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

Andreas Steinmayr†

Preliminary version. Comments are welcome.

Abstract

How does adult migration affect the educational attainment of children who stay behind? This paper revisits this question by addressing the problem of *double selection* – the decision to migrate and the subsequent decision whether only some or all members of a household migrate. In the latter case, the household will usually not be included in cross-sectional data at all. We tackle the resulting sample-selection problem by modeling the behavior of the household members using principal stratification. This allows identifying bounds on the effects of migration on school attendance of children who stay behind in Mexico using observational data. The results suggest that adult migration reduces school attendance rates of boys between 26 and 14 percentage points while the direction of the effect is ambiguous for girls. However, sensitivity analysis point to the fact that results are sensitive with respect to the share of all-move households.

Keywords: Migration, sample selection, principal stratification
JEL classification: C21, O15, J61

*I thank Xavier D'Haultfoeuille, Martin Huber, Michael Lechner, Toru Kitagawa, David McKenzie, Giovanni Mellace, and Steven Stillman for helpful discussions. I thank Christina Felfe and Conny Wunsch for helpful comments on the draft and seminar participants at the Labor Economics Seminar in Lech for helpful comments. All remaining errors are my own.

†Swiss Institute for Empirical Economic Research, University of St. Gallen. Email: andreas.steinmayr@unisg.ch

1 Introduction

With more than 215 million international migrants worldwide (World Bank, 2010) the costs and benefits of migration are a highly relevant policy question. In the past, research has addressed these question mainly from a destination country perspective. Recent research has given more attention to the effects of international migration on the sending country, especially in the context of migration from poor to rich countries. Of particular interest has been the question whether international migration can improve the welfare of the migrant's household and family members who stay behind in the sending country. Research has focused on investigating the effects of emigration on educational attainment of children, children's health, labor supply of spouses, and household poverty, among others.[1]

This paper focuses on the effects of migration on the educational attainment of children who stay behind. The direction of the effects is a priori unclear. While migrant-sending households benefit from remittance inflows, the absence of the migrant may have negative effects as well: the migrant is not earning local income and does not contribute to household production. In particular the migrant is absent as a caregiver for the children in the household and children may be required to undertake house-, farm-, or market-work. Furthermore, children from migrant families may be more likely to migrate in the future, which changes their incentives to invest in human capital.[2] The direction of the overall effects is therefore mainly an empirical question. Existing studies come to different conclusions about the direction of the effects of migration on educational attainment of children who stay behind. For example, Cox-Edwards and Ureta (2003); Yang (2008); Alcaraz, Chiquiar, and Salcedo (2012) find positive effects of living in a migrant household, while Lahaie, Hayes, Piper, and Heymann (2009); Giannelli and Mangiavacchi (2010); McKenzie and Rapoport (2011) find negative effects. One reason for the heterogeneous findings might be that these studies investigate different settings (e.g. different source and destination countries, different types of migration, short- vs. long-run effects). Another reason might be differences in the methodological approach and unresolved endogeneity problems.

[1] Antman (2013) provides a comprehensive overview of the literature on the effects of migration on the remaining household members.

[2] McKenzie and Rapoport (2011) show that the incentives to invest in education increase for children in migrant households if the returns to education in the potential destination country are higher than in the source country. The opposite is true if the returns to education are relatively lower in the potential destination country .

Research in this field is primarily based on data from source-country household surveys. The treatment *living in a migrant household* is usually defined as having at least one household member who has emigrated. The main identification problem that this literature has tried to tackle is the non-random selection of households into migration. However, recent research has pointed to other possible sources of endogeneity (Gibson, McKenzie, and Stillman, 2010, 2011a). Most importantly, among households involved in migration, some send a subset of members with the rest staying behind while other households migrate as a whole, i.e. some households migrate with their children, while other households leave the children behind. As the second decision is most likely also influenced by factors that are related to educational attainment, this second form of selection also leads to biased estimates of the effects of migration. Even worse, if all household members migrate, the household will usually not be included in cross-sectional survey data at all, as no household member is left to respond to the survey. This problem has been acknowledged for estimating the overall number of emigrants based on source-country survey data (including population censuses) (e.g. Ibarraran and Lubotsky, 2007). In the existing literature on the effects of migration on remaining household members, this form of endogeneity has largely been ignored. One of the reasons might be that the problem that arises for identification of causal effects is not obvious at first sight. A common argument is that if the interest is in the effects on remaining household members, the households that leave no members behind are not of interest anyway. Without further assumptions this argument is misguided, as we will explain in Section 2.

This paper contributes to the literature in two ways. The first contribution is methodological. We clearly structure the identification problem in the presence of these two forms of endogeneity by using the notation of principal stratification to model the behavior of the household members. We show the assumptions implicitly made about the selection process if the second form of endogeneity is ignored and discuss the consequences of a violation of these assumptions. We then derive nonparametric bounds for the effects of adult migration on children who stay behind under a transparent set of behavioral assumptions. The second contribution is substantive. We revisit the effect of migration on the educational attainment of children left behind in Mexico. We take into account that the observational data misses all-move households and derive bounds under different sets of assumptions. For the main scenario, we find a negative effect of adult migration on school attendance of boys that ranges between 26 and 14 percentage points. The direction of the effects for girls is ambiguous. However, sensitivity analysis point to the fact that the bounds are sensitive with respect to the share of households that migrate as a whole.

This paper connects to the econometric and statistical literature on partial identification in the presence of sample selection that goes back to Manski (1989, 1994). In particular it builds on the relatively recent approach of principal stratification that was introduced by Frangakis and Rubin (2002) to deal with post-treatment complications such as sample selection. Several recent papers have used principal stratification to derive bounds on the effects of policy interventions in the presence of post-treatment complications (see for example Zhang and Rubin, 2003; Mattei and Mealli, 2007; Zhang, Rubin, and Mealli, 2008; Huber and Mellace, 2013). Using principal stratification has the advantage that it allows us to characterize the potential – not the observed – behavior of household members. This allows us to be very transparent about the assumptions needed for the identification of causal effects and thus helps to reveal often hidden but crucial and sometimes not innocuous assumptions.

The remainder of the paper is structured as follows. Section 2 discusses the double-selection problem. Section 3 introduces an econometric framework to structure the identification problem, first under the assumption of randomly assigned adult migration (Section 3.2). In a second step, we extend this framework to an instrumental variable setting and derive bounds under two different sets of assumptions (Section 3.3). Section 4 illustrates the approach for the effects of adult migration on school attendance of children in Mexico. Section 5 concludes.

2 The effect of migration on children left behind and the double-selection problem

The effect of parental migration on the educational attainment of children has been investigated by various studies. Usually researchers investigate the case when one parent (or another adult member of the household) migrates and the child remains in the source location. Equation (1) displays a stylized version of a linear model as common in this 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 abroad (for simplicity of the argument, assume that households have only one adult individual). u_{ij} is an error term.

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

The selection problem addressed in most cases is the non-random selection of households into migration. Households who send a migrant may for example be wealthier and therefore find it easier to finance the cost of migration. Members of these households may also differ in terms of education, demographic characteristics or preferences from members of non-migrant households. Many of the factors that drive the migration decision may also influence the decision to invest in the human capital of the child, leading to an endogeneity problem. Thus, the main concern is that the error term is correlated with the variable of interest ($E[hmig_j u_{ij}] \neq 0$). Various strategies have been implemented to address this endogeneity, such as 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 an overview of the various approaches used in the literature see Antman (2013).

A second form of selection arises as in some households, which decide to engage in migration, not only one individual migrates but several or even all household members migrate (see Gibson, McKenzie, and Stillman, 2010, 2011a, for a related discussion). Also the children might be among the migrants. This gives rise to two problems. First, we usually do not observe the outcomes for the children who migrate. The children who stay behind and for whom we observe the outcome are a selected group. This complication becomes even worse by the way the data are normally collected. Household surveys in emigration countries usually ask the respondent whether one or several household members are currently abroad. Households that answer with yes to this question are referred to as migrant households (*treated*). Households that answer with no are referred to as non-migrant (*control*) households. However, if the whole household migrates, no individual is left to answer the survey and these households are therefore not included in cross-sectional datasets. We can therefore only estimate 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.

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

Instead of assuming that $hmig_j$ is uncorrelated with the error, this model requires that ($E[hmig_j s_j u_{ij}] = 0$). In other words, $hmig_j$ needs to be uncorrelated with the error in the sample of households which do not migrate as a whole and are therefore observed. However, this assumption is not enough. We furthermore require $E[s_j u_{ij}] = 0$. Assume for the moment that the migration status of the adult household member is randomly assigned and therefore $hmig_j$ is uncorrelated with u_{ij} and that the true effect of $hmig_j$ on Y_{ij} is zero. After households learn about their assigned $hmig_j$, they decide whether the children should migrate ($s_j = 0$) or stay ($s_j = 1$). It is reasonable to assume that migration of the adult increases the likelihood of migration of the children. If migration is costly, then only those households that can afford migration of the children will migrate with them. Those households that are observed are therefore on average poorer than those households that are not observed any more. At the same time, household wealth has a positive influence on educational attainment of the children (Leibowitz, 1974; Blau, 1999; Case, Lubotsky, and Paxson, 2002; Currie, 2009; Almond and Currie, 2011) and is thus in the error term u_{ij}. In the observed sample, $hmig_j$ is therefore negatively correlated with u_{ij} and a researcher who estimates Equation (2) would wrongly conclude the migration has a negative effect.

This particular form of *invisible* sample selection is usually ignored in existing studies that investigate the effects of migration on remaining household members. However, it is acknowledged by papers that estimate overall migrant numbers (Ibarraran and Lubotsky, 2007) or migrant selectivity (McKenzie and Rapoport, 2007). 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 may not always be possible to distinguish between migration and other forms of attrition.

Sample selection is only one problem that arises if children could potentially also be among the migrants. Assume that we could observe child outcomes, even if all household members migrate, e.g. by collecting data from peers in other households. In this case we could obtain unbiased estimates from Equation 1. Now we could estimate the overall effect of adult migration. This overall effect also includes the possibility that the child is among the migrants. However, migrating as a family from one country to another is obviously a different treatment as migration of an adult when the children stay behind. If interest is in the effect of adult migration on children staying behind, collecting data on all-move households does not solve the problem.

Recently, the use of (quasi-) experiments for future research on migration has been strongly encouraged (McKenzie and Yang, 2010; McKenzie, 2012). However, as randomization usually only addresses the first source of endogeneity (which households engage in migration), the second form (who and how many 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 explicitly address the second form of endogeneity has been to define a different parameter of interest and to estimate the effect only for those household (members) that can be identified as never migrants based on observable characteristics. Gibson, McKenzie, and Stillman (2010, 2011a) use the 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 all eligible individuals comply with the visa and join the principal migrant in case she migrates. It is thus possible to restrict the sample in the control and the treatment group to household members, who are not eligible to join the principal migrant and are therefore always observed. This subgroup consists primarily of siblings, nephews, nieces and parents of the migrant – individuals who are not in the migrant's nuclear family. The fact that in this setting all migrants take their children with them makes it impossible to identify the effect of parental migration on children's outcome, which is one of the most important parameters for policy makers. In a similar setting, Mergo (2011) drops all households from the control group, where the household head filed the visa application and thus it seems possible that all household members would have joined the household head in case she would have won in the visa lottery.

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 explicitly address it (Yang, 2008; Antman, 2011).

3 Econometric framework

3.1 Setup and parameter of interest

Following the treatment evaluation literature, we use a potential outcome framework initially developed by Rubin (1974). The idea of this approach 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 it lives in a migrant household and whether the same child would attend school if it does not live in a migrant household. The obvious 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 empirical application, we will refer to these individuals as adult (I_1) and child (I_2). While this might seem to be a strong simplification, it does not limit the applicability of this framework to only this type of households. We will discuss the consequences of this simplification when introducing behavioral assumptions and in the empirical example.

$M_j = m_j \epsilon \{0, 1\}$ denotes the migration status of individual j. I_1 is the principal migrant who makes the first migration decision and chooses either to stay ($M_1 = 0$) or migrate ($M_1 = 1$). We will first discuss the general selection problem under the simplifying assumption of randomly assigned M_1. This assumption will be relaxed in a second step. I_2 chooses either to stay ($M_2 = 0$) or migrate ($M_2 = 1$) depending on the choice of I_1. It is important to note that this need not necessarily be a sequential decision process. The decision regarding the migration of the child could also be made by the adult simultaneously with her own decision to migrate. The resulting sample selection problem is identical. The central problem is that the migration of the child depends on the migration of the adult but not vice versa.

If migration of an adult household member is considered the treatment of interest, then the migration of children may 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 (stayers) is observed are endogenously selected and the treatment influences the selection.

We observe the outcome Y at some point in time after M_1 and M_2 have realized. In the empirical application Y is school attendance of the child. We define a set of potential outcomes for Y and M_2. Y depends on the migration state of the adult and the child and therefore is a function of M_1 *and* M_2. Y depends on M_1 as migration of an adult household member is likely to affect the educational attainment of the child. Furthermore, Y depends on M_2 as migration of the child itself also influences 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.

We assume to have a *random* sample of n households from the population in the source country, which was drawn after the households were treated, i.e. households engaged in migration. The sample, and also the population, at this point in time do not include any households with $M_1 = 1$ *and* $M_2 = 1$. Although the sample is representative for the population at that given point in time, the population we observe is different from the population before households engaged in migration and this change in the composition of the population is a function of migration.

We rule out interaction effects between units of different households, an assumption which is commonly referred to as Stable Unit Treatment Value Assumption (SUTVA) (Rubin, 1980). In most applications SUTVA implies that the potential outcomes of a unit are independent of treatment status of any other units. In the application in this paper, it implies that potential outcomes of a child are not affected by the treatment of units in other households. In other words, school attendance does not depend on the migration state of other households but it depends on the migration state of other household members.

In this setting we can distinguish between several different 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. Researches might also be interested in $Y(1,1) - Y(0,0)$, the effect if the child migrates with the adult, compared 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 compared to a situation in which the child remains behind while the adult migrates (e. g. Gibson, McKenzie, and Stillman, 2011b). We will focus on $Y(1,0) - Y(0,0)$ as this effect has received most attention in the literature. If we do not assume that the effects of migration are homogenous for all individuals (treatment effect homogeneity), we furthermore need to define the population for which we want to identify the effect. We will 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 therefore whether a particular individual belongs to this group is not observable as only either $M_2(0)$ or $M_2(1)$ can be observed but not both. We focus on this group as it is the only group for which the outcome is observed under both migration states of the adult. Furthermore, in countries with predominantly labor migration, where 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

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

3.2 Identification with randomly assigned adult migration status

In order to focus on the identification problem induced by the migration of I_2, we will assume random assignment of the migration status of I_1. In a second step, we will relax the assumption of random assignment of M_1. From the random assignment of M_1 it follows that all potential outcomes are independent of M_1 (Assumption 1). However, the actual outcomes are not independent of M_1. If M_1 affects the migration status of the child and the outcome variable, then the observed outcomes differ for households with $M_1 = 0$ and $M_1 = 1$.

Assumption 1. *Randomly assigned migration status of* I_1

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

Stratification on potential migration behavior

Consider now 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) we refer to these groups as PRINCIPAL STRATA. Principal strata are sub-populations of units (in our case households) that share the same potential values of intermediate variables under different treatment states. We can distinguish four different possible combinations of potential migration behavior of I_2 (Table 1). Note that this 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 the potential behavior of units with respect to an instrumental variable. In our setting the types describe the potential migration behavior of the children with respect to the migration status of the adult. With reference to the LATE framework we refer to the types (G) as *always migrants*, *compliers*, *defiers*, and *never migrants*. Children characterized as always migrants would migrate, irrespective of the migration status of the adult. Compliers would migrate if the adult migrates, but would stay if the adult stays. Defiers would migrate if the adult stays and would 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. In reality not all strata must necessarily exist.

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

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

The idea of principal stratification is to compare units within common principal strata. As treatment assignment does not affect membership to a particular principal stratum, the estimated effects are causal effects (Frangakis and Rubin, 2002). A principal stratum carries only the information whether a child would migrate or stay depending whether the adult migrates or stays, irrespective of the actual migration status of the adult. Conditional on the principal strata, potential outcomes $Y(m_1, m_2)$ are independent of the treatment M_1. Conditioning on principal strata would be equivalent to conditioning on the characteristics reflected in the post-treatment variable. 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 identification problems become more obvious from 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 composed of compliers and never migrants (Column (1)). Only for these two principal strata is it possible to observe this combination of M_1 and M_2. Similar for the other observed groups: the observed group $O(0,1)$ is composed of always migrants and defiers, the observed group $O(1,0)$ is composed of defiers and never migrants, and the observed group $O(1,1)$ is composed of always migrants and compliers.

A researcher ignoring the second selection problem might estimate the difference $E[Y|M_1 = 1, M_2 = 0] - E[Y|M_1 = 0, M_2 = 0]$. However, this would mean comparing strata D and N under treatment with strata C and N under control. This difference does not reflect a causal effect as individuals/households with different characteristics are compared. The assumption one would have to make in order 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 very

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

strong assumption.

As explained above, a principal effect within a stratum is a well-defined causal effect. One can therefore estimate the effects within each stratum and then aggregate to obtain the effect for the population of interest. For policy makers in the source country, the children who stay behind are of particular interest. If interest is in $Y(1,0) - Y(0,0)$, the only stratum for which both potential outcomes can be observed are never migrants.[3] The average partial effect for never migrants is defined as

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

Note that this is identical to the effect defined in Equation 3. Subsequently we will focus on the 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.

Bounds on the treatment effect

Additional behavioral and distributional assumptions can be used to derive bounds for the effect of interest. One relatively weak behavioral assumption in the setting, where I_2 is a child, is that I_2 would not migrate alone. If the household would have more than one adult, then this assumption means that: the child would not migrate if not at least one adult migrates. This assumption rules out the existence of always migrants and defiers, as children in these two strata would migrate if the adult would not migrate.

[3] Note that never migrants are not equal to the group with $M_2 = 0$. This observed group also includes compliers.

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. As Assumption 2 rules out defiers and always migrants we should not observe any households with the combination $M_1 = 0$ and $M_2 = 1$, meaning any household where all adult members stay and only a child migrates.[4] Given Assumption 2, group $O(1,0)$ corresponds directly to the stratum of never migrants under treatment. Therefore the outcome under treatment for never migrants is directly identified

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

Group $O(0,0)$ is a mixture of compliers and never migrants. The observed outcome is therefore a mixture of the potential outcomes of these two strata 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

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

The share of compliers and never migrants could be directly obtained from $\pi_C = P\left(M_2=1|M_1=1\right)$ and $\pi_N = P\left(M_2=0|M_1=1\right)$ if at least the existence of households where all individuals migrated is known. This might be the case in a panel dataset where households dissolve between two waves but the information about their migration is available from other sources. Information about the existence of these households is usually not available in cross-sectional datasets. In this case strata proportions have to be estimated by using other

[4] Note that for the bounds derived below also a weaker monotonicity assumption that rules out defiers would be sufficient. We still use Assumption 2 as it is necessary for identification in the setting where migration of the adult is not random. Furthermore it is not rejected by our data.

data sources or have to be based on assumptions. In the empirical application we obtain the number of all-move households from comparing migrant numbers from the source country population census with the migrant numbers from the destination country census. We calculate the ratio of the number of children not included in the data relative to the observed number of children in migrant households (γ). Based on this information we can calculate the strata proportions $\pi_N = {}^1\!/_{1+\gamma}$ and $\pi_C = 1 - \pi_N$.

Following Zhang and Rubin (2003); Lee (2009) we can derive sharp[5] bounds for $E[Y(0,0)|G = N]$ and θ_N. The idea behind these bounds is simple. We know that the observed group of households where neither the adult nor the 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 that 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 the 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 $Y|M_1 = 0, M_2 = 0$. $E[Y(0,0)|G = C]$ can be bounded from above by the mean of Y in the upper $1-\pi_C$ quantiles of the distribution in the cell $O(0,0)$ and from below by the mean in the lower π_C quantiles.[6] To directly obtain bounds for $E[Y(0,0)|G = N]$ we take take the mean in the lower $1-\pi_C$ quantiles for the lower bound and in the upper π_C quantiles for the upper bound (see Appendix B for the calculations).

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

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

[6] Note that if Y is discrete, the occurrence of mass points with equal outcome values cause the quantile function to be not unique. For this reason we replace the non-unique quantile function with a modified version as suggested in Kitagawa (2009) and Huber and Mellace (2013). Intuitively, we use a rank function instead of a quantile function to break ties. We sort the data in the observed cell $M_1 = 0, M_2 = 0$ on the outcome. For the lower bound we then estimate the mean in the subsample of the first $\pi_C * N_{00}$ observations, where N_{00} denotes the number of observations with $M_1 = 0, M_2 = 0$. For the upper bound we estimate the mean in the subsample of the last $\pi_C * N_{00}$ observations.

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

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 adult migration

In practice many empirical studies use an instrument for the migration decision of the principal migrant (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). We therefore drop the assumption of random assignment of M_1 and assume that a binary instrument $Z=z\epsilon\{0,1\}$ exists, which is randomly assigned and affects the migration decision of the adult. $M_1(z)$ denotes the potential migration of I_1 as a function of the value of the 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 a function of Z. In the presence of the second selection problem, we have to modify the classical IV assumptions (Imbens and Angrist, 1994; Angrist, Imbens, and Rubin, 1996). Specifically we make the following assumptions. We assume that the instrument is randomly assigned and therefore independent of all potential outcomes (Assumption 3).

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 states that the effect of Z on the potential outcomes Y must be via an effect of Z on M_1 and M_2 (the effect of Z on M_2 is indirect via M_1). In other words, the instrument may affect the educational outcomes of the children only through its effect on the migration status of the household members. Assumption 5 states that the effect of the instrument on the potential migration status of I_2 must be via an effect of Z on M_1. In other words, the decision of the household whether only the adult migrates or the whole household migrates, does not depend on the value of the instrument. In a later step we will derive alternative bounds for the case that this assumption is violated by replacing Assumption 5 with an additional monotonicity assumption. Assumptions 5 and 4 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 the migration status only.

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. *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 states that the instrument has a non-zero average effect on the migration of I_1. For the moment we 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 needs to satisfy Assumptions 3, 4, 5, and 6 simultaneously (Imbens and Angrist, 1994; Angrist, Imbens, and Rubin, 1996). An important difference with respect to the exclusion restriction is, that we require Z to be a valid instrument for Y and M_2. In this sense our setting is very similar to Chen and Flores (2012). However, there are two differences to their setting. First, in our setting M_2 is both an indicator whether the individual is observed and a treatment in itself. In Chen and Flores (2012) the outcome is not a function of the selection indicator. Second, in our setting the probability to observe a household decreases with adult migration as this increases the probability that the whole household migrates. In the setting under study in Chen and Flores (2012) the probability to observe the outcome increases for treated individuals.

We now distinguish principal strata with respect to the *instrument*. We can differentiate the types of adults with respect to 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 on the value of one but not if it takes on the value of zero; a defier would migrate if the instrument is zero but not if the instrument is one; and a never migrant would not migrate irrespective of the value of the instrument. We can also distinguish these four types of children. Note that we define the types of the children also with respect to the instrument, even though we assume that the effect works only indirectly via M_1. Combining the four strata of adults with the four strata of children gives in total $4 \times 4 = 16$ principal strata (Table 4 in Appendix A). We refer to the strata (household types) 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 where the adult would migrate if $Z = 1$ and would not migrate if $Z = 0$ and the child would never migrate.

Assumption 5 rules out the existence of strata AC, AD, NC, ND. In these strata the instrument has a direct effect on M_2, as I_1 does not react to the instrument in these strata. Furthermore, we continue to assume that the child would only migrate if the adult migrates (Assumption 2). This assumption rules out the existence of the strata $CA, CD, DA, DC, NA, NC, ND$.[7] Again, this has the empirically testable assumption that no households with $M_1 = 0$ and $M_2 = 1$ should be observed. Additionally, we 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 states that every adult is at least as likely to migrate if $Z = 1$ as she would be if $Z = 0$.

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

$$M_{i1}(0) \leq M_{i1}(1)$$

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

The outcome Y is observed under treatment and control only 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. The causal effect for this stratum is therefore the local

[7]The existence of some strata is ruled out by more than one assumption.

average treatment effect (LATE) for children who are never migrants. In what follows we will concentrate on the identification of this effect under the proposed set of assumptions.

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

Bounds on the treatment effect

Identification of the strata proportions is necessary in order to bound the treatment effect. If all-move households are not observed in the data, the identification of strata proportion requires again external information about the ratio of the number of children not observed to the observed number of children in migrant households (γ) (see Section 4.3 for an explanation how we calculate γ using information from other data sources). Strata proportions cannot just be estimated as conditional probabilities but need to be adjusted due to the fact that not the entire sample is observed. For example, while π_{AN} is $P\left(M_1 = 1, M_2 = 0|Z = 0\right)$ if the entire sample is observed, π_{AN} would be overestimated if we ignore the fact that we do not observe households with $M_1 = 1, M_2 = 1, Z = 0$. For this reason we calculate adjustment factors based on γ. The adjustment factor in the sub-sample with $Z = 0$ is $\lambda_0 = N_0/(N_0 + N_{010} * \gamma)$ and in the sub-sample with $Z = 1$ it is $\lambda_1 = N_1/(N_1 + N_{110} * \gamma)$. N_z denotes the number of observations with $Z = z$, N_{z10} the number of observations with $Z = z, M_1 = 0, M_2 = 0$. The terms $(N_0 + N_{010} * \gamma)$ and $(N_1 + N_{110} * \gamma)$ correspond to the numbers of households in the subsamples with $Z = 0$ and $Z = 1$, that we would observe if all-move households were also observable. Given this information, strata proportions are identified as

$$\begin{aligned}
\pi_{AN} &= P\left(M_1 = 1, M_2 = 0|Z = 0\right) * \lambda_0 \\
\pi_{CN} &= P\left(M_1 = 1, M_2 = 0|Z = 1\right) * \lambda_1 - \pi_{AN} \\
\pi_{NN} &= P\left(M_1 = 0, M_2 = 0|Z = 1\right) * \lambda_1 \\
\pi_{CC} &= \pi_{CN} * (\gamma) \\
\pi_{AA} &= \pi_{AN} * (\gamma)
\end{aligned}$$

To simplify notation, we denote $\overline{Y}^{zm_1m_2} \equiv E\left[Y|Z = z, M_1 = m_1, M_2 = m_2\right]$ for the observed outcomes. We denote $\alpha_{CN} \equiv \pi_{CN}/\left(\pi_{CN} + \pi_{NN} + \pi_{CC}\right)$ and

$\alpha_{CC} \equiv \pi_{CC}/\left(\pi_{CN}+\pi_{NN}+\pi_{CC}\right)$ for the conditional probabilities in the observed group $O(0,0,0)$.

The potential outcome of CN under treatment, $Y(1,0)|G = CN$, is observed as part of the mixture distribution in the observed group $O(1,1,0)$.

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

which can be reformulated to

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

Under treatment stratum AN corresponds directly to the observed group $O(0,1,0)$ and the 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 expected outcome under treatment for stratum CN is point identified as

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

The approach to derive bounds for the potential outcome of CN under control stems from Chen and Flores (2012). They derive bounds for a situation where the potential outcome of interest is part of a mixture of three strata and the expected outcome of one stratum is point identified. In our setting, the observed outcome for the group $O(0,0,0)$ is a mixture of the outcomes of strata CN, NN, and CC and the outcome of stratum NN is point identified (see below).

We introduce additional notation to describe the bounds. 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

$$\overline{Y}(y_{a'}^{000} \leq Y \leq y_a^{000}) \equiv E\left[Y|Z=0, M_1=0, M_2=0, y_{a'}^{000} \leq Y \leq y_a^{000}\right] \tag{13}$$

The idea behind these bounds is to find the lowest and highest possible values for $E\left[Y(0,0)|G=CN\right]$ subject to the constraint $\overline{Y}^{100} = E\left[Y(0,0)|G=NN\right]$. In the unconstrained case, the upper and lower bound for $E\left[Y(0,0)|G=CN\right]$ can be derived in a similar way as in the scenario with randomly assigned

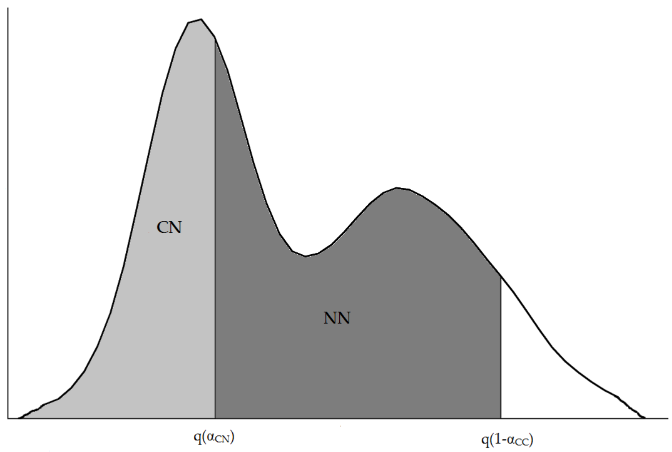

Figure 1: Unconstrained lower bound for $E\left[Y(0,0)|G=CN\right]$

M_1. We can bound $E\left[Y(0,0)|G=CN\right]$ from below by the expected value of Y for the α_{CN} fraction of smallest values of Y in the group $O(0,0,0)$. Now we check whether this unconstrained solution can satisfy the constraint that $\overline{Y}^{100} = E\left[Y(0,0)|G=NN\right]$. Under the assumptions that the smallest values in group $O(0,0,0)$ are only from CN observations, the lower bound for $E\left[Y(0,0)|G=NN\right]$ is given by $\overline{Y}(y^{000}_{\alpha_{CN}} \le Y \le y^{000}_{1-\alpha_{CC}})$, 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.

If the constraint is not satisfied, the lower bound can be derived from the mixture distribution of CN and NN in the lower $1-\alpha_{CC}$ quantiles of the distribution of Y in the cell $\{Z=0, M_1=0, M_2=0\}$ (Chen and Flores, 2012).

$$E^{L}_{CN}\left[Y(0,0)|G=CN\right] = \begin{cases} \overline{Y}(Y \le y^{000}_{\alpha_{CN}}),\ if\ \overline{Y}(y^{000}_{\alpha_{CN}} \le Y \le y^{000}_{1-\alpha_{CC}}) \le \overline{Y}^{100} \\ \overline{Y}(Y \le y^{000}_{1-\alpha_{CC}}) * \frac{\pi_{NN}+\pi_{CN}}{\pi_{CN}} - \overline{Y}^{100} * \frac{\pi_{NN}}{\pi_{CN}},\ otherwise \end{cases} \tag{14}$$

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

Bounds for the causal effect θ_{CN} can be constructed by combining the point identified potential outcomes under treatment with the bounds for the potential outcomes of stratum CN under control.

$$\theta^U_{CN} = E\left[Y(1,0)|G=CN\right] - E^L_{CN}\left[Y(0,0)|G=CN\right] \tag{16}$$

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

Alternative bounds without exclusion restriction of Z on M_2

Assumption 5 may be controversial in some settings. For example, if the proposed instrument shifts the cost of migration, one could imagine that this does not only influence the migration decision of the adult but of all household members. In this case the exclusion restriction would be violated. As stated above, Assumption 5 rules out the existence of strata AC, AD, NC, ND and thus allows identification of the bounds as described above. However, Assumption 2 rules out the existence of strata NC and ND as well. By imposing monotonicity of M_2 in Z we can also rule out the existence of stratum AD. Assumption 8 states that the probability to migrate for children must be strictly higher if the instrument is one compared to the situation where the instrument is zero, which is most likely the case if the instrument reduces the cost of migration.

Assumption 8. *Individual-level monotonicity of M_2 in Z*

$$M_{i2}(0) \leq M_{i2}(1)$$

The difference to the situation before is that we cannot rule out the existence of stratum AC. Identification of the bounds for $E\left[Y(0,0)|G=CN\right]$ is unaffected by this change, except that identification of the strata proportion requires additional assumptions. In the scenario where the migration of the child was independent of the instrument, it was enough to calculate the ratio of the number of not observed children to the observed number of children in migrant households using other data sources. We continue to use the overall ratio γ. However, we need to make two additional assumptions.

Assumption 9. *γ independent of household type*

$$\pi_{CN} * \gamma = \pi_{CC} \text{ and } \pi_{AN} * \gamma = \pi_{AC} + \pi_{AA}$$

Assumption 10. *The shares of compliers and always migrants among children are equal in households where the adult is an always migrant*

$$\pi_{AC} = \pi_{AA}$$

Assumption 9 states that γ is equal for households where the adult is a complier and households where the adult is an always migrant. We make this assumption in the absence of reliable data on potential differences in γ for different types of households. Note that this assumption implies a trade-off. If all all-move households were households where the adult is a complier, then π_{CC} would be large and thus the bounds on $E\left[Y(0,0)|G=CN\right]$ would be large but we could still point identify $E\left[Y(1,0)|G=CN\right]$. On the contrary, if all all-move households were households where the adult is an always-migrant, then we could point identify $E\left[Y(0,0)|G=CN\right]$ but we would get larger bounds on $E\left[Y(1,0)|G=CN\right]$.

We also assume that the share of compliers among children is identical to the share of always migrants among households where the adult is an always migrant (Assumption 10). The case that would lead to the widest bounds on $E\left[Y(1,0)|G=CN\right]$ would be to assume that $\pi_{AA}=0$ and therefore $\pi_{AN} * \gamma = \pi_{AC}$. However, as long as π_{AC} is small compared to π_{AN}, the bounds will only slightly increase compared to the ones under Assumption 5. The formulas for the strata proportions using Assumptions 9 and 10 are given in Appendix B.

Under Assumptions 8, 9, and 10 it is no longer possible to point-identify $E\left[Y(1,0)|G=AN\right]$ and in further consequence $E\left[Y(1,0)|G=CN\right]$. However, it is possible to derive sharp bounds on $E\left[Y(1,0)|G=CN\right]$.[8] $Y(1,0)|G=AN$ is observed in the groups $O(0,1,0)$ and $O(1,1,0)$. Within each of these cells we can bound $E\left[Y(1,0)|G=AN\right]$ from below by the expected value of Y in the α_{AN}^{z10} fraction of smallest values of Y for $z=0,1$.[9] The sharp lower bound for $E\left[Y(1,0)|G=AN\right]$ is the maximum of the two. For the upper bound we take the α_{AN}^{z10} fractions of largest values of Y in the two cells and then the minimum of the two. We then use an adjusted version of Equation 12. Instead of using

[8] See (Huber and Mellace, 2013) for the proof of sharpness of these bounds.

[9] $\alpha_{AN}^{010} \equiv \pi_{AN}/\left(\pi_{AN}+\pi_{AC}\right)$ denotes the share of AN households in the observed group $O(0,1,0)$ and $\alpha_{AN}^{110} \equiv \pi_{AN}/\left(\pi_{AN}+\pi_{CN}\right)$ in the observed group $O(1,1,0)$.

the point identified expected outcome for stratum CN under treatment, we use the upper and lower bounds for this quantity (Equations 18 and 19).

$$\begin{aligned} E^L_{CN}\left[Y(1,0)|G=CN\right] &= \overline{Y}^{110} * \frac{\pi_{AN}+\pi_{CN}}{\pi_{CN}} \\ &- \ min\left\{\overline{Y}(Y \geq y^{010}_{1-\alpha^{010}_{AN}}), \overline{Y}(Y \geq y^{110}_{1-\alpha^{110}_{AN}})\right\} * \frac{\pi_{AN}}{\pi_{CN}} \end{aligned} \tag{18}$$

$$\begin{aligned} E^U_{CN}\left[Y(1,0)|G=CN\right] &= \overline{Y}^{110} * \frac{\pi_{AN}+\pi_{CN}}{\pi_{CN}} \\ &- \ max\left\{\overline{Y}(Y \leq y^{010}_{\alpha^{010}_{AN}}), \overline{Y}(Y \leq y^{110}_{\alpha^{110}_{AN}})\right\} * \frac{\pi_{AN}}{\pi_{CN}} \end{aligned} \tag{19}$$

Bounds for the causal effect θ_{CN} can be constructed by combining the bounds for the potential outcomes of stratum CN under control with the bounds for potential outcomes of CN under treatment.

$$\theta^U_{CN} = E^U_{CN}\left[Y(1,0)|G=CN\right] - E^L_{CN}\left[Y(0,0)|G=CN\right] \tag{20}$$

$$\theta^L_{CN} = E^L_{CN}\left[Y(1,0)|G=CN\right] - E^U_{CN}\left[Y(0,0)|G=CN\right] \tag{21}$$

3.4 Estimation and inference

All bounds in the setting with imperfect compliance include at least one minimum or maximum operator.[10] 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 the min and max functions can be severely biased in finite samples and confidence intervals can neither be estimated using standard asymptotics nor bootstrap methods. Chernozhukov, Lee, and Rosen (2012) derive a method to obtain conservative half-median unbiased estimates and confidence intervals for the bounds. The main idea of their approach is to apply the min (max) function not directly on the bounding function but on a precision corrected version of it. Precision is adjusted by adding to each estimated bounding function its pointwise standard error times

[10] Equations 18 and 19 are written in terms of min and max operators. But also equations 14 and 15 involve these operators. For example, equation 14 can be rewritten as $E^L_{CN}\left[Y(0,0)|G=CN\right] = max\left\{\overline{Y}(Y \leq y^{000}_{\alpha_{CN}}), \overline{Y}(Y \leq y^{000}_{1-\alpha_{CC}}) * \frac{\pi_{NN}+\pi_{CN}}{\pi_{CN}} - \overline{Y}^{100} * \frac{\pi_{NN}}{\pi_{CN}}\right\}$.

an appropriate critical value. Estimates with higher standard errors therefore require larger adjustments. The estimated bounds are rather conservative. 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. A detailed description of the implementation of the procedure based on Huber and Mellace (2013); Chen and Flores (2012) is provided in Appendix B.

4 Empirical example: Migration and educational attainment in Mexico

Mexico has been the most studied source country in migration research. In the empirical application we follow McKenzie and Rapoport (2011) (henceforth MR) and estimate the effect of migration on school attendance in Mexico. MR use historical migration rates as an instrument for current migration to identify the effects for four distinct groups: boys and girls in the age groups of 12 to 15 and 16 to 18 years. They find that migration of an adult household member reduces school attendance rates of 12 to 15 year old boys by 16 percentage points and by 9 percentage boys for girls, the latter effect is however not significantly different form zero. MR also estimate the impact on years of education attained and for children aged 16 to 18 years. In the current paper we will only focus on the effect of migration on school attendance and only on the sample of children aged 12 to 15 years. Restricting the analysis in this way has two advantages for the proposed research. 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. Second, compared to years of education, school attendance is the more natural outcome for children and adolescents who have not yet completed their education. Using years of education gives rise to additional problems due to censoring.

4.1 Data

This paper uses the same Mexican dataset as MR. The 1997 Encuesta Nacional de la Dinàmica Demográfica (ENADID) is a nationally representative survey with a total sample of 73,412 households. This corresponds to roughly 2,300 households in each of the 32 states.[11] In order to allow for comparability of the results with MR, we restrict the sample in a similar way 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. We 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 U.S. to work or who has moved to the U.S. in the last 5 years for any other reason. It should be noted that this is not the optimal migrant definition to illustrate the effect of endogeneity due to migrants sample selection as it also includes return migrants. However, to maximize comparability of the results with previous research, we follow the definition of MR.[12] MR argue that also prior migration episodes of adult household members influence the educational attainment of the children. If the instrument also affects these prior migration episodes, then defining the treatment in a way that households with former migration episodes are considered as untreated would lead to a violation of the exclusion restriction. The instrument would have an effect on the outcome that is not via the defined treatment.

The outcome of interest is school attendance. Even as school attendance in Mexico is compulsory up to the age of 16 years[13], attendance rates at the time of the survey were significantly below 100 percent. Attendance rates in the estimation sample are 74% for boys and 66% for girls. The overall attendance rate drops from 87% at age 12 to 49% at age 15. Table 6 shows that boys in migrant households have lower attendance rates (71%) than boys in non-migrant households (75%). There is hardly any difference for girls. This is especially remarkable as mothers in migrants households have on average about 0.3 years more schooling than mothers in non-migrant households. It is important to note that these statistics are only based on children who stay in Mexico. We do not know the characteristics of the households that migrated as a whole.

In the sample we have no children who are categorized as current migrants. This is strong evidence in support of Assumption 2, that children would not migrate alone. [14]

[11] Mexico has 31 states and one federal district. For simplicity we will only refer to these entities as *states*.

[12] For a discussion on the advantages and disadvantages of this migrant definition, please refer to MR.

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

[14] There are 14 children who report prior migration episodes. Out of these 14 children, six come from non-migrant households. However, the questionnaire includes several questions on migration and the answers for these observations are not consistent. This indicates that data

4.2 Historical migration networks to instrument for selection of households into migration

To overcome the problem of self-selection into migration, a number of recent studies (e.g. McKenzie and Hildebrandt, 2005; McKenzie and Rapoport, 2011) have used historical state-level migration rates as an instrument for current migration levels. The argument for this instrument is that existing networks lower the migration cost for subsequent migrants and therefore trigger additional migration. The exclusion restriction is that these historical migration rates do not affect education outcomes today except through current migration of household members. A detailed discussion of this instrument and the exclusion restriction with respect to educational attainment can be found in MR. However, the two versions of the bounds in this paper require additional assumptions about the instrument. Assumption 5 states that the instrument must not influence the migration decision of the child directly. This seems to be a reasonable assumption if the migration network primarily helps the adult migrant to find a job in the destination country. However, if the network provides other help to the migrant as well (e.g. find housing, organize childcare,...) then the instrument might also directly influence the probability that the household migrates with the child. As an alternative, we derive bounds under Assumption 8 instead of Assumption 5. Assumption 8 allows for a direct effect of the instrument on the migration probability of the child but the effect is required to be monotone. If networks indeed lower the cost of migration in general, then assuming that the instrument can have a positive but no negative effect on the migration probability of the child seems to be plausible. However, using this weaker assumption comes at the cost of wider bounds.

As MR we use state-level migration rates to the US from 1924 taken from Woodruff and Zenteno (2007). We recode this continuous measure into a binary one, by defining states as low-migration states ($Z = 0$) if the migration rate was below the state-level median (3.78%) and as high-migration states ($Z = 1$) if migration rate was above (see Frölich 2007 for details on this transformation). We do this to allow stratification on instrument assignment, which would not be possible with a continuous instrument. Figure 3 in Appendix A shows the relation between historical migration rates and the probability of a child to live in a migrant household. One can clearly see a sharp increase in the probability around the median of state-level migration rates, making this cut-off a reason-

problems are the reason for this finding.

able one, as it maximizes the share of compliers among adults. In this setting compliers are those individuals, who would migrate only if they live in a high-migration state. Unlike MR we abstain from including additional covariates in our estimation to ensure instrument validity. Covariates would further complicate the analysis. The main problem is that we do not observe the distribution of covariates for those households, which migrated as a whole. Furthermore, two stage least squares point estimates with a binary instrument without covariates differ only slightly from those using covariates and are similar to the results in MR.[15]

4.3 All-move households in Mexico

While the ENADID dataset provides rich information on individual migration histories, it misses information on households which migrate as a whole. If no household member is left to answer the survey questions, not even the existence of the household is recorded. In order to get an understanding of how widespread the phenomenon of all-move households is in Mexico, we build on previous work that used data from the origin and the destination country of migrants. Ibarraran and Lubotsky (2007) estimate the size of the Mexican immigrant population in the United States a) based on the 2000 Mexican Census and b) based on the 2000 U.S. Census. As the Mexican Census is conducted as a household survey, it is very likely to miss migrants who migrated with their whole household. The estimate of the size of the Mexican born population living in the U.S. based on the Mexican Census is 1,221,598[16], while the estimate 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 by far lower for female migrants (33.6%) than for male migrants (69.9%).

The authors argue that this difference is primarily due to married couples who have migrated as a whole household and are therefore most likely not counted in the Mexican Census. Once married couples with both spouses present in the U.S. are removed from the U.S. Census estimates, the remaining migrant

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

[16] This number excludes migrants who have returned to Mexico.

number is 1,492,111 and thus by far 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. They find that 14.4% of male and 48% of female recent Mexican migrants are married with their spouse present in the U.S. and also conclude that these individuals are likely not covered in Mexico-based surveys.

The discrepancies between the numbers from the Mexican and the U.S. Census are even larger for children. In the age group 0-13 years, the number of migrants in the Mexican Census in only about 10% of the number of migrants in the U.S. Census. In the age group 12 to 15 years the ratio is about 50%. Overall the U.S. Census counts 82,240 Mexican born children in this age group.[17] Again, the reason is most likely that children migrate with their whole family and are therefore not counted in the Mexican Census any more.

The ratio of the total number of migrants and the number of migrants net of married couples with both spouses present in the U.S. may be seen as a rough estimate for the share of migrants missed due to all-move households. This ratio of 1.46 suggests that for every two migrants counted in a Mexican household survey, one additional migrant is missed. However, the relevant number for assessing the potential bias from sample-selection due to all-move households is the ratio of the number of children missed due to migration of whole households to the number of observed children in migrant households. To calculate this number we use information from two sources. First, we use the ENADID to calculate the total number of children in migrant households in Mexico. Using the definition of a migrant household as described above and the expansion factors provided with the data, we calculate the total number of children aged 12 to 15 years who live in a migrant household to be 1,516,924. Second, from the year 2000 U.S. census we know the number of Mexican born children in this age group to be 82,240. From these two numbers we can approximate γ to be 0.054. Of course this is only a very crude calculation. This ratio appears to be rather low but this is due to the fact that we use a rather broad definition of a migrant household and thus have a large denominator. We therefore test the sensitivity of the results to different ratios ranging from 0 to 0.5. For the main analysis we will use a ratio of $\gamma = 0.054$.

[17]Thanks to Darren Lubotsky for providing this estimate.

4.4 Results

We bound the effect of living in a migrant household on school attendance for children aged 12 to 15 years in Mexico. To be more precise, the effect of interest is for the group of children who would never migrate but who live in a household where adult migration is induced by the availability of historic migration networks. Using a parametric framework, MR found a significant negative effect for boys (-16 percentage points) and an insignificant effect for girls (-9 percentage points). However, MR did not address the sample selection due to all-move households. To get a comparison to these previous results, we ignore the sample selection and estimate the effects using a simple Wald estimator without covariates. The results are similar to the findings of MR. The estimated effect for boys is -19.5 and significant, the effect for girls is 8.1 and not significantly different from zero (Table 3).[18]

Next, we assess the sensitivity of these results to the sample selection induced by the migration of whole households. The first rows of Table 3 present the estimated strata proportions. The proportion of stratum CN for boys is 0.260 (s.e. 0.045), the proportion of stratum CC is 0.014 (s.e. 0.002). The biggest stratum with a proportion of 0.596 (s.e. 0.037) is stratum NN. Strata proportions are very similar for girls. The next three rows display the point identified expected 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 63.3% for boys and 60% for girls. School attendance is slightly higher for boys than for girls in all strata. The lower and upper bounds for school attendance rates for stratum CN under control are 87.1 and 100%. For girls they are substantially lower with 55.4% and 69.7% respectively.

The lower and upper bounds on the average effect for the CN stratum for boys are -26.3 and -14.4 percentage points. For girls the respective numbers are 0.2 and 14.5. However, the confidence intervals are rather wide for both groups. For boys the 95% confidence intervals includes zero by a small margin.

While it appears that the effect of living in a migrant household for boys is negative even if the sample selection is taken into account, the opposite is true for girls. The estimated bounds suggest that the effect might even be positive. This result is in line with the arguments and empirical findings of a series of recent papers. Antman (2012) suggests that paternal migration is associated

[18]While the estimates for boys with and without covariates are very similar to the results of MR, there is a bigger discrepancy in the estimates for girls. Not reported estimates using the continuous instrument are much closer to the result of MR.

with a shift in decision making power towards the mother and that mothers choose to spend more on the education of girls. Antman (2011) finds that in the short run boys have to respond to paternal absence with an increase in work and a decrease in study hours. Both channels may contribute to the fact that boys experience a negative effect on school attendance by migration of adult household members and no or even a positive effect exists for girls.

	Boys		Girls	
π_{AN}	0.123	(0.028)	0.121	(0.027)
π_{AA}	0.007	(0.002)	0.007	(0.001)
π_{CN}	0.260	(0.045)	0.250	(0.042)
π_{CC}	0.014	(0.002)	0.013	(0.002)
π_{NN}	0.596	(0.037)	0.609	(0.034)
$E\left[Y(0,0)\|G=NN\right]$	0.727***	(0.022)	0.699***	(0.024)
$E\left[Y(1,0)\|G=AN\right]$	0.678***	(0.020)	0.637***	(0.024)
$E\left[Y(1,0)\|G=CN\right]$	0.633***	(0.037)	0.600***	(0.041)
$\overline{Y}(y^{000}_{\alpha_{CN}} \leq Y \leq y^{000}_{1-\alpha_{CC}})$	1.000***	(0.016)	0.904***	(0.055)
$\overline{Y}(y^{000}_{\alpha_{CC}} \leq Y \leq y^{000}_{1-\alpha_{CN}})$	0.670***	(0.026)	0.516***	(0.039)
Bounds on $E\left[Y(0,0)\|G=CN\right]$	[0.871	1.000]	[0.554	0.697]
Wald estimate	-0.195**	(0.084)	0.081	(0.118)
Bounds on θ_{CN}	[-0.263	-0.144]	[0.002	0.145]
CLR 95% confidence interval	(-0.418	0.003)	(-0.210	0.358)
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. Clustered standard errors in parenthesis from 1999 bootstrap replications. Numbers in parentheses in the bottom row are 95% confidence intervals calculated by the procedure suggested by Chernozhukov, Lee, and Rosen (2012), while numbers in square brackets are identified sets determined by the half-median unbiased estimators. * denotes that estimate is statistically different from zero at the 10%, ** at 5%, and *** at 1% significance level.

Table 3: Table of results

The results for the bounds with monotonicity assumption A8 instead of exclusion restriction A5 can be found in Table 7 in Appendix A. The bounds on the effect for the CN stratum for boys are -28.6 and -12.6 percentage points and therefore only slightly wider than before. The corresponding bounds for girls are -3.2 and 16.7. The small difference in the widths of the bounds comes from the fact that $E[Y(1,0)|G = CN]$ is not point identified but also bounded.

4.5 Sensitivity with respect to γ

The ratio of the number of children not included due to migration of the whole household to the number of children observed in migrant households is a crucial parameter for the analysis. For the main results presented above we calculated this parameter to be 0.054. However, as there is substantial uncertainty involved how close this number is to the true value, we test the behavior of the bounds for different ratios. We repeat the analysis for γ between 0 (there are no children in all-move households) and 0.5 (for every two children observed in a migrant household we miss one child in an all-move household).

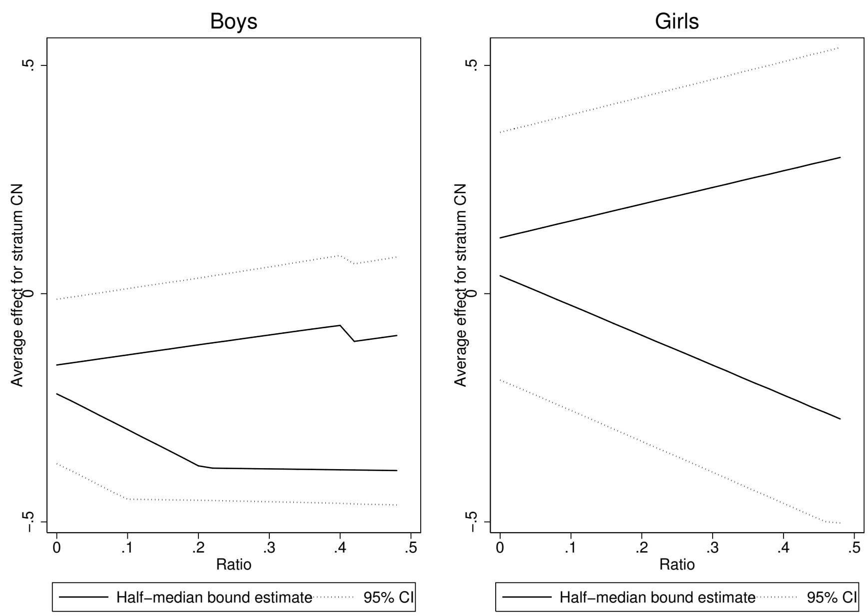

Figure 2: Sensitivity of bounds for different ratios unobserved to observed children in migrant households

Figure 2 displays the resulting bounds on the effects for boys and girls. What becomes apparent is that 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 ratio of about 0.22 and decreases only slightly thereafter. Up to a ratio of 0.22 the estimate for the upper bound on $E\left[Y(0,0)|G = CN\right]$ stems from the constrained solution. Once π_{CN} is sufficiently small so that the constraint is not binding any more, the estimate stems from the unconstrained solution. The slight decrease after this threshold results from the steady decline in π_{CN} and the fact that the expected value is computed in an ever decreasing fraction of the largest values of the outcome Y. For girls we only observe the constrained solution over the whole range of ratios. The kink in the upper bound for boys at a ratio of 0.4 stems from the precision adjustment.

One insight from Figure 2 is that the ratio of γ is not the only determinant of the behavior of the bounds. For example, at a ratio of 0.3 the width of the bounds is 0.29 for boys and 0.39 for girls, which is due to the different distributions of the outcome variable.

Overall, this sensitivity check points to the importance of the double selection problem. Zero is included in the confidence intervals for boys and girls already for rather small levels of γ. For boys the bounds do not include a zero effect up to a γ of 0.5. For girls the bounds contain a zero effect already for $\gamma = 0.055$. Ignoring the problem could therefore lead to false conclusions, even if the share of all-move households is small.

Figure 4 in the appendix compares the bounds under the different sets of assumptions for different values of γ. Overall the behavior of the different bounds is rather similar.

5 Conclusion

This paper examines the identification of the causal effects of migration on the remaining household members in the presence of *double-selection*. The first selection problem that complicates the empirical analysis is the non-random selection of households into migration. The second selection problem arises from the decision of households of whether to send only one or a subset of individuals or to migrate as a whole. Households that migrate as a whole are usually not included in cross-sectional household survey data at all. The second form of selection has largely been ignored in the previous literature, as it is not apparent immediately.

We use principal stratification to model the migration decision of the household members and to structure the identification problem if both selection issues are present. This allows us to derive bounds on the effect of migration on school attendance of children left behind under a transparent set of assumptions. Based on Mexican ENADID data we show that the effect of migration of an adult household member on the school attendance of boys is likely to be negative, even if we take the second form of selection into account. The direction of the effect is ambiguous for girls. However, sensitivity analysis with respect to the share of all-move households indicate that the results are sensible to the second selection problem. Ignoring it can lead to false conclusions even if the share of all-move households is small.

More general, this paper uses a novel approach to identify the effects of migration on remaining household members. While the current paper primarily focuses on the selection problem induced by all-move households, using partial identification could be a reasonable way to go in a research field that has struggled to come up with credible empirical strategies to point identify the effects of interest.

References

Alcaraz, C., D. Chiquiar, and A. Salcedo (2012): "Remittances, schooling, and child labor in Mexico," *Journal of Development Economics*, 97(1), 156–165.

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. (2013): "The Impact of Migration on Family Left Behind," *forthcoming in International Handbook on the Economics of Migration.*

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.

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*.

Chernozhukov, V., S. Lee, and A. M. Rosen (2012): "Intersection Bounds: Estimation and Inference," *forthcoming in Econometrica*.

Cox-Edwards, A., and M. Ureta (2003): "International migration, remittances, and schooling: evidence from El Salvador," *Journal of Development Economics*, 72(2), 429–461.

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.

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.

Giannelli, G. C., and L. Mangiavacchi (2010): "Children's Schooling and Parental Migration: Empirical Evidence on the 'Left-behind' Generation in Albania," *Labour*, 24(s1), 76–92.

Gibson, J., D. McKenzie, and S. Stillman (2010): "Accounting for selectivity and duration-dependent heterogeneity when estimating the impact of emigration on incomes and poverty in sending areas," *Policy Research Working Paper Series*, 5268.

——— (2011a): "The Impacts of International Migration on Remaining Household Members: Omnibus Results from a Migration Lottery Program," *The 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.

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," *mimeo.*

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

Huber, M., and G. Mellace (2013): "Sharp IV bounds on average treatment effects under endogeneity and noncompliance," *mimeo.*

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.

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

Kitagawa, T. (2009): "Identification region of the potential outcome distributions under instrument independence," *cemmap working paper CWP30/09*, pp. 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.

Lahaie, C., J. A. Hayes, T. M. Piper, and J. Heymann (2009): "Work and family divided across borders: the impact of parental migration on Mexican children in transnational families," *Community, Work & Family*, 12(3), 299–312.

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

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

Manski, C. F. (1989): "Anatomy of the Selection Problem," *The 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.

Mattei, A., and F. Mealli (2007): "Application of the principal stratification approach to the Faenza randomized experiment on breast self-examination.," *Biometrics*, 63(2), 437–46.

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

McKenzie, D., and H. Rapoport (2007): "Self-selection patterns in Mexico-U.S. migration: the role of migration networks," *The 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*.

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

Mergo, T. (2011): "The Effects of International Migration on Migrant-Source Households in Developing Countries: Evidence from Ethiopian Diversity-Visa (DV) Lottery Migrants to the US," *mimeo*.

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.

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

World Bank (2010): *Migration and Remittances Factbook 2011*. World Bank Publishers, Washington.

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_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	1	AC	A5
	1	1	D	0	0	AD	A5, A8
	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, A8
	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, A8
	0	1	N	0	0	DN	A7
N	0	0	A	1	1	NA	A2
	0	0	C	0	0	NC	A2, A5
	0	0	D	1	1	ND	A2, A5, A8
	0	0	N	0	0	NN	

Table 4: 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	CC, 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	AC, AN
$O(0,1,1) = \{Z = 0, M_1 = 1, M_2 = 1\}$		AA, AD, DA, DD	AA	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	AA, AC, 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. Column (3) shows the remaining strata if Assumption 5 is replaced with Assumption 8.

Table 5: Correspondence between observed groups and latent strata

	Boys		Girls	
Migrant household (0/1)	0.24		0.24	
	$M_1 = 1$	$M_1 = 0$	$M_1 = 1$	$M_1 = 0$
School attendance (0/1)	0.71	0.75	0.66	0.67
Age	13.44	13.43	13.51	13.45
Years of schooling of mother	3.97	3.64	3.93	3.68
State migration rate in 1924	0.10	0.04	0.09	0.04
Binary instrument (0/1)	0.72	0.36	0.71	0.36
Observations	1,954	6,039	1,810	5,853

Table 6: Descriptive statistics for non-migrant children

	Boys		Girls	
π_{AN}	0.120	(0.028)	0.118	(0.027)
π_{AC}	0.003	(0.001)	0.003	(0.001)
π_{AA}	0.003	(0.001)	0.003	(0.001)
π_{CN}	0.263	(0.044)	0.253	(0.042)
π_{CC}	0.014	(0.002)	0.014	(0.002)
π_{NN}	0.596	(0.037)	0.609	(0.034)
Bounds on θ_{CN}	[-0.286	-0.126]	[-0.032	0.167]
CLR 95% confidence interval	(-0.435	0.016)	(-0.232	0.370)
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 and $\pi_{AA} = \pi_{AC}$. Clustered standard errors in parenthesis from 1999 bootstrap replications. Numbers in parentheses in the bottom rows are 95% confidence intervals calculated by the procedure suggested by Chernozhukov, Lee, and Rosen (2012), while numbers in square brackets are identified sets determined by the half-median unbiased estimators. * denotes that estimate is statistically different from zero at the 10%, ** at 5%, and *** at 1% significance level.

Table 7: Table of results without exclusion restriction

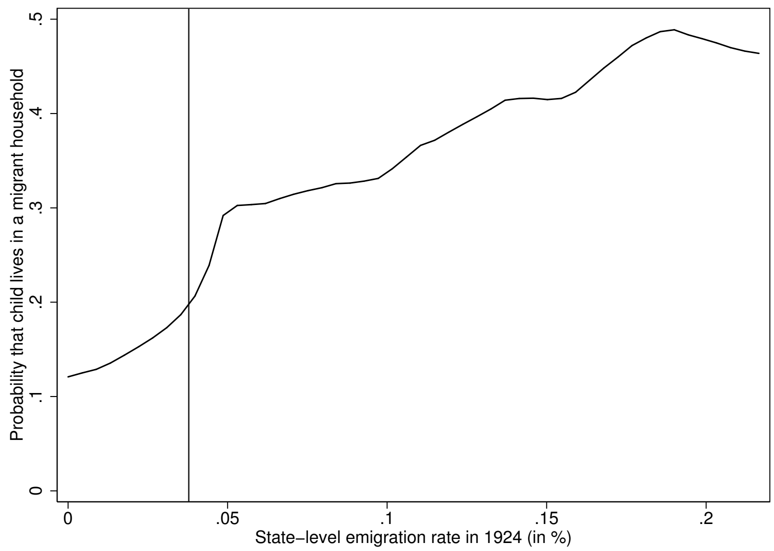

Figure 3: Cut-off for binary instrument

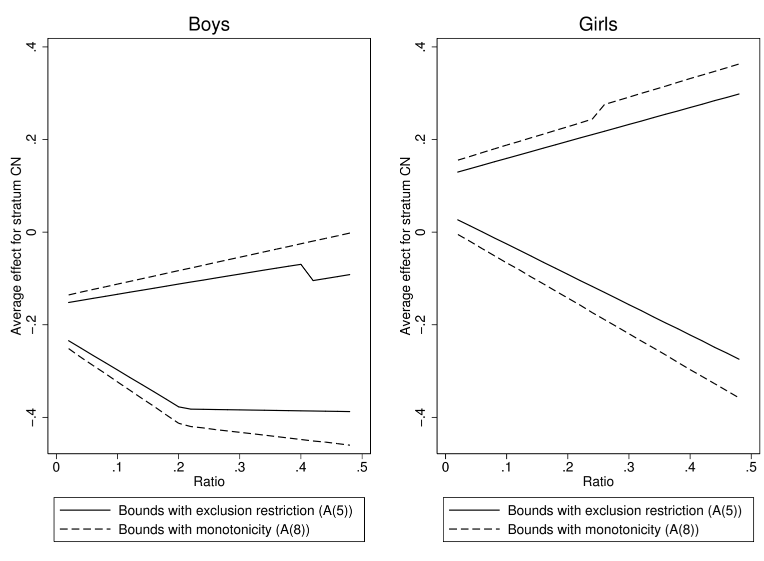

Figure 4: Comparison of bounds derived under different sets of assumptions over different ratios of unobserved to observed children in migrant households

B Technical Appendix

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

Group $O(0,0)$ is a mixture of compliers and never migrants. The observed outcome is therefore a mixture of the potential outcomes of these two strata under control

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

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

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

The upper bound for $E\left[Y(0,0)|G=C\right]$ can be obtained from taking the upper π_C quantiles in the observed group $O(0,0)$:

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

The respective lower bound can be obatined from taking the lower π_C quantiles. Thus the lower and upper bound for $E\left(Y(0,0)|G=N\right)$ can be rewritten as:

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

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

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

B.2 Strata proportions without exclusion restriction of Z on M_2

We continue to assume an overall ratio γ of observed children in migrant household to children not observed due to migration of the whole household. Furthermore we assume that this ratio is equal for $\pi_{CN} * \gamma = \pi_{CC}$ and $\pi_{AN} * \gamma = \pi_{AC} + \pi_{AA}$. In the absence of any reliable information on this strata proportions we assume $\pi_{AC} = \pi_{AA}$. Again the other strata proportions cannot just be estimated as conditional probabilities but need to be adjusted due to the fact that not the entire sample is observed. The adjustment factor in the sub-sample with $Z = 0$ is $\lambda_0 = N_0 / \left(N_0 + N_{010} * \left(\gamma / \left(2 + \gamma\right)\right)\right)$ and in the sub-sample with $Z = 1$ it is $\lambda_1 = N_1/(N_1 + N_{110} * \gamma)$. N_z denotes the number of observations with $Z = z$, N_{z10} the number of observations with $Z = z, M_1 = 0, M_2 = 0$. Given this information, the strata proportions are identified as

$$
\begin{aligned}
\pi_{AN} &= P\left(M_1 = 1, M_2 = 0 | Z = 0\right) * \lambda_0 * {}^{2}/_{(2+\gamma)} \\
\pi_{CN} &= P\left(M_1 = 1, M_2 = 0 | Z = 1\right) * \lambda_1 - \pi_{AN} \\
\pi_{NN} &= P\left(M_1 = 0, M_2 = 0 | Z = 1\right) * \lambda_1 \\
\pi_{CC} &= \pi_{CN} * \gamma \\
\pi_{AC} &= \pi_{AA} = \pi_{AN} * (\gamma/2)
\end{aligned}
$$

B.3 Inference based on Chernozhukov, Lee, and Rosen (2009)

We will explain the estimation procedure for $E^L_{CN}\left[Y(0,0)|G = CN\right]$. 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 $\Delta^L(0) = \overline{Y}(Y \leq y^{000}_{\alpha_{CN}})$ and $\Delta^L(1) = \overline{Y}(Y \leq y^{000}_{1-\alpha_{CC}}) * \frac{\pi_{NN}+\pi_{CN}}{\pi_{CN}} - \overline{Y}^{100} * \frac{\pi_{NN}}{\pi_{CN}}$. Let $\mathbf{\Delta}^L = [\Delta^L(0)\Delta^L(1)]'$ be the vector containing the two bounding functions. We subsequently discuss the estimation of the lower bound along with its confidence region (the proceeding for the upper bound is analogous). We 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 and Mellace (2013). 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.[19]
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[20] 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 $\hat{\mathcal{V}} = \{v \in \mathcal{V} : \hat{\Delta}_{01}^{UB}(v') \leq \max_{v' \in V}\{[\hat{\Delta}^L(v') + \tilde{k}(c) \cdot \hat{s}(v')] + 2 \cdot \tilde{k}(c) \cdot \hat{s}(v')\}\}$.
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 $\tau = (\rho \cdot \log(n))^{-1}$. Compute $\hat{a} = 1 - \Phi(\tau \cdot \Gamma) \cdot \alpha$, where α is the chosen confidence level.

[19] In the empirical part we use 1'999 bootstrap replications.
[20] We set R=1'000'000.

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