When a Random Sample is Not Random. Bounds on the Effect of Migration on Household Members Left Behind*

Andreas Steinmayr†

Job Market Paper. This version 2/3/2015

Abstract

A key problem in the literature on the economics of migration is how emigration of an individual affects households left behind. Answers to this question must confront a problem I refer to as *invisible sample selection*: when entire households migrate, no information about them remains in their source country. Since estimation is typically based on source country data, invisible sample selection yields biased estimates if all-move households differ from households that send only a subset of their members abroad. I address this identification problem and derive nonparametric bounds within a principal stratification framework. Instrumental variables estimates are biased, even if all-move households do not differ in their potential outcomes. For this case, I show identification of the local average treatment effect. I illustrate the approach using individual and household data from widely cited, recent studies. Potential bias from invisible sample selection can be large, but transparent assumptions regarding behaviors of household members and selectivity of migrants allow identification of informative bounds.

Keywords: Sample selection, migration, selectivity, principal stratification

JEL classification: C21, F22, J61, O15

*I thank Dan Black, Stéphane Bonhomme, Xavier D'Haultfoeuille, Alfonso Flores-Lagunes, Jeff Grogger, Martin Huber, Michael Lechner, Toru Kitagawa, David McKenzie, Giovanni Mellace, Elisabeth Sadoulet, and Steven Stillman for helpful discussions. I thank Christina Felfe, Rebecca Lordan, Damon Jones, Elie Murard, Toman Omar Mahmoud, Mathias Reynaert, André Richter, Anthony Strittmatter, Petra Thiemann, and Conny Wunsch for helpful comments on the draft, and participants at the Labor Economics Seminar in Lech, SOLE, 6th Migration & Development Conference, DAGStat, NOeG, Norface Migration Conference and seminars at Chicago, Syracuse, and St. Gallen for helpful comments. All errors are my own.

†Harris School of Public Policy Studies, University of Chicago. *asteinmayr@uchicago.edu*. Most of this paper was written during my doctoral studies at the University of St. Gallen. This study was generously funded by the Swiss National Science Foundation. An earlier version of the paper was titled "When a Random Sample is Not Random. Bounds on the Effect of Migration on Children Left Behind" (first version: May 2013).

1 Introduction

With more than 232 million international migrants worldwide (United Nations, 2013), migration implies high costs and benefits for both source and destination countries. Economic research has only recently devoted more attention to the source countries of migrants. An intense academic and policy debate goes on whether the costs of migration outweigh the benefits for source countries, for example with respect to gains and losses in human capital (Docquier and Rapoport, 2012).[1] For households in developing countries, migration is a potential strategy to increase household income and diversify income sources. Migration often splits households, with some household members migrating and remitting money to members left behind in the source country. Although income effects are widely accepted to be positive (McKenzie, Gibson, and Stillman, 2010), the overall effect of migration on household welfare is ambiguous since absence of household members might affect household welfare negatively. A burgeoning literature investigates effects of migration on various aspects of welfare of household members left behind, such as children's health and educational attainment, labor supply of spouses, and household poverty.[2]

Identification of the effects of migration on remaining household members faces two major selection problems. Most of the literature address the non-random selection of households into migration (i.e., which households send migrants). However, selection within households poses another source of endogeneity that has almost entirely been ignored (i.e., which members of the household migrate). The identification problem is further complicated if all household members migrate (*all-move households*). In that case, the household will not be included in data collected in the source country, i.e., the household is not included in household surveys since no household member was left to respond to a survey. I refer to this issue as *invisible sample selection*.

The following stylized research design illustrates the complications these selection problems create for identification. For simplicity, consider one-adult-one-child households. Assume the adult participates in a visa lottery, and migrates if he wins and stays if he does not. Due to random assignment of the visa, adult migration is unrelated to household characteristics, and thus selection of households into migration does not pose a problem for identification. The decision to take the child along, however, is not random. It is taken by the household, and depends on household characteristics. Assume for example that only wealthier households can take the child, and poorer households leave the child behind.[3] Children left behind in the sample of lottery winners live, on average, in less wealthy households than children in non-migration households. This negative correlation

[1] See media coverage in *The Economist* (2011, May 28). "Drain or gain?"

[2] Antman (2013) provides a comprehensive overview of the literature on the effects of migration on remaining household members and the empirical strategies employed in this literature.

[3] The direction of the selectivity is irrelevant for the resulting sample selection problem.

biases estimates obtained by comparing outcomes of children left behind in migrant and non-migrant households even though adult migration is randomly assigned. Consider further that outcome data are collected by a household survey in the source country after the lottery occurs. Wealthier households disappear from the sample of lottery winners since no individual remains to respond to the survey. Sample selection becomes invisible since, in the absence of data on the initial population, no information is recorded on these all-move households.

One reason this issue has not received much attention might be that the problem that arises for identification of causal effects is not obvious since sample selection in cross-sectional data is *invisible*. It is not apparent that households that leave no members behind are relevant for identification of causal effects of migration on *remaining* household members. However, as the example above illustrates, selection within households and related migration of whole households constitute sample selection problems that can lead to biased estimates.[4]

This paper contributes to the literature in several ways. First, I address the identification problems induced by selection between and within households and invisible sample selection by applying the statistical concept of *principal stratification* (Frangakis and Rubin, 2002) to model behaviors of household members and selectivity of migrants. This approach allows a clear discussion of the assumptions made implicitly about the selection process if the second form of selection is ignored. I derive non-parametric bounds on the effects of migration on household members left behind under transparent sets of behavioral and distributional assumptions. Second, I show that results from previous studies that ignore invisible sample selection might suffer from substantial bias. I replicate results from a study with well-identified point estimates of migration effects on household composition and household assets (Gibson, McKenzie, and Stillman, 2011a). I compare these results to the bounds derived in my paper, which identify effects for a broader population and can be applied if non-migrants cannot be identified based on observable characteristics. The bounds suggest that estimates not taking into account invisible sample selection might understate the true magnitude of the effects of migration for some outcomes (e.g., agricultural assets) and indicate significant effects on outcomes for which the bounds can not reject a zero effect (e.g., number of elderly individuals living in a household). In a second application, I revisit the effect of migration on the educational attainment of children left behind in Mexico (McKenzie and Rapoport, 2011). I consider that observational data miss migrat-

[4]The problem of whole households moving and not being included in source country data has been acknowledged in studies that estimate the overall number of emigrants (Ibarraran and Lubotsky, 2007) or migrant selectivity (McKenzie and Rapoport, 2007) based on source country data.

Gibson, McKenzie, and Stillman (2011a) exploit specific visa regulations, which allow identification of the effects of migration for household members who are not eligible to migrate and therefore always stay behind. This approach addresses selection within the household and also the case of all-move households at the cost of focusing on a relatively specific subpopulation.

ing children in all-move households, and derive bounds under varying assumptions. The bounds indicate that unadjusted point estimates are likely lower bounds of the true effects. Overall, the empirical examples show that a combination of behavioral and distributional assumptions provides substantial identifying power to derive informative bounds.

This paper also contributes to literature on sample selection that derives from Gronau (1974) and Heckman (1974), and in particular to literature on partial identification (Manski, 1989, 1994). It extends the literature on sample selection by studying a situation during which for some units, not only the outcome is unobserved, but also the units are not included in cross-sectional data at all. A random sample of households drawn after migration began is not representative of the original population. An important finding is that instrumental variable estimates are biased even if there is no systematic selection within the household. It suffices that complier households differ in their potential outcomes from households that do not comply with the instrument (always and never migrating households) for estimates to be biased. For this scenario, I propose an alternative estimator. The paper also contributes to related literature on mediation analysis, which decomposes individual causal mechanisms from overall causal effects.[5] In the previous example, migration of the child does not only create a sample selection problem but is a treatment in itself. Migration of the adult can be seen as the main treatment and migration of the child as a mediator, a channel through which adult migration affects child outcomes. Identifying the effect of adult migration with the child staying behind corresponds to identifying the direct effect of adult migration and ruling out indirect effects through child migration.

The paper builds on the approach of principal stratification, introduced by Frangakis and Rubin (2002) to deal with post-treatment complications in bio-medical literature (e.g., death of patients during drug evaluation). Principal stratification allows characterizing the potential – not the observed – behaviors of household members, which in turn allows transparency regarding assumptions needed for identification of causal effects. Several recent papers use principal stratification to derive bounds on the effects of policy interventions in the presence of post-treatment complications.[6] To derive bounds in a setting with non-compliance and sample selection, I use an approach from Chen and Flores (2012). To my best knowledge, principal stratification has neither been used to model interactions between units, nor in a migration context.

Invisible sample selection can also appear if other sources of data are used or in completely different settings. Administrative school data could, for example, provide outcome

[5]For rather general mediation models, see for example Pearl (2001) and Albert and Nelson (2011). Recent studies in economics mostly evaluate mechanisms through which active labor market policies work (Flores and Flores-Lagunes, 2009, 2010; Huber, 2014; Huber, Lechner, and Mellace, 2014). Heckman, Pinto, and Savelyev (2013) investigate mechanisms through which the Perry Preschool program affects outcomes later in life.

[6]Recent papers include Zhang and Rubin 2003; Mattei and Mealli 2007; Zhang, Rubin, and Mealli 2008; Huber, Laffers, and Mellace 2014; Chen and Flores 2012; Blanco, Flores, and Flores-Lagunes 2013a,b.

measures for children. Children who migrate are not enrolled and therefore not included in the data. More generally, often researchers only obtain data from a population that is already affected by a treatment, with the treatment affecting not only the outcome but potentially also the composition of the population. For example, Almond (2006) investigates the long-term effects of the 1918 influenza pandemic using U.S. census data from the second half of the 20th century. Selective attrition due to death constitutes a potential selection problem that leads the sampled population to be different from the "treated" population.[7] Studies concerning intergenerational effects are particularly prone to this type of sample selection, as a treatment affecting the parent generation might not only affect child outcomes but also fertility of the parent generation and thus the composition of observed children (for a discussion of endogenous fertility decisions see Heckman and Mosso, 2014). Consider another example, how selective outmigration affects the composition of rural populations. Observed changes in poverty rates between two census waves can either be driven by changing poverty rates of a stable population or a change in the composition of the populations (World Bank, 2008).

The remainder of the paper is structured as follows. Section 2 discusses intra-household selection and invisible sample selection. Section 3 introduces an econometric framework to structure the identification problem and bound the effects of interest, relating to the introductory example. To focus on the second selection problem, I first assume randomly assigned migration of the principal migrant. In a second step, I investigate the problem in an instrumental variables setting. Section 4 illustrates the approach for the effects of migration on households in Tonga and the effect of adult migration on school attendance of children in Mexico. Section 5 discusses possible extensions and variations of this approach. Section 6 concludes.

2 The effect of migration on household members left behind and the invisible sample selection problem

I illustrate the selection problems and the proposed approach using the example from the introduction, in which interest is in the effect of adult migration on child outcomes, which several studies address. Researchers usually investigate the case in which one parent (or another adult member of the household) migrates and the child remains in the source location. Equation (1) displays a stylized linear model common in migration literature. Y_{ij} denotes an outcome of child i in household j. $hmig_j$ is a binary indicator whether the household has at least one adult member living abroad (for simplicity, assume households

[7] In an earlier version, Almond provides an extensive discussion about the direction – though not the exact magnitude – of potential bias (Almond, 2005).

have only one adult individual), and u_{ij} is an error term.

$$Y_{ij} = \beta_0 + \beta_1 hmig_j + u_{ij} \tag{1}$$

The selection problem addressed in most studies is non-random selection of households into migration. Households that send a migrant might, for example, be wealthier and therefore find it easier to finance the costs of migration. Members of these households might also differ in terms of education, demographics, or preferences from members of non-migrant households. Many factors that drive a migration decision might also influence decisions of monetary and time investments in child-raising, which lead to an endogeneity problem. Thus, the concern is whether the error term correlates with the variable of interest ($E[hmig_j u_{ij}] \neq 0$). Various strategies have been implemented to address this endogeneity, including selection on observables (e.g., Kuhn, Everett, and Silvey, 2011), instrumental variables (e.g., Hanson and Woodruff, 2003; McKenzie and Hildebrandt, 2005; McKenzie and Rapoport, 2011), or fixed-effects approaches (e.g., Antman (2012) uses family fixed-effects). For a discussion of the various approaches used in the literature, see Antman (2013).

However, in some households that migrate, not only one individual migrates, but several or even all household members migrate (see Gibson, McKenzie, and Stillman, 2013, 2011a, for a related discussion). Also, the child might be among the migrants, which gives rise to two problems. First, outcomes for children who migrate are not observed or not well defined. Second, children who stay behind and for whom we observe the outcome are a selected group that might differ in its characteristics from children who migrate. This complication worsens regarding the way data are normally collected; household surveys in source countries ask respondents whether one or several household members are currently abroad. Households that answer yes are referred to as migrant (*treated*) households. Households that answer no are referred to as non-migrant (*control*) households. However, if the whole household migrates, no individual is left to answer the survey, and those households are not included in cross-sectional datasets. Thus, we can estimate only Equation (2), where s_j is a binary selection indicator, which is one if the household is observed and zero if the household is not observed (i.e., if all household members migrated).[8]

$$s_j Y_{ij} = \beta_0 s_j + \beta_1 s_j hmig_j + s_j u_{ij} \tag{2}$$

Instead of assuming $hmig_j$ to be uncorrelated with the error, this model requires that $E[hmig_j s_j u_{ij}] = 0$. The migration status of the adult ($hmig_j$) needs to be uncorrelated with the error in the sample of observed children. Assume the migration status of the

[8] I will elaborate on the selection problem in the case of households with more than two members in more detail in Section 3.1.

adult household member is assigned randomly, and migration of the child is the choice of the household, whereby children would not migrate without the adult. Due to random assignment, $hmig_j$ is uncorrelated with u_{ij}. After households learn about their assigned $hmig_j$, they decide whether the child should migrate ($s_j = 0$) or stay ($s_j = 1$). If households have a disutility from being separated and if migration is costly, richer households are more likely to migrate with the child. Observed children in the treated group are therefore, on average, poorer than children in households that are unobserved. At the same time, household wealth has a positive influence on child welfare (Almond and Currie, 2011). In the observed sample, $hmig_j$ correlates negatively with u_{ij} and estimates of Equation (2) are biased negatively.

In panel data, when entire households migrate between two waves of data collection, the existence of the household is at least documented in the earlier wave. However, it might not always be possible to distinguish migration and other forms of attrition.

Vast econometric literature that dates back to Gronau (1974) and Heckman (1974) deals with the problem of sample selection for the identification of causal effects. The literature addresses the problem that outcomes (e.g., wages) are not observed for part of the population (e.g., the unemployed). Broadly, this literature developed two solutions to the selection problem. The first approach uses latent variable models as the Heckman selection model (Heckman, 1979) to correct for selection bias, but requires strong parametric assumptions or a valid instrument. The second approach, based on Manski (1989, 1994), derives bounds on the quantities of interest. Such bounds can be derived under various – usually weaker – assumptions. Two complications set the current paper apart from existing literature on sample selection. First, the unit of analysis is defined less clearly in the context of the effect of migration on remaining household members. The treatment is the migration of one or several household members. This treatment changes the composition of the remaining household members; in the counterfactual situation, migrants are among household members. Second, literature on sample selection assumes researchers have a random sample of units with observed treatment state and covariate values, where for some units, the outcome is unobserved. All-move households however are not included in data collected after migration begins. In the example above, a random sample of households drawn from the population after migration starts is unrepresentative of the initial population since all-move households are not included in the data.

Sample selection is only one problem that arises if children are among migrants. Assume we observe child outcomes even if all household members migrate (e.g., by collecting data on the child in the destination country from peers in other households). We could obtain unbiased estimates from Equation (1), which would correspond to the overall effect of adult migration, with the outcome for some children measured in the source and for others in the destination country. However, migrating as a family from one country to another is a different treatment than migration of an adult when the child stays behind. Child migration is both, a selection indicator and a treatment/mediator. This paper focuses on the direct effect of adult migration, ruling out indirect effects through child migration.

Recently, use of (quasi-) experiments for future research on migration has been encouraged strongly (McKenzie and Yang, 2010; McKenzie, 2012). However, as randomization usually addresses only the first source of endogeneity (which households engage in migration), the second form (who and how many household members migrate) is a problem in experimental settings as well. The solution of the few papers that use visa lotteries to account for the first form of endogeneity and address the second form of endogeneity has been to define a different parameter of interest and estimate the effect for only those households and household members that can be identified as *never migrants* based on observable characteristics. Gibson, McKenzie, and Stillman (2011a) and Gibson, McKenzie, and Stillman (2013) use visa rules that dictate which household members are allowed to migrate with the principal migrant. In their setting of migration from Tonga and Samoa to New Zealand, they removed all eligible individuals from the estimation sample. They therefore estimate the effect for individuals who are ineligible to join the principal migrant and are therefore always observed. This subgroup consists primarily of siblings, nephews, nieces, and parents of the migrant. Estimating the effect of migration on the migrant's nuclear family is not possible using their approach. In studies based on observational panel data, several papers recognize the second form of endogeneity and provide some discussion on how severe the problem could be, but do not address it (Yang, 2008; Antman, 2011).

3 Econometric framework

This section introduces the econometric approach to structure the identification problem. First, I introduce the setup and parameters of interest. In a second step, I concentrate on the second selection problem by assuming randomly assigned migration status of the principal migrant. In a third step, I show identification in a setting with an instrumental variable for migration of the principal migrant and sample selection problems induced by migration of other household members.

3.1 Setup and parameter of interest

Following treatment evaluation literature, I use a potential outcome framework developed by Rubin (1974). The idea is to compare the outcome of interest in two hypothetical states of the world: one in which a unit receives the treatment, and one in which the same unit does not. In the setting under investigation, we might ask whether a particular child would attend school if he lived in a migrant household and whether the same child would attend school if he did not live in a migrant household. The problem is that only one of these two situations can be observed in the real world. Suppose that households consist of two individuals (I_1, I_2). With reference to the second empirical application, I refer to these individuals as principal migrant/adult (I_1) and accompanying migrant/child (I_2). In empirical applications, I_1 and I_2 may refer to different entities, not only individuals. In the Tongan application, I_1 refers to the individual, who applied for a visa, and I_2 refers to the household as a whole. The unit of analysis is the household. In the Mexican application, I_1 refers to any adult in the household, and I_2 refers to an individual child. The unit of analysis is the child. The setup can be applied to different forms of intra-household selection, including a scenario in which the household dissolves as all members migrate.

$M_j = m_j \epsilon \{0, 1\}$ denotes the migration status of individual j. I_1 makes the first migration decision and chooses either to stay ($M_1 = 0$) or migrate ($M_1 = 1$). I discuss the general selection problem under the simplifying assumption of randomly assigned M_1. I_2 chooses to stay ($M_2 = 0$) or migrate ($M_2 = 1$) depending on the choice of I_1. This does not necessarily have to be a sequential decision process nor the decision of I_2, but can also be a household decision. Crucial is that M_2 is a function of M_1.

If migration of the principal migrant is considered the treatment of interest, migration of children might be considered a post-treatment complication. The econometric literature usually refers to this type of complication as *endogenous sample-selection* (Gronau, 1974; Heckman, 1974); those for whom the outcome (i.e., stayers) is observed are endogenously selected, and this selection is a function of treatment.

I observe the outcome Y (e.g., school attendance of the child) at some point after M_1 and M_2 have been realized. I define a set of potential outcomes for Y and M_2. Y is a function of M_1 *and* M_2. Y depends on M_1 since migration of an adult household member is likely to affect the educational attainment of the child. Y depends on M_2 since migration of the child also affects educational attainment. $Y(m_1, m_2)$ denotes the potential values of the outcome. $Y(0, 0)$ is the outcome of the child in case no member of the household migrates; $Y(1, 0)$ is the outcome in case the adult migrates and the child stays behind; $Y(0, 1)$ is the outcome in case the adult stays and the child migrates; and $Y(1, 1)$ is the outcome if the adult migrates and takes the child with her. Similarly, $M_2(m_1)$ denote the potential migration state of I_2 as a function of migration of I_1. $M_2(0)$ is the migration state of the child if the adult stays, and $M_2(1)$ is the migration state of the child if the adult migrates.

I assume having a *random* sample of households from the population in the source country, drawn after the households were treated (i.e., individuals migrated). The sample and population do not include households in which both adult and child migrate ($M_1 = 1$, $M_2 = 1$). Although the sample is representative of the population at that point, the observed population is different from the population at the time the treatment was assigned.

If only a subset of household members migrate, the household is still observed but intra-household selection problems prevail if a child is among the migrants.

In this setting we distinguish several effects. The difference $Y(1,0)-Y(0,0)$ is the effect of adult migration if the child stays (i.e., the partial effect of M_1 on Y for M_2 being zero). I focus on $Y(1,0)-Y(0,0)$ since this effect is most policy relevant and has received the most attention in the literature.[9] If we do not assume treatment effect homogeneity, we must define the population for which we want to identify the effect. I focus on children who would always stay behind even if the adult migrates (i.e., children for whom $M_2(0) = M_2(1) = 0$). This is a latent group, and whether an individual belongs to this group is unobservable since only either $M_2(0)$ or $M_2(1)$ can be observed, but not both. I focus on this group since it is the only group for which the outcome is observed under both migration states of the adult. In countries with predominantly labor migration in which only a small fraction of households migrates with the children, it is also quantitatively the most important group. The average partial effect of M_1 for children who would never migrate is defined as

$$\theta \equiv E\left[(Y(1,0) - Y(0,0))\,|M_2(0) = 0, M_2(1) = 0\right]. \tag{3}$$

This definition disqualifies interactions between units of households, an assumption referred to commonly as Stable Unit Treatment Value Assumption (SUTVA) (Rubin, 1980). In most applications, SUTVA implies that potential outcomes of a unit are independent of treatment statuses of any other units. In this setting, it implies that potential outcomes of a child are unaffected by treatment of units in other households.

3.2 Identification with randomly assigned migration of principal migrant

To focus on the identification problem induced by the migration of I_2, I assume random assignment of the migration status of I_1. From the random assignment of M_1, it follows that all *potential* outcomes are independent of M_1 (Assumption 1).

Assumption 1. *Randomly assigned migration status of* I_1

$$\{Y(m_1, m_2), M_2(m_1)\} \perp M_1 \textit{ for all } m_1, m_2 \,\epsilon\, \{0,1\}$$

3.2.1 Stratification on potential migration behavior

Consider the potential migration behavior of I_2. Based on the joint value of the potential migration behavior $(M_2(0), M_2(1))$, children can be stratified into four latent groups (Table 1). Following Frangakis and Rubin (2002), I refer to these groups as *principal strata*, sub-populations of units (in this case, households) that share the same potential values

[9] I discuss several other effects in Section 5.

of intermediate variables under various treatment states. We can distinguish four combinations of potential migration behaviors of I_2 (Table 1). These four types correspond to the classification in the Local Average Treatment Effects (LATE) framework (Imbens and Angrist, 1994; Angrist, Imbens, and Rubin, 1996). In the LATE framework, the types describe potential behaviors of units regarding an instrumental variable. In my setting, the types describe the potential migration behaviors of the child concerning the migration status of the adult. With reference to the LATE framework, I refer to the types (G) as *(A)lways migrants*, *(C)ompliers*, *(D)efiers*, and *(N)ever migrants*. Children characterized as *always migrants* would migrate, irrespective of the migration status of the adult. *Compliers* would migrate if the adult migrates, but stay if the adult stays. *Defiers* would migrate if the adult stays, and stay if the adult migrates. *Never migrants* would always stay. These four principal strata are hypothetically possible combinations of the potential values of M_2. Not all strata necessarily exist in reality.

[Table 1 about here]

Principal stratification compares units within principal strata. Since treatment assignment does not affect membership to a principal stratum, the estimated effects are causal effects (Frangakis and Rubin, 2002). A principal stratum carries the information whether a child would migrate or stay if the adult migrates or stays. Conditional on the principal strata, potential outcomes $Y(m_1, m_2)$ are independent of the treatment M_1. This implication is substantially different from the notion that potential outcomes are independent of treatment M_1 given the observed migration status of I_2. The problems for identification become more obvious in Table 2, which shows the correspondence between observed groups and latent strata. The observed group $O(0,0)$ with $M_1 = 0$ and $M_2 = 0$ is comprised of compliers and never migrants (Column (1)). Similar for the other observed groups: the observed group $O(0,1)$ is comprised of always migrants and defiers, the observed group $O(1,0)$ is comprised of defiers and never migrants, and the observed group $O(1,1)$ is comprised of always migrants and compliers.

Ignoring the second selection problem leads to estimation of $E\left[Y|M_1 = 1, M_2 = 0\right] - E\left[Y|M_1 = 0, M_2 = 0\right]$. However, this implies taking the difference between strata D and N under treatment and strata C and N under control. The assumption one would have to make to give this difference a causal interpretation is that the potential outcomes under control are equal for compliers and never migrants, and that they are equal under treatment for defiers and never migrants, which is a strong, and in most scenarios, implausible assumption.

[Table 2 about here]

As explained above, a principal effect within a stratum is a well-defined causal effect. One can estimate the effect for a specific stratum, and the average partial effect for never migrants is defined as

$$\theta_N \equiv E\left[\left(Y\left(1,0\right)-Y\left(0,0\right)\right)|G=N\right]. \tag{4}$$

This is identical to the effect defined in Equation 3, and I focus on identification of this effect. To complete the notation, let π_A denote the share of always migrants, π_C the share of compliers, π_D the share of defiers, and π_N the share of never migrants.

3.2.2 Bounds on the treatment effect

To derive bounds on θ_N, I impose additional behavioral assumptions. One weak assumption in the setting where I_2 is a child is that I_2 would not migrate alone. If the household has more than one adult, this assumption means that the child would not migrate unless at least one adult migrates. This assumption disqualifies the existence of always migrants and defiers since children in these two strata would migrate if the adult would not migrate.

Assumption 2. *I_2 only migrates if I_1 migrates*

$$M_2(0) = 0$$

Column (2) in Table 2 shows the correspondence between observed groups and latent strata under Assumption 2. This assumption has empirically testable implications. Since Assumption 2 disqualifies defiers and always migrants we should not observe any households with the combination $M_1 = 0$ and $M_2 = 1$, meaning any household in which the adult stays and only the child migrates.[10] Given Assumption 2, group $O(1,0)$ corresponds directly to the stratum of never migrants under treatment. Therefore, the mean potential outcome under treatment for never migrants is identified as

$$E\left[Y(1,0)|G=N\right] \quad = \quad E\left[Y|M_1=1,M_2=0\right]. \tag{5}$$

The observed outcome in group $O(0,0)$ is a mixture of the potential outcomes of compliers and never migrants under control

$$E\left[Y|M_1=0,M_2=0\right] \quad = \quad E\left[Y(0,0)|G=C\right]\pi_C + E\left[Y(0,0)|G=N\right]\pi_N. \tag{6}$$

This expression can be transformed to obtain the potential outcome of never migrants under control

$$E\left[Y(0,0)|G=N\right] \quad = \quad \frac{E\left[Y|M_1=0,M_2=0\right]-E\left[Y(0,0)|G=C\right]\pi_C}{\pi_N}. \tag{7}$$

[10] For the bounds derived below, a weaker monotonicity assumption that rules out defiers would be sufficient. Only strata proportions in Equation 6 need to be adjusted by dividing by $(\pi_C + \pi_N)$. I use Assumption 2 since it is necessary for identification in the setting in which migration of the adult is not random, and it is a plausible assumption in this setting.

The share of compliers and never migrants could be obtained directly from $\pi_C = P(M_2 = 1|M_1 = 1)$ and $\pi_N = P(M_2 = 0|M_1 = 1)$ if the existence of households in which all individuals migrated is known and the absence of individuals in remaining households is recorded. However, information about the existence of these households is usually unavailable in cross-sectional datasets. In this case, strata proportions must be calculated using other data sources as demonstrated in the empirical examples.

I calculate γ, the ratio of the number of all-move households $O(1,1)$ to the observed number of migrant households $O(1,0)$. If no external information on all-move households is available, γ can be used to investigate sensitivity of results with respect to sample selection. I.e., it can be tested which values of γ allow ruling out certain values of θ_N. Based on γ , I can calculate strata proportions $\pi_N = {}^1/_{1+\gamma}$ and $\pi_C = {}^\gamma/_{1+\gamma}$.

Following Zhang and Rubin (2003) and Lee (2009), sharp bounds[11] on $E[Y(0,0)|G = N]$ and θ_N can be derived. The observed group of households in which neither the adult nor child migrated ($O(0,0)$) consists of the two latent groups of never migrants and compliers with proportions π_N and π_C. The two extreme scenarios we can imagine are a) the outcome of the worst complier is better than the outcome of the best never migrant. In this case we can remove the upper π_C quantiles from the distribution of Y in cell $O(0,0)$ and estimate the average outcome for the remaining individuals, which gives us the lowest possible outcome for never migrants under control. The opposite scenario b) would be that the outcome of the best complier is worse than the outcome of the worst never migrant. Removing the lower π_C quantiles from the distribution and estimating the mean gives us the upper bound for the outcome of never migrants under control. Let $q(a)$ be the a-quantile of the distribution of $Y|M_1 = 0, M_2 = 0$. $E[Y(0,0)|G = N]$ can be bounded from above by the mean of Y in the upper $1 - \pi_N$ quantiles of the distribution in the cell $O(0,0)$, and from below by the mean in the lower π_N quantiles[12] (see Appendix B.1).

The lower and upper bounds on $E[Y(0,0)|G = N]$ are

$$\begin{aligned} E_N^L[Y(0,0)|G = N] &= E[Y|M_1 = 0, M_2 = 0, Y < q(\pi_N)] \\ E_N^U[Y(0,0)|G = N] &= E[Y|M_1 = 0, M_2 = 0, Y > q(1 - \pi_N)] \end{aligned}$$

[11] Bounds are sharp if they are the tightest bounds one could obtain given the available data and assumptions made.

[12] If Y is discrete, the occurrence of mass points with equal outcome values cause the quantile function to be non-unique. For this reason, I replace the non-unique quantile function with a modified version as Kitagawa (2009) and Huber, Laffers, and Mellace (2014) suggest. Intuitively, I use a rank function instead of a quantile function to break ties. I sort data in observed cell $M_1 = 0, M_2 = 0$ on the outcome. For the lower bound, I estimate the mean in the subsample of the first $\pi_N * N_{00}$ observations, where N_{00} denotes the number of observations with $M_1 = 0, M_2 = 0$. For the upper bound, I estimate the mean in the subsample of the last $\pi_N * N_{00}$ observations.

and for the corresponding causal effects

$$\begin{aligned}\theta_N^U &= E\left[Y|M_1=1,M_2=0\right]-E_N^L\left[Y(0,0)|G=N\right]\\ \theta_N^L &= E\left[Y|M_1=1,M_2=0\right]-E_N^U\left[Y(0,0)|G=N\right].\end{aligned}$$

3.3 Identification with non-random migration of the principal migrant

Many empirical studies use an instrument for the migration decision of the principal migrant.[13] Therefore, I study identification with an instrumental variable in more detail. Assume a binary instrument $Z = z\,\epsilon\,\{0,1\}$ exists, which is assigned randomly and affects the migration decision of the principal migrant. $M_1(z)$ denotes the potential migration of I_1 as a function of the value of instrument Z. Let us for the moment also write the potential values of migration of the child $M_2(m_1,z)$ and the outcome $Y(m_1,m_2,z)$ as functions of Z. In the presence of the second selection problem, we must modify the classical IV assumptions (Imbens and Angrist, 1994; Angrist, Imbens, and Rubin, 1996). Assumption 3 suggests that the instrument is assigned randomly and therefore independent of all potential outcomes.

Assumption 3. *Randomly assigned instrument*

$$\{Y(m_1,\, m_2,\, z),\, M_2(m_1,\, z),\, M_1(z)\} \perp Z \; for\; all\; z, m_1, m_2\,\epsilon\,\{0,1\}$$

Assumption 4 suggests that the effect of Z on the potential outcomes Y must be through an effect of Z on M_1 and M_2 (the effect of Z on M_2 is indirect through M_1). The instrument affects outcomes only through its effect on the migration status of the household members.

Assumption 4. *Exclusion restriction of Z with respect to Y*

$$Y(m_1,m_2,z) = Y(m_1,m_2,z') = Y(m_1,m_2)\; for\; all\; m_1,\; m_2,\; z\,\epsilon\,\{0,1\}$$

Assumption 5 suggests that the effect of the instrument on the potential migration status of I_2 must be through an effect of Z on M_1. The decision of the household of whether only the adult or also the child migrates does not depend on the value of the instrument. Assumptions 4 and 5 allow us to use the previous notation of potential outcomes and write the potential variables $M_2(m_1)$ and $Y(m_1,m_2)$ as a function of migration status only.

[13] See for example Hanson and Woodruff (2003); McKenzie and Hildebrandt (2005); Yang (2008); Amuedo-Dorantes, Georges, and Pozo (2010); McKenzie and Rapoport (2011); Antman (2011); Gibson, McKenzie, and Stillman (2011b).

Assumption 5. *Exclusion restriction of Z with respect to M_2*

$$M_2(m_1, z) = M_2(m_1, z') = M_2(m_1)\; for\; all\; m_1, m_2, z\,\epsilon\,\{0, 1\}$$

Assumption 6 suggests that the instrument has a non-zero average effect on the migration of I_1. For the moment, I do not assume anything about the direction of the effect.

Assumption 6. *Non-zero average effect of Z on M_1*

$$E\left[M_1(1) - M_1(0)\right] \neq 0$$

A valid instrument must satisfy Assumptions 3, 4, 5, and 6 simultaneously (Imbens and Angrist, 1994; Angrist, Imbens, and Rubin, 1996). An important difference regarding the exclusion restriction is that I require Z to be a valid instrument for Y and M_2 (similar to Imai (2007) and Chen and Flores (2012)). However, there are two differences to the settings in these papers. First, in my setting, M_2 is both an indicator of whether the individual is observed and a treatment in itself. The identified effect can therefore be seen as the net or direct effect of adult migration. In Imai (2007) and Chen and Flores (2012), the outcome is not a function of the selection indicator. Second, in my setting, the probability to observe a household decreases with adult migration since this increases the probability that the entire household migrates. In the studies mentioned above, the probability of observing the outcome increases for treated individuals. However, since this is a symmetric problem, it does not affect identification.

I distinguish principal strata with respect to the *instrument.* We can differentiate the types of adults regarding the instrument as always migrants (A), compliers (C), defiers (D), and never migrants (N). An adult who is an always migrant would migrate irrespective of the value of the instrument. A complier would migrate if the instrument takes a value of one but not if it takes zero. A defier would migrate if the instrument is zero but not if the instrument is one. A never migrant would not migrate irrespective of the value of the instrument. We can also distinguish these four types of children. I define the types of children also with respect to the instrument, even though I assume that the effect works only indirectly through M_1. Combining the four strata of adults with the four strata of children gives in total $4 \times 4 = 16$ principal strata (latent household types) (see Table 3 in Appendix A). I refer to the strata using a two-letter system; the first letter refers to the type of I_1, the second to the type of I_2 (e.g., CN refers to a household in which the principal migrant is a complier and the child would never migrate).

Assumption 5 disqualifies the existence of strata AC, AD, NC, ND. In these strata, the instrument has a direct effect on M_2 since I_1 does not react to the instrument in these strata. I continue to assume that the child would only migrate if the principal migrant migrates (Assumption 2). This assumption disqualifies the existence of strata CA, CD,

DA, DC, NA, NC, ND.[14] Again, this has the empirically testable assumption that no households with $M_1 = 0$ and $M_2 = 1$ should be observed. I assume a monotone effect of the instrument on migration of I_1, which is a standard assumption in the instrumental variables literature (Imbens and Angrist, 1994; Angrist, Imbens, and Rubin, 1996). This assumption suggests that every principal migrant is at least as likely to migrate if $Z = 1$ as he would be if $Z = 0$.

Assumption 7. *Individual-level monotonicity of M_1 in Z*

$$M_{i1}(0) \leq M_{i1}(1) \, for \, all \, i$$

Assumption 7 disqualifies defiers among adults and therefore eliminates strata DA, DC, DD, DN. Assumptions 2, 5, and 7 combined disqualify the existence of 11 of the 16 principal strata (last column, Table 3 in Appendix A). Table 4 in Appendix A shows the correspondence between observed groups and latent strata. Column (1) presents the corresponding strata without Assumptions 2, 5, and 7, Column (2) the remaining strata if these assumptions are imposed.

I concentrate on the effect for stratum CN. In this stratum, M_1 is induced to change from 0 to 1 by the instrument, and M_2 is always zero. This is the only stratum for which outcomes are observed for both children in non-migrant and migrant households. The effect for this stratum can be identified without making assumptions about unobserved outcomes. The causal effect for this stratum is therefore the local average treatment effect (LATE) for children who are never migrants. In the absence of always migrating adults, this effect is also the average treatment effect on the treated (ATET) for children who are never migrants.

$$\theta_{CN} \equiv E\left[\left(Y_i\left(1,0\right) - Y_i\left(0,0\right)\right)|G = CN\right] \tag{8}$$

3.3.1 Latent types of all-move households

Evident from Column (2) in Table 4 in Appendix A, all-move households ($O(0,1,1)$ and $O(1,1,1)$) could belong either to stratum AA or CC under the proposed assumptions. If information about migration of children is available, all strata proportions are identified (Appendix B.2). If this information is not available, external information on the number of unobserved children can be used. I define γ, the ratio of unobserved children in all-move households to the observed number of children in migrant households.[15] However, to point identify strata proportions, we need further assumptions about the existence of strata AA and CC. Information on the institutional setting could help to disqualify the existence of one of these strata. I discuss identification for two extreme scenarios. First, all all-move

[14] The existence of some strata is disqualified by more than one assumption.

[15] See Section 4.2.1 for an explanation of how I calculate γ using information from other data sources.

households belong to stratum AA, and second, all all-move households belong to stratum CC.

In the simpler case, all all-move households belong to stratum AA ($\pi_{CC} = 0$). In this situation, the process that motivates migration of whole households is independent of the instrument. The observed samples with $Z = 0$ and $Z = 1$ both contain only households of type CN, NN, AN (Column (2), Table 4). Sample selection is not a problem, and the conventional Wald estimator yields consistent estimates of θ_{CN} (see calculations in Appendix B.3).

More problematic is if all-move households belong to stratum CC ($\pi_{AA} = 0$). In this scenario, observed group $O(0,0,0)$ contains households of type CC, which are not observed in the sample with $Z = 1$.

In a first step, I calculate the number of households missed based on external information or assumptions about γ. If $\pi_{AA} = 0$, all all-move households are in group $O(1,1,1)$ (Column (3), Table 4). Denote $N^{zm_1m_2}$ the number of observations in each cell. The number of missed observations in group $O(1,1,1)$ is $N_{111} = \gamma * (N_{010} + N_{110})$. With this information, all strata proportions are identified:

$$\begin{aligned}
\pi_{AN} &= \frac{N_{010}}{N_{000} + N_{010}} \\
\pi_{CC} &= \frac{N_{111}}{N_{100} + N_{110} + N_{111}} \\
\pi_{NN} &= \frac{N_{100}}{N_{100} + N_{110} + N_{111}} \\
\pi_{CN} &= 1 - \pi_{AN} - \pi_{NN} - \pi_{CC}.
\end{aligned}$$

To simplify notation, I denote $\overline{Y}^{zm_1m_2} \equiv E\left[Y|Z = z, M_1 = m_1, M_2 = m_2\right]$ for the observed mean outcomes. The potential outcome of CN under treatment, $Y(1,0)|G = CN$, is observed as part of the mixture distribution in the observed group $O(1,1,0)$.

$$\overline{Y}^{110} = \frac{E\left[Y(1,0)|G = CN\right]\pi_{CN} + E\left[Y(1,0)|G = AN\right]\pi_{AN}}{\pi_{CN} + \pi_{AN}} \tag{9}$$

can be reformulated to

$$E\left[Y(1,0)|G = CN\right] = \frac{\overline{Y}^{110}(\pi_{CN} + \pi_{AN}) - E\left[Y(1,0)|G = AN\right]\pi_{AN}}{\pi_{CN}}. \tag{10}$$

Stratum AN corresponds directly to the observed group $O(0,1,0)$, and the mean potential outcome under treatment for this stratum is identified as

$$E\left[Y(1,0)|G = AN\right] = \overline{Y}^{010}. \tag{11}$$

Using Equations 10 and 11, the mean potential outcome under treatment for stratum

CN is identified as

$$E[Y(1,0)|G=CN] = \frac{\overline{Y}^{110}(\pi_{CN}+\pi_{AN}) - \overline{Y}^{010}\pi_{AN}}{\pi_{CN}}. \tag{12}$$

The mean potential outcome under control for stratum CN is part of the mixture distribution in group $O(0,0,0)$, which consists of strata CN, NN, and CC:

$$\overline{Y}^{000} = \frac{E[Y(0,0)|G=CN]\,\pi_{CN} + E[Y(0,0)|G=NN]\,\pi_{NN} + E[Y(0,0)|G=CC]\,\pi_{CC}}{\pi_{CN}+\pi_{NN}+\pi_{CC}}. \tag{13}$$

$E[Y(0,0)|G=NN]$ is identified since group $O(1,0,0)$ corresponds directly to stratum NN. However, the two conditional means of strata CN and CC are not. I derive bounds on $E[Y(0,0)|G=CN]$ following the procedure Chen and Flores (2012) suggest. For simplification, I introduce additional notation. Let y_a^{000} be the a-th quantile of Y in the observed group $\{Z=0, M_1=0, M_2=0\}$, and let the mean outcome in this cell for those outcomes between the a'-th and a-th quantiles of Y be

$$\overline{Y}(y_{a'}^{000} \leq Y \leq y_a^{000}) \equiv E\left[Y|Z=0, M_1=0, M_2=0, y_{a'}^{000} \leq Y \leq y_a^{000}\right] \tag{14}$$

The idea behind these bounds is to find the lowest and highest possible values for $E[Y(0,0)|G=CN]$, subject to the constraint $\overline{Y}^{100} = E[Y(0,0)|G=NN]$. In the unconstrained case, the upper and lower bound of $E[Y(0,0)|G=CN]$ can be derived similarly as in the scenario with randomly assigned M_1. We can bound $E[Y(0,0)|G=CN]$ from below by the expected value of Y for the $\pi_{CN}/(\pi_{CN}+\pi_{NN}+\pi_{CC})$ fraction of smallest values of Y and from above by the expected value of Y for the $\pi_{CN}/(\pi_{CN}+\pi_{NN}+\pi_{CC})$ fraction of largest values of Y in group $O(0,0,0)$.

I assess whether this unconstrained solution satisfies the constraint $\overline{Y}^{100} = E[Y(0,0)|G=NN]$. Under the assumption that the smallest values in group $O(0,0,0)$ are only from CN observations, the lower bound for $E[Y(0,0)|G=NN]$ is given by $\overline{Y}(y_{\alpha_{CN}}^{000} \leq Y \leq y_{1-\alpha_{CC}}^{000})$, the mean estimated in the central area in Figure 1. In case this estimated lower bound is lower than $\overline{Y}^{100}$, the unconstrained solution is identical to the solution of the constrained problem (upper line in Equation 15).

[Figure 1 about here]

If the constraint is unsatisfied, we can derive a lower bound from the mixture distribution of CN and NN in the lower $1-\pi_{CC}/(\pi_{CN}+\pi_{NN}+\pi_{CC})$ quantiles of the distribution of Y in cell $\{Z=0, M_1=0, M_2=0\}$ by assuming all CC observations are at the top of the distribution, and the remaining lower part is a mixture of CN and NN (lower line in

Equation 15). The upper bound can be derived similarly (Equation 16).[16]

$$E^L_{CN}\left[Y(0,0)|G=CN\right]=\begin{cases}\overline{Y}(Y\leq y^{000}_{\alpha_{CN}}),\ if\,\overline{Y}(y^{000}_{\alpha_{CN}}\leq Y\leq y^{000}_{1-\alpha_{CC}})\leq\overline{Y}^{100}\\ \overline{Y}(Y\leq y^{000}_{1-\alpha_{CC}})*\frac{\pi_{NN}+\pi_{CN}}{\pi_{CN}}-\overline{Y}^{100}*\frac{\pi_{NN}}{\pi_{CN}},\ otherwise\end{cases}\tag{15}$$

$$E^U_{CN}\left[Y(0,0)|G=CN\right]=\begin{cases}\overline{Y}(Y\geq y^{000}_{1-\alpha_{CN}}),\ if\,\overline{Y}(y^{000}_{\alpha_{CC}}\leq Y\leq y^{000}_{1-\alpha_{CN}})\geq\overline{Y}^{100}\\ \overline{Y}(Y\geq y^{000}_{\alpha_{CC}})*\frac{\pi_{NN}+\pi_{CN}}{\pi_{CN}}-\overline{Y}^{100}*\frac{\pi_{NN}}{\pi_{CN}},\ otherwise\end{cases}\tag{16}$$

Bounds on the causal effect θ_{CN} can be constructed by combining the point identified potential outcomes under treatment with the bounds on potential outcomes under control:

$$\theta^U_{CN} = E\left[Y(1,0)|G=CN\right]-E^L_{CN}\left[Y(0,0)|G=CN\right]\tag{17}$$

$$\theta^L_{CN} = E\left[Y(1,0)|G=CN\right]-E^U_{CN}\left[Y(0,0)|G=CN\right]\tag{18}$$

A scenario during which the group of all-move households are a mixture of strata AA and CC would, for a given γ, lead to smaller bounds in comparison to a situation in which they are all of type CC. The smaller π_{CC}, the narrower the bounds on $E\left[Y(0,0)|G=CN\right]$. If π_{CC} becomes zero, we are back to the point-identified case.

3.3.2 Distributional assumptions to tighten the bounds

In addition to the behavioral assumptions presented above, distributional assumptions can further tighten the bounds (Chen and Flores, 2012). These distributional assumptions are specific to the setting under study and might vary by outcome variable. Such assumptions can be derived from theoretical arguments about migrant selectivity. For example, a standard assumption is that migration is costly and that costs increase with the number of migrating individuals. Assume that being separated generates disutility for households (Agesa and Kim, 2001). In such a scenario, households that can afford taking their children with them will be selected positively. Better-off households are also likely to invest more in the education of children (Leibowitz, 1974; Blau, 1999; Case, Lubotsky, and Paxson, 2002; Currie, 2009; Almond and Currie, 2011). Therefore, children in CC households will, on average, have more favorable outcomes (e.g. greater school attendance) than children in CN households, as formalized in Assumption 8.

[16] Alternative formulations for Equation 15 and 16 are given by

$$E^L_{CN}\left[Y(0,0)|G=CN\right]=max\left\{\overline{Y}(Y\leq y^{000}_{\alpha_{CN}}),\overline{Y}(Y\leq y^{000}_{1-\alpha_{CC}})*\frac{\pi_{NN}+\pi_{CN}}{\pi_{CN}}-\overline{Y}^{100}*\frac{\pi_{NN}}{\pi_{CN}}\right\}$$

$$E^U_{CN}\left[Y(0,0)|G=CN\right]=min\left\{\overline{Y}(Y\geq y^{000}_{1-\alpha_{CN}}),\overline{Y}(Y\geq y^{000}_{\alpha_{CC}})*\frac{\pi_{NN}+\pi_{CN}}{\pi_{CN}}-\overline{Y}^{100}*\frac{\pi_{NN}}{\pi_{CN}}\right\}.$$

Assumption 8. *Mean dominance 1*

$$E[Y(0,0)|G = CC] \geq E[Y(0,0)|G = CN]$$

Assumption 8 tightens the bound on $E[Y(0,0)|G = CN]$. We can write

$$\overline{Y}^{000} = \frac{E[Y(0,0)|G = CN, CC](\pi_{CN} + \pi_{CC})}{\pi_{CC} + \pi_{CN} + \pi_{NN}} + \frac{E[Y(0,0)|G = NN]\pi_{NN}}{\pi_{CC} + \pi_{CN} + \pi_{NN}} \tag{19}$$

where $E[Y(0,0)|G = NN]$ is identified. Assumption 8 implies that $E[Y(0,0)|G = CN] \leq E[Y(0,0)|G = CN, CC]$. $E[Y(0,0)|G = CN, CC]$ therefore provides an *upper* bound on $E[Y(0,0)|G = CN]$ that is lower or equal as the one in Equation (16):

$$E_{CN}^{U,MD}[Y(0,0)|G = CN] = \frac{\overline{Y}^{000}(\pi_{NN} + \pi_{CN} + \pi_{CC}) - \overline{Y}^{100}\pi_{NN}}{\pi_{CN} + \pi_{CC}} \tag{20}$$

Consider as another example the question of how migration affects the number of children living in origin households. Assume the mean number of children in all-move households (CC) under control is lower or equal to households in which children stay behind (CN). This assumption appears plausible since each additional child living in a household increases the chance that at least one child stays behind. The mean number of children in CN households is higher than in CC households (Assumption 9).

Assumption 9. *Mean dominance 2*

$$E[Y(0,0)|G = CC] \leq E[Y(0,0)|G = CN]$$

We can derive exactly the same quantity as in Equation (20), which in this scenario provides a *lower* bound on $E[Y(0,0)|G = CN]$:

$$E_{CN}^{L,MD}[Y(0,0)|G = CN] = \frac{\overline{Y}^{000}(\pi_{NN} + \pi_{CN} + \pi_{CC}) - \overline{Y}^{100}\pi_{NN}}{\pi_{CN} + \pi_{CC}} \tag{21}$$

Bounds on causal effect θ_{CN} can be constructed by combining point identified potential outcomes under treatment (Equation 12), with the alternative bounds on the potential outcomes of stratum CN under control:

$$\theta_{CN}^{U,MD} = E[Y(1,0)|G = CN] - E_{CN}^{L,MD}[Y(0,0)|G = CN] \tag{22}$$
$$\theta_{CN}^{L,MD} = E[Y(1,0)|G = CN] - E_{CN}^{U,MD}[Y(0,0)|G = CN] \tag{23}$$

3.3.3 Instrumental variables bias without systematic intra-household selection

The bias in the setting with randomly assigned migration status of the principal migrant came solely from differences in potential outcomes under control between children in complier and never migrant households (see Section 3.2). Uncorrected instrumental variables estimates, however, can be biased even if the potential outcomes under control are identical for household types CN and CC. Assumption 10 states the the mean potential outcomes under control are equal for households of type CN and CC.[17]

Assumption 10. *No systematic selection of accompanying migrant*

$$E\left[Y(0,0)|G=CC\right]=E\left[Y(0,0)|G=CN\right]$$

The effect on stratum CN can be point identified as Assumption 10 implies that $E_{CN}^{U/L,MD}\left[Y(0,0)|G=CN\right]$ (derived in Section 3.3.2) corresponds now to the identified mean potential outcome under control. Combining the identified outcome under treatment with the identified outcome under control identifies the causal effect for stratum CN:

$$\theta_{CN}^{NS}=\frac{\overline{Y}^{110}(\pi_{CN}+\pi_{AN})-\overline{Y}^{010}\pi_{AN}}{\pi_{CN}}-\frac{\overline{Y}^{000}(\pi_{NN}+\pi_{CN}+\pi_{CC})-\overline{Y}^{100}\pi_{NN}}{\pi_{CN}+\pi_{CC}} \tag{24}$$

For comparison, consider a Wald estimator in the sample of observed households:

$$\theta_W=\frac{E[Y|Z=1,M_2=0]-E[Y|Z=0,M_2=0]}{E[M_1|Z=1,M_2=0]-E[M_1|Z=0,M_2=0]} \tag{25}$$

The four quantities in Equation (25) can be formulated as means of observed outcomes weighted by strata proportions (for calculations and a more detailed discussion refer to B.4)

$$\theta_W=\frac{\left[\frac{\pi_{NN}*\overline{Y}^{100}+(\pi_{CN}+\pi_{AN})*\overline{Y}^{110}}{\pi_{NN}+\pi_{CN}+\pi_{AN}}\right]-\left[(\pi_{CC}+\pi_{CN}+\pi_{NN})*\overline{Y}^{000}+\pi_{AN}*\overline{Y}^{010}\right]}{\left[\frac{\pi_{CN}+\pi_{AN}}{\pi_{NN}+\pi_{CN}+\pi_{AN}}\right]-\left[\pi_{AN}\right]}.$$

Subtracting θ_{CN}^{NS} from θ_W gives the bias of the Wald estimator:

$$b_W=\theta_W-\theta_{CN}^{NS}=\frac{\pi_{CC}\left[\left(\overline{Y}^{100}-\overline{Y}^{000}\right)a+\left(\overline{Y}^{110}-\overline{Y}^{010}\right)b\right]}{c}, \tag{26}$$

where $a=(\pi_{CN}\pi_{NN})\left(\pi_{CN}+\pi_{CC}+\pi_{NN}\right)$, $b=\pi_{AN}\left(\pi_{CN}^2+\pi_{CN}\pi_{CC}+\pi_{CN}\pi_{AN}+\pi_{CC}\pi_{AN}\right)$, and $c=\left(\pi_{CN}^2+\pi_{CN}\pi_{CC}\right)\left(\pi_{CN}\pi_{CC}+\pi_{CC}\pi_{AN}+\pi_{CN}^2+\pi_{CN}\pi_{AN}+\pi_{CN}\pi_{NN}\right)$. The bias is zero if there are no all-move households ($\pi_{CC}=0$) or if $\overline{Y}^{100}=\overline{Y}^{000}$ and $\overline{Y}^{110}=\overline{Y}^{010}$. The latter

[17] Mean potential outcomes under treatment are irrelevant in this setting.

conditions imply that mean potential outcomes under control are equal for latent groups CN, CC, NN and under treatment for latent groups AN and CN. θ_{CN}^{NS} instead of θ_W should be used in settings, in which Assumption 10 is credible.

3.4 Estimation and inference

Bounds derived in section 3.3.1 include minimum or maximum operators. These operators create several problems for estimation and inference. Hirano and Porter (2012) show that for non-differentiable parameters such as min and max operators, no asymptotically unbiased estimators exist. Therefore, estimators for bounds that use min and max functions can be severely biased in finite samples, and confidence intervals cannot be estimated using standard asymptotics or bootstrap methods. Chernozhukov, Lee, and Rosen (2012) derive a method to obtain half-median unbiased estimators for the lower and upper bound, and confidence intervals for the true parameter. The idea is to apply the min (max) function not directly to the bounding function, but to a precision-corrected version of it. Precision is adjusted by adding to each estimated bounding function its point-wise standard error times an appropriate critical value. Estimates with higher standard errors therefore require larger adjustments. The estimated bounds are conservative, and the half-median unbiased estimator of the upper bound exceeds the true value of the upper bound with the probability of at least 0.5 asymptotically. The estimator of the lower bound falls below the true bound with probability 0.5. Appendix B.5 provides a detailed description of the implementation of the procedure based on Huber, Laffers, and Mellace (2014) and Chen and Flores (2012).[18]

Confidence intervals for bounds that do not involve min and max operators are based on the results from Imbens and Manski (2004), which include the treatment effect of interest with probability 95%:

$$\left(\hat{\theta}^L - 1.654\hat{\sigma}^L, \hat{\theta}^U + 1.654\hat{\sigma}^U\right)$$

$\hat{\theta}^L$ and $\hat{\theta}^U$ denote the estimated bounds and $\hat{\sigma}^L$ and $\hat{\sigma}^U$ the respective estimated standard errors which are obtained by bootstrap with 999 replications.

4 Empirical applications

This section presents two empirical applications of the bounds. The first applies bounds to data from a visa lottery in Tonga used by Gibson, McKenzie, and Stillman (2011a) (henceforth GMS) to study the effects of migration on remaining household members. I apply the bounds to a set of outcomes at the household level – household composition and

[18] Due to precision adjustment, bounds and confidence intervals can be outside the support for outcomes with limited support if the unadjusted estimate is close to the limits of the support. If estimates or confidence intervals of the upper/lower bound of $E[Y(0,0)|CN]$ are larger/smaller than the support of Y, I replaced the estimate with the upper/lower limit of the support of Y.

household assets. This application allows comparing the bounds to well-identified point estimates. The unit of analysis (I_2) is the household as a whole, while the principal migrant (I_1) is the individual, who applied for the visa lottery. The second application is based on a paper from McKenzie and Rapoport (2011) (henceforth MR) that studies the effect of migration on school attendance in Mexico and does not address the issue of invisible sample selection. I test sensitivity of results to various assumptions regarding all-move households. The unit of analysis (I_2) is the individual child and the principal migrant (I_1) is any adult household member.

4.1 Effect of migration on remaining households in Tonga

GMS study the effects of migration from Tonga to New Zealand on household members left behind in Tonga. New Zealand allows a quota of 250 Tongans to immigrate to New Zealand each year without going through the usual migration categories. Among eligible registrants (Tongan citizens aged 18 to 45 years who meet English, health, and character requirements), a random ballot decides who receives a visa to migrate. These registrants are the principal migrants in my framework. Ballot winners must provide a job offer in New Zealand within six months after the lottery to have their application to migrate approved. Ballot winners can apply for visas for their immediate family (spouses and dependent children up to age 24).[19]

GMS use data from a household survey in Tonga that was designed to capture the effects of migration. The survey does not contain households of ballot winners in which all household members join the principal migrant. GMS use the random ballot to instrument for migration of the principal migrant. Instrumental variable estimation is necessary since 15% of ballot winners (among observed households) do not comply with the lottery and do not move to New Zealand (GMS refer to concerns regarding this non-compliance as *dropout bias*). GMS argue that substitution bias is of little concern in this context since the chances of eligibility to migrate under another migration channel are low.

I use household-level data from GMS that include only households that participated in the visa lottery. This sample consisted of 124 households that were unsuccessful in the lottery and have no migrants ($O(0,0,0)$), 26 households that were successful in the lottery but where nobody migrated ($O(1,0,0)$), and 61 households that were successful in the lottery and where the principal migrant (and potentially other household members) migrated to New Zealand, but at least one person stayed behind ($O(1,1,0)$). These observed patterns have two important implications. First, the data do not contain households in which all individuals migrated. Second, the data contain no households with migrants who were unsuccessful in the lottery $O(0,1,0)$. The second observation is evidence that no

[19] For a more extensive description of this visa lottery, refer to GMS.

households of type AN exist, and it is a strong indication that no other households with always migrants or defiers exist, as GMS assume.

4.1.1 All move-households

GMS use visa regulations and define all-move households as households in which all individuals would be eligible to join the principal migrant in case he wins the lottery and migrates. Since these visa rules are based on observable characteristics (i.e., age, relationship with the principal migrant), GMS identified these households and removed them from their data. I refer to these households as *visa all-move households.* GMS therefore identified the effect for households that leave more distant relatives behind. This concerns 75 of 124 households in the observed group $O(0,0,0)$, and 18 of 26 households in the observed group $O(1,0,0)$ (see Table 5, Panel A). Finding visa all-move households in group $O(1,0,0)$ shows that the definition of all-move households based on observable characteristics is not identical to latent stratum CC. All households in group $O(1,0,0)$ belong to the latent stratum NN - households in which no individual would migrate (see Table 4). However, we can use the information on visa all-move households to estimate π_{CC}.

Using only information from observed group $O(0,0,0)$ leads to the conclusion that $\hat{\pi}_{CC}$ equals $N_{000}^{VAM}/N_{000} = 75/124 = 0.61$. However, in observed group $O(1,0,0)$, which consists only of stratum NN, 18 of 26 identify as visa all-move households ($N_{100}^{VAM}/N_{100} = 18/26 = 0.7$). The overall share of visa all-move households in $O(0,0,0)$ is therefore a combination of the 70% visa all-move households in stratum NN and 100% visa all-move households in stratum CC (Equation 27). The ratio of $\pi_{CN}/\pi_{NN} = N_{110}/N_{100} = 2.34$.[20] Since no households of type AN exist, it holds that $\pi_{NN}+\pi_{CN}+\pi_{CC} = 1$ (Equation 29). Combining this information gives a system of three equations, which we can solve to obtain the strata proportions:

$$N_{100}^{VAM}/N_{100} * \hat{\pi}_{NN} + \hat{\pi}_{CC} = N_{000}^{VAM}/N_{000} \tag{27}$$

$$N_{110}/N_{100} * 1.368 = \hat{\pi}_{CN}/\hat{\pi}_{NN} \tag{28}$$

$$\hat{\pi}_{NN} + \hat{\pi}_{CN} + \hat{\pi}_{CC} = 1 \tag{29}$$

Panel b of Table 5 shows the estimated strata proportions. No households belong to stratum AN, 35% to stratum CN, 53% to stratum CC, and 11% to stratum NN. The ratio of unobserved to observed migrant households $\gamma = 1.5$.

[Table 4 about here]

[20] This ratio must be adjusted by a factor of 1.368 due to differential sample weights (Equation 28). Group $O(0,0,0)$ has an expansion factor of 37.9, group $O(1,0,0)$ of 2.5, and group $O(1,1,0)$ of 3.4. For the estimation of the bounds this is relevant only when calculating the ratio π_{CN}/π_{NN}.

4.1.2 Assessment of assumptions

Assumption 2 holds since we do not see any migrants in households in which the principal migrant does not migrate. Assumption 3 holds through the random ballot that decides who receives a visa. The two exclusion restrictions, Assumption 4, and 5 are also likely to hold. Obtaining a visa has no direct effect on household welfare, except through its effect on migration. Other household members can only obtain a visa if the principal migrant takes up his visa and migrates. Therefore, the visa affects migration of other household members only through its effect on migration of the principal migrant. Assumption 7 is very likely to hold since it appears unreasonable that a visa makes a person less likely to migrate. Another question is whether all-move households are of type AA or CC. GMS argue that it is difficult to obtain another type of visa, which strongly suggests that all-move households must be of type CC. We do not observe any households with migrants, who are not lottery winners, which is further evidence that all-move households are of type CC. I discuss distributional assumptions in the results section for individual outcome variables.

4.1.3 Results

Table 6 shows the bounds on the effect of migration on household composition. Household composition is a particularly appropriate outcome to study the problem of invisible sample selection since it relates strongly to the propensity of households to migrate as a whole.[21] An approach that ignores selection due to all-move households would conclude that migration reduces total household size by 0.85 persons, not statistically different from zero. Removing all visa all-move households from the estimation sample, GMS conclude that the effect is -2.26.

I find that the point identified household size of CN households under treatment is 4.69. Without a mean dominance assumption, the size of CN households under control can be bounded between 2.17 and 8.56, which leads to bounds on the effect between -3.87 and 2.54. Assuming CC households are on average smaller than CN households (Assumption 9), increases the lower bound on $E[Y(0,0)|G = CN]$ to 5.38 and decreases the upper bound on the effect to -0.69, which shows that this assumption has substantial identifying power.

[Table 6 about here]

The bounds analysis confirms the negative effect of migration on household size, driven by a reduction in the number of children and prime-age individuals in the households. The number of elderly is not reduced.

[21] The results in this section correspond to Tables 4 and 6 in GMS.

A second set of results examines the effects of migration on household assets (Table 7). 48% of CN households under treatment own a home. Without distributional assumptions, home ownership under control can only be bounded between 0 and 1. These bounds correspond to the bounds Manski (1994) suggests that use only the fact that the outcome variable is bounded. The resulting bounds on the treatment effect are -0.52 and 0.48. The bounds can be narrowed substantially by assuming CC households are less likely to own a home since home ownership increases the probability that someone stays behind.[22] The upper bound of $E[Y(0,0)|G = CN]$ lowers to 0.4, and the effect can be bounded between -0.52 and 0.08. The unadjusted estimate (0.12) lies outside of these bounds.

I invoke the same mean dominance assumption when identifying the effect on agricultural assets and livestock owned; households that leave someone behind (CN) own, on average, more livestock than households that leave nobody behind (CC). The bounds under this assumption suggest negative effects on the probability to own any agricultural assets, as well as the number of pigs, chickens, and cattle owned, though the confidence intervals do not disqualify zero effects.

[Table 7 about here]

The difference between the point identified effects under assumption 10, θ_{CN}^{NS}, and the unadjusted effect θ_W is the bias described in Section 3.3.3. For example, for total household size, the bias is -0.16, which corresponds to 23% of the corrected estimate. The effect on the number of pigs owned is biased by -0.11, more than 100% of the corrected estimate.

Analysis of the Tongan data shows that bounds with a monotonicity and mean dominance assumption have significant identifying power, even in situations with a high share of all-move households. However, they also reveal that instrumental variables estimates ignoring invisible sample selection can be biased substantially.

4.2 Effect of migration on school attendance in Mexico

The second empirical application follows MR, estimating the effect of migration on school attendance in Mexico. MR use historic migration rates as an instrument for current migration, finding that migration of an adult household member reduce school attendance rates for 12 to 15 year old boys by 16 percentage points, and by 9 percentage points for girls; however, the latter effect is not significantly different from zero.[23]

[22] Home ownership is generally seen as having an impeding effect on migration. See, for example, Massey and Espinosa (1997) and Nivalainen (2004).

[23] Unlike MR, I focus only on the effect of migration on school attendance in the sample of children aged 12 to 15 years. Restricting analysis in this way offers two advantages. First, children in this age group are unlikely to migrate without their parents, which is required by Assumption 2. This assumption does not hold for 16 to 18 year old adolescents, the second group that MR consider. Second, in comparison to years of education, school attendance is the more natural outcome for children and adolescents who have not yet completed their education.

Data stem from the Mexican 1997 Encuesta Nacional de la Dinàmica Demográfica (ENADID).[24] I follow MR and define a child as *living in a migrant household* if the household has a member aged 19 and over who has ever been to the United States to work, or who has moved to the United States in the last 5 years for any other reason.[25]

The outcome of interest is school attendance. Although school attendance in Mexico is compulsory up to the age of 16 years,[26] attendance rates at the time of the survey were significantly below 100% (74% for boys and 66% for girls in the estimation sample).

4.2.1 Sample selection due to child migration

No children in the sample are categorized as current migrants.[27] Potential sample selection therefore arises from migration of whole households. Although the ENADID dataset provides rich information on individual migration histories, it lacks information on households that migrate as a whole. To gain an understanding of how widespread the phenomenon of all-move households is in Mexico, I build on existing research that uses census data from the source and destination countries of migrants. Ibarraran and Lubotsky (2007) estimate the size of the Mexican immigrant population in the United States based on a) the 2000 Mexican census and b) the 2000 U.S. census. Since the Mexican census was conducted as a household survey, it ignored all-move households. The estimated size of the Mexican-born population living in the United States based on the Mexican census is 1,221,598,[28] and based on the U.S. census is 2,205,356. Thus, the total migrant population in the Mexican census is only 55.4% the size of the population in the U.S. census. This rate is lower for female (33.6%) than for male migrants (69.9%).

The authors argue the difference is primarily due to married couples that dissolved their household in Mexico and are therefore not counted in the Mexican census. Once married couples with both spouses present in the United States are removed from U.S. census estimates, the remaining migrant number is 1,492,111, closer to the number from the Mexican census. In a similar analysis, McKenzie and Rapoport (2007) use the U.S. census 5% public use sample to analyze the marital status of recent Mexican immigrants.

[24] The ENADID is a nationally representative household survey, with a sample of 73,412 households. This corresponds to roughly 2,300 households in each of the 32 states. To allow comparability of results with MR, I restrict the sample similarly to households in municipalities outside of cities with more than 50,000 inhabitants. The estimation sample consists of 15,665 children aged 12 to 15 years in 11,160 households.

[25] This is not the optimum migrant definition to study sample selection since it also includes return migrants. To maximize comparability of results with existing research, I follow the definition from MR, who argue that prior migration episodes of adult household members also influence the education outcomes of children. For a more extensive discussion on the advantages and disadvantages of this migrant definition, refer to MR.

[26] http://www.sep.gob.mx/en/sep_en/Basic_Education_a

[27] Fourteen children reported prior migration episodes, and of these, six come from non-migrant households. However, the questionnaire included several questions on migration, and the answers to these observations are inconsistent. Therefore, data problems seem to be the reason for this finding.

[28] This number excludes migrants who returned to Mexico.

They find that 14.4% of male and 48% of female recent Mexican migrants are married, with their spouses present in the United States, concluding these individuals are likely not counted in Mexico-based surveys.

Discrepancies between numbers from the Mexican and U.S. censuses are even larger for children. In the age group 12 to 15 years, the number of migrants in the Mexican census is only about 50% of the number of migrants in the U.S. census. Again, the reason is most likely that children migrate with their whole families and are therefore not counted in the Mexican census anymore. Overall, the U.S. census counts 82,240 Mexican-born children in this age group, which are most likely not included in Mexican data.[29]

I use the ENADID to calculate the total number of children in migrant households in Mexico. Using the definition of a migrant household described above and the expansion factors provided with the data, I calculate the total number of children aged 12 to 15 years who live in a migrant household to be 1,516,924. Dividing the number of children missed by the observed number of children in migrant households, I calculate γ to be 0.054.[30] This ratio appears low, but this is because of the broad definition of a migrant household in the ENADID, and thus the large denominator. I test the sensitivity of results to various values of γ ranging from zero to 0.5. For the main analysis, I use a ratio of $\gamma = 0.054$.

4.2.2 Assessment of assumptions

To overcome the problem of self-selection into migration, a number of recent studies (e.g. McKenzie and Hildebrandt, 2005; McKenzie and Rapoport, 2011) use historic migration rates to measure current migration. Existing networks lower migration costs for subsequent migrants, and therefore trigger additional migration. The exclusion restriction is that these historic migration rates do not affect educational outcomes today except through current migrations of household members (Assumption 4). A detailed discussion of this instrument and the exclusion restriction regarding educational attainment can be found in MR. However, the bounds in this paper require an additional assumption about the instrument. Assumption 5 suggests the instrument must not influence the migration decision of a child directly, which appears reasonable if migration networks primarily help adult migrants find a job in the destination country.

Like MR, I use state-level migration rates to the United States from 1924 taken from Woodruff and Zenteno (2007). I recode this continuous measure into a binary one by defining states as low-migration states ($Z = 0$) if the migration rate is below the state-level median (3.78%) and as high-migration states ($Z = 1$) if the migration rate is above (see Frölich (2007) for details on this transformation). I do this to allow stratification on

[29] Thanks go to Darren Lubotsky for providing this estimate.

[30] I calculate this ratio for Mexico as a whole since I can not identify the origin region within Mexico of migrants in the U.S. census. Therefore I have to assume that this ratio is equal between rural and urban regions.

instrument assignment, which would not be possible with a continuous instrument. Figure 2 in Appendix A shows the positive relationship between historic migration rates and the probability of a child living in a migrant household (Assumption 6). In this setting, compliers are individuals who would migrate only if they live in a high-migration state. Unlike MR, I abstain from including additional covariates in my estimation to ensure Assumption 3 holds. Covariates would substantially complicate the analysis since I do not observe the distribution of covariates for all-move households. Two-stage, least-squares point estimates with a binary instrument without covariates differ only slightly from those using covariates, and are similar to results from MR.[31]

No children in the sample are categorized as current migrants. This is strong evidence in support of Assumption 2, that children would not migrate alone. Whether all-move households react to the instrument (CC) or are always migrants (AA) cannot be clarified with these data. Assuming they are all type CC leads to conservative bounds.

4.2.3 Results

The bounds are on the effect of migration on school attendance for the group of children who would never migrate but live in a household in which adults react to the instrument. Ignoring sample selection and estimating effects using a simple Wald estimator suggests the effect for boys is -0.19 and significant, and the effect for girls is 0.08 and not significant (Table 8).[32]

The first rows of Table 8 show the estimated strata proportions. The proportion of stratum CN for boys is 0.26, and the proportion of stratum CC is 0.02. NN is the largest stratum, with a proportion of 0.6. Strata proportions are similar for girls. The next three rows display the point identified mean potential outcomes for stratum NN under control, and for strata AN and CN under treatment. The expected school attendance rate under treatment for stratum CN is 0.63 for boys and 0.6 for girls. School attendance is slightly higher for boys than girls in all strata. The bounds under monotonicity for school attendance rates of boys in stratum CN under control are 0.78 and 0.89. For girls, they are substantially lower, 0.49 and 0.57, respectively.

The lower and upper bounds on θ_{CN} for boys are -0.28 and -0.14. For girls, the respective numbers are -0.01 and 0.15. However, confidence intervals are wide for both groups. For boys, the 95% confidence intervals exclude zero.

[31] For the model with controls, I use the following state-level control variables: number of schools per 1,000 inhabitants in 1930, literacy rate in 1960, and male and female attendance rates in 1930. These are not the same controls used by MR since those controls could not be reconstructed. Including these covariates changes the point estimate in a two-stage, least-squares estimation for boys from -19.5 to -14.7 percentage points and for girls from 8.2 to 7.4 percentage points.

[32] Although the estimates for boys with and without covariates are similar to results from MR, there is a larger discrepancy in the estimates for girls. Unreported estimates using the continuous instrument are much closer to MR's result.

Imposing the additional mean dominance assumption that the mean outcome under control of stratum CC is weakly greater than of stratum CN, narrows the bounds by increasing the lower bound on θ_{CN}. For boys, the lower bound increases to -0.19 and for girls to 0.07.

The effect of living in a migrant household for boys is negative even when sample selection is considered. The opposite is true for girls, and the estimated bounds suggest that the effect might even be positive. This result accords with arguments and empirical findings from a series of recent papers. Antman (2012) suggests paternal migration associates with a shift in decision-making power toward the mother, and that mothers choose to spend more on the education of girls. Antman (2011) finds that in the short run, boys must respond to paternal absence with an increase in work and decrease in study hours. Both channels might contribute to the fact that boys experience a negative effect on school attendance and no, or even a positive, effect exists for girls.

[Table 8 about here]

4.3 Sensitivity with respect to γ

For main results, I compute γ to be 0.054. However, due to the substantial uncertainty in the computation of this number, I repeat analysis for γ between zero (i.e., no children in all-move households) and 0.5 (i.e., for every two children observed in a migrant household, one child in an all-move household is missed).

[Figure 3 about here]

Figure 3 shows the resulting bounds on the effects for boys and girls. The width of the bounds does not increase constantly over the range of the observed ratios. For boys, the lower bound decreases steeply up to a value of γ of about 0.14, and only slightly thereafter. Up to a ratio of 0.14, the estimate of the upper bound on $E[Y(0,0)|G = CN]$ stems from the constrained solution. Once π_{CN} is sufficiently small so the constraint no longer binds, the estimate stems from the unconstrained solution. A slight decrease after this threshold results from a steady decline in π_{CN} and the fact that the expected value is computed in a decreasing fraction of largest values of the outcome Y in group $O(0,0,0)$. For girls, we observe a constrained solution up to a ratio of 0.41.[33]

Two additional insights can be gained from Figure 3. First, γ is not the only determinant of the behaviors of the bounds. For example, at a value of 0.3, the width of the bounds is 0.33 for boys and 0.59 for girls, which is due to the various distributions of the

[33] The kink in the upper bound for boys at a ratio of about 0.3 stems from the precision adjustment. Above this ratio, the unrestricted solution is not considered "close" enough to influence the asymptotic distribution of the upper bound.

outcome variable. Second, the lower bound under monotonicity and mean dominance is insensitive to variations in γ.

5 Discussion

This section discusses extensions of the proposed approach and avenues for future research.

An alternative to the bounds derived from Chen and Flores (2012) would be an approach outlined in Imai (2007). Imai suggests subtracting the full distribution of $Y(0,0)|G = NN$ derived from the observed group $O(1,0,0)$ from the distribution of Y in the observed group $O(0,0,0)$, and employing the trimming procedure in the remaining distribution. Although this approach might tighten the bounds somewhat, estimating and subtracting distributions instead of means creates complications for estimation. Since this approach is not developed fully, I abstain from including those bounds here.

Many applications rely on covariates to ensure instrument validity or conditional independence. Incorporating covariates in the principal stratification framework is an active field of research. Frangakis and Rubin (2002) suggest conducting the analysis within cells defined by observed pre-treatment variables. Lee (2009) shows that this strategy can be used to narrow the bounds. Two issues complicate this approach in the migration setting. First, when it is desirable to condition on multiple variables, cells might become too small due to the curse of dimensionality. However, computing a propensity score and conducting the analysis within strata of the propensity score might circumvent this problem. The second and more complicated issue is that covariates are usually unobserved for all-move households, and therefore it is impossible to condition on the covariates of this group.

This paper investigates a situation in which the principal migrant can be identified and distinguished from other household members, even if nobody in the household migrates. Extensions should examine situations in which the principal migrant(s) cannot be identified, and allow for more complicated household structures. Especially when interest is in the effect of migration of an adult on other adult household members (i.e., the effect of migration on labor supply of other household members), identifying the principal migrant might be impossible.

The proposed setting is applicable to situations in which sample selection occurs due to all-move households, and situations in which sample selection occurs due to migration of only a subset of household members. However, special attention should be given to situations where sample selection is driven by both as the migration decision of the "last" household member might be driven by a different decision process (e.g., someone has to stay behind to take care of property). Another potential refinement would be to consider situations of stepwise migration. Migration processes often take the form that one individual leaves first and the remaining household members follow with some delay.

While this paper discusses intra-household selection mainly from a sample selection perspective, the proposed approach can also be used to identify various other effects and disentangle mechanisms (as a reference to literature on mediation analysis see for example Pearl (2001); Flores and Flores-Lagunes (2009, 2010); Huber (2014)). Researchers and policymakers might be interested in $Y(1,1) - Y(0,0)$, the effect if a child migrates with an adult, in comparison to a situation in which no household member migrates (e.g., Stillman, Gibson, and McKenzie, 2012) or in $Y(1,1) - Y(1,0)$, which is the effect of migration of the whole household in comparison to a situation in which the child remains while the adult migrates (e.g., Gibson, McKenzie, and Stillman, 2011b). One could identify the effects not only for one latent population, but for various populations. Huber, Laffers, and Mellace (2014) derive bounds for average treatment effects on the treated and other populations in a setting with non-compliance. Chen, Flores, and Flores-Lagunes (2014) derive bounds for population average treatment effects. Such approaches could be extended and applied to the migration setting.

6 Conclusion

This paper examines identification of the causal effects of migration on remaining household members in the presence of selection into migration between and within households. If households migrate as a whole, they are usually not included in source country data, which creates additional problems due to *invisible sample selection*. Households that are observed comprise a selected sample and estimates of migration might be biased. Addressing the selection of migrants within the household and the related problem of invisible sample selection has been largely ignored in existing literature.

This paper derives nonparametric bounds on the effect of migration on remaining household members. Using principal stratification allows structuring the identification problem by making transparent assumptions about migration decisions of household members. This approach allows point or partial identification of effects even in complex settings with multiple selection problems present. An important though less obvious insight from the econometric analysis is that invisible sample selection biases instrumental variables estimates even if intra-household selection is unrelated to potential outcomes.

Two empirical applications illustrate the proposed approach. The first uses data from a visa lottery in Tonga to study the effects of migration on household composition and household assets. The Tongan context allows a comparison of the bounds with a) estimates that ignore the second selection problem and b) estimates for a specific subpopulation that take into account both selection problems (Gibson, McKenzie, and Stillman, 2011b). The second example uses data from a study on the effects of migration on educational attainment in Mexico that does not address invisible sample selection (McKenzie and Rapoport, 2011). I calculate the share of children missed in Mexican data by comparing census data from Mexico and the United States. The results suggest that ignoring the second selection problem can bias estimates in both directions, understating the true magnitude of an effect, or suggesting significant effects where the true effect might be zero. The proposed approach can also be used to disentangle direct and indirect effects of migration. I discuss several possible extensions to identify not only the effects discussed in this paper, but also a variety of other related effects and the effects on other latent and observed populations.

More generally, the issue of invisible sample selection is not specific to migration research. Invisible sample selection changes the composition of a population in unobserved ways and can for example be the result of endogenous fertility decisions, household formation, death, or firm entry and exit. Therefore, the insights from this paper can be adapted and applied to a wider literature in applied economics.

Finally, the paper encourages partial instead of point identification in contexts in which point identification can be achieved only under strong and unrealistic ignorability assumptions, which is often the case in migration research. A strength of this approach is that instead of making strong ignorability assumptions, many weaker assumptions can be combined to derive informative bounds. Instead of assuming selection processes are independent of outcome-generating processes, making assumptions about the direction of selection, positive or negative, might be more appropriate. The approach is especially suited to migration studies since theoretical and empirical literature on migrant selectivity provides a foundation from which to derive credible assumptions.

References

Agesa, R. U., and S. Kim (2001): "Rural to Urban Migration as a Household Decision: Evidence from Kenya," *Review of Development Economics*, 5(1), 60–75.

Albert, J. M., and S. Nelson (2011): "Generalized Causal Mediation Analysis," *Biometrics*, 67(3), 1028–1038.

Almond, D. (2005): "Is the 1918 Influenza Pandemic Over? Long-Term Effects of In Utero Influenza Exposure in the Post-1940 U.S. Population," *Manuscript, NBER, Cambridge, MA*.

——— (2006): "Is the 1918 Influenza Pandemic Over? Long-Term Effects of In Utero Influenza Exposure in the Post-1940 US Population," *Journal of Political Economy*, 114(4), 672–712.

Almond, D., and J. Currie (2011): "Human Capital Development Before Age Five," *Handbook of Labor Economics*, 4(11), 1315–1486.

Amuedo-Dorantes, C., A. Georges, and S. Pozo (2010): "Migration, Remittances, and Children's Schooling in Haiti," *The ANNALS of the American Academy of Political and Social Science*, 630(1), 224–244.

Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996): "Identification of Causal Effects Using Instrumental Variables," *Journal of the American Statistical Association*, 91(434), 444–455.

Antman, F. M. (2011): "The Intergenerational Effects of Paternal Migration on Schooling and Work: What Can We Learn from Children's Time Allocations?," *Journal of Development Economics*, 96(2), 200–208.

Antman, F. M. (2012): "Gender, Educational Attainment and the Impact of Parental Migration on Children Left Behind," *Journal of Population Economics*, 25(4), 1187–1214.

——— (2013): "The Impact of Migration on Family Left Behind," *International Handbook on the Economics of Migration*, pp. 293–308.

Blanco, G., C. A. Flores, and A. Flores-Lagunes (2013a): "Bounds on Average and Quantile Treatment Effects of Job Corps Training on Wages," *Journal of Human Resources*, 48(3), 659–701.

Blanco, G., C. a. Flores, and A. Flores-Lagunes (2013b): "The Effects of Job Corps Training on Wages of Adolescents and Young Adults," *American Economic Review*, 103(3), 418–422.

BLAU, D. M. (1999): "The Effect of Income on Child Development," *Review of Economics and Statistics*, 81(2), 261–276.

CASE, A., D. LUBOTSKY, AND C. PAXSON (2002): "Economic Status and Health in Childhood: The Origins of the Gradient," *American Economic Review*, 92(5), 1308–1334.

CHEN, X., AND C. A. FLORES (2012): "Bounds on Treatment Effects in the Presence of Sample Selection and Noncompliance : The Wage Effects of Job Corps," *mimeo*.

CHEN, X., C. A. FLORES, AND A. FLORES-LAGUNES (2014): "Bounds on Population Average Treatment Effects with an Instrumental Variable," *Manuscript*.

CHERNOZHUKOV, V., S. LEE, AND A. M. ROSEN (2012): "Intersection Bounds: Estimation and Inference," *forthcoming in Econometrica*.

CURRIE, J. (2009): "Healthy, Wealthy, and Wise: Socioeconomic Status, Poor Health in Childhood, and Human Capital Development," *Journal of Economic Literature*, 47(1), 87–122.

DOCQUIER, F., AND H. RAPOPORT (2012): "Globalization, Brain Drain, and Development," *Journal of Economic Literature*, 50(3), 681–730.

FLORES, C. A., AND A. FLORES-LAGUNES (2009): "Identification and Estimation of Causal Mechanisms and Net Effects of a Treatment under Unconfoundedness," *IZA Dicussion Paper Series*, 4237.

FLORES, C. A., AND A. FLORES-LAGUNES (2010): "Nonparametric Partial Identification of Causal Net and Mechanism Average Treatment Effects," *California Polytechnic State University at San Luis Obispo*, Manuscript.

FRANGAKIS, C. E., AND D. B. RUBIN (2002): "Principal Stratification in Causal Inference," *Biometrics*, 58(1), 21–29.

FRÖLICH, M. (2007): "Nonparametric IV estimation of local average treatment effects with covariates," *Journal of Econometrics*, 139(1), 35–75.

GIBSON, J., D. MCKENZIE, AND S. STILLMAN (2011a): "The Impacts of International Migration on Remaining Household Members: Omnibus Results from a Migration Lottery Program," *Review of Economics and Statistics*, 93(4), 1297–1318.

——— (2011b): "What Happens to Diet and Child Health When Migration Splits Households? Evidence from a Migration Lottery Program," *Food Policy*, 36(1), 7–15.

——— (2013): "Accounting for Selectivity and Duration-dependent Heterogeneity when Estimating the Impact of Emigration on Incomes and Poverty in Sending Areas," *Economic Development and Cultural Change*, 61(2), 247–280.

Gronau, R. (1974): "Wage Comparisons - A Selectivity Bias," *Journal of Political Economy*, 82(6), 1119–1143.

Hanson, G. H., and C. M. Woodruff (2003): "Emigration and Educational Attainment in Mexico," *Manuscript*.

Heckman, J. (1979): "Sample Selection Bias as a Specification Error," *Econometrica*, 47(1), 153–161.

Heckman, J., R. Pinto, and P. Savelyev (2013): "Understanding the Mechanisms through which an Influential Early Childhood Program Boosted Adult Outcomes.," *American Economic Review*, 103(6), 2052–2086.

Heckman, J. J. (1974): "Shadow Prices, Market Wages, and Labor Supply," *Econometrica*, 42(4), 679–694.

Heckman, J. J., and S. Mosso (2014): "The Economics of Human Development and Social Mobility," *Annual Review of Economics*, 6, 689–733.

Hirano, K., and J. R. Porter (2012): "Impossibility Results for Nondifferentiable Functionals," *Econometrica*, 80(4), 1769–1790.

Huber, M. (2014): "Identifying Causal Mechanisms (Primarily) Based on Inverse Probability Weighting," *Journal of Applied Econometrics*, 29(6), 920–943.

Huber, M., L. Laffers, and G. Mellace (2014): "Sharp IV Bounds on Average Treatment Effects on the Treated and Other Populations under Endogeneity and Noncompliance," *Manuscript*.

Huber, M., M. Lechner, and G. Mellace (2014): "Why Do Tougher Caseworkers Increase Employment? The Role of Programme Assignment as a Causal Mechanism," *SEPS Discussion Paper*, 2014-14.

Ibarraran, P., and D. Lubotsky (2007): "Mexican Immigration and Self-Selection: New Evidence from the 2000 Mexican Census," in *NBER Chapters*, ed. by G. J. Borjas, pp. 159–192. National Bureau of Economic Research, Inc.

Imai, K. (2007): "Identification Analysis for Randomized Experiments with Noncompliance and "Truncation-by-Death"," *Manuscript, Department of Politics, Princeton University*.

Imbens, G. W., and J. D. Angrist (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62(2), 467–475.

Imbens, G. W., and C. F. Manski (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72(6), 1845–1857.

Kitagawa, T. (2009): "Identification Region of the Potential Outcome Distributions under Instrument Independence," *cemmap working paper*, 30/09, 1–37.

Kuhn, R., B. Everett, and R. Silvey (2011): "The Effects of Children's Migration on Elderly Kin's Health: a Counterfactual Approach," *Demography*, 48(1), 183–209.

Lee, D. S. (2009): "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects," *Review of Economic Studies*, 76(3), 1071–1102.

Leibowitz, A. (1974): "Home Investments in Children," *Journal of Political Economy*, 82(2), 111–131.

Manski, C. F. (1989): "Anatomy of the Selection Problem," *Journal of Human Resources*, 24(3), 343–360.

——— (1994): "The Selection Problem," in *Advances in Econometrics: Sixth World Congress*, ed. by C. A. Sims, pp. 143–170. Cambridge University Press, Cambridge, UK.

Massey, D., and K. Espinosa (1997): "What's Driving Mexico-U.S. Migration? A Theoretical, Empirical, and Policy Analysis," *American Journal of Sociology*, 102(4), 939–999.

Mattei, A., and F. Mealli (2007): "Application of the Principal Stratification Approach to the Faenza Randomized Experiment on Breast Self-examination.," *Biometrics*, 63(2), 437–446.

McKenzie, D. (2012): "Learning about Migration through Experiments," *CReAM Discussion Paper Series 1207*.

McKenzie, D., J. Gibson, and S. Stillman (2010): "How Important Is Selection? Experimental vs. Non-Experimental Measures of the Income Gains from Migration," *Journal of the European Economic Association*, 8(4), 913–945.

McKenzie, D., and N. Hildebrandt (2005): "The Effects of Migration on Child Health in Mexico," *Economia*, 6(1), 257–289.

McKenzie, D., and H. Rapoport (2007): "Self-selection Patterns in Mexico-U.S. Migration: The Role of Migration Networks," *Review of Economics and Statistics*, 92(4), 811–821.

McKenzie, D., and H. Rapoport (2011): "Can Migration Reduce Educational Attainment? Evidence from Mexico," *Journal of Population Economics*, 24(4), 1331–1358.

McKenzie, D., and D. Yang (2010): "Experimental Approaches in Migration Studies," *CReAM Discussion Paper Series 1017*.

Nivalainen, S. (2004): "Determinants of Family Migration: Short Moves vs. Long Moves," *Journal of Population Economics*, 17(1), 157–175.

Pearl, J. (2001): "Direct and Indirect Effects," in *Proceedings of the Seventeenth Conference on Uncertainy in Artifcial Intelligence*, pp. 411–420. Morgan Kaufman, San Francisco.

Rubin, D. B. (1974): "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies.," *Journal of Educational Psychology*, 66(5), 688–701.

——— (1980): "Randomization Analysis of Experimental Data: The Fisher Randomization Test Comment," *Journal of the American Statistical Association*, 75(371), 591–593.

Stillman, S., J. Gibson, and D. McKenzie (2012): "The Impact of Immigration on Child Health: Experimental Evidence from a Migration Lottery Program," *Economic Inquiry*, 50(1), 62–81.

United Nations (2013): *International Migration Report 2013*. Department of Economic and Social Affairs, Population Division.

Woodruff, C., and R. Zenteno (2007): "Migration Networks and Microenterprises in Mexico," *Journal of Development Economics*, 82(2), 509–528.

World Bank (2008): *World Development Report 2008. Agriculture for Development*. World Bank Publishers, Washington DC.

Yang, D. (2008): "International Migration, Remittances and Household Investment: Evidence from Philippine Migrants' Exchange Rate Shocks," *Economic Journal*, 118(528), 591–630.

Zhang, J. L., and D. B. Rubin (2003): "Estimation of Causal Effects via Principal Stratification when Some Outcomes are Truncated by "Death"," *Journal of Educational and Behavioral Statistics*, 28(4), 353–368.

Zhang, J. L., D. B. Rubin, and F. Mealli (2008): "Evaluating the Effects of Job Training Programs on Wages through Principal Stratification," in *Modelling and Evaluating Treatment Effects in Econometrics (Advances in Econometrics, Volume 21)*, ed. by T. Fomby, R. C. Hill, D. L. Millimet, J. A. Smith, and E. J. Vytlacil, pp. 117–145. Emerald Group Publishing.

A Tables and figures

Type I_2	$M_2(1)$	$M_2(0)$	Description
A (lways migrant)	1	1	I_2 always migrates, irrespective of M_1
C (omplier)	1	0	I_2 migrates if I_1 migrates but not otherwise
D (efier)	0	1	I_2 migrates if I_1 stays but not otherwise
N (ever migrant)	0	0	I_2 never migrates, irrespective of M_1

Table 1: Principal strata with randomly assigned migration status of I_1

Observed subgroups $O(m_1, m_2)$	Outcome Y	Latent strata	
		(1)	(2)
$O(0,0) = \{M_1 = 0, M_2 = 0\}$	observed	C, N	C, N
$O(0,1) = \{M_1 = 0, M_2 = 1\}$		A, D	-
$O(1,0) = \{M_1 = 1, M_2 = 0\}$	observed	D, N	N
$O(1,1) = \{M_1 = 1, M_2 = 1\}$		A, C	C

Note: Column (1) shows all latent strata. Column (2) shows the remaining strata after Assumption 2 has been imposed.

Table 2: Correspondence between observed groups and latent strata

Type I_1	$M_1(1)$	$M_1(0)$	Type I_2	$M_2(1)$	$M_2(0)$	Household type	Exclusion criterion
A	1	1	A	1	1	AA	
	1	1	C	1	0	AC	A5
	1	1	D	0	1	AD	A5
	1	1	N	0	0	AN	
C	1	0	A	1	1	CA	A2
	1	0	C	1	0	CC	
	1	0	D	0	1	CD	A2
	1	0	N	0	0	CN	
D	0	1	A	1	1	DA	A2, A7
	0	1	C	0	1	DC	A7
	0	1	D	1	0	DD	A7
	0	1	N	0	0	DN	A7
N	0	0	A	1	1	NA	A2
	0	0	C	1	0	NC	A2, A5
	0	0	D	0	1	ND	A2, A5
	0	0	N	0	0	NN	

Table 3: Principal strata with imperfect compliance of the principal migrant

Observed subgroups $O(z, m_1, m_2)$	Outcome Y	Latent strata		
		(1)	(2)	(3)
$O(0,0,0) = \{Z = 0, M_1 = 0, M_2 = 0\}$	observed	CC, CN, NC, NN	CN, NN	CC, CN, NN
$O(0,0,1) = \{Z = 0, M_1 = 0, M_2 = 1\}$		CA, CD, NA, ND	-	-
$O(0,1,0) = \{Z = 0, M_1 = 1, M_2 = 0\}$	observed	AC, AN, DC, DN	AN	AN
$O(0,1,1) = \{Z = 0, M_1 = 1, M_2 = 1\}$		AA, AD, DA, DD	AA	-
$O(1,0,0) = \{Z = 1, M_1 = 0, M_2 = 0\}$	observed	DD, DN, ND, NN	NN	NN
$O(1,0,1) = \{Z = 1, M_1 = 0, M_2 = 1\}$		DA, DC, NA, NC	-	-
$O(1,1,0) = \{Z = 1, M_1 = 1, M_2 = 0\}$	observed	AD, AN, CD, CN	AN, CN	AN, CN
$O(1,1,1) = \{Z = 1, M_1 = 1, M_2 = 1\}$		AA, AC, CA, CC	AA	CC

Note: Column (1) shows the principal strata without assumptions. Column (2) shows the remaining strata after Assumptions 2, 5, and 7 have been imposed and all all-move households are of type AA. Column (3) shows the remaining strata under the same assumptions if all all-move households are of type CC.

Table 4: Correspondence between observed groups and latent strata

Panel a: Observed groups		
$O(z, m_1, m_2)$	Number of households	
	all	visa all-move
$O(0,0,0)$	124	75
$O(0,0,1)$	-	
$O(0,1,0)$	-	
$O(0,1,1)$	-	
$O(1,0,0)$	26	18
$O(1,0,1)$	-	
$O(1,1,0)$	61	
$O(1,1,1)$	-	

Panel b: Latent strata proportions	
share	standard error
π_{AN}	0
π_{CN}	0.35
π_{CC}	0.53
π_{NN}	0.11
γ	1.5

Note: The left panel presents the observed number of households in the dataset of GMS by value of the instrument and migration status. Visa all-move households are households where all individuals would be eligible to join the principal migrant. The right panel shows the estimated ratio of unobserved to observed migrant households and the estimated strata proportions. Standard errors in parentheses from 999 bootstrap replications.

Table 5: Tonga: observed groups and latent strata

	Total Household Size		Adults Aged 18 to 45		Children Aged under 18		Adults Aged over 45	
$E[Y(0,0)\|G=NN]$	4.42***	(0.52)	2.15***	(0.21)	2.08***	(0.39)	0.19**	(0.08)
$E[Y(1,0)\|G=CN]$	4.69***	(0.34)	1.62***	(0.16)	1.92***	(0.24)	1.10***	(0.11)
Bounds under assumptions 2-7								
Bounds on $E[Y(0,0)\|G=CN]$	[2.17	8.56]	[1.34	3.51]	[0.22	5.03]	[0.00	1.04]
CLR 95% confidence interval	(1.58	9.28)	(1.14	3.94)	(0.00	5.62)	(0.00	1.25)
Bounds on θ_{CN}	[-3.87	2.54]	[-1.89	0.29]	[-3.11	1.72]	[0.06	1.10]
CLR 95% confidence interval	(-4.77	3.36)	(-2.39	0.61)	(-3.80	2.16)	(-0.22	1.29)
+ mean dominance assumption (9)								
Bounds on $E[Y(0,0)\|G=CN]$	[5.38	8.56]	[2.35	3.51]	[2.61	5.03]	[0.42	1.04]
CLR 95% confidence interval	(4.83	9.28)	(2.11	3.94)	(2.21	5.62)	(0.29	1.25)
Bounds on θ_{CN}	[-3.87	-0.69]	[-1.89	-0.73]	[-3.11	-0.69]	[0.06	0.68]
CLR 95% confidence interval	(-4.77	0.11)	(-2.39	-0.37)	(-3.86	-0.13)	(-0.22	0.91)
Unadjusted IV θ_W	-0.85	(0.54)	-0.76***	(0.24)	-0.78**	(0.38)	0.64***	(0.15)
θ_{CN}^{NS}	-0.69	(0.49)	-0.73***	(0.22)	-0.69**	(0.34)	0.68***	(0.14)
IV GMS	-2.26***	(0.62)	-1.53***	(0.34)	-0.86**	(0.42)	0.08	(0.18)
Observations	211		211		211		211	

Note: Results are based on the estimated strata proportions in Table 5. Standard errors in parentheses from 999 bootstrap replications. For the bounds without mean dominance, numbers in parentheses in the bottom rows are 95% confidence intervals calculated using the procedure suggested by Chernozhukov, Lee, and Rosen (2012), while numbers in square brackets are identified sets determined by the half-median unbiased estimators. For the bounds with mean dominance, the 95% confidence interval is calculated using the procedure suggested by Imbens and Manski (2004). * denotes that estimate is statistically different from zero at the 10%, ** at 5%, and *** at 1% significance level.

Table 6: Tonga: effect on household composition

	Home ownership		Any agricultural assets		Number of pigs		Number of chickens		Number of cattle	
$E[Y(0,0)\|G=NN]$	0.62***	(0.10)	0.85***	(0.07)	4.27***	(0.75)	3.54***	(0.79)	0.65***	(0.21)
$E[Y(1,0)\|G=CN]$	0.48***	(0.06)	0.79***	(0.05)	4.82***	(0.61)	4.33***	(0.72)	0.93***	(0.25)
Bounds under assumptions 2-7										
Bounds on $E[Y(0,0)\|G=CN]$	[0.00	1.00]	[0.63	1.00]	[1.05	8.97]	[0.00	13.59]	[0.00	3.16]
CLR 95% confidence interval	(0.00	1.00)	(0.45	1.00)	(0.44	10.02)	(0.00	15.96)	(0.00	4.05)
Bounds on θ_{CN}	[-0.52	0.48]	[-0.21	0.16]	[-4.14	3.81]	[-9.24	4.33]	[-2.23	0.93]
CLR 95% confidence interval	(-0.63	0.58)	(-0.30	0.36)	(-5.56	4.92)	(-11.83	5.52)	(-3.23	1.35)
+ mean dominance assumption (9)										
Bounds on $E[Y(0,0)\|G=CN]$	[0.40	1.00]	[0.86	1.00]	[4.92	8.97]	[5.94	13.58]	[1.27	3.16]
CLR 95% confidence interval	(0.31	1.00)	(0.80	1.00)	(4.22	10.02)	(4.53	15.96)	(0.87	4.05)
Bounds on θ_{CN}	[-0.52	0.08]	[-0.21	-0.07]	[-4.14	-0.10]	[-9.24	-1.62]	[-2.23	-0.34]
CLR 95% confidence interval	(-0.63	0.21)	(-0.30	0.04)	(-5.56	1.12)	(-11.83	0.18)	(-3.23	0.25)
Unadjusted IV θ_W	0.12	(0.09)	-0.07	(0.07)	-0.21	(0.82)	-2.01*	(1.22)	-0.44	(0.39)
θ_{CN}^{NS}	0.08	(0.08)	-0.07	(0.06)	-0.10	(0.74)	-1.62	(1.08)	-0.34	(0.35)
IV GMS	-0.01	(0.10)	-0.19***	(0.06)	-1.26	(0.80)	-4.58***	(1.64)	-0.85*	(0.51)
Observations	211		211		211		211		211	

Note: Results are based on the estimated strata proportions in Table 5. Standard errors in parentheses from 999 bootstrap replications. For the bounds without mean dominance, numbers in parentheses in the bottom rows are 95% confidence intervals calculated using the procedure suggested by Chernozhukov, Lee, and Rosen (2012), while numbers in square brackets are identified sets determined by the half-median unbiased estimators. For the bounds with mean dominance, the 95% confidence interval is calculated using the procedure suggested by Imbens and Manski (2004). * denotes that estimate is statistically different from zero at the 10%, ** at 5%, and *** at 1% significance level.

Table 7: Tonga: effect on household assets

	Boys		Girls	
π_{AN}	0.12***	(0.03)	0.12***	(0.03)
π_{CN}	0.26***	(0.04)	0.25***	(0.04)
π_{CC}	0.02***	(0.00)	0.02***	(0.00)
π_{NN}	0.60***	(0.04)	0.61***	(0.03)
$E\left[Y(0,0)\vert G=NN\right]$	0.73***	(0.02)	0.70***	(0.02)
$E\left[Y(1,0)\vert G=AN\right]$	0.77***	(0.03)	0.71***	(0.02)
$E\left[Y(1,0)\vert G=CN\right]$	0.63***	(0.04)	0.60***	(0.04)
Bounds under assumptions 2-7				
Bounds on $E\left[Y(0,0)\vert G=CN\right]$	[0.78	0.89]	[0.49	0.57]
CLR 95% confidence interval	(0.70	0.99)	(0.32	0.73)
Bounds on θ_{CN}	[-0.28	-0.14]	[-0.01	0.15]
CLR 95% confidence interval	(-0.40	-0.04)	(-0.16	0.31)
+ mean dominance assumption (8)				
Bounds on $E\left[Y(0,0)\vert G=CN\right]$	[0.78	0.82]	[0.49	0.53]
CLR 95% confidence interval	(0.70	0.92)	(0.32	0.68)
Bounds on θ_{CN}	[-0.19	-0.14]	[0.07	0.15]
CLR 95% confidence interval	(-0.32	-0.04)	(-0.11	0.31)
Unadjusted IV θ_W	-0.19**	(0.08)	0.08	(0.12)
θ_{CN}^{NS}	-0.19**	(0.08)	0.07	(0.11)
Observations	7,993		7,663	

Note: Results based on the assumption that the ratio of the number of children not included in the sample due to migration of the whole household to the number of children observed in migrant households is 0.054. Standard errors in parentheses from 999 bootstrap replications clustered at the state level. For the bounds without mean dominance, numbers in parentheses in the bottom rows are 95% confidence intervals calculated using the procedure suggested by Chernozhukov, Lee, and Rosen (2012), while numbers in square brackets are identified sets determined by the half-median unbiased estimators. For the bounds with mean dominance, the 95% confidence interval is calculated using the procedure suggested by Imbens and Manski (2004). * denotes that estimate is statistically different from zero at the 10%, ** at 5%, and *** at 1% significance level.

Table 8: Mexico: effect on school attendance

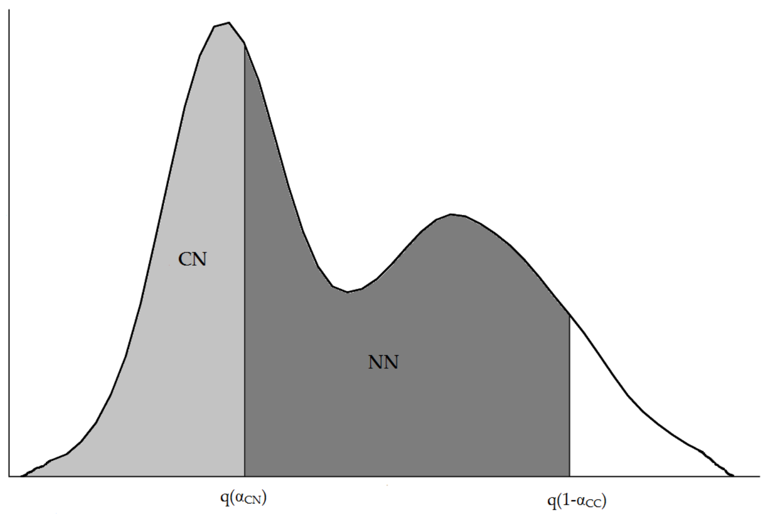

Figure 1: Unconstrained lower bound for $E[Y(0,0)|G = CN]$

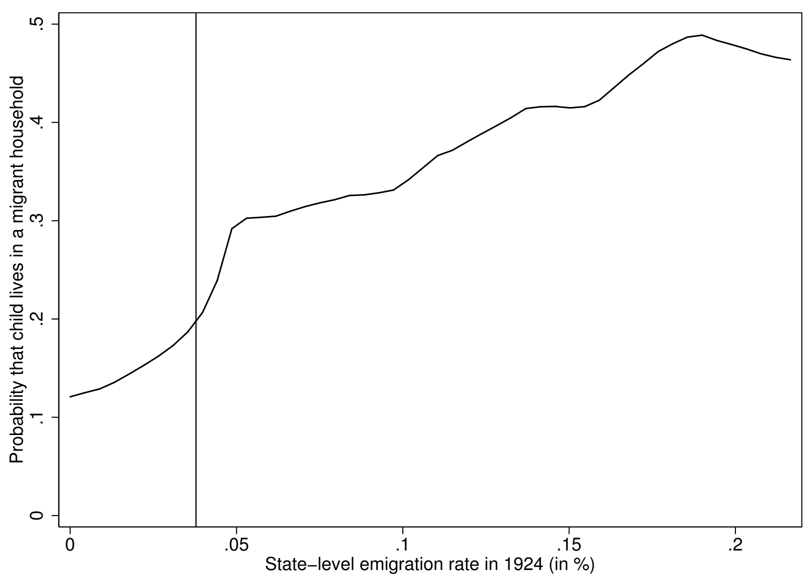

Figure 2: Cut-off for binary instrument

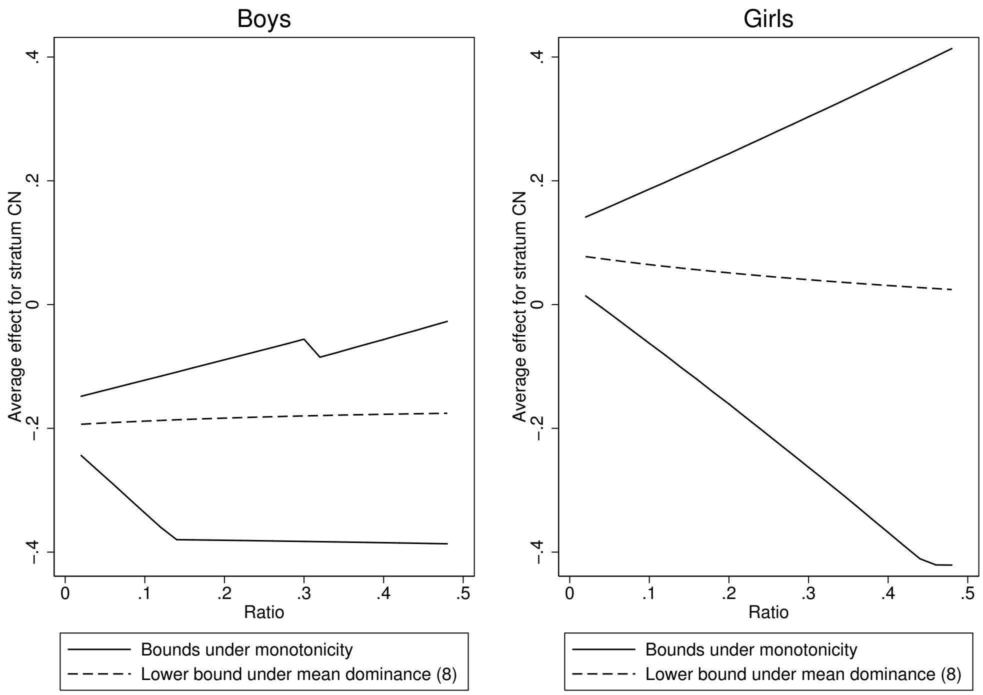

Figure 3: Sensitivity of bounds for different ratios of unobserved to observed children in migrant households

B Technical Appendix

B.1 Bounds on $E[Y(0,0)|G=N]$ with randomly assigned M_1

The observed outcome in group $O(0,0)$ is therefore a mixture of the potential outcomes of compliers and never migrants under control

$$E[Y|M_1=0,M_2=0] \quad = \quad E[Y(0,0)|G=C]\,\pi_C + E[Y(0,0)|G=N]\,\pi_N$$

This expression can be transformed to obtain the potential outcome of never migrants under control

$$E(Y(0,0)|G=N) \quad = \quad \frac{E[Y|M_1=0,M_2=0] - E[Y(0,0)|G=C]\,\pi_C}{\pi_N}$$

The upper bound on $E[Y(0,0)|G=C]$ can be obtained by taking the upper π_C quantiles in the observed group $O(0,0)$

$$E_N^U[Y(0,0)|G=C] = E[Y|M_1=0,M_2=0,Y \geqslant q(1-\pi_C)]$$

The respective lower bound can be obtained by taking the lower π_C quantiles. Thus the lower and upper bound for $E(Y(0,0)|G=N)$ can be rewritten as

$$\begin{aligned} E_N^L[Y(0,0)|G=N] &= \frac{E[Y|M_1=0,M_2=0]}{\pi_N} - \frac{E[Y|M_1=0,M_2=0,Y \geqslant q(1-\pi_C)]\,\pi_C}{\pi_N} \\ &= E[Y|M_1=0,M_2=0,Y < q(1-\pi_C)] \end{aligned}$$

$$\begin{aligned} E_N^U[Y(0,0)|G=N] &= \frac{E[Y|M_1=0,M_2=0]}{\pi_N} - \frac{E[Y|M_1=0,M_2=0,Y \leq q(\pi_C)]\,\pi_C}{\pi_N} \\ &= E[Y|M_1=0,M_2=0,Y > q(\pi_C)] \end{aligned}$$

The simplifications presented in these two equations make use from the fact that subtracting the weighted mean of Y in the upper (lower) π_C quantiles is equivalent of taking the mean in the lower (upper) $1-\pi_C$ quantiles.

B.2 Identification of strata proportions if migration of I_2 is observed

$$\begin{aligned}
\pi_{AN} &= P(M_1 = 1, M_2 = 0 | Z = 0) \\
\pi_{AA} &= P(M_1 = 1, M_2 = 1 | Z = 0) \\
\pi_{NN} &= P(M_1 = 0, M_2 = 0 | Z = 1) \\
\pi_{CC} &= P(M_1 = 1, M_2 = 1 | Z = 1) - P(M_1 = 1, M_2 = 1 | Z = 0) \\
\pi_{CN} &= P(M_1 = 1, M_2 = 0 | Z = 1) - P(M_1 = 1, M_2 = 0 | Z = 0)
\end{aligned}$$

B.3 Point identification if $\pi_{CC} = 0$

This section shows that the Wald estimator in the sample of observed households provides an unbiased estimate of $E\left[(Y_i(1,0) - Y_i(0,0))\,|G = CN\right]$ if $\pi_{CC} = 0$ and $\pi_{AA} \geq 0$.

$E\left[Y(1,0)|G = CN\right]$ is identifed in Equation (10)

$$E\left[Y(1,0)|G = CN\right] \quad = \quad \frac{\overline{Y}^{110}(\pi_{CN} + \pi_{AN}) - \overline{Y}^{010}\pi_{AN}}{\pi_{CN}}.$$

$E\left[Y(0,0)|G = CN\right]$ is identified from $\overline{Y}^{000} = \frac{E[Y(0,0)|G=CN]\pi_{CN} + E[Y(0,0)|G=NN]\pi_{NN}}{\pi_{CN}+\pi_{NN}}$ and $E\left[Y(0,0)|G = NN\right] = \overline{Y}^{100}$

$$E\left[Y(0,0)|G = CN\right] = \frac{(\pi_{CN} + \pi_{NN})\,\overline{Y}^{000} - \overline{Y}^{100}\pi_{NN}}{\pi_{CN}}.$$

Therefore, the causal effect is identfied as

$$\theta_{CN} = \frac{\left(\overline{Y}^{110}(\pi_{CN} + \pi_{AN}) - \overline{Y}^{010}\pi_{AN}\right) - \left((\pi_{CN} + \pi_{NN})\,\overline{Y}^{000} - \overline{Y}^{100}\pi_{NN}\right)}{\pi_{CN}}.$$

For comparison, consider a Wald estimator in the sample of observed households:

$$\theta_W = \frac{E[Y|Z = 1, M_2 = 0] - E[Y|Z = 0, M_2 = 0]}{E[M_1|Z = 1, M_2 = 0] - E[M_1|Z = 0, M_2 = 0]} \tag{30}$$

The four quantities in Equation (30) can be formulated as weighted means of observed outcomes and strata proportions

$$\begin{aligned}
E[Y|Z=1,M_2=0] &= \frac{\pi_{NN}*\overline{Y}^{100}+(\pi_{CN}+\pi_{AN})*\overline{Y}^{110}}{\pi_{NN}+\pi_{CN}+\pi_{AN}} \\
E[Y|Z=0,M_2=0] &= \frac{(\pi_{CN}+\pi_{NN})*\overline{Y}^{000}+\pi_{AN}*\overline{Y}^{010}}{\pi_{CN}+\pi_{NN}+\pi_{AN}} \\
E[M_1|Z=1,M_2=0] &= P(M_1=1|Z=1,M_2=0)=\frac{\pi_{CN}+\pi_{AN}}{\pi_{NN}+\pi_{CN}+\pi_{AN}} \\
E[M_1|Z=0,M_2=0] &= P(M_1=0|Z=1,M_2=0)=\frac{\pi_{AN}}{\pi_{CN}+\pi_{NN}+\pi_{AN}}
\end{aligned}$$

The Wald estimator

$$\theta_W=\frac{\left[\frac{\pi_{NN}*\overline{Y}^{100}+(\pi_{CN}+\pi_{AN})*\overline{Y}^{110}}{\pi_{NN}+\pi_{CN}+\pi_{AN}}\right]-\left[\frac{(\pi_{CN}+\pi_{NN})*\overline{Y}^{000}+\pi_{AN}*\overline{Y}^{010}}{\pi_{CN}+\pi_{NN}+\pi_{AN}}\right]}{\left[\frac{\pi_{CN}+\pi_{AN}}{\pi_{NN}+\pi_{CN}+\pi_{AN}}\right]-\left[\frac{\pi_{AN}}{\pi_{CN}+\pi_{NN}+\pi_{AN}}\right]}$$

simplifies to

$$\theta_W=\frac{\left(\pi_{NN}*\overline{Y}^{100}+(\pi_{CN}+\pi_{AN})*\overline{Y}^{110}\right)-\left((\pi_{CN}+\pi_{NN})*\overline{Y}^{000}+\pi_{AN}*\overline{Y}^{010}\right)}{\pi_{CN}},$$

which equals the effect for stratum CN:

$$\theta_W=\frac{\left(\overline{Y}^{110}(\pi_{CN}+\pi_{AN})-\overline{Y}^{010}\pi_{AN}\right)-\left((\pi_{CN}+\pi_{NN})\,\overline{Y}^{000}-\overline{Y}^{100}\pi_{NN}\right)}{\pi_{CN}}=\theta_{CN}.$$

B.4 Bias of instrumental variables estimate withouth systematic intra-household selection

Assumption 10 ($E[Y(0,0)|G=CC]=E[Y(0,0)|G=CN]$) allows point identification of θ_{CN} even if $\pi_{CC}>0$. $E[Y(1,0)|G=CN]$ is identifed in Equation (10)

$$E[Y(1,0)|G=CN] \quad = \quad \frac{\overline{Y}^{110}(\pi_{CN}+\pi_{AN})-\overline{Y}^{010}\pi_{AN}}{\pi_{CN}}.$$

$E[Y(0,0)|G=CN]$ can be identified using $\overline{Y}^{000}=\frac{E[Y(0,0)|G=CN,CC](\pi_{CN}+\pi_{CC})+E[Y(0,0)|G=NN]\pi_{NN}}{\pi_{CN}+\pi_{NN}+\pi_{CC}}$ and $E[Y(0,0)|G=NN]=\overline{Y}^{100}$. Assumption 10 implies that ($E[Y(0,0)|G=CN,CC]=E[Y(0,0)|G=CN]$) and therefore

$$E[Y(0,0)|G=CN]=\frac{\overline{Y}^{000}(\pi_{CN}+\pi_{NN}+\pi_{CC})-\overline{Y}^{100}\pi_{NN}}{\pi_{CN}}.$$

Therefore, the causal effect is identfied as

$$\theta_{CN}^{NS}=\frac{\overline{Y}^{110}(\pi_{CN}+\pi_{AN})-\overline{Y}^{010}\pi_{AN}}{\pi_{CN}}-\frac{\overline{Y}^{000}(\pi_{NN}+\pi_{CN}+\pi_{CC})-\overline{Y}^{100}\pi_{NN}}{\pi_{CN}+\pi_{CC}}.$$

For comparison, consider a Wald estimator in the sample of observed households

$$\theta_W = \frac{E[Y|Z=1, M_2=0] - E[Y|Z=0, M_2=0]}{E[M_1|Z=1, M_2=0] - E[M_1|Z=0, M_2=0]} \tag{31}$$

The four quantities in Equation 31 can be formulated as weighted means of observed outcomes and strata proportions, making use of $\pi_{CC} + \pi_{CN} + \pi_{NN} + \pi_{AN} = 1$.

$$\begin{aligned}
E[Y|Z=1, M_2=0] &= \frac{\pi_{NN} * \overline{Y}^{100} + (\pi_{CN} + \pi_{AN}) * \overline{Y}^{110}}{\pi_{NN} + \pi_{CN} + \pi_{AN}} \\
E[Y|Z=0, M_2=0] &= \frac{(\pi_{CC} + \pi_{CN} + \pi_{NN}) * \overline{Y}^{000} + \pi_{AN} * \overline{Y}^{010}}{\pi_{CC} + \pi_{CN} + \pi_{NN} + \pi_{AN}} = (\pi_{CC} + \pi_{CN} + \pi_{NN}) * \overline{Y}^{000} + \pi_{AN} * \overline{Y}^{010} \\
E[M_1|Z=1, M_2=0] &= P(M_1 = 1|Z=1, M_2=0) = \frac{\pi_{CN} + \pi_{AN}}{\pi_{NN} + \pi_{CN} + \pi_{AN}} \\
E[M_1|Z=0, M_2=0] &= P(M_1 = 0|Z=1, M_2=0) = \frac{\pi_{AN}}{\pi_{CC} + \pi_{CN} + \pi_{NN} + \pi_{AN}} = \pi_{AN}
\end{aligned}$$

The Wald estimator is

$$\theta_W = \frac{\left[\frac{\pi_{NN} * \overline{Y}^{100} + (\pi_{CN} + \pi_{AN}) * \overline{Y}^{110}}{\pi_{NN} + \pi_{CN} + \pi_{AN}}\right] - \left[(\pi_{CC} + \pi_{CN} + \pi_{NN}) * \overline{Y}^{000} + \pi_{AN} * \overline{Y}^{010}\right]}{\left[\frac{\pi_{CN} + \pi_{AN}}{\pi_{NN} + \pi_{CN} + \pi_{AN}}\right] - [\pi_{AN}]},$$

which can be simplified to

$$\theta_W = \frac{\pi_{NN} * \overline{Y}^{100} + (\pi_{CN} + \pi_{AN}) * \overline{Y}^{110} - (\pi_{NN} + \pi_{CN} + \pi_{AN}) * \left((\pi_{CC} + \pi_{CN} + \pi_{NN}) * \overline{Y}^{000} + \pi_{AN} * \overline{Y}^{010}\right)}{\pi_{CN} + \pi_{AN} - \pi_{AN} * (\pi_{NN} + \pi_{CN} + \pi_{AN})}$$

Subtracting θ_{CN}^{NS} from θ_W gives the bias of the Wald estimator

$$b_W = \theta_W - \theta_{CN}^{NS} = \frac{\pi_{CC}\left[\left(\overline{Y}^{100} - \overline{Y}^{000}\right) a + \left(\overline{Y}^{110} - \overline{Y}^{010}\right) b\right]}{c}$$

where

$$\begin{aligned}
a &= (\pi_{CN}\pi_{NN})(\pi_{CN} + \pi_{CC} + \pi_{NN}) \\
b &= \pi_{AN}\left(\pi_{CN}^2 + \pi_{CN}\pi_{CC} + \pi_{CN}\pi_{AN} + \pi_{CC}\pi_{AN}\right) \\
c &= \left(\pi_{CN}^2 + \pi_{CN}\pi_{CC}\right)\left(\pi_{CN}\pi_{CC} + \pi_{CC}\pi_{AN} + \pi_{CN}^2 + \pi_{CN}\pi_{AN} + \pi_{CN}\pi_{NN}\right).
\end{aligned}$$

B.5 Inference based on Chernozhukov, Lee, and Rosen (2009)

I explain the estimation procedure for $E_{CN}^L[Y(0,0)|G = CN]$. Recall that the lower bound of the expected value of stratum CN under control is given by $\Delta^L = \max_{v \in \mathcal{V} = \{0,1\}}[\Delta^L(v)]$, with $\Delta^L(0) = \overline{Y}(Y \leq y_{\alpha_{CN}}^{000})$ and $\Delta^L(1) = \overline{Y}(Y \leq y_{1-\alpha_{CC}}^{000}) * \frac{\pi_{NN} + \pi_{CN}}{\pi_{CN}} - \overline{Y}^{100} * \frac{\pi_{NN}}{\pi_{CN}}$. Let $\boldsymbol{\Delta}^L = [\Delta^L(0)\Delta^L(1)]'$ be the vector containing the two bounding functions. I subsequently discuss the estimation of the lower bound along with its confidence region (the proceeding for the upper bound is analogous). I use the procedure of Chernozhukov, Lee, and Rosen (2012) to obtain a half-median-unbiased estimator of $max_{v \in V}\left[\Delta^L(v)\right]$. This appendix is based on similar descriptions of this method in Chen and Flores (2012); Huber, Laffers, and Mellace (2014). The main idea is that instead of taking the maximum of the estimated $\hat{\Delta}^L(v)$ directly, one uses the following precision adjusted version, denoted by $\tilde{\Delta}^L(p)$, which consists of the initial estimate plus $s(v)$, a measure of the precision of $\hat{\Delta}^L(v)$, times an appropriate critical value $k(p)$:

$$\tilde{\Delta}^L(p) = \max_{\pi_{01,i}}[\hat{\Delta}^L(v) + k(p) \cdot s(v)].$$

As outlined below, $k(p)$ is a function of the sample size and the estimated variance-covariance matrix of $\sqrt{n}(\hat{\mathbf{\Delta}}^L - \mathbf{\Delta}^L)$, denoted by $\hat{\Omega}$. For $p = \frac{1}{2}$, the estimator $\tilde{\Delta}^L(p)$ is half-median-unbiased, which implies that the estimate of the upper bound exceeds its true value with probability at least one half asymptotically.

The following algorithm briefly sketches the estimation of Δ^L along with its upper confidence band based on the precision adjustment.

1. Estimate the vector $\hat{\mathbf{\Delta}}_{01}^{UB}$ by its sample analog. Estimate its variance-covariance matrix $\hat{\Omega}$ by bootstrapping B times.[34]

2. Denoting by $\hat{g}(v)^\top$ the v-th row of $\hat{\Omega}^{\frac{1}{2}}$, estimate $\hat{s}(v) = \frac{\|\hat{g}(v)\|}{\sqrt{n}}$, where $\|\cdot\|$ is the Euclidean norm.

3. Simulate R[35] draws, $H_1, \ldots, H_R$ from a $N(\mathbf{0}, I_2)$, where $\mathbf{0}$ and I_2 are the null vector and the identity matrix of dimension 2, respectively.

4. Let $H_r^*(v) = \hat{g}(v)^\top Z_r / \|\hat{g}(v)\|$ for $r = 1, \ldots, R$.

5. Let $\tilde{k}(c)$ be the c-th quantile of $\max_{v \in \mathcal{V}} H_r^*(v), r = 1, \ldots, R$, where $c = 1 - \frac{0.1}{\log(n)}$.

6. Compute the set estimator $\hat{\mathcal{V}} = \{v \in \mathcal{V} : \hat{\Delta}_{01}^{UB}(v') \le \max_{v' \in V}\{[\hat{\Delta}^L(v') + \tilde{k}(c) \cdot \hat{s}(v')] + 2 \cdot \tilde{k}(c) \cdot \hat{s}(v')\}\}$.

7. Estimate the critical value $\hat{k}(p)$ by the p-th quantile of $\max_{v \in \hat{\mathcal{V}}} H_r^*(v), r = 1, \ldots, R$.

8. For half-median-unbiasedness, set $p = \frac{1}{2}$ and compute $\tilde{\Delta}^L(\frac{1}{2}) = \max_{v \in \mathcal{V}}[\hat{\Delta}^L(v) + \hat{k}(\frac{1}{2}) \cdot \hat{s}(v)]$.

9. To obtain the upper confidence band, estimate the half-median-unbiased lower bound $\tilde{\Delta}^U(p)$.

10. Let $\Gamma = \max(0, \tilde{\Delta}^U(\frac{1}{2}) - \tilde{\Delta}^L(\frac{1}{2}))$, $\rho = \max(\tilde{\Delta}^L(\frac{3}{4}) - \tilde{\Delta}^L(\frac{1}{4}), \tilde{\Delta}^U(\frac{3}{4}) - \tilde{\Delta}^U(\frac{1}{4}))$ and $\tau = (\rho \log(n))^{-1}$. Compute $\hat{a} = 1 - \Phi(\tau \cdot \Gamma)\alpha$, where α is the chosen confidence level.

11. The lower confidence band for the estimate of Δ^L is obtained by $\tilde{\Delta}^L(\hat{a})$.

[34] In the empirical part I use 1,999 bootstrap replications.
[35] I set R=1,000,000.