 1/(1-\beta_0 c_{ij})$, together with the Theorem 2.2.1 weights, in equations (A.14) - (A.15), we obtain the FS weights as follows:

$$\begin{aligned}
w_{ii,1} &= \gamma_0 + \frac{\beta_0}{n_P(i)}\sum_{k\in N_P(i)}\frac{\beta_0\gamma_0\lambda_{ki}1\{i\in N_P(k)\}}{n_P(k)(1-\beta_0 c_{kk})},\\
w_{ii,\varepsilon} &= 1 + \frac{\beta_0}{n_P(i)}\sum_{k\in N_P(i)}\frac{\beta_0\lambda_{ki}1\{i\in N_P(k)\}}{n_P(k)(1-\beta_0 c_{kk})},\\
w_{ii,2} &= \frac{\beta_0}{n_P(i)}\sum_{k\in N_P(i)}\frac{\delta_0\lambda_{ki}1\{i\in N_P(k)\}}{n_P(k)},
\end{aligned}$$

and for each $j \in N_{P,2}(i)$,

$$\begin{aligned}
w_{ij,1} &= \frac{\beta_0}{n_P(i)}\sum_{k\in N_P(i)} w_{kj,1}^i 1\{j\in \overline{N}_P(k)\}\\
&= \frac{\beta_0}{n_P(i)}\left(\sum_{k\in N_P(i)} w_{kj,1}^i 1\{j\in N_P(k)\} + w_{jj,1}^i 1\{j=k\}\right)\\
&= \frac{\beta_0}{n_P(i)}\sum_{k\in N_P(i)}\frac{\beta_0\gamma_0\lambda_{kj}1\{j\in N_P(k)\}}{n_P(k)(1-\beta_0 c_{kk})} + \frac{\beta_0\gamma_0 1\{j\in N_P(i)\}}{n_P(i)(1-\beta_0 c_{jj})}.
\end{aligned}$$

Analogously,

$$w_{ij,\varepsilon} = \frac{\beta_0 1\{j\in N_P(i)\}}{n_P(i)(1-\beta_0 c_{jj})} + \frac{\beta_0}{n_P(i)}\sum_{k\in N_P(i)}\frac{\beta_0\lambda_{kj}1\{j\in N_P(k)\}}{n_P(k)(1-\beta_0 c_{kk})},$$

and as for $w_{ij,2}$, if $j \in N_P(i)$,

$$w_{ij,2} = \delta_0/n_P(i) + \frac{\beta_0}{n_P(i)} \sum_{k \in N_P(i)} \frac{\delta_0 \lambda_{kj} 1\{j \in N_P(k)\}}{n_P(k)},$$

and if $j \in N_{P,2}(i) \backslash N_P(i)$,

$$w_{ij,2} = \frac{\beta_0}{n_P(i)} \sum_{k \in N_P(i)} \frac{\delta_0 \lambda_{kj} 1\{j \in N_P(k)\}}{n_P(k)}.$$

Next, using the definitions of $\bar{\lambda}_{ij}$ and $\tilde{\lambda}_{ij}$ in the theorem, we write

$$\begin{aligned} w_{ii,1} &= \gamma_0 + \beta_0^2 \gamma_0 \tilde{\lambda}_{ii}, \\ w_{ii,\varepsilon} &= 1 + \beta_0^2 \tilde{\lambda}_{ii}, \end{aligned}$$

and $w_{ii,2} = \beta_0 \delta_0 \bar{\lambda}_{ii}$. Lastly, for each $j \in N_{P,2}(i)$, we have

$$\begin{aligned} w_{ij,1} &= \frac{\beta_0 \gamma_0 \lambda_{jj} 1\{j \in N_P(i)\}}{n_P(i)} + \beta_0^2 \gamma_0 \tilde{\lambda}_{ij}, \\ w_{ij,\varepsilon} &= \frac{\beta_0 \lambda_{jj} 1\{j \in N_P(i)\}}{n_P(i)} + \beta_0^2 \tilde{\lambda}_{ij}, \end{aligned}$$

and

$$w_{ij,2} = \begin{cases} \beta_0 \delta_0 \bar{\lambda}_{ij}, & j \in N_{P,2}(i) \backslash N_P(i) \\ \delta_0/n_P(i) + \beta_0 \delta_0 \bar{\lambda}_{ij}, & j \in N_P(i). \end{cases}$$

Substituting these weights back into the best response function for FS types, we obtain the desired result. ■