

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

Andreas Steinmayr[†]

August 6, 2020

Abstract

A key problem in the literature on the economics of migration is how emigration of an individual affects households left behind. Answers to this question must confront a problem I refer to as *invisible sample selection*: when entire households migrate, no information about them remains in their source country. This problem occurs in addition to the question of selection of households into migration, i.e., which households send any migrants. This paper uses principal stratification to address both identification problems simultaneously for a setup with an instrumental variable for the migration of the principal migrant. I derive nonparametric bounds on the effect of migration on remaining household members under varying assumptions and discuss the assumptions implicitly made when the problem is ignored. Building on two widely cited studies, I show that transparent assumptions regarding the behavior of household members and selectivity of migrants allow identification of informative bounds but also that not adjusting for invisible selection may result in substantial bias.

Keywords: Sample selection, migration, selectivity, principal stratification

JEL classification: C21, F22, J61, O15

*I thank Dan Black, Stéphane Bonhomme, Xavier D'Haultfoeulle, Alfonso Flores-Lagunes, Jeffrey Grogger, Martin Huber, Michael Lechner, Charles Manski, Toru Kitagawa, David McKenzie, Giovanni Mellace, Elisabeth Sadoulet, and Steven Stillman for fruitful discussions. Christina Felfe, Rebecca Lordan, Damon Jones, Elie Murard, Toman Barsbai, Anthony Strittmatter, Petra Thiemann, and Conny Wunsch provided helpful comments on the draft. I also thank participants at various conferences and seminars for helpful comments. Emre Oral provided excellent research assistance. All errors are my own.

[†]Department of Economics, University of Munich (LMU), CESifo, Immigration Policy Lab, and IZA. Ludwigstr. 33, 81543 Munich, andreas.steinmayr@econ.lmu.de. This study was generously funded by the Swiss National Science Foundation (P2SGP1_148963). An earlier version of the paper was titled “When a Random Sample is Not Random. Bounds on the Effect of Migration on Children Left Behind” .

1 Introduction

Migration often splits households, with some household members migrating and remitting money to members left behind in the source country. Although income effects are widely accepted to be positive (see for example McKenzie, Gibson, and Stillman, 2010), the overall effect of migration on household welfare is ambiguous since absence of household members might affect household welfare negatively. A burgeoning literature investigates effects of migration on various aspects of welfare of household members left behind, such as children’s health and educational attainment, labor supply of spouses, and household poverty.¹

Identification of the effects of migration on remaining household members faces two major selection problems. First, the non-random selection of households into migration, i.e., which households send migrants. Various strategies have been implemented to address this endogeneity, including selection on observables (e.g., Kuhn, Everett, and Silvey, 2011), instrumental variables (e.g., Hanson and Woodruff, 2003; McKenzie and Hildebrandt, 2005; McKenzie and Rapoport, 2011), or fixed-effects approaches (e.g., Antman, 2012). Second, selection within households, i.e., which members of the household migrate, poses another selection problem that has almost entirely been ignored (see Gibson, McKenzie, and Stillman, 2011a, 2013, for a related discussion). Selection within the household gives rise to two problems. To start with, outcomes for household members who migrate are not observed or not well defined. Moreover, household members who stay behind are a selected group that might differ in its characteristics from migrating household members. The identification problem is further complicated if all household members migrate (*all-move households*). In that case, the household will not be included in data collected in the source country, i.e., the household is not included in household surveys since no household member is left to respond to a survey. I refer to this issue as *invisible sample selection*.

The following stylized research design illustrates the complications these selection problems create for identification. For simplicity, consider one-adult-one-child households. The adult participates in a visa lottery, and migrates if she wins and stays if she does not. Due to the random assignment of the visa, selection of households into migration does not pose a problem for identification. The decision to take the child along, however, is endogenous and depends on household characteristics. Assume for example that only wealthier households can take the child, and poorer households leave the child behind. Consider further that outcome data are collected by a household survey in the source country after the lottery occurs. Children left behind in the *observed* sample of migrant households live, on average, in initially poorer households than children in non-migration households. This negative correlation biases estimates obtained by comparing outcomes of children left behind in migrant and non-migrant households even though adult migration is

¹Antman (2013) provides a comprehensive overview of the literature on the effects of migration on remaining household members and the empirical strategies employed in this literature.

randomly assigned. Wealthier households disappear from the sample of lottery winners since no individual remains to respond to the survey. Sample selection becomes invisible since, in the absence of data on the initial population, no information is recorded on these all-move households.

One reason invisible sample selection has not received much attention might be that the problem that arises for identification of causal effects is not obvious since sample selection in cross-sectional data is *invisible*. It is not apparent that households that leave no members behind are relevant for identification of causal effects of migration on *remaining* household members. However, as the above example illustrates, selection within households and related migration of whole households constitute sample selection problems that can lead to inconsistent estimates.²

In this paper, I address the identification problems induced by selection between and within households and, relatedly, invisible sample selection. The paper extends the literature on sample selection by studying a situation in which for some units, not only is the outcome unobserved, but the units are not included in cross-sectional data at all. I apply *principal stratification* (Frangakis and Rubin, 2002) to model the behavior of household members and selectivity of migrants. Principal stratification allows characterizing the potential – not the observed – behavior of household members, which in turn allows transparency regarding assumptions needed for point and partial identification of causal effects. I do so for a setting with a binary instrument for the migration decision of the principal migrant to address the between-household selection problem.³ I derive bounds on the effect of migration for the stratum of never migrating remaining household members, where the principal migrant complies with the instrument. In my two applications, this stratum accounts for 35% and 25% of the underlying populations respectively. The paper further shows that in the presence of invisible sample selection, instrumental variable estimates are inconsistent even if selection of the accompanying migrant is unrelated to potential outcomes. For this scenario, I propose an alternative strategy to point identify the effect for the stratum mentioned before.

In the empirical part of the paper, I show that results from previous studies that ignore invisible sample selection might suffer from substantial bias. In a first application, I replicate results of a study with well-identified point estimates of migration effects on household composition and household assets (Gibson, McKenzie, and Stillman, 2011a). I compare these results to the bounds derived in my paper, which identify effects for a broader population and can be applied if non-migrants cannot be identified based on observ-

²The problem of whole households moving and not being included in source country data has been acknowledged in studies that estimate the overall number of emigrants (Ibarraran and Lubotsky, 2007) or migrant selectivity (McKenzie and Rapoport, 2007) based on source country data.

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

³Continuous instruments can be discretized to apply the approach outlined in this paper. See Frölich (2007) for further discussion on this topic. The approach does not apply to situations where an instrument is valid only conditional on a set of covariates.

able characteristics. In a second application, I revisit the effect of migration on the educational attainment of children left behind in Mexico (McKenzie and Rapoport, 2011). Unlike the original study, I consider that observational data miss migrating children in all-move households, and derive bounds under varying assumptions. The bounds indicate that point estimates ignoring the sample selection are likely lower bounds of the true effects. Overall, the empirical examples show that a combination of behavioral and distributional assumptions provides substantial identifying power to derive informative bounds.

This paper contributes to two strands of literatures. First, the literature on the identification of treatment effects in the presence of multiple identification problems and the related literature on mediation analysis. Several recent papers use principal stratification to derive bounds on the effects of policy interventions in the presence of post-treatment complications.⁴ To derive bounds in a setting with non-compliance and sample selection, I use an approach from Chen and Flores (2015). What sets the current paper apart from existing literature on sample selection is that it allows for units to be completely unobserved while the existing literature treats the outcome as unobserved while the unit itself is observed. Mediation analysis decomposes individual causal mechanisms from overall causal effects.⁵ In the previous example, migration of the child does not only create a sample selection problem but is a treatment in itself. Migration of the adult can be seen as the main treatment and migration of the child as a mediator, a channel through which adult migration affects child outcomes. Identifying the effect of adult migration with the child staying behind corresponds to identifying the direct effect of adult migration and ruling out indirect effects through child migration.

Second, it contributes to an applied literature on the effects of migration on remaining households. This literature has focused almost exclusively on selection between households. Recently, use of (quasi-) experiments for research on migration has been encouraged strongly (McKenzie and Yang, 2012; McKenzie, 2015). However, as randomization usually addresses only the first source of endogeneity (which households engage in migration), the second form (who and how many household members migrate) is a problem in experimental settings as well. The solution of the few papers that use visa lotteries to account for the first form of endogeneity and address the second form of endogeneity has been to define a different parameter of interest and estimate the effect for only those households and household members that can be identified as *never migrants* based on observable characteristics. Gibson, McKenzie, and Stillman (2011a) and Gibson, McKenzie, and Stillman (2013) use visa rules that dictate which household members are allowed to migrate

⁴Recent papers include Zhang and Rubin 2003; Mattei and Mealli 2007; Zhang, Rubin, and Mealli 2008; Chen and Flores 2015; Blanco, Flores, and Flores-Lagunes 2013a,b; Huber, Laffers, and Mellace 2017.

⁵For rather general mediation models, see for example Pearl (2001) and Albert and Nelson (2011). Recent studies in economics mostly evaluate mechanisms through which active labor market policies work (Flores and Flores-Lagunes, 2009, 2010; Huber, 2014; Huber, Lechner, and Mellace, 2017). Heckman, Pinto, and Savelyev (2013) investigate mechanisms through which the Perry Preschool program affects outcomes later in life.

with the principal migrant. In their setting of migration from Tonga and Samoa to New Zealand, they removed all eligible individuals from the estimation sample. They therefore estimate the effect for individuals who are ineligible to join the principal migrant and are therefore always observed. This subgroup consists primarily of siblings, nephews, nieces, and parents of the migrant. Estimating the effect of migration on the migrant's nuclear family is not possible using their approach. In studies based on observational panel data, several papers recognize the second form of endogeneity and provide some discussion on how severe the problem could be, but do not address it (Yang, 2008; Antman, 2011).

The remainder of the paper is structured as follows. Section 2 introduces an econometric framework and the identification results. Section 3 illustrates the approach for the effects of migration on households in Tonga and the effect of adult migration on school attendance of children in Mexico. Section 4 discusses possible extensions and variations of this approach. Section 5 concludes.

2 Econometric framework

2.1 Setup and parameter of interest

Following treatment evaluation literature, I use a potential outcome framework developed by Rubin (1974). Suppose that households consist of two individuals (I_1, I_2). With reference to the empirical applications, I refer to these individuals as principal migrant/adult (I_1) and accompanying migrant/child (I_2). In empirical applications, I_1 and I_2 may refer to different entities, not only individuals. For example, in the Tongan application, I_1 refers to the individual, who applied for a visa, and I_2 refers to the household as a whole. The unit of analysis is the household. However, in the Mexican application, I_1 refers to any adult in the household, and I_2 refers to an individual child. The unit of analysis is the child. The setup can be applied to different forms of intra-household selection, including a scenario in which the household dissolves as all members migrate.

$M_j = m_j \in \{0, 1\}$ denotes the migration status of individual j . I_1 makes the first migration decision and chooses either to stay ($M_1 = 0$) or migrate ($M_1 = 1$). I_2 chooses to stay ($M_2 = 0$) or migrate ($M_2 = 1$) depending on the choice of I_1 . $M_2(m_1)$ denotes the potential migration state of I_2 as a function of migration of I_1 . This does not necessarily have to be a sequential decision process nor the decision of I_2 , but can also be a household decision. The key feature here is that M_2 is a function of M_1 .

If M_1 is considered the treatment of interest, M_2 constitutes an *endogenous sample-selection problem* (Gronau, 1974; Heckman, 1974); those for whom the outcome is observed, i.e., stayers, are a selected group and the selection is a function of treatment.

The outcome Y is observed at some point after M_1 and M_2 have been realized. Y is a function of both, M_1 and M_2 . Thus, M_2 does not only create a sample selection problem but is a treatment in itself. $Y(m_1, m_2)$ denotes the potential values of the outcome.

I assume to have a *random* sample of households from the population in the source country, drawn after households were treated, i.e., individuals migrated. The sample does not include households in which all members migrate ($M_1 = 1, M_2 = 1$). Although the sample is representative of the population at that point, the observed population is different from the population at the time the treatment was assigned. If only a subset of household members migrate, the household is still observed in the data but the selection problem of which household members join the principal migrant and which stay behind, prevails.

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

Many empirical studies use an instrument for the migration decision of the principal migrant.⁶ Therefore, I study identification with an instrumental variable in more detail.⁷ Assume a binary instrument $Z = z \in \{0, 1\}$ exists, which affects the migration decision of the principal migrant. $M_1(z)$ denotes the potential migration of I_1 as a function of the value of instrument Z . Let us for the moment also write the potential values of migration of the child $M_2(m_1, z)$ and the outcome $Y(m_1, m_2, z)$ as functions of Z . In the presence of the second selection problem, we must modify the classical IV assumptions (Imbens and Angrist, 1994; Angrist, Imbens, and Rubin, 1996). Assumption 1 states that the instrument is assigned randomly and therefore independent of all potential outcomes.

Assumption 1. *Randomly assigned instrument*

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

Assumption 2 states that the effect of Z on the potential outcomes Y must be through an effect of Z on M_1 and M_2 .

Assumption 2. *Exclusion restriction of Z with respect to Y*

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

⁶See for example Hanson and Woodruff (2003); McKenzie and Hildebrandt (2005); Yang (2008); Amuedo-Dorantes, Georges, and Pozo (2010); McKenzie and Rapoport (2011); Antman (2011); Gibson, McKenzie, and Stillman (2011b).

⁷Recent papers that study identification with non-compliance and sample selection are Imai (2007), Frumento, Mealli, Pacini, and Rubin (2012), and Chen, Flores, and Flores-Lagunes (2018).

Assumption 3 states that the effect of the instrument on the potential migration status of I_2 must be through an effect of Z on M_1 . Assumptions 2 and 3 allow us to use the previous notation of potential outcomes and write the potential variables $M_2(m_1)$ and $Y(m_1, m_2)$ as a function of migration status only.

Assumption 3. *Exclusion restriction of Z with respect to M_2*

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

Assumption 4 states that the instrument has a non-zero average effect on the migration of I_1 . For the moment, I do not assume anything about the direction of the effect.

Assumption 4. *Non-zero average effect of Z on M_1*

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

A valid instrument must satisfy Assumptions 1, 2, 3, and 4 simultaneously (Imbens and Angrist, 1994; Angrist, Imbens, and Rubin, 1996). An important difference regarding the exclusion restriction is that I require Z to be a valid instrument for Y and M_2 (similar to Imai, 2007 and Chen and Flores, 2015). There are two differences to the settings in these two related papers. First, in my setting, M_2 is both an indicator of whether the individual is observed and a treatment in itself. In Imai (2007) and Chen and Flores (2015), the outcome is not a function of the selection indicator. Second, in my setting, the probability of observing Y decreases with M_1 since this increases the probability that $M_2 = 1$, i.e, the entire household migrates. In the studies mentioned above, the probability of observing the outcome increases for treated individuals. However, the latter fact only changes the interpretation of the latent strata but not the identification itself.

I distinguish principal strata with respect to the *instrument*. We can differentiate the types of principal migrants regarding the instrument as always migrants (A), compliers (C), defiers (D), and never migrants (N). An always migrant would migrate irrespective of the value of the instrument. A complier would migrate if the instrument takes a value of one but not if it takes zero. A defier would migrate if the instrument is zero but not if the instrument is one. A never migrant would not migrate irrespective of the value of the instrument. We can also distinguish these four types of accompanying migrants. I define the types of accompanying migrants also with respect to the instrument, even though I assume that the effect works only indirectly through M_1 (Assumption 3). Combining the four strata of principal migrants with the four strata of accompanying migrants gives in total $4 \times 4 = 16$ principal strata (latent household types) (see Table 1). I refer to the strata using a two-letter system; the first letter refers to the type of I_1 , the second to the type

Table 1: Principal strata and exclusion criteria

Type I_1	$M_1(1)$	$M_1(0)$	Type I_2	$M_2(1)$	$M_2(0)$	Household type	Exclusion criterion
A	1	1	A	1	1	AA	
	1	1	C	1	0	AC	A3
	1	1	D	0	1	AD	A3
	1	1	N	0	0	AN	
C	1	0	A	1	1	CA	A6
	1	0	C	1	0	CC	
	1	0	D	0	1	CD	A6
	1	0	N	0	0	CN	
D	0	1	A	1	1	DA	A5, A6
	0	1	C	0	1	DC	A5, A6
	0	1	D	1	0	DD	A5
	0	1	N	0	0	DN	A5
N	0	0	A	1	1	NA	A6
	0	0	C	1	0	NC	A3, A6
	0	0	D	0	1	ND	A3, A6
	0	0	N	0	0	NN	

of I_2 (e.g., CN refers to a household in which the principal migrant is a complier and the accompanying migrant is a never migrant). Let π_s denote the proportion of stratum s in the population.

Assumption 3 disqualifies the existence of strata AC , AD , NC , ND . In these strata, the instrument has a direct effect on M_2 since I_1 does not react to the instrument in these strata. This has the empirically testable assumption that no households with $M_1 = 0$ and $M_2 = 1$ should be observed. I assume a monotone effect of the instrument on migration of I_1 , which is a standard assumption in the instrumental variables literature (Imbens and Angrist, 1994; Angrist, Imbens, and Rubin, 1996). This assumption suggests that every principal migrant is at least as likely to migrate if $Z = 1$ as he would be if $Z = 0$.

Assumption 5. *Individual-level monotonicity of M_1 in Z*

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

Assumption 5 disqualifies defiers among principal migrants and therefore eliminates strata DA , DC , DD , DN . Assumption 6 states that I_2 only migrates if I_1 migrates. This assumption disqualifies the existence of strata CA , CD , DA , DC , NA , NC , ND .⁸ Assumptions 3, 5, and 6 combined disqualify the existence of 11 of the 16 principal strata (last column, Table 1).

Assumption 6. *I_2 only migrates if I_1 migrates*

$$M_2(0) = 0$$

⁸The existence of some strata is disqualified by more than one assumption.

In this setting, we can distinguish several effects of potential interest. The difference $Y(1, 0) - Y(0, 0)$ is the partial effect of M_1 on Y for M_2 being zero. I focus on $Y(1, 0) - Y(0, 0)$ since this effect is most policy relevant and has received the most attention in the literature.⁹ I further concentrate on the effect for stratum CN . In this stratum, M_1 is induced to change from 0 to 1 by the instrument, and M_2 is always zero. In countries with predominantly labor migration, in which only a small fraction of households migrates with the children, it is also quantitatively the most important group. This is also the only stratum for which outcomes are observed for I_2 in non-migrant and migrant households. The effect for this stratum can be identified without making assumptions about unobserved outcomes. The average partial effect of M_1 on individuals of type I_2 who would never migrate is defined as

$$\theta_{CN} \equiv E[(Y(1, 0) - Y(0, 0)) | M_2(0) = 0, M_2(1) = 0] = E[(Y(1, 0) - Y(0, 0)) | CN] \quad (1)$$

2.2 Correspondence between observed and latent household types

Table 2 shows the correspondence between observed groups and latent strata. Column (1) presents the corresponding strata without Assumptions 3, 5, and 6, Column (2) the remaining strata if these assumptions are imposed. Evident from Column (2), all-move households, where $M_1 = 1$ and $M_2 = 1$, could belong either to stratum AA or CC under Assumptions 1-6. If information about migration of complete households is available, all strata proportions π_S are identified (see Appendix A.1).

If this information is not available, external information on the number of all-move households can be used. Denote $N_{zm_1m_2}$ the number of observations in each cell. I define $\gamma \equiv \frac{N_{011} + N_{111}}{N_{010} + N_{110}}$, the ratio of unobserved all-move households to the observed number of migrant households.¹⁰ However, to point identify strata proportions, we need further assumptions about the existence of strata AA and CC . Information on the institutional setting could help to disqualify the existence of one of these strata. I discuss identification for the two polar scenarios. First, I rule out the existence of stratum CC ($\pi_{CC} = 0$), so all all-move households belong to stratum AA . Second, I rule out the existence of stratum AA ($\pi_{AA} = 0$), so all all-move households belong to stratum CC .

In the first case ($\pi_{CC} = 0$), the process that motivates migration of whole households is independent of the instrument. The observed samples with $Z = 0$ and $Z = 1$ both contain only households of type CN , NN , AN (Column (2), Table 2). Sample selection is not a problem, and the conventional Wald estimator yields consistent estimates of θ_{CN} (see calculations in Appendix A.2).

The second case, $\pi_{AA} = 0$, is more problematic. In this scenario, the observed group $O(0, 0, 0)$ contains households of type CC , which are not observed in the sample with $Z = 1$. To be conservative, I derive

⁹I discuss several other effects in Section 4.

¹⁰See Section 3.2.1 for an explanation of how I calculate γ using information from other data sources.

Table 2: Correspondence between observed groups and latent strata

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

Note: Column (1) shows the principal strata without assumptions. Column (2) shows the remaining strata after Assumptions 6, 3, and 5 have been imposed and all all-move households are of type *AA*. Column (3) shows the remaining strata under the same assumptions if all all-move households are of type *CC* (Assumption 7).

the bounds under the assumption that no households of type *AA* exist and all-move households are thus all of type *CC* (Assumption 7). See Huber and Mellace (2015) for a related application of this worst-case approach.

Assumption 7. *No households of type *AA* exist*

$$\pi_{AA} = 0$$

In a first step, I calculate the number of households missed, N_{111} , based on external information or assumptions about γ . If $\pi_{AA} = 0$, all all-move households are in group $O(1, 1, 1)$ (Column (3), Table 2). The number of missed observations in group $O(1, 1, 1)$ is $N_{111} = \gamma * (N_{010} + N_{110})$. With this information, all strata proportions are identified under Assumptions 1-7:

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

2.3 Bounds under Assumptions 1-7

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

in the observed group $O(1, 1, 0)$.

$$\bar{Y}^{110} = \frac{E[Y(1, 0)|CN] \pi_{CN} + E[Y(1, 0)|AN] \pi_{AN}}{\pi_{CN} + \pi_{AN}}. \quad (2)$$

This can be reformulated to

$$E[Y(1, 0)|CN] = \frac{\bar{Y}^{110}(\pi_{CN} + \pi_{AN}) - E[Y(1, 0)|AN] \pi_{AN}}{\pi_{CN}}. \quad (3)$$

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

$$E[Y(1, 0)|AN] = \bar{Y}^{010}. \quad (4)$$

Using Equations (3) and (4), the $E[Y(1, 0)|CN]$ is identified as

$$E[Y(1, 0)|CN] = \frac{\bar{Y}^{110}(\pi_{CN} + \pi_{AN}) - \bar{Y}^{010} \pi_{AN}}{\pi_{CN}}. \quad (5)$$

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

$$\bar{Y}^{000} = \frac{E[Y(0, 0)|CN] \pi_{CN} + E[Y(0, 0)|NN] \pi_{NN} + E[Y(0, 0)|CC] \pi_{CC}}{\pi_{CN} + \pi_{NN} + \pi_{CC}}. \quad (6)$$

$E[Y(0, 0)|NN]$ is identified since group $O(1, 0, 0)$ corresponds directly to stratum NN . However, the two conditional means of strata CN and CC are not. I derive bounds on $E[Y(0, 0)|CN]$ following the procedure proposed by Chen and Flores (2015). For simplification, I introduce additional notation. Let y_a^{000} be the a -th quantile of Y in the observed group $\{Z = 0, M_1 = 0, M_2 = 0\}$, and let the mean outcome in this cell for those outcomes between the a' -th and a -th quantiles of Y be

$$\bar{Y}(y_{a'}^{000} \leq Y \leq y_a^{000}) \equiv E[Y|Z = 0, M_1 = 0, M_2 = 0, y_{a'}^{000} \leq Y \leq y_a^{000}] \quad (7)$$

Let α_S be the share of stratum S in group $O(0, 0, 0)$, which is $\alpha_S = \pi_S/(\pi_{CN} + \pi_{NN} + \pi_{CC})$. The idea behind these bounds is to find the lowest and highest possible values of $E[Y(0, 0)|CN]$, subject to the constraint $\bar{Y}^{100} = E[Y(0, 0)|NN]$. In the unconstrained case, we can bound $E[Y(0, 0)|CN]$ from below by the expected value of Y for the α_{CN} fraction of smallest values of Y and from above by the expected value of Y for the α_{CN} fraction of largest values of Y in group $O(0, 0, 0)$.

Under the assumption that the smallest values in group $O(0, 0, 0)$ are only from CN observations, the lower bound for $E[Y(0, 0)|NN]$ is given by $\bar{Y}(y_{\alpha_{CN}}^{000} \leq Y \leq y_{1-\alpha_{CC}}^{000})$, the mean estimated in the central area

in Figure 1. In case this estimated lower bound is lower than $\bar{Y}^{100}$, the unconstrained solution is identical to the solution of the constrained problem (upper line in Equation (10)).

[Figure 1 about here]

If the constraint is not satisfied, we can derive a lower bound from the mixture distribution of CN and NN in the lower $1 - \alpha_{CC}$ quantiles of the distribution of Y in cell $O(0, 0, 0)$ by assuming all CC observations are at the top of the distribution, and the remaining lower part is a mixture of CN and NN (lower line in Equation (10)). The upper bound can be derived similarly (Equation (11)).¹¹

Proposition 1. *If Assumptions 1-7 hold, then $\theta_{CN}^L \leq \theta_{CN} \leq \theta_{CN}^U$, where*

$$\theta_{CN}^U = E[Y(1, 0)|CN] - E_{CN}^L[Y(0, 0)|CN] \quad (8)$$

$$\theta_{CN}^L = E[Y(1, 0)|CN] - E_{CN}^U[Y(0, 0)|CN] \quad (9)$$

$$E_{CN}^L[Y(0, 0)|CN] = \begin{cases} \bar{Y}(Y \leq y_{\alpha_{CN}}^{000}), \text{ if } \bar{Y}(y_{\alpha_{CN}}^{000} \leq Y \leq y_{1-\alpha_{CC}}^{000}) \leq \bar{Y}^{100} \\ \bar{Y}(Y \leq y_{1-\alpha_{CC}}^{000}) * \frac{\pi_{NN} + \pi_{CN}}{\pi_{CN}} - \bar{Y}^{100} * \frac{\pi_{NN}}{\pi_{CN}}, \text{ otherwise} \end{cases} \quad (10)$$

$$E_{CN}^U[Y(0, 0)|CN] = \begin{cases} \bar{Y}(Y \geq y_{1-\alpha_{CN}}^{000}), \text{ if } \bar{Y}(y_{\alpha_{CC}}^{000} \leq Y \leq y_{1-\alpha_{CN}}^{000}) \geq \bar{Y}^{100} \\ \bar{Y}(Y \geq y_{\alpha_{CC}}^{000}) * \frac{\pi_{NN} + \pi_{CN}}{\pi_{CN}} - \bar{Y}^{100} * \frac{\pi_{NN}}{\pi_{CN}}, \text{ otherwise} \end{cases} \quad (11)$$

The proof is analogous to the one in Chen and Flores (2015).

If the group of all-move households is a mixture of strata AA and CC , then, for a given γ , this lead to tighter bounds on $E[Y(0, 0)|CN]$ and, subsequently, θ_{CN}^U in comparison to a situation in which they are all of type CC . This is because for a given γ , the share of CC households, α_{CC} , in the observed cell $O(0, 0, 0)$ decreases. The constraint for deriving $E_{CN}^L[Y(0, 0)|CN]$ depends on the value of $y_{1-\alpha_{CC}}^{000}$, which is weakly larger if α_{CC} is smaller. The expression in the constrained case also becomes weakly larger since $\bar{Y}(Y \leq y_{1-\alpha_{CC}}^{000})$ is weakly larger if α_{CC} is smaller. If π_{CC} , and as a consequence α_{CC} , become zero, we are back to the point-identified case.

¹¹Alternative formulations for Equation (10) and (11) are given by

$$E_{CN}^L[Y(0, 0)|CN] = \max \left\{ \bar{Y}(Y \leq y_{\alpha_{CN}}^{000}), \bar{Y}(Y \leq y_{1-\alpha_{CC}}^{000}) * \frac{\pi_{NN} + \pi_{CN}}{\pi_{CN}} - \bar{Y}^{100} * \frac{\pi_{NN}}{\pi_{CN}} \right\}$$

$$E_{CN}^U[Y(0, 0)|CN] = \min \left\{ \bar{Y}(Y \geq y_{1-\alpha_{CN}}^{000}), \bar{Y}(Y \geq y_{\alpha_{CC}}^{000}) * \frac{\pi_{NN} + \pi_{CN}}{\pi_{CN}} - \bar{Y}^{100} * \frac{\pi_{NN}}{\pi_{CN}} \right\}.$$

2.4 Distributional assumptions to tighten the bounds

In addition to the behavioral assumptions presented above, distributional assumptions can further tighten the bounds (Chen and Flores, 2015). These distributional assumptions are specific to the setting under study and might vary by outcome variable. Such assumptions can be derived from theoretical arguments about migrant selectivity. For example, a standard assumption is that migration is costly and that costs increase with the number of migrating individuals. Assume that being separated generates disutility for households (Agesa and Kim, 2001). In such a scenario, households that can afford taking their children with them will be selected positively. Better-off households are also likely to invest more in the education of children (Leibowitz, 1974; Blau, 1999; Case, Lubotsky, and Paxson, 2002; Currie, 2009; Almond and Currie, 2011). Therefore, children in CC households will, on average, have more favorable outcomes (e.g. greater school attendance) than children in CN households, as formalized in Assumption 8.

Assumption 8. *Mean dominance 1*

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

Assumption 8 tightens the bound on $E[Y(0, 0)|CN]$. To see this, write

$$\bar{Y}^{000} = \frac{E[Y(0, 0)|CN, CC] (\pi_{CN} + \pi_{CC})}{\pi_{CC} + \pi_{CN} + \pi_{NN}} + \frac{E[Y(0, 0)|NN] \pi_{NN}}{\pi_{CC} + \pi_{CN} + \pi_{NN}} \quad (12)$$

where $E[Y(0, 0)|NN]$ is identified. Assumption 8 implies that $E[Y(0, 0)|CN] \leq E[Y(0, 0)|CN, CC]$.

Proposition 2. *If assumptions 1-7 and 8 hold, $E[Y(0, 0)|CN, CC]$ provides an upper bound on $E[Y(0, 0)|CN]$ that is lower or equal as the one in Equation (11):*

$$E_{CN}^{U, MD}[Y(0, 0)|CN] = \frac{\bar{Y}^{000}(\pi_{NN} + \pi_{CN} + \pi_{CC}) - \bar{Y}^{100}\pi_{NN}}{\pi_{CN} + \pi_{CC}} \quad (13)$$

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

Assumption 9. *Mean dominance 2*

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

Proposition 3. *If assumptions 1-7 and 9 hold, $E[Y(0,0)|CN, CC]$ provides a lower bound on $E[Y(0,0)|CN]$:*

$$E_{CN}^{L,MD}[Y(0,0)|G=CN] = \frac{\bar{Y}^{000}(\pi_{NN} + \pi_{CN} + \pi_{CC}) - \bar{Y}^{100}\pi_{NN}}{\pi_{CN} + \pi_{CC}} \quad (14)$$

Bounds on causal effect θ_{CN} can be constructed by combining point identified potential outcomes under treatment (Equation (5)), with the alternative bounds on the potential outcomes of stratum CN under control:

$$\theta_{CN}^{U,MD} = E[Y(1,0)|CN] - E_{CN}^{L,MD}[Y(0,0)|CN] \quad (15)$$

$$\theta_{CN}^{L,MD} = E[Y(1,0)|CN] - E_{CN}^{U,MD}[Y(0,0)|CN] \quad (16)$$

2.5 Inconsistency of the Wald estimator without systematic selection of the accompanying migrant

Researchers might be tempted to assume no systematic selection of the accompanying migrant to achieve point identification. The most obvious assumption would be to strengthen Assumptions 8 or 9 from inequalities to equalities as stated in Assumption 10. The assumption states that the mean potential outcomes under control are identical for household types CN and CC . I refer to this situation as absence of selection of the accompanying migrant. However, the absence of this form of intra-household selection still causes the conventional Wald estimator to be inconsistent for θ_{CN} and an alternative identification result presented here should be used.

Assumption 10. *No systematic selection of accompanying migrant*

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

Proposition 4. *If Assumptions 1-7 and 10 hold, then $E_{CN}^{U/L,MD}[Y(0,0)|CN]$ (derived in Section 2.4) corresponds to $E[Y(0,0)|CN]$.*

Combining the identified outcome under treatment $E[Y(1,0)|CN]$ (see Equation (5)) with the identified outcome under control identifies the causal effect for stratum CN :

$$\theta_{CN}^{NS} = \frac{\bar{Y}^{110}(\pi_{CN} + \pi_{AN}) - \bar{Y}^{010}\pi_{AN}}{\pi_{CN}} - \frac{\bar{Y}^{000}(\pi_{NN} + \pi_{CN} + \pi_{CC}) - \bar{Y}^{100}\pi_{NN}}{\pi_{CN} + \pi_{CC}} \quad (17)$$

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

$$\theta_W = \frac{E[Y|Z = 1, M_2 = 0] - E[Y|Z = 0, M_2 = 0]}{E[M_1|Z = 1, M_2 = 0] - E[M_1|Z = 0, M_2 = 0]} \quad (18)$$

The four quantities in Equation (18) can be formulated as means of observed outcomes weighted by strata proportions (for calculations and a more detailed discussion see Appendix A.3)

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

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

$$b_W = \theta_W - \theta_{CN}^{NS} = \frac{\pi_{CC} \left[(\bar{Y}^{100} - \bar{Y}^{000}) a + (\bar{Y}^{110} - \bar{Y}^{010}) b \right]}{c}, \quad (19)$$

where $a = (\pi_{CN} \pi_{NN}) (\pi_{CN} + \pi_{CC} + \pi_{NN})$, $b = \pi_{AN} (\pi_{CN}^2 + \pi_{CN} \pi_{CC} + \pi_{CN} \pi_{AN} + \pi_{CC} \pi_{AN})$, and $c = (\pi_{CN}^2 + \pi_{CN} \pi_{CC}) (\pi_{CN} \pi_{CC} + \pi_{CC} \pi_{AN} + \pi_{CN}^2 + \pi_{CN} \pi_{AN} + \pi_{CN} \pi_{NN})$. The inconsistency is zero if $\pi_{CC} = 0$ or if $\bar{Y}^{100} = \bar{Y}^{000}$ and $\bar{Y}^{110} = \bar{Y}^{010}$. The latter conditions imply that mean potential outcomes under control are equal for latent groups CN , CC , NN and under treatment, for latent groups AN and CN . Thus, θ_{CN}^{NS} instead of θ_W should be used in settings, in which Assumption 10 is credible.

2.6 Estimation and inference

Bounds derived in Section 2.2 include minimum or maximum operators. These operators create several problems for estimation and inference. Hirano and Porter (2012) show that for non-differentiable parameters such as min and max operators, no asymptotically unbiased estimators exist. Therefore, estimators for bounds that use min and max functions can be severely biased in finite samples, and confidence intervals cannot be estimated using standard asymptotics or bootstrap methods. Chernozhukov, Lee, and Rosen (2013) derive a method to obtain half-median unbiased estimators for the lower and upper bound, and confidence intervals for the true parameter. The idea is to apply the min (max) function not directly to the bounding function, but to a precision-corrected version of it. Precision is adjusted by adding to each estimated bounding function its point-wise standard error times an appropriate critical value. Estimates with higher standard errors therefore require larger adjustments. The estimated bounds are conservative, and the half-median unbiased estimator of the upper bound exceeds the true value of the upper bound with the probability of at least 0.5 asymptotically. The estimator of the lower bound falls below the true value of the bound with probability of at least 0.5 asymptotically. Appendix A.4 provides a detailed description of the implementation of the procedure based on Huber, Laffers, and Mellace (2017) and Chen and Flores

(2015).¹²

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

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

$\hat{\theta}^L$ and $\hat{\theta}^U$ denote the estimated bounds and $\hat{\sigma}^L$ and $\hat{\sigma}^U$ the respective estimated standard errors which are obtained by bootstrap with 1999 replications.

3 Empirical applications

This section presents two empirical applications of the bounds. The first applies bounds to data from a visa lottery in Tonga used by Gibson, McKenzie, and Stillman (2011a) (henceforth GMS) to study the effects of migration on remaining household members. I apply the bounds to a set of outcomes at the household level – household composition and household assets. This application allows comparing the bounds to well-identified point estimates. The unit of analysis is the household as a whole (I_2), while the principal migrant is the individual, who applied for the visa lottery (I_1). The second application is based on a paper from McKenzie and Rapoport (2011) (henceforth MR) that studies the effect of migration on school attendance in Mexico and does not address the issue of invisible sample selection. I test sensitivity of results to various assumptions regarding all-move households. The unit of analysis is the individual child (I_2) and the principal migrant is any adult household member (I_1).

3.1 Effect of migration on remaining households in Tonga

GMS study the effects of migration from Tonga to New Zealand on household members left behind in Tonga. New Zealand allows a quota of 250 Tongans to immigrate to New Zealand each year without going through the usual migration categories. Among eligible registrants (Tongan citizens aged 18 to 45 years who meet English, health, and character requirements), a random ballot decides who receives a visa to migrate. These registrants are the principal migrants in my framework. Ballot winners must provide a job offer in New Zealand within six months after the lottery to have their application to migrate approved. Ballot winners can apply for visas for their immediate family (spouses and dependent children up to age 24).¹³

¹²Due to precision adjustment, bounds and confidence intervals can be outside the support for outcomes with limited support if the unadjusted estimate is close to the limits of the support. If estimates or confidence intervals of the upper/lower bound of $E[Y(0,0)|CN]$ are larger/smaller than the support of Y , I replaced the estimate with the upper/lower limit of the support of Y .

¹³For a more extensive description of this visa lottery, refer to GMS.

GMS use data from a household survey in Tonga that was designed to capture the effects of migration. The survey does not contain households of ballot winners in which all household members join the principal migrant.¹⁴ GMS use the random ballot to instrument for migration of the principal migrant. Instrumental variable estimation is necessary since 15% of ballot winners (among observed households) do not comply with the lottery and do not move to New Zealand (GMS refer to concerns regarding this non-compliance as *dropout bias*). GMS argue that substitution bias is of little concern in this context since the chances of migration under another migration channel are extremely low.

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

3.1.1 All-move households

GMS use visa regulations and define all-move households as households in which all individuals would be eligible to join the principal migrant in case she wins the lottery and migrates. Since these visa rules are based on observable characteristics (i.e., age, relationship with the principal migrant), GMS identified these households and removed them from their data. I refer to these households as *visa all-move households* (*VAM*). GMS therefore identified the effect for households that leave more distant relatives behind. This concerns 75 of 124 households in the observed group $O(0,0,0)$, and 18 of 26 households in the observed group $O(1,0,0)$ (see Table 3, Panel a). Finding visa all-move households in group $O(1,0,0)$ shows that the definition of all-move households based on observable characteristics is not identical to latent stratum *CC*. All households in group $O(1,0,0)$ belong to the latent stratum *NN* - households in which no individual would migrate (see Table 2). However, we can use the information on visa all-move households to estimate π_{CC} .

Using only information from observed group $O(0,0,0)$ leads to the conclusion that $\hat{\pi}_{CC}$ equals $N_{000}^{VAM}/N_{000} = 75/124 = 0.61$. However, in observed group $O(1,0,0)$, which consists only of stratum *NN*, 18 of 26 identify as visa all-move households ($N_{100}^{VAM}/N_{100} = 18/26 = 0.7$). The overall share of visa all-move households

¹⁴For a related but separate project, GMS do interview migrants in New Zealand including all-move households.

in $O(0,0,0)$ is therefore a combination of the 70% visa all-move households in stratum NN and 100% visa all-move households in stratum CC (Equation (20)). The ratio of $\pi_{CN}/\pi_{NN} = N_{110}/N_{100} = 2.34$.¹⁵ Since no households of type AN exist, it holds that $\pi_{NN} + \pi_{CN} + \pi_{CC} = 1$ (Equation (22)). Combining this information gives a system of three equations, which we can solve to obtain the strata proportions:

$$N_{100}^{VAM}/N_{100} * \hat{\pi}_{NN} + \hat{\pi}_{CC} = N_{000}^{VAM}/N_{000} \quad (20)$$

$$N_{110}/N_{100} * 1.368 = \hat{\pi}_{CN}/\hat{\pi}_{NN} \quad (21)$$

$$\hat{\pi}_{NN} + \hat{\pi}_{CN} + \hat{\pi}_{CC} = 1 \quad (22)$$

Panel b of Table 3 shows the estimated strata proportions. No households belong to stratum AN , 35% to stratum CN , 53% to stratum CC , and 11% to stratum NN . The ratio of unobserved to observed migrant households $\gamma = 1.5$.

Table 3: Tonga: observed groups and latent strata

Panel a: Observed groups			Panel b: Latent strata proportions		
$O(z, m_1, m_2)$	Number of households		Stratum	Proportion	Standard error
	all	visa all-move			
$O(0, 0, 0)$	124	75	π_{AN}	0.00	(0.00)
$O(0, 0, 1)$	-		π_{CN}	0.35	(0.04)
$O(0, 1, 0)$	-		π_{CC}	0.53	(0.06)
$O(0, 1, 1)$	-		π_{NN}	0.11	(0.03)
$O(1, 0, 0)$	26	18			
$O(1, 0, 1)$	-		γ	1.5	(0.33)
$O(1, 1, 0)$	61				
$O(1, 1, 1)$	-				

Note: The left panel presents the observed number of households in the dataset of GMS by value of the instrument and migration status. Visa all-move households are households where all individuals would be eligible to join the principal migrant. The right panel shows the estimated ratio of unobserved to observed migrant households (γ) and the estimated strata proportions. Standard errors in parentheses from 1999 bootstrap replications.

3.1.2 Assessment of assumptions

Assumption 1 holds through the random ballot that decides who receives a visa. The two exclusion restrictions, Assumption 2, and 3 are also likely to hold. Obtaining a visa has no direct effect on household welfare, except through its effect on migration (Assumption 2). Other household members can only obtain a visa if the principal migrant takes up his visa and migrates. An effect of the visa on other household members can thus only work through an effect on migration of the visa applicant (Assumption 3). Therefore, the visa

¹⁵This ratio must be adjusted by a factor of 1.368 due to differential sample weights (Equation 21). Group $O(0,0,0)$ has an expansion factor of 37.9, group $O(1,0,0)$ of 2.5, and group $O(1,1,0)$ of 3.4. For the estimation of the bounds this is relevant only when calculating the ratio π_{CN}/π_{NN} .

affects migration of other household members only through its effect on migration of the principal migrant. Assumption 5 is very likely to hold since it appears unreasonable that a visa makes a person less likely to migrate. Assumption 6 holds since we do not see any migrants in households in which the principal migrant does not migrate. Another question is whether all-move households are of type *AA* or *CC*. GMS argue that it is difficult to obtain another type of visa, which strongly suggests that all-move households must be of type *CC*. We do not observe any households with migrants, who are not lottery winners, which is further evidence that all-move households are of type *CC*. Thus, the assumptions required for the bounds in Proposition 1 are likely to be fulfilled. I discuss and assess further distributional assumptions that tighten the bounds or even allow point identification in the results section if these assumptions are specific to the individual outcome variables.

3.1.3 Results

Table 4 shows the bounds on the effect of migration on household composition. Household composition is a particularly appropriate outcome to study the problem of invisible sample selection since it relates strongly to the propensity of households to migrate as a whole.¹⁶ Removing all visa all-move households from the estimation sample, GMS conclude that the effect of migration of the principal migrant on the number of individuals living in the household is -2.26 persons.

The point identified household size of *CN* households under treatment is 4.69 persons. Using the bounds from Proposition 1, the size of *CN* households under control can be bounded between 2.17 and 8.56 persons, which leads to bounds on the effect for stratum *CN* between -3.87 and 2.54. A reasonable assumption might be that under control, *CC* households are on average smaller than *CN* households (Assumption 9). The larger a household, the more likely it is that at least one member stays behind. Invoking this assumption (Proposition 3) increases the lower bound on $E[Y(0,0)|CN]$ to 5.38. It decreases the upper bound on the effect to -0.69, which shows that this assumption has substantial identifying power.

The bounds analysis confirms the negative effect of migration on household size. For comparison, an approach that uses a Wald estimator and ignores selection due to all-move households would conclude that migration reduces total household size by 0.85 persons, not statistically different from zero. Invoking Assumption 10 allows point identification using the results from Proposition 4. The estimate of θ_{CN}^{NS} is -0.69, which is different from the Wald estimator. However, the underlying assumption that under control, *CC* households have on average the same number of members as *CN* households, is unlikely to hold in this context as *CC* households are likely smaller as discussed before. Still, these results are instructive to illustrate the difference between the Wald estimator θ_W and θ_{CN}^{NS} . The remaining columns show that the reduction in household size is driven by a reduction in the number of children and prime-age individuals in

¹⁶The results in this section correspond to Tables 4 and 6 in GMS.

Table 4: Effect of migration on household composition

	Total household size		Adults aged 18 to 45		Children aged under 18		Adults aged over 45	
$E[Y(0,0) NN]$	4.42	(0.51)	2.15	(0.21)	2.08	(0.38)	0.19	(0.08)
$E[Y(1,0) CN]$	4.69	(0.34)	1.62	(0.15)	1.92	(0.24)	1.10	(0.12)
<i>Bounds under Proposition 1</i>								
Bounds on $E[Y(0,0) CN]$	[2.16	8.55]	[1.34	3.51]	[0.22	5.03]	[0.00	1.04]
CLR 95% CI	(1.57	9.26)	(1.14	3.95)	(0.00	5.63)	(0.00	1.25)
Bounds on θ_{CN}	[-3.87	2.54]	[-1.89	0.29]	[-3.11	1.72]	[0.06	1.10]
CLR 95% CI	(-4.74	3.33)	(-2.39	0.60)	(-3.81	2.16)	(-0.22	1.29)
<i>Bounds under Proposition 3</i>								
Bounds on $E[Y(0,0) CN]$	[5.38	8.55]	[2.35	3.51]	[2.61	5.03]	[0.42	1.04]
CLR 95% CI	(4.83	9.26)	(2.11	3.95)	(2.20	5.63)	(0.29	1.25)
Bounds on θ_{CN}	[-3.87	-0.69]	[-1.89	-0.73]	[-3.11	-0.69]	[0.06	0.68]
CLR 95% CI	(-4.74	0.09)	(-2.39	-0.38)	(-3.81	-0.13)	(-0.22	0.91)
Unadjusted IV θ_w	-0.85	(0.53)	-0.76	(0.23)	-0.78	(0.39)	0.64	(0.15)
θ_{CN}^{NS}	-0.69	(0.47)	-0.73	(0.21)	-0.69	(0.34)	0.68	(0.14)
IV GMS	-2.26	(0.61)	-1.53	(0.33)	-0.86	(0.44)	0.08	(0.18)
Observations	211		211		211		211	

Note: Results are based on the estimated strata proportions in Table 3. *Total household size* is the number of individuals living in the household. The remaining columns are the number of individuals in the respective age group. $\theta_{CN} = E[Y(1,0) - Y(0,0)|CN]$. θ_{CN}^{NS} is the point identified effect for stratum CN using the identification result in Proposition 4. *Unadjusted IV θ_w* is the estimate from a Wald estimator ignoring invisible sample selection. *IV GMS* is the point identified effect for the group of visa all-move households as reported in GMS. Standard errors in parentheses from 1999 bootstrap replications. For the bounds without mean dominance, numbers in parentheses in the bottom rows are 95% confidence intervals calculated using the procedure by Chernozhukov, Lee, and Rosen (2013), while numbers in square brackets are identified sets determined by the half-median unbiased estimators. For the bounds with mean dominance, the 95% confidence interval is calculated using the procedure by Imbens and Manski (2004).

the households. The number of elderly is not reduced.

A second set of results examines the effects of migration on household assets (Table 5). 48% of CN households under treatment own a home. Using Proposition 1, home ownership rate of CN households under control can only be bounded to be between 0 and 1. The resulting bounds on the treatment effect are -0.52 and 0.48 and hardly informative. The bounds can be narrowed substantially by assuming that under control, CC households are less likely to own a home (Assumption 9). Home ownership is generally seen as having an impeding effect on migration, see, for example, Massey and Espinosa (1997) and Nivalainen (2004). Invoking this assumption (Proposition 3), the lower bound on $E[Y(0,0)|CN]$ increases to 0.4 and the effect on home-ownership rate can be bounded between -0.52 and 0.08. The effect of 0.12 from the unadjusted Wald estimator θ_w lies outside of these bounds.

I invoke the same mean dominance assumption when identifying the effect on agricultural assets and livestock owned: under control, CN households own on average more livestock than CC households. Again, this appears to be a reasonable assumption as agricultural assets might increase the likelihood that someone

Table 5: Effect of migration on household assets

	Home ownership		Any agricultural assets		Number of pigs		Number of chickens		Number of cattle	
$E[Y(0,0) NN]$	0.62	(0.10)	0.85	(0.07)	4.27	(0.73)	3.54	(0.80)	0.65	(0.22)
$E[Y(1,0) CN]$	0.48	(0.06)	0.79	(0.05)	4.82	(0.63)	4.33	(0.75)	0.93	(0.25)
<i>Bounds under Proposition 1</i>										
Bounds on $E[Y(0,0) CN]$	[0.00	1.01]	[0.63	1.00]	[1.05	8.95]	[0.00	13.58]	[0.00	3.16]
CLR 95% CI	(0.00	1.06)	(0.45	1.00)	(0.45	9.99)	(0.00	16.02)	(0.00	4.01)
Bounds on θ_{CN}	[-0.54	0.48]	[-0.22	0.16]	[-4.14	3.81]	[-9.25	4.33]	[-2.22	0.93]
CLR 95% CI	(-0.65	0.58)	(-0.30	0.36)	(-5.57	4.96)	(-11.98	5.56)	(-3.17	1.35)
<i>Bounds under Proposition 3</i>										
Bounds on $E[Y(0,0) CN]$	[0.40	1.01]	[0.86	1.00]	[4.92	8.95]	[5.94	13.58]	[1.27	3.16]
CLR 95% CI	(0.31	1.06)	(0.80	1.00)	(4.24	9.99)	(4.55	16.02)	(0.89	4.01)
Bounds on θ_{CN}	[-0.54	0.08]	[-0.22	-0.07]	[-4.14	-0.10]	[-9.25	-1.62]	[-2.22	-0.34]
CLR 95% CI	(-0.65	0.21)	(-0.30	0.04)	(-5.57	1.15)	(-11.98	0.26)	(-3.17	0.23)
Unadjusted IV θ_w	0.12	(0.09)	-0.07	(0.07)	-0.21	(0.82)	-2.01	(1.27)	-0.44	(0.37)
θ_{CN}^{NS}	0.08	(0.08)	-0.07	(0.07)	-0.10	(0.75)	-1.62	(1.14)	-0.34	(0.34)
IV GMS	-0.01	(0.10)	-0.19	(0.06)	-1.26	(0.81)	-4.58	(1.76)	-0.85	(0.49)
Observations	211		211		211		211		211	

Note: Results are based on the estimated strata proportions in Table 3. *Any agricultural assets* is a binary indicator for owning any pigs, chicken, or cattle. $\theta_{CN} = E[Y(1,0) - Y(0,0)|CN]$. θ_{CN}^{NS} is the point identified effect for stratum *CN* using the identification result in Proposition 4. *Unadjusted IV θ_w* is the estimate from a Wald estimator ignoring invisible sample selection. *IV GMS* is the point identified effect for the group of visa all-move households as reported in GMS. Standard errors in parentheses from 1999 bootstrap replications. For the bounds without mean dominance, numbers in parentheses in the bottom rows are 95% confidence intervals calculated using the procedure by Chernozhukov, Lee, and Rosen (2013), while numbers in square brackets are identified sets determined by the half-median unbiased estimators. For the bounds with mean dominance, the 95% confidence interval is calculated using the procedure suggested by Imbens and Manski (2004).

stays behind. The bounds under Proposition 3 suggest negative effects on the probability to own any agricultural assets, as well as the number of pigs, chickens, and cattle owned, though the confidence intervals do not disqualify zero effects.

As in the previous table, I present estimates for θ_{CN}^{NS} and θ_w that show considerable differences. However, it appears that Assumption 10 is also not credible in this setting and thus both estimates are likely biased.

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

3.2 Effect of migration on school attendance in Mexico

The second empirical application follows MR, estimating the effect of migration on school attendance in Mexico. MR use historical migration rates as an instrument for current migration, finding that migration

of an adult household member reduces school attendance rates for 12 to 15 year old boys by 16 percentage points (pp.), and by 9 pp. for girls; however, the latter effect is not significantly different from zero.¹⁷

Data stem from the Mexican 1997 Encuesta Nacional de la Dinámica Demográfica (ENADID).¹⁸ I follow MR and define a child as *living in a migrant household* if the household has a member aged 19 and over who has ever been to the United States to work, or who has moved to the United States in the last 5 years for any other reason.¹⁹

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

3.2.1 Sample selection due to child migration

No children in the sample are categorized as current migrants.²⁰ Therefore, potential sample selection arises from migration of whole households. Although the ENADID dataset provides rich information on individual migration histories, it lacks information on households that migrate as a whole. To gain an understanding of how widespread the phenomenon of all-move households is in Mexico, I build on existing research that uses census data from the source and destination countries of migrants. Ibarraran and Lubotsky (2007) estimate the size of the Mexican immigrant population in the United States based on a) the 2000 Mexican census and b) the 2000 U.S. census. Since the Mexican census was conducted as a household survey, it ignored all-move households. The estimated size of the Mexican-born population living in the United States based on the Mexican census is 1,221,598, and based on the U.S. census is 2,205,356. Thus, the total migrant population in the Mexican census is only 55.4% the size of the population in the U.S. census. This rate is lower for female (33.6%) than for male migrants (69.9%).

The authors argue the difference is primarily due to married couples that dissolved their household in Mexico and are not counted in the Mexican census. Once married couples with both spouses present in the U.S. are removed from U.S. census estimates, the remaining migrant number is 1,492,111. In a similar

¹⁷Unlike MR, I focus only on the effect of migration on school attendance in the sample of children aged 12 to 15 years. Restricting the analysis in this way offers two advantages. First, children in this age group are unlikely to migrate without their parents, which is required by Assumption 6. This assumption might not hold for 16 to 18 year old adolescents, the second group that MR consider. Second, in comparison to years of education, school attendance is the more natural outcome for children and adolescents who have not yet completed their education.

¹⁸The ENADID is a nationally representative household survey, with a sample of 73,412 households. This corresponds to roughly 2,300 households in each of the 32 states. To allow comparability of results with MR, I restrict the sample similarly to households in municipalities outside of cities with more than 50,000 inhabitants. The estimation sample consists of 15,665 children aged 12 to 15 years in 11,160 households.

¹⁹This is not the optimal migrant definition to study sample selection since it also includes return migrants. However, to maximize comparability of results with existing research, I follow the definition from MR, who argue that prior migration episodes of adult household members also influence the education outcomes of children.

²⁰Fourteen children reported prior migration episodes, and of these, six come from non-migrant households. However, the questionnaire included several questions on migration, and the answers to these observations are inconsistent. Therefore, data problems seem to be the reason for this finding.

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 spouses present in the United States, concluding these individuals are likely not counted in Mexico-based surveys.

Discrepancies between numbers from the Mexican and U.S. censuses are even larger for children. In the age group 12 to 15 years, the number of migrants in the Mexican census is only about 50% of the number of migrants in the U.S. census. Again, the reason is most likely that children migrate with their whole families and are therefore not counted in the Mexican census anymore. Overall, the U.S. census counts 82,240 Mexican-born children in this age group, which are most likely not included in Mexican data.²¹

I use the ENADID to calculate the total number of children in migrant households in Mexico. Using the definition of a migrant household described above and the expansion factors provided with the data, I calculate the total number of children aged 12 to 15 years who live in a migrant household to be 1,516,924. Dividing the number of children missed by the observed number of children in migrant households, I calculate γ to be 0.054.²² This ratio appears low, but this is because of the broad definition of a migrant household in the ENADID, and thus the large denominator. I test the sensitivity of results to various values of γ ranging from zero to 0.5. For the main analysis, I use a ratio of $\gamma = 0.054$.

3.2.2 Assessment of assumptions

To overcome the problem of self-selection into migration, several recent studies (e.g. McKenzie and Hildebrandt, 2005; McKenzie and Rapoport, 2011) use historical migration rates to instrument for current migration. Existing networks lower migration costs for subsequent migrants, and therefore trigger additional migration. The exclusion restriction is that these historical migration rates do not affect educational outcomes today except through current migration of household members (Assumption 2). This assumption has been criticized and might be violated if, for example, local economy effects induced by historical migration influence educational outcomes. A detailed discussion of this instrument and the exclusion restriction regarding educational attainment can be found in MR. However, the bounds in this paper require an additional assumption about the instrument. Assumption 3 suggests the instrument must not influence the migration decision of a child directly, which appears reasonable if migration networks primarily help adult migrants to find a job in the destination country. However, it might be violated if adults are more likely to take the child with them if they expect their network to help them settle and adjust with the child.

Like MR, I use state-level migration rates to the United States from 1924 taken from Woodruff and

²¹Thanks go to Darren Lubotsky for providing this estimate.

²²I calculate this ratio for Mexico as a whole since I can not identify the origin region of migrants in the U.S. census.

Zenteno (2007). I recode this continuous measure into a binary one by defining states as low-migration states ($Z = 0$) if the migration rate is below the state-level median (3.78%) and as high-migration states ($Z = 1$) if the migration rate is above. I do this to allow stratification on instrument assignment, which would not be possible with a continuous instrument. Figure 2 in Appendix 5 shows the positive relationship between historical migration rates and the probability of a child living in a migrant household (Assumption 4). In this setting, compliers are individuals who would migrate only if they live in a high-migration state. Unlike MR, I abstain from including additional covariates in my estimation to ensure Assumption 1 holds. Covariates would substantially complicate the analysis since I do not observe the distribution of covariates for all-move households. Thus, while my approach allows taking into account invisible sample selection, I have to impose the stronger assumption that the instrument is unconditionally valid. However, two-stage-least-squares point estimates with a binary instrument without covariates differ only slightly from those using covariates and are similar to results from MR.²³

No children in the sample are categorized as current migrants. This is strong evidence in support of Assumption 6, that children would not migrate alone. Whether all-move households react to the instrument (CC) or are always migrants (AA) cannot be identified with the available data. Assuming they are all type CC leads to conservative bounds.

We might impose the additional mean dominance assumption that under control, children in CC households are more likely to attend school than children in CN households (Assumption 8). Since migration is costly, richer households are more likely to take their children with them. Richer households may also have more resources to send the children to school. Thus, it is likely that this assumption holds.

3.2.3 Results

The first rows of Table 6 show the estimated strata proportions. The proportion of stratum CN for boys is 0.26, and the proportion of stratum CC is 0.02. NN is the largest stratum, with a proportion of 0.6. Strata proportions are similar for girls. The next three rows display the point identified mean potential outcomes for stratum NN under control, and for strata AN and CN under treatment. The expected school attendance rate under treatment for stratum CN is 0.63 for boys and 0.6 for girls. School attendance is slightly higher for boys than girls in all strata. The bounds under monotonicity for school attendance rates of boys in stratum CN under control are 0.78 and 0.89. For girls, they are substantially lower, 0.49 and 0.57, respectively.

²³For the model with controls, I use the following state-level control variables: number of schools per 1,000 inhabitants in 1930, literacy rate in 1960, and male and female attendance rates in 1930. These are not the same controls used by MR since those controls could not be reconstructed. Including these covariates changes the point estimate in a two-stage, least-squares estimation for boys from -19.5 to -14.7 pp. and for girls from 8.2 to 7.4 pp.

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

Imposing the additional mean dominance assumption that the mean outcome under control of stratum CC is weakly greater than of stratum CN , narrows the bounds by increasing the lower bound on θ_{CN} (Proposition 2). For boys, the lower bound increases to -0.19 and for girls to 0.07.

For comparison, ignoring sample selection and estimating effects using a Wald estimator suggests the effect of migration on school attendance of boys is -0.19 and for girls is 0.08 but not significant. In this case, results for θ_{CN}^{NS} are almost identical to the one from the unadjusted Wald estimator.²⁴

Table 6: Effect of adult migration on school attendance

	School attendance (0/1)			
	Boys		Girls	
π_{AN}	0.12	(0.03)	0.12	(0.03)
π_{CN}	0.26	(0.04)	0.25	(0.04)
π_{CC}	0.02	(0.00)	0.02	(0.00)
π_{NN}	0.60	(0.04)	0.61	(0.04)
$E[Y(0,0) NN]$	0.73	(0.02)	0.70	(0.02)
$E[Y(1,0) AN]$	0.77	(0.03)	0.71	(0.02)
$E[Y(1,0) CN]$	0.63	(0.04)	0.60	(0.04)
<i>Bounds under Proposition 1</i>				
Bounds on $E[Y(0,0) CN]$	[0.78	0.89]	[0.49	0.57]
CLR 95% CI	(0.70	0.99)	(0.31	0.75)
Bounds on θ_{CN}	[-0.28	-0.14]	[-0.02	0.16]
CLR 95% CI	(-0.40	-0.04)	(-0.18	0.32)
<i>Bounds under Proposition 2</i>				
Bounds on $E[Y(0,0) CN]$	[0.78	0.82]	[0.49	0.53]
CLR 95% CI	(0.72	0.92)	(0.36	0.69)
Bounds on θ_{CN}	[-0.19	-0.14]	[0.07	0.16]
CLR 95% CI	(-0.32	-0.04)	(-0.13	0.32)
Unadjusted IV θ_w	-0.19	(0.08)	0.08	(0.13)
θ_{CN}^{NS}	-0.19	(0.08)	0.07	(0.12)
Observations	7993		7663	

Note: Results based on the assumption that $\gamma = 0.054$. $\theta_{CN} = E[Y(1,0) - Y(0,0)|CN]$. θ_{CN}^{NS} is the point identified effect for stratum CN using the identification result in Proposition 4. *Unadjusted IV θ_w* is the estimate from a Wald estimator ignoring invisible sample selection. Standard errors in parentheses from 1999 bootstrap replications clustered at the state level. For the bounds without mean dominance, numbers in parentheses in the bottom rows are 95% confidence intervals calculated using the procedure by Chernozhukov, Lee, and Rosen (2013), while numbers in square brackets are identified sets determined by the half-median unbiased estimators. For the bounds with mean dominance, the 95% confidence interval is calculated using the procedure suggested by Imbens and Manski (2004).

²⁴Although the estimates for boys with and without covariates are similar to results from MR, there is a larger discrepancy in the estimates for girls. Unreported estimates using the continuous instrument are much closer to MR's result.

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

3.2.4 Sensitivity with respect to γ

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

[Figure 3 about here]

Figure 3 shows the resulting bounds on the effects for boys and girls. The width of the bounds does not increase constantly over the range of the observed ratios. For boys, the lower bound decreases steeply up to a value of γ of about 0.14, and only slightly thereafter. Up to a ratio of 0.14, the estimate of the upper bound on $E[Y(0,0)|CN]$ stems from the constrained solution. Once π_{CN} is sufficiently small so the constraint no longer binds, the estimate stems from the unconstrained solution. A slight decrease after this threshold results from a steady decline in π_{CN} and the fact that the expected value is computed in a decreasing fraction of largest values of the outcome Y in group $O(0,0,0)$. For girls, we observe a constrained solution up to a ratio of 0.41.²⁵

Two additional insights can be gained from Figure 3. First, γ is not the only determinant of the behavior of the bounds. For example, at a value of 0.3, the width of the bounds is 0.33 for boys and 0.59 for girls, which is due to the various distributions of the outcome variable. Second, the lower bound under monotonicity and mean dominance is insensitive to variations in γ .

4 Discussion

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

An alternative to the bounds derived from Chen and Flores (2015) would be an approach outlined in Imai (2007). Imai suggests subtracting the full distribution of $Y(0,0)|NN$ derived from the observed group

²⁵The kink in the upper bound for boys at a ratio of about 0.3 stems from the precision adjustment. Above this ratio, the unrestricted solution is not considered “close” enough to influence the asymptotic distribution of the upper bound.

$O(1, 0, 0)$ from the distribution of Y in the observed group $O(0, 0, 0)$, and employing the trimming procedure in the remaining distribution. Although this approach might tighten the bounds somewhat, estimating and subtracting distributions instead of means creates complications for estimation. Since this approach is not developed fully, I abstain from including those bounds here.

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

This paper investigates a situation in which the potential principal migrant can be identified and distinguished from other household members. Extensions should examine situations in which the principal migrant(s) cannot be identified and allow for more complicated household structures. Especially when interest is in the effect of migration of an adult on other adult household members (i.e., the effect of migration on the labor supply of other household members), identifying the principal migrant might be impossible.

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

While this paper discusses intra-household selection mainly from a sample selection perspective, the proposed approach can also be used to identify various other effects and disentangle mechanisms. Researchers and policymakers might be interested in $Y(1, 1) - Y(0, 0)$, the effect if a child migrates with an adult, in comparison to a situation in which no household member migrates (e.g., Stillman, Gibson, and McKenzie, 2012) or in $Y(1, 1) - Y(1, 0)$, which is the effect of migration of the whole household in comparison to a situation in which the child remains while the adult migrates (e.g., Gibson, McKenzie, and Stillman, 2011b). One could identify the effects not only for one latent population, but for various populations. Huber, Laffers, and Mellace (2017) derive bounds on average treatment effects on the treated and other populations in a

setting with non-compliance. Chen, Flores, and Flores-Lagunes (2014) derive bounds on population average treatment effects. Such approaches could be extended and applied to the migration setting.

5 Conclusion

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

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

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

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

Finally, the paper encourages partial instead of point identification in contexts in which point identifica-

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

References

- AGESA, R. U., AND S. KIM (2001): “Rural to Urban Migration as a Household Decision: Evidence from Kenya,” *Review of Development Economics*, 5(1), 60–75.
- ALBERT, J. M., AND S. NELSON (2011): “Generalized Causal Mediation Analysis,” *Biometrics*, 67(3), 1028–1038.
- ALMOND, D., AND J. CURRIE (2011): “Human Capital Development Before Age Five,” *Handbook of Labor Economics*, 4(11), 1315–1486.
- AMUEDO-DORANTES, C., A. GEORGES, AND S. POZO (2010): “Migration, Remittances, and Children’s Schooling in Haiti,” *The ANNALS of the American Academy of Political and Social Science*, 630(1), 224–244.
- ANGRIST, J. D., G. W. IMBENS, AND D. B. RUBIN (1996): “Identification of Causal Effects Using Instrumental Variables,” *Journal of the American Statistical Association*, 91(434), 444–455.
- ANTMAN, F. M. (2011): “The Intergenerational Effects of Paternal Migration on Schooling and Work: What Can We Learn from Children’s Time Allocations?,” *Journal of Development Economics*, 96(2), 200–208.
- ANTMAN, F. M. (2012): “Gender, Educational Attainment and the Impact of Parental Migration on Children Left Behind,” *Journal of Population Economics*, 25(4), 1187–1214.
- (2013): “The Impact of Migration on Family Left Behind,” *International Handbook on the Economics of Migration*, pp. 293–308.
- BLANCO, G., C. A. FLORES, AND A. FLORES-LAGUNES (2013a): “Bounds on Average and Quantile Treatment Effects of Job Corps Training on Wages,” *Journal of Human Resources*, 48(3), 659–701.
- BLANCO, G., C. A. FLORES, AND A. FLORES-LAGUNES (2013b): “The Effects of Job Corps Training on Wages of Adolescents and Young Adults,” *American Economic Review*, 103(3), 418–422.
- BLAU, D. M. (1999): “The Effect of Income on Child Development,” *Review of Economics and Statistics*, 81(2), 261–276.
- CASE, A., D. LUBOTSKY, AND C. PAXSON (2002): “Economic Status and Health in Childhood: The Origins of the Gradient,” *American Economic Review*, 92(5), 1308–1334.
- CHEN, X., AND C. A. FLORES (2015): “Bounds on Treatment Effects in the Presence of Sample Selection and Noncompliance: The Wage Effects of Job Corps,” *Journal of Business and Economics Statistics*, 33(4), 523–540.
- CHEN, X., C. A. FLORES, AND A. FLORES-LAGUNES (2014): “Bounds on Population Average Treatment Effects with an Instrumental Variable,” *Manuscript*.
- (2018): “Going Beyond LATE: Bounding Average Treatment Effects of Job Corps Training,” *Journal of Human Resources*, 53, 1050–1099.

- CHERNOZHUKOV, V., S. LEE, AND A. M. ROSEN (2013): “Intersection Bounds: Estimation and Inference,” *Econometrica*, 81(2), 667–737.
- 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.
- FLORES, C. A., AND A. FLORES-LAGUNES (2009): “Identification and Estimation of Causal Mechanisms and Net Effects of a Treatment under Unconfoundedness,” *IZA Discussion Paper Series*, 4237.
- FLORES, C. A., AND A. FLORES-LAGUNES (2010): “Nonparametric Partial Identification of Causal Net and Mechanism Average Treatment Effects,” *California Polytechnic State University at San Luis Obispo*, Manuscript.
- FRANGAKIS, C. E., AND D. B. RUBIN (2002): “Principal Stratification in Causal Inference,” *Biometrics*, 58(1), 21–29.
- FRÖLICH, M. (2007): “Nonparametric IV estimation of local average treatment effects with covariates,” *Journal of Econometrics*, 139(1), 35–75.
- FRUMENTO, P., F. MEALLI, B. PACINI, AND D. B. RUBIN (2012): “Evaluating the Effect of Training on Wages in the Presence of Noncompliance, Nonemployment, and Missing Outcome Data,” *Journal of the American Statistical Association*, 107(498), 450–466.
- GIBSON, J., D. MCKENZIE, AND S. STILLMAN (2011a): “The Impacts of International Migration on Remaining Household Members: Omnibus Results from a Migration Lottery Program,” *Review of Economics and Statistics*, 93(4), 1297–1318.
- (2011b): “What Happens to Diet and Child Health When Migration Splits Households? Evidence from a Migration Lottery Program,” *Food Policy*, 36(1), 7–15.
- (2013): “Accounting for Selectivity and Duration-dependent Heterogeneity when Estimating the Impact of Emigration on Incomes and Poverty in Sending Areas,” *Economic Development and Cultural Change*, 61(2), 247–280.
- GRONAU, R. (1974): “Wage Comparisons - A Selectivity Bias,” *Journal of Political Economy*, 82(6), 1119–1143.
- HANSON, G. H., AND C. M. WOODRUFF (2003): “Emigration and Educational Attainment in Mexico,” *Manuscript*.
- HECKMAN, J., R. PINTO, AND P. SAVELYEV (2013): “Understanding the Mechanisms through which an Influential Early Childhood Program Boosted Adult Outcomes,” *American Economic Review*, 103(6), 2052–2086.
- HECKMAN, J. J. (1974): “Shadow Prices, Market Wages, and Labor Supply,” *Econometrica*, 42(4), 679–694.
- HIRANO, K., AND J. R. PORTER (2012): “Impossibility Results for Nondifferentiable Functionals,” *Econometrica*, 80(4), 1769–1790.
- HUBER, M. (2014): “Identifying Causal Mechanisms (Primarily) Based on Inverse Probability Weighting,” *Journal of Applied Econometrics*, 29(6), 920–943.
- HUBER, M., L. LAFFERS, AND G. MELLACE (2017): “Sharp IV Bounds on Average Treatment Effects on the Treated and Other Populations Under Endogeneity and Noncompliance,” *Journal of Applied Econometrics*, 32(1), 56–79.
- HUBER, M., M. LECHNER, AND G. MELLACE (2017): “Why Do Tougher Caseworkers Increase Employment? The Role of Program Assignment as a Causal Mechanism,” *Review of Economics and Statistics*, 99(1), 180–183.
- HUBER, M., AND G. MELLACE (2015): “Sharp Bounds on Causal Effects under Sample Selection,” *Oxford Bulletin of Economics and Statistics*, 77(1), 129–151.
- IBARRARAN, P., AND D. LUBOTSKY (2007): “Mexican Immigration and Self-Selection: New Evidence from the 2000 Mexican Census,” in *NBER Chapters*, ed. by G. J. Borjas, pp. 159–192. National Bureau of

- Economic Research, Inc.
- IMAI, K. (2007): "Identification Analysis for Randomized Experiments with Noncompliance and "Truncation-by-Death"," *Manuscript, Department of Politics, Princeton University*.
- IMBENS, G. W., AND J. D. ANGRIST (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62(2), 467–475.
- IMBENS, G. W., AND C. F. MANSKI (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72(6), 1845–1857.
- KUHN, R., B. EVERETT, AND R. SILVEY (2011): "The Effects of Children's Migration on Elderly Kin's Health: a Counterfactual Approach," *Demography*, 48(1), 183–209.
- LEE, D. S. (2009): "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects," *Review of Economic Studies*, 76(3), 1071–1102.
- LEIBOWITZ, A. (1974): "Home Investments in Children," *Journal of Political Economy*, 82(2), 111–131.
- MASSEY, D., AND K. ESPINOSA (1997): "What's Driving Mexico-U.S. Migration? A Theoretical, Empirical, and Policy Analysis," *American Journal of Sociology*, 102(4), 939–999.
- MATTEI, A., AND F. MEALLI (2007): "Application of the Principal Stratification Approach to the Faenza Randomized Experiment on Breast Self-examination," *Biometrics*, 63(2), 437–446.
- McKENZIE, D. (2015): "Learning about migration through experiments," in *Migration: Economic change, social challenge*, ed. by C. Dustmann, pp. 68–82. Oxford University Press, USA.
- McKENZIE, D., J. GIBSON, AND S. STILLMAN (2010): "How Important Is Selection? Experimental vs. Non-Experimental Measures of the Income Gains from Migration," *Journal of the European Economic Association*, 8(4), 913–945.
- McKENZIE, D., AND N. HILDEBRANDT (2005): "The Effects of Migration on Child Health in Mexico," *Economia*, 6(1), 257–289.
- McKENZIE, D., AND H. RAPOPORT (2007): "Self-selection Patterns in Mexico-U.S. Migration: The Role of Migration Networks," *Review of Economics and Statistics*, 92(4), 811–821.
- McKENZIE, D., AND H. RAPOPORT (2011): "Can Migration Reduce Educational Attainment? Evidence from Mexico," *Journal of Population Economics*, 24(4), 1331–1358.
- McKENZIE, D., AND D. YANG (2012): "Experimental Approaches in Migration Studies," in *Handbook of Research Methods in Migration*, ed. by C. Vargas-Silva, pp. 249–272. Edward Elgar Publishing.
- NIVALAINEN, S. (2004): "Determinants of Family Migration: Short Moves vs. Long Moves," *Journal of Population Economics*, 17(1), 157–175.
- PEARL, J. (2001): "Direct and Indirect Effects," in *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, pp. 411–420. Morgan Kaufman, San Francisco.
- RUBIN, D. B. (1974): "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies," *Journal of Educational Psychology*, 66(5), 688–701.
- (1980): "Randomization Analysis of Experimental Data: The Fisher Randomization Test Comment," *Journal of the American Statistical Association*, 75(371), 591–593.
- STILLMAN, S., J. GIBSON, AND D. McKENZIE (2012): "The Impact of Immigration on Child Health: Experimental Evidence from a Migration Lottery Program," *Economic Inquiry*, 50(1), 62–81.
- WOODRUFF, C., AND R. ZENTENO (2007): "Migration Networks and Microenterprises in Mexico," *Journal of Development Economics*, 82(2), 509–528.
- YANG, D. (2008): "International Migration, Remittances and Household Investment: Evidence from Philippine Migrants' Exchange Rate Shocks," *Economic Journal*, 118(528), 591–630.
- ZHANG, J., AND D. 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. Vytlačil, pp. 117–145. Emerald Group Publishing.

Figures

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

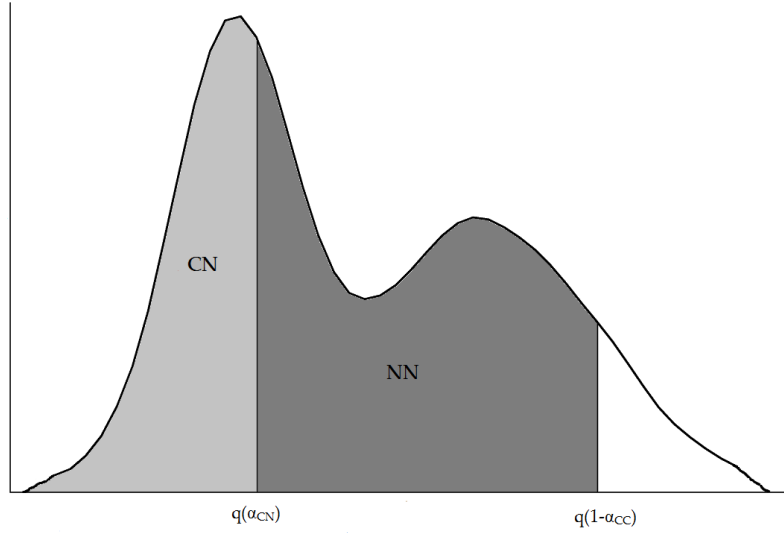
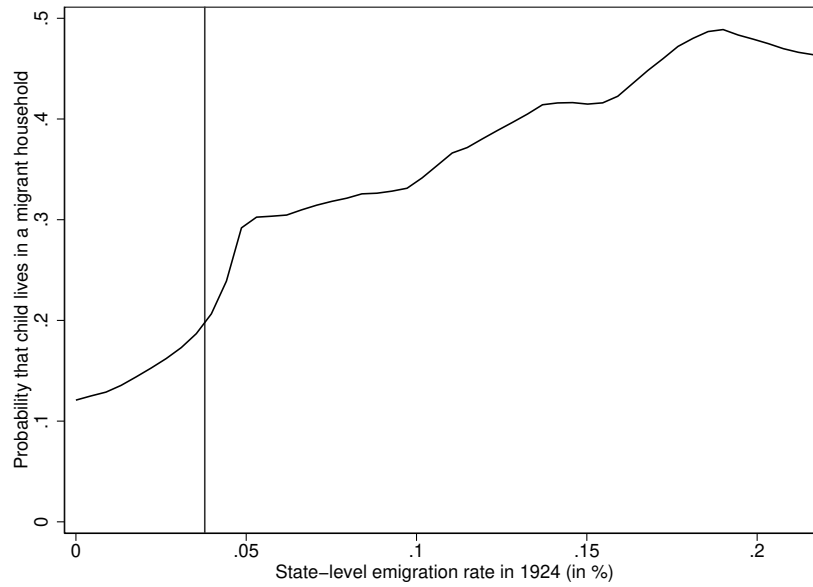
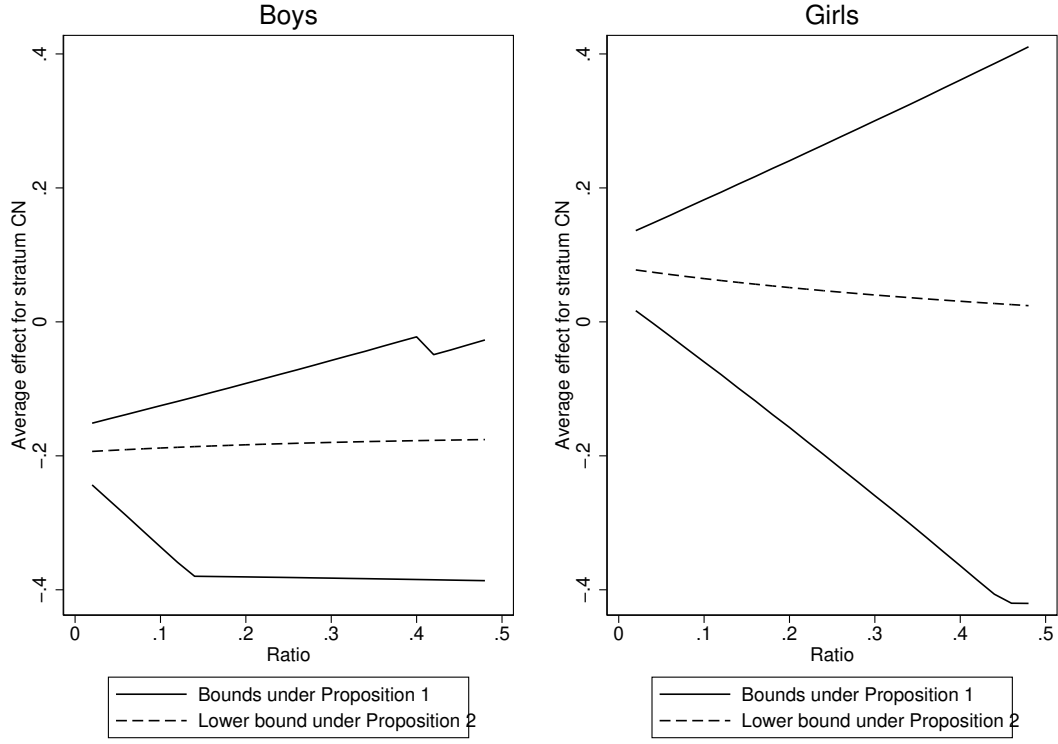


Figure 2: Cut-off for binary instrument: historical and contemporaneous migration



Notes: The figure shows the relationship between the Mexican state-level emigration rate in 1924 and the probability for an observed child to live in a migrant household. Local polynomial estimate with Epanechnikov kernel (bandwidth 0.021). The vertical line shows the median value that serves as the cut-off for the construction of the binary instrument.

Figure 3: Sensitivity of bounds on θ_{CN} to different ratios of unobserved to observed children in migrant households (γ)



Notes: The figure shows bounds on θ_{CN} for different values of γ .

A Appendix

A.1 Identification of strata proportions if migration of I_2 is observed

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

A.2 Point identification if $\pi_{CC} = 0$

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

$E[Y(1, 0) | G = CN]$ is identified in Equation (3)

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

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

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

Therefore, the causal effect is identified as

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

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

$$\theta_W = \frac{E[Y | Z = 1, M_2 = 0] - E[Y | Z = 0, M_2 = 0]}{E[M_1 | Z = 1, M_2 = 0] - E[M_1 | Z = 0, M_2 = 0]} \quad (23)$$

The four quantities in Equation (23) can be formulated as weighted means of observed outcomes and strata proportions

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

The Wald estimator

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

simplifies to

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

which equals the effect for stratum CN :

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

A.3 Inconsistency of the Wald estimator without systematic selection of the accompanying migrant

Assumption 10 ($E[Y(0,0)|G=CC] = E[Y(0,0)|G=CN]$) allows point identification of θ_{CN} even if $\pi_{CC} > 0$. $E[Y(1,0)|G=CN]$ is identified in Equation (3)

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

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

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

Therefore, the causal effect is identified as

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

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

$$\theta_W = \frac{E[Y|Z = 1, M_2 = 0] - E[Y|Z = 0, M_2 = 0]}{E[M_1|Z = 1, M_2 = 0] - E[M_1|Z = 0, M_2 = 0]} \quad (24)$$

The four quantities in Equation (24) can be formulated as weighted means of observed outcomes and strata proportions, making use of $\pi_{CC} + \pi_{CN} + \pi_{NN} + \pi_{AN} = 1$.

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

The Wald estimator is

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

which can be simplified to

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

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

$$\begin{aligned} b_W &= \theta_W - \theta_{CN}^{NS} = \frac{\pi_{CC} \left[(\bar{Y}^{100} - \bar{Y}^{000}) a + (\bar{Y}^{110} - \bar{Y}^{010}) b \right]}{c} \\ a &= (\pi_{CN} \pi_{NN}) (\pi_{CN} + \pi_{CC} + \pi_{NN}) \\ b &= \pi_{AN} (\pi_{CN}^2 + \pi_{CN} \pi_{CC} + \pi_{CN} \pi_{AN} + \pi_{CC} \pi_{AN}) \\ c &= (\pi_{CN}^2 + \pi_{CN} \pi_{CC}) (\pi_{CN} \pi_{CC} + \pi_{CC} \pi_{AN} + \pi_{CN}^2 + \pi_{CN} \pi_{AN} + \pi_{CN} \pi_{NN}). \end{aligned}$$

A.4 Inference based on Chernozhukov, Lee, and Rosen (2013)

I explain the estimation procedure for $E_{CN}^L[Y(0, 0)|G = CN]$. Recall that the lower bound of the expected value of stratum CN under control is given by $\Delta^L = \max_{v \in V = \{0, 1\}} [\Delta^L(v)]$, with $\Delta^L(0) = \bar{Y}(Y \leq y_{\alpha_{CN}}^{000})$ and $\Delta^L(1) = \bar{Y}(Y \leq y_{1-\alpha_{CC}}^{000}) * \frac{\pi_{NN} + \pi_{CN}}{\pi_{CN}} - \bar{Y}^{100} * \frac{\pi_{NN}}{\pi_{CN}}$. Let $\mathbf{\Delta}^L = [\Delta^L(0), \Delta^L(1)]'$ be the vector containing the two bounding functions. I subsequently discuss the estimation of the lower bound along with its confidence region (the proceeding for the upper bound is analogous). I use the procedure of Chernozhukov, Lee, and Rosen (2013) to obtain a half-median-unbiased estimator of $\max_{v \in V} [\Delta^L(v)]$. This appendix is based on similar descriptions of this method in Chen and Flores (2015); Huber, Laffers, and Mellace (2017).

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 minus $s(v)$, a measure of the precision of $\hat{\Delta}^L(v)$, times an appropriate critical value $k(p)$:

$$\tilde{\Delta}^L(p) = \max_{v \in V} [\hat{\Delta}^L(v) - k(p) \cdot s(v)].$$

As outlined below, $k(p)$ is a function of the sample size and the estimated variance-covariance matrix of $\sqrt{n}(\hat{\Delta}^L - \Delta^L)$, 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 lower bound is below its true value with probability at least one half asymptotically.

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

1. Estimate the vector $\hat{\Delta}^L$ by its sample analog. Estimate its variance-covariance matrix $\hat{\Omega}$ by bootstrapping B times.²⁶
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 ²⁷ draws, $H_1, \dots, H_R$ 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 $r = 1, \dots, R$.
5. Let $\tilde{k}(c)$ be the c -th quantile of $\max_{v \in \mathcal{V}} H_r^*(v)$, $r = 1, \dots, R$, where $c = 1 - \frac{0.1}{\log(n)}$.
6. Compute the set estimator $\hat{\mathcal{V}} = \{v \in \mathcal{V} : \hat{\Delta}_{01}^L(v') \leq \max_{v' \in \mathcal{V}} [\hat{\Delta}^L(v') - \tilde{k}(c) \cdot \hat{s}(v')] - 2 \cdot \tilde{k}(c) \cdot \hat{s}(v')\}$.
7. Estimate the critical value $\hat{k}(p)$ by the p -th quantile of $\max_{v \in \hat{\mathcal{V}}} H_r^*(v)$, $r = 1, \dots, R$.
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. Estimate the lower bound Δ^L by $\hat{\Delta}^L = \tilde{\Delta}^L(\frac{1}{2})$.
10. To obtain the lower confidence band, estimate the half-median-unbiased lower bound $\tilde{\Delta}^L(p)$.
11. Let $\Gamma = \max(0, \tilde{\Delta}^U(\frac{1}{2}) - \tilde{\Delta}^L(\frac{1}{2}))$, where $\tilde{\Delta}^U$ is the upper bound derived from steps 1 to 9 in a symmetric way, $\rho = \max(\tilde{\Delta}^L(\frac{3}{4}) - \tilde{\Delta}^L(\frac{1}{4}), \tilde{\Delta}^U(\frac{3}{4}) - \tilde{\Delta}^U(\frac{1}{4}))$ and $\tau = (\rho \log(n))^{-1}$. Compute $\hat{a} = 1 - \Phi(\tau \cdot \Gamma)\alpha$, where α is the chosen confidence level.
12. The lower confidence band for the estimate of Δ^L is obtained by $\tilde{\Delta}^L(\hat{a})$.

²⁶In the empirical part I use 1,999 bootstrap replications.

²⁷I set $R=1,000,000$.