Université de Montréal

Analyse en identification partielle de la décision d'émigrer des étudiants africains

par

Romuald Méango

Département de sciences économiques
Faculté des Arts et des Sciences

Thèse présentée à la Faculté des Arts et des Sciences
en vue de l'obtention du grade de Philosophiae Doctor (Ph.D.)
en sciences économiques,

Mai 2013

Université de Montréal
Faculté des Arts et des Sciences

Cette thèse intitulée :

Analyse en identification partielle de la décision d'émigrer des étudiants africains

par

Romuald Méango

a été évaluée par un jury composé des personnes suivantes :

Président-rapporteur : .....
Directeur de recherche : .....
Examinateur externe : .....
Membre du jury : .....
Représentant du Doyen : .....

Résumé

La migration internationale d'étudiants est un investissement couteux pour les familles dans plusieurs pays en voie de développement. Cependant, cet investissement est susceptible de générer des bénéfices financiers et sociaux relativement importants aux investisseurs, tout autant que des externalités pour d'autres membres de la famille. Cette thèse s'intéresse à deux aspects importants de la migration des étudiants internationaux : (i) Qui part ? Quels sont les déterminants de la probabilité de migration ? (ii) Qui paie ? Comment la famille s'organise-t-elle pour couvrir les frais de la migration ?

Entreprendre une telle étude met le chercheur en face de défis importants, notamment, l'absence de données complètes et fiables ; la dispersion géographique des étudiants migrants en étant la cause première. La première contribution importante de ce travail est le développement d'une méthode de sondage en « boule de neige » pour des populations difficiles à atteindre, ainsi que d'estimateurs corrigeant les possibles biais de sélection. A partir de cette méthodologie, nous avons collectées des données incluant simultanément des étudiants migrants et non-migrants du Cameroun en utilisant une plateforme internet. Un second défi relativement bien documenté est l'endogénéité de la variable mesurant le niveau final d'éducation de l'étudiant. Nous tirons avantage des récents développements théoriques dans le traitement des problèmes d'identification dans les modèles de choix discrets avec variables instrumentales pour résoudre cette difficulté, tout en conservant la simplicité des hypothèses nécessaires. Ce travail constitue l'une des premières applications de cette méthodologie à des

questions de migration.

Le premier chapitre de la thèse étudie la décision prise par la famille d'investir dans la migration étudiante. Il propose un modèle structurel empirique de choix discret qui reflète à la fois le rendement brut de la migration et la contrainte budgétaire liée au problème de choix des agents. Nos résultats démontrent que le choix du niveau final d'éducation, les résultats académiques et l'aide de la famille sont des déterminants importants de la probabilité d'émigrer, au contraire du sexe de l'étudiant qui ne semble pas affecter très significativement la décision familiale.

Le second chapitre s'efforce de comprendre comment les agents décident de leur participation à la décision de migration et comment la famille partage les profits et décourage le phénomène de « passagers clandestins ». D'autres résultats dans la littérature sur l'identification partielle nous permettent de considérer des comportements stratégiques au sein de l'unité familiale. Les résultats empiriques suggèrent que le modèle « unitaire », où un agent représentatif maximise l'utilité familiale, est incompatible avec nos données. Les aidants extérieurs à la famille nucléaire subissent un coût strictement positif pour leur participation, ce qui décourage leur implication. Les obligations familiales et sociales semblent expliquer les cas de participation d'un aidant, mieux qu'un possible altruisme de ce dernier.

Finalement, le troisième chapitre présente le cadre théorique plus général dans lequel s'imbriquent les modèles développés dans les précédents chapitres. Les méthodes d'identification et d'inférence présentées sont spécialisées aux jeux finis avec information complète. Avec nos co-auteurs, nous proposons notamment une procédure combinatoire pour une implémentation efficace du bootstrap pour effectuer l'inférence dans les modèles cités ci-dessus. Nous en faisons une application sur les déterminants du choix familial de soins à long terme pour des parents âgés.

Mots clés : Mobilité étudiante, Technique d'échantillonage de réseau, Modèles structurels incomplets, Identification partielle.

Summary

International migration of students is a costly investment for family units in many developing countries. However, it might yield substantial financial and social return for the investors, as well as externalities for other family members. This thesis addresses primarily two aspects of international student migration : (i) Who goes ? What are the determinants of the probability of migration ? (ii) Who pays ? How does the family organize to bear the cost of the migration ?

Engaging in this study, one faces the challenge of data limitation, a direct consequence of the geographical dispersion of the population of interest. The first important contribution of this work is to provide a new chain-referral sampling methodology for hard-to-reach populations, along with estimators to correct for selection biases. We collected data which include both migrant and non-migrant students from Cameroon, using an online platform. A second challenge is the well-documented problem of endogeneity of the educational attainment. We take advantage of recent advances in the treatment of identification problems in instrumental variable discrete choice models to solve this issue while keeping assumptions at a low level. In particular, validity of the partial identification methodology does not rest on the existence of an instrument. To the best of my knowledge, this is the first empirical application of this methodology to migration related issues.

The first chapter studies the decision made by a family to invest in student migration. It proposes an empirical structural decision model which reflects the importance of both the return on the investment and the budget constraint in agent choices.

Our results show that the choice of level of education, the past academic results in secondary school are significant determinants of the probability to migrate, unlike the gender which does not seem to play any role in the family decision.

The objective of the second chapter is to understand how agents decide to be part of the migration project and how the family organizes itself to share profits and discourage free riding-behavior. Further results on partial identification for games of incomplete information, allow us to consider strategic behavior of family. The results suggest that models with a representative individual are not compatible with our data. Helpers (significant extended family members) incur a non-zero cost of participation that discourages involvement in the migration process. Kinship obligation and obligation for male members to participate, and not altruism, appear as the main reason of their participation.

Finally, the third chapter presents the more general theoretical framework in which the preceding models are embedded. The method presented is specialized to finite games of complete information, but is of interest for application to the empirical analysis of instrumental variable models of discrete choice (Chapter 1), cooperative and non-cooperative games (Chapter 2), as well as revealed preference analysis. With our co-authors, we propose an efficient combinatorial bootstrap procedure for inference in games of complete information that runs in linear computing time and an application to the determinants of long term elderly care choices.

Keywords : Student mobility, Network sampling, Incomplete structural models, partial identification.

Table des matières

Table des figures

Liste des tableaux

Remerciements

Je tiens à remercier mon directeur de recherche, Marc Henry, qui est allé bien au-delà du simple devoir qui accompagne cette responsabilité. Je suis honoré d'avoir pu évoluer à ses côtés.

Je remercie les laboratoires de recherche et les institutions qui m'ont accueillie et financée : le Centre Interuniversitaire de Recherche en Economie Quantitative (CIREQ) et le département d'économie de l'Université de Montréal, tout spécialement son corps professoral, le "Centre for Microdata Methods And Practice" (CeMMAP). Cette thèse a également bénéficiée du financement de la Bourse de fins d'études doctorales de la Faculté des Études Supérieures et Postdoctorales (FESP) de l'Université de Montréal, ainsi que de l'assistance technique de Daniel Stubbs et Joël Cornuz.

Je suis également redevable aux participants des différentes conférences, séminaires, groupes de travail, où ces travaux ont été présentés, pour leurs discussions et commentaires.

Enfin, mes remerciements vont à l'endroit de mon épouse, de ma famille et de mes amis pour leur soutien et leurs prières.

Introduction générale

La croissance importante du nombre d'étudiants internationaux dans les dernières décennies a suscité un intérêt particulier de la littérature en sciences économique pour cette forme de migration. Deux caractéristiques importantes justifie cette spécialisation. En premier, la migration étudiante peut être considérée comme une voie pour une migration de plus long-terme. Comme mis en évidence par Rosenzweig (2008), à la différence de plusieurs types de visas, le visa étudiant ne fait pas l'objet de restrictions de quotas. De plus, la probabilité pour un étudiant ayant complété ses études dans un pays tiers de rester dans son pays hôte domine largement la probabilité d'émigration de celle d'un étudiant formé uniquement dans son pays d'origine. Finalement, la différence de qualité entre universités, ainsi que la différence de rémunération pour un niveau de qualification donné entre les pays à revenu par habitants élevés et travilleurs de pays en développement, sont autant de facteurs qui attirent les migrants. Entrer dans un pays hôte avec un statut étudiant apparaît naturellementcomme une stratégie dominante pour une migration éventuelle.
En second, du fait du coût élevé de la vie et des frais de scolarité, la contrainte budgetaire joue un rôle clé. Les ressources de toute la famille (et pas seulement celle du candidat à la migration) sont à considérer. Selon des données colloectées en 2006 par l'international de l'éducation, le financement des études à l'étranger provient premièrement de fonds « personnelles et familiaux » pour 64% des étudiants internationaux.

Cependant, surtout du fait de l'absence d'études quantitatives sur les ménages, la

littérature s'est concentrée sur une analyse macroéconomique des déterminants de la mobilité étudiante. L'article de Batista, Lacuesta, and Vicente (2012) constitue une exception notable. Les récentes contributions de Rosenzweig, Irwin, and Williamson (2006), Rosenzweig (2008), Beine, Noel, and Ragot (2012) et Perkins and Neumayer (2011) parmi d'autres étendent les résultats de la littérature plus établie sur la migration internationale des individus hautement qualifiés. Ils étudient notamment les déterminants dans le pays d'origine et/ou de migration qui encouragent, attirent ou retiennent les étudiants internationaux. ces déterminants sont connus sous le nom de « push-factors » et « pull-factors ». L'aspect ayant reçu le moins d'attention jusque là est la perspective microéconomique de cette décision d'émigration ; des questions importantes telles que celle de l'influence du choix de niveau d'éducation sur la probabilité de migration, des caractéristiques propres à l'étudiant et à sa famille qui augmentent ou détériorent ses chances de migration, de l'organisation de l'unité familiale pour couvrir les frais des études à l'étranger.

Cette thèse s'intéresse à deux aspects importants de la migration des étudiants internationaux : (i) Qui part ? Quels sont les déterminants de la probabilité de migration ? (ii) Qui paie ? Comment la famille s'organise-t-elle pour couvrir les frais de la migration ?

Comme mentionné plus haut, la famille nucléaire est grandement impliquée dans l'investissement pour l'émigration d'un étudiant et ses caractéristiques doivent être prises en compte dans notre analyse. De plus, comme documenté dans la littérature sur la migration étudiante (Perkins and Neumayer (2011)), les migrants reçoivent un soutien non-négligeable de la diaspora dans le pays de destination ; par exemple, un membre de la famille éloigné pourvoie au logement de l'étudiant migrant en l'accueillant sous son toît. Cet effet, connu dans la littérature comme un effet de réseau, procure à la famille un capital implicite, ce qui influe sur la contrainte budgétaire. Les caractéristiques de cet aidant devraient également faire partie de notre analyse.

Entreprendre une telle étude met le chercheur en face de défis importants, no-

tamment, l'absence de données complètes et fiables ; la dispersion géographique des étudiants migrants en étant la cause première. La première contribution importante de ce travail est le développement d'une méthode de sondage en « boule de neige » pour des populations difficiles à atteindre, ainsi que d'estimateurs corrigeant les possibles biais de sélection,. Elle bâtit sur les contributions de Heckathorn (1997) et Thompson (2006). A partir de cette méthodologie, nous avons collectées des données incluant simultanément des étudiants migrants et non-migrants du Cameroun en utilisant une plateforme internet. Cette méthodologie est developpée en Annexe du chapitre 1.

Dans le premier chapitre, nous nous intéressons à la question des déterminants de la migration étudiante, particulièrement, les caractéristiques de la famille et du candidat à la migration. Nous développons un modèle structurel de choix discrets, dans lequel la famille (au sens large) décide d'un investissement dans la migration étudiante. En accord avec la littérature sur la formation du capital humain, nous supposons que le choix d'éducation est un investissement de la famille pour maximiser le revenu pendant la durée de vie. La famille décide du pays dans lequel se feront les études supérieures de l'étudiant ; ce choix est toutefois contingent à celui du niveau d'éducation final. La migration a lieu lorsque les revenus espérés sont positifs. Des modèles similaires sont développés par Brezis and Soueri (2011) et Beine, Noel, and Ragot (2012), à la différence notable que nous introduisons une contrainte budgétaire liée au capital disponible à la famille. Une difficulté inhérente à notre modèle et relativement bien documenté (voir la discussion de Batista, Lacuesta, and Vicente (2012)) est l'endogénéité du choix d'éducation, due notamment à la simultanéité des décisions de migration et d'éducation. Nous tirons avantage des récents développements théoriques dans le traitement des problèmes d'identification dans les modèles de choix discrets avec variables instrumentales pour résoudre cette difficulté, tout en conservant la simplicité des hypothèses nécessaires. Ce travail constitue l'une des premières applications de cette méthodologie à des questions de migration. Nos résultats démontrent que le choix du niveau final d'éducation, les résultats académiques et le soutien financier familial sont des déterminants importants de la

probabilité d'émigrer, au contraire du sexe de l'enfant qui ne semble pas affecter très significativement la décision familiale. La contribution de l'aidant semble se limiter à une contribution au capital de la famille.

Le second chapitre s'efforce de comprendre comment les agents décident de leur participation à la décision de migration et comment la famille partage les profits et décourage le phénomène de « passagers clandestins ». En effet, couvrir les frais attenants à la migration étudiante (frais de voyage, coût de la vie, frais de scolarité) dans un pays qui a potentiellement un niveau de vie plus élevé que le pays d'origine, est un investissement très important. Deux explications pourraient être fournies à la participation des membres de la famille : soit cet investissement produit pour les investisseurs des bénéfices financiers et sociaux, soit, ces investisseurs sont altruistes et tirent un certain profit de l'utilité de l'enfant. L'objectif de ce chapitre est précisement de distinguer entre les motivations des agents qui sont le plus impliqués dans le financement de la migration étudiante. Le modèle structurel proposé décrit un jeu non-coopératif de participation entre les membres de la famille élargie. La littérature sur les comportements stratégiques dans d'autres contextes de décision familiale est riche de modèles similaires Manser and Brown (1980), McElroy and Horney (1981), Lundberg and Pollak (1994), Engers and Stern (2002). A l'intérieur de l'unité familiale, les parents et l'aidant sont considérés comme des investisseurs potentiels pour la migration étudiante. Cependant, chacun d'eux à des incitations à ne pas participer à la décision puisque l'investissement produit des externalités pour toute la famille. Les joueurs décident de leur participation en maximisant leur profit à l'équilibre, l'équilibre de Nash étant ici considéré. Etant donné les profits de chaque joueur, notre modèle prédit des équilibres de Nash, possiblement multiples. C'est la présence de ces équilibres multiples qui justifie l'utilisation d'une procédure en identification partielle. Nous faisons appel à des résultats dans la littérature sur l'identification partielle qui nous permettent de considérer des comportements stratégiques au sein de l'unité familiale (Beresteanu, Molchanov, and Molinari (2011), Galichon and Henry (2011)).

Les résultats empiriques suggèrent que le modèle « unitaire », où un agent représentatif maximise l'utilité familiale, est incompatible avec nos données. Les aidants extérieurs à la famille nucléaire subissent un coût strictement positif pour leur participation, ce qui décourage leur implication. Les obligations familiales et sociales semblent expliquer les cas de participation d'un aidant, mieux qu'un possible altruisme de ces derniers.

Finalement, le troisième chapitre présente le cadre théorique plus général dans lequel s'imbriquent les modèles développés dans les précédents chapitres. Les méthodes d'identification et d'inférence présentées sont spécialisées aux jeux finis avec information complète. Beresteanu, Molchanov, and Molinari (2011) et Galichon and Henry (2011) montrent que de tels modèles sont équivalents à une collection d'inégalités de moment dont le cardinal augmentent exponentiellement avec le nombre d'alternatives discrètes. Avec nos co-auteurs, nous proposons une caractérisation équivalente, basée sur des méthodes d'analyse combinatoires classiques, qui réduit les contraintes computationelles. Nous proposons notamment une procédure combinatoire pour une implémentation efficace du bootstrap pour effectuer l'inférence statistique dans les modèles cités ci-dessus. Nous en faisons une application sur les déterminants du choix familial de soins à long terme pour des parents âgés.

Chapter 1

The determinants of International Student Migration: A partial identification analysis

Abstract

This paper studies the decision made by a family to invest in student migration and the consequences of the migration of skilled individuals for the sending country. We propose an empirical structural decision model which reflects the importance of both the return of the investment and the budget constraint in agent choices. Taking advantage of recent advances in the treatment of identification problems in IV-discrete choice models, we circumvent the problem of endogeneity of the educational attainment and conduct inference for the parameters of interest. The data are collected on students from Cameroon, using a new snowball sampling procedure, which allow the inclusion of both migrants and non-migrants in the sample. We propose bias corrected estimators for this procedure. We study the characteristics of potential candidates to migration that increase or decrease their probability to migrate. We also test for existence of a brain gain effect in Cameroon. The results show that educational attainment in tertiary education is not positively correlated with the prospect of migration, invalidating the brain gain hypothesis.

1.1 Introduction

As the number of students choosing to study abroad has considerably increased since 1970's, the interest of the economic literature in this specific form of migration has also grown in the recent years. There is at least to features that make the specialization to this topic relevant. First, student migration may be seen as a route to permanent emigration. As pointed out by Rosenzweig (2008), unlike for many visas, there are no country ceilings or kinship requirements for student visas. In addition, the probability that a foreign-trained student will remain in the host country is higher than the overall emigration probability for a domestically schooled student. Finally, universities are often of better quality and for comparable skills levels, remuneration are higher in high-income countries. Entering the host country as a student appears therefore as a dominant strategy for prospecting migrants. Second, due to higher costs of living and potentially higher fees in foreign countries, the budget constraint plays an important role. The family (and not only the student) resources matter. According to data released by the Institute of International Education (IE) in 2006, the primary source of funding is "personal and family" for about 64 percent of foreign students.

However, mainly because of data constraints, the focus of the literature has been largely directed to macroeconomic analyzes of determinants of student mobility, Batista, Lacuesta, and Vicente (2012) (BLV, hereafter) being one notable exception. The recent contributions from Rosenzweig, Irwin, and Williamson (2006), Rosenzweig (2008), Beine, Noel, and Ragot (2012) and Perkins and Neumayer (2011) among others build on the findings of the more established literature on international migration to study determinants in origin and/or destination countries that encourage, attract or retain student migrants, the so-called push and pull factors. What so far received less attention are the determinants of the choice of location of tertiary studies from a microeconomic perspective. Most importantly, how influential are the choice of final level of education, the characteristic of the candidate to migration, as well as family's socio-economic characteristics?

As stated earlier, the nuclear family is highly involved in migration choices and there characteristics should be included in any analysis of the decision process. Moreover, as documented by survey results and the literature of student migration (see for example Perkins and Neumayer (2011)), migrants might receive some support from the diaspora in their destination country, e.g a distant relative of the student's family living in the destination country provides some help by hosting the migrant, providing food and accomodation. This effect, known in the migration literature as a network effect, provides some implicit capital to the family. The characteristics of this helper should therefore be relevant determinants within our analysis.

But engaging in this study, one faces major challenges, the main of them being the data limitations. Most existing empirical studies are based on country-level samples which fail to reflect the heterogeneity within a specific community. Conversely, as the population of interest is often scattered around the globe, micro-level samples are easily in danger of missing part of the population of interest. Furthermore, the pervasive problem of endogeneity renders the econometric analysis difficult. In that strong structral assumption are required for identification and estimation of the structural parameters, the robustness of the methods in use so far (in BLV for example) is debatable. These challenges explain the scarcity of the literature on the microeconomic determinants of international student migration.

The objective of this paper is to provide a rigorous framework to study the investment decision made by the family of the candidate to emigration, while addressing the above concerns. Throughout the paper, we work with a structural discrete choice model of private investment of the family in the student migration. Reminescent of the human capital litterature, our framework assume education of a child to be an investment of the family unit which seeks to maximize it lifetime earning. The family decides of the location of tertiary study of the sudent; however, this choice is contingent on the choice of final education level (hence the problem of endogeneity). Migration occurs when the expected return of the investment is positive. Our model

disentangles the liquidity constraints effects from the expected return of the investment. We are interested in the determinants of the return component for the family, as well as the deterring effect of the financial constraint component. The novelty of our approach is that, albeit a monotonicity condition, we do not impose any restriction on the relationship of the educational attainment and migration, neither do we rely on a identifying instrument.

The contributions to the literature on the topic are three-fold: first, we take advantage of recent developments in network sampling to construct a novel dataset on a population of students from Cameroon, migrants and non-migrants. Second, we show that our structural model of investment in student migration offers itself quite nicely to an inference procedure in an incomplete (partially identified) model framework. This has two major advantages: frist, the structural assumptions are fairly parsimonious because mainly driven by the economic analysis and not impose to solve technical shortcomings. Furthermore, we do not appeal to the use of a identifying instrument, hence, circumventing the problems of validity and relevance. Following the proposal in Chesher, Rosen, and Smolinski (2011), we use sharp bounds for our discrete choice models, with endogenous right-hand side regressor. These bounds translate in moment inequalities and a large literature has developed in this field. A challenge though remains the computational burden of these methods, which is often increased when considering models with relatively high number of covariates. Henry, Méango, and Queyranne (2011) (Chapter 3, HMQ hereafter) proposes a combinatorial approach to solve the above problem and their method is the best suited to inference in our framework. As a final contribution of this paper, we propose a methodology to preestimate a number of parameters of interest, decreasing significantly the computational requirement. This step appears crucial, since the sample size required to achieve informative inference, grows rapidly with the number of covariates.

For the purpose of the study, a survey has been conducted on the population of Cameroonian, aged 18 or more, having completed secondary school by obtaining

the "Baccalauréat"[1]. We gathered information on 418 respondents. Our survey data show that close to three quarters of migrant rely on themselves or on family capital to finance their study abroad. More than half of the respondants reports the existence of a potential or effective helper in the migration process. The "typical" helper is a male, an uncle or a brother who have a university degree. Unsurprinsingly, families where at least one of the parent have higher level of education, and families who possess higher physical capital, have less difficulty to meet the budget constraint for migration. Concerning determinants of the return of investment in migration, we find that a higher choice of educational attainment along with better candidate's results during secondary school significantly increase returns to migration. This finding suggests a positive selection of migrants. Interestingly, first born-child has lower probability of migration than the subsequent children, while the gender does not seem to affect the probability of migration.

This paper is mainly related to three strands of literature, the Network Sampling literature, the student-mobility and the more general migration literature and the partial identification literature. With the recent expansion of social networking services and advances in computational capabilities, a renewed interest has grown for methods of data collection over networks. The procedure we propose has been inspired by the Respondant-Driven Sampling methodology proposed by Heckathorn (1997, 2007) and Wejnert and Heckathorn (2008). To estimate inclusion probabilities and correct for oversampling of some population members, we use estimators first applied by Thompson (2006) in the context of network sampled data.

The incentives for international migration of skilled individuals have been extensively studied by a number of early key contributions for which the surveys by Borjas(1989, 1994) serve as good references. Rosenzweig (2008) studies out-migration of (Asian) students. One of his contributions is to distinguish the effects of the return

[1]Similar to the french educational system, "Baccalauréat" is a compulsory state exam for completion of secondary school.

of out-migration and budgetary constraint on a candidate decision. Both appear important in the final decision of migration. Beine, Noel, and Ragot (2012) and Perkins and Neumayer (2011) study the determinants of migration in multi-origin multi-destination framework. Among interesting results, they find a strong effect of the number of migrants in the destination country on the probability of migration.

Micro-level studies of the question of the incentives of international student migration are still rare. BLV studies the case of Cape Verde by taking advantage of a specifically tailored survey on households in the country. They are interested in testing the brain gain effect, which in their framework, amounts to testing for a significant linear correlation between the own future probability of migration and the schooling decisions. To achieve identification, their econometric model relies on a distributional assumption on the joint behavior of the latent variable and on some exclusion restrictions. Additional, though not exhaustive, reference on international student migration includes Dreher and Poutvaara (2011), Bessey (2011), Brezis and Soueri (2011), Thissen and Ederveen (2006), Van Bouwel and Veugelers (2009).

Our work is also related to an increasing literature on partially identified models, following the seminal works of Manski (1993) and Jovanovic (1989). Closely related to our framework, the papers of Beresteanu, Molchanov, and Molinari (2011), Galichon and Henry (2011) and Chesher, Rosen, and Smolinski (2011), which relax the requirement for point identification in structural models and derive sharp bounds on the parameters of these models. The first two explore the case of strategic games in complete information, while the last one is concerned with instrumental variable model of discrete choice. As the bounds translate into moment inequalities, our inference procedure is related to a large literature which has developed on inference in moment inequality models since the seminal contribution of Chernozhukov, Hong, and Tamer (2007). A major challenge is the computational burden of these methods, here aggravated by the relatively high number of covariates.

In Section 1.2, we present a family investment decision model in a human capital

framework. Section 1.3 is devoted to the inference procedure. The data collection procedure is presented in Section 1.4. Finally, we gather the results of the inference on the parameters of the structural model in Section 1.5, before we conclude. Proofs are collected in the appendix.

1.2 A structural Model of Private Investment in Student Migration

Student migration displays two important characteristics. First, the choice of location of tertiary education is the result of an arbitrage between the schooling and employment opportunities available in the origin country and in the host country. Second, the amount of capital that the *family* can invest in the process is key to the migration process. The model we present mirrors these two essential characteristics.

Following Beine, Noel, and Ragot (2012) (see also Brezis and Soueri (2011)), we consider a framework based on the human capital literature, where "Education is considered as an investment in future earnings and employment for rationale [families] who seek to maximize the lifetime earnings." The decision of a student migration is then between (1) obtaining further education in a foreign country or (2) studying or starting professional activities at home. Of course, this decision does not preclude further international mobility when education is completed, although the above choice affects significantly the probability of later migration - student migrants have better opportunities on the labor market of their host country (some evidences in Rosenzweig (2008)). Of equal importance, student migrants might return to their origin country to work. Our explicit assumption is that the choice of location of tertiary education and the choice of location of work are taken sequentially. Note that the benefits discussed here are not only pecuniary. For example, the utility enjoyed by having a child entering the marriage market of the host country, or the disutility of having a child living to a greater distance (e.g for provision of health care to elderly parents).

To be more specific, the evaluation of benefits depend on two main components: one that depends on the characteristics of the child. Of particular interest for us, his/her final educational attainment. The unobserved innate ability of the child also influences how likely he is to adjust to life in a foreign culture or how much he prefers staying in his country[2].

The other important part, on which these benefits depend on, is the wealth of the family. Indeed, obtaining tertiary education in a foreign country which has often higher income per-capita entails significant costs. Education costs, travel costs and living costs. The family could possess some implicit capital that will lower the financial input required for the student migration investment. In particular, if a relative lives in the destination country. This relative could provide material and financial support. Family characteristics will therefore influence the amount of money that the familly needs to borrow for the student migration investment, if any. The availability of a scholarship for the student might also significantly relax the budget constraint of the family.

The families form myopic expectations on the return on studying abroad or staying home, observing skill prices and probabilities of later settlement in a foreign country. The family compares returns on both alternatives, and invest in migration when the migration option yields the largest expected benefits. We introduce more formal notation in the following.

Consider a family indexed by i. Here we think of an extended family (parents, child and some relative of the family who might provide some support) with one member (a child) who is a potential migrant. The family possesses a capital K_i^0 (explicit or implicit, as discussed above). The child has observable characteristics X_i (e.g. education, gender and previous academic result of the child) and characteristics ε_i, known to all family members but unobservable for the econometrician (innate ability of the child, cultural factors influencing evaluation of the migration alternative,etc.). When

[2]This component could be extend to specific preferences of the family.

the child complete secondary school, the family has the opportunity of making an investment of level I, known and fixed across families, by financing further education of the child in a foreign country. The reader should understand I as the minimum financial input enabling student migration to the country of destination. It includes travel, education and living expenses. We call $I_f(Ed_i, X_i, \varepsilon_i, \theta)$ the *g*ross revenue from the investment and $r(Ed_i, X_i, \varepsilon_i, \theta)$, its return, which both depend on the observable and unobserved characteristics of the family. Note that the two variables are linked by the following equation :

$$r(Ed_i, X_i, \varepsilon_i, \theta) = \frac{I_f(Ed_i, X_i, \varepsilon_i, \theta)}{I} - 1$$

θ are here the parameters of interest. Ed_i is the final education level. The choice of education will affect the salary the agent expect after graduation but also the probability of subsequent migration[3]. The endogeneity problem will arise primarily from the potential correlation between the unobservable variable ε_i and the choice of final educational attainment. If $K_i^0 < I$, the family must borrow capital to be able to make the investment. Denote r_0 the interest rate for borrowing. In the following, we will reason in term of return or interest rate.

As mentioned above, the student has the alternative to remain in his origin country, either to obtain further education or to work. Again, later on, the individual decides to migrate or not and his/her prospect of migration enters the valuation of this given alternative. We will assume that this alternative, along with other potential investment alternatives in the origin country, yield an interest rate on the family capital K_i^0, that we will denote r_1. By assuming r_1 to be constant, we implicitly assume that the return on the investment of the family capital in the origin country, is not (or only mildly) affected by the choice of final education. This would be the case if the return to skills is low in Cameroon and the probability of work migration

[3]Dreher and Poutvaara (2011) points out that " host countries are interested in educating foreign students, partly to attract human capital benefiting the domestic economy". Educational attainment and probability of settlement in the host country increase together for a migrant student.

when schooling has been completed in the origin country is low and not very sensitive to the choice of education. As this seems to be case in the country where we conduct our empirical application, for the level of education we consider (Masters) , we do not expect our result to be very sensitive to this hypothesis (see IOM report, 2010). As the reader might expect, education in many developing country is pretty cheap. However, it yields low expected returns, we impose $r_1 \leq r_0$[4].

The family compares investments with respect to their expected return. Knowing the interest rates above, the family chooses the investment alternative to maximize its expected profit that we will denote $\Pi_i \equiv \Pi_i \left(Ed_i, X_i, K_i^0, \varepsilon_i; r_0, \theta\right)$ (to simplify notation, we drop the subscript i in the subsequent development). Choosing to remain in the origin country yields an expected profit of :

$$\Pi = r_1 K^0$$

While the investment student in migration, along with a choice of final education level of Ed gives:

$$\Pi = r\left(Ed, X, \varepsilon; \theta\right) I + \left(K^0 - I\right)\left(r_1 1\left\{K^0 - I > 0\right\} + r_0 1\left\{K^0 - I < 0\right\}\right)$$

with $1A = 1$ if A is true, and 0 if not. The *n*et profit of the student migration for the family can then be written:

$$\tilde{\Pi} = \left(r\left(Ed, X, \varepsilon; \theta\right) - r_1\right) I + \left(r_0 - r_1\right) \min\left(K^0 - I, 0\right) \tag{1.2.1}$$

Note that $(r_0 - r_1)$ measures the effect of the budget on the net profit of student migration. If $r_0 = r_1$, the family must simply compare its return from the student migration investment to the return of alternative investment in the origin country. All that matters in the decision is the additional return of investment and the budget

[4]For reference, the interest rate spread (the difference between the lending and the borrowing rate) in Cameroon is about 0.2 for "tontines" which are very popular among Cameroonian households for financing even very large investments (Nemb and Jumbo (2011))

constraint plays no role. We can simplify Eq. (1.2.1) to

$$\frac{\tilde{\Pi}}{I} = r\left(Ed, X, \varepsilon; \theta\right) + r_0 \min\left(\frac{K^0}{I} - 1, 0\right) \tag{1.2.2}$$

and $r(.)$ and r_0 can be reinterpreted "net of the interest rate r_1". We assume the following separable formulation of return function:

$$r\left(Ed, X, \varepsilon; \theta\right) = \tilde{r}\left(Ed, X; \theta\right) - \varepsilon$$

and define finally:

$$\tilde{\pi}\left(Ed, X, \frac{K^0}{I}; r_0, \theta\right) \equiv \tilde{r}\left(Ed, X; \theta\right) + r_0 \min\left(\frac{K^0}{I} - 1, 0\right).$$

The investment decision being denoted $Y \in \{0, 1\}$, where 1 means that the family chooses student migration, is then characterized in the following way:

$$Y = 1\left\{\tilde{\pi}\left(Ed, X, \frac{K^0}{I}; r_0, \theta\right) - \varepsilon \geq 0\right\}$$

We will assume later in Section 1.5 a linear return in the characteristics (see Eq. 1.5.1). It is assumed that the variable ε follows a logistic distribution of variance normalized to 1. In an abuse of notation in the following and when the context is clear enough to allow it, we will refer to the structural parameters of the function $\tilde{\pi}$ as θ, while meaning the pair (r_0, θ). The parameter r_0 will receive special attention in Section 1.3.4.

1.2.1 Treatment of endogeneity of the educational attainment

Among the observable characteristics influencing the migration decision, the schooling attainment of the candidate must be seen as an endogeneous variable. The nature of the potential endogeneity is here two-fold. First, the innate ability remains an omitted variable correlated with the educational attainment. This problem is well

known in the labor literature. For example, the signaling model in Spence (1973), agents with high innate ability are assumed to find school less difficult and obtain higher education as a signal of their high ability (see Willis (1985) for survey). Second, as recognize in the migration literature, education choice and migration decision are often simultaneously decided at household level (Hanson and Woodruff (2003), McKenzie and Rapoport (2007)). A change in educational attainment affects the decision to migrate, which affects the working migration probabilities, impacting in turn the return of student migration, and so forth. It is easy in fact to agree with the claim that more education decreases the cost of migration procedures and, therefore, makes it more likely to migrate. Indeed, Docquier and Marfouk (2006) find that emigration propensities are five to ten times higher for workers with more than twelve years of education than for workers with less than twelve years of education. In the other direction, the prospect of migration could be the incentive to acquire more education since it ensures higher returns abroad. Rosenzweig (2008), for example, points out that the choice of location of tertiary education significantly affects the probability that a person can emigrate permanently.

The common methodology to deal with these two issues would amount in our framework to specify further the relationship between the migration and schooling decision, in the form of a simultaneous probability model. Identification and estimation methodology then rely on a structural assumption about the error terms (bivariate normal distribution), as well as strong (identifying) exclusion restriction (see for example Mallar (1977)). Namely, we need variables that affect the choice of final education level and are known at the time the migration decision is made, but which do not directly affect the migration decision. This methodology is used by BLV, albeit the fact that they rather treat the problem of endogeneity of the migration variable in the equation of the schooling choice. Our main critics to this methodology are two-fold: first, the imposed structural assumption on the error terms are made for technical reason rather than sound economic intuition. Second, the exclusion restriction should be teamed with conditions on the support of the excluded

variable to obtain point identification. The argument is similar to the one in **?** (Theorem 2): identification of parameters occur through "independent variation in one regressor while driving another to take extreme values on its support (identification at infinity)". Altogether, these assumptions are arguably very strong.

The strength of the methodology presented in this paper resides in that we need not to model any further the relationship between Educational attainment (Ed) and migration (Y) to conduct inference[5]. In particular, the inference procedure does not require the existence of a valid instrument. The model rests therefore upon fewer structural assumption. It comes at the cost of losing identification of parameters. Indeed, allowing for endogenous explanatory variable in discrete choice models might hinder point identification of parameter of interests. Early treatments of this problem can be found in Manski (1993).

In this paper, we pursue the avenue of set identification rather than point identification. Recent contributions made by Beresteanu, Molchanov, and Molinari (2011), Galichon and Henry (2011) (GH, hereafter) and Chesher, Rosen, and Smolinski (2011) (CRS) show that, without additional hypothesis to the model, bounds can be derived on the structural parameters of an IV-discrete choice model. We present these results in the context of our model and show how to conduct inference, following the proposal of HMQ.

1.3 Inference procedure

In the model described above, we are first interested in the determinants of migration. We build on the work of GH, CRS and HMQ to conduct inference for the parameters of the structural equation (1.2.2). A confidence region is derived by using sharp bounds on the structural parameters. However, because some data are incompletely

[5]We will however impose a condition on the ordering of returns on migration for different level of education, all other variables being controlled

observed, further complications are added to the derivation of the sharp bounds, independently of the inference procedure. In the following, we present sharp bounds on the structural parameters of the model in its simplest form. On this simple model, we show how to conduct inference. Section 1.3.3 then deals with the complications mentioned above.

1.3.1 Sharp bounds

Suppose that we observe $Y \in \mathcal{Y}, Ed \in \mathcal{E}$ and $X \in \mathcal{X}$ on a family i, as defined by the previous section. We suppose in this section that we can also observe $V \equiv \frac{K_0}{I} - 1$, the fraction of the investment that the family needs to borrow. Note that V is a random variable distributed on the real line. We denote the vector $W \equiv (X, V)$. The unobservable latent variable is ε, distributed on the real line. It summarizes the decision shifters that are known to the family but unknown to the econometrician. Our interest is primarly in inference on θ, a finite dimensional parameter which characterizes the return on student migration of the family. With the above notation, the utility of the family can then be rewritten:

$$\tilde{\pi}\left(Ed, W; r_0, \theta\right) - \varepsilon = r\left(Ed, X; \theta\right) + r_0 . \min(V, 0) - \varepsilon \tag{1.3.1}$$

We assume that X and V are exogeneous in the sense that ε and W are stochastically independent. The above assumption implies in particular the innate ability of the child will not be influenced by the capital possessed by his family. Let $F^0_{Ed,Y|W}$, the distribution of the (Ed, Y) given W, and $F_{\varepsilon;\theta}$, the distribution of ε given W on $\mathbb{R}$. We will denote the respective densities accordingly $P_0(.|W)$ and $P_\varepsilon(.|W)$. As noted above, endogeneity of the random variable Ed might preclude point identification as several parameter values θ from the model could be consistent with the data. We define Θ_I this set of parameter. A model which singles Θ_I as the set of possible values for θ is said to be set identified. Point identification occurs when Θ_I is reduced to a singleton. A model is rejected by the data if Θ_I is the empty set.

We give here the sharp bounds induced on the structural parameters for a model with two alternatives of migration and two levels of education.

Theorem 1 (Sharp bounds) *Consider the set of parameters θ for which we have positive return on student migration to education, i.e.*

$$\tilde{\pi}(1, w; \theta) > \tilde{\pi}(0, w; \theta)), \text{ for all } w. \tag{1.3.2}$$

A parameter θ belongs to the identified set, if and only if we have :

$$\begin{aligned}
P_0(Y = 1, Ed = 0 \,|W = w) &\leq F_\varepsilon(\tilde{\pi}(0, w; \theta)), && (1.3.3)\\
P_0(Y = 1, Ed = 1 \,|W = w) &\leq F_\varepsilon(\tilde{\pi}(1, w; \theta)), && (1.3.4)\\
P_0(Y = 0, Ed = 0 \,|W = w) &\leq 1 - F_\varepsilon(\tilde{\pi}(0, w; \theta)), && (1.3.5)\\
P_0(Y = 0, Ed = 1 \,|W = w) &\leq 1 - F_\varepsilon(\tilde{\pi}(1, w; \theta)), && (1.3.6)\\
P_0(Y = 1 \,|W = w) &\leq F_\varepsilon(\tilde{\pi}(1, w; \theta)), && (1.3.7)\\
P_0(Y = 0 \,|W = w) &\leq 1 - F_\varepsilon(\tilde{\pi}(0, w; \theta)), && (1.3.8)\\
& && w \text{ a.e.}
\end{aligned}$$

Theorem 1 is a direct consequence of Propositon 1 in Appendix 1.7.1. The terms in the left-hand side are derived from the Data Generating Process of the observable variables. Note that they do not include the parameter θ. The terms in the right-hand side are derived from the cumulative distribution of the latent variable ε. To understand these inequalities, we can think of the observed outcome (y, e) as part of a multiple equilibrium predicted by the model. For example, let θ_0 be the true parameter. $(Y = 0, Ed = 1)$ can only be observed if $\tilde{\pi}(0, w; \theta_0) \leq \varepsilon$ (even after obtaining education $Ed = 1$, the return of migration are not large enough to make migration attractive). But if the latter is true, the model predicted also that $(Y = 0, Ed = 0)$ was a possible outcome i.e, returns to migration are also too small for those who choose education $Ed = 0$. $\{(Y = 0, Ed = 1); (Y = 0, Ed = 0)\}$ can be understood as a multiple equilibrium prediction. In other words, we cannot observe $(Y = 0, Ed = 1)$ more

often than the model predicted $\{(Y=0,Ed=1);(Y=0,Ed=0)\}$. Hence, (1.3.6). Theorem 1 summarizes this fact: the probability that we observe a given outcome, cannot be greater than the probability that the model predicts at least one of all the multiple equilibria where this outcome is part of.

Under the condition (1.3.2), Θ_I is defined as the set of parameters for which (1.3.3) - (1.3.8) are true. These bounds which can easily be derived as necessary conditions, are also sufficient (see Appendix 1.7.1), hence the term "sharp" bounds. Note that some of these inequalities are redundant, and only a subset of those is enough to characterize Θ_I. We now turn to the inference procedure.

1.3.2 Inference on the structural parameters

Condition (1.3.2) is not a stochastic condition which will be assumed through out the rest of the paper. Here, we seek coverage of the identified set, Θ_I with a prescribed probability, $1-\alpha$. The idea of the procedure is to define a new set of inequalities which relaxes the bounds with a definite probability $1-\alpha$, so that a parameter satisfying (1.3.3) - (1.3.8) will satisfy this new set of inequalities with confidence level $1-\alpha$. The confidence region will simply be the collection of all those parameters satisfying the new set of inequalities. The relaxation occurs through the construction of a function $\underline{P}_n$, dominated by the probability distribution P_0. Suppose indeed that we can construct a function $\underline{P}_n$ such that for all $(y,e)\in\{0,1\}\times\{0,1\}$, and w a.e.:

$$\underline{P}_n\left(Y=y,Ed=e\,|W=w\right)\leq P_0\left(Y=y,Ed=e\,|W=w\right) \tag{1.3.9}$$

with probability $1-\alpha$. A new set of inequalities (1.3.3') - (1.3.8') can be obtained by replacing P_0 by $\underline{P}_n$ in (1.3.3) - (1.3.8). Define now $\hat{\Theta}_n$ has the collection of parameters θ satisfying (1.3.3') - (1.3.8'). Theorem 4 in Appendix 1.7.1 shows that we achieve with $\hat{\Theta}_n$ a proper coverage of the identified set Θ_I.

Construction of a functional satisfying (1.3.9) is proposed by HMQ through a procedure called "efficient combinatorial bootstrap" that runs in linear computing

time. It involves bootstrapping the empirical process of the distribution $P_0(.|W)$, to retrieve the $(1-\alpha)$-quantile of this process, $c_\alpha(W)$. Then, the empirical distribution $\hat{P}_0(.|W)$ is decreased by $c_\alpha(W)$. This decreased quantity offers the desired functional $\underline{P}_n$.

Although HMQ only covers the case of discrete variables, results from Chernozhukov, Lee, and Rosen (2009) (Section 4.1) makes it sensible to use the procedure in the parametric case.

1.3.3 Incomplete observation of Educational attainment

The respondent's educational attainment is not completely observed unless he/she has completed her study. In fact, 48% of the respondents have not. Information are available though on the level of education at the time of response, which we will denote it $\underline{Ed}$. To complement this information for those respondants who had not yet completed there studies, a question in survey was relative to their highest expected educational attainment. this variable is then observed in our dataset and we will denote it $\overline{Ed}$. Under the assumption that a person does not study more than the expected level, i.e. $\underline{Ed} \leq Ed \leq \overline{Ed}$. We then have an interval $\left[\underline{Ed}; \overline{Ed}\right]$ such that:

$$\mathbb{P}\left(Ed \in \left[\underline{Ed}; \overline{Ed}\right]\right) = 1 \tag{1.3.10}$$

This information can be incorporated to the earlier framework as a censored variable problem. The non-redundant sharp bounds become:

$$\begin{aligned}
P_0\left(Y=1, \underline{E}d=0, \overline{E}d=0 \,|W=w\right) &\leq F_\varepsilon\left(\tilde{\pi}\left(0,w;\theta\right)\right), && (1.3.11)\\
P_0\left(Y=0, \underline{E}d=1, \overline{E}d=1 \,|W=w\right) &\leq 1-F_\varepsilon\left(\tilde{\pi}\left(1,w;\theta\right)\right), \\
P_0\left(Y=1 \,|W=w\right) &\leq F_\varepsilon\left(\tilde{\pi}\left(1,w;\theta\right)\right), && (1.3.12)\\
P_0\left(Y=0 \,|W=w\right) &\leq 1-F_\varepsilon\left(\tilde{\pi}\left(0,w;\theta\right)\right), && (1.3.13)\\
& && w \text{ a.e.}
\end{aligned}$$

See Proof in Appendix 1.7.1.

1.3.4 First-step estimation of V: the proportion of the investment that the family needs to borrow

We assumed in Section 1.3.1 that the realization of the random variable V, the ratio of the capital of the family to the amount to invest, was observable for each family. As in many other studies, this information is absent from the dataset. The "natural" method to get around this problem, would be to add to the covariates, all observable variables which are known to influence the capital of the family, and conduct inference with this new set of covariates. This, however, by augmenting the dimensionality of W, increases the size of the data required to achieve informative inference and the computational burden. We use instead additional information provided by the survey to devise a two-step estimation in the same spirit as an Heckit estimation. This simplification will however be at the expense of informativeness of inference relative to the parameter measuring the effect of the budget constraint.

The dataset allows to distinguish families who need to borrow from families who have sufficient funds to cover all the costs of the migration investment. In other words, we observe $1\{V < 0\}$ instead of V. From this information and the observation of other socio-economic characteristics of the family, we construct a confidence region for the realization of V. Indeed, denote L_i, the observable socio-economic characteristics of the family i. We postulate for V_i the following single index functional form:

Assumption 1

$$V_i = \beta L_i + u_i \text{ where } u_i \text{ follows } N(0, \sigma_u) \tag{1.3.14}$$

where u is stochastically independent of ε given (W, L).

See the variables in included in L in Table 1.1.

Since $1\{v < 0\}$ is observed, the parameter of this model can be estimated through a probit estimation, under a scale normalization. One word of caution is therefore required here. If σ_u were known, for a real $0 < \alpha_v < 1$, we would then have an interval

$[\underline{v}; \overline{v}]$ such that:

$$\mathbb{P}\left(V \in [\underline{v}; \overline{v}]\right) = 1 - \alpha_v \tag{1.3.15}$$

Our inference problem becomes one where the covariates are defined by an interval rather than a point. By an appeal to the Composition Theorem of Galichon and Henry (2006a) (Theorem 1), we can redefine our identified set and propose a valid confidence region, following the same procedure as in Sections 1.3.1 and 1.3.2. However, σ_u is a nuisance parameter here and the estimation procedure requires a scale normalization. As the reader might expect, the probit model under the standard normalization $\sigma_u = 1$ gives us not a confidence interval for V, but for a variable $V_{|\sigma=1}$, related to the original variable by the relation:

$$V_{|\sigma=1} = \sigma_u^{-1} V.$$

The nuisance parameter σ_u will affect inference on the parameter r_0. Indeed, using $V_{|\sigma=1}$ Equation (1.3.1) can be rewritten.

$$\tilde{\pi}\left(Ed, W; r_0, \theta\right) - \varepsilon = r\left(Ed, X; \theta\right) + r_0.\sigma_u.min(V_{|\sigma=1}, 0) - \varepsilon \tag{1.3.16}$$

The bounds that we obtain will be informative for the parameter $r_0.\sigma_u$. However, since σ_u remains unrestricted, the model will be uninformative on the parameter r_0.

Note that, the inference methodology is in two steps, reminiscent of the two-steps of an Heckit estimation. Prior to the inference in our partially identified framework, we perform a first step estimation to overcome the computational burden induced by the high number of parameters. This first step estimation consists in a probit of some components of the parameter vector θ. Without it, the partial identification inference would be computationally infeasible, since it would involve searching over a parameter grid with an unreasonable size. The price to pay is uninformativeness of our procedure on the parameter r_0. A mathematically rigorous treatment of the inference procedure in a general case is offered in Appendix 1.7.1 for interested readers.

1.4 Database

We begin this section by providing some facts on Cameroon, which we believe are relevant in understanding the background of the study. The main part of this section (Subsection 1.4.2) is devoted to the core of the snowball sampling methodology used for the online survey. We explicit the sampling design and show, under simplifying assumptions, how to estimate the population proportions by correcting for non-response and bias in the survey sampling design. Finally, in Section 1.4.3, we discuss the descriptive statistics drawn from the dataset.

1.4.1 Snapshot of Cameroon

Cameroon[6] has an estimated population of 19.5 million in 2009, relatively young: an estimated 40.9% are under 15. After a severe recession period from 1985 to 2000, due to the fall in price of raw material exports, the nominal GDP returned to a steady growth from 2001, with an annual rate around 3% from 2004 to 2009 . However, 30% of the population lives with less than 2$ per day in 2007. The educational system in Cameroon is a mixture of British and French precedents. The typical curriculum consists of 6 years in primary school and 7 years in secondary school. Access to university is conditional on passing the state exam, "Baccalauréat", named after the equivalent French exam. The enrollment in first year primary school is estimated at 88.3% in 2008, with a noticeable difference between male (94.3%) and female (82.3%). However, transition and survival rate are fairly low, with 16% of the total of enrolled being repeaters in the same year. Progression to secondary school is 44.4%. The Bac-

[6]The data presented in this section are compounded from different reports released by: (1) international organizations, principally the 2009 Report of the International Organization for Migration (IOM) on the national profile of migration in Cameroon, the database of the UNESCO Institute for statistics, accessible online, and the World Development Indicators as released by the World Bank in 2011 and (2) the National Institute of Statistics (INSEE) of Cameroon and the Cameroonian Ministry of Education.

calauréat exam concerns close to 50,000 candidates each year. The success rate went up from 40 % to 50-55% during the past 5 years increasing the pressure on universities which receive relatively scarce allowances for investment in new infrastructures (see Makosso, 2006). The number of migrants with Cameroonian citizenship was less than 1% of the population. They tend to be long term migrants. In stark contrast, migration of skilled individuals is of relatively high magnitude. The ratio of skilled migrant to the population of skilled non-migrant is 17.2% (Docquier and Marfouk (2006)). Brain drain is, for this reason, a serious concern for the Cameroonian State. In line with the migrating trend of highly educated, the ratio of Cameroonian students enrolled in a foreign country to the total number of Cameroonian students was estimated at 14.5% in 2006. On a final note, in the database collected by Docquier and Marfouk, Cameroon ranks 25th, among countries with a population higher than 4 millions, in term of rate of migration of skilled individual to OECD in 2000 (11 other sub-Saharan African countries are part of the list of the 30 highest rates). Beine, Docquier, and Rapoport (2008) finds evidence of a detrimental Brain Drain effect in Cameroon. According to their measure, the country looses 0.1% of its skilled force relative to the situation of closed economy.

1.4.2 Sampling methodology and corrected estimators

For the purpose of the study, a survey has been conducted on the population of Cameroonian, aged 18 or more, having completed secondary school by obtaining the Baccalauréat. A novelty of our dataset is that it comprises information on both migrants and non-migrants.

Sampling Design

To reach both populations (migrant and non-migrant), we used a snowball sampling procedure through an online platform. The initial sample consisted of 22 individuals

(called "seeds") contacted by the researcher. The seeds where chosen on the base of geographical (country of residence) and demographical (gender, age) to include, as much as possible, all the components of the population. Each seed were asked to answer a questionnaire and to invite as much friends as possible from the population of interest. The invitee would receive an electronic mail from his host with the detail of the survey, and a unique link to access the online questionnaire. If he/she accepted to participate, he/she was required to complete the questionnaire and invite as much friends as possible in the population of interest. Recruitment is said to occur in waves and stops when invitees fail to complete the survey or invite other friends. The wave at which i is invited is the number of recruiters that separates him from the initial sample. Participation in the study was restricted to a prior invitation and each invitee received a unique token which enabled us to retrace the paths of invitation. More information on the survey implementation is available in Appendix 1.7.3.

Estimators

We give here an idea of the methodology to correct for biases induced by non-random sampling and non-response of invitees. The mathematical proofs and the modification suggested to the combinatorial bootstrap procedure advised in HMQ are relegated in Appendix 1.7.2.

Because of the particularities of sampling procedure, Horvitz-Thompson estimators are used as unbiased estimators of the true population mean. To use these estimators, we need to compute the inclusion probability for individuals in the sample. Call q_{ki}, the conditional probability for individual i to be invited to the survey, knowing that the survey reached wave k. Suppose that we are at wave k. The set of people invited during the previous wave, are now the set of recruiters. Call this set the *active set* and denote it a_k. Proposition 4 in Section 1.7.2, shows that under some simplifying assumptions, q_{ki} can be decomposed into two main terms :

- The probability that an invitee i agrees to participate in the study, conditional

on being invited. Under the assumption that non-response occurs at random, we estimate this probability through the ratio of response when the survey reaches wave k.

- And the probability that individual i is invited by a recruiter in the active set. To estimate this second term, we need (1) to know the probability that i knows a recruiter in the active set. To do so, we fit a model of graph, the model of Erdos-Renỳi, to our sample network, in order to estimate the probability that two individuals know each other. We further need to know (2) the probability for a recruiter $j \in a_k$ to invite i, given that the two individuals share a relationship. Under the assumption that j chooses his/her invitees at random among his friends, this probability is given by the ratio of the number of invitees to the number of connections of j in the population, j's *degree*.

1.4.3 Descriptive statistics

The dataset consists in 402 individuals, see Tables 1.3 to 1.4. We discuss some facts on the dataset and sometimes compare characteristics of migrants and other respondants, when differences are statistically significant.

Both populations are similar with regards to their age and marital status. Migrants however, appear to obtain their secondary school degree one year earlier than the others, suggesting higher number of repeaters among the non-migrants. Respondents are predominantly male. OECD countries (especially France and North America) and african countries are the favorite destinations of migrants in our sample. The survey seems to capture too few migrants to Germany compared to existing data, even after applying our correction. We however capture a sizable proportion of migrant returning to Cameroon after their studies. 19% of respondants who once migrated are now residing in Cameroon. Concerning the education, more than 40% have completed their studies. We estimate that close to a quarter of students have

acquired a Master Degree or an equivalent. The most popular fields of study are Social sciences, science and engineering. We estimate that at least 42% leave before obtaining any tertiary degree in Cameroon, and more than a third pursue at least 4 years of study abroad.

Regarding the financement, while Parents are the primary source of financing studies in Cameroon, foreign studies expenses appear to be more often shared by members of the family or left to the charge of the migrant. Parents of OECD-migrants seem to differ from the others in education and their ownership of car. More than half the respondents declare an helper to the process of migration. This helper is in the majority of cases a male, with a university degree, with close link to the family (uncle/aunt, brother/sister) and/or lives abroad.

Hereafter, we will refer to OECD countries of two types, those with high fees (US, UK, Canada, Australia and Ireland) and those with low fees. The countries are classified according to the average tuition required for students from Cameroon, as collected by the author from official documents produced by the consulates of OECD countries. As for the time span of the survey, "High tuitions" range from $8,000 US to $30,000 US in addition of all living costs. Countries with low tuitions, such as Germany and France, require from $0 US to $2000 US each year. Australia and Ireland, which are high tuition countries, do not appear in our sample.

1.5 Empirical results

We present first the results for the preestimation of the family capital. We then turn in Subsection 1.5.2 to the results of our inference procedure for the discrete choice model of student migration investment.

1.5.1 First-step estimation

As detailed in Section 1.3.4, we run a probit regression to retrieve a 95 % confidence interval on the proportion of capital that a family needs to borrow in order to meet the liquidity constraint. The dependent variable is only discretely observed. In the questionnaire, respondents identify who is, will be or would have been responsible for the different costs in the event of a migration. These costs are divided in living costs, tuition costs and travel costs. The possible payers are the candidate himself, the parents, a definite helper, the government or a scholarship, or a mix of all these options. If individuals are paying, the respondents is further asked whether the costs are paid through savings, regular income, borrowing from an individual or from an institution, a mix of the previous options or from other means. We will consider that the budget constraint is binding, if one of the payer provides the funds through the means of a loan, or if a non-migrant expects the government to pay the fees for the migration.

For a first regression, including the whole sample, results are counter-intuitive, as the capital of the family is unrelated to its physical capital (ownership of a car or a house) and negatively correlated to the parents education. This suggests that non-migrants differ significantly from migrants in their evaluation of the costs. As they do not have full information on migration costs, non-migrants appear to provide unreliable estimates of the ability of the family to meet the costs of the migration. We therefore only use the subsample of 132 migrants to the OECD. Results are displayed in Table 1.1 for several specifications.

The results show that obtention of a scholarship significantly decreases the liquidity constraints (Scholarships available for students in Cameroon are mainly on the basis of merit and not on the basis of need). Number of cars owned by the family (the parents and helper when present) as well as maximal family education are also good predictors of the family capital. This latter result concords with the findings of National Institute of Statistics of Cameroon, in their survey on Cameroonian Household

in 2006.

What do we learn about the helper? Interestingly, once we have controlled for the physical capital and the education of a helper, that he/she resides in the country of migration does not appear as a significant source of capital for the family. Our data do not support the idea that a migrant will hold significant implicit capital from the location of a low-educated helper with low capital in the host country. The measure of the support provided even decreases in absolute value when we control for the number of siblings abroad.

1.5.2 Inference on parameters of the structural model of Investment

We conduct inference on the model described in section 1.2 with two levels of education: $Ed = 1$ if the individual has at least a Masters Degree, $Ed = 0$ otherwise.

Model specification

We use the following specification of the return of individual i for choosing alternative j, $j \in \{0, 1\}$ in eq. (1.2.2)

$$\begin{aligned} \tilde{\pi}_0 &= 0 \\ \tilde{\pi}_1 &= \mu + \alpha.Ed_i + X_i.\beta^{'} + r_0.\sigma_u.\min\left(\hat{\tilde{V}}_1(L_i), 0\right) \end{aligned}$$

where:

- $\hat{\tilde{V}}(L_i)$ denotes the estimate of the proportion of the investment that a family needs to borrow. We use Specification 2 in Table 1.1.
- X_i are the characteristics of individuals which might influence the success of the migration.

Table 1.1: Probit regression of the Proportion of capital that a family must borrow, conditional on emigration to OECD countries.

	Spec 1	Spec 2	Spec 3	Spec 4
Intercept	-0.564**	1.187***	1.225***	1.313***
	(0.326)	(0.415)	(0.432)	(0.444)
Grant	-1.112***	-1.927 ***	-2.024***	-2.050***
	(0.410)	(0.517)	(0.515)	(0.487)
OECD High Tuition	0.934**	0.849**	0.745**	0.226
	(0.396)	(0.370)	(0.363)	(0.394)
Family characteristics				
Maximum parent education		-0.370 ***	-0.312 **	-0.291 **
		(0.132)	(0.137)	(0.142)
Number of cars owned		-0.481 **	-0.440 **	-0.417 **
		(0.202)	(0.194)	(0.172)
Helper in Migration Country			-0.469	-0.165
			(0.355)	(0.350)
OCDE high $\times$ Nbr children abroad				0.364
				(0.291)
OCDE low $\times$ Nbr children abroad				-0.259**
				(0.139)
log likelihood	-34.983	-27.340	-26.798	-25.338
AIC	75.966	64.679	65.596	66.676

Number of obs. = 132.

Estimation is made for a normal standard error term.

Standard deviation are in parentheses.

(***) significant at 1%. (**) significant at 5%. (*) significant at 10%.

- Again, ε is assumed to have a logistic distribution with normalized variance (equal to 1).

In this specification, α measures the additional return from holding a Master degree in a foreign country. The characteristics of the individuals that will be of interest are the Gender (with parameter β_G), the primogeniture (β_{prim}), the country of residence of the helper (β_{resid}) and the success of the child in secondary education (β_{qual}). To measure the latter, we construct a dummy variable which is one if the candidate to migration has succeeded the Baccalauréat with honors. We interact this dummy variable with another variable relative to the age at which the candidate passed the Baccalauréat. This variable is zero if the individual was older than the median person, and equal to the difference of age if younger[7]. A positive coefficient suggest positive selection of the migrant. Regarding enrollment in first year primary school, families in Cameroon exhibit a strong preference for male children over female children. It is of much interest to understand whether this preference survives through the process of student migration. The literature on migration of high skilled individuals suggests that such gender gap may not exist (see survey from Docquier and Rapoport (2009)). Related to the fact that helpers are most commonly brothers of the candidate, we expect firstborn children to suffer from a lack of support. A complementary explanation might be provided by the literature on elderly parents care. Firstborn children are often chosen to provide care for elderly parents and distance deters the readiness of the child to provide such care (see for example Engers and Stern (2002), Rainer and Siedler (2009)). Since student migration might ultimately result in permanent migration, family members (mostly parents) might be reluctant to send the eldest child abroad. Residence of the helper in the migration country was found not to add to the capital of the family once we control for the education and physical capital

[7]The educational system displays a high rate of repeaters at every level. According to UNESCO statistics, around 20% in primary school and 14% in each year of Secondary school. Younger students are then expected to have finished primary and secondary school without repeating a class, and to be better than the average.

of the helper. We explore whether it influences in alternative ways the return of migration. We think here of some non-pecuniary return of having a relative living in the same country. The literature on economic migration acknowledges the existence of psychological cost that may arise from separation from one's native that might be reduced by the presence of a previous migrant in the host country (Bauer, Epstein, and Gang (2000), Mahmood and Schömann (2003)). We will also explore whether the fact that parents have a past experience of migration (β_{exp}) influences the migration probability. A number of studies (for example BLV) use variables related to the past experience of migration of parents as exclusion restriction in the equation of migration[8].

Significance of the individual parameters is evaluated by checking whether the hyperplanes defined by $\theta_i = 0$ - where θ_i is a component of θ - intersect the 90% confidence region. We report the range for each parameters in Table 1.2.

Discussion of results

An important result is that the differential in the return from holding a Masters degree has a strong positive effect on incentive to study abroad ($\alpha > 0$). To give and idea of the magnitude of α, we can interpret the variation of the odd ratio defined as the probability to migrate relative to the probability not to migrate. Our finding implies that, controlling for other covariates, the odds of migration will at least double when individual choose to study at least to Master Degree. Although families exhibit a clear preference for studies in Cameroon rather than abroad ($\mu < 0$), their disutility is entirely compensated by the return from holding a foreign Master degree.

Our measure of the success of the child during secondary school appears positively correlated with the chances of migration. We reject $\beta_{qual} < 0$, which means that the best students are more likely to migrate. This suggests a positive selection

[8]We also used in an alternative specification the past experience of student migration of the parent. This variable does not appear to be significant.

Table 1.2: Parameter ranges

	Min	Max
Master in foreign Country (α)	0.88	3.61
Cst. (μ)	-0.38	-0.13
Female (β_G)	-0.90	0.10
Primogeniture (β_{prim})	-0.70	-0.29
Academic results (β_{qual})	0	2.41
Migration experience (β_{exp})	-0.50	0.25
Residence of helper (β_{resid})	-0.41	0.04

$r_0\sigma_u$ is not reported since uninformative.

of migrants. An interesting result is that the firstborn child is less likely to migrate than his younger siblings. Indeed, his/her odds of migration are 25 % smaller. This result is not surprising as a third of the helper reported are (older) brothers or sisters. Investigation of chain migration within families might be of great interest. We find no overwhelming evidence that the male students are favored by the family in the migration process, (we cannot rule out $\beta_G = 0$), a result in line with the main findings in the previous literature on migration. Nevertheless, the confidence region shows a tilt toward negative values of this parameter. Our data are not compatible with models were strong preference is given to migration of female students. When we study instead a dummy variable for the presence of the helper in the migration country (not reported), the range of β_{resid}, includes 0 (and is almost symmetric around this value). In the same line, when we consider instead residence of the helper in Cameroon, we cannot rule out $\beta_{resid} = 0$, and models where there exists a positive effect from having an helper residing in Cameroon are almost always rejected. Whether the helper is from diaspora or living in Cameroon, does not seem to be a significant variable for the final outcome. The role of the helper seems then to be limited to provision of (explicit or implicit) capital to the migration process. To reconcile this result with the

findings of the literature about the network effect of the diaspora, we must conlude that the lessening of psychological costs is then not related to a sole individual but to a larger network that the migrant can build. Finally, the past migration experience of the parents does not appear as a significant determinant of the migration of the child, a finding warranting caution in the use of this variable as exclusion restriction.

1.6 Conclusion

This paper presented a framework to analyze the determinants of the choice of migration from a microeconomic perspective. We do so while overcoming the main obstacles to a rigorous study, namely: the lack of data about population of migrants and non-migrants and the technical difficulty raised by the endogeneity of the educational attainment in our discrete-choice model. We deal with the first of these issues using a novel chain-referral sampling procedure, run through an online platform. The procedure allowed to overcome the geographical challenge posed by the type of population of interest. We proposed population mean estimators that correct for biases induced by non-response and non-random sampling. In practice, strong assumptions on recruitment behaviors are needed to retrieve the inclusion probabilities of sampled individuals. We also need to fit a random graph model to represent the relation pattern between individuals in the population. The choices made in this paper were for computational reason.

We propose an empirical structural decision model which reflects the importance of both the return of the investment and the budget constraint in agent choices. An important contribution of this paper is the appeal to recent results in the literature about incomplete, partially identified models to circumvent the problem of endogeneity. In doing so, we relax the point identification condition for the parameters of our structural model of investment in student migration. We conduct inference using sharp bounds and retrieve confidence regions which appear quite informative. An-

other contribution of the paper is the novel two-steps inference methodology that takes into account censored or incompletely observed variables and considerably reduce the computational burden.

From our sample, we find that a higher educational level and better results during secondary education of the candidate increase significantly the returns to migration. We find also that male students are not favored by the family in the migration process and the firstborn child is less likely to migrate than his younger siblings. Our interpretation is that they suffer from lack support from an elder brother/sister. Our survey data are actually quite informative about the helper who appears to be in the majority of the cases a male relative of the parent and the child, with tertiary education. When present, he influences primarily the budget constraint of the family.

1.7 Appendix

1.7.1 Mathematical treatment Inference procedure

Derivation of the sharp bounds

Suppose that we observe $Y \in \mathcal{Y}, Ed \in \mathcal{E}$ and $W \in \mathcal{W}$. The random vectors Y, Ed, W and ε are defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. With the notation of section 1.3.1, we provide now a characterization of Θ_I. To state the characterization, we define additional objects.
Define:

$$G\left(\varepsilon|W;\theta\right) \equiv \{(y,e) : y = 1\{\tilde{\pi}\left(e,W;\theta\right) \geq \varepsilon\}\}$$

as a correspondance from $\mathbb{R}$ to $\mathcal{Y} \times \mathcal{E}$ which, for given values of the exogenous regressors, associates to each ε a duple (y,e) predicted by the model, for given values of W. The following example provides an insight of the type of object represented by the multi-valued mapping G.

Example 1 (Correspondence G with three level of school attainment) *Suppose Ed is a discrete variable which takes the ordered values $\{0,1,2\}$. Suppose further that $r(0) < r(1) < r(2)$ which amounts to a positive return of additional years of schooling. The multi-valued mapping G can then be written in the following form:*

$$\begin{aligned} G: \quad \varepsilon \in [-\infty; \tilde{\pi}(0;\theta)] \quad &\rightrightarrows \{(1,0);(1,1);(1,2)\} && (1.7.1)\\ \varepsilon \in [\tilde{\pi}(0;\theta);\tilde{\pi}(1;\theta)] \quad &\rightrightarrows \{(0,0);(1,1);(1,2)\} && (1.7.2)\\ \varepsilon \in [\tilde{\pi}(1;\theta);\tilde{\pi}(2;\theta)] \quad &\rightrightarrows \{(0,0);(0,1);(1,2)\} && (1.7.3)\\ \varepsilon \in [\tilde{\pi}(2;\theta);+\infty] \quad &\rightrightarrows \{(0,0);(0,1);(0,2)\} && (1.7.4) \end{aligned}$$

Note that $\varepsilon > r(2)$ (resp. $\varepsilon < r(0)$) implies that migration is never (resp. always) profitable to the family.

From GH, we draw the characterization of the identified set that we take to be our definition of the identified set:

Definition 1 (Characterization of the identified set.) *A parameter value θ belongs to the identified set, if and only if there exists a probability distribution p defined on $(\Omega, \mathcal{F}, \mathbb{P})$ with marginal $F^0_{Ed,Y|W}$ and $F_{\varepsilon|W,\theta}$, such that :*

$$E_p\left(1\{(Y,Ed) \in G(\varepsilon;\theta)\}|W=w\right) = 1, \ w \ a.e. \tag{1.7.5}$$

This definition is not operational since inference on parameters induces finding a possibly infinite dimensional probability distribution. We can write an alternative characterization, more useful in practice. It summarizes the problem to one of checking a finite set of inequalities which involves compact sets on the real line. We define the following object:

$$T(y,e|W;\theta) \equiv \{\varepsilon : (y,e) \in G(\varepsilon|W;\theta)\}$$

a subset of $\mathcal{E}$. For $Ed = e$ and for a given θ and W, $T(y,e|W;\theta)$ gives all the values of ε that deliver the value y of Y .

Proposition 1 (Alternative characterization of the identified set, Theorem 1 CRS) *Let $K(\mathbb{R})$ be the set of closed set of $\mathbb{R}$.*

$$\Theta_I = \{\theta \in \Theta : P_0(T(Y, Ed|W;\theta) \subseteq S|W = w) \leq P_\varepsilon(S|W = w;\theta), \\ w \text{ a.e. } \forall S \in K(\mathbb{R})\} \quad (1.7.6)$$

Theorem 2 in CRS shows that it is enough to check the inequalities for the sets S which are connected and for union of sets on the support of $T(Y, Ed;\theta)$. Continued example shows the kind of inequalities involved.

Example 1 continued Let $\mathcal{S}$ be the subset of $K(\mathbb{R})$ which collects all the relevant compact set S, i.e., which are connected and union of sets on the support of $T(Y, Ed;\theta)$.

$$\mathcal{S} = \{[-\infty;\tilde{\pi}(0;\theta)];[-\infty;\tilde{\pi}(1;\theta)];[-\infty;\tilde{\pi}(2;\theta)]; \\ [\tilde{\pi}(0;\theta);+\infty];[\tilde{\pi}(1;\theta);+\infty];[\tilde{\pi}(2;\theta);+\infty]\}$$

For each element $S \in \mathcal{S}$, we test

$$P_0(T(Y, Ed;\theta) \subseteq S) \leq P_\varepsilon(S;\theta)$$

For example, when $S = [-\infty;\tilde{\pi}(0;\theta)]$, the inequality becomes:

$$\mathbb{P}(Y = 1, Ed = 0) \leq F_\varepsilon(\tilde{\pi}(0;\theta))$$

Confidence Region

We proposed a method to compute the object of interest Θ_I in the limit case, where the true distribution of dependent variables $\mathbb{P}(y, e|w, z)$, $(y, e) \in \mathcal{Y} \times \mathcal{E}$, is known. Of most interest, is the problem of inference on Θ_I based on a sample of observations $((Y_1, Ed_1, W_1), \ldots, (Y_n, Ed_n, W_n))$ from an ergodic sequence. We seek coverage of the identified set with prescribed probability $1 - \alpha$, for some $\alpha \in [0, 1]$.

Definition 2 (Confidence region) *A confidence region of asymptotic level* $1-\alpha$ *for the identified set is defined as a region* Θ_n *satisfying*

$$\liminf_n \mathbb{P}(\Theta_I \subseteq \Theta_n) \geq 1-\alpha.$$

In HMQ, the authors proposed a computationally feasible procedure which is based on the idea of dilation of the inequality and makes use of the resampling bootstrap. We give here a sketch of the idea. We are interested in coverage of the set of values of the parameter θ such that

$$P_0\left(T\left(Y, Ed; \theta\right) \subseteq S\left|w\right.\right) \leq P_\varepsilon\left(S|w; \theta\right), \tag{1.7.7}$$

for all values of w, and all subset S of $K(\mathbb{R})$. P_ε is determined from the model, but P_0 is unknown. However, if we can construct lower probabilities from the sample of observations, i.e. random functions $\underline{P}_n(S|w,z)$ that are dominated by the probabilities $P_0(S|w,z)$ for all values of w, and all subsets S of $K\left(\mathbb{R}\right)$, then in particular,

$$\underline{P}_n\left(T\left(Y, Ed\left|W\right.; \theta\right) \subseteq S\left|w\right.\right) \leq P_0\left(T\left(Y, Ed\left|W\right.; \theta\right) \subseteq S\left|w\right.\right)$$

for each w, and each subsets S of $K\left(\mathbb{R}\right)$ (Remark that the dominated functions are constructed independently of the parameter θ). Hence any θ satisfying (1.7.7) for each w, and each subset S of $K\left(\mathbb{R}\right)$, also satisfies

$$\underline{P}_n\left(T\left(Y, Ed\left|W\right.; \theta\right) \subseteq S\left|w\right.\right) \leq P_\varepsilon\left(S|w; \theta\right)$$

for all values of w, and all subset S of $K(\mathbb{R})$. It remains to control the level of confidence of the covering region, which is achieved by requiring that $\underline{P}_n$ dominate P_0 with probability asymptotically no less than the desired confidence level. The fundamental feature of the procedure is that it dissociates search in the parameter space from the statistical procedure necessary to control the confidence level. The lower "probabilities" $\underline{P}_n$ can be determined independently of θ in a procedure that is performed once and for all using only sample information. Hence the following theorem, similar to Theorem 1 in HMQ.

Assumption 2 *Let the random functions $S \mapsto \underline{P}_n(S|w)$, $S \in K(\mathbb{R})$, satisfy*

$$\liminf_n \mathbb{P}\left(\max_{1\leq j\leq n}\max_{S\in K(\mathbb{R})}[\underline{P}_n(S|W_j) - P(S|W_j)] \leq 0\right) \geq 1-\alpha.$$

Theorem 2 (Confidence region) *Under Assumption 2, the region*

$$\Theta_n(\underline{P}_n) \;=\; \left\{\theta\in\Theta:\ \max_{1\leq j\leq n}\max_{S\in K(\mathbb{R})}\underline{P}_n\left(T\left(Y,Ed\,|W_j\,;\theta\right)\subseteq S\,|W_j\,\right) - P_\varepsilon\left(S|W_j;\theta\right)\leq 0\right\}$$

is a confidence region of asymptotic level $1-\alpha$ for Θ_I.

Note that because of the limited size of our dataset, we use a logit approximation to the joint distribution of Y and Ed given the observable covariates. Hence, the bootstrap procedure in HMQ remains valid even in presence of continuous covariates because of the parametric assumption.

Incomplete observation of Educational attainment

Define J to be a correspondence that maps $\mathcal{Y}\times\mathcal{E}$ into $\mathcal{Y}\times\mathcal{E}\times\mathcal{E}$ and associates to the duple (Y,Ed), the triple $\left(Y,\underline{Ed},\overline{Ed}\right)$. By the composition theorem (Theorem 1, Galichon and Henry (2006a)), the identified set can now be characterized the following way:

Proposition 2 (The identified set with incomplete observation of Ed) *A parameter value θ belongs to the identified set, if and only if there exists a probability distribution p defined on $(\Omega,\mathcal{F},\mathbb{P})$ with marginal $F^0_{Ed,Y|W}$ and $F_{\varepsilon|W;\theta}$, such that :*

$$E_p\left(1\left\{\left(Y,\underline{Ed},\overline{Ed}\right)\in J\circ G\left(\varepsilon\,|W=w\,;\theta\right)\right\}\right)=1,\ \ w\ a.e. \tag{1.7.8}$$

Replacing our previous definition of the identified set (1.7.5) in Definition 1 by (1.7.8), we can construct a confidence region as advised in section 1.7.1

First-step estimation of V

Suppose there exists an interval $[\underline{v};\overline{v}]$ such that:

$$\mathbb{P}\left(V \in [\underline{v};\overline{v}]\right) = 1 - \alpha_v \tag{1.7.9}$$

Note now that $\tilde{\pi} \geq \varepsilon$ if and only if $r \geq \varepsilon - r_0.v1\{v < 0\}$. Define $\tilde{\varepsilon}$ such that $\tilde{\varepsilon} = \varepsilon - r_0.v1\{v < 0\}$. We have that $\tilde{\varepsilon} \in [\varepsilon - r_0\overline{v}1\{v < 0\}; \varepsilon - r_0\underline{v}1\{v < 0\}]$ w.p. $1 - \alpha_v$.Define finally:

$$K : (\tilde{\varepsilon}\,|(\underline{v}1\{v < 0\}, \overline{v}1\{v < 0\})) \mapsto [\tilde{\varepsilon} + r_0\underline{v}1\{v < 0\}; \tilde{\varepsilon} + r_0\overline{v}1\{v < 0\}]$$

We have that: $\mathbb{P}\left(\varepsilon \in K\left(\tilde{\varepsilon}\,|(\underline{v}1\{v < 0\}, \overline{v}1\{v < 0\})\right)\right) = 1 - \alpha_v$. Applying again the composition theorem cited above, we can characterize the identified set as follow:

Proposition 3 (The identified set with incomplete observation of Ed and preestimation of V) *A parameter value θ belongs to the identified set, if and only if there exists a probability distribution p defined on $(\Omega, \mathcal{F}, \mathbb{P})$ with marginal $F^0_{Ed,Y|W}$ and $F_{\varepsilon|W;\theta}$, such that :*

$$E_p\left(1\left\{\left(Y, \underline{Ed}, \overline{Ed}\right) \in J \circ G \circ K\left(\tilde{\varepsilon}\,|W = w\,;\theta\right)\right\}\right) = 1 - \alpha_v, \;\; w \; a.e. \tag{1.7.10}$$

1.7.2 Sampling Methodology: Correction for biases induced by non random sampling and non-response

A natural problem impedes the use of traditional techniques for simple random sampling with a chain-referral sample. Because of the combined effect of peer recruitment and non-response, people reached by the survey might be different from those absent in the survey. Our approach to this problem is to assume enough structure on both the network and the recruitment behavior, so that characteristics of invitees are independent of characteristics of their host. Then, we can recover the selection probabilities of each individual in the sample. We then use an Horvitz-Thompson estimator to correct for selection biases and compute proportions in the population. First, we introduce some notation.

Notations

We suppose that the population of interest is a undirected[9] graph $\mathcal{E}$, given by a set of nodes N with label $\{1, 2, \ldots, N\}$ and values $\mathbf{w} = (\mathbf{w}_1, \ldots, \mathbf{w}_N)$ where $\mathbf{w}_i = (y_i, x_i)$ and an $N \times N$ matrix $\mathbf{L}$ indicating relationships or links between nodes. An element L_{ij} of $\mathbf{L}$ is one if there is a link from node i to node j and zero otherwise. We assume that $L_{ii} = 0$. L_{ij} determines the graph structure of population but will be observed only partially, even for the sampled units. We denote $D_i = \sum_j L_{ij}$, i's *d*egree, i.e. the number of people i is linked to in the population. A subset s is a subset of vertices and edges from the population. $s = \left(s^{(1)}, s^{(2)}\right)$, where $s^{(1)}$ are nodes, on which $\mathbf{w_i}$ is observed and $s^{(2)}$ are pairs of nodes, for which the value of the link L_{ij} is observed.

Sampling design

Start from the initial sample d_0, selected with probability π_0. Each unit is asked to complete a questionnaire and refer as many individuals as possible. The variable $\mathbf{w_i}$ and D_i are revealed by the respondent and also L_{ij} if j is invited by i. We will note $I_{ij} = 1$ if i invites j, and 0 if not. The researcher then send invitation to all the invitees of i. Each new respondent is asked to complete the same task: fill out a questionnaire and refer as many friends as possible. Recruitment occurs in waves and stops when invitees fail to complete the survey or invite other friends. The wave at which i is invited is the number of recruiters that separates him from the initial sample.

Assumptions

We detail here two sets of assumptions, the first set on the graph, which is assumed to be a realization of a stochastic process and the second set on the recruitment

[9]This hypothesis seems quite reasonable since invitation is restricted to *friends*. It amounts to say if individual i invites j, then j could have also invited i.

behavior, which is assume to exhibit independence with respect to our variable of interest.

Assumption 3 (Network) *We assume the following about the graph $\mathcal{E}$:*

1. *The graph $\mathcal{E}$ is the realization of a stochastic process $G(N,p)$ (Erdős-Rényi model), indexed by a parameter (N,p), where p is the probability that there exists an edge between any vertex i and vertex j.*
2. *The size of the graph N is known.*
3. *The graph is almost surely connected, i.e. $\exists\ \varepsilon > 0$ such that $\left(p > \dfrac{(1+\varepsilon)\log N}{N}\right)$.*
4. *A respondent reports accurately his/her degree.*

Assumption 4 (Recruitment behavior) *We assume the following about recruitment behavior of each person in the population of interest:*

1. *An host i invites at random an observed proportion of its network, V_i.*
2. *Non-response is random ("missing-at-random" data).*

The latter assumption together with the assumed stochastic model for the graph, ensures that we have independence between the characteristics of hosts and invitees. This assumption are not innocuous, but are the relatively standard in the literature on chain-referral sampling (see for example Heckathorn (1997) for a thorough discussion). In practice, we checked that the correlation between respectively the migration status, the educational attainment and degree of invitee and host is low.

We now turn to the computation of the inclusion probability of each individual. We will actually be interested in the probability of inclusion of an individual, conditional on the sampling process having reached wave k. The interested reader is strongly encouraged to read Thompson (2006) for a motivation and detailed exposition of the estimator.

Conditional inclusion probability

We introduce some more notations borrowed from Thompson (2006). Suppose that we have reached the k^{th} wave. And define:

- $\mathcal{S}$, the sample resulting from the procedure,
- s_k, the sample selected at step k,
- s_{ck}, the current sample, i.e.
$$s_{ck} = \bigcup_{j=0}^{k-1} s_j$$
- a_k, the current active set as subset of the current set consisting of individual making invitation.
- q_{ki}, the probability that i is selected at step k.

At step $k \geq 1$, we are interested in q_{ki}, the probability that $i \in s_k$, conditional on observation of a_k. We will denote $I_{ik} = 1$ if i is invited as opposed to $I_{ik} = 0$, if i is not invited. In addition, $I_{ijk} = 1$ will be the special case where i is invited by j. The event $\{i \in s_k\}$ happens if (1) i is invited, $I_{ik} = 1$, and (2) and $\{i$ answers$\}$. We can then write:

$$q_{ki} = P\left(\{i \text{ answers}\} \,|I_{ik} = 1, a_k\right) \times P\left(I_{ik} = 1 \,|a_k\right) \tag{1.7.11}$$

We treat separately each of the two terms in the right-hand side of the equation.

$P\left(\{i \text{ answers}\} \,|I_{ik} = 1, a_k\right)$ is the probability of response of unit i given that he has been invited by a member of the active set. By Assumption 4, this is independent of our variable of interest[10]. In practice, the response changed with time, decreasing

[10]If the response behavior is independent of the population characteristics and of the identity of the recruiter, the response rate in the sample can be used as an unbiased estimator. If it depends solely on characteristics of the recruiter, the empirical response rate conditional on recruiters characteristics provides a consistent estimate. The more challenging case is when non-response differs from one

sharply for the last few waves. We will then use the fraction of invitees answering the questionnaire at wave k, p_{a_k}, as a proxy for this conditional response probability.

$P(I_{ik} = 1 \,|a_k)$ is the probability that an individual would be invited given observation of the active set. By Assumption 4, i is chosen by $j \in a_k$ with probability V_j if $L_{ij} = 1$. We have then:

$$P(I_{ijk} = 1 \,|a_k) \quad = \quad P(L_{ij} = 1 \,|a_k) \,.V_j = p.V_j \tag{1.7.12}$$

where the second equality comes from Assumption 3.

Proposition 4 (Conditional inclusion probability) *Under Assumption 3 and 4, the conditional probability of inclusion of individual i at wave k is*

$$q_{ki}(p) = p_{a_k} \times \left(1 - \prod_{j \in a_k} (1 - p.V_j)\right) \tag{1.7.13}$$

Proposition 4 gives us an exact analytic expression for the conditional probability of inclusion.

Estimator of population proportions

If one is interested in a characteristic y of the population, Thompson advises the use of the following estimator:

$$\bar{y} = \frac{1}{N(K+1)} \left(\sum_{i=1}^{n_0} \frac{y_{i0}}{\pi_{i0}} + \sum_{k=1}^{K} \sum_{i \in s_{ck}} \frac{y_i}{q_{ki}(p)} \right) \tag{1.7.14}$$

and provides variance estimators. See section 4.6 and 5 in Thompson (2006).

group to another. In the case of sampling a network of migrant and non-migrant students, migrants might enjoy an easier access to Internet which makes them more willing to complete a 15 minutes questionnaire than non-migrants. In such case, assuming a similar non-response rate for migrants and non-migrants will overestimate the proportion of non-migrant participating. Comparing our estimates to existing institutional data, we do not detect such pattern

In (1.7.13), p is unknown. We appeal one more time to Assumption 3 which gives us an expression for p. Since the distribution of degrees in an Erdős-Rényi graph is a Poisson distribution with parameter Np, we have for D_i, the degree of individual i, $\mathbb{E}(D_i) = Np$. We obtain then a consistent estimator of p by solving for $\hat{p}$, the following moment condition:

$$N\hat{p} = \frac{1}{N(K+1}\left(\sum_{i=1}^{n_0}\frac{D_{i0}}{\pi_{i0}} + \sum_{k=1}^{K}\sum_{i\in s_{ck}}\frac{D_i}{q_{ki}(\hat{p})}\right) \tag{1.7.15}$$

To account for this sequential two-step estimation, one can use the bootstrap to retrieve variance estimators [11]. The next subsection show however how we use a bootstrap procedure to account for this additional uncertainty.

How does this affect our inference procedure?

We introduce weights for the individual in the sample. We further need to account for the fact that these weights are estimated from the degree observed. Finally, because of the relatively small sample size, we approximate the conditional joint distribution of migration and education by a parametric (logistic) distribution, $P\left((Y, Ed)|W; \rho\right)$. Recall that we observe for each individual (Y_i, Ed_i, W_i, D_i), we show here how to modify the inference procedure recommended by HMQ :

- Compute $\hat{p}$ as in equation (1.7.15), and a normalized weight for each individual i, defined by

$$\hat{\omega}_i = \frac{\eta_i}{\sum_{i\in S}\eta_i}$$

where

$$\eta_i = \frac{1}{N(K+1)}\left(\sum_{j=1}^{n_0}\frac{1\{i=j\}}{\pi_{i0}} + \sum_{k=1}^{K}\sum_{j\in s_{ck}}\frac{1\{i=j\}}{q_{ki}(\hat{p})}\right) \tag{1.7.16}$$

[11] see Rao and Wu (1988) for a methodology for resampling bootstrap procedure when the sample observations are drawn with different weights. Since the descriptive statistics are merely for indicative purpose, we abstract from the first-step estimation and assume that the weights are accurately estimated. See results in Appendix 1.7.3

- Estimate the parameter ρ, characterizing the logistic distribution, by minimizing the weighted (by $\hat{\omega}$) objective function. Let $\hat{\rho}$ be the solution to this minimization.

- For each bootstrap replication $b = 1, \ldots, B$:

  - draw n realizations of D^* from a Poisson distribution with parameter $N\hat{p}$, and n replications of (Y^b, Ed^b) from the conditional logistic $p$df $P(.|W;\rho)$.
  - Compute p^b as in (1.7.15), where D is replaced by D^b, and ω^b accordingly.
  - Estimate the parameter characterizing the logistic distribution, by minimizing the weighted (by ω^b) objective function. Let ρ^b be the solution to this minimization.
  - Use the empirical process $P\left(.|W;\rho^b\right) - P(.|W;\hat{\rho})$ in the procedure advised by HMQ to retrieve the generalized $(1-\alpha)$ of interest.

1.7.3 Additional information on the Survey and the dataset

The survey has been conducted from March 27, 2011 to May 8, 2011 under the title "Migration des jeunes Camerounais après le baacalauréat" (Migration of young Cameroonians after high-School). The online platform is accessible at the address: www.migration-cameroun.com. The population of interest was Cameroonian aged 18 or more, having completed secondary school by obtaining the "Baccalauréat". Our dataset comprises information on both migrants and non-migrants. The definition used for the questionnaire is the following: A migrant is an individual who has studied for more than 6 months in a foreign country. We retain as migrants in the calculation, only those having acquired one year or more of education in a foreign country.

We applied the methodology presented in section 1.4.2. Participants were compensated by being registered each week to a lottery with four prizes of equal value ($50 CAD). After 6 weeks, 418 respondents (1710 individuals have been invited to the

survey) provided information about (i) their education, (ii) their migration historic or plans, and the way migration is or would be financed; (iii) socio-economic characteristics of their parents and siblings and, finally, (iv) socio-economic characteristics of a member of the large family who could be designated as a helper in a (potential or effective) migration process. 16 individuals are excluded for the dataset because after we detected severe inconsistency in their answers. Table 1.3 presents the usual average statistics computed on the remaining sample. Because of the particularities of sampling procedure, Horvitz-Thompson estimators are used as unbiased estimators of the true population mean. Details of this adjustments and the estimators are presented in 1.4.2 and Appendix 1.7.2. The results of these adjustments are presented in Table 1.4 below. Note that we use, for the future parametric estimations in section 1.5.1, Weighted Maximum Likelihood Estimators (see for reference Cameron and Trivedi (2005), p.479).

For the inference procedure, we exclude from the analysis 44 individuals who obtained their Baccalauréat degree outside a reference period, namely before 1996 and after 2006. The main reason for the lower bound is that the Purchasing Parity Power has significantly changed in the previous period, due to a 50 percent devaluation of the Cameroonian currency in January 1995. Observations in this period are too few to allow for an efficient analysis. The upper bound excludes individuals for which we are unable to observe whether they have migrated or will migrate before achieving the minimum number of years to get a master degree. We further exclude 22 units who are currently attempting migration, for we cannot observe the outcome of this attempt.

Table 1.3: Descriptive Statistics

	Non Mig.		Mig.		Pop.	
Population	176		226		402	
	Aver.	s.d.	Aver.	s.d.	Aver.	s.d.
Female	0.38	0.49	0.47	0.50	0.43	0.50
Age	27.90	4.13	27.34	5.10	27.58	4.70
Age at Bac	19.31	2.12	18.70	1.75	18.97	1.94
Married	0.17	0.38	0.19	0.39	0.18	0.39
Residence Country						
Cameroon	0.89	0.32	0.19	0.39	0.50	0.50
Africa	0.02	0.13	0.15	0.36	0.09	0.29
France	0.01	0.08	0.19	0.39	0.11	0.31
Germany	0.02	0.13	0.06	0.24	0.04	0.20
North America	0.06	0.23	0.26	0.44	0.17	0.38
Other OECD	0.02	0.13	0.12	0.33	0.08	0.27
Other	0.00	0.00	0.03	0.17	0.02	0.13
Educational Attainment at the time of the survey						
Univ 0 - Baccalauréat	0.18	0.39	0.22	0.42	0.20	0.40
Univ 1 - Bachelor	0.65	0.48	0.36	0.48	0.49	0.50
Univ 2 - Master	0.14	0.34	0.35	0.48	0.25	0.44
Univ 3 - Doctorate	0.03	0.17	0.08	0.26	0.05	0.23
Completed study	0.49	0.50	0.38	0.49	0.43	0.50
Field of study						
Medical sciences	0.06	0.23	0.12	0.32	0.09	0.29
Economics	0.29	0.45	0.33	0.47	0.31	0.46
Science and Engineering	0.36	0.48	0.35	0.48	0.36	0.48

Continued on next page

Table 1.3 – *Continued from previous page*

	Non Mig.		Mig.		Pop.	
Population	176		226		402	
	Aver.	s.d.	Aver.	s.d.	Aver.	s.d.
Migration Country[2]						
Africa	0.02	0.13	0.31	0.46	0.18	0.39
France	0.10	0.30	0.23	0.42	0.17	0.38
Germany	0.01	0.08	0.09	0.28	0.05	0.22
North America	0.41	0.49	0.15	0.36	0.27	0.44
Other OECD	0.11	0.32	0.18	0.38	0.15	0.36
Other	0.00	0.00	0.04	0.20	0.02	0.15
Education attained at time of departure						
Univ 0 - Baccalaureat	-	-	0.48	0.50	-	-
Univ 1 - Bachelor	-	-	0.23	0.42	-	-
Univ 2 - Master	-	-	0.29	0.45	-	-
Univ 3 - Doctorate	-	-	0.00	0.07	-	-
Number of years of study in a foreign country						
0	-	-	0.12	0.33	-	-
1	-	-	0.16	0.37	-	-
2	-	-	0.17	0.38	-	-
3	-	-	0.18	0.39	-	-
4 and more	-	-	0.36	0.48	-	-
Paying for education in foreign country[3]						
Self	0.24	0.43	0.13	0.34	0.18	0.38
Father and/or mother	0.16	0.37	0.34	0.47	0.26	0.44
Helper	0.06	0.24	0.04	0.21	0.05	0.22
Shared	0.20	0.40	0.27	0.44	0.24	0.43
Grant	0.31	0.46	0.20	0.40	0.25	0.43

Continued on next page

Table 1.3 – *Continued from previous page*

	Non Mig.		Mig.		Pop.	
Population	176		226		402	
	Aver.	s.d.	Aver.	s.d.	Aver.	s.d.
Paying for education in Cameroon[4]						
Self	0.09	0.29	0.07	0.26	0.08	0.27
Father and/or mother	0.60	0.49	0.57	0.50	0.58	0.49
Helper	0.03	0.18	0.04	0.20	0.04	0.19
Shared	0.03	0.18	0.06	0.24	0.05	0.22
Grant	0.22	0.41	0.11	0.31	0.15	0.36
Mother						
Absent Mother	0.16	0.37	0.13	0.34	0.14	0.35
Level of Education						
Primary	0.47	0.50	0.33	0.47	0.39	0.49
Secondary	0.28	0.45	0.35	0.48	0.31	0.46
Tertiary (Univ 0 and 1)	0.18	0.39	0.24	0.43	0.22	0.42
Tertiary (Univ 2 and 3)	0.07	0.26	0.08	0.28	0.08	0.27
Have lived in foreign country	0.15	0.36	0.13	0.34	0.13	0.34
Father						
Absent Father	0.22	0.41	0.19	0.39	0.20	0.40
Level of Education						
Primary	0.39	0.49	0.29	0.45	0.33	0.47
Secondary	0.19	0.40	0.15	0.35	0.17	0.37
Tertiary (Univ 0 and 1)	0.22	0.41	0.25	0.44	0.24	0.43
Tertiary (Univ 2 and 3)	0.20	0.40	0.31	0.47	0.27	0.44
Have lived in foreign country	0.14	0.35	0.26	0.44	0.21	0.41

Continued on next page

Table 1.3 – *Continued from previous page*

	Non Mig.		Mig.		Pop.	
Population	176		226		402	
	Aver.	s.d.	Aver.	s.d.	Aver.	s.d.
Family Capital						
Owns a car	0.37	0.48	0.56	0.50	0.48	0.50
Owns a house	0.64	0.48	0.70	0.46	0.67	0.47
Owns a Field	0.50	0.50	0.59	0.49	0.55	0.50
Helper[5]						
Declare and Helper	0.52	0.52	0.55	0.55	0.53	0.53
Female	0.29	0.45	0.31	0.47	0.30	0.46
is a brother/sister	0.41	0.49	0.31	0.47	0.35	0.48
is an oncle/aunt	0.33	0.47	0.40	0.49	0.37	0.48
other link	0.26	0.44	0.29	0.46	0.28	0.45
Level of education						
Primary	0.07	0.25	0.06	0.23	0.06	0.24
Secondary	0.07	0.25	0.06	0.23	0.06	0.24
Tertiary (Univ 0 and 1)	0.45	0.50	0.30	0.46	0.36	0.48
Tertiary (Univ 2 and 3)	0.42	0.50	0.59	0.49	0.52	0.50
Have lived in foreign country	0.66	0.48	0.56	0.50	0.60	0.49
Lives in the migration country	0.40	0.49	0.50	0.50	0.46	0.50
Owns a car	0.36	0.48	0.48	0.50	0.43	0.50
Owns a house	0.33	0.47	0.44	0.50	0.39	0.49
Owns a Field	0.26	0.44	0.33	0.47	0.30	0.46

[1]These statistics are computed without re-weighting and represent the average and standard errors of the sample.

[2]For non-migrant, the country where they will attempt migration or would choose to migrate if they were given the opportunity to do so.

[3]For non-migrant, the way a temptative migration would be financed.

[4]Some individuals in the population did not study at all at the tertiary level in Cameroon.

[5]Statistics on the helper characteristics are computed conditional to the declaration of an helper.

Table 1.4: Weighted Descriptive Statistics for variables used in the study

	Non Mig.		Mig.		Pop.	
Population	176		226		402	
	Aver.	s.d.	Aver.	s.d.	Aver.	s.d.
Female	0.46	0.50	0.47	0.50	0.46	0.50
Age	25.86	3.97	24.59	4.16	25.23	4.12
Age at Bac	18.87	1.58	17.95	1.77	18.41	1.74
Educational Attainment at the time of the survey						
Univ 0 - Baccalaureat	0.24	0.42	0.41	0.49	0.32	0.47
Univ 1 - Bachelor	0.56	0.50	0.32	0.47	0.44	0.50
Univ 2 - Master or more	0.20	0.42	0.27	0.48	0.24	0.46
Migration Country[2]						
Africa	0.15	0.35	0.31	0.46	0.23	0.42
France	0.05	0.21	0.15	0.36	0.10	0.30
Germany	0.00	0.06	0.04	0.19	0.02	0.14
North America	0.33	0.47	0.06	0.24	0.20	0.40
Other OECD	0.06	0.23	0.40	0.49	0.23	0.42
Other	0.04	0.19	0.03	0.16	0.03	0.18
Paying for education in foreign country[3]						
Self	0.10	0.30	0.08	0.28	0.09	0.29
Father and/or mother	0.06	0.24	0.34	0.47	0.20	0.40
Helper	0.04	0.18	0.02	0.14	0.03	0.17
Shared	0.23	0.42	0.30	0.46	0.27	0.44
Grant	0.56	0.50	0.25	0.43	0.40	0.49

Continued on next page

Table 1.4 – *Continued from previous page*

	Non Mig.		Mig.		Pop.	
Population	176		226		402	
	Aver.	s.d.	Aver.	s.d.	Aver.	s.d.
Mother						
Absent Mother	0.19	0.39	0.09	0.28	0.14	0.35
Level of Education						
Primary	0.48	0.50	0.29	0.46	0.38	0.49
Secondary	0.31	0.46	0.14	0.34	0.22	0.42
Tertiary (Univ 0 and 1)	0.20	0.40	0.50	0.50	0.35	0.48
Tertiary (Univ 2 and 3)	0.02	0.13	0.07	0.25	0.04	0.20
Father						
Absent Father	0.24	0.43	0.12	0.33	0.18	0.39
Level of Education						
Primary	0.31	0.46	0.32	0.47	0.31	0.46
Secondary	0.11	0.31	0.06	0.24	0.08	0.28
Tertiary (Univ 0 and 1)	0.22	0.42	0.14	0.34	0.18	0.38
Tertiary (Univ 2 and 3)	0.36	0.48	0.48	0.50	0.42	0.49
Family Capital						
Owns a car	0.46	0.50	0.76	0.43	0.61	0.49
Helper[4]						
Declare and Helper	0.51	0.50	0.61	0.49	0.56	0.50
Female	0.15	0.36	0.13	0.33	0.14	0.34
is a brother/sister/oncle/aunt	0.89	0.31	0.76	0.43	0.82	0.39
Level of education						
Primary	0.30	0.46	0.02	0.14	0.15	0.35
Secondary	0.04	0.19	0.02	0.14	0.03	0.17
Tertiary (Univ 0 and 1)	0.47	0.50	0.40	0.49	0.43	0.50

Continued on next page

Table 1.4 – *Continued from previous page*

	Non Mig.		Mig.		Pop.	
Population	176		226		402	
	Aver.	s.d.	Aver.	s.d.	Aver.	s.d.
Tertiary (Univ 2 and 3)	0.19	0.39	0.56	0.50	0.39	0.49
Lives in the migration country	0.21	0.41	0.73	0.45	0.49	0.50
Owns a car	0.17	0.38	0.31	0.46	0.25	0.43

[1]These statistics are computed with re-weighting and represent estimates of the average and standard errors of the population. See Appendix 1.7.2 for details.

[2]For non-migrant, the country where they will attempt migration or would choose to migrate if they were given the opportunity to do so.

[3]For non-migrant, the way a student migration *w*ould be financed.

[4]Statistics on the helper characteristics are computed conditional to the declaration of an helper.

1.7.4 Intent to return

We complement our study with a logistic regression to explain the "intent to return" to Cameroon of the respondents who have completed their study abroad[12]. The proportion of returning migrants measures the extent to which outsourcing the tertiary education could be a successful strategy for a developing country. Citing Docquier in his discussion of Rosenzweig (2008):"The outsourcing of tertiary education partly reduces the fiscal burden supported by the country of origin. Nevertheless, migrants who stay permanently in the host country impose an important fiscal cost on their country of origin." Although, a potentially important source of beneficial brain circulation, few empirical studies attempt this exercise because of lack of data.

The dependent variable is measured by the answer to the question: "Do you intend to settle in Cameroon?". The option were "yes", "no", "not sure". When the answer is "not sure", we recode as "no" for individual who are still living outside Cameroon and "yes" for those who have already returned. The underlying assumption is that uncertainty about the establishment decision means that the individual will stay in the place where he lives. We analyze the effect of additional education, migration to OECD country, number of years of study, family capital and number of siblings living abroad. The results are reported in Table 1.5. In this analysis, the baseline category are students who have completed their schooling with a bachelor degree in a non-OECD country and with less than two years of study. Migrating to an OECD country has a significant strong negative effect on the intent to return. Estimates suggest that studying in the US, Canada or the UK, decreases by almost 90 percent the odds ratio of return relative to stay. The decrease for low tuition OECD countries is of the order of 70 percent. Once we control for the number of year of study and the capital of the family, the education does not have a significant effect on the intent of return. If anything, an individual with a higher degree express more willingness

[12]A probit regression returns similar log likelihood. We use the logistic specification to analyze the odds ratio

to return to Cameroon. One additional sibling living abroad makes an individual 50 percent more likely to return, while an increase of one percent of the capital of the family, results in a one percent increase in the odds of returning.

Table 1.5: Logit regression of the intent to return.

	Spec 1		Spec 2		Spec 3	
	B	exp(B)	B	exp(B)	B	exp(B)
Intercept	-0.66	0.52	-0.43	0.65	0.25	1.29
	(0.72)		(0.95)		(0.98)	
Education						
PhD	1.28*	3.60	1.15	3.17	1.08	2.93
	(0.79)		(1.13)		(1.10)	
Master	2.16***	8.65	1.72*	5.60	1.29	3.62
	(0.64)		(1.00)		(0.96)	
Destination Country						
OECD High	-2.01***	0.13	-2.25***	0.11	-2.19***	0.11
	(0.56)		(0.62)		(0.63)	
OECD low	-0.66	0.52	-1.41***	0.24	-1.29***	0.28
	0.46		0.41		0.41	
Other controls						
More than one year of study	0.17	1.19			-0.81*	0.45
	(0.48)				(0.45)	
Nbr Children Abroad			0.31**	1.36	0.39**	1.48
			(0.16)		(0.16)	
Family capital			0.55**	1.74	0.73***	2.07
			(0.27)		(0.25)	

Number of obs. = 84.

Estimation is made for standardized logistic distribution.

Individuals included have migrated and completed their studies abroad.

Standard deviation in parentheses.

(***) significant at 1%. (**) significant at 5%. (*) significant at 10%.

Chapter 2

Strategic interactions in student migration decisions

Abstract

The involvement of several economic agents in the process of international migration of students has been acknowledged by the literature, yet not studied with micro-level data. This paper analyzes the incentives for strategic agents to participate in the student migration investment decision, with a focus on the organization of the family unit. We model the migration decision as the result of a participation game between "extended" family members. We use sharp bounds for partially identified parameters of discrete games in complete information to conduct inference on the preference parameters. Our data are drawn from a survey specifically tailored for this study on Cameroonian students, migrants and non-migrants. The results suggest the benefit of this investment are equally shared within the family. Participation of parents to the migration process is driven to some extent by a concern for the child's interest, while helpers' participation is submitted to a set of social norms, namely kinship and gender obligations.

2.1 Introduction

The empirical and theoretical literature on student mobility suggests that students originating from developing countries are very likely to rely on the help of several members of their family, their community and of the diaspora during their migration. For example, the importance of the family unit is underlined by Boyd (1989). A survey conducted by Institute of International Education (IE) in 2006 reveals that the primary source of funding is "personal and family" for about 64 percent of foreign students. Additional support from a preexisting social network of migrants in the destination country has also recently been documented (Beine, Noel, and Ragot (2012), Perkins and Neumayer (2011)). However important, very little is actually known about the characteristics and incentives of these helpers. Indeed, covering travel and living expenses of a student in a foreign country, which might have higher per capita income, entails significant cost. Two competing explanations could be provided: either the investment generates for the investors economic and/or social benefits, or they are altruistic individuals (or "caring individuals" to use the terminology employed in Chiappori (1992)) who derive some benefit from the student utility. The objective of this paper is precisely to understand the incentives to participate in the migration investment of those agents who provide the most significant help.

We study a social interaction model describing the decision made by a group of individuals that we call the extended family, to invest in the education of a student in a foreign country. Our definition of the family unit includes three types of individuals, the parents, a child and a potential helper, who should be seen by the reader as a representative of a community to which the child belong. In addition to the incentives of investors, we investigate how this extended family unit organize itself to share profit and discourage free riding-behavior. We are precisely interested in the following questions: How does the family share the returns of investment? How does the private benefits compare with the shared return? Do family members exhibit some altruism? What kind of social obligations are at play? Among other results, we

find some evidence for altruism of the parents. We can also identify some social norms: egalitarianism in profit sharing, kinship and gender obligations influencing participation decisions.

Our structural model describes a participation game between family members. Within the family unit, parents and a representative of the extended family are seen as potential investors in the student migration and education. However, each of them have incentives not to participate in this costly decision process, as family members can enjoy private benefits generated by the investment independently of their participation decision. But failure to participate in the decision implies that the absent member cannot influence the investment decision of participants. Given individual payoff, the model predicts possibly multiple Nash Equilibrium (NE) for a participation game that take place between family members. Here arises an important challenge. The actual outcome (as observed by the researcher) should be interpreted as resulting from a selection mechanism on these multiple equilibria. However, this selection mechanism remains unobserved, while different equilibrium selection mechanisms could be reasonably invoked in our framework. Players might systematically choose the equilibrium maximizing total family's utility. They might, on the other hand, prefer an equilibrium involving the largest number of players to all others. Further, parents might systematically play the equilibrium that ensure migration. Or the selection mechanism mechanism could be a mixture of all the precedents. Having little guidance on this question in the context of student migration investment, we choose to make no assumption on the equilibrium selection mechanism. There is a cost associated to this choice: set identification rather than point identification for our structural parameters of family members' preferences. Building on recent contributions made by Beresteanu, Molchanov, and Molinari (2011) and Galichon and Henry (2011) (BMM-GH, hereafter), we derive sharps bounds for our structural parameters. We conduct inference in the context of our model, following the proposal in Section 1.3.

In addition to the related literature already cited in the previous chapter, this work is specifically related to the literature about family negotiations and migration networks. Since the seminal work of Becker (1973), numerous contributions have been made to address the criticisms regarding the shortcomings of the common-preferences models. This alternative strand of literature aims at explicitly taking into account the individualistic element of the household decision process (see for example Manser and Brown (1980), McElroy and Horney (1981), Lundberg and Pollak (1994)). Our model is closely linked to the one postulated in Engers and Stern (2002). The authors study a bargaining game for longterm care for elderly parents.

Although closely linked, the scope of interest differs between our paper and the current literature on migration network (for example Beine, Docquier, and Özden (2011)). The latter mainly refers to social ties between migrants and other migrants of similar origin within the destination country (Here the stock of migrants or ratio of migrants to the population is often used as proxy, see Dreher and Poutvaara (2011)). Our emphasis is on the parents and an agent that is reported by the student as potential significant contributor to the migration investment. There is therefore an overlapping of both populations of interest; however, none includes the other.

In section 2.2, we propose a participation game between family members to model the interactions during the decision process. Section 2.3 is devoted to the inference procedure. We derive sharp bounds on structural parameters. We briefly present in section 2.4 relevant summary statistics from the survey and the specification for our inference in section 2.5. Finally, we gather the results of the inference on the structural model and conclude with their discussion in Section 2.6.

2.2 A structural Model of Private Investment in Student Migration in a non-cooperative framework

We have argued that the family, as a whole, faces the issue of allowing a child of the family to study (often at university level) in a more developed country. Indeed, due to higher costs of living and potentially higher tuitions in foreign countries, the budget constraint plays an important role in international student migration. Family resources matter. According to data released by the Institute of International Education (IE) in 2006, the primary source of funding is "personal and family" for about 64 percent of foreign students. Our survey data show that close to three quarters of migrant rely on themselves or on family capital to finance their study abroad. It is also of interest to note that for a family, having a close connection in the diaspora enables her to send abroad more relatives in the future (Banerjee (1983)). Therefore, we find plausible to assume that this decision is taken at a family level.

Again, our definition of *family* encompasses more than the strict nuclear family (parents and children). We allow, as it is common in many developing countries, for a representative individual, outside of the nuclear family (often belonging to the extended family) to be part of the decision process. Boundaries of *family* are different from one culture to another. In several developing countries, the child is said to "belong to the whole community", understand village, extended family, etc. A related phenomenon is child fosterage (transfer or exchange of children among family) in Subsaharian-Africa (see for example Isiugo-Abanihe (1985)). To study individual incentives for participation in the sending of the student, we allow for a non-cooperative framework. This entails that agents maximizes their individual utility. Each family member observes the outcomes associated with all possible participation profile. He/she makes the decision to participate in order to maximize their utility in Nash Equilibrium (NE).

Before we proceed, note that student migration holds two additional characteristics[1]. First, the decision of migration results from an arbitrage between the schooling and employment opportunities available in the origin country and in the host country. The family will *simultaneously* choose an education level when deciding or not to finance a potential migration. Therefore, the decision is about an education and migration investment. Second, the amount of capital that the *family* can invest in the process is key to the decision process. Each additional participant contributes to increase the capital available for the investment. In the following, we shall distinguish participants from non-participants. We model interactions among agents using the following assumptions:

- An agent might be subject to a cost in case of non-participation which should be understood as some form of retaliation available for the participants against the non-participants. This cost reflects the existence of social and cultural obligations among family members.

- By participating in the decision process, the agent also endures a sunk-cost, whatever the outcome of the process. The reader should think of time spent in meetings, costs for information search, opportunity costs from loosing or delaying other existing investment opportunities, etc. These costs might be shared among participants, but do not affect non-participants.

- Successful investment ensures some direct returns to the family (financial such as remittances, higher salary for the child, or social such as benefit from extending the family network in the diaspora, etc.). These returns are divided among the participants.

- Successful investment also generates some private benefits for each family member, who receives some return proportional to the return on the investment. Together with the cost of participation, these "externalities" provide incentives

[1]For a thorough discussion of both points, see 1.2

for free-riding behavior, as some family members could, for example, enjoy the benefits of a close connection in the diaspora but avoid participation costs.

- As we are concerned with a measure of altruism of the family members, we consider that parents and helper might account for the child's utility function in their own utility function (see for example Li, Rosenzweig, and Zhang (2010)). They "care" for the utility of the child. However, if the utility of the child is proportional to the size of the profit generated by the investment, we will be unable to distinguish altruism from "externalities" described above. We need to assume that an investment which generates a positive profit for the family, also generates some positive private benefit for the child. The magnitude of this private benefit will be independent of the expected return on investment. In the case of migration for example, the student can enjoy additional benefit by entering the marriage market of the host country.

With respect to these features, we describe the different players, the available strategies, their valuations and pay-off in the following subsections. Note that we will model the decision process as a family meeting as in Engers and Stern (2002).

2.2.1 The players and strategies

We allow for three types of players. Parents, children and a representative individual for the extended family.

- The Child (indexed by 0) enters university and is candidate to migration. He has a set of characteristics X which will partially determine the benefits of the migration. To simplify, we consider that a candidate is *de facto* participating in the process.

- The parent or parents are treated together as one individual and referred to by the index 1. The parent decides whether or not to attend the meeting.

- As discussed previously, a representative individual of the community has the opportunity of attending the family meeting. She faces the same choice as the parent. We index her by the subscript 2.

2.2.2 Timing of the game

At the beginning of the period, each family member is informed that an education/migration investment opportunity exists, and that a meeting will take place to decide if this investment is profitable for the family. Each potential participant observes the child characteristics, the capital that will be available to invest for each possible set of participants, and the expected return on investment for all possible set of participants and all possible decisions taken during the meeting. Finally, each player can observe the size of the private benefits expected, the benefit sharing rule and the different costs associated to his/her attendance decision. Given this knowledge, parents and representative of the extended family decide simultaneously whether to attend or not (we will model the participation decision of family members as the result of a Nash Equilibrium in a strategic game). Then, the meeting takes place among the participants. They decide if they should undertake the investment, given the characteristics of the participants. Two decisions are taken by the attendants: they choose simultaneously the migration option and the education level that maximize the expected return of the participants. When they expect a negative net return, the participants decide collectively not to invest. Thus, no return (and no externality) is generated. On the other hand, if positive net returns are expected, the family agrees to share them among the participants according to a predetermined sharing rule, while the non-participants will only enjoy some private benefits. Whatever their choice of investment is, the participants incur a sunk-cost of attending the meeting.

2.2.3 Players' valuation and payoff

We introduce here some notation.

Participation decision. Regarding the participation game, we will denote by P a decision to attend the meeting, and by N a decision not to attend. Let A_i be the attendance decision of player $i \in \{1, 2\}$ and (A_1, A_2) the pair of participation decision made by the players. The pair (A_i, A_{-i}) denotes the attendance decision of player i, when the decision of the other player is A_{-i}.

Benefit and benefit sharing rule. Denote π the expected profit of the investment. $\alpha_i(A_i, A_{-i})$ will be the share of the benefit for individual $i \in \{0, 1, 2\}$, with $\sum \alpha_i = 1$. Note that $\alpha_i(N, A_{-i}) = 0$, i.e., if i decides not to participate, he/she receives no direct compensation. This expected profit will depend on the observable characteristics of the candidates and the capital gathered by the participants (which also depend on their observed characteristics).

Costs. By participating, i incurs the non-refundable cost $c_i(P, A_{-i})$, while non-participation induces for i the costs $c_i(N, A_{-i})$. Note that these costs depend on the participation of other family members. We expect for example, the sunk-cost of participation to be divided between players if both attend, while a player would be alone to bear the cost in case he/she is the only participant. Since non-participation cost are interpreted as retaliation from the other player, it may be that they differ for player i depending on the participation of player j.

Private (indirect) benefits. In addition, let $\beta_i\tilde{\pi}$ be the private, non-transferable return of individual i once investment is chosen by the participants. This will be the "indirect benefit" of member i. Finally, we allow for the possibility that a family member derives some indirect utility from the fact that the participant choose to invest, whatever is the expected return. Let γ_i be this additional utility that we interpret as a measure of altruism in the case of parents and helper. We expect for example altruistic parents to be concerned by an effective investment in the child as well as by the size of its return.

Suppose that i takes the participation of the other family member as given, A_{-i}. $\tilde{\pi}(N, A_{-i})$ is the net profit of the family resulting from the investment choice in the absence of i. $\tilde{\pi}(P, A_{-i})$ is, analogously, the net profit when i attends the meeting. By attending, i gets net payoff:

$$w_i^{P,A_{-i}} \equiv \left((\alpha_i + \beta_i)\tilde{\pi}(P, A_{-i}) + \gamma_i\right).1_{\{\tilde{\pi}(P,A_{-i})>0\}} - c_i(P, A_{-i})$$

And by not attending:

$$w_i^{N,A_{-i}} \equiv \left(\beta_i.\tilde{\pi}(N, A_{-i}) + \gamma_i\right).1_{\{\tilde{\pi}(N,A_{-i})>0\}} - c_i(N, A_{-i})$$

The second term in the payoff is a random benefit from participation, ψ_i which is 0 for family members who decide not to participate and distributed according to an absolutely continuous distribution $\nu(.|\theta)$, for each family member who participates. This variable summarizes the factors influencing the participation decision that are unobservable to the researcher, mainly the history of relationships within the family. All members observe the realizations of ψ, whereas the analyst only knows its distribution. The Payoff matrix can then be written as in the following matrix:

	Player 2 (Helper)	
Player 1 (Parent)	N	P
N	w_1^{NN}, w_2^{NN}	$w_1^{NP}, \psi_2 + w_2^{NP}$
P	$\psi_1 + w_1^{PN}, w_2^{PN}$	$\psi_1 + w_1^{PP}, \psi_2 + w_2^{PP}$

Example 2 *Suppose a family for which the investment in education/migration will yield the following net profit for the family:*

$$\tilde{\pi} = \begin{cases} \pi_{PP} & \textit{if both } 1 \textit{ and } 2 \textit{ participate,} \\ \pi_{NP} & \textit{if only one player, } 1 \textit{ or } 2, \textit{ participates,} \\ \pi_{NN} & \textit{otherwise.} \end{cases}$$

To simplify, $\pi_{NN} < 0$. The cost of the meeting is $c \geq 0$ and will be split equally among participants and the child. The returns on the investment are also split equally among participants and the child. There is no cost associated to a choice not to attend the meeting, but the non-participant will enjoy an externality equivalent to β times the total of the generated return. We will restrict $\beta \in [0; 1)$ and $\gamma_1 = \gamma_2 = 0$. Let ψ be the unobservable shocks with cumulative distribution F_ψ. The payoff matrix of the family, given unobservable shocks of participation (ψ_1, ψ_2) is:

	Player 2 (Helper)	
Player 1 (Parent)	N	P
N	$0, 0$	$\beta\pi_{NP}, \psi_2 + (\frac{1}{2} + \beta)\pi_{NP} - \frac{c}{2}$
P	$\psi_1 + (\frac{1}{2} + \beta)\pi_{PN} - \frac{c}{2}, \beta\pi_{NP}$	$\psi_1 + (\frac{1}{3} + \beta)\pi_{PN} - \frac{c}{3}$, $\psi_2 + (\frac{1}{3} + \beta)\pi_{PN} - \frac{c}{3}$

Here the result "no investment is made" occurs when the child receives no support from the family, i.e, if (N, N) is the NE (neither the parent, nor the Helper participates in the decision process).

The payoff matrix leads to the following Nash Equilibria in pure strategies:

$$\begin{aligned}
\text{- } (N,N) &\Leftrightarrow \psi_1 \leq w_1^{NN} - w_1^{PN} \text{ and } \psi_2 \leq w_2^{NN} - w_2^{NP} \\
\text{- } (P,P) &\Leftrightarrow \psi_1 \geq w_1^{NP} - w_1^{PP} \text{ and } \psi_2 \geq w_2^{PN} - w_2^{PP} \\
\text{- } (N,P) &\Leftrightarrow \psi_1 \leq w_1^{NP} - w_1^{PP} \text{ and } \psi_2 \geq w_2^{NN} - w_2^{NP} \\
\text{- } (P,N) &\Leftrightarrow \psi_1 \geq w_1^{NN} - w_1^{PN} \text{ and } \psi_2 \leq w_2^{PN} - w_2^{PP}
\end{aligned}$$

We can define an equilibrium correspondence $\psi \rightrightarrows G(\psi|x;\theta)$, which associate to every value of the latent variable a set of Nash equilibrium in pure strategies. Note that G is a correspondence and not a one-to-one mapping, since multiple Nash equilibrium can be predicted. This feature is the fundamental reason that justifies the inference methodology used in this paper. More on this in the next section.

This correspondence depends on the rankings of the terms $w_i^r - w_i^s$, $i \in \{1,2\}$ and $r,s \in \{NN, NP, PN, PP\}$. Although we will restrict ourselves to pure strategy NE, it can also be shown that there exists a unique Nash Equilibrium in mixed strategies as follows. A mixed profile $(\eta_1 P + (1-\eta_1)N; \eta_2 P + (1-\eta_2 N))$ is a Nash equilibrium if and only if

$$\eta_1 = \frac{\psi_2 - \left(w_2^{NN} - w_2^{NP}\right)}{(w_2^{PN} - w_2^{PP}) - (w_2^{NN} - w_2^{NP})} \text{ and } \eta_2 = \frac{\psi_1 - \left(w_1^{NN} - w_1^{PN}\right)}{(w_1^{NP} - w_1^{PP}) - (w_1^{NN} - w_1^{PN})},$$

the denominators are non zero and

$$\min\left\{w_1^{NN} - w_1^{PN}; w_1^{NP} - w_1^{PP}\right\} < \psi_1 < \max\left\{w_1^{NN} - w_1^{PN}; w_1^{NP} - w_1^{PP}\right\}$$
$$\min\left\{w_2^{NN} - w_2^{NP}; w_2^{PN} - w^{PP}\right\} < \psi_2 < \max\left\{w_2^{NN} - w_2^{NP}; w_2^{PN} - w_2^{PP}\right\}$$

The equilibrium correspondence can therefore be extended to account for these mixed strategies (as in Chapter 3).

Example 2 continued The NE for the game describe earlier are characterized by:

$$\begin{aligned}
\text{- } (N,N) &\Leftrightarrow \psi_1 \leq \frac{c}{2} - (\frac{1}{2} + \beta)\pi_{NP} \text{ and } \psi_2 \leq \frac{c}{2} - (\frac{1}{2} + \beta)\pi_{NP} \\
\text{- } (P,P) &\Leftrightarrow \psi_1 \geq \frac{c}{3} - (\frac{1}{3} + \beta)\pi_{PP} + \beta\pi_{NP} \text{ and } \psi_2 \geq \frac{c}{3} - (\frac{1}{3} + \beta)\pi_{PP} + \beta\pi_{NP} \\
\text{- } (N,P) &\Leftrightarrow \psi_1 \leq \frac{c}{3} - (\frac{1}{3} + \beta)\pi_{PP} + \beta\pi_{NP} \text{ and } \psi_2 \geq \frac{c}{2} - (\frac{1}{2} + \beta)\pi_{NP} \\
\text{- } (P,N) &\Leftrightarrow \psi_1 \geq \frac{c}{2} - (\frac{1}{2} + \beta)\pi_{NP} \text{ and } \psi_2 \leq \frac{c}{3} - (\frac{1}{3} + \beta)\pi_{PP} + \beta\pi_{NP}
\end{aligned}$$

In particular, (NP, PN) is a (multiple) NE in pure strategy
for $\psi \in \left[\frac{c}{2} - (\frac{1}{2} + \beta)\pi_{NP}; \frac{c}{3} - (\frac{1}{3} + \beta)\pi_{PP} + \beta\pi_{NP}\right]^2$ and

$$\eta_i = \frac{\left(\psi_{(1-i)} - \frac{c}{2} + (\frac{1}{2} + \beta)\pi_{NP}\right)}{(\beta\pi_{NP} - c/6) + (\frac{1}{3} + \beta)\pi_{PP} - (\frac{1}{2} + \beta)\pi_{NP}}$$

characterizes the NE in mixed strategy under the above existence conditions.

2.3 Inference procedure

In this section, we are interested in the construction of a confidence region for the structural parameters of the non-cooperative investment model described above. In

particular, in example 2, we would wish to conduct inference on β, c and the parameters of the distribution of ψ. However, as evidenced by the example, multiple equilibria can easily arise as predictions of the model. But in the data, we always observe a single outcome. Unless we are willing to make further assumption on the equilibrium selection mechanism, identification of the structural parameter is not guaranteed. Berry and Tamer (2006) and Ackerberg, Benkard, Berry, and Pakes (2007) give an account of the various way this identification problem was approached in the literature. Identification of structural parameters is usually achieved through equilibrium refinements, shape restrictions, informational assumption or the specification of equilibrium selection mechanism. An alternative approach is to characterize a set of compatible parameters rather than a point. An idea explored by Andrews, Berry, and Jia (2003) in the context of oligopoly entry is to base inference purely on the identified features of the models with multiple equilibria, which are sets of values rather than a single value of the structural parameter vector. Without imposing further assumption on strategic games of interaction with discrete set of strategies, BMM-GH provide sharp bounds to characterize the set of all parameters, the identified set, for which there is a selection mechanism such that the observed outcomes are compatible with the predicted equilibrium. All parameters in this set are observationally equivalent. In the following, we first present those bounds in the context of example 2. We refer the reader to GH for mathematical proofs. Inference in this context is conducted following the proposal of in Section 1.3. Then, in section 2.3.2, we address the issue that the return of the investment is not observed in the data, although we observe determinants of this return.

2.3.1 Sharp bounds

We use the sharp bounds provided by GH to conduct inference on the structural parameters of the model described in section 2.2. A rigorous characterization of the identified set can be found in their Theorem 2. Here, we use our running example to

give a flavor of the type of bounds we derive.

Example 2 continued Consider only the NE in pure strategy and the case where $\beta > c/12$. We will denote θ the vector which collect the parameters of interest. The set of (possibly multiple) NE predicted by the model consists of

$$(N,N),(N,P),(P,N),(P,P) \text{ and } (\{(P,N);(N,P)\})$$

. The sharp bounds induced by Theorem 2 of GH give:

$$\begin{aligned}
\mathbb{P}(N,N) &\leq F_\psi\left(\psi_1 < \frac{c}{2} - (\frac{1}{2}+\beta)\pi_{NP}, \psi_2 < \frac{c}{2} - (\frac{1}{2}+\beta)\pi_{NP}\right) \\
&\equiv \mathcal{L}((N,N);\theta) \\
\mathbb{P}(N,P) &\leq F_\psi\left(\psi_1 < \frac{c}{3} - (\frac{1}{3}+\beta)\pi_{PP} + \beta\pi_{NP}, \psi_2 > \frac{c}{2} - (\frac{1}{2}+\beta)\pi_{NP}\right) \\
&\equiv \mathcal{L}((N,P);\theta) \\
\mathbb{P}(P,N) &\leq F_\psi\left(\psi_1 > \frac{c}{2} - (\frac{1}{2}+\beta)\pi_{NP}, \psi_2 < \frac{c}{3} - (\frac{1}{3}+\beta)\pi_{PP} + \beta\pi_{NP},\right) \\
&\equiv \mathcal{L}((P,N);\theta) \qquad (2.3.1) \\
\mathbb{P}(P,P) &\leq F_\psi\left(\psi_1 > \frac{c}{3} - (\frac{1}{3}+\beta)\pi_{PP} + \beta\pi_{NP}, \psi_2 > \frac{c}{3} - (\frac{1}{3}+\beta)\pi_{PP} + \beta\pi_{NP}\right) \\
&\equiv \mathcal{L}((P,P);\theta) \\
\mathbb{P}(P,N) + \mathbb{P}(N,P) &\leq 1 - F_\psi\left(\psi_1 < \frac{c}{2} - (\frac{1}{2}+\beta)\pi_{NP}, \psi_2 < \frac{c}{2} - (\frac{1}{2}+\beta)\pi_{NP}\right) \\
&\quad - F_\psi\left(\psi_1 > \frac{c}{3} - (\frac{1}{3}+\beta)\pi_{PP} + \beta\pi_{NP}, \psi_2 > \frac{c}{3} - (\frac{1}{3}+\beta)\pi_{PP} + \beta\pi_{NP}\right) \\
&\equiv \mathcal{L}((\{(P,N);(N,P)\});\theta)
\end{aligned}$$

The terms in the left-hand side are derived from the observable outcome of the game. Note that they do not include the parameter θ. The terms in the right-hand side are derived from the cumulative distribution of the latent variable ψ. These inequalities can be understood, heuristically, in a multiple equilibria framework. They mean that, if a model is compatible with the data, the probability that we observe a given outcome, cannot be greater than the probability that the model predicts at least one of the (possibly multiple) equilibria where this outcome is part of. For example, (N,P) can only be observed if $\psi_1 < \frac{c}{3} - (\frac{1}{3}+\beta)\pi_{PP} + \beta\pi_{NP}$ and $\psi_2 > \frac{c}{2} - (\frac{1}{2}+\beta)\pi_{NP}$. But the latter is true for: $\psi \in \left[\frac{c}{2} - (\frac{1}{2}+\beta)\pi_{NP}; \frac{c}{3} - (\frac{1}{3}+\beta)\pi_{PP} + \beta\pi_{NP}\right]^2$, in which case

the model predicted also that (P, N) is possible outcome. In other words, we cannot observe (N, P) more often than the model predicted (N, P) as a single equilibrium or as part of the multiple equilibrium $(\{(P, N); (N, P)\})$.

Θ_I is defined as the set of parameters for which the above inequalities are true. These bounds which can easily be derived as necessary conditions, are also sufficient, hence the term "sharp" bounds.

2.3.2 Incomplete observation of the return of migration

We assumed in section 2.3.1 that $\tilde{\pi}$ the expected benefit of the investment, was observable for a given family. However, this information is absent from the dataset. We use the same idea as in in Section 1.3, where the capital of the family is missing to tackle this problem. Suppose that we can bound $\tilde{\pi}$ with a given probability, so that there exit an interval $\left[\underline{\tilde{\pi}}; \overline{\tilde{\pi}}\right]$ which containing $\tilde{\pi}$ with a given probability. Once this interval constructed, the problem collapses to a problem where the covariates are defined by an interval rather than a point. By appealing to the composition theorem from Galichon and Henry (2006a) (Theorem 1), we can redefine our identified set accordingly and propose a valid confidence region, following the same procedure. We illustrate in example 2 how the bounds are changed.

Example 2 continued Suppose now that the investment in migration will yield the following net profit for the family:

$$\tilde{\pi} = \begin{cases} \pi_{PP} & \text{if both 1 and 2 participate,} \\ \pi_{NP} & \text{if only one player, 1 or 2, participates,} \\ \pi_{NN} & \text{otherwise.} \end{cases}$$

To simplify, $\pi_{NN} < 0$. However, we are unable to observe $\tilde{\pi}$, but we know that with

a given probability $(1-\alpha_\pi)$:

$$\tilde{\pi} \in \begin{cases} [\underline{\pi}_{PP}; \overline{\pi}_{PP}] & \text{if both 1 and 2 participate,} \\ [\underline{\pi}_{NP}; \overline{\pi}_{NP}] & \text{if only one player, 1 or 2, participates,} \\ \pi_{NN} < 0 & \text{otherwise.} \end{cases}$$

The central terms in equations (2.3.1) will become:

$$\begin{aligned}
&(i) \qquad F_\psi\left(\psi_1 < \tfrac{c}{2} - (\tfrac{1}{2}+\beta)\underline{\pi}_{NP}, \psi_2 < \tfrac{c}{2} - (\tfrac{1}{2}+\beta)\underline{\pi}_{NP}\right) \\
&(ii) \qquad F_\psi\left(\psi_1 < \tfrac{c}{3} - (\tfrac{1}{3}+\beta)\underline{\pi}_{PP} + \beta\overline{\pi}_{NP}, \psi_2 > \tfrac{c}{2} - (\tfrac{1}{2}+\beta)\overline{\pi}_{NP}\right) \\
&(iii) \qquad F_\psi\left(\psi_1 > \tfrac{c}{2} - (\tfrac{1}{2}+\beta)\overline{\pi}_{NP}, \psi_2 < \tfrac{c}{3} - (\tfrac{1}{3}+\beta)\underline{\pi}_{PP} + \beta\overline{\pi}_{NP}\right) \\
&(iv) \qquad F_\psi\left(\psi_1 > \beta\underline{\pi}_{NP} + \tfrac{c-\overline{\pi}_{PP}}{3}, \psi_2 > \tfrac{c}{3} - (\tfrac{1}{3}+\beta)\overline{\pi}_{PP} + \beta\underline{\pi}_{NP}\right) \\
&(v) \qquad 1 - F_\psi\left(\psi_1 < \tfrac{c}{2} - (\tfrac{1}{2}+\beta)\overline{\pi}_{NP}, \psi_2 < \tfrac{c}{2} - (\tfrac{1}{2}+\beta)\overline{\pi}_{NP}\right) \\
&\qquad\quad - F_\psi\left(\psi_1 > \beta\overline{\pi}_{NP} + \tfrac{c-\underline{\pi}_{PP}}{3}, \psi_2 > \tfrac{c}{3} - (\tfrac{1}{3}+\beta)\underline{\pi}_{PP} + \beta\overline{\pi}_{NP}\right)
\end{aligned}$$

A parameter θ will belong to our confidence region if it satisfies the transformed inequality with probability $(1-\alpha_\pi)$.

The challenge now is to construct such an interval. Section 2.5.1 details the procedure.

2.4 Data

For the purpose of the study, a survey has been conducted on the population of Cameroonian, aged 18 or more, having completed secondary school by obtaining the "Baccalauréat". The data show that the family is greatly involved in financing the costs of studies (local or foreign). In more than half of the cases, these expenses are borne (or expected to be) by the family. There is however a stark contrast between investments in local and foreign studies when it comes to the identity of the payers. The costs of studies in Cameroon are shared in only 5% of families while for education abroad, they are shared between family members in 40% of the cases. It is therefore of great interest to understand how families organize to realize this investment.

We define here the concept of participation in the migration decision. During the survey, individuals where asked if they were migrants, if they had ever attempted migration or had never attempted migration. In each of the cases, they were asked to identify the person who would (or did) pay for the expenses related to the migration. They were also asked to identify (if applicable) a helper who could (or did) help in the process of migration, either with financial or material assistance. We estimate that 56% of the respondants have a potential helper. We therefore distinguish families with a helper from families without a helper. A helper or a parent is said to participate to the migration process if he pays (alone or with other people) the expenses related to the migration.

We will only consider families where at least one of the parents (father or mother) is present. Within 139 families without helper, parents participate in 52.52% of the cases. When both parent and helper are present (174 families), we observe the outcomes: no participation (N, N) in 24.71% of families, participation of parent alone (P, N) in 26.44% of families, participation of the helper alone (N, P) in only 8.62% of observations, and both participation (P, P) in 40.23% of observations. Table 2.1 summarizes the main characteristics of the helper. The "typical" helper is a male, an uncle or a brother who have a university degree. We refer the reader to Section 1.7 for a thorough discussion of the dataset and estimators of averages in the population.

2.5 Specification

2.5.1 Specification for the First-step estimation of the return to migration

We propose to use an ordered probit on the joint decision of education and migration to preestimate the return on investment. We assume two types of education (Masters, for those whose expected education is higher or equivalent to a Masters degree, No

Table 2.1: Characteristics of the helper in proportions of the population of migrants, non-migrants, and of the total population.

Helper	Non-Mig.		Mig.		Total	
	Aver.	s.d.	Aver.	s.d.	Aver.	s.d.
Declare and Helper	0.51	0.50	0.61	0.49	0.56	0.50
Female	0.15	0.36	0.13	0.33	0.14	0.34
is a brother/sister/oncle/aunt	0.89	0.31	0.76	0.43	0.82	0.39
Level of education						
Primary	0.30	0.46	0.02	0.14	0.15	0.35
Secondary	0.04	0.19	0.02	0.14	0.03	0.17
Tertiary (Univ 0 and 1)	0.47	0.50	0.40	0.49	0.43	0.50
Tertiary (Univ 2 and 3)	0.19	0.39	0.56	0.50	0.39	0.49
Lives in the migration country	0.21	0.41	0.73	0.45	0.49	0.50
Owns a car	0.17	0.38	0.31	0.46	0.25	0.43

Number of observations = 174.

These statistics are computed with re-weighting and represent estimates of the average and standard errors of the population. See Appendix 1.7.2 for details.

Masters otherwise) and the outcome migration or non-migration. Under the joint assumption of positive return of schooling and migration, we use the following ordering of alternatives: (1) non-migration without a Masters, (2) non-migration with a Masters, (3) migration without a Masters, (4) migration with a Masters, the first option being the reference option. We then construct bounds on this return using the corresponding confidence regions. The ordering of the alternatives is reasonable given the assumption of a positive return on education, that is however smaller in the origin country than the return of migration. Reports from institutional data (as in Njike Njikam, Lontchi Tchoffo, and Fotzeu Mwaffo (2005)) show that the country of interest in this study, Cameroon, exhibits such characteristics.

Note that this pre-estimation step is up-to a scale parameter that we treat as a nuisance parameter. To avoid unnecessary complication, we will use the mormalization that the variance of the unobservable private costs of participation is the same as the variance of the decision shifter affecting the choice of migration and education. As the latter variance will be normalized to one in the following, this choice leads to interpret the magnitude of preferences' parameters as the reponse of the utility of agents to a variation of the covariates equivalent to one unit of standard deviation of the return of investment. This assumption is not completely innocuous; it can be easily shown that this nuisance parameter affects the magnitude of the parameters of interest α and β, however not their sign. Caution is therefore warranted in interpreting the magnitude of β_1 and β_2 in Table 2.2.

The explanatory variables for our probit estimation include the characteristics of the child (age at which the child passed the state exam "Baccalauréat" and the results at this exam, the presence of a scholarship, whether the migration country has high tuition) and characteristics of the family (the maximum education in the family, the presence of a helper in the migration country, Numbers of cars owned by the family). This variables where found to explain the return on migration and education in Section 1.5. Once the regression is performed, we obtain prediction of the return of migration in absence (or in presence) of an individual, by changing accordingly the value of the covariates. We then construct confidence interval of the benefit $\tilde{\pi}$ for each attendance profile.

2.5.2 Specification of the players' utlity functions

Recall that our parameters of interest are:

- α_i, the share of the profit received by the player i. In the following, we will explore different sharing rules: (1) a "single-preference" sharing rule where all profits go to the child, (2) an egalitarian sharing rule where the profit are equally

divided among the players, (3) a "fair" sharing rule where the profits are divided according to the shapley value.

- $\beta_i \in [0;1]$, the proportion of the return that individual i receives as indirect benefits of the investment.

- γ_i measures a utility enjoyed by individual i from the fact that the family realize a successful investment and independant of the size of this return. We wil interpret this parameter as a measure of altruism.

- c_i, is the cost of attendance, net of retaliation costs. We will study how this cost varies with characteristics of the player. We postulate:

$$\begin{aligned} c_1 &= c_{01} \qquad \text{(Cost for parents)} \\ c_2 &= c_{02} + c_F FAM + c_D DIA + c_G GENDER \qquad \text{(Cost for helper)} \end{aligned}$$

  where FAM, DIA and $GENDER$ are dummy variables which equals one respectively when the individual is an oncle/aunt or a brother/sister, when the individual lives outside Cameroon (i.e. in the diaspora), when the individual is a female.

- We assume that ψ_i, the unobserved benefits of participation of player i, has a logit distribution, with mean zero and variance $\sigma_\psi^2 = 1$.

We discuss our results in the next section.

2.6 Results and discussion

We are first interested in the sharing of profits. Our dataset readily rejects the single-preference and the fair sharing rule. The confidence region is non empty only for the egalitarian sharing rule. This suggests the type of social norms that govern the decision process. Regardless of the amount each individual invests in the migration,

the common returns seem to be split in equal proportion among all members of the family participating in the decision.

For an egalitarian sharing rule, we construct a confidence region under the normalization that the variance of the unobservable private costs of participation is the same as the variance of the decision shifter affecting the choice of migration and education. We fail to reject the hypothesis that the helper derives no additional utility from the realization of the migration project ($\gamma_2 = 0$). We then construct a constrained confidence region under this additional restriction. The ranges of the remaining parameters of interest are displayed in Table 2.2. With regard to the utility

Table 2.2: Range of preference parameters

		Min	Max
Parent	β_1	0.78	1.00
	γ_1	0.50	5.00
	c_{01}	0.86	1.15
Helper	β_2	0.00	1.00
	γ_2	0.00	0.00
	c_{02}	0.12	3.06
	c_F	-2.52	-0.58
	c_G	-3.50	-0.67
	c_D	-0.40	5.00

Results are displayed for $\sigma = 1$,$\gamma_2 = 0$
and an egalitarian sharing rule

function of the parents, the table shows that β_1 is relatively close to 1. The private utility for the parent appears to be quite high and to dominate the share of return that he receives while participating in decision process. Furthermore, unlike for the helper, this concern for the familial return appears to be related also to the realization of the investment ($\tilde{\pi} > 0$), and not only to the size of the return, suggesting some

degree of altruism. Indeed, γ_1 can be quite large, even when the size of the return remains modest, to the point where this additional utility can compensate for the costs of attendance.

With regard to the utility function of the helpers, the data are uninformative on the size of the private benefit enjoyed by the helper. In particular, it remains unclear whether the parent can exclude free-riders from any profit. However, we can identify the sign and bound the magnitude of the participation costs. As is the case for parents, these constant costs are significantly positive for helpers, a fact that discourages participation to the decision process. However, this disutility can be compensated for, when the helper is a "close family member", understand brother/sister of the child or brother/sister of the parent. We interpret this effect as kinship obligation. Indeed, the negative sign of c_F implies that the retaliation costs that a family member suffers in case of non-participation are higher than his/her cost of attendance. Similar interpretations can be made for the parameter c_G: the retaliation costs in case of non-participation to the decision process will be higher for a male representative of the extended family than for a female. In other words, social norms are more stringent for males than for females when it comes to providing help in the migration process. Finally, we cannot reject the hypothesis $H_0 : c_D = 0$, however, the confidence region shows a tilt toward positive values of the parameter, suggesting models with higher cost of participation for helpers in the diaspora. Two (possibly complementary) interpretations are available: (1) it may be that the means of retaliation are geographically limited and become less effective for individuals living abroad, (2) or that the logistic cost of participation are relatively higher for those not residing in the origin country.

2.7 Conclusion

To sum up, the extended family involved in the migration process should not be viewed as an homogeneous entity where individuals maximize the sole familial interest, nor should it be viewed as a compound of egoistic individuals. The results suggest that participation to the migration process is driven to some extent by a concern for the familial interest, while family interaction and helpers participation are submitted to a set of social norms.

Chapter 3

Combinatorial bootstrap inference in partially identified incomplete structural models

with Marc Henry and Maurice Queyranne.

Abstract

We propose a computationally feasible inference method in finite games of complete information. Galichon and Henry (2011) and Beresteanu, Molchanov, and Molinari (2011) show that such models are equivalent to a collection of moment inequalities that increases exponentially with the number of discrete outcomes. We propose an equivalent characterization based on classical combinatorial optimization methods that alleviates this computational burden and allows the construction of confidence regions with an efficient combinatorial bootstrap procedure that runs in linear computing time. The method can also be applied to the empirical analysis of cooperative and noncooperative games, instrumental variable models of discrete choice and revealed preference analysis. We propose an application to the determinants of long term elderly care choices.

3.1 Introduction

With the conjoined advent of powerful computing capabilities and rich data sets, the empirical evaluation of complex structural models with equilibrium data is becoming prevalent, particularly in the analysis of social networks and industrial organization. However, in such models, multiple equilibria are the norm rather than the exception. Though multiplicity of equilibria and identifiability of the model's structural parameters are conceptually distinct, the former often leads to a failure of the latter, thereby invalidating traditional inference methods. This is generally remedied by imposing additional assumptions to achieve identification, such as imposing an equilibrium selection mechanism or a refinement of the equilibrium concept. Manski (1993) and Jovanovic (1989) were among the first to advocate a new inference approach that dispenses with identification assumptions and delivers confidence regions for partially identified structural parameters. A large literature has developed on the general problem of inference on partially identified parameters defined as minimizers of objective functions or more specifically as solutions to moment inequality restrictions, following the seminal work of Chernozhukov, Hong, and Tamer (2007).

In structural estimation using equilibrium conditions, the partial identification approach was initially applied, as in Haile and Tamer (2003), to achieve simple and robust inference from implications of the model in the form of a small number of moment inequalities. This partial identification approach was applied to inference in games by Andrews, Berry, and Jia (2003), Pakes, Porter, Ho, and Ishii (2004), Ciliberto and Tamer (2009), Jia (2008) among others. However, this approach brings only part of the empirical content of the model to bear on the estimation, resulting in unnecessary loss of informativeness. In models with multiple equilibria and no additional prior information, nothing is known of the equilibrium selection mechanism. If a particular equilibrium selection mechanism is posited, the model likelihood can be derived and inference based on it. Jovanovic (1989) characterizes compatibility of an economic structure with the true data generating process as the existence of some

(unknown) equilibrium selection mechanism, for which the likelihood is equal to the true data generating mechanism. Berry and Tamer (2006) define the identified set as the collection of structural parameter values for which the structure is compatible with the data generating mechanism in the sense of Jovanovic (1989). This definition of the identified set is not directly conducive to inference, as it involves an infinite dimensional (nuisance) parameter (the equilibrium selection mechanism). However, in the case of finite non cooperative games of complete information, Galichon and Henry (2011) and Beresteanu, Molchanov, and Molinari (2011) show equivalence of the Jovanovic (1989) definition with a system of inequalities. Hence, they show that the empirical content of such models is characterized by a finite collection of moment inequalities.

A large literature has developed on inference in moment inequality models since the seminal contribution of Chernozhukov, Hong, and Tamer (2007). We discuss and review it in Section 3.4. However, a major challenge in the framework of this paper is that the number of inequalities characterizing the empirical content of the model grows exponentially with the number of equilibrium strategy profiles. Hence the combinatorial optimization approach that we propose in this paper is to the best of our knowledge the only computationally feasible inference procedure for empirically relevant incomplete economic structures. The growing literature on "inference with many moment inequalities" addresses theoretical issues relating to the case, where the number of inequalities grows with sample size and does not alleviate the computational burden mentioned here. This problem of exponential complexity goes a long way towards explaining the dearth of empirical studies using partial identification in such models. However, abandoning this partial identification approach would mean abandoning robust inference not only in non cooperative games of perfect information but also in large classes of models that share exactly the same feature, and fall into the framework of this paper. They include cooperative games, such as matching games and network formation games, revealed preference analysis of spacial preferences and matching markets and instrumental variable models of discrete choice.

The objective of this paper is to propose a combinatorial solution to this problem, where the number of inequality restrictions grows exponentially with the number of strategy profiles or discrete outcomes. Ekeland, Galichon, and Henry (2010) have shown that generic partial identification problems can be formulated as optimal transportation problems. Developing ideas in Galichon and Henry (2011), we exploit the special structure of discrete choice problems and show that correct specification can be formulated as a problem of maximizing flow through a network, and that the identified set can be obtained from the Max-Flow Min-Cut Theorem. The dual problems of maximizing flow through a network and finding a minimum capacity cut are classics in combinatorial optimization and operations research, with applications in many areas such as traffic, communications, routing and scheduling; see, for example Schrijver (2004) for the theory and history, and Ahuja, Magnanti, and Orlin (1993) for numerous applications. To our knowledge this is the first application of the Max-Flow Min-Cut Theorem to statistical inference for equilibrium models. We apply this powerful combinatorial method to the problem of constructing confidence regions for structural parameters. We construct a functional quantile for the bootstrap process using a linear computing time algorithm and replace the unknown empirical process by this quantile in the system of moment inequalities to obtain the least relaxation of the moment inequalities, hence maximum informativeness, while controlling the confidence level of the covering region. Since the procedure involves bootstrapping the empirical process only, it does not suffer from the problems of bootstrap validity in partially identified models described in Chernozhukov, Hong, and Tamer (2007) and Bugni (2010). We illustrate and assess our procedure on a very simple full information game with 2 players and 3 strategies, easily derived equilibria and yet a large number of inequalities to characterize its empirical content (namely 127). We simulate the game under a variety of parameter values and assumptions on the data generating process and with explanatory variables. Finally, we illustrate the approach, the procedure and the interpretation of results on an application to the determinants of long term elderly care choices of American families.

In summary, the main contributions of this paper are as follows:

1. We present a unified approach to inference in incomplete structural models.

2. We provide a simplified and insightful new proof for a characterization of the identified set.

3. We present a computationally efficient, combinatorial procedure that allows feasible inference in empirically relevant incomplete structural models. We demonstrate its practical efficiency in extensive simulations of a simple game.

4. We apply this methodology to an empirical example and demonstrate the type of econometric analysis and insights that it allows.

The paper is organized as follows. The next section introduces the general framework and the object of study. Section 3.3 derives the characterization of the identified set with the Min-Cut Max-Flow Theorem. Section 3.4 describes the combinatorial procedure to efficiently construct the confidence region. Section 3.5 contains the simulation evidence and Section 3.6 the empirical application. The last section concludes. Proofs are collected in an appendix.

3.2 Analytical framework

3.2.1 Model specification

We consider the following model specification.

$$Y \in G(X, \varepsilon; \theta), \tag{3.2.1}$$

where Y is an observable outcome variable, which takes values in a finite set $\mathcal{Y} = \{y_1, \ldots, y_K\}$, X is a vector of exogenous explanatory variables with domain $\mathcal{X}$, ε is a vector of unobservable heterogeneity variables with domain $\Xi \subset \mathbb{R}^l$ and $\theta \in$

$\Theta \subset \mathbb{R}^d$ is a vector of unknown parameters. Finally, $G : (X, \varepsilon) \rightrightarrows G(X, \varepsilon; \theta)$ is a multi-valued mapping. The random elements X, Y and ε are defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The sample consists in n observational units $i = 1, \ldots, n$, which are independent and identical in distribution. To each unit i is attached a vector $(Y_i, X_i, \varepsilon_i)$, only the first two elements of which shall be observed. For each potential outcome $y \in \mathcal{Y}$, we denote by $P(y|X)$ the conditional probability $\mathbb{P}(Y = y|X)$. If Z is a subset of $\mathcal{Y}$, $P(Z|X)$ will denote $\sum_{y \in Z} P(y|X)$. It is important to emphasize here the fact that $P(.|X)$ denotes the true outcome data generating process, which is unknown, but can be estimated from the data. It is not a function of the structural parameter vector and cannot be construed as the likelihood from the model. The vector of unobservable variables ε in the economic structure has conditional cumulative distribution function $F(\varepsilon|X; \theta)$ for some known function F parameterized by θ (the same notation is used for the parameters of the model correspondence and for the parameters of the error distribution to indicate that they may have common components). The economic structure is summarized by the multi-valued mapping G. A special case of specification (3.2.1) arises when G is a function, in which case model (3.2.1) is a nonlinear non separable single equation discrete choice model as in Chesher (2010). Here, however, we entertain the possibility of G having multiple values arising from multiple equilibria, data censoring or endogeneity. G is entirely given by the economic structural model, up to an unknown parameter vector θ.

The analytical framework, concepts and procedures proposed throughout the paper will be illustrated and discussed with the following simple example.

Example 3 (Partnership game) *Our example is a simple non cooperative full information game of complementarities.*

- STRATEGIES*: There are two players, who simultaneously decide, whether to invest strongly (strategy H), weakly (strategy L) or not at all (strategy O) in a partnership.*

- PAYOFFS: *Players pay a cost $c \geq 0$ (respectively $2c$) for a weak (respectively strong) investment. Benefits that accrue to players depend on the overall level of investment in the partnership and explanatory variables J_i, $i = 1, 2$, where $J_i = 1$ if player i is female, and zero otherwise. The benefits for player i are $3c(1 + \beta J_i)$ in case both players invest strongly, $2c(1 + \beta J_i)$ in case one player invests weakly and the other strongly and $c(1 + \beta J_i)$ in case both players invest weakly. Finally player i also experiences an idiosyncratic random participation payoff ε_i, $i = 1, 2$ with a density with respect to Lebesgue measure. The payoff matrix for the game is given in the following Table.*

Table 3.1: Payoff matrix for the partnership game.

		Player 2 : H	Player 2 : L	Player 2 : O
Player 1:	H	$3c(1+\beta J_i) - 2c + \varepsilon_1$ $3c(1+\beta J_i) - 2c + \varepsilon_2$	$2c(1+\beta J_i) - 2c + \varepsilon_1$ $2c(1+\beta J_i) - c + \varepsilon_2$	$-2c + \varepsilon_1$ 0
	L	$2c(1+\beta J_i) - c + \varepsilon_1$ $2c(1+\beta J_i) - 2c + \varepsilon_2$	$c(1+\beta J_i) - c + \varepsilon_1$ $c(1+\beta J_i) - c + \varepsilon_2$	$-c + \varepsilon_1$ 0
	O	0 $-2c + \varepsilon_2$	0 $-c + \varepsilon_2$	0 0

In each cell, the top expression is player 1's payoff and the bottom term is player 2's payoff.

- EQUILIBRIUM CONCEPT: *We assume that outcomes are Nash equilibria in pure strategies. Other equilibrium concepts could be entertained, in particular with mixed strategies, as will be discussed in Section 3.4.1 and illustrated in the empirical application.*

The strategies, payoffs and equilibrium concept together define the economic structure.

Y is an observed equilibrium strategy profile. $J = (J_1, J_2)$ is also observed by the analyst. The idiosyncratic participation benefit $\varepsilon = (\varepsilon_1, \varepsilon_2)$ is not, but it is common knowledge to the players. The structural parameter vector is $\theta = (c, \beta)$. The equilibrium correspondence, i.e., the set of equilibria for each value of ε, J and θ, can be easily derived, and defines the multi-valued mapping G in model specification (3.2.1), which is represented in the $(\varepsilon_1, \varepsilon_2)$ space in Figure 3.1 for the case $\beta = 0$. Since we assume that ε has absolutely continuous distribution with respect to Lebesgue measure, we do not include zero probability predictions, such as $\{OO, OL\}$ when $\varepsilon_2 = c$ and $\varepsilon_1 < -c$ for instance.

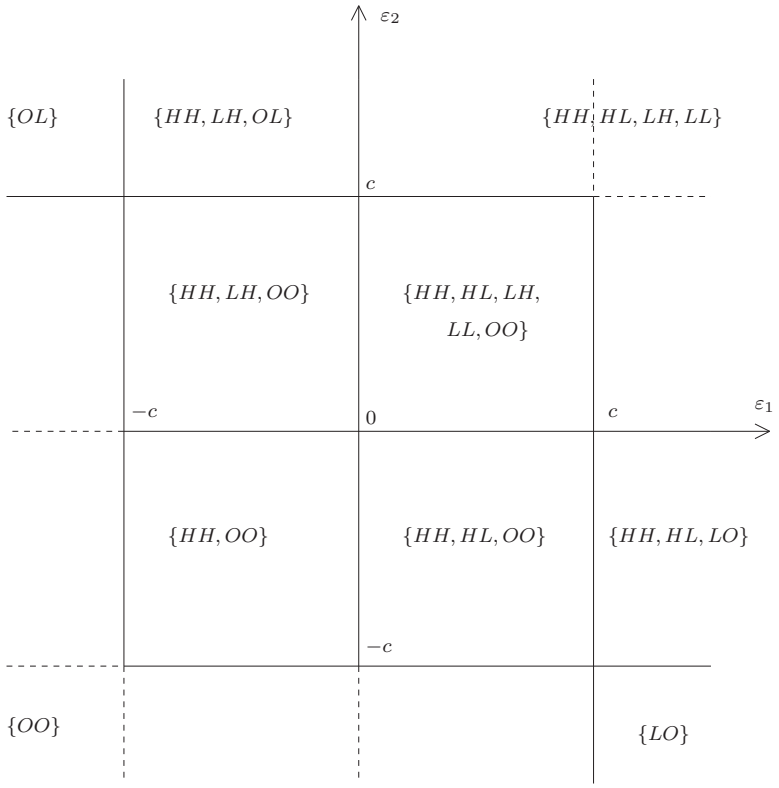


Figure 3.1: Representation of the equilibrium correspondence $G(J, \varepsilon; \theta)$ in the $(\varepsilon_1, \varepsilon_2)$ space, when $\beta = 0$.

3.2.2 Object of inference

Model (3.2.1) has the fundamental feature that G is multi-valued (because of multiple equilibria in the example above for instance). For a given value of (X, ε, θ), the model predicts a set of possible outcomes $G(X, \varepsilon; \theta)$. Only one of them, namely Y, is actually realized, but the economic structure is silent about how that particular Y was selected among $G(X, \varepsilon; \theta)$. In other words, the economic structure holds no information about the equilibrium selection mechanism. If the true (unknown) equilibrium selection mechanism is denoted $\pi_0(y|\varepsilon, X)$, which is a probability on $G(X, \varepsilon; \theta)$, then the likelihood of observation y can be written

$$L(\theta|y, X) = \int_{\Xi} \pi_0(y|\varepsilon, X) dF(\varepsilon|X; \theta),$$

and the true parameter θ_0 satisfies

$$P(y|X) = \int_{\Xi} \pi_0(y|\varepsilon, X) dF(\varepsilon|X; \theta_0), \ X\text{-a.s.}, \ \text{ for all } y. \tag{3.2.2}$$

Jovanovic (1989) points out that the incomplete model (incomplete because the equilibrium selection is not modeled) is compatible with the true data generating process $P(.|X)$ if and only if there exists a (generally non unique) equilibrium selection mechanism π_0 such that (3.2.2) holds. The identified set is then defined as the set Θ_I of parameter values θ such that model (3.2.1) is compatible in the sense of Jovanovic (1989).

Definition 3 (Identified set) *The identified set Θ_I is the set of parameter values $\theta \in \Theta$ such that there exists a probability kernel $\pi(.|\varepsilon, X)$ with support $G(X, \varepsilon; \theta)$ for which (3.2.2) holds.*

The identified set is empty if no value of the parameter can rationalize the data generating process, in which case the structural model is misspecified. The identified set is a singleton in case of point identification, which occurs if G happens to be single valued under the true parameter values (in case $c = \beta = 0$ in Example 3)

or in very special cases under large support assumptions on X, as in Tamer (2003). The identified set is totally uninformative, i.e., $\Theta_I = \Theta$, in case the model has no empirical content (if for instance $G(X, \varepsilon; \theta_0)$ contains all selected outcome values for almost all ε at the true value θ_0).

3.2.3 Applications of the framework

Specification (3.2.1), hence the inference procedure presented in this paper, has a wide range of applications. Some of the most compelling ones are the empirical analysis of games, instrumental variable models of discrete choice with endogeneity and revealed preference analysis.

- Empirical analysis of games: As illustrated in Example 3, Model (3.2.1) applies to the empirical analysis of noncooperative games of perfect information (normal form games). They include the classic entry game of Bresnahan and Reiss (1990) and Berry (1992) as well as the social interaction game of Soetevent and Kooreman (2007). Noncooperative games of private information make for a less compelling application of this framework as point identification conditions are more easily derived and justified than in their perfect information counterparts (see for instance Aradillas-Lopez (2010) and Bajari, Hahn, Hong, and Ridder (2011) for a discussion). Finally, some cooperative games can be analyzed and estimated within the present framework, in particular matching and social network formation games, where the equilibrium correspondence is characterized by pairwise stability. Uetake and Watanabe (2011) present an empirical analysis of entry by merger, where the present inference procedure can be applied.

- Discrete choice models with endogeneity: Chesher, Rosen, and Smolinski (2011) show that instrumental variable models of discrete choice fall under model (3.2.1) and they use Theorem 1 of Galichon and Henry (2011) or equiv-

alently Theorem 3.2 of Beresteanu, Molchanov, and Molinari (2011) to characterize the identified set. The present work complements Chesher, Rosen, and Smolinski (2011) in proposing the first feasible inference procedure for such models.

- Revealed preference analysis: Henry and Mourifié (2011) apply the inference procedure proposed here to analyze voting behaviour from a revealed preference standpoint. The same approach can be applied to revealed preference testing in matching markets as in Echenique, Lee, Shum, and Yenmez (2011) or the revealed preference approach to games taken in Pakes, Porter, Ho, and Ishii (2004).

3.3 Operational characterization of the identified set

As noted in Berry and Tamer (2006), Definition 3 is not an operational definition of the identified set, as it includes the equilibrium selection mechanism as an infinite dimensional parameter. Galichon and Henry (2006b,2011) and Beresteanu, Molchanov, and Molinari (2011) show a characterization of the identified set with a finite collection of moment inequalities. In this section, we give an equivalent characterization of the identified set, whose proof is much simpler and relies on the Min-Cut Max-Flow Theorem, which brings classical efficient combinatorial optimization methods to bear on the problem. This will prove crucial for the feasibility of the inference procedure in realistic and relevant empirical examples.

First, we set out the main heuristic for the operational characterization of the identified set. Model specification (3.2.1) is a discrete choice model, hence the set $\mathcal{Y}$ of outcomes is finite and the correspondence G takes only a finite number of values, which we label $\mathcal{U} = \{u_1, \ldots, u_J\}$. Each u is a set (possibly singleton) of outcomes in

$\mathcal{Y}$. Because the model is incomplete, it does not predict the probabilities of individual outcomes in $\mathcal{Y}$, but it predicts the probability of each combination of equilibria listed in $\mathcal{U}$. We denote these probabilities $Q(u|X;\theta)$ as they depend on the structural parameter value.

Definition 4 (Predicted probabilities) *For each $u \in \mathcal{U}$, we define $Q(u|X;\theta) := \mathbb{P}(G(X,\varepsilon;\theta) = u|X,\theta)$. If V is a subset of $\mathcal{U}$, we write $Q(V|X;\theta) = \sum_{u\in V} Q(u|X;\theta)$.*

In most applications, it will be difficult to obtain closed forms for $Q(u|X;\theta)$. However, ε can be randomly generated. Given a sample $(\varepsilon^r)_{r=1,...,R}$ of simulated values, $Q(u|X;\theta)$ can be approximated by $\sum_{r=1}^{R} 1\{u = G(X,\varepsilon^r;\theta)\}/R$. Bajari, Hong, and Ryan (2010) propose an importance sampling procedure that greatly reduces the computational burden of this stage of the inference. The simulation procedure is now standard and cannot be avoided if one wishes, as we do here, to exhaust the empirical content of the structural model.

Example 3 continued: In the partnership example with $\beta = 0$, the model predicts the following values for the equilibrium correspondence: $\mathcal{U} = \{\ \{OL\}, \{LH, OL, HH\}, \{HH, LH, OO\},\ \{OO\},\ \{HH, OO\},\ \{HH, LL, HL, LH\},\ \{HH, LL, OO, HL, LH\}, \{HH, OO, HL\},\ \{HH, HL, LO\},\ \{LO\}\}$. The set $\mathcal{Y}$ of equilibrium strategy profiles (that may be observed) is $\{HH,\ HL,\ LH,\ LL,\ LO,\ OL,\ OO\}$ with 7 elements, while the set of predicted collections of equilibria (possible values of the equilibrium correspondence) $\mathcal{U}$ has 10 elements. The predicted probabilities can be computed in the following way. For instance, $Q(\{OL\}|c) = \mathbb{P}(\varepsilon_1 \leq -c \text{ and } \varepsilon_2 \leq c)$ and $Q(\{HH, LH, OL\}|c) = \mathbb{P}(-c \leq \varepsilon_1 \leq 0 \text{ and } \varepsilon_2 \leq c)$ and the remaining 8 probabilities are determined similarly from Figure 3.1.

The model structure imposes a set of restrictions on the relation between the predicted probabilities of equilibrium combinations and the true probabilities of outcomes. For instance, the predicted probability $Q(\{HH, LH, OL\}|X;\theta)$ in the above example cannot be larger than the sum $P(HH) + P(LH) + P(OL)$ of probabilities of

occurrence of each individual equilibrium in u, since Y is either HH, LH or OL, when $u = \{HH, LH, OL\}$ is predicted. More generally, since P and Q are the marginals of the joint distribution of (Y, U) given X, we must have for all $u \in \mathcal{U}$:

$$Q(u|X;\theta) = \sum_{y \in u} \mathbb{P}(Y = y \text{ and } U = u|X;\theta) \leq \sum_{y \in u} \mathbb{P}(Y = y|X;\theta) = \sum_{y \in u} P(y|X) \quad (3.3.1)$$

Note that $Q(u|X;\theta)$ may be strictly smaller than $\sum_{y \in u} P(y|X)$ when some outcome $y \in u$ also belongs to other combinations u' that may arise under different values of ε, as its (marginal) probability $P(y|X)$ must then be split between $Q(u|X;\theta)$ and the probabilities $Q(u'|X;\theta)$ of such other combinations $u' \in \mathcal{U}$ containing y. However, inequalities (3.3.1) do not exhaust the information in the structure. They may all be satisfied and yet the structure may be incompatible with the data generating process as the following example shows. Hence more inequalities will be needed as derived below.

Example 3 continued: In the partnership example with $\beta = 0$, suppose that the true equilibrium selection mechanism is such that $Q(\{OL\}|\theta) = P(OL) > 0$ and $Q(\{HH, LH, OL\}|\theta) = P(HH)+P(LH)+P(OL)$. Then $Q(\{OL\} \cup \{HH, LH, OL\}|\theta) = Q(\{OL\}|\theta) + Q(\{HH, LH, OL\}|\theta) > P(HH) + P(LH) + P(OL)$ so that $\theta \notin \Theta_I$.

Extending this observation, consider a subset $V \subseteq \mathcal{U}$ and define

$$V^{\cup} := \{y \in Y : y \in u \text{ for some } u \in V\} = \bigcup_{u \in V} u.$$

Then we must have

$$\begin{aligned} Q(V|X;\theta) &= \sum_{u \in V} \sum_{y \in u} \mathbb{P}(Y = y \text{ and } U = u|X;\theta) \\ &= \sum_{y \in V^{\cup}} \sum_{u \in V : y \in u} \mathbb{P}(Y = y \text{ and } U = u|X;\theta) \\ &\leq \sum_{y \in V^{\cup}} \sum_{u \in \mathcal{U}} \mathbb{P}(Y = y \text{ and } U = u|X;\theta) \\ &= \sum_{y \in V^{\cup}} P(y|X) \end{aligned}$$

where the inequality is again due to the fact that some $y \in V^{\cup}$ may also belong to some $u' \notin V$. Since this inequality holds for every $V \subseteq \mathcal{U}$, we must have

$$\max_{V \subseteq \mathcal{U}} \left(\sum_{u \in V} Q(u|X;\theta) - \sum_{y \in V^{\cup}} P(y|X) \right) \leq 0.$$

This inequality must also hold for every realization x of X in the domain $\mathcal{X}$ of the explanatory variables, implying that every θ in the identified set Θ_I must satisfy

$$\sup_{x \subseteq \mathcal{X}} \max_{V \subseteq \mathcal{U}} \left(\sum_{u \in V} Q(u|x;\theta) - \sum_{y \in V^{\cup}} P(y|x) \right) \leq 0.$$

So far, we have shown implications of the model. It is far more difficult to show that these implication actually exhaust all the empirical content of the model, i.e., that they involve no loss of information and constitute sharp bounds. In Theorem 3 below, we will show this with an appeal to the classical Max-Flow Min-Cut Theorem of combinatorial optimization, providing our characterization (3.3.2) of the identified set. We thereby provide, for the case of a finite set of possible outcomes, a new and simpler proof of the characterization of the identified set with a finite collection of inequalities, without the complicated apparatus of the theory of random sets. This allows us to emphasize the combinatorial optimization formulation of our inference problem, which is key to its tractable solution in empirically relevant instances. Theorem 3 below also provides an alternative characterization (3.3.3) of the identified set from the "dual" perspective of outcome subsets $Z \subseteq \mathcal{Y}$, in addition to the preceding characterization (3.3.2) based on combination subsets $V \subseteq \mathcal{U}$, with the notation

$$Z^{\cap} := \{u \in \mathcal{U} : \ u \subseteq Z\} \ \text{ and } \ Z^{-1} := \{u \in \mathcal{U} : \ u \cap Z \neq \varnothing\}.$$

This alternative characterization may be useful in situations where the number of possible outcomes is much smaller than the number of possible combinations (as is the case in Example 3, where the number of equilibrium outcomes (cardinality of $\mathcal{Y}$) is 7, so the corresponding number of inequalities to be checked is $2^7 - 1 = 127$, whereas the number of predicted equilibrium combinations (cardinality of $\mathcal{U}$) is 10, so

the corresponding number of inequalities to check would be $2^{10} - 1 = 1023$). Finally, it is also equivalent to the characterization of the identified set derived in Galichon and Henry (2006b), which we give in (3.3.4) in our notation.

Theorem 3 *The identified set is*

$$\Theta_I = \left\{\theta \in \Theta : \sup_{x \in \mathcal{X}} \max_{V \subseteq \mathcal{U}} \left(Q(V|x;\theta) - P(V^{\cup}|x)\right) \leq 0\right\} \tag{3.3.2}$$

$$= \left\{\theta \in \Theta : \sup_{x \in \mathcal{X}} \max_{Z \subseteq \mathcal{Y}} \left(Q(Z^{\cap}|x;\theta) - P(Z|x;\theta)\right) \leq 0\right\} \tag{3.3.3}$$

$$= \left\{\theta \in \Theta : \sup_{x \in \mathcal{X}} \max_{Z \subseteq \mathcal{Y}} \left(P(Z|x;\theta) - Q(Z^{-1}|x;\theta)\right) \leq 0\right\}. \tag{3.3.4}$$

Theorem 3 gives three characterizations of the identified set Θ_I, sometimes called *sharp identified region* in the literature. Θ_I contains all the values of the parameter such that (3.2.1) holds and only such values. Moreover, all elements of Θ_I are observationally equivalent. Hence no value of the parameter vector θ contained in Θ_I can be rejected on the basis of the information available to the analyst. Thus, Θ_I completely characterizes the empirical content of the model.

Example 3 continued: To illustrate the computation of the identified set, consider the case, where it is known that $\beta = 0$. Assume that the true parameter value is $c_0 = 1/4$ and the idiosyncratic shocks are independent and uniformly distributed over $[-1/2, 1/2]$. Suppose further that the true data generating process is equal to the distribution implied by a uniform equilibrium selection rule, whereby all equilibrium strategy profiles within the equilibrium correspondence are selected with equal probability. For example, when $\varepsilon_1 \geq c_0 = 1/4$ and $-1/4 = -c_0 \leq \varepsilon_2 \leq 0$, each strategy profile within the equilibrium correspondence $\{HH, HL, LO\}$ is equally likely. The probability distribution of the true data generating process in this case is defined by $P(HH) = 167/960$, $P(OO) = 191/480$, $P(OL) = P(LO) = 1/12$, $P(LL) = 19/320$ and $P(HL) = P(LH) = 97/960$. The identified set is derived as the set of values of c such that the $2^7 - 1 = 127$ inequalities of the form $P(Z) \geq Q(Z^{\cap}|c)$, all

$Z \subseteq \{HH, HL, LH, LL, LO, OL, OO\}$, are satisfied. For instance, one of those inequalities is $59/320 = P(LO \text{ or } HL) \geq Q(\{LO\}|c) = (1/2 - c)^2$ if $c \leq 1/2$ and zero otherwise. The identified set can be computed using a Min-Cut Max-Flow algorithm, which yields $[1/2 - 1/\sqrt{12}, 1/3] \simeq [0.2113, 0.3333]$ where the lower bound of the interval happens to be the smallest value of $c > 0$ for which the inequality in (3.3.3) with $Z = \{LO, OL\}$ is satisfied, and the upper bound happens to be the largest value for which that with $Z = \{HH, HL, LH, LL, OO\}$ is satisfied.

As illustrated in Example 3, even in simple examples, where the equilibria are very easy to compute, the exponential size of the characterization of the identified set is a severe computational burden that is best approached with combinatorial optimization techniques, as developed in the next section.

3.4 Confidence region

3.4.1 Objective

We now turn to the problem of inference on Θ_I based on a sample of observations $((Y_1, X_1), \ldots, (Y_n, X_n))$. We seek coverage of the identified set with prescribed probability $1 - \alpha$, for some $\alpha \in (0, 1)$. It would be tempting to appeal to the large literature on inference in moment inequality models. This includes several proposals for the construction of confidence regions covering each point in the identified set, which are generally preferred on account of the fact that they may be more informative (although this may sometimes be misleading as pointed out in Henry and Onatski (2012)). Such proposals include Section 5 of Chernozhukov, Hong, and Tamer (2007), Romano and Shaikh (2008), Rosen (2008), Galichon and Henry (2009) and Andrews and Soares (2010) among others. All of the above propose to construct confidence regions by inverting specification tests. Hence, the confidence region is constructed through a search in the parameter space, with a computationally demanding testing

procedure at each parameter value visited in the search. This becomes computationally infeasible for realistic parameter vector dimensions. With a reasonably precise grid search and 5 parameters (for example), the number of points to be visited is in the tens of billions. If the identified set is known to be convex, the search can be conducted from a central point with a dichotomy in polar coordinates, yet it remains computationally impractical to conduct a statistical procedure for each point in the search.

Hence, each parameter value in the search must be accepted or rejected based on a deterministic criterion. This means the significance of the confidence region must be controlled independently of the parameter value. This will automatically produce a confidence region that covers the identified set. Proposals for the construction of confidence regions covering the identified set include Chernozhukov, Hong, and Tamer (2007), Romano and Shaikh (2010), Galichon and Henry (2006a) and Bugni (2010) among others. These can be applied to realistic models defined by a small number of moment inequality restrictions. However, a major challenge in the framework of this paper is that the number of inequalities characterizing the empirical content of the model in Theorem 3 grows exponentially with the cardinality of $\mathcal{Y}$, which in the case of games is the number of equilibrium strategy profiles (in the very simple partnership game of Example 3, the number of inequalities is 127). Hence the combinatorial optimization approach that we propose in this paper is to the best of our knowledge the only computationally feasible inference procedure for empirically relevant economic structures defined by finite games and other models of discrete choice with endogeneity.

Definition 5 (Confidence region) *A confidence region of asymptotic level* $1-\alpha$ *for the identified set* Θ_I *is defined as a sequence of regions* Θ_n, $n \in \mathbb{N}$, *satisfying* $\liminf_n \mathbb{P}(\Theta_I \subseteq \Theta_n) \geq 1-\alpha$.

We seek coverage of the set of values of the parameter θ such that $Q(V|x,\theta) \leq P(V^{\cup}|x)$ for all values of x and all subset V of $\mathcal{U}$. Q is determined from the model,

but P is unknown. However, if we can construct random functions $\overline{P}_n(A|x)$ that dominate the probabilities $P(A|x)$ for all values of x and all subsets A of $\mathcal{Y}$ with high probability, then in particular, $\overline{P}_n(V^{\cup}|x) \geq P(V^{\cup}|x)$ for each x and each subset V of $\mathcal{U}$. Hence any θ satisfying $Q(V|x,\theta) \leq P(V^{\cup}|x)$ for all values of x and all subsets V of $\mathcal{U}$ also satisfies $Q(V|x,\theta) \leq \overline{P}_n(V^{\cup}|x)$ for all values of x and all subsets V of $\mathcal{U}$. There remains to control the level of confidence of the covering region, which is achieved by requiring that $\overline{P}_n$ dominate P with probability asymptotically no less than the desired confidence level. Equivalently, when working from characterization (3.3.4), we impose the same requirement for dominated functions $\underline{P}_n$. Hence the following assumption.

Assumption 5 *Let the random functions $A \mapsto \overline{P}_n(A|x)$, $A \subseteq \mathcal{Y}$, satisfy*

$$\liminf_n \mathbb{P}\left(\sup_{x\in\mathcal{X}} \max_{A\subseteq\mathcal{Y}} \left[P(A|x) - \overline{P}_n(A|x)\right] \leq 0\right) \geq 1-\alpha. \tag{3.4.1}$$

Suppose now a value θ_0 of the parameter vector belongs to the identified set Θ_I. Then, by Theorem 3, for all x and $V \subseteq \mathcal{U}$, $Q(V|x;\theta_0) \leq P(V^{\cup}|x)$, so that with probability tending to no less than $1-\alpha$, $Q(V|x;\theta_0) \leq \overline{P}_n(V^{\cup}|x)$, hence Theorem 4.

Theorem 4 (Confidence region) *Under Assumption 5, the sets*

$$\Theta_I(\overline{P}_n) = \left\{\theta \in \Theta : \sup_{x\in\mathcal{X}} \max_{V\subseteq\mathcal{U}} (Q(V|x;\theta) - \overline{P}_n(V^{\cup}|x)) \leq 0\right\} \tag{3.4.2}$$

define a confidence region of asymptotic level $1-\alpha$ for Θ_I (according to Definition 5).

Theorem 4 has the fundamental feature that it dissociates search in the parameter space (or even possibly search over a class of models) from the statistical procedure necessary to control the confidence level. The upper probabilities $\overline{P}_n$ can be determined independently of θ in a procedure that is performed once and for all using only sample information, i.e. fully nonparametrically. Once the upper probabilities are determined, probabilities Q over predicted sets of outcomes are computed for particular

chosen specifications of the structure and values of the parameter, and such specifications and values are tested with inequalities defining $\Theta_n(\overline{P}_n)$. This dissociation of the statistical procedure to control confidence level from the search in the parameter space is crucial to the computational feasibility of the proposed inference procedure in realistic examples (i.e. sample sizes in the thousands, two-digit dimension of the parameter space and two-digit cardinality of the set of observed outcomes, as in the application to teen behavior in Soetevent and Kooreman (2007), or to entry in the airline market in Ciliberto and Tamer (2009)). The latter consider only equilibria in pure strategies, as we have until now. If equilibria in mixed strategies are also considered, as in Bajari, Hong, and Ryan (2010) and in the family bargaining application below, we can appeal to results in Beresteanu, Molchanov, and Molinari (2011) and Galichon and Henry (2011). In particular, Galichon and Henry (2011) show that if the game has a *Shapley regular core* (which is the case in the family bargaining application, by Lemma 2 of Galichon and Henry (2011)), then the identified set is characterized by (3.3.3) of Theorem 3 with the caveat that the set function $Z \mapsto Q(Z^{\cap}|x)$ is replaced by

$$\mathcal{L}(Z|x) = \int \min_{\sigma \in G(\varepsilon|X;\theta)} \sigma(Z) d\nu(\varepsilon), \tag{3.4.3}$$

where $G(\varepsilon|X;\theta)$ is now a set of mixed strategies, i.e. a set of probabilities on the set of outcomes, as opposed to a subset of the set of outcomes. Hence the methodology is be easily adapted, as in the application of Section 3.6.

3.4.2 Control of confidence level

We now turn to the determination of random functions satisfying Assumption 5. First, for each $y \in \mathcal{Y}$, let $\hat{P}_n(y|x)$ be the empirical analog (or more generally a nonparametric estimator) of $P(y|x)$ and $\hat{P}_n(A|x) = \sum_{y \in A} \hat{P}_n(y|x)$ for each $A \subseteq \mathcal{Y}$. A simple way of achieving (3.4.1) is by considering the random variable

$$M_n := \sup_{x \in \mathcal{X}} \max_{A \subseteq \mathcal{Y}} [P(A|x) - \hat{P}_n(A|x)].$$

Denoting by c_n^α the $(1-\alpha)$-quantile of the distribution of M_n, we have $\mathbb{P}(M_n \leq c_n^\alpha) = 1-\alpha$ by construction, hence

$$\mathbb{P}\left(\sup_{x\in\mathcal{X}} \max_{A\subseteq\mathcal{Y}} [P(A|x) - \hat{P}_n(A|x) - c_n^\alpha] \leq 0\right) \geq 1-\alpha, \tag{3.4.4}$$

and the desired result with $\overline{P}(A|x) = \hat{P}_n(A|x) + c_n^\alpha$. However, by construction, c_n^α is independent of A and x, so that the region obtained by plugging $\overline{P}(A|x) = \hat{P}_n(A|x) + c_n^\alpha$ into (3.4.2) of Theorem 4 will be unnecessarily conservative. We propose, instead, to replace c_n^α by a function $\beta_n(A|x)$ of A and x, which we interpret as a *functional quantile* of the distribution of the random function $P(A|x) - \hat{P}_n(A|x)$. Analogously to (3.4.4), we require it to satisfy

$$\mathbb{P}\left(\sup_{x\in\mathcal{X}} \max_{A\subseteq\mathcal{Y}} [P(A|x) - \hat{P}_n(A|x) - \beta_n(A|x)] \leq 0\right) \geq 1-\alpha. \tag{3.4.5}$$

We first give a heuristic description of our proposed functional quantile before precisely spelling out the bootstrap procedure involved in approximating it. If $\mathcal{X}$ is finite, the random matrix $P(A|x) - \hat{P}_n(A|x)$, with $A \subseteq \mathcal{Y}$ and $x \in \mathcal{X}$ has a finite population of possible realizations, at most one for each possible sample draw. These realizations can be ordered according to the maximum entry in the matrix $\max_{x\in\mathcal{X}} \max_{A\subseteq\mathcal{Y}} [P(A|x) - \hat{P}_n(A|x)]$. Now take all realizations that never exceed the $(1-\alpha)$-quantile c_n^α of $\max_{x\in\mathcal{X}} \max_{A\subseteq\mathcal{Y}} [P(A|x) - \hat{P}_n(A|x)]$ and define $\overline{P}_n(A|x) = \hat{P}_n(A|x) + \beta_n(A|x)$, where $\beta_n(A|x)$ is the pointwise maximum over all realizations that never exceed c_n^α. This guarantees that the resulting confidence region obtained in (3.4.2) of Theorem 4 with $\overline{P}_n(A|x) = \hat{P}_n(A|x) + \beta_n(A|x)$ will be valid and will be contained in the region obtained with $\overline{P}_n(A|x) = \hat{P}_n(A|x) + c_n^\alpha$ (hence more informative than the latter). In case the conditioning variables are finitely supported, it is well known (see Singh (1981) and Bickel and Freedman (1981)) that the nonparametric bootstrap version of c_n^α is a valid approximation, which in turns guarantees the validity of the bootstrap procedure described below. In case X has continuous components, Chernozhukov, Lee, and Rosen (2009) derive the asymptotic distribution of the supremum (over $\mathcal{X}$) of the conditional empirical process, but nothing is known of its nonparametric bootstrap approximation.

Definition 6 (Nonparametric Bootstrap) *Let $\mathbb{P}_n^*$ denote probability statements relative to the bootstrap distribution and conditional on the original sample $((Y_1, X_1), \ldots, (Y_n, X_n))$. A bootstrap sample takes the form $((Y_1^*, X_1), \ldots, (Y_n^*, X_n))$, where the explanatory variable is not resampled and for each i, Y_i^* is drawn from distribution $\hat{P}_n(.|X_i)$. Let $((Y_1^b, X_1), \ldots, (Y_n^b, X_n))$, $b = 1, \ldots, B$ be a sequence of B bootstrapped samples. Denote by $\hat{P}_n^*(.|.)$ the bootstrap version (i.e., constructed identically from a bootstrap sample) of $\hat{P}_n(.|.)$ and $\hat{P}_n^b$, $b = 1, \ldots, B$ its values taken on the B realized bootstrap samples. Finally, for each $A \subseteq \mathcal{Y}$ and $1 \leq j \leq n$, denote $\zeta_n^*(A|X_j) = \sum_{y \in A}[\hat{P}_n(y|X_j) - \hat{P}_n^*(y|X_j)]$ and define $\zeta_n^b(A|X_j)$ analogously.*

In the bootstrap version of the problem, we are seeking functions β_n satisfying

$$\mathbb{P}_n^* \left(\max_{1 \leq j \leq n} \max_{A \subseteq \mathcal{Y}} [\hat{P}_n(A|X_j) - \hat{P}_n^*(A|X_j) - \beta_n(A|X_j)] \leq 0 \right) \geq 1 - \alpha \ *\text{-a.s.}$$

If there was a total order on the space of realizations of ζ_n^*, we could choose β_n as the quantile of level $1 - \alpha$ of the distribution of ζ_n^*. However, the $\zeta_n^*(\cdot, X_j)$'s are random functions defined on $2^{\mathcal{Y}} \times \{X_1, \ldots, X_n\}$, hence there is no such total order. We propose to determine β_n from a subset of $\lceil B(1-\alpha) \rceil$ bootstrap realizations determined as follows (where $\lfloor x \rfloor$ is the largest integer below x).

Step 1: Draw bootstrap samples $((Y_1^b, X_1), \ldots, (Y_n^b, X_n))$, for $b = 1, \ldots, B$.

Step 2: For each $b \leq B$, $j \leq n$ and $A \subseteq \mathcal{Y}$, compute $\zeta_n^b(A|X_j) = \hat{P}_n(A|X_j) - \hat{P}_n^b(A|X_j)$.

Step 3: Discard at most a proportion α of the bootstrap indices, and compute $\beta_n(A|X_j)$ as the maximum over the remaining bootstrap realizations $\zeta_n^b(A|X_j)$.

Discarding at most $B\alpha$ among the bootstrap realizations guarantees the control of the level of confidence, and we wish to choose the set $D \subseteq \{1, \ldots, B\}$ of discarded indices so as to make β_n *as small as possible, to maximize informativeness* of the resulting confidence region. Again, if there was a total order, we would be similarly

discarding the $B\alpha$ largest realizations of ζ_n^b, effectively choosing β_n as the quantile of the distribution of ζ_n^b, $b = 1\ldots, B$. Instead, we discard all realizations of the matrix $\zeta_n^b(A|X_j)$ that have at least one entry that strictly exceeds the $(1-\alpha)$-quantile of $w_b = \max_{1\leq j\leq n}\max_{A\subseteq\mathcal{Y}}\zeta_n^b(A|X_j)$. Hence, we choose D solving the optimization problem

$$\min\left\{\max_{b\notin D} w_b \;:\; D\subseteq\{1,\ldots,B\},\ |D|\leq B\alpha\right\}. \tag{3.4.6}$$

The procedure is explained graphically in Figure 3.2.

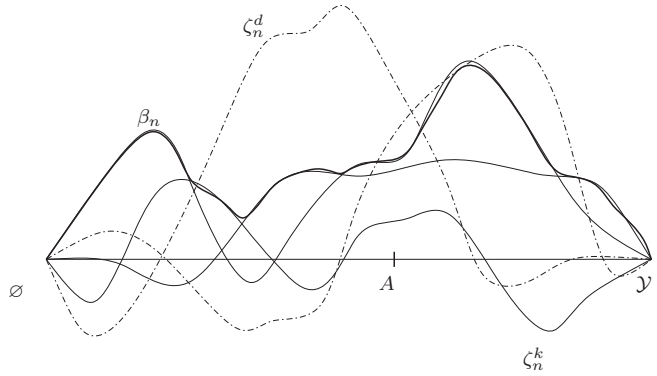


Figure 3.2: Stylized representation of the determination of the functional quantile β_n in a case without explanatory variables.

The subsets A of $\mathcal{Y}$ are represented on the horizontal axis, ranging from $\varnothing$ to $\mathcal{Y}$. ζ_n^d is one of two discarded realization of the empirical process (dotted lines), whereas ζ_n^k is one of three realizations that are not discarded (solid lines). β_n is the pointwise maximum over the realizations that were not discarded (thick line).

Problem (3.4.6) can be solved by the following *Bootstrap Realization Selection (BRS) algorithm*:

BRS Step 1: For each $b \leq B$, set $w_b' = \max_{1\leq j\leq n}\sum_{y\in\mathcal{Y}}\max\{0, \hat{P}_n(y|X_j) - \hat{P}_n^b(y|X_j)\}$.

BRS Step 2: Let D be the set of indices b of the $\lfloor B\alpha\rfloor$ largest w_b'.

Proposition 5 *The BRS algorithm determines an optimal solution to problem (3.4.6) in $O(nB|\mathcal{Y}|)$ time.*

Remark 1 *Problem (3.6) may have alternate optimum solutions. As observed by a referee, this may arise when the sample size n is small, since $\hat{P}_n(y|X_j)$ and $\hat{P}_n^b(y|X_j)$ are multiples of $1/n$ and thus distinct w_b's are more likely to have the same value when the sample size n is small. In case of ties,* any *optimum solution D to Problem (3.6) may be used to discard bootstrap realizations and determine functions β_n. If one desires a specific tie-breaking rule, e.g., for robustness or reproducibility, then we suggest the following lexicographic selection rule as a refinement to BRS Step 2: let w^b be the vector with components $w_j^b = \sum_{y\in\mathcal{Y}} \max\{0, \hat{P}_n(y|X_j) - \hat{P}_n^b(y|X_j)\}$ for $j = 1, \ldots, n$; and let $[w]^b$ be the vector w^b with its components sorted in nonincreasing order, i.e., with $[w]_1^b = w_b \geq [w]_2^b \geq \cdots \geq [w]_n^b = \min_j w_j^b$; then discard the $\lfloor B\alpha \rfloor$ bootstrap realizations b with the lexicographically largest vector $[w]^b$. In other words, we refine problem (3.6) as lexmin $\{lexmax_{b\notin D}[w]^b : D \subseteq \{1, \ldots, B\}, |D| \leq B\alpha\}$ where lexmin and lexmax denote the minimum and maximum relative to the lexicographic total order of vectors with n components. This rule aims at simultaneously minimizing all the values $\beta(A|X_j)$ without going through extensive additional computations.*

In problem (3.4.6), we chose to minimize the maximum, over all $j \in \{1, \ldots, n\}$ and $A \subseteq \mathcal{Y}$, of the non-discarded bootstrap realizations $\zeta_n^b(A|X_j)$. Other objectives are possible, for example the $\mathbb{L}^1$ objective $\sum_{b\notin D} w_b$. The main justification for the $\mathbb{L}^\infty$ norm objective $\max_{d\notin D} w_b$ in (3.4.6) is that it leads to a problem solvable in linear time. In contrast, the problem with an $\mathbb{L}^1$ objective is computationally difficult, namely NP-hard in the strong sense, as shown in the next result.

Proposition 6 *Minimization of $\{\sum_{b\notin D} w_b \ : \ |D| \leq \lfloor B\alpha \rfloor, \ D \subseteq \{1, \ldots, B\}\}$ is NP-hard in the strong sense.*

This result implies that unless $P = NP$, there exists no algorithm for this problem that runs in polynomial time. This is a severe computational drawback relative to

the linear-time algorithm achieved with with BRS.

3.4.3 Search in the parameter space

Once the functional quantile has been computed, there remains to search in the parameter space for the values of θ that satisfy (3.4.2). As shown in the Lemma 1, the function to be optimized in characterization (3.3.2) of the identified set is supermodular.

Definition 7 (Supermodular function) *A set function $\rho : A \mapsto \rho(A) \in \mathbb{R}$ is called supermodular (resp. submodular) if for all pairs of sets (A, B), $\rho(A \cup B) + \rho(A \cap B) \geq$ (resp. $\leq$) $\rho(A) + \rho(B)$.*

Lemma 1 *The function $V \mapsto P(V^{\cup}|x)$ is submodular for all $x \in \mathcal{X}$.*

In the computation of $\Theta_n(\overline{P}_n)$, it may be desirable to require $\overline{P}_n(V^{\cup}|x)$ to also be submodular as a function of $V \subseteq \mathcal{U}$, so that the function to be maximized in (3.4.2) can be maximized using submodular optimization techniques. This can be achieved by adding the following additional linear constraints (see Schrijver (2004)): $\forall u \neq v \in \mathcal{U}, \forall V \subseteq \mathcal{U} \backslash \{u, v\},\ j = 1, \ldots, n,$

$$\begin{aligned} \overline{P}_n([V \cup \{u\} \cup \{v\}]^{\cup}|X_j) - \overline{P}_n([V \cup \{u\}]^{\cup}|X_j) \\ -\overline{P}_n([V \cup \{v\}]^{\cup}|X_j) + \overline{P}_n([V]^{\cup}|X_j) \leq 0. \end{aligned} \tag{3.4.7}$$

The problem of checking whether θ is in the confidence regions can then be solved in polynomial time. Moreover, since submodular optimization has far ranging applications in all areas of operations research, many extremely efficient algorithms and implementations are readily available.

3.5 Simulation based on Example 3

We now illustrate and assess the performance of our procedure on the game described in Example 3. Throughout the experiment, we assume that $(\varepsilon_1, \varepsilon_2)$ is uniformly distributed on $[-1/2, 1/2]^2$ and $J = (J_1, J_2)$ is a vector of independent Bernoulli(1/2) random variables. True values for the parameters are indicated with a 0 subscript. We consider the following true parameter specifications: $(\beta_0, c_0) = (0, 0)$ (point identified case) and $(\beta_0, c_0) = (0, 1/4)$ (which corresponds in some sense to the greatest possible indeterminacy). For the true data generating process, we consider two distinct equilibrium selection rules (which, like the true parameter values, are of course supposed unknown in the inference procedure). The first rule specifies that in case of multiplicity, all equilibrium strategy profiles in the equilibrium correspondence are selected with equal probability: we call this case "uniform selection". The second selection rule specifies that in case of multiplicity, the equilibrium with largest aggregate investment is selected; suppose for instance that the equilibrium correspondence takes the value $\{HH, HL, LO\}$, then equilibrium strategy profile HH is realized: we call this case "maximal selection". In the case of maximal selection with $c_0 = 0.25$, $\beta_0 = 0$ is assumed known a priori by the analyst performing inference (to avoid an unbounded identified set in the simulations). In the remaining 3 cases, β_0 is unknown a priori. The experiment is run as follows. We calculate in each of the 4 cases above the distribution of the true data generating process. With the latter, we compute the identified set. In the point identified case, the identified set is equal to the true value. In the case $c_0 = 0.25$, with $\beta = 0$ known a priori and maximal selection, the identified set is $[0.2113, 0.3333]$ as explained in the example at the end of Section 3.3. In case $(c_0 = 0.25, \beta_0 = 0)$ with uniform selection, the identified set projects to $[0, 0.375]$ on the c coordinate and to $[0, 0.320]$ on the β coordinate. We then simulate 5000 samples of sizes $n = 100$, $n = 500$ and $n = 1000$ from this distribution and construct confidence regions for the identified set using lower probabilities $\underline{P}_n$ (based on characterization 3.3.4), which turned out to have better coverage properties. We use 999

bootstrap replications for the first two sample sizes, and 399 bootstrap replications for $n = 1000$. We consider confidence levels 90%, 95% and 99%. Coverage probabilities of the true value and of the identified set by the confidence region, as computed from the 5000 samples, are displayed in Table 3.2 for the data generating process obtained with maximal selection and Table 3.3 for the data generating process obtained with uniform selection. Alongside coverage of the identified set and of the true value, we report the effective level at which Condition (3.4.5) is satisfied to directly assess the bootstrap functional quantile approximation. Monte Carlo coverage of the identified set is close to the theoretical level in the case of maximal selection and tends to be very high in case of uniform selection. In cases of maximal and uniform selection alike, coverage of Condition (3.4.5) is almost identical to point coverage in the point identified case ($c_0 = 0$), but lower in the set identified case ($c_0 = 0.25$). Overall the procedure over rejects in all but 13 out of a possible 90 cases. Improvements with sample size occur only in 21 cases (out of a possible 60). These improvements tend to occur when going from $n = 500$ to $n = 1000$ and given the nonparametric procedure, there are doubt as to the accuracy of the procedure for $n = 100$. Finally, the coverage of the true value (as opposed to the whole identified set) is only marginally greater than the coverage of the whole identified set.

3.6 Application to long term elderly care decisions

We estimate the determinants of long term care option choices for elderly parents in American families. The model we use closely follows the one proposed by Engers and Stern (2002) who present these choices as the result of a non family participation game. The family members decide simultaneously whether to participate in a family reunion where the care option maximizing the participants' utility is chosen. Profits are then split among these participants according to some benefit-sharing rule. The data consists of a sample of $1,212$ elderly Americans with two children drawn from

Table 3.2: Coverage probabilities of (α_0, β_0) and of the identified set by the confidence region.

(α_0, β_0)	n	Level	Point Coverage	Set Coverage	Condition (3.4.5)
$(0, 0)$	100	0.99	-	0.9826	0.9792
		0.95	-	0.9574	0.9564
		0.90	-	0.9324	0.9392
	500	0.99	-	0.9894	0.9894
		0.95	-	0.9770	0.9760
		0.90	-	0.9592	0.9584
	1000	0.99	-	0.9714	0.9712
		0.95	-	0.9564	0.9554
		0.90	-	0.9362	0.9352
$(0.5, 0)$	100	0.99	0.9364	0.9364	0.9286
		0.95	0.9356	0.9354	0.9122
		0.90	0.9232	0.9220	0.8830
	500	0.99	0.9906	0.9902	0.9656
		0.95	0.9810	0.9804	0.9518
		0.90	0.9640	0.9632	0.9330
	1000	0.99	0.9878	0.9870	0.9772
		0.95	0.9746	0.9730	0.9532
		0.90	0.9594	0.9570	0.9210

As computed from 5000 samples.

The last column shows the level at which Condition (3.4.5) is satisfied.

Case, where the data generating process obtained with maximal selection.

Table 3.3: Coverage probabilities of (α_0, β_0) and of the identified set by the confidence region.

(α_0, β_0)	n	Level	Point Coverage	Set Coverage	Condition (3.4.5)
$(0, 0)$	100	0.99	-	0.9872	0.9846
		0.95	-	0.9784	0.9746
		0.90	-	0.9680	0.9652
	500	0.99	-	0.9950	0.9944
		0.95	-	0.9886	0.9872
		0.90	-	0.9814	0.9794
	1000	0.99	-	0.9790	0.9738
		0.95	-	0.9706	0.9640
		0.90	-	0.9628	0.9548
$(0.5, 0)$	100	0.99	0.9998	0.9986	0.9850
		0.95	0.9998	0.9984	0.9792
		0.90	0.9998	0.9978	0.9704
	500	0.99	1.0000	0.9996	0.9850
		0.95	1.0000	0.9974	0.9664
		0.90	1.0000	0.9980	0.9382
	1000	0.99	1.0000	0.9964	0.9792
		0.95	1.0000	0.9956	0.9694
		0.90	1.0000	0.9938	0.9578

As computed from 5000 samples.

The last column shows the level at which condition (3.4.5) is satisfied.

Case where the data generating process is obtained with uniform selection.

the National Long Term Care Survey, sponsored by the National Institute of Aging and conducted by the Duke University Center for Demographic Studies under Grant number U01-AG007198, Duke (1999). Elderly people were interviewed in 1984 about their living and care arrangements. The survey questions include gender and age of the children, the distance between homes of the elderly parent and each of the children, the disability status of the elderly parent (where disability is referred to as problems with "Activities of Daily Living or Instrumental Activities of Daily Living (ADL)") and the number of days per week each of the children devotes to the care of the elderly parent. The dependent variable is the care provision for the parent. The parent is asked to list children (either at home or away from home) and how much each provides help. If only one child is listed as providing significant help, that child is designated the primary care giver. If more than one child is listed, the one providing the most time is designated the primary care giver. If the elderly parent lives in a nursing home, then the nursing home is the primary care giver. If no child is listed and the parent does not live in a nursing home, then the parent is designated as "living alone". Table 3.4 presents the list of variables used in the analysis. They include parent characteristics, characteristics of the children and the care option chosen. A more detailed discussion and summary statistics and additional results can be found in the supplementary material.

3.6.1 The game

The observable choice of care option is modeled as in Engers and Stern (2002) as the outcome of a family bargaining game. We index family members as follows. Parent: 0, Firstborn child: 1 and Second born child: 2. The payoff to family member i, $i = 0, 1, 2$, is the sum of three terms. The first term V_{ij} is the value to parent 0 and to child i of care option j, where $j \in 1, 2$ means child j becomes the primary care giver, $j = 0$ means the parent remains self-reliant and $j = 3$, the parent is moved to a nursing home. The matrix $V = (V_{ij})_{ij}$ is known to both children and the parent.

Table 3.4: List of variables.

Variables	**Equal to 1 if:**	**Percentage of sample**
Care Option		
	Living with child 1	26.81
	Living with child 2	6.75
	Living in nursing home	19.92
	Living home alone	46.54
Parent Variables		
DA	Highly disabled	33.81
DM	Living with the spouse	40.36
Children Variables		
DD	Living with parent	11.55
$DD1$	Distance from parent: 31 min and more	49.45
DS	Female	49.26

We suppose it takes the form

$$V_{ij} = \gamma_{ij} + W\beta_{ij} + Z_j\psi_{ij}$$

where W indicates the characteristics of the parents (DA and DM), and Z_j indicates the characteristics of care option j (DS, DD1 and DD2) and $X = (W, Z)$. $\theta = (\gamma_{ij}, \beta_{ij}, \psi_{ij})'$ is unknown to the analyst and the object of inference.

Example 4 *Consider the following family, in which the matrix where given value of X and θ result in V that takes the form:*

$$V = \begin{bmatrix} 0 & 0 & 0 & -1 \\ 0 & 4 & -1 & 1 \\ 0 & -1 & 4 & 1 \end{bmatrix}$$

Rows indicate family member $i = 0, 1, 2$, and columns represents care giving options $j = 0, 1, 2, 3$, in that order. In this example, the parent is indifferent between all the care options, except the one where she has to move to the Nursing home. Each child prefers to be the primary care giver to any other care option, followed by the parent living in a nursing home, living at home and being taken care of by the other child, in that order.

The second term in the payoff results from the family bargaining process as follows. We assume that it is always in the interest of the parent to attend the family reunion. However, child i ($i = 1, 2$) can refrain from participating in the meeting. By choosing not to participate, a member of the family agrees on whatever is decided but can neither assume the role of primary care giver, nor can he be involved in any side payment. Both children simultaneously decide whether or not to participate in the long term care decision. Suppose M is the set of children who participate. The option chosen is option $j \in M \cup \{0, 3\}$ which maximizes the participants's total utility $\sum_{i \in M} V_{ij}$. It is assumed that participants abide by the decision and that benefits are then shared equally among parent and children participating in the decision through

a monetary transfer s_i, which is the second term in the children's payoff. The third term ϵ_i in the payoff is a random benefit from participation, which is 0 for children who decide not to participate and distributed according to absolutely continuous distribution $\nu(.|\theta)$ for each child who participates. All children observe the realizations of ϵ, whereas the analyst only knows its distribution. The Payoff matrix is given in Table 3.5, where overall benefit shares w_i^{IJ}, $i = 1, 2$, $I, J = N, P$ are defined and derived in the supplementary appendix. Multiple Nash equilibria in pure and mixed strategies are also derived in the appendix. Each equilibrium action profile results in a (almost surely) unique care option choice, hence for each participation shock ϵ, we can derive $G(\epsilon|X;\theta)$ as the set of probability measures on the set of care options $\{0, 1, 2, 3\}$ induced by mixed strategy profiles, which are probabilities on the set of participation profiles $\{NN, NP, PN, PP\}$.

Table 3.5: Payoffs for the family participation game.

	Child 2	
Child 1	N	P
N	w_1^{NN}, w_2^{NN}	$w_1^{NP}, \varepsilon_2 + w_2^{NP}$
P	$\varepsilon_1 + w_1^{PN}, w_2^{PN}$	$\varepsilon_1 + w^{PP}, \varepsilon_2 + w^{PP}$

3.6.2 Specification

We provide estimates for the following utility specification (an alternative with altruistic utility specification was estimated and results are reported in the supplementary

material).

$$V(X;\theta) = \begin{bmatrix} \begin{pmatrix} \beta_{00} \\ +\beta_m DM \\ +\beta_{ah} DA \end{pmatrix} & \begin{pmatrix} \alpha \\ +\psi_s DS_1 \end{pmatrix} & \psi_s DS_2 & 0 \\ \begin{pmatrix} \beta_m DM \\ +\beta_{ah} DA \end{pmatrix} & \begin{pmatrix} \beta_{11} \\ +\psi_1 DD_1 \\ +\beta_{ac} DA \end{pmatrix} & 0 & 0 \\ \begin{pmatrix} \beta_m DM \\ +\beta_{ah} DA \end{pmatrix} & 0 & \begin{pmatrix} \beta_{11} \\ +\psi_1 DD_2 \\ +\beta_{ac} DA \end{pmatrix} & 0 \end{bmatrix}$$

Recall that the columns indicate the options, in the following order $\{0, 1, 2, 3\}$, and the rows represent each member of the family, in the following order Parent, Child 1, Child 2. For example, the value the first born child (family member 1) living less than 30 minutes away from the parent's home attaches to the fact that she takes care of a non disabled, non-married parent is measured by β_{11}, whereas for a disabled parent, it is $\beta_{11} + \beta_{ac}$.

3.6.3 Estimation methodology

The methodology proposed in the paper allows the construction of the identified set based on the hypothetical knowledge of the true distribution of the data. As described in Section 3.4, we account for sampling uncertainty and control the level of confidence by constructing set functions $A \mapsto \overline{P}(A|X)$, which dominate $P(A|X)$ (uniformly over $A \subseteq \{0, 1, 2, 3\}$ and X) with probability $1 - \alpha$ (the chosen level of confidence, here 0.95). We implement the method detailed in Section 3.4 (except that the *pairs* or *cases* bootstrap was used instead of the nonparametric bootstrap advocated above) with a number of bootstrap replications $B = 2500$. Second, we obtain the model likelihood by simulating the valuation matrix and computing the Equilibrium correspondence

from the payoff matrix, for given values of X and θ. The procedure is as follow. For a given X and θ,

- We generate and store R draws of ε from the distribution ν_θ. Here, $R = 5000$ and ν_θ is normally distributed with mean μ and variance σ_ε^2, where $(\mu, \sigma_\varepsilon^2)$ belong to the parameter θ.

- For each value ε^r, we compute the valuation matrix $V(X, \varepsilon^r, \theta)$ and the corresponding payoff matrix.

- Then, we determine the equilibrium correspondence $G(X, \varepsilon; \theta)$ from the analytical results derived in the preceding section. The Gambit software provides an alternative for computing numerically the set NE for more complex games.

- The last step of the simulation is to compute an estimator of the model likelihood $\mathcal{L}$ defined in (3.4.3) as follows: $\hat{\mathcal{L}}(A|X;\theta) = \frac{1}{R}\sum_{r=1}^{R} \min\{\sigma(A) : \sigma \in G(X, \varepsilon^r; \theta)\}$.

Having constructed those two elements, the identified set comprises all values of θ such that for all observed values of the explanatory variables, the minimum over $A \subseteq \{0, 1, 2, 3\}$ of the function $\overline{P}(A|X;\theta) - \hat{\mathcal{L}}(A|X;\theta)$ is non negative, as explained in Section 3.4. We construct an n-dimensional grid to conduct the search over the parameter space. Each value of the parameter can be tested in a fraction of a second on a standard laptop, and a region of small dimensionality (1 to 4) can be constructed in a few hours, again on a standard laptop without parallel processing. However, estimation time grows exponentially with the number of parameters induced by the model. In our case, each specification involves a 12-dimensional parameter space. Parallel processing becomes therefore necessary. We use an Open-MP procedure for parallel processing, which is perfectly suited to the method we propose. The computation resources have been provided by the Réseau Québécois de Calcul de Haute Performance (RQCHP). All computation where made under the system "Cottos" which provides

up to 128 computation nodes (1024 CPU cores) equipped with two Intel Xeon E5462 quad-core processors at 3 GHz. Under 1 node, approximately 10^7 parameters points can be tested in 24 hours.

3.6.4 Results

We perform the estimation under different values of the mean and variance of the error term. To alleviate the computational burden, we first test the significance of some of the individual parameters by checking whether the hyper planes defined by $\theta_i = 0$ - where θ_i is a component of θ - intersect the 95% confidence region. We fail to reject the Null Hypothesis if the estimation procedure returns a non-empty set. We then obtain a constrained confidence region for the remaining parameters. For each value of mean and variance of the error term, we find a non empty intersection between the confidence region and the hyperplane defined by $\beta_{11} = 0$. This means we fail to reject (at the 5% level) the null hypothesis that there is no additional constant disutility for a child to take care of an elderly parent. Since, this hypothesis is not rejected, we obtain a constrained confidence region for the remaining parameters. We then obtain confidence regions for different values of β_{11} and discuss the latter's effect on the regions. We note that the Null hypothesis $H_0 : \beta_{00} = 0$ is always rejected. Hence, when we control for all other effects, parents are not indifferent between the first two options. They show a clear preference in favor of living in their own home (option called "living alone") instead of living in a nursing home (β_{00} is always positive). The results we present are then for given values of β_{00}. We provide an insight of how different values of this parameter change the results. We report the range for each parameters in Table 3.6. Note that the identified set is not a compact set. In particular, β_{ac}, β_{ah}, β_m and ψ are allowed to diverge to $-\infty$. Results are generally consistent with expectations and previous results on the subject. Namely:

1. The existence of several problems with the parent's functional ability is a key determinant of the decision to enter a nursing home. β_{ah} and β_{ac} are both

Table 3.6: Parameters Range for estimation of Specification 1 at $\beta_{11} = 0$, $\beta_{ac} = -\beta_m$ and for different values of the error terms and of β_{00}.

Parameters	Min	Max	Min	Max	Min	Max	Min	Max
β_{00}	2	2	3	3	1	1	1	1
β_{11}	0	0	0	0	0	0	0	0
β_{ah}	$-\infty$	-3.57	$-\infty$	-3.57	$-\infty$	-2.86	$-\infty$	-2.14
$\beta_{ac} = -\beta_m$	$-\infty$	-2.86	$-\infty$	-2.86	$-\infty$	-3.57	$-\infty$	-3.57
α	0.00	8.00	0.00	8.00	1.00	5.00	0.00	4.00
ψ_s	0.00	5.00	0.00	4.00	1.00	4.00	0.00	2.00
ψ_d	$-\infty$	-2.86	$-\infty$	-2.14	$-\infty$	-1.43	$-\infty$	-3.57
μ	-1	-1	0	0	-1	-1	0	0
σ_ε	1	1	1	1	1	1	1	1
σ_u^2	1	1	1	1	0.25	0.25	0.25	0.25
p_ξ	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1

negative and can both be (very) large. The negative sign of β_{ah} captures the fact that a parent's disability increases the value of care provided by the family or a specialized institution. In addition, $\beta_{ac} < 0$ means that the disability entails a utility cost for the child if he is chosen as primary care giver.

2. Parameter β_m associated with the parent living with a spouse is positive and large. This implies that married parents are more likely to remain self-reliant. In families where the parent is disabled, the effect of living with the spouse compensates the disutility of disability and preserves the incentive for parents to live at home.

3. While we cannot rule out parents being indifferent to the gender or birth order of their primary care giver, estimation shows a tilt of the confidence interval toward positive values for both parameters, with a possible positive and large magnitude of the parameter α. In case $\mu = -1$ and $\sigma_u^2 = 0.25$, the data reveal that parents exhibit a preference for an older and for a female care giver.

4. Children living more than 30 minutes from the parents are less likely to provide care than those living closer to the parents. Distance has a (possibly strong) disutility effect on children's incentives to participate in the care decision.

The shape of the confidence region also conveys a considerable amount of information. Figure 3.3 shows two dimensional dimensional projections and cuts of the confidence region for column 2 of Table 3.6, i.e $\mu_\varepsilon = 0$, $\sigma_\varepsilon^2 = 1$, $\sigma_u^2 = 1$. Of great interest is the projection of the identified set in the plan β_{ah}, β_m. Figure 3.3(a) reveals an almost linear relation between the two parameters of the type $\beta_{ah} = -\beta_m$. The estimation rejects models for which the absolute value of the two parameters are significatively different. The data suggest therefore that the disutility induced by the disability of the parent can be entirely compensated by the presence of a spouse in the same household. Notice the triangular shape of the region plotted in Figure 3.3(b) which entails that simultaneous large values of ψ_s and α are rejected. This finding means

that only one of the effects (gender or birth order) can be large, not both. In other words, firstborn daughter are not the only possible care givers. Note also that both effects can be very small, though not jointly insignificant. We observe similar types of constraints for the pairs (α, β_{ah}), (α, β_{ac}), (α, ψ_d), (ψ_s, β_{ac}), (ψ_s, ψ_d) as large values of parameters α or ψ_s are only permitted when the other parameters are jointly large (see Figure 3.3(c) to 3.3(f)). For example, we obtain a constrained confidence region at $\beta_{ac} = -3.5$. The ranges for the two parameters, α and ψ_s, are tighter, as $\alpha \in [1, 2]$ and $\psi \in [0, 1]$. Figure 3.4 shows the effect of the variation of parameter β_{11} on ψ_s and α. Recall that β_{11} represents a fixed cost or benefit for the child chosen as care giver. We observe negative relations between β_{11} and ψ_s, and β_{11} and α. Negative values of ψ_s and α are only admissible for positive values of β_{11}. Hence a model where parents exhibit no favoritism for a daughter and/or a firstborn, or favoritism for a son and/or a second born, will be consistent with our data if and only if there exist a strictly positive constant benefit for a child to be caregiver.

3.7 Conclusion

We have considered the problem of statistical inference in incomplete partially identified structural models, such as models of discrete choice with interactions and other forms of endogeneity. A characterization of the identified set for structural parameters was given, with an appeal to a classical theorem in combinatorial optimization, the Max-Flow Min-Cut Theorem, thereby emphasizing the optimization formulation of the problem of inference in such models. Finally, we have shown how to apply combinatorial optimization methods within a bootstrap procedure in order to compute informative confidence regions very efficiently, hence feasibly in empirically relevant applications. An application of the methodology was carried out on a family bargaining example and it was shown that most findings in the literature on the determinants of long term elderly care by American families were supported in this more robust

framework, where the effects of interaction are accounted for. This procedure applies to very general classes of models and its efficiency and coverage properties could no doubt be improved, when tailored to more specific applications. In particular, the application to matching games and revealed preference testing of stability in matching still poses considerable challenges. Other perspectives for further work include the application of Max-Flow Min-Cut algorithms to the detection of redundant inequalities at the identification stage, to improve the performance at the inference stage, possibly by appealing to other existing procedures if the number of non redundant inequalities is small enough.

3.8 Appendix

3.8.1 Proofs of results in the main text

Proof 1 (Proof of Theorem 3) *By Proposition 1 of Galichon and Henry (2011), a value θ of the parameter vector belongs to Θ_I if and only if $\mathbb{P}(Y \in G(X, \varepsilon; \theta)) = 1$, X-a.s. (which we drop from the notation from this point on). Hence if there exists a pair (Y, U) of random vectors on $\mathcal{Y} \times \mathcal{U}$ such that Y has probability mass $P(y|X)$, $y \in \mathcal{Y}$, U has probability mass $Q(u|X;\theta)$, $u \in \mathcal{U}$, and $\mathbb{P}(Y \in U|X) = 1$. This is equivalent to the existence of non negative weights π_y^u, $(y,u) \in \mathcal{Y} \times \mathcal{U}$, such that $\sum_{u \in \mathcal{U}} \pi_y^u = P(y|X)$, $\sum_{y \in \mathcal{Y}} \pi_y^u = Q(u|X)$, and $\pi_y^u = 0$ when $y \notin u$. The latter is equivalent to the following programming problem with auxiliary variables a_y, $y \in \mathcal{Y}$ and a^u, $u \in \mathcal{U}$ having zero as a solution. The programming problem is the following:* $\min(\sum_{y \in \mathcal{Y}} a_y + \sum_{u \in \mathcal{U}} a^u)$ *subject to the constraints $\sum_{u \in \mathcal{U}} \pi_y^u + a_y \leq P(y|X)$, $\sum_{y \in \mathcal{Y}} \pi_y^u + a^u \leq Q(u|X;\theta)$, a_y, a^u, $\pi_y^u \geq 0$, and $\pi_y^u = 0$ when $y \notin u$. Since $\sum_{y \in \mathcal{Y}} a_y + \sum_{u \in \mathcal{U}} a^u \leq \sum_{y \in \mathcal{Y}} P(y|X) + \sum_{u \in \mathcal{U}} Q(u|X;\theta) - 2\sum_{y \in \mathcal{Y}} \sum_{u \in \mathcal{U}} \pi_y^u = 2 - 2\sum_{y \in \mathcal{Y}} \sum_{u \in \mathcal{U}} \pi_y^u$, the latter is also equivalent to $\max \sum_{y \in \mathcal{Y}} \sum_{u \in \mathcal{U}} \pi_y^u \geq 1$ subject to the constraints $\sum_{u \in \mathcal{U}} \pi_y^u \leq P(y|X)$, $\sum_{y \in \mathcal{Y}} \pi_y^u \leq Q(u|X)$, $\pi_y^u \geq 0$ and $\pi_y^u = 0$ when $y \notin u$. This is called a maximum flow problem, i.e. the problem of maximizing quantity flowing through a network under capacity*

constraints. A network is a collection of nodes, including a source S *and a* sink T*, and directed edges between the nodes. For instance,* (N_1, N_2) *is an edge leading from node* N_1 *to node* N_2*. Here the network involved in the maximum flow problem is comprised of a source* S*,* K *nodes corresponding to the* K *elements of* $\mathcal{Y}$*,* J *nodes corresponding to the* J *elements of* $\mathcal{U}$ *and a sink* T*. The source* S *is connected to each of the nodes* $y_1, \ldots, y_k$ *in* $\mathcal{Y}$*. A node* $y \in \mathcal{Y}$ *is connected to a node* $u \in \mathcal{U}$ *if and only if* $y \in u$*. All nodes* $u_1, \ldots, u_J$ *in* $\mathcal{U}$ *are connected to the sink* T*. To each edge is attached a capacity, which is the maximum amount that can flow through it. Capacity is constrained to* $P(y|X)$ *between* S *and node* y*. Capacity is unconstrained (i.e. infinite) between node* y *and node* u *such that* $y \in u$*. The capacity of edges between a node* u *and the sink* T *is constrained to* $Q(u|X;\theta)$*.*

We have shown that $\theta \in \Theta_I$ *if and only if the maximum flow in the network described above is equal to* 1*. We now appeal to a classical result in combinatorial optimization called the* Max-Flow Min-Cut Theorem*, see for instance Theorem 10.3 page 150 of Schrijver (2004). A* cut *through a network is partition of the nodes into two sets separating the source from the sink. The* capacity *of a cut is defined as the sum of the capacities of edges in the network that cross the cut from the source side to the sink side. Let a cut be defined by the set* V *of elements of* $\mathcal{U}$ *and the set* Z *of elements of* $\mathcal{Y}$ *on the sink side of the cut. Since the capacity of an edge from* y *to* u *such that* $y \in u$ *is infinite, the cut defined by* V *and* Z *has finite capacity if and only if* $y \in u$ *and* $u \in V$ *jointly imply* $y \in Z$*. Such a cut has capacity* $C(Z, V) = \sum_{y \in Z} P(y|X) + \sum_{u \in \mathcal{U} \setminus V} Q(u|X;\theta) = \sum_{y \in Z} P(y|X) + 1 - \sum_{u \in V} Q(u|X;\theta)$*. A cut has minimum capacity if no node can be moved between the source side of the cut and the sink side of the cut without increasing capacity, hence if* $y \notin u$ *and* $u \in V$ *jointly imply* $y \notin Z$*, hence if* $Z = V^{\cup} = \bigcup \{u : u \in V\}$*. Therefore, the capacity of a minimum cut is* $C(V^{\cup}, V) = \sum_{y \in V^{\cup}} P(y|X) + 1 - \sum_{u \in V} Q(u|X;\theta) = P(V^{\cup}|X) + 1 - Q(V|X;\theta)$*. By the Max-Flow Min-Cut Theorem, the capacity of any minimum cut is equal to the maximum flow through the network, hence* $\theta \in \Theta_I$ *if and only if for all subset* V *of* $\mathcal{U}$*,* $P(V^{\cup}|X) + 1 - Q(V|X;\theta) \geq 1$*, i.e.* $Q(V|X;\theta) \leq P(V^{\cup}|X)$*, and the result follows.*

Proof 2 (Proof of Lemma 1) *Take an $x \in \mathcal{X}$. Take any $u \in \mathcal{U}$ and $V \subseteq \mathcal{U}\backslash\{u\}$. We have $P([V \cup \{u\}]^{\cup}|x) - P(V^{\cup}|x) = \sum_{y \in \bigcup_{v \in V \cup \{u\}} v} P(y|x) - \sum_{y \in \bigcup_{v \in V} v} P(y|x) = \sum_{y \in u \backslash V^{\cup}} P(y|x) = P(u\backslash V^{\cup}|x)$, which is non-increasing in V, hence the result.*

Proof 3 (Proof of Theorem 4) *Given a value $\theta \in \Theta_I$, by Theorem 3, we have $\sup_{x \in \mathcal{X}} \max_{V \subseteq \mathcal{U}} (Q(V|x;\theta) - P(V^{\cup}|x)) \leq 0$. Under Assumption 5, $\sup_{x \in \mathcal{X}} \max_{V \subseteq \mathcal{U}} (P(V^{\cup}|x)) - \overline{P}_n(V^{\cup}|x)) \leq 0$, with limiting probability larger than $1-\alpha$. Hence, with probability at least $1-\alpha$, $\sup_{x \in \mathcal{X}} \max_{V \subseteq \mathcal{U}} (Q(V|x;\theta) - \overline{P}_n(V^{\cup}|x)) \leq 0$, and thus $\theta \in \Theta_I(\overline{P}_n)$.*

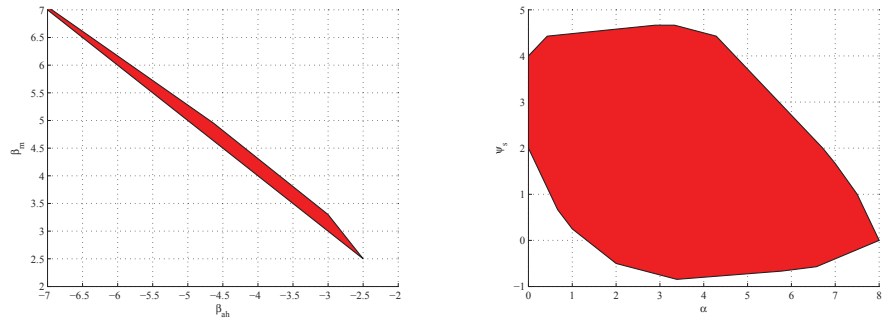
Proof 4 (Proof of Proposition 5) *We first justify the BRS Step 1 by showing that $w_b = w'_b$ for all b. Indeed observe that for any $j \in \{1, \ldots, n\}$ and $A \subseteq \mathcal{Y}$, we have*

$$\zeta_n^b(A|X_j) = \sum_{y \in A} \hat{P}_n(y|X_j) - \sum_{y \in A} \hat{P}_n^b(y|X_j) = \sum_{y \in A} [\hat{P}_n(y|X_j) - \hat{P}_n^b(y|X_j)]$$

and thus $\max_{A \subseteq \mathcal{Y}} \zeta_n^b(A|X_j)$ is attained by selecting all the elements $y \in \mathcal{Y}$ with $\hat{P}_n(y|X_j) - \hat{P}_n^b(y|X_j) > 0$. It follows that $w'_b = \max_{1 \leq j \leq n} \max_{A \subseteq \mathcal{Y}} [\sum_{y \in \mathcal{Y}} \hat{P}_n(y|X_j) - \sum_{y \in A} \hat{P}_n^b(y|X_j)]$ and therefore $w_b = w'_b$. To justify BRS Step 2, let w^{opt} denote the optimum objective value of problem (3.4.6). If D fails to include any b such that $w_b > w^{opt}$ then $\max_{b \notin D} w_b > w^{opt}$, therefore an optimal D must include all b such that $w_b > w^{opt}$. On the other hand, if D is any optimal subset and some $b' \in D$ satisfies $w_{b'} \leq w^{opt}$ then discarding b' from D yields a feasible subset $D\backslash\{b'\}$ (since $|D\backslash\{b'\}| < |D| \leq \overline{d}$) such that $\max_{b \in D\backslash\{b'\}} w_b \leq \max_{b \in D} w_b$ hence $D\backslash\{b'\}$ is an alternate optimal solution. Therefore an optimal D consists of all indices b such that $w_b > w^{opt}$. Concerning the running time, BRS Step 1 requires $O(nB|\mathcal{Y}|)$ time, and BRS Step 2 requires $O(B)$ time using a linear time selection (or median-finding) algorithm (see Blum, Floyd, Pratt, Rivest, and Tarjan (1973)).

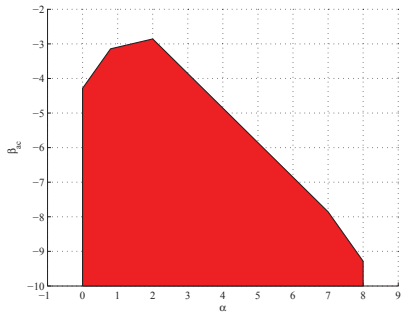
Proof 5 (Proof of Proposition 6) *The problem corresponds to the following decision problem: given an $n \times m$ matrix H, an integer k and a target value t, can one find a subset $S \subseteq \{1, \ldots, n\}$ such that $|S| \geq k$ and $\sum_{i=1}^{m} \max_{j \in S} H_{ij} \leq t$? Denote (H, k, t) an instance of the latter problem. Consider the well-known NP-hard*

decision problem CLIQUE (see for instance section 4.8 page 43 of Schrijver (2004)): given a graph $G = (V, E)$ and an integer q satisfying $2 \leq q \leq |V|$, does there exist a subset $Q \subseteq V$ such that $|Q| \geq q$ and for all $i, j \in V$, $ij \in E$ (i.e. Q is a clique*). To any instance (G, q) of the problem CLIQUE, we associate an instance (H, k, t) of our decision problem, where lines of H corresponds to vertices of G (elements of V), columns of H corresponds to edges in G (elements of E) and $H_{ij} = 1$ if vertex i belongs to edge j, and 0 otherwise. For any subset $S \subseteq E$ of edges in G, we have for all $i \in E$, $\max_{j \in S} H_{ij} = 1$ if i belongs to at least one element of S, and 0 otherwise. Hence, $\sum_{i \in E} \max_{j \in S} H_{ij}$ is the number of vertices that belong to at least one edge in S. Define $k = q(q-1)/2$ and $t = q$. Then, a set S of k edges involves at least (hence exactly) q vertices if and only if S is the set of edges of a CLIQUE. Hence the answer to the decision problem (H, k, t) thus defined is YES if and only if G contains a CLIQUE with q vertices. Since CLIQUE is NP-complete, it follows that our decision problem is NP-hard. Since $k = O(|V|^2)$ and $t = O(|V|)$, the input size (in unitary notation) of such instances of our problem is polynomially bounded by the input size (in unitary or binary notation) $\Omega(|V|)$ of the corresponding instance of CLIQUE. Hence our decision problem is NP-hard in the strong sense.*

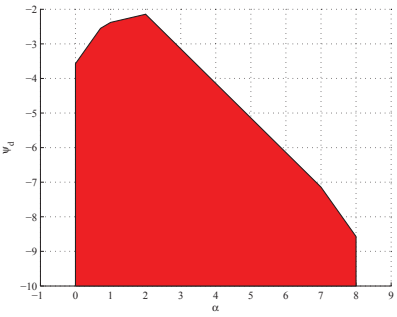


(a) (β_{ah}, β_m) region

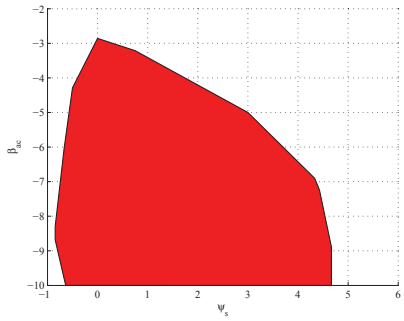
(b) (α, ψ_s) region



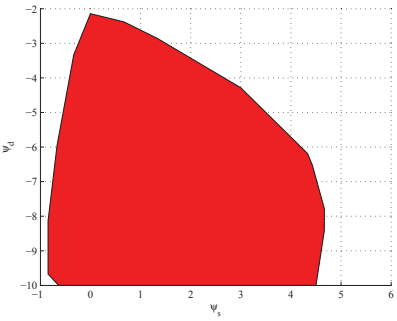
(c) (α, β_{ac}) region



(d) (α, ψ_d) region

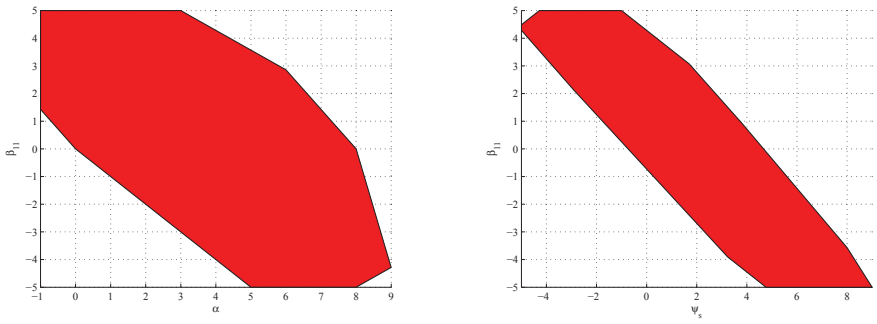


(e) (ψ_s, β_{ac}) region



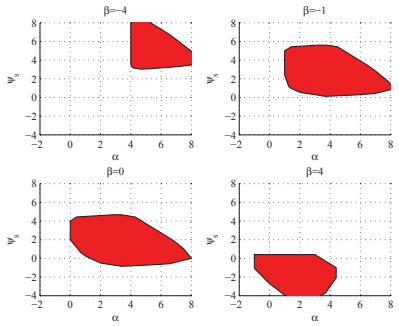
(f) (ψ_s, ψ_d) region

Figure 3.3: Two dimensional representations of the confidence region at $\beta_{00} = 3$, $\beta_{11} = 0$, $\mu = 0$, $\sigma_\varepsilon = 1$, $\sigma_u = 1$, $p_\xi = 0.1$



(a) (α, β_{11}) region

(b) (ψ_s, β_{11}) region



(c) (α, ψ_s) regions for different values of β_{11}

Figure 3.4: Parameter β_{11} in relation with other parameters: $\beta_{00} = 3$, $\mu = 0$, $\sigma_\varepsilon = 1$, $\sigma_u = 1$, $p_\xi = 0.1$

Conclusion générale

Cette thèse s'intéresse à deux aspects importants de la migration des étudiants internationaux : les déterminants de la probabilité de migration et l'organisation de la famille afin de couvrir les frais de la migration.

Le Chapitre 1 présente un cadre de travail pour analyser avec une perspective microéconomique les déterminants du choix de migration étudiante. Nous y arrivons en surmontant les difficultés présentes jusqu'à présent dans la littérature, notamment : l'absence de données microéconomiques comprennant une population d'étudiants migrants et non-migrants, et les problème d'identification des paramètres structurels induit par l'engonéité du niveau final d'éducation dans notre modèle à choix discret. Pour résoudre la première de ses difficultés, nous avons recours à un sondage de type boule de neige, qui utilise une plateforme internet. La procédure proposée permet de surmonter le défi de la dispersion géographique des unités au sein de la population d'intérêt. Nous proposons un estimateur de la moyenne de la population qui permet de corriger les biais liés à la non-réponse et à la selection des répondants. En pratique, des hypothèses fortes sont nécessaires pour calculer les probabilités d'inclusion des individus échantillonnés. Nous devons aussi estimer un modèle de graphe aléatoire pour représenter les liens entre les individus de la population d'intérêt. Les choix dans ce chapitre ont été faits pour des raisons computationnelles.

Par ailleurs, nous proposons une modèle structurel de décision d'investissement dans la migration étudiante. Ce modèle reflète à la fois le rendement de l'investissement et la contrainte budgétaire qui y est attachée. Une importante contribu-

tion de cette étude est l'usage des récents résultats de la littérature au sujet des modèles incomplets et partiellement identifiés pour contourner le problème d'endogénéité cité plus-haut. Ce faisant, nous relaxons la condition d'identification ponctuelle des paramètres de notre modèle. Nous effectuons l'inférence en utilisant des inégalités de moments et calculons des intervalles de confiance relativement informatifs. Une contribution supplémentaire de notre étude est le développement d'une procédure d'inférence qui prend en compte les données censurées ou incomplètement observées. Cette procédure a notamment l'avantage de réduire considérablement la contrainte computationnelle.

A partir de notre échantillon d'étudiants Camerounais, nous trouvons qu'un diplôme Masters et de bons résultats au cours des études secondaires augmentent les chances de migration étudiante. Il ne semble pas que les étudiants de sexe masculin soient favorisés par la famille au détriment de ceux de l'autre sexe. Nous trouvons aussi que le premier-né a une probabilité plus faible d'émigrer que ces plus jeunes frères et soeurs. Notre interprètation est qu'il souffre d'un manque de soutien familial.

Le chapitre 2 porte son attention au processus de décision au sein de la famille. En particulier, nous nous interrogeons sur les incitations des membres de la famille à participer à ce processus coûteux. Nos données sont très informatives sur la distribution des frais de l'investissement entre les membres de sa famille et sur l'identité des contributeurs. Les parents sont sollicités dans la grande majorité des ménages. Plus de la moitié des répondants reporte le soutien potentiel d'un aidant dans le processus de migration. Le modèle structurel proposé décrit un jeu non-coopératif de participation entre les membres de la famille élargie. Nous utilisons encore une fois les résultats de la littérature en identification partielle pour déduire des inégalités de moments qui servent ensuite à l'inférence sur les paramètres structurels. Les résultats suggèrent que les personnes impliqués dans le financement de la migration étudiante ne devraient pas être considérés comme ayant des préférences homogènes. Toutefois, il serait tout autant erroné d'y voir un rassemblement d'individus égoïstes. Les préférences des

parents exhibent de l'altruisme en faveur de l'enfant. Les aidants extérieurs à la famille nucléaire subissent un coût strictement positif pour leur participation, ce qui décourage leur implication. Les obligations familiales et sociales semblent expliquer les cas de participation d'un aidant, mieux qu'un possible altruisme de ce dernier.

Finalement, le troisième chapitre présente le cadre théorique plus général dans lequel s'imbriquent les modèles développés dans les précédents chapitres. Avec nos co-auteurs, nous y avons considéré le problème d'inférence statistique de modèles structurels incomplets et partiellement identifiés, tels que le modèle à choix discret developpé au Chapitre 1. Une caractérisation de l'ensemble identifié pour les paramêtres structurels est fournie, en s'appuyant sur un théorème classique d'optimisation combinatoire, le théorème "Max-Flow Min-Cut". Finalement, nous montrons comment appliquer les méthodes d'optimisation combinatoires au sein d'une procédure bootstrap afin de calculer des régions de confiance informatives de façon efficace. Une application de cette méthodologie sur l'étude des déterminants des soins apportés aux parents agés dans des familles américaines conduit à la plupart des résultats de la littérature, avec cette procédure plus robuste.

La procédure d'échantillonnage proposé semble une solution efficace au défi de la collecte de données sur des poulations de migrants. L'étude des performances sous des hypothèses alternatives de cette procédure et des estimateurs qui l'accompagnent apparaît comme une question de recherche intéressante. Les résultats sur l'importance de l'implication de la famille dans le processus de décision suggère qu'une attention plus poussée soit accordée au phénomène de migration en chaine et à ses conséquences sur la qualité des migrants.

Bibliographie

Ackerberg, D., L. Benkard, S. Berry, and A. Pakes (2007) : "Econometric tools for analyzing market outcomes," *Handbook of Econometrics*, Volume 6A.

Ahuja, R. K., T. L. Magnanti, and J. B. Orlin (1993) : *Network Flows : Theory, Algorithms, and Applications*. Prentice Hall.

Andrews, D., S. Berry, and P. Jia (2003) : "Placing bounds on parameters of entry games in the presence of multiple equilibria," unpublished manuscript.

Andrews, D., and G. Soares (2010) : "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 78, 119–157.

Aradillas-Lopez, A. (2010) : "Semiparametric estimation of a simultaneous game with incomplete information," *Journal of Econometrics*, 157, 409–431.

Bajari, P., J. Hahn, H. Hong, and G. Ridder (2011) : "A note on semiparametric estimation of finite mixtures of discrete choice models with applications to game theoretic models," *International Economic Review*, 52, 807–824.

Bajari, P., H. Hong, and S. Ryan (2010) : "Identification and estimation of a discrete game of complete information," *Econometrica*, 78, 1529–1568.

Banerjee, B. (1983) : "Social Networks in the Migration Process : Empirical Evidence on Chain Migration in India," *The Journal of Developing Areas*, 17(2), pp. 185–196.

Batista, C., A. Lacuesta, and P. C. Vicente (2012) : "Testing the 'brain gain' hypothesis : Micro evidence from Cape Verde," *Journal of Development Economics*, 97(1), 32–45.

Bauer, T. K., G. Epstein, and I. Gang (2000) : "What are migration networks ?," IZA Discussion paper.

Becker, G. S. (1973) : "A Theory of Marriage : Part I," *Journal of Political Economy*, 81(4), 813–46.

Beine, M., F. Docquier, and Ç. Özden (2011) : "Diasporas," *Journal of Development Economics*, 95(1), 30–41.

Beine, M., F. Docquier, and H. Rapoport (2008) : "Brain drain and human capital formation in developing countries : winners and losers," *Economic Journal*, 118(528), 631–652.

Beine, M., R. Noel, and L. Ragot (2012) : "The determinants of international mobility of students," CESifo Working Paper : Economics of Education, No. 3848.

Beresteanu, A., I. Molchanov, and F. Molinari (2011) : "Sharp identification regions in models with convex predictions," *Econometrica*, 79, 1785–1821.

Berry, S. (1992) : "Estimation of a model of entry in the airline industry," *Econometrica*, 60, 889–917.

Berry, S., and E. Tamer (2006) : "Identification in models of oligopoly entry," in *Advances in Economics and Econometrics*, pp. 46–85. Cambridge University Press.

Bessey, D. (2011) : "International student migration to Germany," *Empirical Economics*, 42(1), 345–361.

Bickel, P., and D. Freedman (1981) : "Some asymptotic theory for the bootstrap," *Annals of Statistics*, 9, 1196–1217.

Blum, M., R. Floyd, V. Pratt, R. Rivest, and R. Tarjan (1973) : "Time bounds for selection," *Journal of Computational Systems Science*, 7, 448–461.

Borjas, G. J. (1989) : "Economic Theory and International Migration," *International Migration Review*, 23(3), pp. 457–485.

Borjas, G. J., and B. Bratsberg (1994) : "Who Leaves ? The Outmigration of the Foreign-Born," *SSRN eLibrary*.

Boyd, M. (1989) : "Family and personal networks in international migration : recent developments and new agendas," *International Migration Review*, pp. 638–670.

Bresnahan, T., and P. Reiss (1990) : "Entry in monopoly markets," *Review of Economic Studies*, 57, 531–553.

Brezis, E., and A. Soueri (2011) : "Why students migrate ? Where do they migrate to ?," Working Papers 25, AlmaLaurea Inter-University Consortium.

Bugni, F. (2010) : "Bootstrap inference in partially identified models defined by moment inequalities : coverage of the identified set," *Econometrica*, 78, 735–753.

Cameron, A. C., and P. K. Trivedi (2005) : *Microeconometrics : Methods and Applications*. New York : Cambridge University Press.

Chernozhukov, V., H. Hong, and E. Tamer (2007) : "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75, 1243–1285.

Chernozhukov, V., S. Lee, and A. Rosen (2009) : "Inference on intersection bounds," CEMMAP Working Paper CWP19/09.

Chesher, A. (2010) : "Instrumental variable models for discrete outcomes," *Econometrica*, 78, 575–601.

Chesher, A., A. Rosen, and K. Smolinski (2011) : "An instrumental variable model of multiple discrete choice," CEMMAP Working Paper CWP06/11.

Chiappori, P.-A. (1992) : "Collective Labor Supply and Welfare," *Journal of Political Economy*, 100(3), pp. 437–467.

Ciliberto, F., and E. Tamer (2009) : "Market structure and multiple equilibria in airline markets," *Econometrica*, 70, 1791–1828.

Docquier, F., and A. Marfouk (2006) : "International migration by educational attainment, 1990 - 2000," *Caglar Ozden, Maurice Schiff, Editors , International Migration, Remittances, and the Brain Drain*.

Docquier, F., and H. Rapoport (2009) : "The economics of brain drain," unpublished manuscript.

Dreher, A., and P. Poutvaara (2011) : "Foreign students and migration to the United States," *World Development*, 39(8), 1294–1307.

Duke (1999) : "National Long Term Care Survey," Public use data set produced and distributed by the Duke University Center for Demographic Studies with funding from the National Institute on Aging under Grant No. U01-AG007198.

Echenique, F., S. Lee, M. Shum, and B. Yenmez (2011) : "The revealed preference theory of stable and extremal stable matchings," unpublished manuscript.

Ekeland, I., A. Galichon, and M. Henry (2010) : "Optimal transportation and the falsifiability of incompletely specified economic models.," *Economic Theory*, 42, 355–374.

Engers, M., and S. Stern (2002) : "Long-term care and family bargaining," *International Economic Review*, 43(1), 73–114.

Galichon, A., and M. Henry (2006a) : "Dilation Bootstrap," unpublished manuscript.

——— (2006b) : "Inference in incomplete models," available from SSRN at http ://papers.ssrn.com/sol3/papers.cfm ?abstract_id=886907.

——— (2009) : "A test of non-identifying restrictions and confidence regions for partially identified parameters," *Journal of Econometrics*, 152, 186–196.

——— (2011) : "Set identification in models with multiple equilibria," *Review of Economic Studies*, 78, 1264–1298.

Haile, P., and E. Tamer (2003) : "Inference with an Incomplete Model of English Auctions," *Journal of Political Economy*, 111, 1–51.

Hanson, G. H., and C. Woodruff (2003) : "Emigration and educational attainment in Mexico," *Mimeo*.

Heckathorn, D. D. (1997) : "Respondent-Driven Sampling : A New Approach to the Study of Hidden Populations," *Social Problems*, 44(2), pp. 174–199.

——— (2007) : "Extensions of Respondent-Driven Sampling : Analyzing Continuous Variables and Controlling for Differential Recruitment," *Sociological Methodology*, 37, pp. 151–208.

Henry, M., R. Méango, and M. Queyranne (2011) : "Combinatorial bootstrap inference in partially identified incomplete structural models," unpublished manuscript.

Henry, M., and I. Mourifié (2011) : "Euclidean revealed preferences : testing the spatial voting model," *Journal of Applied Econometrics*, forthcoming.

Henry, M., and A. Onatski (2012) : "Set coverage and robust policy," *Economics Letters*, 115(2), 256–257.

Isiugo-Abanihe, U. C. (1985) : "Child Fosterage in West Africa," *Population and Development Review*, 11(1), 53–73.

Jia, P. (2008) : "What Happens When Wal-Mart Comes to Town : An Empirical Analysis of the Discount Retailing Industry," *Econometrica*, 76, 1263–1316.

Jovanovic, B. (1989) : "Observable implications of models with multiple equilibria," *Econometrica*, 57, 1431–1437.

Li, H., M. Rosenzweig, and J. Zhang (2010) : "Altruism, Favoritism, and Guilt in the Allocation of Family Resources : Sophie's Choice in Mao's Mass Send-Down Movement," *Journal of Political Economy*, 118(1), 1–38.

Lundberg, S., and R. A. Pollak (1994) : "Noncooperative Bargaining Models of Marriage," *The American Economic Review*, 84(2), pp. 132–137.

Mahmood, T., and K. Schömann (2003) : "On the Migration Decision of IT-Graduates : A Two-Level Nested Logit Model," *WZB Markets and Political Economy Working Paper No. SP II*, 22.

Mallar, C. D. (1977) : "The estimation of simultaneous probability models," *Econometrica : Journal of the Econometric Society*, pp. 1717–1722.

Manser, M., and M. Brown (1980) : "Marriage and Household Decision-Making : A Bargaining Analysis," *International Economic Review*, 21(1), pp. 31–44.

Manski, C. (1993) : "Identification of endogenous social effects : the reflection problem," *Review of Economic Studies*, 60, 531–542.

McElroy, M. B., and M. J. Horney (1981) : "Nash-Bargained Household Decisions : Toward a Generalization of the Theory of Demand," *International Economic Review*, 22(2), pp. 333–349.

McKenzie, D., and H. Rapoport (2007) : "Network effects and the dynamics of migration and inequality : Theory and evidence from Mexico," *Journal of Development Economics*, 84(1), 1 – 24.

Nemb, P. S., and U. E. Jumbo (2011) : "Banks and tontines : Complementarity or Competition ? The case of Cameroon," *International Finance and Business Journal*, 1(1), 30–40.

Njike Njikam, G. B., R. Lontchi Tchoffo, and V. Fotzeu Mwaffo (2005) : "Caractéristiques et déterminants de l'emploi des jeunes au Cameroun," unpublished manuscript.

Pakes, A., J. Porter, K. Ho, and J. Ishii (2004) : "Moment inequalities and their application," unpublished manuscript.

Perkins, R., and E. Neumayer (2011) : "Educational mobilities in an age of internationalization : quality, social ties and border controls in the uneven flows of foreign students," in *EUGEO session of the Royal Geographical Society Annual International Conference, London*, vol. 31.

Rainer, H., and T. Siedler (2009) : "O brother, where art thou ? The effects of having a sibling on geographic mobility and labour market outcomes," *Economica*, 76(303), 528–556.

Rao, J. N. K., and C. F. J. Wu (1988) : "Resampling Inference With Complex Survey Data," *Journal of the American Statistical Association*, 83(401), pp. 231–241.

Romano, J., and A. Shaikh (2008) : "Inference for identifiable parameters in partially identified econometric models," *Journal of Statistical Planning and Inference*, 139, 2786–2807.

——— (2010) : "Inference for the identified set in partially identified econometric models," *Econometrica*, 78, 169–211.

Rosen, A. (2008) : "Confidence sets for partially identified parameters that satisfy a finite number of moment inequalities," *Journal of Econometrics*, 146, 107–117.

Rosenzweig, M. R. (2008) : "Higher Education and International Migration in Asia : Brain Circulation," in *Annual World Bank Conference on Development Economics*, pp. 59–100.

Rosenzweig, M. R., D. A. Irwin, and J. G. Williamson (2006) : "Global Wage Differences and International Student Flows [with Comments and Discussion]," in *Brookings Trade Forum*, pp. 57–96. JSTOR.

Schrijver, A. (2004) : *Combinatorial optimization : polyhedra and efficiency*. Springer : Berlin.

Singh, K. (1981) : "On the asymptotic accuracy of Efron's bootstrap," *Annals of Statistics*, 9, 1187–1195.

Soetevent, A., and Kooreman (2007) : "A discrete choice model with social interactions : with an application to high school teen behaviour," *Journal of Applied Econometrics*, 22, 599–624.

Spence, M. (1973) : "Job market signaling," *The Quarterly Journal of Economics*, 87(3), 355–374.

Tamer, E. (2003) : "Incomplete simultaneous discrete response model with multiple equilibria," *Review of Economic Studies*, 70, 147–165.

Thissen, L., and S. Ederveen (2006) : *Higher education : Time for coordination on a European level ?*, no. 68. CPB Netherlands Bureau for Economic Policy Analysis.

Thompson, S. K. (2006) : "Adaptive Web Sampling," *Biometrics*, 62(4), 1224–1234.

Uetake, K., and Y. Watanabe (2011) : "Entry by acquisition : estimates from a two-sided matching model with externality," unpublished manuscript.

Van Bouwel, L., and R. Veugelers (2009) : "The determinants of student mobility in Europe : the quality dimension," *FBE Research Report MSI_0912*, pp. 1–39.

Wejnert, C., and D. D. Heckathorn (2008) : "Web-based network sampling : Efficiency and efficancy of Respondent-Driven Sampling for online research," *Sociological Methods and Research*.

Willis, R. J. (1985) : *Wage determinants : A survey and reinterpretation of human capital earnings functions*. Economics Research Center/NORC.