



**Improving the measurement and policy relevance of migration information  
in multi-topic household surveys\***

Alan de Brauw

Poverty, Health and Nutrition Division  
International Food Policy Research Institute

and

Calogero Carletto

Development Research Group  
World Bank

May 2012

---

\* Thanks to Carlo Azzarri, Kathleen Beegle, Adriana Castaldo, Robert E.B. Lucas, David McKenzie, J. Edward Taylor, and Dean Yang for comments on earlier drafts of this manuscript.

## TABLE OF CONTENTS

<b>1. INTRODUCTION .....</b>	<b>3</b>
<b>2. DEFINING AND MEASURING MIGRATION .....</b>	<b>8</b>
2.1. WHO IS A “MIGRANT”? .....	8
2.2. TYPES OF MIGRATION.....	9
2.3. IDENTIFYING MIGRATION IN SURVEY DATA.....	11
2.4. KEY POINTS .....	13
<b>3. METHODOLOGICAL CONSIDERATIONS.....</b>	<b>15</b>
3.1 MIGRATION AS A RARE EVENT .....	15
3.1.1. <i>Choosing a Sampling Frame</i> .....	15
3.1.2. <i>Sample Design</i> .....	17
BOX 1. ALTERNATIVE METHODS OF SETTING UP TWO STAGE SAMPLING .....	20
3.1.3 <i>When to Consider Using Listing</i> .....	21
3.1.4. <i>Key Points</i> .....	21
3.2 MIGRANT ABSENCE FROM SURVEY LOCATION .....	22
3.2.1. <i>Direct versus Proxy Respondent</i> .....	22
3.2.2. <i>Issues related to Extended Household Membership</i> .....	23
3.2.3. <i>Key Points</i> .....	24
3.3 MIGRATION IS NOT A RANDOM PROCESS .....	25
3.3.1. <i>Strategy 1: Attempt to Ignore the Bias</i> .....	26
3.3.2. <i>Strategy 2: Panel Data</i> .....	27
3.3.3. <i>Strategy 3: Use Matching Methods</i> .....	28
3.3.4. <i>Strategy 4: Instrumental Variables</i> .....	29
3.3.5. <i>How Does the Endogeneity of Migration Affect Study or Questionnaire Design?</i> .....	32
3.3.6. <i>Key Points</i> .....	34
<b>4. QUESTIONNAIRE DESIGN.....</b>	<b>35</b>
4.1. COLLECTING INFORMATION ABOUT MIGRATION EXPERIENCE.....	35
4.1.1. <i>Identifying Migrants Through Screening Questions</i> .....	36
4.1.2. <i>Migration History Module: Tradeoffs</i> .....	38
4.1.3. <i>Designing the Migration History</i> .....	40
4.1.6. <i>Key Points</i> .....	45
BOX 2. TRACKING MIGRANTS OR HOUSEHOLDS MIGRANTS LEFT.....	47
4.2. MEASURING DETERMINANTS OF EMIGRATION .....	48
4.2.1. <i>Demographics and Human Capital Characteristics Determining Migration</i> .....	50
4.2.2. <i>Differences in wage rates between the source and potential destinations</i> .....	50
4.2.3. <i>Migration Networks and Information</i> .....	51
4.2.4. <i>Spreading out Risk through Migration</i> .....	53
4.2.5. <i>Relative Deprivation or Inequality</i> .....	54
4.2.6. <i>Policies Inhibiting or Promoting Migration, at Source or Destination</i> .....	55
4.2.7. <i>Reuniting with Family</i> .....	56
4.2.8. <i>Educational Facilities or Other Local Public Goods</i> .....	57
4.2.9. <i>Violence, Disease, Environmental Disasters or Other Factors</i> .....	58
4.3. MIGRATION, IMPACTS, AND POLICY—SPECIFIC DATA REQUIREMENTS .....	58
4.3.1. <i>Poverty and Income Distribution</i> .....	59
4.3.2. <i>Migration and Employment</i> .....	60
4.3.3. <i>Human Capital Formation</i> .....	62
4.3.4. <i>Brain Drain, Gain, or Waste</i> .....	63
4.3.5. <i>Migration, Gender and Gender Related Policy</i> .....	65
4.3.6. <i>Social Implications of Migration</i> .....	66
4.3.7. <i>Policies Encouraging Return Migration</i> .....	68
4.3.8. <i>Migration and Social Protection</i> .....	69
4.3.9. <i>Migration, Infrastructure, and the Provision of Public Services</i> .....	70

<i>4.3.10. Migration and Household Investments</i> .....	72
<i>4.3.11. Migration and Environmental Displacement</i> .....	74
<b>5. CONCLUSION</b> .....	<b>75</b>
<b>FIGURE 1. DECISION TREE FOR COLLECTING INFORMATION THROUGH PROXY</b> .....	<b>76</b>
<b>REFERENCES</b> .....	<b>77</b>

## **Improving the measurement and policy relevance of migration information in multi-topic household surveys**

### **1. Introduction**

Although migration has been an important phenomenon shaping the demographic profile of countries for centuries, the past decade has seen migration rapidly rise to become a prominent feature of the world economy. According to the United Nations (2009), the number of people living outside their country of birth is now increasing faster than world population growth. Since 1975, the global stock of migrants has more than doubled to more than 213 million, representing approximately 3.1 percent of the world's population. Corresponding to the increase in international migration, remittances from migrants residing in developed countries to households in developing countries have increased significantly. Formal remittance flows – which grossly underestimate total remittances – totaled \$328 billion in 2008 and were second only to foreign direct investment (FDI) as a capital flow into developing countries (IFAD, 2007; Ratha et al., 2009). For many developing countries, remittance flows are larger than FDI and Official Development Assistance (ODA) combined, and in at least 42 countries migrant remittances account for 10 percent or more of GDP (IFAD, 2007; World Bank, 2011).

As international migration and migrant remittances have increased, migration has once again emerged as a contentious policy issue in both developed and developing countries (Global Forum on Migration and Development, 2007; Massey, 1999). Policies to further liberalize migration flows into developed countries often face strident opposition from several fronts. Proponents of liberalization tend to argue that increased migration offers opportunities to advance the economic development of both sending and receiving countries. Meanwhile, critics emphasize economic, political, and social drawbacks to increased population movements. These issues are relevant in regards to both migration from the global south to the north, as well as migration within the global south itself. Despite the strongly held opinions on both sides of the migration policy debate, little solid empirical information exists on the ways in which various policies affect migration and migrant households.

The resultant need to better understand the determinants and impact of migration has generated a growing demand for better migration data, as reflected in the emergence of new data collection efforts. While existing data on migration and remittance flows is partial and imperfect, there are nonetheless a number of sources that currently produce information on these topics. Formal international remittance flows are often reported by central banks, and the number of legal international immigrants in a given country is tracked by national governments through

various sources, including border and other administrative records. Estimates of internal migration can often be obtained from population censuses. However, these sources are inadequate for a number of reasons. Aside from failing to distinguish and account for various types of migration (for example, international emigration, including long-term, circular and seasonal migration), sources such as these do not typically include the types of variables that can be used to understand the determinants and impact of migration.

Labor force surveys and specialized migration surveys also fall short of including sufficient information to study the determinants and impact of migration. The International Labor Organization has recently developed specific modules focused on international migration to add to labor force surveys, such as the quarterly labor force survey conducted in Thailand (2007), which asks questions about international origins of workers. Specialized migration surveys such as the Migration from Africa to Europe Project (MAFE) can also be used to analyze issues related to migration. The MAFE surveys include a number of socio-economic and welfare indicators at both the individual and household level, in addition to migration histories, employment histories, residence permit status, and other issues. However, both of these types of surveys are primarily limited to simply characterizing migrants and the driving forces behind migration. As minimal information on topics beyond migration is collected in these surveys, they often offer little insight into the relationship between migration, sources of income and welfare outcomes. Most significantly, these surveys seldom contain the information necessary to construct an acceptable welfare measure, thus precluding the use of the survey for any poverty or distributional analysis. Furthermore, the lack of information on other sources of income limits the analytical use of these surveys for understanding the role of migration in household livelihoods. Finally, specialized surveys are often too costly for most developing countries to carry out on a large scale and sustainable basis.

Collecting information about migration as part of a larger multi-purpose data collection effort represents a good opportunity to learn more about migration in a cost effective manner. However, as discussed by Lucas (2000), collecting detailed information on migration has not traditionally been a priority for most household surveys, including the Living Standards Measurement Study (LSMS) surveys. Nonetheless, in response to the strong demand for more high-quality migration data, an increasing number of LSMS surveys now include extensive modules on different types of migration, including the latest LSMS surveys carried out in Albania and Tajikistan.

Although at irregular frequencies, most countries now implement nationally representative, multi-topic household surveys *à la* LSMS for poverty monitoring and analysis.

LSMS surveys are based on multi-topic questionnaires designed to study numerous aspects of household welfare and behavior, and integrating migration information into this data collection effort can be an efficient way to collect migration data. The extent of this integration will depend on a number of reasons, including the perceived importance of the subject by the survey implementers and policymakers, as well as the necessary trade-offs dictated by the multi-purpose nature of the instrument. Using LSMS surveys to study migration offers specific advantages; most importantly, one can analyze the relationship between migration and a number of variables to a far greater degree than would be possible with other types of surveys such as a Household Budget Survey (HBS) or a Labor Force Survey (LFS). The main drawback of using LSMS-style surveys to study migration is that migrant populations are often too small relative to the entire population and tend to be clustered. Consequently, a typical multi-stage cluster sample of a multi-topic survey like an LSMS may not be able to capture a sufficient number of migrants to enable analysts to make accurate statistical inferences about those sub-populations. Additionally, as mentioned, multi-topic surveys are designed to cover a large number of issues, resulting in potential compromises in terms of the level of depth with which these issues can be explored. Despite these limitations, the increased insight that multi-topic surveys can provide into the complex interaction of migration with issues of policy and welfare makes them a worthwhile method for the study of migration in developing country contexts.

This paper discusses how LSMS surveys can be adapted in order to study a range of issues surrounding migration. To date, the LSMS has been under-utilized as a tool to study migration; only a handful of previous LSMS surveys have included detailed migration modules regarding current household members or previous household members who have out-migrated. There are several reasons that migration may have been overlooked in previous surveys. First, migration has only recently increased in prominence, and as a result, it may not have previously seemed to be a high-priority topic for inclusion in a questionnaire. Second, migration is a rare event from a statistical perspective, and the relatively small sample size of most LSMS surveys often makes them unsuitable for the study of migration. Third, as migration is not a random event, there are methodological issues with identifying causal relationships between welfare and migration that may need to be addressed by making additional changes to questionnaire content. However, even if causal relationships cannot be identified, multi-topic household surveys can be used to further the understanding of correlations between migration and policy-relevant variables not well-understood or established prior to the survey.

The main objective of this chapter is therefore to provide basic guidelines to researchers interested in studying migration for collecting migration information as part of a multi- topic

household survey. We begin with a brief discussion of definitional and measurement issues with regards to various types of migration. We then address key methodological considerations that should be taken into account, focusing particularly on overcoming the rare aspects of migration through alternative sampling designs. Lastly, we consider the measurement of migration and other variables of interest from the perspective of questionnaire design, acknowledging the necessary trade-offs between length, content, and accuracy inherent to the implementation of a multi- topic LSMS- style survey.

It is important to note that all of the policy issues related to migration and its potential impacts discussed in the section on questionnaire design can apply to both internal and international migration. The main difference between internal and international migration can be thought of as a difference in costs and benefits, as internal migration is significantly less costly than international migration, but also often bears lower returns. Migrants who leave their country of origin legally must have passports, which often cost 10% of GDP per capita or more (McKenzie, 2007). In many cases, migrants must also have a plane ticket and a visa to migrate legally. All of these formalities combine to make an initial international migration quite costly. If individuals choose to or must migrate illegally to their intended destination, they may incur other risks or costs, or may be forced to pay a broker to facilitate their move (e.g. Dolfin and Genicot, 2006; Hanson, 2007). As a result, international migration typically excludes the poorest (e.g. Chiswick and Hatton, 2003). Moving to another country can also be costly in other ways; individuals who move must adapt to the culture and language of their destination country. In contrast, although it may involve a change from living in a small village to a large metropolis with different customs and attitudes, internal migration typically does not involve as many non-monetary costs, unless undertaken by ethnic minorities or in countries with restrictions to internal migration.<sup>1</sup>

Since the up-front costs of migration differ so greatly for internal relative to international migration, it is reasonable to expect the relative return to labor in international migration to be considerably higher than internal migration. As a result, the types of individuals undertaking internal versus international migration may differ significantly, and therefore international and internal migration should be differentiated clearly within a survey and analyzed separately.<sup>2</sup>

It is also worth mentioning that our focus is on migration and its impacts on source households and communities. Although many of the expected impacts of migration may flow

---

<sup>1</sup> The best studied example of internal migration restrictions is China's *hukou* system, which was relaxed throughout the 1990s and early 2000s (Mallee, 1995).

<sup>2</sup> Furthermore, internal migration was largely covered by Lucas (2000).

through remittances, we do not cover them in this paper. Therefore, when we consider the effects of migration on households in this paper, we refer either to the effects of the individual leaving the household or the net effects of the individual leaving and any remittances they may send back. However, by design, we do not cover the collection of information on remittances.

## 2. Defining and Measuring Migration

Unlike mortality and fertility, migration is not related to a tangible biological occurrence, making it one of the more difficult demographic phenomena to measure. Quantifying this inherently subjective concept is further complicated by the fact that there are several modes of migration, and that these modes of migration can in turn be partially determined by the varying motivations for migration (Zlotnick, 1987a, 1987b).

### 2.1. WHO IS A “MIGRANT”?

In order to properly quantify migration within the context of a household survey, one must first identify the specific characteristics that define a migrant. Although no consensus currently exists on the exact definition of a migrant in its many forms, five individual characteristics can help analysts to determine which individuals should be considered migrants. These characteristics are: (i) the place of birth; (ii) whether or not the individual resides in the place of birth; (iii) household membership; (iv) the duration of any stays away from the residence; and (v) a time period of reference. From the perspective of a household survey, in some contexts it may be important to measure all five of these characteristics in order to be able to classify survey respondents accurately and consistently.

Depending upon their relationship with the community being surveyed, migrants can either be classified as *immigrants* or *emigrants*. Individuals who were originally located elsewhere but who now reside in the surveyed community are defined as immigrants. Individuals who originally resided in the surveyed community but are now located elsewhere are defined as emigrants. The concept of location in this case refers to the local geographic area, which can depend upon context, but might best be considered the primary sampling unit used for the survey. With regards to international migration, individuals with citizenship in another country who were not born in the destination country are defined as immigrants. Emigrants who remain within the country where the survey is being conducted are defined as *internal emigrants*; those who now reside elsewhere in the world are known as *international emigrants*.

Immigrants can be readily identified within household surveys and censuses simply by asking about the country of birth and/or citizenship, since these surveys typically enumerate all household members currently residing within the surveyed community. Meanwhile, emigrants associated with the household are more difficult to identify and are generally not enumerated. As

a result, although it is crucial to understanding the relationship between migration and policy, emigration is generally more difficult to quantify using household survey data or censuses. Generally, households in surveys (and censuses) are defined on the basis of the individuals who eat together and sleep under the same roof on a regular basis.<sup>3</sup> For the purposes of measuring and analyzing migration, this definition lends itself well to the study of migration experiences among the surveyed population, including the prevalence of immigration. However, in order to study the impact of emigration, the survey needs to ask a range of questions regarding a group of individuals who are not resident at the time of the survey. In this context, the standard definition of a household member is less useful, as it excludes individuals who maintain ties to the household but clearly do not eat with other household members nor regularly sleep in the household. Therefore, when emigration is a focus of the survey, it becomes necessary to expand the definition of household members to include information about additional individuals who do not fit into the primary definition of the household.

## 2.2. TYPES OF MIGRATION

There are several types of migration that must be accounted for when conducting a survey on the topic. More specifically, it is necessary to define the various forms that migration can take and to delineate the potential rules that can be used to differentiate them from one another.

Short-term (*temporary*) and long-term (*permanent*) migrants may have different motivations and as a result the impact of each type of migration may differ. Since temporary and permanent migration may have different determinants, each type of migration may also differ in its effect upon the households left behind (i.e. the *source households*). Beyond this, either type of decision to migrate is not irreversible. A temporary migrant may decide to overstay and remain permanently in the host country or, vice versa, a permanent or long-term migrant may decide to go back home. Migrants may also leave seasonally; *seasonal* migrants are similar to temporary migrants, but a seasonal migrant leaves the household for a short period of time annually at the same time of year. The implications of seasonal migration for households can also differ from temporary migration, and therefore it is worth studying separately in a number of contexts.

When migration has taken place for some time from a geographic area, it is also important to consider identifying *return migrants*, who can potentially catalyze development at the origin

---

<sup>3</sup> Additional rules are generally imposed to more clearly define household membership, including a duration of residence rule, e.g. usually, that all household members must have resided in the household for a minimum of 3 to 6 months over the previous 12 months. Special provisions are generally made for the household head, newborns and new household members, such as those who marry into the household.

through the use of skills or capital that they obtained as migrants. Return migrants must in turn be distinguished from *circular migrants*, who are similar in that they may be resident in the household at the time of the survey, but differ in that they either come and go seasonally or migrate repeatedly, leaving every few years.

In household surveys, these concepts can be difficult to delineate and measure clearly. Temporary migration and permanent migration are usually distinguished by what can be considered an arbitrary threshold. For example, the United Nations (1998) uses a threshold of twelve months to separate temporary from permanent migrants. Return migration, seasonal migration, and circular migration can also be difficult to distinguish from one another. Seasonal or circular migrants can be mistakenly categorized as return migrants if the survey form is not specifically designed to capture repeat instances of migration either within the calendar year, in the case of seasonal migrants, or over several previous years, which would be needed to study circular migrants. Still others who have recently returned from a long migration spell might not be truly circular migrants, but may have the intention or willingness to migrate again if the right opportunity arose. These distinctions could eventually result in differing policy outcomes – for example, migrants who have returned temporarily or have not fully committed to their return might be less likely to make investments at home.

Differentiating migrant *stocks*, or the population present or missing from a country or area at a given time, from migration *flows*, or the number of individuals leaving or returning to a country or area in a given period of time, is also important for measurement purposes. There are complications in measuring both of these phenomena. In regards to migration stocks, if a multi-topic household survey is used to estimate the international emigrant stock, one might count all emigrants from households in the sample, yet some of these emigrants may already have returned home to live in a different household and/or an urban area. A survey that attempts to estimate the emigrant stocks in this manner is therefore likely to double count migrants, as the same out-migrant may be claimed by two or more households in the sample, leading to an overestimate of out-migration. Migration flows are equally difficult to measure accurately. Within a specific time period, there can be a difference between the gross migration flow (including multiple migration spells for one migrant) and the net migration flow. If care is not taken to measure coming and going during the time period in question, net migration flows can miss instances of migration by seasonal or temporary migrants.

Although our discussion of migration so far has implicitly assumed international migration, *internal* migration is usually an even more common phenomenon. As an economy develops, migration naturally occurs from rural to urban areas of a country. Urbanization has been an

inevitable part of the development process ever since the Industrial Revolution (e.g. Williamson, 1988; Taylor and Martin, 2001). Moreover, there is a distinct lack of high quality, nationally representative data on internal migration (Foster and Rosenzweig, 2008). One conceptual difficulty with measuring internal migration is that the concept of “urban” can be difficult to define from the perspective of a survey designed to learn about internal migration. Its definition differs from country to country and may also differ within the same country across time. As a result, defining rural-urban migration consistently across countries over a period of time may well be impossible. Extending the classification to include peri-urban areas may be more relevant to study internal migration, as most rural migrants tend to concentrate in the outskirts of large cities where environmental conditions and access to services may be somewhat different than in an urban setting.<sup>4</sup>

Finally, a large number of international migrants are *undocumented*, making them more difficult to identify in survey data, in part because proxy respondents may not want to discuss their whereabouts or any remittances they receive from undocumented migrants in fear of negative repercussions for the migrant (Heckmann, 2004; Massey and Capoferro, 2004). Similarly, surveys collecting information about immigrants may miss undocumented immigrants, either because undocumented immigrants are less likely to appear in the sampling frame or because they are less likely to accept being interviewed if they are selected. From the perspective of emigration, it is a worthwhile investment to train enumerators to first make respondents comfortable before asking sensitive questions about potential undocumented migration.<sup>5</sup>

### 2.3. IDENTIFYING MIGRATION IN SURVEY DATA

An LSMS survey designed to study migration must gather information on past or current migration events for all household members, as well as, even if only partially, for former household members and for some individuals who fail to qualify as household members based on the residency rule adopted by the survey. One potential problem for studying migration is that such residency rules are often too stringent, resulting in the exclusion of individuals of interest. Alternatively, collecting information about individuals who are away or are no longer members of

---

<sup>4</sup> The main problem is that countries do not differentiate peri-urban areas in their classification of census enumeration areas. Based on some experimental work carried out by the Albania Institute of Statistics (INSTAT), the 2005 Albania LSMS contained an oversample of peri-urban areas from the city of Tirana, the capital (INSTAT, 2005).

<sup>5</sup> Understanding attitudes about undocumented emigration or immigration is also helpful when weighing the sensitivity of the issue. For example, almost all emigration from Guatemala is undocumented and as such there is no social stigma against answering questions about undocumented emigration. On the other hand, families of emigrants from Eastern Europe may find the issue particularly sensitive.

the household may be difficult and ultimately of little use, particularly for individuals who left the household long ago. The main trade-off faced by the analyst relates to collecting enough information about migrants to ensure that information about people connected to the household is included, while avoiding obtaining inaccurate information and significantly extending the length of the interview.

To determine individuals for whom additional information should be enumerated, a first step is to consider relaxing the residency rule, so as to include more individuals in the main questionnaire. For example, one may want to include all individuals who have lived in the household in the past twelve months in the roster, and gather information on the months of presence over that period. Collecting information on more individuals, including the months of residence, would then enable the analyst to adjust the definition *ex-post* according to the specific objective of the analysis. Once a more exhaustive list has been created in the roster, a rule must be established regarding the individuals for whom the full questionnaire should be administered and the individuals for whom one should ask a limited number of questions through proxy respondents. In summary, the survey must strike a balance between attempting to ask for more information than can be accurately recalled by proxy respondents, and asking for too little information, meaning that relationships with previous members or members temporarily away would be missing. The key is to achieve this balance while maintaining the ability to calculate household size based on the national criteria of household membership.

One must next decide how to broadly identify migrants, with an eye to identifying both migrants who are currently away as well as migrants who left the household in the past and have now returned. To categorize individuals either as migrants or non-migrants, one should use a residency rule specific to migration in order to distinguish the household members who are to be considered migrants for the purposes of the survey. For example, if the rule is an absence of three months, members of the household who have been absent for more than three months for reasons other than health or family visits would be considered migrant household members, whether or not they were considered household members by the residency rule. Meanwhile, individuals who were away for a shorter period of time might not be considered migrants.

For return migration one would be interested in knowing about all current household members with past migration experience over a given period, and particular to international migration, all former household members with past international migration experience who now live in the source country. If migrants have returned to the household, it is also important to include a line of questioning to learn about whether they intend to migrate again, or if they have migrated for short periods repeatedly in the past, in order to determine whether they would be

better categorized as seasonal or circular migrants. There are challenges in collecting information on each of these broader groups, since long recall periods can create errors in the memories of migration experience among return migrants, whereas information on current out-migrants or return migrants living elsewhere must be collected from a proxy respondent and is therefore subject to its own set of problems.

Individuals who can be considered migrants may have left the household 20 or 30 years earlier or, conversely, they may be a member of that sub-group of individuals who were excluded because they did not pass the residency restriction. The amount of time since an individual left the household can be a key factor in determining which individuals are to be considered migrants from the perspective of a survey. In specifying criteria for individuals considered to be out-migrants, one must further decide whether all former household members should be included (i.e. any individual who lived in the household at any point in time) or whether migrants should be restricted to nuclear family members (i.e. referring exclusively to the sons, daughters, and spouse of the household head). Including all former household members may result in the aggregate double counting of migrants, and may lead to greater inaccuracies in respondents' self-reported definition of household membership. On the other hand, restricting attention to nuclear family members may improve accuracy, but could lead to underestimates of total migration.

#### 2.4. KEY POINTS

Migration is the one of the most difficult demographic phenomena to study within a household survey context. More specifically, while a nationally representative sample with a good sampling framework is sufficient to collect information about immigration, in order to enumerate information about both immigration and emigration, it is important to:

- Learn about immigration, ask about the place of birth of all current household members, and include a module asking about the migration history of each member.
- Include a survey module that collects limited information about additional individuals potentially associated with the household who are not present, but were present in the past.
- Collect data on the following five characteristics of all individuals in the household roster, so everyone can be characterized as either migrants or non-migrants:
  - The place of birth;
  - Whether or not the individual resides in the place of birth;
  - Household membership by an expansive definition of the household;
  - The duration of any stays away from the residence; and

- A time period of reference over which these stays might have occurred.
- In designing the survey and questionnaire, be cognizant of the types of migration that are frequent in the specific context:
  - Short-term (*temporary*) and long-term (*permanent*) migrants, which are separated by some time criterion determined by the analyst;
  - *Seasonal* migrants, who leave for a specified period of time each year and should be identified through questions about repeated, short migration spells;
  - *Return* migrants, who had migrated at some time in the past and have returned to the country or household somewhat permanently; and
  - *Circular* migrants, who have returned but plan to leave again for a significant period of time, or migrate repeatedly for long spells. They are differentiated from seasonal migrants by the duration and irregularity of their migration spells.

### 3. Methodological Considerations

Migration is a complex demographic phenomenon, and therefore designing a nationally representative, multi-topic survey to learn about the motivations for migration, the effects of migration, and/or the interactions between migration, policy, and outcomes is not trivial. Before we consider the data requirements for analyzing different types of migration and the interaction of migration with specific policies and outcomes, we discuss three issues related to questionnaire design that are important to consider when designing the study. We first describe the sample frame and design, then issues that arise from the fact that certain types of migrants are more likely to be absent from the household at the time of the interview, and finally the implications for survey design due to the non-random nature of migration.

#### 3.1 MIGRATION AS A RARE EVENT

A methodological consideration that must be taken into account when incorporating a migration component into a multi-topic survey arises from the fact that migration is a *rare event*. Rare events are defined as statistical occurrences that happen infrequently; in this case, we mean that in a random selection of any household for our sample in a particular country, the probability that the household has a migrant is close to zero.<sup>6</sup> As a result, in a normal clustered sample design typical of multi-topic surveys, the expected number of households associated with emigration may be very low. Furthermore, in light of the likely geographic clustering of migration, there is a high chance that the sample may miss these areas altogether. In order to counter this lack of prevalence, it is important to consider various strategies for alternative sampling frames and alternative sample designs that can account for this issue.

##### 3.1.1. Choosing a Sampling Frame

A standard LSMS survey produces data on the distribution of living standards within a country using a nationally representative sample. The typical sample design of an LSMS is a multi-stage cluster design, usually involving two or more stages, and possible stratification based on some administrative or location subdivision. In the case of a two-stage design, primary sampling units

---

<sup>6</sup> Note that the rarity of the event will also depend upon the definition adopted and the type of migrant being studied. Here we generally refer to a “migrant” as an international emigrant associated with the household, who may have left the household within a designated time period. Other types of migration (such as internal migration) might not be as infrequent, particularly when considering the stock of migrants rather than the flow.

(PSUs) are selected in the first stage – also referred to as Enumeration Areas (EAs), these depend on the local context and may constitute villages, sub-villages, urban neighborhoods or even partitions of neighborhoods that have well-marked boundaries, as established by the national statistical agency. Households are then randomly selected within each PSU in the second stage. If each PSU is assigned a non-zero probability of selection in the first stage and there is an accurate household listing of the selected PSUs, one can then use the sample to make inferences about the entire population. In the context of migration, however, the problem is that aside from a few countries with an extremely high prevalence of migration and/or a very large sample size, a sample drawn in this fashion is unlikely to contain a sufficiently large number of migrants. This prevalence will obviously be affected by the chosen migrant definition but, in most cases, irrespective of the definition adopted, a random cluster sample is unlikely to meet the demands of the research.<sup>7</sup> Theoretically, one could simply increase the planned sample size to ensure that there are enough migrants for statistical analysis by either increasing the number of PSUs, the number of households in each PSU, or both. However, it must be emphasized that in most cases, this is not a cost-effective option. Furthermore, beyond the increase in monetary costs, the implications in terms of data quality and measurement errors can be extremely taxing. Given these various issues, alternative sampling designs are a preferable method of ensuring sufficient variation in the characteristics of migrants (or impacts) to come to meaningful conclusions.

One reasonable alternative is to use some other type of probability sampling when choosing PSUs for an LSMS survey with a focus on migration. With this option, households in each of the selected PSU can be selected according to some pre-established stratification assigning migrant households a higher probability of being selected. Methods for oversampling migrant households in each stage of selection, if feasible and properly implemented, can provide an adequate probability sample to study migration within the context of a traditional multi-topic survey. Unfortunately, the sampling frames generally used for drawing the sample, primarily the most recent population census, do not contain any information to allow this type of disproportionate sampling. In the best case scenario, the population census will contain information for the identification of immigrants, but the collection of information on emigrants is extremely rare. Alternative probability sampling techniques must then be adopted to collect information on a sufficiently large number of migrants in a traditional LSMS survey.

---

<sup>7</sup> If projected migration rates are higher, this concern is abated for analyzing the stock of emigrants, though it may still hold for the flow. Some obvious exceptions include international migration from small countries that have a large proportion of their populations out-migrating, such as many Caribbean and Central American countries and small island countries in the South Pacific. Internal migration rates are often much higher than international migration rates, particularly in large countries (Lucas, 1993; Foster and Rosenzweig, 2008).

### 3.1.2. Sample Design

While the literature suggests a number of techniques to better identify rare events such as migration (e.g. Kish, 1965), two alternative sampling designs are particularly appropriate for the study of migration: 1) disproportionate sampling of high migration PSUs and 2) stratified random sampling within PSUs, also known as two-phase sampling. These two methods can either be used individually or in conjunction with one another.

A disproportionate sampling design implies that PSUs with higher migration rates are identified prior to the survey and oversampled. In other words, PSUs known to have a high rate of emigration would be allocated a higher probability of selection as compared to PSUs with lower rates of migration. Representativeness would be regained through weighting. The first challenge to implement such a technique is to have an adequate sampling frame allowing for stratification based on the incidence of migration. Unfortunately, as mentioned, this is notoriously difficult to accomplish, as population censuses do not collect this information. When the information is not readily available in the frame itself, alternative methods can be attempted in order to identify high migration areas<sup>8</sup>.

In addition to, or in lieu of, adopting disproportionate stratified sampling in the first stage of PSU selection, one may also implement some form of disproportionate selection of households within each of the selected PSUs. Even in relatively high migration areas, the prevalence of migration is unlikely to be so high that a random draw of households will be an efficient way to select a sufficiently large number of migrant households. This method, commonly referred to as two-phase sampling, requires a full listing operation in each selected PSU, which must collect information on the migration variable of interest (as defined by the study) to enable the oversampling of migrant households. By carrying out a listing operation, in which information on migrant status is elicited either directly from the household or through proxy respondents, one can clearly identify migrant households so as to select a sample with more migrant households. A household survey conducted in Guatemala used this approach, canvassing selected communities to identify migrant households with the help of a short census-like questionnaire. In this case, the listing operation was part of a broader community census used by a large regional food security project to identify and characterize areas at high risk of food insecurity (Covarrubias and Carletto, 2009). When using this method, it is crucial that the probability of selecting a migrant household is explicitly known, so that weights can be constructed and any

---

<sup>8</sup> One additional consideration to keep in mind when using disproportionate stratified sampling is that one must assume that the determinants in high migration areas are similar to the determinants in low migration areas, which could be a problem, particularly if the final sample ends up omitting most low migration areas.

econometric estimation of the determinants of migration remains consistent (see Cameron and Trivedi, 2005, Chapter 24 for a discussion). Relative to an overall survey budget, listing operations are not very expensive and ultimately may be the most cost effective way to identify migrant households (Muñoz, 2007). One could also use alternative methods of learning about the prevalence of migration among the population prior to using disproportionate sampling (see Box 1 below for an example).

As noted, to use the alternative sampling methods described above, it is important to have information about the prevalence of migration either prior to sampling PSUs or within sampled PSUs after they have been sampled. Ideally, this information could come from a census or a similar large survey; however, if emigration is being studied, the existence of such information is unlikely, because censuses rarely include information about migration.<sup>9</sup> Assuming that such information is not available, one could use alternative sources to learn about migration prevalence. These sources might include expert opinions, qualitative surveys, or surveys in destination areas that contain information about the specific location from which the migrant departed. However, there are a number of potential drawbacks to this method. Firstly, without statistical information on migration to design the sampling frame, it is impossible to come up with correct sampling weights. Furthermore, if one uses destination surveys to learn about migration in source areas, one must be aware that unless all the migrants in the country migrated to the destinations for which one has data, migrants who migrated to other destinations are likely to be poorly represented in the sampling frame. Lastly, by relying on anecdotal or qualitative information rather than on statistical information to choose a sample, one might be concerned that the resulting sample of migrants will not be representative of the target migrant population. A potential alternative could be to apply Small Area Estimation (SAE) techniques to estimate the prevalence of migration by combining survey and census information.<sup>10</sup>

Assuming that migrants can be properly identified in the sampling frame, a further decision must be made as to whether to select based on the proportion of migrants over the population in the reference area or, conversely, based on the proportion of households with migrants out of the total number of households. Since much of the analysis on migration is done at the household level, the second option will likely be preferable.<sup>11</sup> A third option is to base

---

<sup>9</sup> A notable exception is El Salvador, which asked about migration in its 2007 census. Between 2008 and 2010, more countries have begun to include a section on emigrants (household members who moved abroad in the past ten years) in their census questionnaires, including the 2010 censuses of Bermuda and the Bahamas.

<sup>10</sup> For more information about SAE, visit <http://go.worldbank.org/9CYUFEUQ30>.

<sup>11</sup> For an example of a three-stage disproportionate sample of immigrants using a suitable sampling frame, see Chapter 3 of Bilsborrow et al. (1997).

selection on the proportion of households with household heads as a migrant (Bilborrow et al, 1997).<sup>12</sup>

Aside from using disproportionate sampling as part of the actual LSMS sample, the possibility of drawing a booster sample – that is, an oversample – of migrants beyond the original sample households should also be considered. To save on fieldwork costs, the booster sample can consist of migrant households from the same or adjacent EAs that have previously been identified as high migration areas. An added advantage of the use of booster samples is that a large survey also provides a large control group of households for possible matching.

Ideally, for identifying high migration areas or a migrant household a listing operation would not be necessary, if the census asked some basic questions about household migration status. Unfortunately, although immigration status is usually recorded, this is rarely the case for emigration status of current or former household members. Adding a few simple questions in future censuses would allow the census to be used to determine high migration areas and/or migration households.<sup>13</sup> For example, including the following questions in a future census would provide analysts with a sampling frame to study migration:

- Has any current member of the household migrated abroad over the past 5 (10) years?
- Is any former household member presently living abroad?  
and/or
- Are any of the children of the household head presently living abroad?

---

<sup>12</sup> These procedures are similar to those used by the Mexican Migration Project, which purposefully selected high migration areas for its sample. It is worth noting that estimates of migration stocks and flows coming out of such surveys are not representative without careful weighting of the sample.

<sup>13</sup> It is also important to note that the definition of migration may differ from country to country as well, which might affect the viability of using such questions for designing the sample frame.

### Box 1. Alternative Methods of Setting up Two Stage Sampling

To select rare events, other non-probability sampling techniques may be used; for example, multiplicity methods such as snowballing have been widely used in the migration literature. One use of snowballing gathers information on undocumented migrants, using as a starting point or a 'seed' a list of members of a diaspora organization or a list of migrants assisted by an NGO in destination countries. The 'seed' household is used to identify additional migrant households of the same country of origin, and so on until the necessary number of observations is reached. Since snowball sampling does not generally lead to a representative sample, it is useful primarily for description but not for statistical inference. A variant of snowball sampling is "Respondent Driven Sampling," developed by Heckathorn (1997; 2002). The issue with respondent driven sampling is that something must be known about the population in order to properly weight it to mimic a representative sample. If the migrant population is not accurately represented in the secondary information source, then respondent driven sampling is not a panacea either.

Techniques such as random walks using selected households in a community as starting points can also be used to identify additional rare events. A recent survey of the Nikkei population in Brazil used the aggregation point intercept method (McKenzie and Mistiaen, 2009). They found that although the method led to a higher estimated population than the stratified random sample, the estimates became close after reweighting to account for individuals who appear in the sample multiple times. In all cases when using these non-probabilistic methods, it is crucial to collect ancillary information on the implementation of the sample to be able to identify the reference population in an attempt to make 'educated inferences' about a larger population group.

It should be noted that if immigration rather than emigration is a major goal of the survey, disproportionate sampling based on census information may be an effective strategy of ensuring that there is enough coverage of immigrant households in the sample. Immigrant households are also likely to be infrequent in developing countries, making an alternative sampling strategy necessary. Since immigrant status is often asked in a census or another typical sampling frame, one would have information about the proportion of households including immigrants in each primary sampling unit, making disproportionate sampling feasible. As migration between countries in the South (South-South migration) is increasing, perhaps as rapidly as North-South migration (Ratha and Shaw, 2007), immigration may become a more frequent target in future LSMS-type surveys.<sup>14</sup>

---

<sup>14</sup> One might also be interested in the urban component of rural-urban migration within a country. Disproportionate sampling of urban primary sampling units would be the parallel that would provide more information about migrants living in urban areas.

### *3.1.3 When to Consider Using Listing*

While listing operations are not expensive in terms of field cost, careful decisions must be made in regards to the inclusion of relevant questions, whether a listing operation is to be conducted in parallel or sequentially with fieldwork, and other important methodological issues. If a census or a listing operation has been completed within the past 6 to 12 months, it may be possible to avoid conducting a new listing operation entirely, as long as the prevalence of migration in the population is sufficiently high.

Consider a survey with a target for the number of households with migrants to be included. The target could be reached by simply adding households to the overall sample until the expected number of migrant households was equal to the target. Alternatively, one could over-sample migrant households through a listing operation. The tradeoff between the two can be illustrated by a consideration of the cost ratio between performing a listing operation and adding identified migrant households, relative to the total survey costs if households are simply added to the survey to include a target number of migrant households. When the ratio is equal to 1, the analyst is indifferent between performing a listing operation and increasing the total sample size because the costs may be similar. Conditional on the cost of the listing operation, the benefit of increasing the total sample size increases relative to conducting a listing operation to inform the sample as the expected proportion of migrants in the sample increases. When the listing operation is cheap relative to surveying additional households, the listing operation makes sense even if migration is quite prevalent. On the other hand, if local conditions are such that listing operations are relatively expensive, it may make more sense at lower rates of migration prevalence to simply increase the sample size to include the desired number of migrant households in the sample.

### *3.1.4. Key Points*

Because migration is a rare event, one must carefully consider the construction of the sample for an LSMS survey before embarking upon questionnaire design. Some important points to keep in mind regarding sample construction are as follows:

- The use of a standard sampling frame is not recommended as it is likely to lead to a lack of migrant households in the resulting data set. However, if the goal of the survey is to

measure the stock or the flow of migrants rather than assessing the determinants of migration or its impacts, one should not change the sample frame.

- We advocate considering an alternative sampling method, in order to place additional weight on the probability of selecting migrant households; if weights are known, the resulting sample can be re-weighted to ensure representativeness.
- Listing operations that include information about migration can be used to help oversample migrant households, ensuring that enough migrant households are found in the sample to make statistical inferences.

### 3.2 MIGRANT ABSENCE FROM SURVEY LOCATION

A second methodological consideration that affects survey design and questionnaire content is that the migrant, almost by definition, is likely to be absent from the household at the time of the interview. It is therefore necessary to learn about migrants via the information of proxy respondents. Migrants are also likely to fail the residency rule, as the definition of household membership used in the survey for estimating per capita consumption usually excludes migrants. Therefore, it is important to think through what can be learned about additional individuals related to the household who do not reside there, as some may be considered migrants.

#### 3.2.1. *Direct versus Proxy Respondent*

Even if a strategy for sampling enough migrant households has been chosen, another set of issues revolves around *who* in the household answers questions about migration. Ideally, one wants to ask the migrant or the return migrant about his or her personal migration experience. The best quality information will be elicited when the migrant is asked. However, when considering migration in the present period, the migrant is almost by definition unavailable for response, making it necessary to ask migration questions through a proxy respondent.<sup>15</sup>

There are two notable exceptions to this rule. First, since return migrants should be available for an interview, they can and should be asked directly about their migration experience. However, as return migration is highly selective, one should not consider the experiences of return migrants as representative of migrants in general. Second, if seasonal migration is

---

<sup>15</sup> One exception might be when or if migrants are home for holidays; another possibility is when the focus of the survey is to study seasonal migration (e.g. Macours and Vakis, 2010). However, scheduling LSMS visits for holidays to capture migration is not advised, as consumption usually increases during holidays, which would invalidate the consumption measure or make its computation more complicated, as occurred in Vietnam (World Bank, 2000). Furthermore, refusal rates can increase during holiday periods.

prevalent, it is worthwhile structuring the fieldwork to occur during the time of the year when seasonal migrants would be at home. For example, seasonal migrants who work in tourism would likely be home during the off-season, while seasonal agricultural workers are likely to be home during the fallow season.

Proxy respondents have the advantage that they allow the enumerator to acquire information that would not be otherwise. However, there are also disadvantages. The main disadvantage is that the proxy is not the migrant, so they may not know some of the information that the analyst would like to collect. Therefore, only a limited set of questions should be asked. Another drawback to using proxy respondents is that information about the migrant's motivations and experience will both reflect the actual experience of the migrant as well as the perceptions of the migrant's experience held by the proxy respondent. Any bias incorporated into the answer of a proxy respondent is very difficult to sign. To attempt to limit the bias in answers by proxy respondents, one should:

- Ensure that questions about migration are directed to the member of the household most knowledgeable about migrants. This individual could be the household head, the mother of the migrant(s), or perhaps an elder (not the head) living in the household who is familiar with the whereabouts of various family members.
- Realize that answers to any questions asked about the migrant come from the perspective of the proxy respondent rather than from the migrant. For example, a proxy respondent might not know the extent to which the migrant used migrant networks to obtain employment, or even the wages the migrant earns. Properly phrased questions can help guarantee that the response reflects the migrant's perspective rather than that of the proxy.
- Along the same lines, limit questions that require subjective answers or recognize that using a proxy respondent changes the meaning of answers. For example, if one asked a proxy respondent how easy the journey to the destination was or whether or not the migrant would consider moving back home, the answer would be meaningless.

### *3.2.2. Issues related to Extended Household Membership*

As discussed in subsection 2.1.1, the standard definition of household members that is used for other purposes in LSMS surveys—for example, to define the household for the purpose of computing per capita consumption—is too narrow a definition to learn about the exposure of the household to migration, as some migrants linked to the household will not be considered

members according to the residency rule. Therefore, information about individuals not passing the residency rule must be enumerated. In planning to collect this information, one must make decisions regarding the individuals about whom additional information should be collected, and the extent of that information.

There are several ways to potentially identify migrants beyond individuals who are already considered household members. The most expansive definition that one could use would be to initially ask about everyone who had ever lived in the household since its formation. This method would lead to the largest number of households being classified as migrant households – however, reports about individuals who left the household 20 or 30 years ago without returning are likely to be inaccurate. An opposite extreme is to only ask about individuals who did not pass the residency rule, but did live in the household for a portion of the previous 12 months. This definition would potentially exclude individuals who contribute financially or otherwise to the household. Some potential compromises between the two include:

- Fully enumerating basic information and whereabouts of the children of the household head, and potentially the spouse of the household head, even if they do not pass the strict residency rule;
- Collecting basic information on any individuals related to the household head who lived in the household for at least 3-6 months and have left in the previous five years;
- Some combination of the two; for example, one might only enumerate information concerning children of the household head who left in the previous five years.

To make a decision about which extended residency rule to use, one should consider local conditions in the country of study and decide the most appropriate rule for the specific circumstances, potentially turning to local institutions for guidance.

### *3.2.3. Key Points*

Because migrants are not likely to be present for the interview, information about migrants must be gathered through proxy respondents. In deciding how to ensure that accurate information is collected, we recommend that one should:

- Ensure that if the migrant or returned migrant is present in the household, that person will be asked about their migration experience directly.
- Expand the definition of household membership for the purposes of learning about migration.
- Ask for the basic demographic information of any individuals who have lived in the household over the past twelve months.

### 3.3 MIGRATION IS NOT A RANDOM PROCESS

The decision to migrate is not made separately from other decisions within households and may be affected by either observable or unobservable factors that explain household expenditures or other variables of interest. Since the same unobservable characteristics that affect outcome variables of interest likely affect migration, such as ability or risk aversion, measures of migration should almost always be considered endogenous to any outcomes of interest. If one uses migration as an explanatory variable in a regression, the coefficient estimate on migration can be biased via three different paths:

- *Reverse causality.* The outcome variable of interest may affect migration, just as migration may affect the outcome of interest. For example, per capita consumption in the household might lead the household to send out a migrant, or household per-capita consumption might be affected by having an out-migrant. If households tend to send out migrants temporarily in response to consumption shocks, one might find a negative relationship between migration and consumption, even if the opposite were more generally true.
- *Selection bias.* Individuals and households migrate for specific reasons, and because they are willing and/or eager to migrate, individuals who migrate are inherently different than individuals who do not. As a result, migrants are not directly comparable to the entire population of non-migrants; they are selected into migration.<sup>16</sup> And although some indicators predicting selection into migration are observable (age, education level, labor market experience), others generally are not (ambition, risk preferences). Without controlling for selection, migrant households are not directly comparable with non-migrant households. Selection is also an issue when comparing return migrant households with other households, as returning individuals did not randomly choose to return, making it impossible to directly compare them with either non-migrants or migrants who have not returned.

One method that can help limit this form of selection bias is to collect information about what the migrant was doing immediately prior to leaving. The effect of collecting such information likely depends a great deal on the context, but one could imagine that in certain situations a migrant may have lacked employment (relative to others with the same human capital characteristics), which could have served as a catalyst for migration.

---

<sup>16</sup> A literature exists on selection into migration itself, and the determinants of selection. For some examples in the case of Mexican migration, see Chiquiar and Hanson (2005), Ibarrran and Lubotsky (2007), and McKenzie and Rapoport (2007). For the perspective of immigration, see Borjas (1987).

Alternatively, it could be that those who were previously able to find wage employment in the source community saved their earnings to finance later migration.

A second form of selection bias occurs when whole households move, and are therefore missed entirely by the sampling strategy.<sup>17</sup> If these households are not tracked, then any estimation results on the impacts of migration are subject to this selection bias. With a nationally representative survey, however, one could hypothetically ask households that have recently internally migrated about their source community, to understand the characteristics of entire households that move. This selection bias is particularly problematic when using panel data, as the households that are not found in the second and subsequent rounds of the survey may have migrated. Panels, however, do allow the observer to understand which whole households migrate and which do not.

- *Omitted variable bias.* As alluded to above, migrants have characteristics that are unobservable to a researcher, and the decision to migrate depends upon factors that are similarly unobservable. If these unobservables, or observable variables that are omitted from the model, differ systematically for migrant households and non-migrant households, ordinary least squares estimates of the effects of migration will be biased. If panel data is available, one can use household fixed effects to control for time-invariant unobservables, but unobservables may also vary over time, thereby still rendering estimates based on fixed effects regressions potentially biased.

There are several ways that surveys can be designed to attempt to limit these endogeneity problems. The best solution for dealing with endogeneity problems is to use an experimental design, in which the treatment is randomly assigned and information is collected on outcomes before and after the random assignment occurs. However, given the nature of migration, the “treatment” is hardly ever randomly assigned.<sup>18</sup> Given that random assignment is not possible, one can use several approaches to deal with the endogeneity of migration. We describe these methods, their advantages and disadvantages, and their resulting implications for questionnaire and study design below.

### 3.3.1. Strategy 1: Attempt to Ignore the Bias

---

<sup>17</sup> This problem is more acute if panel data is being collected or if internal migration is being studied; because the sampling frame is nationally representative in a typical cross-sectional LSMS, when studying international migration, this statement only applies to whole households that migrate internationally. Generally this is an uncommon phenomenon, and it would be either impossible or highly impractical to try to find such households.

<sup>18</sup> An exception is work by McKenzie, Gibson and Stillman (2006), who take advantage of the random allocation of New Zealand visas to Tongan residents who enter a lottery.

One can, of course, simply ignore the bias of any effect of migration on outcomes, at least in estimation. One might choose to ignore the bias or argue that it is of minimal importance if standard ordinary least squares (OLS) estimates are largely consistent with estimates using other estimation methods. Alternatively, one might choose to simply use the OLS coefficients if the analyst is only interested in:

- Associations between migration and the outcome variable, rather than causal estimates;
- The sign of the coefficient rather than the magnitude of the coefficient, but only if the bias can be signed and will not change the sign of the coefficient estimate.

If both points above are not met, it is not advisable to use OLS to analyze migration, because the bias could change the sign of the estimated coefficient of interest and therefore any inference based on estimates would be invalid. In that case, it is preferable to use one or more of the following three methods in estimating the impacts of migration.

### *3.3.2. Strategy 2: Panel Data*

There are several reasons that panel data is potentially useful for studying migration. First, with more than one observation on each household, one does not have to rely entirely on recall data to study migration, as one can observe changes in the composition of household members who live there over time. By asking for reasons that changes occurred in subsequent panel rounds beyond the first, one can differentiate absence due to death or marriage from absence due to migration. Panel data can also be used to confirm questions related to recall asked in the second and subsequent rounds of data collection.

From the perspective of identification, panel data is also useful for limiting bias. When using panel data to analyze migration, one can include household or individual level fixed effects in regressions, depending upon the level of analysis. Assuming that analysis is being performed at the household level, a household fixed effect will capture any observables or unobservables about the household that do not change over time. The remaining problem is that residual correlation between changes in migration and changes in unobservables that occur over time will potentially still bias estimates of the impacts of migration. Using panel data can potentially limit the bias in estimates of the impacts of migration, but it cannot eliminate the bias by itself. However, panel data can be particularly effective in controlling for endogeneity bias when combined with either matching methods or instrumental variables, which are described below.

Panel data also has its drawbacks. By definition, one needs to survey households twice before one has panel data. Therefore, using panel data is only feasible when results of the survey

related to migration are not required for at least two or three years after the first survey is completed, or where a subsequent round of an existing survey is taking place. If refusal rates are high on repeat visits or if households have moved from the original sample areas, attrition bias may become a further concern (e.g. Wooldridge, 2002).<sup>19</sup> Attrition may also be related to the migration of households out of sample areas, which would lead to a selection bias in estimates that is difficult to overcome without tracking migrants or migrant households (Foster and Rosenzweig, 2008). The households in panels also age, making them less representative of the demographic composition of the population over time. Finally, panel data is expensive to collect, particularly if migrant tracking is attempted.

### *3.3.3. Strategy 3: Use Matching Methods*

To potentially deal with the endogeneity of migration, one could use nearest neighbor or propensity score matching techniques to estimate the effect of migration on outcomes. Matching estimators essentially match “treatment” observations (migrant households) with “control” observations (non-migrant households) that have similar values of explanatory variables, and then assume that unobservables are equal in expectation for the treatment and control observations. The primary concern regarding matching methods is that the lack of bias of coefficient estimates hinges on the assumption of equivalence in unobservables. To ensure that one could not have controlled for these variables when using matching methods, it is important to include all variables that might affect both the propensity to migrate and the outcome variable in the matching exercise. Although somewhat counterintuitive, it is actually important not to include variables that might be considered good instruments for migration in matching exercises, because instruments can decrease the quality of the match (Bhattacharya and Vogt 2007).

If one omits variables that affect both migration and the outcome of interest, the estimator will certainly be biased. As the lack of bias in estimates using matching methods hinges on the assumption that conditional on observables, the expected value of any unobservables are equivalent for both the treatment (migrant) and control (non-migrant) groups, it is therefore particularly important to make sure that specific observables that will affect migration are included in the survey, such as pre-migration income or wealth. McKenzie, Gibson and Stillman (2006) show that when considering the impact of migration on wages among Tongan migrants in New Zealand, a matching estimate is modestly biased relative to an experimental estimate. Only a few authors have attempted to apply matching methods to estimate the impacts of migration;

---

<sup>19</sup> From the perspective of migration, there is a silver lining in attrition if one can learn about whether households internally or internationally migrated between survey rounds.

Ham, Li, and Reagan (2006) apply propensity score matching methods to estimate wage growth among migrants within the United States, and Acosta (2006) uses propensity score matching to estimate the impacts of migration on school enrollment in El Salvador.

To summarize, if one plans to use matching methods to learn about the effects of migration, it is important to collect comprehensive data on variables that will affect both migration and outcomes of interest. If important variables are left out of the survey, then matching estimates may suffer from significant bias. Alternately, the bias in a matching estimator is likely to be modest if all important explanatory variables are included in analysis.

#### 3.3.4. Strategy 4: Instrumental Variables

As migration is an endogenous variable, the most common way to estimate the impacts of migration is to use instrumental variables techniques. For instrumental variables estimates to be theoretically unbiased or minimally biased, instruments used to study migration must be strongly correlated with migration but uncorrelated with outcomes of interest. If instruments are not strongly correlated with migration, then weak instrument bias can occur (e.g. Staiger and Stock, 1997; Stock and Yogo, 2004). Weak instruments bias coefficient estimates towards the OLS estimate, so it is preferable to avoid particularly weak instruments. A weak instrument problem can be created by the use of a large set of instruments (e.g. Angrist and Krueger, 1991; Bound, Jaeger, and Baker, 1995; Staiger and Stock, 1997), so small sets of instruments are preferable to large ones. Furthermore, at least one instrument is necessary for identification.

Planning to use instrumental variables has important implications for data collection. In an overview of the use of instrumental variables for analysis, Murray (2005) discusses the use of statistical arguments to ensure that the instruments being used are plausibly exogenous. From the perspective of analyzing migration, the instruments should only affect the outcome of interest through their effect on migration. One should be able to rule out any other channels through which the proposed instruments might affect the outcome, and doing so empirically is much more convincing than an *ad hominem* argument. Some channels can be analyzed from within the data set, whereas others require external sources of data. For example, de Brauw and Giles (2006) explore whether their instruments for migration, a quartic function of the years since ID cards were issued in China, are correlated with variables that measure other policy changes that took place around the same time as ID issuance, all of which are constructed through the data set they use. McKenzie and Rapoport (2007) examine correlations between their instruments, which are based on historical networks, and state level variables that are similar to their outcomes of interest. Their instruments and their exogeneity defense are both largely derived from external

data sources. Although it is nearly impossible to rule out all potential channels through which such effects work, ruling out such channels adds credence to the argument that the instruments are good ones.

To ensure that potentially good instruments appear in the data set, it is worthwhile to consider what instruments have been used to identify migration in the past. The ideal instruments are either natural experiments or almost natural experiments (Rosenzweig and Wolpin, 2000). However, if one finds a natural experiment that affects migration, one must be careful to ensure that the natural experiment only affects migration and not other variables, including the outcome of interest. As argued by Sasin and McKenzie (2007), natural disasters are not likely to be good instruments for migration. Although earthquakes, hurricanes, mudslides, or severe droughts can lead to more migration, they are also likely affect any outcomes of interest. An example of the type of natural experiment that might affect migration but not other outcomes could be unanticipated government interventions that affect migration but not other components of the labor market. While different instruments for migration have been used in the past, here we highlight some of the better ones:

- *Historical* (e.g. Woodruff and Zenteno, 2007; McKenzie and Rapoport, 2006) or *nascent networks* (de Brauw and Harigaya, 2007; de Brauw, 2009; McCarthy et al. 2006; Miluka et al., 2010). Although the strength of mature networks is endogenously determined (e.g. Munshi, 2003), depending upon the outcome being studied, historical networks may affect the propensity to migrate without affecting contemporaneous outcomes. Either historical or nascent network variables are plausibly exogenous, but it is likely that whatever factors lead to historical or nascent networks also affect other households in the community.
- *The distance of the household to some fixed point*. McKenzie, Gibson and Stillman (2010) show that distance to the NZ embassy is a good instrument for migration from Tonga to New Zealand relative to the experimental estimator they have available. Miluka et al. (2007) use the distance from households in Albania to one of two border crossings with Greece as one of their instruments identify emigration, and several papers use the distance to North-South railways built around the beginning of the 20<sup>th</sup> century to aid identification of migration or migrant networks from Mexico (e.g. McKenzie and Rapoport, 2007; Hildebrandt and McKenzie, 2005; Woodruff and Zenteno, 2007). Variables related to distance may generally affect socioeconomic status through other channels as well; for example, railways are often built where it is easiest to build railways, and hence economic activity might be greater in those places than elsewhere.

It is therefore important to test whether these variables are correlated with other variables that might affect the outcomes of interest.

- *Wage rates in migrant destinations.* Variables could plausibly be constructed at the household or community level that measure shocks or deviations from trends in wage rates at the destination. Such variables should identify labor demand, theoretically allowing one to identify labor supply. One requirement for such variables to be good instruments is differences in migrant destinations across space. Additionally, such a strategy might have quite stringent external data requirements, especially in the case of international migration. Although the exact strategy as outlined here has not been used in the past, Yang and Martinez (2005) and Yang (2008) use differences in changes in exchange rates to identify remittances in the Philippines after the Asian financial crisis, as emigrants in different destination countries suddenly held potential remittances with different values in the Philippines. Antman (2007) applies a similar strategy, using unemployment rates in the US state to which Mexicans migrated to identify the effect of migration on educational outcomes among migrant children. For this strategy to work, it is important to collect the actual city, state, or locality of the destination, rather than just the country or region, to ensure that the wage shocks faced by the migrant can be estimated as accurately as possible.
- *Lagged weather shocks.* Munshi (2003) and Giles and Yoo (2006) have both used distant lagged weather shocks that affect the quality of the migration network. One particular problem with these variables is that they can affect outcomes through persistence. For example, negative agricultural production shocks represented by rainfall shocks could lead to decreased consumption over a long period.
- *Policy variables.* As noted earlier, government policies that affect regions differentially can make good instruments for migration. However, finding such instruments can be quite difficult in a cross-sectional setting, as the policy must have spatially differential effects. Another problem with policy variables is that policies are often endogenously placed in specific regions as pilots, to ensure that they work well. In part because they use panel data, de Brauw and Giles (2006; 2008) are able to use differential timing in the issuance of ID cards that allow legal temporary resident in China's cities to identify migrant networks. As with the distance to fixed points, one must convincingly argue that policy variables are not proxying for other variables that might affect outcomes of interest.

### 3.3.5. How Does the Endogeneity of Migration Affect Study or Questionnaire Design?

If one chooses to analyze migration using either matching methods or instrumental variables, there are direct consequences for both the study and the questionnaire design, specific to the strategy to be used for arguing identification:

- *Matching.* If one is planning to use matching methods, a large set of covariates is necessary. In fact, the larger the set of covariates, the better, as additional covariates help build the case that unobservables are equal in expectation in both the treatment and control groups. Variables included in the questionnaire for matching purposes should affect both migration and outcome of interest; there is no need to attempt to include instruments for migration if one plans to solely use matching methods. In general, if an LSMS survey includes enough observations on migration, it should be well-suited for analyzing migration through matching, as the data set typically includes a large set of covariates.
- *Instrumental Variables.* If one is planning to use instrumental variables techniques and any of the instruments are intended to come from the survey itself, questions that will help build those variables must obviously be included. However, the discussion in the previous subsection also implies that if one has a good candidate instrument or set of candidate instruments, one should consider the defense of those instruments. To do so, one should think through the types of variables that might best represent unobservables that could be correlated with both migration and the outcome or set of outcomes of interest.
  - Instruments are sometimes generated from companion community or PSU level data, so it is useful to include a community-level questionnaire in the study design. Of the potential instruments described above, the two that can be best studied using a community-level questionnaire are migration networks and policy variables. It is therefore worth designing the community-level questionnaires to ask about any community-wide migration networks and any policy variables that might affect local migration, although one should always be wary that potential policy variables could be correlated with any outcomes of interest.
  - Instruments also sometimes come from external data sources; rainfall shocks, for example, must be computed from administrative data. These sources should be carefully documented.

- If one is planning to use wage shocks or wages at the destination as an instrument, one potentially needs external data sources about wages at those destinations, and must include in the survey questions about the city or location of the migrant at the destination. In the case of international migration, this information should include whether the destination is urban or rural, and preferably the name of the province, region, state, or city, so that wage rates or shocks can be estimated more precisely.<sup>20</sup> In fact, asking about the actual destination can also help analysts track migrant networks more precisely, so it is advisable to seek increased precision in the coding of the location of migrants away from the household. If one plans to use distances to a border, railroad, or other geographic features, the location of the household should be determined using a GPS, and one should obtain GIS data on the geographic features in order to calculate distances.<sup>21</sup>

---

<sup>20</sup> It is worth noting that weights that come solely from a household survey could be considered endogenous as endogenous processes would have led to the proportion of migrants from each place in specific locations.

<sup>21</sup> There are ethical issues that one should consider when using GPS data prior to making the data public (Gibson and McKenzie, 2007a).

### 3.3.6. Key Points

The study design, and therefore the data collection strategy, should consider that migration is not a random process, implying that migration must be treated as endogenous. Endogeneity may be created by reverse causality, self-selection, or omitted variables. To attempt to address endogeneity, we recommend instrumental variables despite their potential drawbacks, since the bias in point estimates is likely to be lowest with good instruments. However, matching is a plausible alternative if good instruments are not available or not likely to be found. If using instruments, one should:

- Confirm that, given the context, an instrumental variables strategy is the best way to deal with the endogeneity of migration;
- Carefully consider potential instruments for migration, starting with the list above, but also thinking creatively about migration in the country being studied; and
- Design the survey to collect information about possible instruments, and/or research the types of external data sources that might also give community-level information that would be useful in constructing instruments.

#### **4. Questionnaire Design**

A major advantage of collecting migration information in the context of an LSMS is that it can be analyzed in combination with data from other parts of the LSMS questionnaire. Even if analyzing the impact of migration on specific outcomes of interest is the primary goal of the survey, it is important to first ensure that the survey can identify all of the potential types of migration taking place either within or from the country or context being studied, as well as the determinants of migration. The analyst should also attempt to include information on as many potential determinants of migration as possible, as an omitted determinant of migration might bias the analyses of outcomes of interest.

The focus of this last section is to propose some basic principles in incorporating migration information into an LSMS-type survey to enable researchers to carry out this type of analysis and policy-makers to benefit from it. As has been previously established, one key consideration is the trade-offs that must be made in order to keep the length of the questionnaires manageable. Another important point is that migrants are a more mobile segment of the population and are therefore less likely to be found at the time of interview. As a result, one must assess the extent to which certain questions can be administered through proxy respondents while still collecting reliable information.

We begin the section with a discussion of how to collect information about the migration experience. We recommend using a migration history, and we detail different methods for collecting migration histories, while focusing our discussion on emigration rather than immigration. The second subsection describes the determinants of migration, discussing the specific sections of the questionnaire that would be particularly useful towards that end. Third, we present an examination of the policy questions that can adequately be addressed with an LSMS survey, and we provide an overview of additional modules that are likely to prove useful when analyzed in conjunction with a migration history.

##### **4.1. COLLECTING INFORMATION ABOUT MIGRATION EXPERIENCE**

Broadly speaking, information can be collected concerning the previous migration experiences of two sets of individuals: 1) current household members, or any household members who have been present at some point during the past twelve months and 2) individuals who are still associated with the household but may not have lived in the household over the past twelve months.

By definition, individuals in the latter group are absent and thus a proxy respondent must always be used. Furthermore, some of these individuals may now only have remote links to the households, as they may have left long ago. If studying these individuals is the primary objective of the study, then tracking and interviewing them at the new location would be the only other option. Short of that, asking a minimal set of questions as part of the survey through a proxy respondent would be feasible. These questions should be included in a separate roster/module and must include the age, gender and civil status of the migrant, as well as the year s/he left the household and some basic information about the new location. In addition, questions on the migrant's occupation in the month prior to migrating and the current occupation should also be asked, together with questions concerning the migrant's remittance behavior. As already mentioned, this group of individuals could include all former household members, but it is preferable to limit the list to the household head, his/her spouse, and their children.

With respect to the first group of individuals – i.e. the current household members or people who lived in the household at any point during the previous 12 months – we recommend beginning by asking a short battery of questions as part of the household roster. These describe the household members who were only present for part of the year as well as the household members who are present at the time of the interview(s). If a given individual is present in the household all year, full information should be collected as regards that individual. Full information should also be collected if individuals are present for part of the year and are in the household at the time of the interview, even if they fail the residency rule. However, a standard rule must be established regarding the proxy evidence that should be collected when the individual is not available. The basic household roster module asked at the beginning of the survey can be used to help determine the information that should be collected by proxy. We recommend initially asking about anyone who is considered a household member or has lived in the household over the past 12 months. In this way, limited demographic information is immediately collected about all individuals who have spent some time in the household, including any migrants who are away or who may have left during the course of the year. Using this method, the measure of household size relevant for measuring per capita consumption is clearly available, as information is captured that allows specific individuals to be excluded from the household definition for consumption purposes.

#### *4.1.1. Identifying Migrants Through Screening Questions*

After deciding upon an expanded definition of household membership, one must decide what information to collect by proxy among individuals who are not present. We illustrate a decision

tree for data collection in Figure 1. If individuals are present in the household all year, one should collect full information about that individual. If individuals are present for part of the year and are in the household at the time of the interview, even if they fail the standard residency rule, one should collect full information as well.<sup>22</sup> The real decision the analyst faces concerns the rule to follow regarding the proxy evidence to be collected when the individual is not available.

Some individuals who no longer live in the household may not be migrants. Specifically, some may have left to set up a household elsewhere in the village or community, while others may have left to set up a household in a nearby community – neither of these groups should be classified as migrant individuals. In order to restrict the collection of detailed information to individuals who should be considered migrants, a set of screening questions should be asked to help correctly and consistently distinguish migrants from non-migrants. Some examples of key screening questions related to migration include the number of months each individual was not present, whether or not they are present currently, the individual’s location if not, a reason for absence code, and a question probing whether or not the person intends to return. The location and reason codes should be adapted to be relevant to the specific country context of study.

Of these, the question of a given individual’s status (potential codes include “present,” “absent but returning,” and “absent but not returning”) as well as the number of months the individual was present in the household can be particularly useful in serving to restrict the amount of information collected about specific classes of individuals. For example, one could hypothetically choose to restrict the collection of detailed information to individuals who were either “present” or “absent but returning,” since the household likely has more information about those individuals than those who are absent and will not return. In general, the objective is to create a rule that will ensure that the proxy respondent is sufficiently familiar with the absent individual to answer questions about him or her accurately. In the above scenario, the proxy respondent would not be requested to provide detailed information for any individual who is absent and not returning, assuming a lack of familiarity with the absent individual. As the key issue in this case is familiarity, any number of other screening questions in this vein – for example, the number of months each individual was present in the previous year – could suffice equally well.

Other issues to consider when designing screening questions may include the geographic distance (the distance that an individual must travel) and the length of absence (the length of time

---

<sup>22</sup> If many individuals fall into this category, this rule could significantly extend the length of the interview. If this is the case, one could ask the household head or primary source of the interview to list all such people, and the enumerator could ask about a randomly chosen subset of those listed.

that an individual must be away from the household) necessary to be considered a migrant. In regards to geographic distance, the tradeoff is illustrated as follows. If in the screening question, one asks whether the individual lives outside the province, then any intra-provincial migration will be missed. Meanwhile, if one uses a smaller geographic unit, depending upon the context one might risk classifying some individuals who are seen by the source household on a daily or weekly basis as migrants.<sup>23</sup> Length of absence is also important to consider, as it can help to discern prior migration from extended trips to visit family or potentially even from tourism. These alternative possibilities can be explicitly ruled out within the line of questioning, or a time rule can be defined – for example, the United Nations requires that an individual be absent for 3 months before being considered a migrant.

#### *4.1.2. Migration History Module: Tradeoffs*

Once the group of migrants has been identified, we suggest administering a full migration history module, documenting all moves over a specified reference period. The history should be broad enough to capture every possible type of migration that could be occurring in the country context. To ensure that the migration history captures high quality information, one must make several important decisions about migration history content. With proper training and by anchoring migration episodes to explicit time benchmarks, like major events such as a political election or a natural disaster, reconstructing full migration histories may be feasible and highly productive.

Although one would ideally want to record all household migration episodes in adulthood, it may be advisable to limit the analysis to people who have left the household within a specific period of time. The choice of recall period length is important both in determining the frequency of migration found in the data set, as well as the quality of information available for analysis. While rendering the most information, querying each household member on every migration experience across their lifetime has three clear drawbacks. First, errors in recall will increase as the recall period increases in length. Second, the marginal value of the information added about the household decreases as the recall period increases in length. Third, the portion of the survey dedicated to migration would require an excessive amount of time if it attempted to enumerate every migration event for a large household. On the other hand, it is important to avoid overly minimizing the detailed recall period, as very few people may fit the definition of

---

<sup>23</sup> This rule is meant to rule out two types of individuals. First, one want to rule out commuters, who are people who commute to a job and might stay near the job site during the week, but frequently return home on the weekends. Second, people may have left the household for non-work related reasons (e.g. marriage) and moved nearby, and hence should not be considered migrants.

“migrant” in a short recall period, potentially leading the analysis to miss the effects of migration on the household.

A five-year reference period can be considered as the minimum length between including marginally relevant information and missing information on migration that might have affected the household. In some cases, depending on the incidence and characteristics of migration in a specific country, a longer period may be preferred. Using a longer reference period would also ameliorate the rare event problem, as more individuals would then be classified as migrants. For example, in transition countries in Eastern and Central Europe, at least for the time being, extending the recall period to the early 1990s, the period in which most of these countries relaxed migration laws, is recommended and still feasible. A relevant example is the 2005 Albania Living Standards Measurement Survey, which extended the recall period to 1990 (INSTAT, 2005). The 2005 Albania LSMS provides an example of migration histories (visit [www.worldbank.org/lsm](http://www.worldbank.org/lsm) for more information). Similarly, in other circumstances, starting the recall period in a year clearly marked by a major event or a policy change may be suitable. For example, a survey completed by the Center for Chinese Agricultural Policy (CCAP) measured internal migration and off-farm labor in China with a module commencing the recall period in 1980 (de Brauw et al., 2002), around the time that the Household Responsibility System reform took place. The long recall period was judged to be sound in pre-testing because off-farm employment was rare in 1980, and the first off-farm job for a household member was a major event in their lives that would be recalled accurately.<sup>24</sup> However, the method used in the CCAP survey might miss off-farm work done by individuals who lived in the household but passed away prior to the five years before the survey (all births and deaths in the household in the preceding five years were enumerated). Furthermore, all households that had passed away in entirety in the preceding five years were also omitted from the sample frame, so migration rates estimated from this sample would be biased slightly upward. This critique would also hold for migration histories; however, in both cases the bias is likely limited as older people are less likely to work off-farm or migrate.

Regardless of the length of recall period ultimately chosen, it is worthwhile to collect the year of first migration for everyone in the sample with any migration experience irrespective of the fixed reference period. If one strictly uses the five-year recall period, one might miss information about either longer migration spells that began prior to the recall period, or important information suggesting access to migrant networks. Furthermore, since the composition of migration changes over time, it is useful to know whether households have earlier experience with migration.

---

<sup>24</sup> The EHRUM, collected in Mexico in 2002 to be nationally representative of rural Mexico, also used 1980 as the beginning of the recall period (see Richter and Taylor, 2005, for more details).

A minimum criterion should also be defined for the duration of the migration event – for example, that a migrant must live abroad for three to six months before being considered an international migrant. Bilsborrow et al. (1997) suggest using six months, but this period may be too long as it would miss seasonal migration that might take place for a three to six month season. Varying the criteria to define migration can increase the number of migrants enumerated, but it also may have costs. Shortening the length of migration duration will increase the number of potential migration events, but shorter events that took place in the reasonably distant past will be more difficult for the respondent to recall (e.g. Smith and Thomas, 2003). Even with six-month long stays abroad or at internal destinations, it is not guaranteed that individuals will recall events that occurred in the reasonably distant past, especially if the survey attempts to enumerate migration episodes over a long period of time or if individuals migrate for short periods of time frequently. Finally, if the minimum length of migration duration is shortened to stays of one month or even less, one is more likely to create ambiguities with other types of movements such as visits to family or vacations. Since exceptionally short lengths of stay are both likely to create such ambiguities and are less likely to be recalled, using a length of migration duration of at least three months is advisable. Anchoring the timeline to specific time marks, such as major economy-wide or local events, as well as idiosyncratic events specific to the household, such as births, deaths, or marriages, can help reduce recall errors. Properly training the enumerators to use such cues is essential.

Finally, it is important to know something about the conditions in the household pre-migration, so it is also useful to collect other, limited information using the chosen recall period as a reference. For example, one could collect information on occupations or assets in the household prior to, during, and after the migration spell of any household member who had migration experience. Other information, often subjective, might also be worth collecting for individuals who are potentially return migrants, such as the reasons for return and whether or not they plan to leave the household again. Some potential reasons for return, such as health of household members or job status, can be corroborated in other parts of the survey.

#### *4.1.3. Designing the Migration History*

In this subsection, we detail two methods of collecting a migration history of individuals for whom full information is to be collected. The first version is a longer, more detailed method, while the second version is shorter and can be used when questionnaire space is limited. As we describe each method, we will highlight its advantages and drawbacks.

#### *4.1.4.1. Method One: Long Version*

We present the long form of the sample migration history in Figure 3. After asking a screening question for whether or not household members have ever migrated, we first ask for the year that the individual left for the first time.<sup>25</sup> The third and fourth questions frame the individual's migration experience, asking whether this was the only migration experience, and then about the time of the most recent migration. The third question also asks about the number of migrations between the initial and most recent migration in order to determine the respondent's frequency of migration. Next, the module enumerates details about the most recent migration episode, including the destination country and city, the length of the stay, the occupation, and the reason for return. The third block of questions in the module asks about the first migration episode. The questions in this block include all of the questions asked about the most recent episode, adding a question to learn about the migrant's occupation prior to any migration (question 11), a question about information on how to migrate (question 13), and how the initial migration was financed (question 14). The last three questions ask about the migrant's family remaining in the destination (if any), the migrant's future plans for departure (if any), and the intended financing for the trip. The two questions about plans for the future can imperfectly help the analyst categorize individuals as repeat or return migrants.

The questions proposed do not detail each migration episode, and therefore one can and should expand this list if space is available in the questionnaire. For example, one might add further questions on whether the migrant entered the destination legally (in the case of international migration), how the migrant might have used migrant networks, how the migration episode was financed, and whether the migrant's spouse, partner, or family accompanied them.<sup>26</sup> Lastly, if seasonal migration is prevalent in the context being studied, it is worth including some additional detail about the length of migration spells between the first and last one, in order to capture information on the frequency and length of migration spells. As proposed, for a migrant

---

<sup>25</sup> We focus in the sample module on international migration, but it could easily be modified to also consider internal migration. We would suggest first asking a screening question about international or internal migration, and then essentially repeating the same module for internal migration, so that one receives full information about both internal and international migration.

<sup>26</sup> From the policy-making perspective, it might be particularly important to ask about whether individuals are migrating legally or not. However, households might find this line of questioning sensitive, as the household would not want to reveal illegal migration if they thought the migrant (or their livelihood) could be endangered in any way. Furthermore, any Internal or Institutional Review Board that would review the questionnaire might object to such a line of questioning. One might also be concerned about the veracity of information in any context where illegal migration might give the respondent a reason to hide. Prior to asking such questions, one should ensure that household would not find such questions sensitive, and if included one should have procedures in place to ensure that answering the questions truthfully would not endanger the migrant, legally or otherwise.

who worked for a piece rate in agriculture in a destination country, the module would only capture information about a short period for the first and last migration spells, and the total number of migration spells. If such migration is common, it is worth adding detail about further migration spells. In this case, if one omitted the question on additional migration spells, the analyst would necessarily not realize that the individual was a seasonal migrant. An alternative would be to use an annual labor history, as discussed in subsection 4.1.4.3.

#### *4.1.4.2. Method 2: Short Version*

The shortened version of the migration history in Figure 4 would be appropriate in contexts where little migration occurred, or where the analyst had limited space with which to ask about the migration experience of current household members. The history should be administered to each individual in the household, including those for which proxy information is being obtained. We propose omitting information about the first migration episode, instead asking primarily for details about the most recent migration in the past five years. We suggest retaining, at the minimum, a question about how the migrant obtained information about jobs in the destination country and whether or not the migrant plans to migrate in the next twelve months, to be able to crudely differentiate repeat migration from return migration. We add questions about the occupation prior to the first migration, as well as the way each individual learned about and financed the first migration. The module also includes questions about family in primary destinations.

#### *4.1.4.3. Annual Migration History*

If additional space is available to enumerate migration within the questionnaire, one can add questions about migration spells between the first and most recent spells within the recall period.<sup>27</sup> As discussed above, it is worthwhile also asking about the first migration, even if it is

---

<sup>27</sup> A final method of asking a migration history is to ask an employment history of all household members. This method asks about employment instead of migration and has been used in the 2000 CCAP survey of rural China and in the EHRUM in 2002. The module proposed is similar to the grid method described in the previous subsection, but without requiring the enumerator to record answers to specific questions about separate migration spells. In principle, the grid can be set up in a number of different ways, but the three questions that are of primary importance to learn about in an employment history if migration is an issue are: the main type of employment, the residence while working, and the location if away from home. The enumerator can be instructed to use a similar method to the detailed migration history above - for example, asking the first time that each individual in the household worked off the farm, and then asking about subsequent years. Although this method can provide a complete picture of off-farm employment and

not included in the recall period, and one should make sure that the whole length of the first migration spell occurring during the recall period is understood, even if it began before the recall period. In attempting to enumerate a migration history over an extended period of time, the main difficulty one faces is the potential for recall errors, particularly for short migration episodes (e.g. Smith and Thomas, 2003). To try to eliminate these errors, one suggestion is to follow the example of the enumeration of international migration in the 2005 Albania LSMS. Individuals were first asked about their most recent migration episode and their first migration episode since the migration recall period began. The enumerator was instructed to fill in both the first and the last migration spell on a grid with each year of the migration recall period listed. The grid folded out from the body of the questionnaire, so that the enumerator could simultaneously see the grid and the questionnaire. After asking about every migration spell, enumerators were instructed to record on the grid the pre-coded migration destination and the number of months the individual was away during that year. After asking about the first migration spell, the grid becomes a reference for the enumerator. If, for example, the respondent recalls their first migration in 2000 and a second migration that occurred in 2003, the enumerator observes on the grid that 2001 and 2002 had no migration, and can ask to confirm. The grid can thereby help minimize recall error by providing a double bound frame of reference for the respondent.

The detailed method of asking a migration history has the advantage of being particularly complete—if asked correctly, one can learn about several different types of migration, although the analyst must still make decisions about how to define different types of migration. However, because of its detail, the migration history can take a considerable amount of time to ask, particularly where households might have several members with migration experience. In a multi-purpose survey like an LSMS, spending so much time enumerating migration might come at the cost of other informative modules, or limited information about other variables that might be of interest. One might also continue to be concerned about potential recall error.

#### *4.1.5. Collecting Information about Current Migrants*

Before discussing the questions that one might want to ask of current migrants, one must consider who should be included as current migrants from the household. One ideally wants to capture information about household members and former household members, as well as anyone who might be linked to the households and is currently overseas. As no method is perfect, we discuss several possible alternatives.

---

migration, it is also time consuming, and as a result both surveys that have used it only asked a subsample of households.

If a fertility module is used, it is worth taking advantage of the information (Bhushan and Oliver, 2000). The fertility module includes information on all children ever born from all female members of reproductive age. If possible, we recommend changing the reference group for current migrants to all sons and daughters of women in the households age 21 and above who are still alive and no longer live in the household, as was done in the 2002 Albania LSMS (INSTAT, 2002). This method excludes any children of women who are no longer present or who have passed away. As an advantage, however, this approach abstracts from endogenous social processes that influence household formation.

A similar approach, implemented in the 2002 EHRUM and the subsequent 2005 Albania LSMS, is to list in a separate module all adult children of the head of the household and/or his/her spouse, if no longer living in the household, regardless of when they left. The list should include all sons and daughters that failed the household membership criterion in the household roster and include them even if the mother is absent or no longer alive (INSTAT, 2005, Carletto and Azzarri, 2007).<sup>28</sup>

One might instead consider listing every previous household member since the household was set up, whether living there presently or not. These people would almost certainly include children of the household head who moved away, as well as other family members who had lived there. The advantage of this method is that the people enumerated would have links with the household, allowing the analyst to capture migration information about the largest possible group of individuals. The primary drawback of this method is that it is time-consuming and some of the information collected might be of particularly poor quality if links are no longer strong. Furthermore, the current migration status of some individuals who had previously lived in the household might be unknown. If the method we propose below cannot be implemented, we suggest using all former household members as an alternative.

We suggest using a modified version of the fertility module to capture information about the spouse of the household head as a frame for learning about migration. This module is linked with the first block of questions in Figure 2, which asks in question (I+2) about the number of children no longer residing in the household of each adult in the household, including the household head and spouse, even if absent. The top block of Figure 5 then asks a few questions about each of these children, to determine their status as a migrant. By following the children of each adult in the household (including the head and spouse), the analyst can abstract from the

---

<sup>28</sup> A slight alternative is to expand the roster only to include all sons and daughters of the household head, regardless of whether or not they still live in the household and whether or not they are adults. This method has been applied to internal migration in a nearly nationally representative survey collected in China (de Brauw et al., 2002).

definition of the household to capture most migrants with strong associations with the household.<sup>29</sup> The bottom block of the questionnaire asks specifically about the household head and spouse, to capture information about them if they have migrated. The enumerator learns in question (I+1) of Figure 2 whether the head or spouse is an internal or an international migrant.

The remainder of Figure 5 illustrates a set of questions that can be asked about potential and current migrants. After listing the gender, age, and location of each child's mother and father, the questionnaire asks about the year of the migrant's departure and their location. If he or she does not live outside the district or immediate geographical area, one skips to the next person. Other than the time of departure and the location, these questions are skipped for the head and the spouse. Second, the module asks when the individuals reached their current location, a reason code, and about the individual's occupation prior to migration. The next set of questions asks about the migrant's education level and marital status, whether or not the migrant's spouse and children accompanied the migrant, the migrant's employment status, and the migrant's present occupation. Fourth, the module asks whether the individual has been away from the household before, and if so how many times. These questions should be modified if the module is being asked of internal or international migrants exclusively. Fifth, the module asks about remittances, including questions on whether or not remittances were sent, the frequency of remittances, the total value of all remittances, and the method of sending remittances. Finally, the last few questions ask about communications between the household being interviewed and the migrant. These questions help further describe the ties between migrant and household. The questions can be modified for the context of the survey, but we always recommend asking about the age, gender, relationship, location, and occupation of the migrant. For occupation, we always recommend asking about both the present occupation as well as the occupation prior to departure.

#### *4.1.6. Key Points*

Several points must be carefully considered when deciding how to collect information about migration experience among household members and other individuals closely linked with sample households. One must consider:

- The types of migration that are frequent in the context being studied, and how to define each type of migration that will be enumerated in the survey;

---

<sup>29</sup> The method misses siblings of the head if the head's parents do not live in the household, for example. However, asking about all siblings of all adult household members would be costly in terms of time, and would not likely yield much additional information about migrants. Moreover, if a goal of the survey was to develop an estimate of the number of emigrants from the country, expanding the list in this section would add to the probability of double counting migrants.

- Decisions about collecting information on individuals failing the residency rule, such as:
  - For which individuals associated with the household who fail the residency rule information about migration will be collected, and
  - How much information to collect about those individuals, considering that this information must largely be collected through proxy respondents.
- A rule for collecting information on individuals who were not present for the past full twelve months. We advocate collecting as much information as possible about those individuals, using the modules described in the next sub-section, if time allows.
- Collecting information on the initial migration, and then asking about details for each migration spell occurring during the past 5 years. The detailed recall period can be expended if framed by a notable event in the recent past and the present.
- The tradeoffs between the quantity of information learned about migration against both the quality of information obtained and the interview length.

## Box 2. Tracking Migrants or Households Migrants Left

A way to collect more exhaustive information about emigrants would be to track them in the destination country. To allow for this possibility, detailed contact information must be collected from the original households, including addresses and telephone numbers where the migrant can be reached. Ideally, the tracking survey should occur within weeks of the survey in the sending country, given the high mobility of these migrants. Tracking surveys of this type have been carried out in some countries, including between Mexico and the US, and between Albania and Greece. Alternatively, one can first carry out a survey of migrants in destination countries and, using a similar approach, track down the original household in the sending countries. An example is the aforementioned study by McKenzie et al (2006), between New Zealand and Tonga.

Tracking surveys can also be used as validation of information being gathered in the original household through proxy respondents, as well as to measure differences in perceptions between migrants and household members left behind. However, while allowing for direct interviews with the migrants, tracking presents a number of problems which often outweigh the benefits. Tracking surveys are generally too costly, and are characterized by high levels of attrition, particularly when the share of illegal migration is high. Minimizing systematic non-response by particular groups, e.g., illegal migrants, is crucial to a successful tracking survey, because systematic biases are notoriously difficult to correct.<sup>30</sup> In a recent survey of Albanian migrants to Greece, a list of contact information of migrants was created from the 2005 Albania LSMS. The phone numbers collected from the original households in Albania were first used by a team in Greece to make a first attempt at contacting the migrant and arranging for an in-person interview. If that failed, either because of refusal or because the migrant could not be found, the team based in Albania re-visited the original households and asked their assistance to contact the migrant by providing a phone card. In case of positive response, the Greek team would be informed to immediately contact the migrant and arrange for an interview. Despite much effort, however, the survey was successful in locating fewer than 50 percent of migrants in the original list.<sup>31</sup>

Other novel approaches have recently linked information about source households with migrants, although there is no way to extend either method to a nationally representative survey. First, Osili (2007) built a sampling framework of Nigerian immigrants to the United States using the Chicago telephone book to identify Igbo surnames. After randomly selecting from that frame, she then interviewed the immigrants in Chicago and tracked their source households in Nigeria. In studying fostered children in Burkina Faso, Akresh (2009) first surveyed households in rural areas, and then tracked any children fostered from those households to as far as Cote D'Ivoire, to compare educational outcomes between the children who left with the children who stayed.

---

<sup>30</sup> Although Korinek, Mistiaen, and Ravallion (2007) describe a method for correcting for systematic non-response, it requires that all variables that systematically affect non-response are observable.

<sup>31</sup> To reduce data collection costs in Greece, the original list only included migrants who lived clustered in groups of at least 5 observations, resulting in the exclusion of a considerable number of widely dispersed observations in small islands of the Greek archipelago.

#### 4.2. MEASURING DETERMINANTS OF EMIGRATION

Even if the goal of the study is to learn about the impacts of migration on a specific outcome or the interaction between migration and policy, it is useful to include modules in the questionnaire that will allow for the analysis of the determinants of migration. Several reviews of the determinants of migration are available, so we do not attempt to characterize all of the important motivations here.<sup>32</sup> It is important to note, however, that without a careful consideration of these determinants, any identification strategy used for migration may suffer from omitted variable bias. In general, information will be required on both migrant (treatment) and non-migrant (control) individuals and their households, both before and after the migration “treatment.” Clearly, some information on pre-migration conditions is needed. Assuming a dichotomous model of migration in which a migrant is identified based on an occurrence within a year, the pre-migration timing for migrants corresponds to the year prior to migration.<sup>33</sup> For longer reference periods, ideally one would want to collect information for each single year as the factors affecting migration are likely to have changed over time; however, that may not be feasible or too costly in most surveys. For non-migrants, when migration refers to a longer reference period (say 5 years prior to the survey) it is recommended to collect information relative to the mid-point of the chosen reference period, e.g. 2.5 years preceding the survey (Bilsborrow et al., 1997).<sup>34</sup>

Furthermore, the determinants of migration may differ by migration type. For example, factors motivating internal and international migration from the same country may differ. The differences may result from the fact that international migration is often more constrained. Formal constraints against migration in the form of anti-migration policies or discrimination in society and labor markets alter incentive schemes for migration. These schemes usually attempt to hinder the total amount of migration that occurs and as a result, many migrants enter destination countries illegally and without documentation. Whereas migrant networks may play a role in reducing the information and search costs of finding jobs in destinations, the type of networks available (internal or international) may further shape the characteristics of migration from a specific community. As a result, one might expect to observe significant differences between the characteristics of internal and international migrants.

---

<sup>32</sup> Examples of reviews that include discussions of the motivations of migration include Williamson (1988), Lucas (1997), Lalonde and Topel (1997), de Haan (1999), and Taylor and Martin (2001).

<sup>33</sup> It is worth emphasizing that different types of migration may have different motivations; for an example of potential determinants of seasonal migration, see Stark and Fan (2007).

<sup>34</sup> One chooses the middle of the detailed recall period because non-migrants implicitly chose not to migrate in each year during the detailed recall period.

The determinants of return migration are also notable and have received less attention in the literature. Household theories of migration can help analysts understand the rationale for return migration. Migrants may return home when they perceive that the discounted value of expected future returns to their labor at the source exceed those at the destination. This may occur because migrants have learned new skills at the destination that can be applied at home; they may have acquired capital that enhances or complements their human capital more at home than at the destination; or the wage differential may change over time (Dustmann and Kirchkamp, 2002; Dustmann, 2003). Purchasing power may also be higher for migrants at home than it was abroad (Dustmann, 1997). There are several further costs to migration that might lead migrants home. For example, migrants might have difficulty integrating at the destination, dealing with cultural issues in the host country or in the city, and/or they might simply become homesick. Household members, particularly aging parents, may become sick, drawing migrants home (e.g. Giles and Mu, 2007). Weather shocks, earthquakes, or other unexpected natural events might catalyze a return home. Migrants may also return home to marry locally. All or some of these factors may alter purely economic calculations of expected future returns, thus leading to migrants to return. From the household perspective, migrants may return as part of a household income generation strategy when financial or liquidity constraints no longer bind. Variables that may affect migration include:<sup>35</sup>

- Demographic and human capital characteristics, such as age, education, gender, or marital status.
- Differences in wage rates and employment opportunities between the home location and potential destinations.
- The availability of migration networks, including information about prior movements of family members, whether nuclear or extended, or the prior movements of peers within the source community.
- Explicit or implicit costs related to moving, as well as the distance to potential destinations.
- Household risk coping strategies focused on migration.
- The household's relative economic standing in the community, meaning that households that observe other households as relatively better off might have additional incentives to migrate to improve their living standards (Stark and Taylor, 1991).

---

<sup>35</sup> Portions of this section in particular are adapted from Lucas (2000).

- Explicit policies that either foster or inhibit migration, which are usually aimed at international migration.
- Moving to live with a new family, through marriage, or moving to reunite a family in a new place.
- The health of family members either who have migrated or at the place of origin.
- The relative availability or quality of facilities, such as educational facilities, at home and at potential destinations.
- Violence, disease, environmental disasters, or other environmental factors.

#### *4.2.1. Demographics and Human Capital Characteristics Determining Migration*

Identifiable characteristics of individuals, including age, gender, educational attainment or literacy, and marital status, are all typically correlated with migration status at the individual level. Correlations between these variables and migration are shaped by economic opportunities available to the relatively young, as well as the opportunities provided by educational attainment. Individuals who are young, relatively well-educated, and unmarried are more likely to be migrants than others. The relationship between migration and gender is more complicated, as international migration from some countries is dominated by women (Stecklov et al., 2007), whereas from other places migration is dominated by men. Internal migration may act similarly within large countries (e.g., China), though evidence is lacking on the validity of this hypothesis. Finally, ethnic minorities are typically less likely to be migrants.

The relationship between demographic variables and migration may be shaped by the analyst's definition of migration (Lucas, 2000). For example, if one is studying return migrants and a return migrant is defined as any individual who has ever migrated in the past, return migration will increase with age. However, the relationship will almost undoubtedly change if return migration is defined as anyone who has returned within the past five years; people in their late 20s or early 30s will likely be classified as return migrants more frequently.

Variables related to demographics and human capital characteristics are usually collected in the household roster (Glewwe, 2000a) and education modules (Glewwe, 2000b). In the migration section, it is important to collect demographic and schooling information about migrants and other individuals included in the survey, but excluded from the household roster.

#### *4.2.2. Differences in wage rates between the source and potential destinations*

When individuals or households make decisions about where to work, they almost certainly compare the wages and other attributes of employers at home against wages and attributes of employers elsewhere, including the probability of finding employment at the destination. While data on employment and wages are normally collected in an LSMS survey (see Schaffner, 2000, for a description of issues related to the employment module), two issues cannot be well addressed with this module. First, it is not possible to collect data from individuals regarding what prospective wages would have been elsewhere, or even what potential wages were available to each individual when choices were made about employment or migration. Moreover, if wage work was not available to the individual, one would want to be able to estimate their marginal value product on the farm. Therefore, information about differences in wage rates will always be incomplete. Second, for people who have out-migrated, one does not generally need data on potential employment opportunities, but one does need this data at the time that the migration decision is made. Since gathering such data involves collecting a lot of information via incomplete recall in a cross-sectional survey, panel data is likely required to begin to investigate this hypothesis. In the case of internal migration, a nationally representative panel (or repeated cross-section) could hypothetically be used to construct wage rates at potential destinations in the past.<sup>36</sup>

#### *4.2.3. Migration Networks and Information*

Migrants consistently use family and community networks to arrange migration, whether for internal or international migration, and the role of networks in migration has been well studied (e.g. Massey and Espinosa, 1997; Davis and Winters, 2001; Winters et al., 2001; Munshi, 2003; and Dolfin and Genicot, 2010). Networks can provide information that increases the potential net returns to migration in several ways. First, networks can provide information about jobs at potential destinations, or network members can arrange jobs for migrants prior to their arrival. Second, networks can provide migrants with information about the moving process, which is particularly important for illegal border crossings (Dolfin and Genicot, 2010; Hanson, 2007). Third, migrants often rely on network members when they initially arrive at the destination, receiving food, shelter, and advice on different customs and language in an unfamiliar country or city. Furthermore, gender may play an important role in networks, as women and men may learn about migrant employment separately, particularly as male and female migrants tend to perform different types of jobs in destinations (e.g. Stecklov et al, 2007).

---

<sup>36</sup> See Lucas (2000) for further discussion.

Family and community migration networks are often measured through proxy variables, such as the share of a household with migration experience or potentially the share of the village or community population that is emigrating at any given time. The migration history described above would be sufficient for constructing a household migration network proxy, but it would miss any extended family members that might also act as a migrant network. To ask whether migrants from the household might have used extended family networks to initially find work, one can add a short set of questions about the first migration experience, asking how the individual decided upon the destination, and if the migrant decided through referral, whether the referral took place through immediate family, extended family, or another community member.

Community network variables are best measured either through community-level surveys or through external data sources, such as a census. Unless accurate population registers are kept, variables constructed from data collected at the community level should be asked of a community administrator (e.g., a village leader or mayor) or another key informant who would best know about migration and other aspects of the local economy.<sup>37</sup> However, any measures of migration derived through community questionnaires are subject to measurement error. If large, statistically representative, labor force type surveys are available for each specific region and include migration, one could also attempt to measure a community network variable using such an external data source. Such information could also hypothetically come from the census. However, census data seldom include information about migration. One such variable that has been constructed from a census is found in McCarthy et al (2006), who use the census in Albania to construct a demographic pyramid for Albania as a whole, and then measure the migration network as the difference in the pyramid between each individual district and the countrywide demographic pyramid. To summarize, the use of external survey information may be more practical in cases when a time lag is desired in the variable, if the analyst plans to attempt to use it as an instrument. However, one should use such variables as instruments with extreme caution, as discussed below.

It is further important to note that the strength of migration networks should be considered endogenously determined (Carrington, Detragiache, and Vishnawath, 1996; Munshi, 2003). As more people leave a specific community, the cost of leaving that community decreases, since more and more information becomes available about potential migration paths and destinations. Unobservables that affect the information flow might also affect the outcome of

---

<sup>37</sup> One example of a place where accurate population registers are kept in rural areas is in China, where village leaders keep records of household populations, births, and deaths. See Rozelle et al. (1999) for an application of learning about migration in China through community surveys, and further discussion.

interest, which could be household-level migration. Even if the analyst decides to look at the net effects of migration on the local source economy by only measuring the effects of the community migration network variable, one must treat the variable as endogenous and find instruments for it (Munshi, 2003; de Brauw and Giles, 2006; 2008). This strategy averages the net effects of migration on migrant and non-migrant households, so one could argue that it understates the impact of migration on specific households that actually participate in migration. However, the advantage to this strategy is that one captures direct and indirect effects of migration on the source household, without having to worry about the selectivity of migration embedded in access to family networks, which most methods miss.

Finally, our discussion of the measurement of migrant networks treats migrant networks as something of a black box—it is unclear how the networks actually facilitate migration. Dolfin and Genicot (2010) find that households with stronger family networks are more likely to use *coyotes*, or migrant smugglers, to facilitate their illegal border crossings into the United States, perhaps because they have more information about employment. The internal workings of migrant networks are potentially a fruitful area for future research.

#### *4.2.4. Spreading out Risk through Migration*

Income risk may contribute to incentives to migrate, particularly in rural areas of developing countries. Agriculture is an inherently risky activity, and when rural households rely on agriculture as a primary source of income, households have incentives to diversify their income sources in order to reduce expected income variability (Binswanger and Rosenzweig, 1986). Migration can also be risky, especially if prospects of employment are unknown at the destination, but the risk is uncorrelated with agricultural risk. Therefore either internal or international migration is a good diversification strategy, because risks to agricultural income are not likely to be correlated with risks to income in urban areas or in other countries, whereas local employment can be affected by severe negative agricultural shocks. Unsurprisingly, there is a large literature demonstrating that risk is a determinant of migration. Rosenzweig and Stark (1989) show that by marrying individuals from other villages, rural residents in India use migration to smooth consumption. Paulson (2003) shows that migration from rural Thailand can be partially explained by an insurance motive; Gubert (2002) finds similarly for migration from the Kayes area of Mali; and Yang and Choi (2007) demonstrate that remittances have an insurance motive in the Philippines. Giles (2006) shows that risk to income decreases in China as access to both local and migrant labor markets increase, and Giles and Yoo (2007) show that as access to migrant labor markets improve, precautionary savings by households in rural China

decrease. Halliday (2006) studies how households mitigate risk after a shock in El Salvador. Finally, migration can be negatively affected by specific types of risk; using panel data, Giles and Mu (2007) demonstrate that poor health among elderly household members has a negative impact on migration.

To further investigate the relationship between migration and risk, one must measure risk, typically through data collected on negative shocks to the household. To incorporate information on shocks into a study of migration, one can either use an additional survey module that asks households to self-report shocks, a module or set of questions that asks about community-level shocks in a companion community questionnaire, or data from an external source. Self-reported shocks modules have not frequently been used in LSMS surveys, but they have been used recently, in the LSMS surveys in Malawi and Albania. Although an advantage of asking about shocks over the past three or five years is that one learns about intra-community variation in exposure to shocks, a potential drawback of using self-reported shocks is that they can be misreported. People tend to recall major shocks, but a potential problem is that people may have different internal definitions for the word “major” in defining shocks, so variables based on self-reported shocks are subjective.<sup>38</sup> Furthermore, it can be time-consuming to ask about shocks, particularly if households recall them in great detail. An alternative is to either use data collected at the community level about shocks or an external data source. Community-level data would necessarily mask any variation within the community in exposure to shocks, as would external data, such as weather station level data on rainfall. In the latter case, one might find it difficult to assign data collected at appropriate weather stations to the PSUs in the sample, as the distance between the nearest weather station and the PSU could potentially be quite large.

#### *4.2.5. Relative Deprivation or Inequality*

Whereas policy makers are particularly interested in the effects of migration on intra-community inequality, inequality within a community might alternatively be a push factor affecting the decision to migrate. Stark (1984) argued that one motivation of rural-urban migration might be to improve the relative standing of one’s household within a community. Relative deprivation hypotheses have primarily been tested in Mexico. Stark and Taylor (1989; 1991) find that in a sample of households surveyed in Michoacán, Mexico, relative deprivation is positively

---

<sup>38</sup> It is also worth noting that one can argue that certain types of self-reported shocks may be endogenous to migration. For example, a certain crop might do poorly if it is neglected after the migrant has left the household. Careful training of enumerators can help avoid this problem, by teaching probes about reasons that agricultural shocks occurred.

associated with migration to the US, but not internal migration. Quinn (2006) uses data from the Mexican Migration Project and argues that while relative deprivation motivates domestic migration, migration to the US appears to increase relative deprivation by his measures.

To construct measures of relative deprivation, one needs to construct a variable from a measure of wealth in the sending community.<sup>39</sup> LSMS surveys have a particular advantage in this regard, as they normally have a multi-stage cluster design, and so one can use information about the wealth of all other households in the sample to measure relative deprivation. The best available measure from an LSMS survey would be the distribution of consumption among members of the sending community. Since any such variable would clearly be measured with error, it would be worth constructing two different measures of relative deprivation in order to use one as an instrument for measurement error in the other; for example, one could also construct asset relative deprivation to use as an instrument. However, learning about asset holdings is also a challenge, as discussed below. Therefore, data requirements for investigating this hypothesis are significant, although surmountable.<sup>40</sup>

#### *4.2.6. Policies Inhibiting or Promoting Migration, at Source or Destination*

Policies that inhibit or promote migration at either the source or destination can affect the decisions of household members by changing the net returns to migration. These policies primarily affect international migration, although they have also affected internal migration in some countries, primarily in the past—China is likely the most studied example. Policies that can affect migration include procedures for obtaining official documents such as passports and birth certificates at the source, and the ease of obtaining employment visas for working in likely destination countries. Other policies may facilitate migration, such as licensing migration contractors, or information campaigns against trafficking. Whereas it is simple to add questions on the first few points to a migration module, the latter are likely to be very context-specific. However, it is certainly worth exploring whether migration policies have the intended effects (e.g., Gibson and McKenzie, 2007b).

From the perspective of the destination, many countries limit the number of migrants who can enter and/or work in any given year, making any further immigration illegal. In the case of

---

<sup>39</sup> See Stark and Taylor (1991) for a description of their relative deprivation measure.

<sup>40</sup> This discussion assumes that measurement error in the consumption aggregate can be assumed to be classical; in other words, one might be concerned that the measurement error in total consumption or consumption per capita is not mean zero with a constant variance. Gibson (2002) shows that the measurement error on consumption cannot be assumed to be classical. The measurement of food expenditures using a seven day recall, an important component of consumption, is correlated with household size.

migration from Mexico to the United States, Hanson (2007) argues that the additional illegal migration is more flexible and apt to respond to labor market based incentives than legal migration. In response, the Mexican ENADID asks whether migrants went to the United States undocumented, on a tourist visa, or on a work visa. As policies inhibiting migration to a specific destination have largely the same effect on every individual within a sending country, one either needs to identify the effect of policies through variation over time, different destinations, or potentially through variation in enforcement across space (e.g. Hanson and Spilimbergo, 1999). The data requirements for investigating such hypotheses are often beyond the scope of a typical LSMS survey, unless the country sends migrants to several different destinations and panel data is available on migration to those destinations.

On the other hand, host counties such as the Gulf States actually encourage temporary migration due to shortages of workers in relatively low skill sectors. In countries where active recruiting of migrants might take place, it is worthwhile including questions about whether or not migrants used recruiting services during their first or most recent migration episode. However, the addition of these questions does not make it possible to identify the impact of these programs on migration without additional data concerning the access of non-migrants to these services (Lucas, 2000).

#### *4.2.7. Reuniting with Family*

Migration does not solely occur for the purpose of finding employment; migrants often follow other migrants, sometimes to reunite with family members who previously migrated. In fact, the majority of visas given for legal immigration to the United States in recent years are for family members who follow previous migrants (Hanson, 2007). Rosenzweig and Stark (1989) argue that in India, women migrate internally to marry as a risk-coping device, finding that the correlation between agricultural risk in two places decreases as distance increases. The gender ratio of typical migrants in a specific migration flow may also change as the migration flow matures; if families follow initial migrants, men often make up the majority of the migration flow in early stages, while women follow men later, constituting the majority of the flow as it matures. One can infer whether or not an individual moved to follow family through the migration history, as it should indicate that another family member moved prior to the individual in question. If the migration took place in the too distant past to capture, it is worth supplementing this approach by asking about the primary reason for the most recent migration in the migration history. This question should list as potential responses both “for marriage” and “because parents moved” as potential answers. It is worth identifying which migrants left to reunite with other family

members, as the economic implications of migrating to reunite with family differ from labor migration.

#### *4.2.8. Educational Facilities or Other Local Public Goods*

Individuals sometimes migrate in order to take advantage of better educational institutions or other local public goods that are not available in their local area. Migration can be therefore be affected by the presence or lack of local public goods in the respective destination and source countries. Migration can also serve as a catalyst for the improvement of local infrastructure; schools that become overcrowded through immigration may respond by improving the school through the construction of additional classrooms,.

If one wants to understand whether individuals migrate as a response to a lack of available local public goods, it may be worthwhile to collect data on local facilities using a community-level or local facility-level questionnaire. However, unless facility census data are available, it is not possible to collect data on other facilities that migrants would potentially use, because the full choice set of migrants is unobservable. One could ask about the availability of certain types of schools; for example, individuals might migrate to board at a secondary or tertiary school that is not available locally. Therefore the community-level questionnaire should at the least include information about whether facilities such as secondary or tertiary schools exist. Since one cannot observe information about the menu of distant educational facilities available to any given individual, the best possible alternative is to ask about the primary motive for the most recent migration, including “for schooling” as an answer, in addition to inquiring about the specific location to which the individual moved. An alternative strategy is to consult secondary sources of information for the availability of services. If the educational module of the questionnaire does not include all of the individuals being asked about migration, one might also ask whether the migrant is still enrolled in school and the level of his/her schooling as follow-up questions if the individual indicates migrating for educational purposes.

Based on the discussion above, this topic may be more relevant for rural-urban migration, particularly with respect to migration for secondary or tertiary school. One could further investigate such hypotheses by including questions on the quality of local secondary schools, if they exist, or even the availability of or the distance to both primary and secondary schools. If schools are poor or non-existent, the availability of schools can be an important determinant of internal family migration. Although international migration does occur for tertiary schooling, it is unlikely that enough such migrants would be included in an LSMS to make statistical generalizations.

It is also difficult to identify migration for other purposes, although one might be comforted that poor health facilities, for example, are likely to be a relatively rare motivation for migration in a household and are unlikely to be picked up in a household survey. The same problems apply to health facilities as educational facilities.

#### *4.2.9. Violence, Disease, Environmental Disasters or Other Factors*

The primary motivation for migration may be local violence or conflict, the local spread of disease, or environmental disasters, which can include everything from mudslides to desertification. While these factors cause significant amounts of migration, it is unlikely that migration at the source would be captured in an LSMS survey unless the survey takes place soon after an event or as violence subsides. It is unlikely that a survey would be planned in an area and during a period in which extensive local violence or disease is a concern. Furthermore, environmental disasters tend to be localized, so any migration to move away from environmental factors would constitute a small portion of an LSMS survey, unless the planned sample size was relatively large. However, one can potentially learn about such migration at the destination, as the inclusion of questions regarding why immigrants to urban areas left their home areas can shed light on the amount of immigration that takes place within countries due to violence or environmental factors. That said, specialized migration surveys that take place in or near affected areas seem to be a better method of learning about migration in response to environmental factors.

### 4.3. MIGRATION, IMPACTS, AND POLICY—SPECIFIC DATA REQUIREMENTS

As noted, in addition to being able to properly measure migration, other parts of the LSMS survey are necessary to examine the relationship between migration, outcomes of interest, and policy. Many modules are always included, but to analyze some of the questions that one would like to address, the survey instrument may need to be modified, or entire modules included to help answer specific questions. In this subsection, we first highlight key questions of particular relevance to current policy debates involving migration and discuss how answering these questions can either make use of existing sections of a typical LSMS questionnaire, or require modifications and additions to the survey instrument. Where appropriate, we discuss further questions that might be appropriate to add to specific modules and point out the types of questions that an LSMS survey is not well suited to study.

#### *4.3.1. Poverty and Income Distribution*

The interactions between migration, poverty, and changes in the income distribution are of primary interest to researchers and policy makers, and have been well studied in the past (e.g. Lipton, 1980; Stark, Taylor, and Yitzhaki, 1986; de Haan, 1999). Yet because migration is difficult to identify statistically, few authors have been able to convincingly demonstrate a causal relationship between migration and either poverty or inequality. Furthermore, the relationship can potentially be studied in either direction; poverty may hinder migration, but migration may also lead to poverty reduction or changes in the income distribution. In studying how migration affects poverty or inequality, it is important to note that different types of migration may have differing effects on poverty and inequality. Furthermore, measurements at the household level must take into account the migrant's absence (Barham and Boucher, 1998). As a result of data requirements and the difficulties involved in identifying migration, research on the dynamics of the relationship between migration and inequality or income growth remains relatively limited (exceptions include McKenzie and Rapoport, 2007, and de Brauw and Giles, 2008). If emigrants are not accounted for in poverty estimates, poverty for some original group of individuals can also be over-estimated (Clemens and Pritchett, 2008). Some questions that one might use to motivate a migration study about migration, poverty, and income distribution include:

- Does participation in migration vary by poverty status?
- How does the distribution of migration opportunity (meaning the opportunity to migrate either internally or internationally) affect the income distribution?
- Does the absence of migrants in rural areas affect agricultural output and thus welfare, either at the household level or in the aggregate?
- Does migration create general equilibrium effects; e.g. when migrants leave, do local wages increase?
- If migrants do not send remittances, is poverty exacerbated (due to the loss of labor) or mitigated (because of reduced household consumption requirements) among those left behind?
- If pre-migration data on poverty is available, are poor households more likely to send out internal migrants than international migrants?

Questions about migration and poverty that specifically relate to internal migration include:

- Can rural/urban migration help decrease the income/consumption differential between rural and urban areas?
- If so, does rural/urban migration have a noticeable effect on overall poverty? Inequality?

To study migration, poverty, and income distribution, the primary requirement is an accurate measure of well-being. The preferred measure for wellbeing is consumption, which is an integral part of any LSMS questionnaire. Deaton and Grosh (2000) provide a detailed description of the issues one faces in measuring consumption, and Deaton and Zaidi (2002) discuss issues related to the computation of a consumption aggregate in household survey data in detail. In order to measure the causal relationship between migration and welfare, one would ideally want to know what the migrant household's per capita consumption level would have been had the migrant remained within the household. Panel data with information pre- and post- migration can be useful for studying migration and distribution as more is known or can be inferred about the counterfactual. However, panel data are seldom available and often cross-sectional datasets must be relied upon. Although the counterfactual is difficult (if not impossible) to ascertain in a cross-sectional study, one can potentially attempt to learn about it by collecting information on pre-migration conditions, such as measures of asset holdings, that can be reconstructed using recall methods. These measures can also be useful for understanding the relationship between migration and, for example, relative deprivation.

Migration might also have general equilibrium effects on the within-community income distribution. When migration occurs, one expects local wages, either explicitly or implicitly, to rise for the types of workers most likely to migrate. Therefore, migration may have more complex effects on the income distribution than the direct effects of a migration variable on inequality or income. In a cross-section, one could use questions about wages found in the labor allocation module to investigate whether communities with more migration have higher increases in wages than other communities, if information on wages in the previous period are also collected. Otherwise, panel data are necessary to investigate general equilibrium effects.

#### *4.3.2. Migration and Employment*

Migration for employment can have several effects on the economy beyond the direct change in location of some workers. First, migration may change the household labor composition. In Albania, Miluka et al (2007) found that international migration decreases overall labor inputs into farming. Similarly, in the Kayes area of Mali, Azam and Gubert (2006) report that individuals left behind by international migrants can exhibit shirking behavior in terms of working in agriculture. Within China, de Brauw et al (2008) find that as internal migration increased, overall agricultural labor inputs decreased, but the average share of agricultural labor inputs provided by women did not change. Migration can also affect child labor; on one hand, households with out-migrants may become wealthier, decreasing the need for child labor; on the other hand, migration

could lead to labor shortages on the farm, potentially increasing the need for child labor. Migration need not only affect the households that migrants leave; at the source, emigration can have general equilibrium effects on wages (Hanson, 2005; Mishra, 2007). Finally, migrants obtain skills during their time away which can prove useful upon return to the household.

Migration itself can be affected by changing opportunities at potential destinations. As economies develop, agriculture becomes a less prominent component of the economy, and new jobs in the industrial and service sectors become available. Additional jobs in the emergent sectors of the economy can change the incentives underlying migration. Trade liberalization through multilateral institutions such as the WTO can also change the returns to labor within an economy, potentially affecting migration. Governments sometimes create programs in order to stimulate the demand for labor, such as training programs or rural works programs. In designing research on migration, some of the following questions might be of interest:

- Are children or women in rural households that migrants leave more likely to work on the farm? Are there specific tasks that women or children begin to complete on the farm that they did not do before?
- Does child labor use decline or increase in households with access to migration or migrant earnings?
- Does migration, through skills acquisition, promote upward mobility upon return?
- In a growing economy, does internal migration occur as industrial or service sector jobs are created, and if so, how does it affect wage rates?
- Similarly, in a growing economy how does international migration affect job creation? Wage rates?
- What is the magnitude of the impact of emigration on wages or employment levels? Does migration lead to worker shortages in specific employment sectors?
- Do urban job creation programs or training induce migration from the rural sector?
- If rural development programs are successful, do they hinder migration? To what extent?

LSMS surveys are well-suited to study the impacts of emigration on wages in specific sectors, labor mobility across sectors, and whether or not skills or capital acquired through migration foster upward labor mobility among individuals. Aside from migration information and data collected for potential identification strategies, the only requirement is to include wage rates and job types in the employment module. For people with migration experience, it is particularly useful to collect information on occupation prior to migration and occupation at the destination, as well as the present occupation for return migrants (see Schaffner, 2000, for examples). Here, a

survey that includes oversampling of migrants might be particularly useful from the perspective of statistical power, particularly as the number of job types enumerated increases.

LSMS surveys are not ideal to evaluate job creation or rural development programs, but to the extent that variation exists in their implementation, an LSMS survey designed to study migration can shed light on how such programs affect migration or how the results of such programs are affected by migration. In general, the only data requirement to investigate relationships between migration and such programs is to include questions about participation in any programs.

#### *4.3.3. Human Capital Formation*

In some countries, costs of schooling are quite high relative to per capita income. As households also face capital and labor constraints on their activities, individuals often leave school before they would otherwise if schooling was cheaper or if capital markets functioned perfectly. Migration can help loosen the capital constraint through remittances that allow households to finance education (e.g., Mansuri, 2006a), although it simultaneously tightens any labor constraints on household activities (e.g., Cox-Edwards and Ureta, 2003). Another factor influencing the relationship between migration and human capital formation is opportunity cost. For individuals approaching the age for leaving school (or the mandatory minimum level of schooling), the opportunity cost of schooling increases as migration becomes more common because of the availability of migration as an outside option (e.g., de Brauw and Giles, 2006). Research questions about the link between migration and human capital formation include:

- Does migration lead to higher education enrollment levels, especially among younger children?
- Does migration increase the likelihood that individuals leave higher levels of schooling?
- Do individuals move to access better educational or health services? Or do whole households move?

Migration may have differential impacts on human capital formation at different points in the life cycle. From the perspective of a teenager deciding whether or not to continue schooling, the possible necessity of migration can add to the opportunity cost of continuing schooling. To generally understand the impacts of migration on school enrollment or educational attainment, one should use questions on human capital (see Glewwe, 2000b). To learn whether migration affected the opportunity cost of schooling, information queried may include the age at which individuals left school.

Some educational outcomes upon which migration might have interesting impacts are not well-enumerated in the standard LSMS format. For example, migration could potentially affect school achievement, drop-out rates, or school repetition, either negatively through parental absence, or positively through wealth effects. To study such outcomes, one would need either to modify the survey to ensure that questions of interest can be answered (see Glewwe, 2000b) or administer specific tests to children as part of the survey. One can also potentially use administrative data on test scores, combined with household survey data, to answer questions about migration and variables related to child achievement. Panel data again can be useful for understanding the causal effects of migration on human capital formation as one can then control for unobserved child ability.

#### *4.3.4. Brain Drain, Gain, or Waste*

A set of major topics revolving around migration and government policy can be characterized as issues related to “brain drain” (e.g., Bhagwati, 1977). From the perspective of international migration, if highly skilled migrants leave a specific country to migrate, when they do not return, positive externalities due to the human capital investments subsidized by the source country are lost. Skilled individuals who migrate often lack strong incentives to return because skill prices are much higher in the developed world than in their home country. In fact, incentives not to return are often embedded in the policies of developed countries pertaining to migrants. For example, in the United States, immigration policy makes it relatively easy to obtain a visa for family members following an immigrant who is working, whereas an immigrant who leaves the United States after completing a job will not be able to obtain another work visa without first arranging another job. One can also localize the concept of brain drain by considering the economic implications of rural areas losing their more productive members as a result of internal migration. Brain drain is a pervasive issue, and it has led to shortages among classes of workers in specific countries (Adams, 2003). A prominent example is Guyana, from which 89 percent of skilled workers have migrated (Docquier and Marfouk, 2005).<sup>41</sup> Such shortages are particularly important in the public sector, where wage rates may not be competitive with potential earnings overseas. In order to avoid brain drain, several developing countries have scholarships available to their citizens to pay for education abroad, which are forfeited if the individual does not return upon completion of the degree. Alternatively, some authors have demonstrated potential positive

---

<sup>41</sup> Clemens and Petersson (2007) describe a new data source that demonstrates the broad migration of African health personnel to developed countries.

externalities to brain drain. Mountford (1997) shows that a brain drain can increase average productivity in an economy, particularly when migration is not certain.

More recently, the theoretical concept of a “brain gain” from migration has been hypothesized. In general, the theoretical work requires specific conditions to be met before a brain gain occurs. Stark and Wang (2002) require a social return to education to show that migration may induce additional human capital investment by raising the expected returns to additional schooling, whether the returns would be realized locally or in migrant destinations. Boucher, Stark, and Taylor (2005)’s findings suggest that some brain gain in rural Mexico is a result of increased human capital investment. On the other hand, Schiff (2005) shows that when the partial equilibrium results that drive brain gain theories are extended into a static general equilibrium framework, the gains are much smaller. One might also argue that brain gain can occur from skills acquired while abroad, assuming that migrants return; Dustmann and Kirchkamp (2002) can be construed as showing brain gain occurring in Turkey due to the return of migrants from Germany. Gang, Co, and Yun (2000) show positive returns to migration among female migrants returning to Hungary, but not among males.

Lastly, “brain waste” can occur from two different perspectives. Consider the notion that each country has an optimal share of skilled, relative to unskilled, labor for its specific capital endowment and/or managerial capacity. If a small country over-invests in a specific type of human capital (for example, engineers) and also restricts labor movement, unemployment rates will be relatively high among workers who acquire that type of human capital. The source economy would then theoretically benefit by allowing those individuals to out-migrate and realize returns to their human capital elsewhere, as they might be expected to send remittances back to the source economy.

Brain waste might also occur when highly skilled migrants cannot find employment commensurate to their skills after migrating. Mattoo, Neagu, and Ozden (2008) show that highly skilled immigrants to the United States from Eastern Europe and Latin America tend not to be able to find employment for which they are trained, whereas immigrants from OECD countries and Asia are more likely to find such employment. This serves as an example of brain waste, as these migrants could use their human capital more effectively were they employed in the sector for which they are trained. Some questions of interest related to brain drain, brain gain, and brain waste that will almost certainly vary from country to country include:

- To what extent do relatively skilled workers migrate?
- Are there shortages of specific types of skilled workers in important industries, such as health care?

- Is the migration of highly skilled individuals correlated with increasing enrollment in higher levels of education?
- Are specific types of (skilled) workers absent from or scarce within the source economy?
- Are highly skilled workers underemployed in the source country? At the destination?

LSMS surveys are well-suited to learn about brain drain, brain gain, and brain waste due to migration from the source country perspective. Through the education module one can characterize the educational background of a country's (or region's) out-migrants relative to the remaining population (Glewwe, 2000b). One can attempt to relate past migration variables constructed from the migration history to school enrollment to study brain gain. If the survey includes a detailed employment module, brain waste can be studied from the source country perspective by examining job types held by individuals in different educational and experience cohorts. An LSMS survey that oversamples migrant households could be particularly informative in studying either brain drain or brain gain, as the distribution of educational attainment among migrants would be measured with less error.

On the other hand, LSMS surveys are likely to be less informative about whether brain drain is occurring in specific occupations (such as nursing or health care), or about whether brain waste is occurring in destination countries. For learning about the brain drain issue from the perspective of specific occupations, a specialized survey targeted at individuals with specific types of training (e.g. nursing school) would be more informative. It is further unlikely that job types in destination countries can be enumerated with enough detail to learn about brain waste among out-migrants.

#### *4.3.5. Migration, Gender and Gender Related Policy*

Whereas there is a rich tradition of studying the relationship between gender and migration in the social sciences, the differences between the motivations and implications of migration among men and women has not received the same amount of emphasis in economics (Martin, 2004).<sup>42</sup> Male and female migrants may have very different motivations for migration. They may use different networks for migration (Davis and Winters, 2001; Stecklov et al., 2007). As gender roles may differ within households, the migration of men and women may affect household level outcomes very differently (Pfeiffer et al, 2007). The same types of policies and migration laws may affect men and women differently (Richter and Taylor, 2007). Remittance behavior among male and female migrants may differ. Yet these differences and their underlying rationale are

---

<sup>42</sup> An exception is the volume edited by Schiff, Morrison, and Sjoblom (2007).

largely under-researched. All of these potential effects are likely to be context-specific, as gender roles differ widely between and even sometimes within countries. Some potential questions that could motivate research in this area include:

- Do wage differentials between the source and the destination have the same effect on migration among men and women?
- Do men and women respond the same way to push and pull factors influencing migration?
- Are women or men more likely to return? What factors determine return migration, and do they differ by gender?
- Does household production of non-tradeables or “z-good” production, including healthy and well-nourished children, suffer more when women migrate relative to men? Or does it increase due to the increase in household income and potential differences in propensities to remit?
- Do women migrate to follow men, or to follow economic opportunities? Does the rationale for migration by women relative to men affect household production or welfare? Does it affect decision making about production among rural households?
- Do male and female migrant networks affect male and female migration differently? Do separate male and female migration networks even exist?
- How do customs or patriarchal societies affect the gender composition of migrants? Do they affect the use of remittances?

Two particularly interesting lines of questions related to gender can be answered using LSMS surveys. First, with enough migrants in the sample, one can study the determinants for different types of migration by gender. The determinants almost certainly differ by gender and the data requirements for investigating this hypothesis are no different than data requirements for studying the general determinants. Second, migration almost certainly affects intra-household time allocation. From a policy perspective, such questions can be important if households begin to spend less time working in agriculture. The inclusion of a time use module, which is asked of each individual present, could shed light on the way that migration affects time allocations (Harvey and Taylor, 2000). As unobservables about household decision-making processes could also affect time allocation within the household, panel data might be particularly useful for studying how migration affects time use.

#### *4.3.6. Social Implications of Migration*

Migration can have further social implications beyond the issues already discussed. For example, when individual migrants leave home, couples and families are separated. In the aggregate, lower fertility rates may result. Child development may also be affected as children are largely raised by one parent—or if both parents migrate away, no parents. Among such children, school performance or psychosocial health may be affected by the lack of parental presence. The potential hardship of raising a child may cause the family members left behind to try to join the migrant at the destination if resources become available.

Second, as migration networks mature, cumulative causation can occur (e.g., Massey et al., 1998). In communities where migration has taken place over a long period of time, both the social and economic costs of migration decline, as connections between the source community and destination build. Furthermore, migration may become part of a “rite of passage” among young adults in a community. Questions to be considered include:

- Does the absence of a parent negatively affect educational achievement? Social and psychological development among children?
- Do areas with more patterned, historical migration have lower fertility rates?
- Does migration become a part of a “rite of passage” in communities with long-standing migration? Does this rite differ by gender?
- Do families follow migrants? If so, do public resources, in the form of infrastructure investments, need to follow migrant families to cities in order to provide enough schooling for migrant children?

Migration can have a number of wide-ranging social implications that have not been carefully studied in the past by economists, and LSMS surveys can be structured or modified to examine several of them. For example, removing a parent from the household can potentially affect children by hindering their psychosocial health or affecting their cognitive development. More generally, migration can affect mental health among any household members exposed to migration (Stillman, Gibson, and McKenzie, 2009). These concepts can be measured using tests targeted at the individuals potentially affected by migration. If such tests are used, it is important to validate them prior to field implementation.

Fertility histories, often included in LSMS surveys, can also help analyze social implications of migration. In combination with migration histories, a fertility history can be used to demonstrate whether or not migration has an impact on fertility, as is often suggested. However, fertility changes take place over a long period of time, and therefore changes in fertility are difficult to capture in a single survey or even in a panel of reasonable length.

#### *4.3.7. Policies Encouraging Return Migration*

As the populations of developed countries age, some developed country governments are realizing that unskilled migrants can help fill gaps in the labor market that native workers would not fill, particularly for prime age workers. As individuals who migrate usually keep close ties to the households and communities they leave, policies that encourage temporary migration followed by a return can help the economies in both the source and destination countries. Although in the case of international migration, migrants who initially left rural areas might decide to return to an urban area in their home country instead of their home village or community, migrants still bring back ideas and capital. As more migrants choose to return, interest in their potential role in the development process has increased. Some research questions of interest include:

- How much return migration actually occurs, relative to out-migration, from either an internal or international perspective?
- What are the characteristics of return migrants and of those migrants who decide to invest back home?
- What factors motivate migrants to return either home to the rural area or to the source country at large? What are the human capital characteristics of return migrants?
- Do migrants who return invest in housing or microenterprises? Or do they wait to leave at the first chance they have to leave again?
- Do the least or most successful return, and what are the implications of this selectivity for source and destination communities?

An LSMS survey should cover several important issues in order to study the implications of return migration. First, it can be useful to directly ask whether individuals with migration experience plan to migrate again. Although plans are not the same as actual migration after the survey, individuals who answer negatively can be labeled as truly returned migrants. To better learn about whether former migrants have returned long-term, one can further ask about what steps have been taken towards planning the next move, and about links (such as the presence of family and friends or investments) that the migrant might still have with the destination country or area, either of which would facilitate the next move and make it more likely. Second, it is important to ask about the activities of a returned migrant, in order to establish that the individual is truly a returned migrant committed to staying; for example, if someone has been home for several years and/or has established a business, they would be less likely to migrate again.

Governments may want to encourage return migration as it returns valuable human capital that may have been further enhanced since the migrant left. If programs exist to encourage return migration, then the questionnaire should explicitly ask about whether migrants took part in any such programs upon returning. However, one might also want to know about the likelihood of return for current out-migrants, which is difficult to know without a survey in the destination country. One can add questions to the survey about the integration of the migrant in the destination country or area. Specifically, one can ask whether or not the spouse lives with the migrant, if the migrant has children who were born or live at the destination, how often the migrant returns home, and whether or not the migrant has investments at home. If children live with migrants away from home, in cases of international migration or countries where multiple languages are spoken, one should also ask whether or not children living at the destination speak the native language.

#### *4.3.8. Migration and Social Protection*

Migration may have particularly interesting interactions with social protection programs. More and more governments are using cash transfer programs as part of their anti-poverty policy, and migration behavior may change in response to social protection payments, or potentially in response to the conditions placed on the receipt of those payments. One can imagine that the relationship between migration and public cash transfers could be either positive or negative. From the positive perspective, cash transfers could be used to help finance migration (Angelucci, 2005); on the other hand, cash transfers lower the risk associated with income, which might reduce the probability of migration (Stecklov et al., 2005). When conditioned on school enrollment, cash transfers can also increase the opportunity cost of migration among specific age groups, decreasing the probability of migration (Angelucci, 2005). It is further possible that private transfers from out-migrants are “crowded out” by public transfers, though the evidence is mixed (see also Cox and Jimenez, 1993).

A second set of issues related to social protection relates to old age insurance or pensions and health insurance. Many governments provide some benefits for elderly residents, and some governments provide health insurance to residents. The portability of benefits for migrants has become an important issue, particularly in Europe (Holzmann et al. 2005). Migrants receiving portable benefits might be more likely to return home if they have paid into pension schemes; if they will not receive benefits, the lack thereof might dampen the probability of migration. Furthermore, measuring these benefits may be important from the perspective of remittances. Some questions that might interest analysts include the following:

- Do social protection payments make it more likely for households to send out migrants because they are better able to finance the fixed costs of migration?
- Do social protection payments make migration less likely because local income becomes less risky or some other migration determinant changes?
- If transfers are conditional on school enrollment, does social protection delay migration, because potential migrants become more valuable to the household when they stay in school?
- Do migrant opportunities decrease participation in social protection programs?
- Do public transfers crowd out migrant remittances, either directly or indirectly by making households ineligible for social protection?
- If public transfers are available, is the quality of remittance data affected, as incentives for underreporting potentially increase? Is the quality of other data on potential migration impacts affected, such as data on investments?

To analyze the interactions between migration and social protection with an LSMS style survey, one may include a short module in the questionnaire on whether or not households received any type of public transfers, and amounts of transfers if so. It is important to use specific names of all potential transfer programs, and to know some details about all cash transfer programs that might exist; these questions can be included in the community-level form. If such questions are necessary, it is essential to pilot-test them prior to including them in a survey. Because many cash transfer programs have undergone or are undergoing rigorous impact evaluations, these surveys may be a source of interesting data to investigate such hypotheses in countries with significant migration, although any inference would lack national representativeness. From the perspective of learning about pension portability or the relationship between benefits abroad and migration decisions, such questions can be included in the migration module.

#### *4.3.9. Migration, Infrastructure, and the Provision of Public Services*

The provision of new public services, such as the improvement of irrigation structures or building roads, schools, or health facilities, may shape the probability of migration in specific areas. Since causality between migration and incremental changes in infrastructure or public services can run both ways, it is particularly difficult to demonstrate causal relationships, but the study of such associations is worth discussing nonetheless.

For example, investments in roads or other transportation networks have a potentially complicated relationship with migration. Better transportation infrastructure to a sending region

lowers the costs of migrating, and as a result one might expect an increase in migration. However, local economic activity might also be catalyzed by better transportation, resulting in higher returns to labor and mitigating the effect of lower transportation costs. Furthermore, while they are being completed, road projects create jobs that may be filled locally or by migrants from other areas; either scenario creates multipliers in the local economy. Therefore, the decision to migrate can be catalyzed by the completion of newer or better transportation networks, but the building of roads or transportation networks might also cause migration into the area. The simultaneity problem would be a difficult one to disentangle with a household survey.

Decisions about the placement of public goods can affect and be affected by internal migration. Policy makers are often concerned about the effects of migration on congestion around or in the use of public works in urban areas, such as roads, electric grids, and sanitation. If individuals or families are free to migrate between rural and urban areas, policy makers need to project the amount of migration that will take place to optimally plan for public investments, as high migration rates necessitate additional urban investments in public works. Alternatively, if international migrants are leaving specific areas of a country more than other areas, one might want to adjust public investments to best serve the population.

Another type of infrastructure that can have a complicated relationship with migration is education and the provision of schooling. Individuals often board at secondary and tertiary schools distant from home, in order to take advantage of better schools or simply because a place was offered at that school while the supply of post-primary schooling was constrained. In fact, secondary and tertiary schooling is often nonexistent in rural areas of less developed countries, and individuals who want to continue schooling must leave home. Children may even be sent to live with relatives to go to school if schools in their local area are inadequate. Particularly in more developed countries, parents may choose to move their families in order to ensure high quality schooling for their children. Alternatively, parents may send their children to either live with relatives and enroll in school in a distant area, or board at a distant school; such individuals are likely picked up in multi-topic survey data as migrants. Some potentially important research questions include:

- Does investment in irrigation lead to decreased migration? If so, what is the magnitude of the decline?
- How does road development affect migration, or how does migration affect road development?
- What is the relationship between the magnitude of migrant stocks and public goods investments? Do investments in public goods change migration flows?

- Are children lacking access to secondary or tertiary schooling less likely to enroll? How does the lack of access affect their schooling potential?

In general, LSMS surveys are not always well-suited to learn about interactions between migration and infrastructure programs, because such programs tend to be geographically specific. However, if community-level questionnaires include information about public services, LSMS surveys can shed light on the influence of migration on public services at the community level, especially if those surveys include innovative measures of public services that are pre-tested by analysts. LSMS surveys that focus on migration can also help inform the development of policies relating to the location of public services, because they hypothetically allow analysts to more completely characterize migration.

#### *4.3.10. Migration and Household Investments*

Migration is inherently a dynamic process, taking time for migrants to reach their destination and get settled. Migrants often send or bring home their remittances, making them “lumpy” transfers, in part because transaction costs for sending money home can be quite high for international and internal migrants. Therefore, households may be able to use migration to overcome a credit constraint keeping them from expanding their economic activities (e.g., Mesnard, 2004). Migrants or migrant households might alternatively wait to finance any investment until returning home, choosing to then invest in building or purchasing a better house, or durable goods that improve their standard of living (de Brauw and Giles, 2008). In several contexts, return migrants have been shown to start businesses upon return (Mesnard and Ravallion, 2006; Woodruff and Zenteno, 2007; Kilic et al, 2009). As a result of human capital potentially acquired by migrants abroad, such investments might result in firms that are more likely to survive (Lucas, 2005) or in firms with more start-up capital on average, as found in Egypt (McCormick and Wahba, 2003). Firms begun by return migrants also generate local employment (Lucas, 2005).

Other types of investment are also possible; for example, return migrants might invest in higher value crops than were possible without the capital earned while migrating. Migration might facilitate investments in education, positively affecting school enrollment; through the infusion of liquidity, households might be able to send more children to school. Migration might also foster investments in the health of very young children (Mansuri, 2006b). Several policy related questions arise:

- Does migration foster investment in productive activities? If so, what types of businesses tend to be owned by migrants?

- Do skills acquired at the migrant destination help migrants in the labor market or as entrepreneurs upon return? Does migration foster upward labor mobility?
- If return migrants are entrepreneurs, do their businesses generate more or less employment than other businesses, and are such firms more likely to survive? What general equilibrium effects do they create?
- Do general equilibrium impacts of migration encourage or discourage investment, and through what channels?
- Do migrant households or households with return migrants invest in more housing or durable goods?
- Do migrant remittances help increase the number of children attending school, or allow them to attain higher levels of schooling?
- Do transaction costs for migrant remittances hinder rural investment?

To fully study the correlations between migration and investment using an LSMS survey, several modules must be included in the survey form. If some modules are not included, one would be concerned about selection bias, as the household would have potentially selected one type of investment over another. To fully study these potential investments, one must ensure that productive asset holdings are fully enumerated, along with recent purchases of assets (for agriculture, see Reardon and Glewwe, 2000). If studying non-agricultural assets, it is important to include questions that ask enumerators to record the date of purchase of all recent and large purchases, as well as to include a question about any large assets that might have been discarded in the reference period. One should include a measure on housing (Malpezzi, 2000), which again asks about timing of any large improvements, and a reasonably detailed form on self-employment (Vijverberg and Mead, 2000), which enumerates the timing of investments into the business. For the study of investments in health, one should include anthropometry, which can also be useful as an alternative indicator of overall well-being (Alderman, 2000). In all of these modules other than anthropometry, it is crucial that all questions about asset purchases and investments are asked that allow the analyst to discern the timing of investments relative to migration spells. In a cross-sectional survey, including all of these modules and asking about the timing of purchases can be quite burdensome for the respondent, and one might be concerned about recall bias affecting the answers to questions. As a result, for several issues, panel data is usually better suited to studying the effects of migration on investments, as changes in assets or microenterprise activity become more obvious and can be more convincingly linked to migration.

#### *4.3.11. Migration and Environmental Displacement*

Environmental changes can be a primary push factor for migration.<sup>43</sup> In areas that are affected by desertification, the returns to labor on the farm may be in steady decline, despite any adjustments the household might make towards its new environment. New risk environments might also cause migration based on environmental displacement; for example, if the risk of flooding has recently increased, farmers may determine that moving out of their homes to new areas is their best alternative for maintaining their livelihood. Finally, projects such as dams can cause governments to resettle people to other locations, making compensation and their livelihoods at their new locations important issues for poverty policy (Dreze, Samson, and Singh, 1997; Duflo and Pande, 2005). Several policy questions arise when considering migration due to environmental displacement:

- What types of individuals or households are more likely to migrate as a result of environmental changes?
- Do the environmental impacts of migration mitigate or reinforce the incentives to migrate over time?
- How can governments better organize the process of migration due to environmental displacement so as not to overcrowd public resources in other areas?
- Can land be found to give to the environmentally displaced? How does that affect land rights among others?
- What are the implications of advocating policies to push those who are at severe risk of being environmentally displaced to new areas?
- If people are displaced by projects, how much compensation must they be given to ensure they do not become poor (or more poor)?

Environmental displacement is usually localized, so LSMS studies may not be particularly well suited for answering questions about environmental displacement. As both migration and environmental displacement are rare events, one would not expect much overlap between the two in a standard nationally representative sampling frame. Therefore, localized surveys are likely to be more effective at answering questions related to the interactions between the two.

One could use an LSMS survey to inform policy related to environmental displacement. For example, consider a relocation program that must take place because of soil degradation. From information enumerated about land area, plot sizes, yields, and land rental rates in the

---

<sup>43</sup> Localized violence can also motivate migration; for example, see Funkhouser (1997).

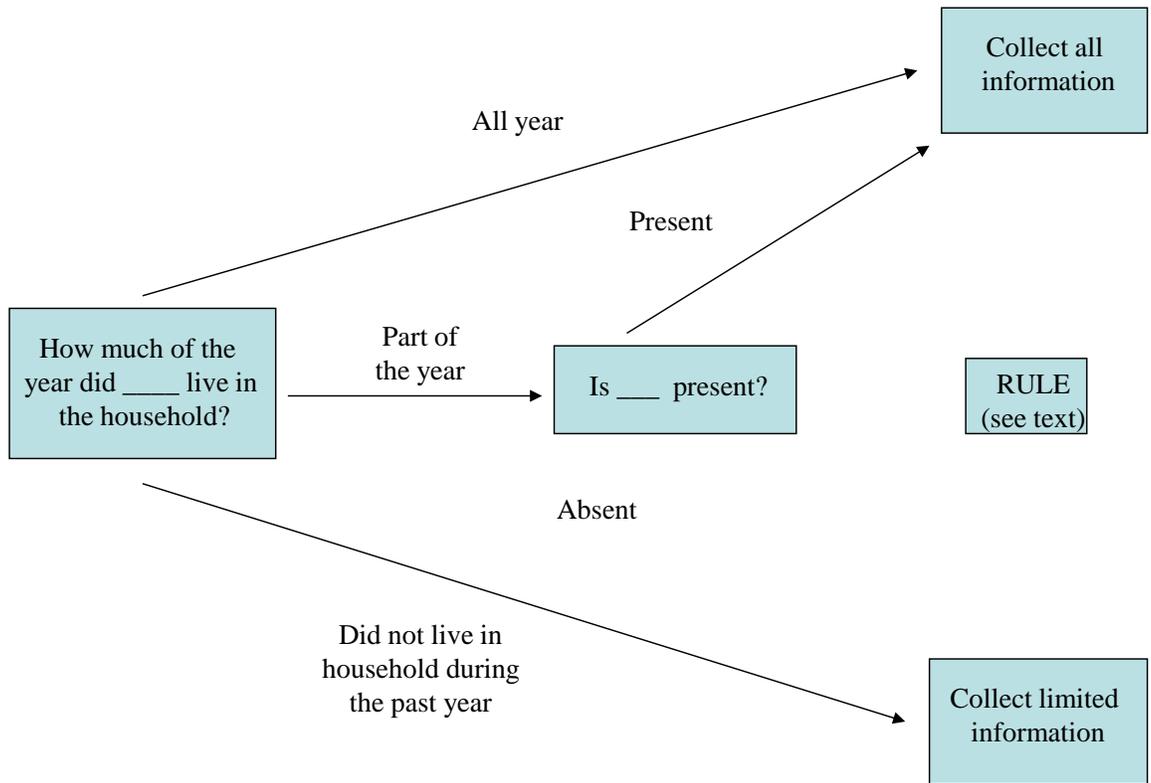
LSMS, one could estimate the cost of a policy to relocate individuals to other lands, if such areas were available.

## **5. Conclusion**

Migration is an increasingly important facet of many developing economies, yet little solid empirical information on the relationship between migration, policy and welfare is available in the existing literature. Research on migration within the multi-topic structure of LSMS surveys can help establish the determinants of migration, inform the creation of future migration policy, and elucidate the impacts of migration on source households. In order to properly study migration with an LSMS survey context, researchers must first address a number of methodological considerations, particularly in relation to sampling, and issues relating to definitions and questionnaire design.

As migration is a rare event, the standard LSMS multi-stage cluster design is unlikely to sample a sufficiently large number of households with migrants to make significant statistical inferences. To account for the small number of households with migration experience in the general population, two alternative sample designs are therefore suggested: 1) disproportionate sampling of high migration PSUs and 2) stratified random sampling within PSUs, or two-phase sampling. Furthermore, using newly constructed listing of the selected PSUs allowing identification of migrant households, and the use of booster samples, to be carried out either in parallel or sequentially to the main survey, is advocated.

Researchers also face a number of tradeoffs in the designing of survey modules to study migration. These tradeoffs typically pertain to the number of questions related to migration for the different types of migrants, the period of recall, the expected accuracy of answers and the overall length of the survey. We have attempted to highlight some of these trade-offs and to help guide the analyst through the choices that must be made in designing effective migration modules. We hope that the data resulting from the thoughtful incorporation of migration components into LSMS-style surveys will enrich future debate on the complex relationship between migration, welfare and policy.



**Figure 1. Decision Tree for Collecting Information Through Proxy**

## References

- Acosta, Pablo (2006). "Labor Supply, School Attendance, and Remittances from International Migration: The Case of El Salvador," World Bank Policy Research Working Paper 3903.
- Adams, Richard Jr. (2003). "International Migration, Remittances, and the Brain Drain: A Study of 24 Labor-Exporting Countries," World Bank Policy Research Working Paper 3069.
- Akresh, Richard (2009). "Flexibility of Household Structure: Child Fostering Decisions in Burkina Faso," *Journal of Human Resources* 44(4): 976-997.
- Alderman, Harold (2000). "Anthropometry," in M. Grosh and P. Glewwe, eds., *Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 years of the Living Standards Measurement Study Volume 1*. World Bank, Washington, DC, pp. 251-272.
- Angelucci, Manuela (2004). "Aid and Migration: An Analysis of the Impact of Progresa on the Timing and Size of Labour Migration," IZA Discussion Paper no. 1187.
- Angrist, Josh, and Alan Krueger (1991). "Does Compulsory School Attendance Affect Schooling and Earnings?" *Quarterly Journal of Economics* 106(4): 979-1014.
- Antman, Francisca (2007). "The Intergenerational Effects of Paternal Migration on Schooling and Work: What can we Learn from Children's Time Allocations?" working paper, University of Colorado and Boulder.
- Azam, Jean-Paul, and Flore Gubert (2006). "Migrants' Remittances and the Household in Africa: A Review of the Evidence," *Journal of African Economies* 15(2): 426-462.
- Barham, Bradford, and Steve Boucher (1998). "Migration, Remittances, and Inequality: Estimating the Net Effects of Migration on Income Distribution." *Journal of Development Economics* 55: 307-31.
- Battacharya, Jay, and William B. Vogt (2007). "Do Instrumental Variables Belong in Propensity Scores?" NBER Technical Working Paper 343.
- Bhushan, Indu, and Raylynn Oliver (2000). "Fertility," in M. Grosh and P. Glewwe, eds., *Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 years of the Living Standards Measurement Study Volume 2*. World Bank, Washington, DC, pp. 31-47.
- Bilsborrow, Richard, A.S. Oberai and Guy Standing (1984). *Migration Surveys in Low-Income Countries: Guidelines for Survey and Questionnaire Design*. London: Croom Helm.
- Bilsborrow, Richard, Graeme Hugo, A.S. Oberai and Hania Zlotnick (1997). "International Migration Statistics: Guidelined for Improving Data Collection Systems", ILO, Geneva.
- Binswanger, Hans, and Mark Rosenzweig (1986). "The Behavioral and Material Determinants of Production Relations in Agriculture," *Journal of Development Studies* 22(3).

- Borjas, George (1987). "Self-Selection and the Earnings of Immigrants," *American Economic Review* 77(4): 531-553.
- Boucher, Steve, Oded Stark, and J. Edward Taylor (2005). "A Gain with a Drain? Evidence from Rural Mexico on the New Economics of the Brain Drain," ZEF – Discussion Papers on Development Policy no. 99.
- Bound, John, D.A. Jaeger and Michael Baker (1995). "Problems with Instrumental Variables Estimation when the Correlation between the Instruments and the Endogenous Explanatory Variable is Weak," *Journal of the American Statistical Association* 90: 443-450.
- Carletto, Gero, and Carlo Azzarri (2007). "Migration in Albania: Trends, Determinants and Impact on Poverty", background paper to the 2007 Albania Poverty Assessment, mimeo.
- Carrington, William, Enrica Detragiache, and Tara Vishnawath (1996). "Migration with Endogenous Moving Costs," *American Economic Review* 86(4): 909-30.
- Chiquiar, Daniel, and Gordon Hanson (2005). "International Migration, Self-Selection, and the Distribution of Wages: Evidence from Mexico and the United States," *Journal of Political Economy* 113(2): 239-281.
- Chiswick, Barry, and Timothy Hatton (2003). "International Migration and the Integration of Labor Markets," in Michael Bordo, Alan Taylor and Jeffrey Williamson, eds., *Globalization in Historical Perspective*, Cambridge: National Bureau of Economic Research: 65-119.
- Clemens, Michael, and Gunilla Petersson (2008). "New Data on African Health Professionals Abroad," *Human Resources for Health* 6(1).
- Clemens, Michael, and Lant Pritchett (2008). "Income per Natural: Measuring Development for People Rather than Places," *Population and Development Review* 34(3): 395-434.
- Cox, Donald and Emmanuel Jimenez. (1993) "Private Transfers and the Effectiveness of Public Income Redistribution in the Philippines," in Dominique van de Walle and K. Nead (eds.) *Public Spending and the Poor: Theory and Evidence*. Baltimore and London: Johns Hopkins University Press.
- Cox-Edwards, Alejandra, and Manuelita Ureta, 2003, "International migration, remittances, and schooling: evidence from El Salvador," *Journal of Development Economics*, 72(2) 429-461.
- Davis, Benjamin, and Paul Winters (2001). "Gender, Networks and Mexico-US Migration," *Journal of Development Studies* 38 (2): 1-26.
- de Brauw, Alan (2010). "Seasonal Migration and Agriculture in Vietnam," *Journal of Development Studies* 46(1): 114-139.
- de Brauw, Alan, and John Giles (2006). "Migrant Opportunity and the Educational Attainment of Youth in Rural China," IZA Discussion Paper no. 2326.

- de Brauw, Alan, and John Giles (2008). "Migrant Labor Markets and the Welfare of Households in the Developing World: Evidence from China," World Bank Policy Research Working Paper no. 4585.
- de Brauw, Alan, and Tomoko Harigaya (2007). "Seasonal Migration and Living Standards in Vietnam," *American Journal of Agricultural Economics* 89(2): 430-447.
- de Brauw, Alan, Qiang Li, Chengfang Liu, Scott Rozelle, and Linxiu Zhang (2008). "Is Feminization of Agriculture Occurring in China? Myths Surrounding Women's Participation in Farming," *China Quarterly* 194: 327-348.
- de Brauw, Alan, Scott Rozelle, Linxiu Zhang, Jikun Huang, and Yigang Zhang (2002). "The Evolution of China's Labor Market During the Reforms," *Journal of Comparative Economics* 29, 329-353.
- de Haan, Arjan (1999). "Livelihoods and Poverty: The Role of Migration—A Critical Review of the Migration Literature," *Journal of Development Studies* 36(2), 1-47.
- Deaton, Angus, and Margaret Grosh (2000). "Consumption," In M. Grosh and P. Glewwe, eds., *Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 years of the Living Standards Measurement Study Volume 1*. World Bank: Washington, DC, pp. 91-134.
- Deaton, Angus, and Salman Zaidi (2002). "Guidelines for Constructing Consumption Aggregates for Welfare Analysis," World Bank.
- Docquier, Frederic, and Abdeslan Marfouk (2005). "International migration by Educational Attainment, 1990-2000," in M. Schiff and C. Ozden, eds., *International Migration, Remittances, and the Brain Drain*. Washington, DC: World Bank.
- Dolfin, Sarah and Garance Genicot (2010). "What do Networks do? The Role of Networks on Migration and Coyote Use," *Review of Development Economics* 14(2): 343-359.
- Dreze, Jean, Meera Samson, and Satyajit Singh (1997). *The Dam and the Nation: Displacement and Resettlement in the Narmada Valley*. Delhi: Oxford University Press.
- Duflo, Esther, and Rohini Pande (2007). "Dams," *Quarterly Journal of Economics* 122(2): 601-646.
- Dustmann, Christian. (1997). "Return Migration, Uncertainty, and Precautionary Savings," *Journal of Development Economics* 52: 295-316.
- Dustmann, Christian. (2003). "Return Migration, Wage Differentials, and the Optimal Migration Duration," *European Economic Review* 47: 353-369.
- Dustmann, Christian, & Kirchkamp, Oliver (2002). The optimal migration duration and economic activities after return migration. *Journal of Development Economics*, 67: 351–372.
- Foster, Andrew, and Mark Rosenzweig (2008). "Economic Development and the Decline of Agricultural Employment," in T.P. Schultz and J. Strauss, eds., *Handbook of Development Economics* volume 4, Chapter 47, pp. 3051-3083.

- Funkhouser, Edward. (1997). "Labor Market Adjustment to Political Conflict: Changes in the Labor Market in El Salvador during the 1980s," *Journal of Development Economics* 52: 31-64.
- Gang, Ira, Catherine Co, and Myeong-So Yun (2000). "Returns to Returning," *Journal of Population Economics* 13: 57-80.
- Gibson, John (2002). "Why does the Engel method work? Food demand, economies of size and household survey methods." *Oxford Bulletin of Economics and Statistics* 64: 341-360.
- Gibson, John, and David McKenzie (2007a). "Using the Global Positioning System in Household Surveys for Better Economics and Better Policy," *World Bank Research Observer* 22(2): 217-241.
- Gibson, John, and David McKenzie (2007b). "The Impact of an Ex-Ante Job Offer Requirement on Labor Migration: The New Zealand-Tonga Experience," in C. Ozden and M. Schiff, eds., *International Migration, Economic Development, and Policy*. World Bank: Washington, DC.
- Giles, John (2006). "Is Life More Risky in the Open? Household Risk Coping and the Opening of China's Labor Markets," *Journal of Development Economics* 81(1): 25-60.
- Giles, John, and Kyeongwon Yoo (2007). "Precautionary Behavior, Migrant Networks, and Household Consumption Decisions: An Empirical Analysis Using Household Panel Data from Rural China," *Review of Economics and Statistics* 89(3): 534-551.
- Giles, John, and Ren Mu (2007). "Elder Parent Health and the Migration Decision of Adult Children: Evidence from Rural China," *Demography* 44(2): 265-288.
- Glewwe, Paul (2000a). "Household Roster." In M. Grosh and P. Glewwe, eds., *Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 years of the Living Standards Measurement Study Volume 1*. World Bank: Washington, DC, pp. 135-142.
- Glewwe, Paul (2000b). "Human Capital." In M. Grosh and P. Glewwe, eds., *Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 years of the Living Standards Measurement Study Volume 1*. World Bank: Washington, DC, pp. 143-176.
- Global Forum on Migration and Development (2007), "Final Conclusions and Recommendations of the Chair, Ambassador Regine de Clercq," available at <http://www.gfmd-fmmd.org/en/conclusions-and-recommendations>
- Gubert, Flore (2002). "Do Migrants Insure Those Left Behind? Evidence from the Kayes Area (Western Mali)." *Oxford Development Studies* 30(3): 267-287.
- Halliday, Timothy (2006). "Migration, Risk, and Liquidity Constraints in El Salvador," *Economic Development and Cultural Change* 54(4): 893-925.
- Ham, John, Xianghong Li, and Patricia Reagan (2006). "Propensity Score Matching, a Distance Based Measure of Migration, and the Wages of Young Men," working paper, University of Southern California.

- Hanson, Gordon, and Antonio Spilimbergo (1999). "Illegal Immigration, Border Enforcement, and Relative Wages: Evidence from Apprehensions at the U.S.-Mexico Border," *American Economic Review* 89(5): 1337-1351.
- Hanson, Gordon (2005). "Emigration, Labor Supply, and Earnings in Mexico," in George Borjas, ed., *Mexican Immigration*. Chicago: University of Chicago Press and the National Bureau of Economic Research.
- Hanson, Gordon (2007). "The Economic Logic of Illegal Migration," Council of Foreign Relations Council Special Report no. 26.
- Harvey, Andrew, and Maria Elena Taylor (2000). "Time Use." In M. Grosh and P. Glewwe, eds., *Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 years of the Living Standards Measurement Study Volume 2*. World Bank: Washington, DC, pp. 249-274.
- Heckathorn, Douglas D. (1997). "Respondent Driven Sampling: A New Approach to the Study of Hidden Populations,," *Social Problems* 44(2): 174-199.
- Heckathorn, Douglas D. (2002). "Respondent Driven Sampling II: Deriving Valid Population Estimates from Chain-Referral Samples of Hidden Populations," *Social Problems* 49(1): 11-34.
- Heckmann, Frederich. (2004). "Illegal Migration: What Can We Know and What Can We Explain? The Case of Germany," *International Migration Review* 38(3): 1103-1125.
- Hildebrandt, Nicole, and David McKenzie (2005). "The Effects of Migration on Child Health in Mexico." *Economia* 6: 257-281.
- Holzmann, Robert, Johannes Koettl, and Taras Chernetsky (2005). "Portability Regimes of Pension and Health Care Benefits for International Migrants: An Analysis of Issues and Good Practices," World Bank Social Protection Discussion Paper Series no. 0519.
- Ibarraran, Pablo, and Darren Lubotsky (2007). "Mexican Immigration and Self-Selection: New Evidence from the 2000 Mexican Census," in *Mexican Immigration to the United States*, George Borjas, ed. Cambridge: National Bureau of Economic Research.
- International Fund for Agricultural Development (2007). "Sending Money Home: Worldwide Remittance Flows to Developing Countries," mimeo.
- INSTAT (2002). Albania Living Standards Measurement Survey, <http://www.worldbank.org/lsm>
- INSTAT (2005). Albania Living Standards Measurement Survey, <http://www.worldbank.org/lsm>
- Kilic, Talip, Gero Carletto, Benjamin Davis, and Alberto Zezza (2009). "Investing Back Home," *Economics of Transition* 17(3): 587-623.
- Kish, Leslie. (1965) *Survey Sampling*. New York: John Wiley.

- Korinek, Anton, Johan Mistiaen, and Martin Ravallion (2007). "An Econometric Method of Correcting for Unit Non-Response in Surveys," *Journal of Econometrics* 136: 213-235.
- Lalonde, Robert, and Robert Topel (1993). "Economic Impact of International Migration and the Economic Performance of Migrants," Chapter 14 in M.R. Rosenzweig and O. Stark, eds., *Handbook of Population and Family Economics* vol. 1, pp. 799-850.
- Lipton, Michael (1980). "Migration from Rural Areas of Poor Countries: The Impact on Rural Productivity and Income Distribution," *World Development* 8: 1-24.
- Lucas, Robert E.B. (1993). "Internal Migration in Developing Countries," Chapter 13 in M.R. Rosenzweig and O. Stark, eds., *Handbook of Population and Family Economics* vol 1., pp. 721-798.
- Lucas, Robert E.B. (2000). "Migration." In M. Grosh and P. Glewwe, eds., *Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 years of the Living Standards Measurement Study Volume 2*. World Bank: Washington, DC, pp. 49-82.
- Lucas, Robert E.B. (2005). *International Migration and Economic Development: Lessons from Low Income Countries*. Boston: Edward Elgar Publishing.
- Macours, Karen, and Renos Vakis (2010). "Seasonal Migration and Early Childhood Development," *World Development* 38(6): 857-869.
- Mallee, Hein (1995). "China's Household Registration System Under Reform," *Development and Change* 26(1): 1-29.
- Malpezzi, Stephen (2000). "Housing." In M. Grosh and P. Glewwe, eds., *Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 years of the Living Standards Measurement Study Volume 1*. World Bank: Washington, DC, pp. 293-314.
- Mansuri, Ghazala. (2006). "Migration, School Attainment, and Child Labor: Evidence from Rural Pakistan," World Bank Policy Research Working Paper Series 3945.
- Mansuri, Ghazala. (2006b). "Migration, Sex Bias, and Child Growth in Rural Pakistan," World Bank Policy Research Working Paper Series 3946.
- Martin, Susan (2004). "Women and Migration," paper prepared for United Nations Division for the Advancement of Women Consultative Meeting on "Migration and Mobility and how this Movement Affects Women."
- Massey, Douglas (1999). "International Migration at the Dawn of the 21<sup>st</sup> Century: The Role of the State," *Population and Development Review* 25(2): 303-322.
- Massey, Douglas, Arango, Graeme Hugo, Ali Kouaouci, Adela Pellegrino and J. Edward Taylor (1998). *Worlds in Motion: Understanding International Migration at the Turn of the Century*, Oxford: Oxford University Press.
- Massey, Douglas, and Chiara Capoferro (2004). "Measuring Undocumented Migration," *International Migration Review* 38(3): 1075-1102.

- Massey, Douglas, and K. Espinosa (1997). "What's Driving U.S.-Mexico Migration? A Theoretical, Empirical, and Policy Analysis," *American Journal of Sociology* 102: 939-999.
- Mattoo, Aaditya, Ileana Cristina Neagu, and Caglar Ozden (2008). "Brain waste? Educated Immigrants in the US Labor Market," *Journal of Development Economics*, 87(2): 255-269.
- McCarthy, Nancy, Gero Carletto, Benjamin Davis, and Irini Maltsoğlu (2006). "Assessing the Massive Impact of Out-Migration on Agriculture," FAO-ESA Working Paper no. 06-14.
- McCormick, Barry, and Jackline Wahba (2003). "Return International Migration and Geographical Inequality: The Case of Egypt," *Journal of African Economies* 12(4): 500-532.
- McKenzie, David (2007). "Paper Walls are Easier to Tear Down: Passport Costs and Legal Barriers to Emigration", *World Development* 35(11): 2026-2039.
- McKenzie, David, John Gibson, and Steven Stillman (2006). "How Important is Selection? Experimental versus Non-Experimental Measures of the Income Gains from Migration," World Bank Policy Research Working Paper Series no. 3906.
- McKenzie, David, and Johan Mistiaen (2009). "Surveying Migrant Households: A Comparison of Census-Based, Snowball, and Intercept Point Surveys," [\*Journal Of The Royal Statistical Society Series A\* 172\(2\): 339-360.](#)
- McKenzie, David, and Hillel Rapoport (2007). "Network effects and the dynamics of migration and inequality: theory and evidence from Mexico." *Journal of Development Economics* 84(1): 1-24.
- Mesnard, Alice (2004). "Temporary Migration and Capital Market Imperfections," *Oxford Economic Papers* 56: 242-262.
- Mesnard, Alice, and Martin Ravallion (2006). "The Wealth Effect on New Business Startups in a Developing Country," *Economica* 73(291): 367-392.
- Miluka, Juna, Gero Carletto, Benjamin Davis, and Alberto Zezza (2010). "The Vanishing Farms? The Impact of International Migration on Family Farming," *Journal of Development Studies* 46(1): 140-161.
- Mishra, Prachi (2007). "Emigration and Wages in Source Countries: Evidence from Mexico," *Journal of Development Economics* 82(1): 180-199.
- Mountford, Andrew (1997). "Can a Brain Drain be Good for Growth in the Source Economy?" *Journal of Development Economics* 53(2): 287-303.
- Muñoz, Juan (2007). "Sampling: What You Don't Know Can Hurt You," presentation at the Multi-topic Household Survey course, January 2007, World Bank.
- Munshi, Kaivan (2003). "Networks in the Modern Economy: Mexican Migrants in the U. S. Labor Market." *Quarterly Journal of Economics* 118:549-597.
- Murray, Michael (2005). "The Bad, the Weak, and the Ugly: Avoiding the Pitfalls of Instrumental Variables Estimation," working paper, Bates College.

- Osili, Una Okonkwo. (2007). "Remittances and savings from international migration: Theory and evidence using a matched sample," *Journal of Development Economics* 83(2):446-465.
- Paulson, Anna (2003). "Insurance Motives for Migration: Evidence from Thailand," working paper, Kellogg School of Management, Northwestern University.
- Pfieffer, Lisa, Susan Richter, Ed Taylor, and Peri Fletcher (2007). "Gender and International Migration: Literature Review," in Andrew Morrison, Maurice Schiff, and Mirja Sjöblom, eds., *The International Migration of Women*. New York: Palgrave MacMillan and the World Bank.
- Quinn, Michael A. (2006). "Relative Deprivation, Wage Differentials, and Mexican Migration," *Review of Development Economics* 10(1): 135-153.
- Ratha, Dilip, and William Shaw (2007). "South-South Migration and Remittances," World Bank working paper no. 102.
- Reardon, Thomas, and Paul Glewwe (2000). "Agriculture." In M. Grosh and P. Glewwe, eds., *Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 years of the Living Standards Measurement Study Volume 2*. World Bank: Washington, DC, pp. 139-182.
- Richter, Susan, and J. Edward Taylor. (2005). "Policy Reforms and the Gender Dynamics of Rural Mexico-to-U.S. Migration," University of California at Davis Department of Agricultural and Resource Economics working paper 05-007.
- Rosenzweig, Mark, and Oded Stark (1989). "Consumption Smoothing, Migration, and Marriage: Evidence from Rural India," *Journal of Political Economy* 97(4): 905-926.
- Rosenzweig, Mark, and Ken Wolpin (2000). "Natural 'Natural Experiments' in Economics," *Journal of Economic Literature* 38(4): 827-874.
- Rozelle, Scott, Li Guo, Minggao Shen, Amelia Hughart, and John Giles (1999). "Leaving China's Farms: Survey Results of New Paths and remaining Hurdles to Rural Migration," *China Quarterly* 158: 367-393.
- Sasin, Marcin and David McKenzie (2007). Migration, Remittances, Poverty, and Human Capital. World Bank Policy Research Working Paper 4272.
- Schaffner, Julie Anderson (2000). "Employment," in M. Grosh and P. Glewwe, eds., *Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 years of the Living Standards Measurement Study Volume 1*. World Bank, Washington, DC, pp. 217-250.
- Schiff, Maurice (2005). "Brain Gain: Claims about Its Size and Impact on Welfare and Growth are Greatly Exaggerated," IZA Discussion Paper No. 1599.
- Schiff, Maurice, Andrew Morrison, and Mirja Sjöblom (2007). *Women in International Migration*. Washington, DC: World Bank.

- Smith, James, and Duncan Thomas (2003). "Remembrances of Things Past: Test-Retest Reliability of Retrospective Migration Histories," *Journal of the Royal Statistical Society, Series A (Statistics in Society)* 166(1): 23-49.
- Som, R.K., (1973). *Recall lapse in demographic inquiries*. New York: Asia Publishing House.
- Staiger, Doug, and James Stock (1997). "Instrumental Variables Regressions with Weak Instruments," *Econometrica* 65(3): 557-586.
- Stark, Oded. (1984). "Rural-to-Urban Migration in LDCs: A Relative Deprivation Approach," *Economic Development and Cultural Change* 32(3): 475-486.
- Stark, Oded, and C. Simon Fan. (2007). "The Analytics of Seasonal Migration," *Economics Letters* 94(2): 304-312.
- Stark, Oded, and J. Edward Taylor. (1989). "Relative Deprivation and International Migration," *Demography* 26(1): 1-14.
- Stark, Oded, and J. Edward Taylor. (1991). "Migration Incentives, Migration Types: The Role of Relative Deprivation," *Economic Journal* 101: 1163-1178.
- Stark, Oded, J. Edward Taylor, and Shlomo Yitzhaki. (1986). "Remittances and Inequality," *Economic Journal* 101: 1163-1178.
- Stark, Oded, and Yong Wang (2002). "Inducing Human Capital Formation: Migration as a Substitute for Subsidies," *Journal of Public Economics* 86: 29-46.
- Stecklov, Guy, Paul Winters, Marco Stampini, and Benjamin Davis. (2005). Can Public Transfers Reduce Mexican Migration? A Study based on Randomized Experimental Data, *Demography*, 42(4): 769-790.
- Stecklov, Guy, Gero Carletto, Carlo Azzari, and Benjamin Davis (2007). "Agency, Education, and Networks: Gender and International Migration from Albania," mimeo, World Bank.
- Stillman, Steven, John Gibson, and David McKenzie (2009). "Migration and Mental Health: Evidence from a Natural Experiment." *Journal of Health Economics* 28(3): 677-687.
- Stock, James, and Motohiro Yogo (2005). "Testing for Weak Instruments in IV Regression," in *Identification and Inference for Econometric Models: A Festschrift in Honor of Thomas Rothenberg*, Cambridge: Cambridge University Press.
- Taylor, J. Edward and Philip L. Martin (2001). "Human capital: Migration and rural population change," in: G. Rausser and B. Gardner, eds., *Handbook of Agricultural Economics* (Elsevier Science Publishers, New York), 458-511.
- Thailand (2007). *The Labor Force Survey: Whole Kingdom, Quarter 4: October - December 2006. Report in Thai and English*. Bangkok: NSO, Ministry of Information and Communication Technology.

- United Nations (1998). Recommendation on Statistics of International Migration: Revision 1, Statistical Papers Series M, no. 58, rev. 1, Department of Economic and Social Affairs, Statistics Division, New York, NY
- United Nations (2006). "International Migration." Population Division, New York.
- Vijverberg, Wim, and Donald Mead (2000). "Household Enterprises," In M. Grosh and P. Glewwe, eds., *Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 years of the Living Standards Measurement Study Volume 2*. World Bank: Washington, DC, pp. 105-138.
- Williamson, Jeffrey G. (1988). "Migration and urbanization," in: Hollis Chenery and T.N. Srinivisan, eds., *Handbook of Development Economics* (Elsevier Science Publishers, New York), pp. 426-446.
- Winters, Paul, Alain de Janvry, and Elizabeth Sadoulet (2001). "Family and Community Networks in Mexico-U.S. Migration," *Journal of Human Resources* 36(1): 159-184.
- Woodruff, Chris, and Rene Zenteno (2007). "Migration networks and microenterprises in Mexico," *Journal of Development Economics* 82: 509–528.
- Wooldridge, Jeffrey (2002). *Econometrics of Cross-Section and Panel Data*. Cambridge, MA: MIT Press.
- World Bank (2000). "Vietnam Living Standards Survey (VLSS) 1997–8 Basic Information." Mimeo, Poverty and Human Resources Division.
- Yang, Dean (2008). "International Migration, Remittances, and Household Investment: Evidence from Philippine Migrants' Exchange Rate Shocks," *Economic Journal* 118(528): 1-40.
- Yang, Dean, and Claudia Martinez (2006). "Remittances and Poverty in Migrants' Home Areas: Evidence from the Philippines," in Caglar Ozden and Maurice Schiff, eds., *International Migration, Remittances, and the Brain Drain*, Washington: World Bank.
- Yang, Dean, and HwaJung Choi (2007). "Are Remittances Insurance? Evidence from Rainfall Shocks in the Philippines," *World Bank Economic Review* 21(2): 219-248.
- Zlotnick, Hania (1987a). "Introduction: Measuring International Migration: Theory and Practice," *International Migration Review*, Vol. 21, No. 4: v-vii.
- Zlotnick, Hania. (1987b). "The Concept of International Migration as Defined in Data Collection Systems," *International Migration Review* 21(4): 925-946.