The Costs and Benefits of Longitudinal Data: Three Applications from the Mexican Family Life Survey

by

Andrea P. Velásquez G.

Department of Economics
Duke University

Date:_____

Approved:

_____
Duncan Thomas, Supervisor

_____
Erica Field

_____
Amar Hamoudi

_____
Seth Sanders

Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Economics in the Graduate School of Duke University

2014

ABSTRACT

The Costs and Benefits of Longitudinal Data: Three Applications from the Mexican Family Life Survey

by

Andrea P. Velásquez G.

Department of Economics
Duke University

Date:_____

Approved:

_____
Duncan Thomas, Supervisor

_____
Erica Field

_____
Amar Hamoudi

_____
Seth Sanders

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Economics in the Graduate School of Duke University

2014

Copyright © 2014 by Andrea P. Velásquez G.
All rights reserved except the rights granted by the
Creative Commons Attribution-Noncommercial Licence

Abstract

Longitudinal surveys have revolutionized empirical research and our understanding of the dynamic processes that affect the economic prosperity, health and well-being of the population. This dissertation explores and provides evidence, through three empirical applications, on the costs and benefits of designing, implementing and using data from a new, innovative longitudinal survey, the Mexican Family Life Survey (MxFLS). The survey, which is representative of the Mexican population living in Mexico in 2002, is designed to follow movers within Mexico and also those who move to the United States. This design lies at the center of the contributions of my research to the scientific literature.

Attrition is the Achilles heel of longitudinal surveys. The first essay of the dissertation focuses on the cost of attrition for scientific knowledge. Following the same individual through time allows a researcher to trace the evolution of a respondent's behaviors and outcomes in a dynamic framework; however, if attrition is selected on unobserved characteristics, the advantage of using panel data could be severely hindered. Exploring different methods to adjust for attrition, this essay provides evidence of limitations of standard post-survey adjustments strategies that are the standard in the literature. These approaches, exploit only baseline characteristics of the respondents and, conditional on those characteristics, treat attriters as missing at random. I provide evidence that this assumption is substantively important and rejected

in the MxFLS in spite of the fact that attrition in that survey is low relative to other nationally-representative surveys conducted in the United States and abroad.

The second essay in this dissertation exploits the fact that MxFLS follows movers within Mexico and those who move across the Mexico-US border to provide new insights into the mechanisms that underlie the selectivity of migrants within Mexico, how they differ from migrants who move from Mexico to the U.S. and how those who return contrast with the migrants who remain in the U.S. more permanently. The results provide evidence that human capital is predictive of migration within Mexico and to the United States, but that there is little indication that the decision to stay in the United States is highly correlated with education. In contrast, having relatives in the United States is not only a powerful predictor of migration to the United States, but it is also predictive of successful economic assimilation.

The third essay exploits a different dimension of the longitudinal survey in order to address an important question regarding the impact of unanticipated crime and violence on population well-being. To wit, the essay rigorously examines the impact of the recent surge in violent crime in Mexico on the labor market outcomes, migration, and wealth of the Mexican population. The timing of the last two waves of the MxFLS paired with the panel nature of the survey, allows the comparison of outcomes of the same individual in periods of low and high violence, which removes the potentially endogenous time-invariant unobserved heterogeneity between respondents. Moreover,

due to the fact that the MxFLS was designed to follow migrant respondents, this study is able to directly test whether there is a systematic migratory response to crime. The results from this analysis find that crime predicts migration and it negatively affects the labor outcomes of self-employed individuals. In addition, the negative effects on the labor outcomes have translated into reductions in per capita expenditure at the household level, which suggests that the recent wave of violence in Mexico may have long-term consequences on the wealth and well-being of Mexican households.

Dedication

To my parents, Martha P. and Pedro A., my greatest inspiration.

Contents

Abstract iv

List of Tables xi

List of Figures xiii

Acknowledgements xiv

1. Introduction 1

2. Attrition in Longitudinal Surveys: Evidence from the Mexican Family Life Survey 5

2.1 Introduction 5

2.2 Conceptual Framework 9

2.3 Data: The Mexican Family Life Survey 14

2.4 Analysis of Attrition 16

2.4.1 Descriptive Statistics 17

2.4.2 Prediction of Attrition 23

2.4.3 Attrition in a Model of Earnings 29

2.4.4. Methods to Handle Missing Data due to Attrition 35

2.5 Multiple Imputation 36

2.5.1 Imputation of Earnings at Baseline 37

2.5.2 Imputation of Earnings in MxFLS2 42

2.5.3 Imputation of Earnings in MxFLS3 43

2.6 Conclusions 46

2.7 Tables and Figures 48

3. Selection and Assimilation of Mexican Migrants to the U.S. 66

3.1 Introduction 66
3.2 Data 70
3.2.1 Migration and Characteristics at Baseline 73
3.2.2 Assimilation Outcomes 76
3.3 Migration to the U.S.: Who Migrates, Who Stays, and Who Returns to Mexico 77
3.4 Assimilation 84
3.5 Discussion and Future Work 92
3.6 Tables and Figures 94
4. The Economic Burden of Crime: Evidence from Mexico 110
4.1 Introduction 110
4.2 Background 117
4.3 The Economic Effects of Crime and Violence 123
4.4 Data: Mexican Family Life Survey 128
4.4.1 Labor Force Participation 131
4.5 Conceptual Framework 133
4.6 Empirical Strategy 137
4.7 Results 144
4.7.1 Labor Market Outcomes 144
4.7.1.1 Females 144
4.7.1.2 Males 150
4.7.2 Per Capita Expenditure (PCE) 154
4.8 Robustness Checks 158

4.8.1 Attrition 158
4.8.2 Prediction of Homicide Rates 159
4.8.3 Placebo Test 160
4.9 Conclusions 162
4.10 Tables and Figures 164
4.11 Supplementary Tables 181
References 184
Biography 191

List of Tables

Table 1: Attrition Rates 48

Table 2: Characteristics Lost vs. Found Age 15+ 49

Table 3: Characteristics Lost vs. Found - By Gender Age 15+ 51

Table 4: Probability of Attrition - Basic Logit Model Age15+ 53

Table 5: Probability of Attrition - Extended Logit Model Age15+ 55

Table 6: Quartic-root of Earnings in 2002 with Interactions - Age [21-65] 57

Table 7: Original and Imputed Models - Quartic-root of Earnings in 2002- Age [21-65] . 58

Table 8: Original and Imputed Models with Interactions - Quartic-root of Earnings in 2002- Age [21-65] 59

Table 9: Original and Imputed Models - Quartic-root of Earnings in 2005- Age [21-65] . 61

Table 10: Original and Imputed Models - Quartic-root of Earnings in 2009- Age 15+ 62

Table 11: Sample Sizes and Recontact Rates in MxFLS 94

Table 12: Migration Between Baseline and MxFLS3 95

Table 13: U.S. Networks Reported at Baseline by MX/U.S. Migration Status 95

Table 14: U.S. Networks Reported at Baseline by U.S. Migration Status. 96

Table 15: Characteristics U.S. Migrants - Panel Members Age15+ 97

Table 16: Baseline Characteristics that Predict Migration Since 2002 - Males 98

Table 17: Baseline Characteristics that Predict Migration Since 2002 - Females 100

Table 18: Baseline Characteristics Predictors of Assimilation in U.S. - MxFLS2 102

Table 19: Assimilation of Current U.S. Migrants - Economic Variables 104

Table 20: Assimilation of Current U.S. Migrants - Use of English 106

Table 21: Assimilation of Current U.S. Migrants - Relatives in U.S. 108

Table 22: Sample Sizes and Recontact Rates in MxFLS 168
Table 23: Employment Transitions 169
Table 24: Prediction of Migration 170
Table 25: Labor Outcomes of Women Age [18-75] 171
Table 26: Labor Outcomes of Self-Employed Women Age [18-75] 172
Table 27: Self-Employed Women by Occupation Age [18-75] 173
Table 28: Labor Outcomes of Men Age [18-75] 174
Table 29: Labor Outcomes of Self-Employed Men Age [18-75] 175
Table 30: Self-Employed Men by Occupation - Age [18-75] 176
Table 31: Log(Per Capita Expenditure) 177
Table 32: Placebo Test - Labor Outcomes of Women Age [18-75] 178
Table 33: Placebo Test - Labor Outcomes of Men Age [18-75] 179
Table 34: Placebo Test - Log(Per Capita Expenditure) 180
Table 35: Prediction of Attrition from MxFLS3 - Men Age 18+ 181
Table 36: Prediction of Attrition from MxFLS3 - Women Age 18+ 182
Table 37: Prediction of Changes on Homicide Rate between 2005 and 2009 183

List of Figures

Figure 1: Distribution of the Quartic-root of Earnings in 2002 63

Figure 2: Distribution of the Quartic-root of Earnings in 2009 63

Figure 3: Distribution of the Quartic-root of Earnings in 2002 – Males 64

Figure 4: Distribution of the Quartic-root of Earnings in 2009 – Males 64

Figure 5: Distribution of the Quartic-root of Earnings in 2002 – Females 65

Figure 6: Distribution of the Quartic-root of Earnings in 2009 – Females 65

Figure 7: INEGI and SNSP - Monthly Homicide Rate (per 100,000) 164

Figure 8: Timing of MxFLS and INEGI - Monthly Homicide Rate (per 1000,000) 165

Figure 9: INEGI Annual Homicide Rate - 2002 165

Figure 10: INEGI Annual Homicide Rate - 2005 166

Figure 11: INEGI Annual Homicide Rate - 2007 166

Figure 12: INEGI Annual Homicide Rate - 2009 167

Figure 13: INEGI Annual Homicide Rate - 2010 167

Acknowledgements

I am very grateful to my advisor Duncan Thomas for his exceptional guidance and support in every step of this journey. His constant encouragement and advice, as well as his passion for the science inspired and made this work possible. He has been an extraordinary example of a scholar and advisor, and I will be forever grateful for his mentorship.

I would also like to thank the other members of my committee Erica Field, Amar Hamoudi, Seth Sanders, and Alessandro Tarozzi for their invaluable feedback and suggestions through different stages of the development of this dissertation. I have benefitted from an incredibly rich academic environment at Duke University and having the opportunity to present my work at the Duke Economics Labor and Development seminar series, Duke University Population Research Institute seminar series, and the International Population Health and Development Seminar, enriched my work.

This dissertation would not have been possible without having had the opportunity to be part of the MxFLS. I am deeply thankful to Duncan Thomas, Graciela Teruel and Luis Rubalcava for giving me the opportunity to be part of such an incredible team. The journey would not have been the same without Maria Genoni and Gabriela Farfan. I could not have had better colleagues and friends to learn from throughout this project. Also my deepest admiration goes out to the MxFLS team of surveyors, their passion and hard work has been an inspiration.

My training at Duke benefitted immensely from the many academic interactions, friendships, and support from each member of the Thomas-Frankenberg lab. Ryan Brown, Gabriela Farfan, Nick Ingwersen, Daniel LaFave, Maria Genoni, Veronica Montalva, Evan Peet, Peter Katz, and Michael Burrows, thanks for all the insightful comments and for your friendship.

To all my friends from Pa' las que sea, thanks for your unconditional friendship and for helping to make Durham a home. To my Colombian friends who, even being many miles away, are always present in my life. I am also thankful to Domenico Ferraro for his friendship trough the tough times, and to all my wonderful friends at Duke that made these six years a great adventure.

I will also always be grateful to Ana Maria Ibáñez. Having the opportunity to work with her on projects related to forced displacement in Colombia were the inspiration for my pursuit of graduate studies.

I would not be here without the support of my entire family. To my parents who have always supported me and taught me, through their example, to be passionate about my dreams. To Lucho and Mimi, for being the example of what a family means. To Ita, my angel in heaven. To Nana, Koke, Nena, and Lea, for being the older siblings that I never had.

My infinite thanks are due to Ryan Brown. His unconditional love and support enlightened my life. I cannot be thankful enough for his friendship, his editorial eye, and

for the long hours spent discussing our research ideas. I feel very blessed to have you as my partner in life.

Overall, thanks to God, for giving me so many blessings.

1. Introduction

The Mexican Family Life Survey (MxFLS) is ideally suited to address an analysis of the costs and benefits of designing, implementing and using data from longitudinal surveys. The MxFLS is a longitudinal survey representative of the Mexican population living in Mexico in 2002, and while the level of respondent attrition is very low compared to similar surveys, there is not perfect sample retention in subsequent waves. One of the challenges in many longitudinal surveys is the successful tracking of migrants. As a consequence, in many longitudinal surveys conducted in developing countries, the vast majority of attrition is driven by loss to follow-up of migrants so that attrition is selected on characteristics associated with migration. This type of attrition is particularly problematic as it would lead to a sample that is endogenously selected on characteristics associated with migration behavior. To address this concern the MxFLS was specifically designed to follow migrants both within Mexico and into the U.S and to continue to track individuals even if they were not interviewed in a previous wave.

The first essay of this dissertation, *"Attrition in Longitudinal Surveys: Evidence from the Mexican Family Life Survey"*, joint work with Maria Genoni, Luis Rubalcava, Graciela Teruel and Duncan Thomas, is a methodological chapter focused on providing analysis of the costs of attrition in longitudinal surveys. When attrition is selected on characteristics that are not observed in the data, post-survey adjustments based on observed characteristics may not properly compensate for the lost respondents. Data

from the MxFLS is used to examine attrition between the 2002 baseline and the first follow up in 2005. In addition to uncovering the observed attributes that predict attrition, we also explore how different methods for making adjustments for attrition, including a multiple imputation approach, would affect the estimates of a standard model of earnings. Using this rich data set, this essay posits that attrition is selective and discusses the potential bias it may have on an uncorrected model of earnings. Moreover, exploiting information from individuals that had not been interviewed in the second wave of the MxFLS but were surveyed in the third wave, we test the effectiveness of a standard post-survey adjustment strategy in a model of earnings. The results suggest that assuming that attrition is ignorable based solely on baseline characteristics, and not accounting for changes in the lives of respondents that occurred after the baseline interview, may lead to biased estimates.

In addition to being ideally suited for exploring the costs of attrition in longitudinal surveys, the MxFLS's design and timing make it possible to explore many other topics of great relevance in developing countries including the labor market impact of crime and violence and the selection and assimilation process of migrants. Specifically, the two additional essays of my dissertation focus first, on Mexican migration to the United States, and second, on the economic consequences of the recent surge of Mexican drug war-related violence.

In the second essay, *"Selection and Assimilation of Mexican Migrants to the U.S"*, joint work with Gabriela Farfan, Maria Genoni, Luis Rubalcava, Graciela Teruel and Duncan Thomas, we exploit data from the U.S. component of the MxFLS to provide evidence on the selectivity and assimilation of recent migrants from Mexico to the United States. The empirical specification exploits information on Mexican migrants measured at baseline, prior to the migration decision, to assess the characteristics that predict selection into migration. Then, using information measured in the U.S., during the two follow-up surveys conducted in 2005 and 2009, we explore the determinants of successful assimilation, measured by labor market outcomes, per capita expenditure, use of English, and living arrangements while in the United States.

The third essay of this dissertation, *"The Economic Burden of Crime: Evidence from Mexico"*, is an empirical analysis of the effect the rapid, unanticipated, and unprecedented rise in violent conflict in Mexico since 2007 has had on individual labor market outcomes and household expenditures. Exploiting information from the MxFLS during periods of both low and high levels of violent crime, this study estimates an individual fixed effect model that controls for time-invariant unobserved heterogeneity that could potentially be correlated with both the level of exposure to violence and labor market outcomes. Moreover, by using the MxFLS this analysis is able to examine the impact of violence on migratory behavior and then utilize an "intent-to-treat" approach

in order to shield the estimates from the potential bias of systematic migratory response to crime.

The results of this study show that increasing violence in Mexico had a particularly strong and multidimensional impact on the labor market outcomes of self-employed women. I find that exposure to violent conflict leads self-employed women to leave the labor market and, amongst those that remain employed, reduce their labor intensity. Similarly, for males, the negative impact is also strongest for the self-employed. A major difference between the genders, though, is that for males, labor market participation is not significantly reduced and their earnings are not sensitive to lower levels of violence. Additionally, I provide evidence of a negative effect of violence on per capita expenditure at the household level, which, for female lead households, is related to a reduction in spending on education.

2. Attrition in Longitudinal Surveys: Evidence from the Mexican Family Life Survey

2.1 Introduction

Innovative longitudinal surveys have revolutionized empirical research in economics and been the foundation for many important contributions to the understanding of many processes that affect population health and well-being. Following the same individuals over time has enabled research to trace the evolution of behaviors and outcomes in a dynamic framework and to better understand the role that different sources of unobserved heterogeneity play in economic models of behavior.

However, the contributions of studies that use longitudinal data depend critically on the extent to which attrition is selected on characteristics that are both correlated with outcomes of interest and not observed. If attrition depends only on characteristics that are observed, then it is straightforward to re-weight the respondents who are followed and adjust estimates for attrition. In general, however, attrition is likely to also depend on characteristics that are not observed in which case there is a risk that re-weighted estimates will be biased. Determining the extent to which attrition contaminates inferences in specific models of behavior remains an unresolved issue in studies that use longitudinal survey data.

This chapter addresses this question using data from three waves of the Mexican Family Life Survey (MxFLS). Along with detailed information on socioeconomic characteristics that are collected in many broad-purpose socio-economic surveys, the

MxFLS collects extensive information not usually included in these surveys that is potentially related to attrition. These include, for example, expectations and intentions about the future, perceptions of current and future health and perceptions of quality of life. We explore the extent to which these data are predictive of attrition. Further, we explore how different methods for making adjustments for attrition affect estimates of the impact of standard covariates in a model of earnings.

While high rates of attrition may result in sample sizes that lack power to detect effects of interest, the rate of attrition is not enough to determine the potential bias due to attrition in behavioral models. If attrition is completely random, it is ignorable in any model (Rubin, 1987). In practice, such happenstance is unlikely and it is inherently difficult to test.

If attrition is selected on observed characteristics, estimates of interest may not be biased. If it is possible to construct post-survey weights to reconstruct a sample that mimics the baseline sample from the perspective of the behavioral model of interest, then attrition can be treated as ignorable, conditional on these characteristics. (See, for example, Fitzgerald et al., 1998; Alderman et al., 2001; Abraham et al., 2006.) Such a situation may arise in at least two situations. First, if attrition is selected only on characteristics that are observed at baseline then it is straightforward to reconstruct the original sample with weights. Second, if attrition is selected on characteristics that are not observed at baseline and also not correlated with any of the characteristics that are

included in a specific model of interest then, in the context of that model, attrition can be treated as missing at random and is thus ignorable.[1] In this situation, the conclusion that attrition is ignorable is model-dependent and is not a general result for a particular longitudinal survey. While it is straightforward to test the first case, it is difficult to construct empirical tests for the second case since it involves assumptions about the correlation between attrition and characteristics that are not observed.

In general, attrition is likely to be related to characteristics that are observed at baseline and characteristics that are not observed at that time and may not even be observable at baseline. For example, unanticipated changes in the life of a respondent that occurs after baseline cannot be measured at that time. If respondents attrite from the survey because of post-baseline changes in their lives that are not observed, then attrition will, by definition, depend on unobserved characteristics. This poses a challenge for tests of the ignorability of attrition that is predicated on information collected at baseline unless, of course, the changes that occur after baseline can be predicted by baseline characteristics. This suggests that attrition may be of greater concern in panels that span a longer time frame, among populations that are undergoing large changes in their lives (such as adolescents entering adulthood) and in study

[1] Without loss of generality, the second case can be extended to include attrition that is correlated with observed characteristics.

settings that involve unanticipated changes in the lives of respondents (such as large economic or natural shocks or experimental interventions).

The next section places this research in context and describes a general conceptual framework to structure the analysis. This is followed by a description of the Mexican Family Life Survey and by a detailed examination of the attrition rates in the second wave. We then present a model that predicts attrition based on a rich set of outcomes measured at baseline. In addition to a broad array of socio-economic characteristics, we include an additional set of variables, for example information about future expectations and migration that are not usually included in these models. The results show that these characteristics are predictive of subsequent attrition, which provides a first signal that attrition may be selected on characteristics that are potentially unobserved at baseline. These analysis is followed by an estimation of a standard model of earnings measured at baseline. We explore the impact of excluding respondents who attrited after baseline and we then test the effectiveness of a multiple imputation strategy to adjust for attrition. We conclude that a multiple imputation strategy, based on measured characteristics at baseline, may correct for attrition. However, evidence of whether attrition is ignorable and/or if attrition is predicted by changes in the lives of the respondents that occurred after the baseline interview cannot be provided by standard post-survey adjustment strategies. In order to shed light on this matter we test a multiple imputation strategy, exploiting information from respondents who had not

been found in the second wave of the MxFLS but who were recovered in the third wave. While most longitudinal surveys do not track individuals that had not been found in previous rounds, this unique aspect of the MxFLS makes it possible to compare a model of earnings using data measured in the third wave of the MxFLS with data calculated using a post-survey adjustment strategy. The results suggest that, in this case, the multiple imputation strategy is not capable of successfully replicating the information of attritors. This analysis provides evidence that changes in the respondent's characteristics between waves are important predictors of attrition and highlights the limitation of post-survey adjustments based solely on baseline characteristics.

2.2 Conceptual Framework

The consequences of non-response are a relevant issue in any study that relies on household and individual level data. Longitudinal surveys at the baseline and cross-sectional studies may suffer from non-response bias if sampled respondents are not assessed. The issue gets more complicated in following waves of longitudinal surveys because maintaining the initial representativeness of the sample depends on finding the same individuals interviewed at the baseline.

Studies that have examined the effects of non-response in longitudinal surveys treat the impact of attrition as a selection bias problem. Seminal work by Fitzgerald et al. (1998) provides a model of attrition that distinguishes selection on observed and

unobserved characteristics. This is a key distinction since attrition that is selected on observed factors may be taken into account with post-survey weights, while attrition that is associated with characteristics that are not observed potentially poses a larger concern. Alderman et al. (2001) follows the Fitzgerald et al. (1998) methodology in their analysis of attrition in panel data from Bolivia, Kenya and South Africa. These studies conclude that, conditional on observed characteristics, attrition is ignorable. This essay follows the approach of Fitzgerald et al. and develops an imputation strategy to adjust for attrition using data from MxFLS. In addition, we extend the approach by taking into account uncertainty in the imputation strategy by adopting multiple imputation methods suggested by Rubin (1987).

Thomas et al. (2012) note that in addition to this literature on attrition in panel data, a different line of survey research has highlighted that non-response is related to the structure of the interview and characteristics of enumerators' characteristics on attrition rates. Integrating these perspectives in a study of attrition in the Indonesia Family Life Survey (IFLS), the authors argue that it is difficult to rule out that attrition in IFLS is correlated with characteristics that are not observed at baseline (in the context of an earnings function) and show that characteristics of interviewers are predictive of attrition and may serve as instruments for attrition in behavioral models.[2] While our

[2] With regard to the role of survey design, Olsen (2005) highlights the importance of persistence during interview follow-ups and the significance of explaining the survey in a clear way to both interviewers and respondents. See also Maluccio (2000) and Zabel (1998).

work on MxFLS follows the first strand of the literature, on-going research will integrate the second perspective into the analyses.

Following Fitzgerald et al. (1998), we describe our conceptual framework in this section. Assume that the outcome of interest is a conditional population density $f(y|x)$, where y is the dependent variable of interest and x is a vector of independent variables. Let A be a latent index of attrition equal to one if y is missing because of attrition, and assume x as a set of covariates that is observed for the entire sample (this assumption is valid for the models developed in the following sections. One can think of this vector as time invariant characteristics or as lagged variables). Given that y is not observed over the entire sample, the econometrician can estimate $g(y|x, A = 0)$. The challenge is to infer the function f from the estimation of g.

One solution is to impose restrictions on the attrition function (defined in equations 2.2 and 2.3 below) to infer the real density f from the estimated density g. Define the probability of attrition as $\Pr(A = 0|y, x, z)$ where x and z are observed for all the sample. The restrictions imposed in this function in order to test for attrition bias depend on whether attrition is selected on observed or unobserved variables. Formally, attrition is selected on observed characteristics if $\Pr(A = 0|y, x, z)$ can be reduced to $\Pr(A = 0|x, z)$ and:

$$y_{it} = \beta_0 + \beta_1 x_{it} + \varepsilon_{it} \tag{2.1}$$

$$A_{it}^{*} = \delta_0 + \delta_1 x_{it} + \delta_2 z_{it} + \mu_{it} \tag{2.2}$$

$$A_{it} = 1 \quad if \quad A_{it}^{*} \geq 0 \tag{2.3a}$$

$$A_{it} = 0 \quad if \quad A_{it}^{*} < 0 \tag{2.3b}$$

where μ is an unobserved factor that may affect attrition. When attrition is a selection on observed characteristics then $\varepsilon|x$ is independent from μ but $\varepsilon|x$ is not independent from z. The variable z is an endogenous variable that affects both attrition and the outcome y but affects y only through A. In a model of earnings, for example, basic demographics such as age, gender and marital status have been found to be determinants of both the earnings and attrition equations (x variables). When attrition is selected on observed factors, the real density function f can be computed using weighted least squares with the weights from the conditional joint density of the outcome of interest observed only for non-attritors and the characteristic z observed for the entire sample.

Selection on unobserved characteristics occurs when $\varepsilon|x$ is independent from z but $\varepsilon|x$ is not independent from μ. For example, in the context of the model of earnings suppose that the variable z is a measure of the length of the interview and suppose that in the residuals in the model of earnings are unobserved measures of ambition. Unobserved variables that are endogenous in a behavioral model may influence the willingness to move to places with higher wages, and individual's or household's migration may difficult tracking, which affects attrition rates. In this case, residuals in the model of earnings are not correlated with z but are correlated with μ. One of the

standard methodologies to address selection on unobserved characteristics is based on the Heckman selection model.[3]

The Heckman selection model is a two-step equation model, where the first equation is a regression based on equation (2.1). The model is truncated because the outcome y is not observed for the attritors' sample. Equation (2.2) represents the estimation of the selection model; in this case the attrition function.

If the correlation between ε and μ is zero, then sample selection is not an issue and OLS estimates are consistent. When the correlation is different from zero, the Heckman correction allows the use of information from the attritors' sample in the regression model. However, identification of $\beta's$ requires an exclusion restriction, a variable z that explains attrition and is not correlated with the residual on the earnings equation. Finding such a variable (an instrument) is a challenge as any variable related with behavioral choices may influence both models and be invalid.

In the framework of models of attrition in longitudinal surveys the ideal $z's$ that satisfy the exclusion restriction are characteristics of the survey. Such variables explain attrition and are exogenous to the respondents' behavior. Variables such as the characteristics of the surveyors (as long as surveyors are allocated randomly to

[3] Along with selection correction methods, an additional test proposed for Fitzgerald et al. (1988) is the comparison of the survey of interest with an external data set. In their study they compare the PSID with the Current Population Survey (CPS). The next section of this chapter discusses possible tests of attrition in the context of the MxFLS.

respondents) are good examples of possible z variables. As mentioned before, such variables are used by Thomas et al. (2012) as instruments for attrition in behavioral models using the IFLS.

2.3 Data: The Mexican Family Life Survey

The Mexican Family Life Survey (MxFLS) is an ongoing longitudinal data set that is representative at the national level and includes information on approximately 8,440 households and 35,600 individuals among 150 communities throughout Mexico. The baseline survey was conducted in 2002, the first follow-up started in 2005 (MxFLS2), and the third wave started in 2009 (MxFLS3) and it is currently in the final stages of fieldwork. The MxFLS is designed to follow all baseline respondents and their children who were born after baseline, regardless of migration.

Table 1 shows attrition rates in 2005. About 11% of the sample was lost to follow-up. A fundamental characteristic of the survey is that every respondent in 2002 is a target respondent for follow-up regardless of migration, such that Mexican migrants to the U.S. and return migrants from U.S. to Mexico are tracked and interviewed as well as movers within Mexico. This is the first longitudinal survey that attempts to follow-up Mexican migrants in the U.S. In 2005, about 2.4% of the baseline respondents were known to have moved to the U.S. Of those, 91% were tracked and interviewed in the U.S.

The first two columns of Table 1 show the number of individuals tracked in 2005 for the entire sample. The percentages are very similar between the complete sample and the sample of individuals older than 15. The analysis of this chapter will focus on adults (members older than 15 years of age in 2002) since they were individually assessed by an enumerator whereas a primary care-giver provided information about each child (under 15).

Of the original sample, 543 people had died by 2005, 525 of which were adults. All the original members except those who died comprise the eligible sample in 2005. Panel A of Table 1 shows that of the 11% attrition rate, 8.8% of the cases involve entire households that are lost. Approximately 83% of the 2002 sample is found in their original 2002 household, and 6 percent is found either in a new household in Mexico or in the U.S. In 2005, 4,529 individuals are new respondents because they were living with a target respondent. Panel B shows that from the people found, 86% complete the interview and almost 4 percent are found but are not interviewed. From the found but non-interviewed, less than 1 percent are directly classified as refusals. Although there is not complete data for these individuals, the reasons for non-response in these cases may be quite different from those who are lost to follow up.[4] Individuals who refuse to answer take a conscious decision about cooperating with the study. In the MxFLS the

[4] Groves and Couper (1998) highlight the importance of distinguishing between different types of attrition.

group of individuals found and without an interview may be considered as refusals. An analysis distinguishing different types of attrition will be explored in future research.

The rich information collected in the MxFLS allows to explore a rich set of covariates as predictors of attrition. The MxFLS contains information about the economic, social and health status of each member of a surveyed household. The questionnaire for adults includes sections on education, labor supply and earnings, migration history, marriage and fertility history, health status, and use of health care. In addition, one member is interviewed about information at the household level. This questionnaire includes a complete household roster including basic socio-demographic characteristics of each household member, and information of household expenditure, and assets ownership (Rubalcava and Teruel, 2006).

As well as variables that are traditionally found in household surveys, the MxFLS contains detailed information on characteristics that are not standard in broad-purpose socio-economic surveys. The MxFLS questionnaire includes information about expectations about the future, perceptions of current and future health and well-being as well as information about crime and victimization.

2.4 Analysis of Attrition

With this background, this section draws comparisons between the group of attritors and non-attritors across multiple variables measured at the baseline. A model of

attrition is also estimated as a function of variables measured at the baseline in order to examine whether differences between attritors and non-attritors hold after controlling for diverse socio-demographic characteristics. In the spirit of Becketti et al. (1988) and Fitzgerald et al. (1998), a test of attrition is performed using a model of the quartic root of monthly earnings in 2002. The model of earnings is chosen because it is a common dependent variable used in economic studies and testing whether attrition may bias its results is an important contribution. After testing for attrition, possible methods of correction are discussed. The limitations of the Heckman model represent a challenge in the context of the MxFLS, so an alternative procedure will be discussed in the next section.

2.4.1 Descriptive Statistics

This section examines the observed correlates of attrition in the MxFLS. Tables 2 and 3 show comparisons between the sample of attritors and non-attritors. The comparisons are made across a number of characteristics measured at baseline (2002), therefore are observed for all the sample of respondents. Column three of Table 2 shows the differences in the mean for all the sample of individual older than 15, and Table 3 shows the differences in the mean for the sample stratified by gender. Most of the results hold for both males and females; however, some of the dissimilarities that exist by gender are highlighted at the end of this section.

The variables are classified in two broad groups. The basic model includes variables measured in most surveys and which are considered as standard determinants of attrition. This group includes individual demographics, participation in the labor force, household composition and resources, and the rural-urban characteristic of the locality where respondents were living in 2002.

The variables in the extended model consist of non-typically measured variables. Along the lines of Abraham et al. (2006), measures of community integration are included. Respondents better integrated to the community have more contacts, which could facilitate their tracking. Migration history and plans about moving may also be determinants of attrition. We will test whether past migration and expectation about migration in the future are important predictors of attrition. This of great relevance since, as mentioned in the introduction, most of attrition in developing countries is driven by loss to follow-up of migrants.

The extended model also includes a set of variables that measure perceptions of quality of life, crime and victimization and wellbeing. The index of emotional wellbeing is built from 21 questions comprised in a section specially designed to measure welfare. The possible values of the index vary between 9 and 63, being 9 the person that felt better and 63 the person that felt worst in terms of emotional welfare.[5]

[5] As an illustration, consider the following question, in parenthesis next to each question is the value that the answer receives to build the wellbeing index: In the last 4 weeks, have you felt lonely? 1. Yes, sometimes (1) 2. Yes, a lot of times (2) 3. Yes, all the time (3) 4. No (0).

The results in Table 2 show that the sample of attritors are on average younger and more likely to be single. On the other hand, attritors are more likely to be more educated and to have more educated parents; they are taller and show higher average cognitive scores. The general results of the variables that measure households' resources show that homeowners and farm-business owners are less likely to attrite. However, individuals with higher levels of wealth are more likely to be in the attritors' sample. Higher levels of per capita expenditure (PCE) also increase the odds of not being interviewed; but it is important to underline that the effect of the per capita expenditure may be driven by the fact that smaller households are lost with higher probability.

Respondents who are tracked in the first follow-up are more likely to live in households with more children and to have at least one co-resident parent, although the difference with the non-tracked individuals is not significant. Abraham et al. (2006) maintains that the number of children may be a proxy for integration in the community. In the same line of analysis, having at least one co-resident parent alive may augment the probability of tracking. If older people are less mobile, their contact information may be crucial to relocate moving households.

Healthier people may be more willing to incur the physical costs of migration. Two types of variables are considered as a measure of health status, a perception of respondents' health status relative to their age and gender demographic group, and objective health indicators. Self-assessed health is difficult to interpret as it shows the

information each person has of her own health rather than objective health. Poorer people that do not have access to health facilities may be less healthy but also may not be aware of it. To complement the analysis of health, measures of hypertension and the body mass index (BMI) are also included. The set of health related variables show that attritors have better measures of both hypertension and BMI, and perceive themselves as healthy relative to their age and gender group. In the set of variables measuring participation in the labor force, attritors are more likely to be employed and to be non-agricultural workers, which is consistent with the result that show that respondents in rural areas have lower probability of being lost during follow-up.

A higher opportunity cost to participate in the survey and a busier working schedule can also significantly affect the likelihood of survey response. For the MxFLS, the numbers support this hypothesis. For every employment category, individuals who were not interviewed had significantly higher total earnings at baseline and worked on average two additional hours per week.

Among the variables that measure the level of integration with the community the key is to consider characteristics that may facilitate the contact of eligible interviewees in 2005. Non-attritors are more likely to be beneficiaries of welfare programs, which is consistent with the fact that in order to keep receiving these program's benefits individuals are not allowed to move away from their original locality. Interestingly, the descriptive statistics show that attritors seem more tied to the

community. They are more likely to participate in activities outside their household and to take care of vulnerable people. Even though these variables may increase the probability of being more involved with the community, they could also diminish the probability of finding respondents at home. Moreover, attritors are more likely to know people or institutions that could lend them money. This result is consistent with the fact that they are more likely to have thought about migrating. Migration requires an important initial investment that the poorest people cannot afford. Attritors are, in fact, more likely to have moved by age 12, and 9.9 percentage points more likely to think about moving in the future.[6]

The last group of characteristics measure perceptions of wellbeing, crime and victimization, and will allow us to condition the probability of attrition on a large set of observed variables. Because these variables are not often collected in longitudinal surveys, if the group of attritors and non-attritors are different in this set of variables, this could be a first proof of selection on characteristics that are usually unobserved. Moreover, these variables may work as a proxy for the quality of life in the place of residence in 2002. Lower cooperation between neighbors, higher perceptions of insecurity and lower perceptions of improvement in the quality of life may affect the desire to change residence. The results show that attritors are more likely to have

[6] Thomas et al. (2012) finds that respondents that moved before 12 are less likely to be re-interviewed.

perceived improvements in their qualities of life but feel less safe and have been victims of robbery with higher probability. Respondents that felt less safe in 2002 might have moved without leaving any contact information, which could hinder their tracking. The index of wellbeing shows that attritors are more likely to have higher levels of emotional status.

Many significant differences exist between the sample of attritors and non-attritors and the general findings are similar for the stratified sample by gender. The results in Table 3 show that the differences in health measures between the attritors' and non- attritors' sample hold only for females. The most remarkable differences between females and males are the measures of participation in the labor force. Females that worked the week before the interview are more likely to be lost in follow-up, while for males this difference is not significant. In addition, male agricultural workers are more likely to be found while for females the probability of being involved in agricultural activities is not different between attritors and non-attritors. Moreover, when disaggregating the measure of earnings by employment sector, the differences are only significant for men at the 5% confidence level. These results may reflect some differences in the patterns of attrition between males and females.

At this point of the analysis, attritors and non-attritors look different in a significant number of observed characteristics and in the set of variables that measure expectations, perceptions of safety and quality of life. Logit models in the next section

allow controlling for a different set of socio-demographic variables to analyze whether these differences persist.

2.4.2 Prediction of Attrition

This section shows results of estimating logistic models that predict the probabilities of attrition between baseline and the first follow-up using the same characteristics presented in the previous section. The analysis is based on data collected at baseline.

As in the previous section the sample consists of all 2002 interviewees older than 15 years of age. The analysis of the descriptive statistics showed evidence of differences in many key socio-economic variables between those who attrite and those who are found, which should generate some concern in ignoring attrition. The logit model allows analyzing these differences controlling for other socio-economic variables. Identifying the characteristics of attritors contributes to improving the design of follow-ups and to determine whether attrition may bias estimation results.

The model is based on the function of attrition explained in section 2.2. The goal is to estimate $\Pr(A = 0|y, x)$, where y is the outcome of interest observed only for non-attritors and x is the set of covariates measured in 2002. The outcome y can be estimated from the set of covariates observed for the entire targeted sample, therefore $E(y|x) = \beta_0 + \beta_1 x$. When $\Pr(A = 0|y, x) = \Pr(A = 0|x)$, the selection of attrition is based only on

observed characteristics and controlling for these should take into account the differences between the sample of attritors and non-attritors.

Two models that estimate the probability of attrition are examined. The basic model estimates the probability as a function of 2002 characteristics considered in standard models of attrition. This model includes as covariates demographic characteristics at the individual level, household's composition, measures of households' assets and socio-economic status, individual health measures, participation in the labor market and whether the place of residence in 2002 is rural or urban (Table 4). An extended model adds variables that measure migration experience and expectations about future migration, integration with the community and perceptions of quality of life (Table 5).

Tables 4 and 5 show the results for the basic and extended logit regressions, where the dependent variable equals one if the respondent was not found in 2005. The results are shown as odd-ratios that indicate the odds of not being tracked relative to being tracked in 2005 for each of the covariates. All the models control for municipality of residence at baseline and the standard errors are clustered at the municipality level. The results of the first column show the odds for the entire sample, and columns two through five present the results after stratifying the sample by gender and by rural and urban areas. Special interest should be given to the set of variables of quality of life, crime and victimization, and migration expectations. These variables are not usually

measured in standard surveys or are not usually considered in attrition models (an exception is Fitzgerald et al. that controls for expectations of migration). The significance of these variables in a model of attrition could give support to the fact that attrition may be determined by unobserved characteristics.

The results for the basic model (Table 4) show that gender, human capital, household's resources, household composition and whether the place of residence is urban or rural are significant predictors of attrition, and their significance holds when looking across the columns in the stratified samples. The results in the first column for the entire sample show that being a female increases the odds of being lost in follow-up, and this significance persists in urban places.

Looking at the different set of characteristics across the different samples, it is evident that human capital is an important predictor of attrition, but there is an important level of heterogeneity depending on gender and place of residence. Better educated respondents as well as those with better educated parents are less likely to be found, and this result is being driven by male respondents and individuals living in urban places. Both the respondent's and their father's education have a non-linear effect on the probability of attrition. Male respondents with complete high school are 54 percent more likely to be lost in follow-up relative to men with incomplete primary and those with college or more are 45 percent more likely to be attritors from the sample. If, for example, unobserved characteristics correlated with education, like ambition, drive

attrition, and, it is the most ambitious individuals within each education group that have the highest earnings after baseline, a post-survey adjustment based only on education would underestimate the returns to education in future follow-ups. Moreover, higher levels of education increase geographical mobility especially in urban areas where the expected return of education is higher. Movers are more difficult to track, and variables that explain mobility may also be determinants of attrition.

An additional difference in the stratified versions is the effect of marital status in urban and rural areas. While in urban areas being married does not have a significant effect on the probability of being lost, in rural areas married individuals are 60 percent less likely to be interviewed. In the category of household characteristics, homeowners are on average 65 percent more likely to be interviewed in follow-up and this is consistent across the different samples (except in rural places where the probability decreases to 25 percent). This result may indicate that those with higher investments at the place of residence may be less likely to be lost in follow-up because of the lower probability associated with migration. The measure of wealth per capita shows that the probability of being found is lower among the households above the 25th percentile of the wealth distribution, and the results are significant in the different samples with the exception of rural localities. This is consistent with the fact that households with more assets are less mobile and therefore easier to track.

The effects of household composition are as expected. Households with more children and co-resident parents have lower probabilities of being lost. This result supports the finding of Abraham et al. (2006) that households with more children are better integrated into the community. However, the set of variables that measure health are not significant predictors of attrition, and these results persist across the different sub-samples.

After controlling for a rich set of characteristics the effects of variables related with the labor market are only significant for the sample of males. Men whose earnings are in the second quartile of the distribution are more likely to not be found relative to those in the bottom quartile, but males at the top of the 50th percentile of the distribution are around 40 percent less likely to attrit from the sample. People with better jobs and higher earnings could be less willing to move, although these people could also have higher opportunity costs. We will discuss in detail a model of earnings in the next section. Moreover, working more hours per week increases the probability of being lost, although the coefficient is very close to one.

Finally, being from a rural community greatly reduces the odds of being lost, particularly in the sample of females. Rural communities rely usually on closer social networks in the community, which facilitates tracking in follow-ups.

Table 5 shows the results for the extended model. The significance and magnitude of the variables of the basic model persist when adding the additional

variables of the extended model. In the set of variables that measure migration, on the one hand, individuals who have migrated by age 12 seem to be more likely to be lost in follow-up, however, the effects are not significant in any of the sub-samples. The variable that measures expectations about migration in the future, on the other hand, is a significant predictor of attrition in every model, except for individuals living in rural areas. Women who expect to migrate in the future are 26 percent more likely to be lost in follow-up, this number increases to 30% for males and to 34% in urban areas. An important result is the lack of significance of this variable in rural areas. Rural areas had a higher follow-up rate (in urban areas the attrition rate was 15.07% while in rural areas was only 4.35%) and one of the reasons could be the lower mobility rates in these areas. After controlling for these rich set of variables, the initial significant effects found for the characteristics that measure emotional well-being or perceptions on security lose their significance in the different sub-samples.

The results of the logit models show that, in addition to the standard socio-economic characteristics, expectation about future migration play a significant role as determinant of the probability of attrition. Moreover, it seems like the attritors' sample has similar characteristics to the standard variables that characterize migrants. Less educated people, individuals more likely to be married and homeowners show lower geographic mobility (Rosenzweig, 1986; Smith and Thomas, 1998; Thomas et al., 2001).

Ignoring attrition or performing corrections based solely on observed characteristics may bias the estimations when attrition is selected on unobservables that are related with the outcome of interest. Unobserved factors like, for example, ambition and aversion towards risk, may determine the decision to participate in certain sectors of the labor market. If such participation requires migrating or working more hours, the probability of attrition may be affected, as well as earnings. Ignoring factors that explain both models cause biased estimations.

Assessing whether attrition is ignorable when is selected on unobserved characteristics remains an empirical question and the methods to correct for attrition impose important assumptions on the distribution of the missing data. Assuming that attrition is ignorable when it is not can seriously compromise the validity of the estimations based on longitudinal data. The next section explores the effect of attrition in a model of earnings.

2.4.3 Attrition in a Model of Earnings

This section presents a model for the quartic-root of monthly earnings and a test for attrition bias in the spirit of Becketti et al. (1988) and Fitzgerald et al. (1998). The test for attrition proposed in this section complements the discussion held in previous sections.

The test originally proposed by Becketti et al. (1988) estimates the outcome of interest on a set of covariates x and on whether the respondent is found in subsequent

waves. While in the previous section the model estimated the probability of attrition based on a set of covariates including y_0, the model in this section estimates the expected value of y_0 conditioned on x and A. Equation (2.4) shows the attrition function, where A^* is a latent index for attrition explained in previous sections, x is a set of 2002 covariates and y_0 is the quartic-root of monthly earnings in 2002. In this model 2005 information is used only to distinguish the sample of attritors and non-attritors.

$$A_{it}^* = \delta_0 + \delta_1 x_{it} + \delta_2 y_{0it} + \eta_{it} \tag{2.4}$$

The purpose of the test is to identify whether the conditional expectation of y_0 depends on the attrition function by inverting equation 2.4 and taking expectations to measure $E(y_0|A, x)$. The empirical strategy to determine the effect of attrition on the conditional expected value of y_0 is based on a linear approximation of $E(y_0|A, x)$ using OLS (Becketti et al., 1988).

In the framework of the MxFLS, the idea of the test is to estimate separate regressions for the respondents that attrite and those that did not and compare their coefficients. The comparison is with data in 2002 between the attritors and non-attritors in order to examine how different the parameters are if only the non-attriting sample is considered in a statistical analysis. The following estimation represents the model of earnings, where the subscript equal to one in y and x stands for the sample of respondents found in 2005, the subscript equal to zero stands for the sample of

respondents not found in 2005 and the subscript with 1 and 0 represents the full sample of respondents:

$$qry_1 = \alpha_0 + \alpha_1 x_1 + u \tag{2.5}$$

$$qry_0 = \beta_0 + \beta_1 x_0 + v \tag{2.6}$$

$$qry_{1,0} = \delta_0 + \delta_1 x_{1,0} + \varepsilon$$

$$qry_{1,0} = [\omega\alpha_0 + (1-\omega)\beta_0] + [\omega\alpha_1 + (1-\omega)\beta_1]x + \varepsilon \tag{2.7}$$

Equation (2.5) estimates the quartic-root of monthly earnings at baseline for non-attritors, equation (2.6) estimates the same regression for the sample of attritors and equation 2.7 estimates the full sample, where ω is the weight of non-attritors in the sample.

The test proposed by Becketti et al. is whether $\delta_0 = \alpha_0$ and $\delta_1 = \alpha_1$, whether using only the sample of attritors presents significantly different estimates from using the full sample of targeted respondents. When the hypothesis of equality between coefficients cannot be rejected, attrition is not a significant determinant of y_0 and does not bias the estimates of the model. This test is equivalent to a test of the γ coefficients in the following equation:

$$qry = \alpha_0 + \alpha_1 x + \gamma_0 + \gamma_1 xL + \mu \tag{2.8}$$

where L is a dummy variable equal to one if the respondent was not found in 2005. Therefore,

$$\beta_0 = \alpha_0 + \gamma_0 \tag{2.9}$$

$$\beta_1 = \alpha_1 + \gamma_1 \tag{2.10}$$

and the test can be re-written as:

$$\delta_0 = \alpha_0$$

$$\omega\alpha_0 + (1-\omega)\beta_0 = \alpha_0$$

$$\alpha_0 = \beta_0$$

Using equation (2.9):

$$\gamma_0 = 0 \tag{2.11}$$

which is a test of the γ coefficients on equation (2.8). The same analysis holds for the α_1, β_1 and γ_1 coefficients. The test of differences between the coefficients of equations (2.5) and (2.6) is therefore equivalent to the test that looks at the significance of the interacted terms in equation (2.8). Testing coefficients across separate regression may be more difficult that testing the γ coefficients.

Table 5 shows the results of equation (2.5) for the quartic-root of monthly earnings conditioned on 4-segment spline functions in age[7], education, height and whether the locality of residence was rural or urban. The sample is restricted to respondents in 2002 between 21 and 65 years old that report being working the week before the interview and are not unpaid, includes municipality fixed effects and clusters the standard errors at the municipality level.

[7] Semi-parametric spline functions impose fewer restrictions on the functional form than the quadratic option.

The results in Table 6 show the estimations of equations (2.7) (model 1) and (2.8) (model 2), for the entire sample and stratifying by gender and by urban-rural. Model 1 shows the coefficients without attrition. In model 2 the purpose is to test the significance of the interacted terms. When the coefficients with the interaction are significantly different from zero, then the hypothesis of equality of coefficients between attritors' and non-attritors' is rejected and attrition bias is evident.

Without stratifying by gender (columns 1 and 2), the interaction of the indicator for attrition with the coefficients of gender, incomplete high school, and college or more, are significant. If in the follow-up, the sample of attritors were ignored, the negative effects of the female gender on earnings would be overestimated and the returns to incomplete high school and college, relative to incomplete primary, would be underestimated. These results show that the individuals that are not being tracked are not missing at random. The women in age to work that are not tracked in 2005 have higher earnings, as well as attritors with some high school or college.

Although women lost in follow-up have on average higher earnings, the returns to education between the samples of women who were found and those who were lost in follow-up do not seem to be significantly different. Even though the magnitude of the coefficients of the interaction between the variables of education and the indicator of attrition are not small, the effects are imprecisely estimated. While the results for women are not significant, it is interesting to contrast their sign and magnitude with the results

for males. While for women, ignoring the sample of attritors could potentially overestimate the effect of education on earnings the effect for males show exactly the opposite pattern. In the sample of males, all the coefficients of education are significantly different between the sample of attritors and non-attritors. For levels of education above complete primary, considering only the sample of non-attritors would underestimate the returns of education. These results support the findings of the previous section. Not only are more educated males lost with higher probability, but also those with higher returns to education.

This result supports one of the prior conclusions: men and women have different determinants for attrition, and education is one of the determinants that plays a differing role. While higher education increases the probability of attrition for males, it does not have an effect on females. In addition, while using only the non-attritors' sample underestimates the returns of education in the group of males, the effect is not significant for females.

Previous results also show differences on the attrition patterns in rural and urban localities, and the results in Table 6 support those results. On the one hand, in rural areas education is not a strong predictor of attrition and those with the lowest returns to education are more likely to attrit from the sample. On the other hand, education significantly predicts attrition in urban areas and those without an interview in the follow-up have also the highest returns to education. Ignoring the sample of attritors

would, for example, underestimate the returns to college in around 46 percent for individuals living in urban places.

2.4.4. Methods to Handle Missing Data due to Attrition

The previous sections showed the differences between attritors and non-attritors and the bias that exists in a model of earnings when ignoring attrition. The results of both the model of attrition and the model of earnings highlight two conclusions: first, attritors and non-attritors are different in a significant number of characteristics including a set of non-standard observed variables; second, estimations of the model of earnings using only the non-attritors sample bias the estimations. The standard method to correct attrition for selection on unobservables is based on the exclusion restriction. However, finding a variable that explains attrition and is not related with the residuals of the earnings model is challenging.

Although there had been advancements in the literature to handle the cases with missing data due to attrition, , there is still little guidance in the literature regarding the most effective approaches to mitigating the impact of attrition in empirical studies that use longitudinal data. Some of the most used techniques are re-weighting methods and imputation methods that assigns a value to the variables of the missings respondents. There are at least two different strategies that use re-weighting methods. First, studies re-weight the sample of non-attriters so that it replicates the baseline sample. This 'raking' is usually on a small set of characteristics such as age, gender, education and

location. Second, studies estimate a multivariate model of attrition and calculate the risk or propensity that an individual will attrite from the sample. The propensity score is used to re-weight the sample of non-attritors. This method typically uses a relatively large number of characteristics to predict the propensity. When attrition is selected only on observed factors, post-survey weights can adjust for attrition. The great limitation of this strategy is that without a sample of attritors it is not possible to test the assumption of selection of attrition only on observables measured at baseline.

An alternative methodology to solve for attrition is to impute values for missing information in the outcome of interest. Both a single imputation and a multiple imputation strategy have been largely used in the literature. Multiple imputation, takes one more step and essentially repeats the single imputation process multiple times to take account of the fact that attrition is a stochastic process (Rubin and Zanutto, 2002).

2.5 Multiple Imputation

The imputation strategy is a method that deals with missing variables based on a specific model that results in a valid statistical inference (Rubin, 1987). In longitudinal surveys, the imputation strategy model gives a value to the missing variables as a function of observed characteristics that determine the non-response behavior. In this sense, both multiple and single imputation models allows including hypothesis for the reasons of non-response. Rubin (1987) claims that assumptions about non-response are

not straightforward and a number of possible models for non-response should be considered in the analysis.

The imputation model makes a match between a respondent with complete information and one with missing information. The match is made based on the hypothesis of non-response. The fundamental advantage of multiple imputations against a single imputation strategy is that it imputes several values (m) to the outcome y, which creates m different complete data sets (for a discussion of the optimal number for m see Rubin, 1987). The multiple imputation accounts, therefore, for the variance of estimates and this increases the efficiency of its estimates (Rubin, 1987).

2.5.1 Imputation of Earnings at Baseline

In this section we predict earnings in 2005 based on observed variables, namely those variables measured in 2002 that predict attrition, for all employed respondents, aged 21 to 65, in 2002. Before explaining the model of earnings in 2005 and its results, it is important to conduct an examination of the validity of the model of imputation by performing a test using the baseline data.[8] For this analysis, the sample is restricted to respondents, between 21 and 65 year old at baseline, that report working the week before the interview (self-employed and employees).

[8] The same exercise has been performed using a re-weighting strategy that calculates a propensity score for the probability of attrition to re-weight the sample of non-attritors. For a model of earnings the results from the multiple imputation strategy were more closely matched to the results using the original data.

To carry out this test the same model of earnings is applied to two different samples. One sample is the entire set of respondents that have full information at baseline and the second is the same data set with multiple imputation estimates replacing the 2002 earnings variable for non-tracked individuals in 2005. If the multiple imputation model replicates the estimations found when using the original 2002 data we can be more confident about a multiple imputation strategy to handle missing earnings in 2005 of those who attrited.

However, this exercise has an important limitation. Even if the multiple imputation model replicates the estimations of the original data, we still cannot identify whether non-response is related to events that happened between baseline and follow-up. For example, if respondents attrited because of life changes that occurred in the hiatus of the survey's waves, then attrition will, by definition, depend on unobserved characteristics. We will discuss potential methodologies to test whether differing changes in baseline characteristics may be correlated with attrition at the end of the section.

We start our analysis by examining the imputation model at baseline. In the sample of employed respondents, a small proportion of individuals do not have information for earnings. In this case, missing information on earnings in the targeted sample is not explained by attrition. In order to deal with this different type of missing data a two-stage imputation model is proposed. The first stage imputes earnings to the

respondents that did not answer the question for earnings and reported being employed. This is not an imputation for attrition. The imputation is based on basic demographics, education and characteristics at the locality level. Once the data at baseline is complete the second-stage imputation can be done.

In addition, we examine the extent to which the inclusion of characteristics that are not usually observed in surveys in the imputation algorithms affects the substantive results. This section discusses the results of a basic and an extended strategy for the multiple imputation models based on the two versions of the attrition model.[9] If the set of additional variables, in particular migration expectations, are good predictors of attrition taking those into account should improve the accuracy of the imputation models.

Table 7 shows the results of the test of the imputation model. The first three columns report the results for the full sample and the following columns for the stratified sample by gender, and by urban-rural. For each sample three models are reported: the model with the original data at baseline, the basic imputation model and the extended imputation.

[9] The accuracy of the imputation strategy relies on the model that defines attrition and, following Rubin (1987), more than one model should be specified when dealing with multiple imputation strategies.

The results show that the multiple imputation strategy is a reliable tool to estimate the model for earnings with missing variables.[10] However, in contrast to our intuition, it seems that the extended imputation does not add precision to the imputation model.

As shown in the previous section, estimates of (2.8) provide a test of whether using only the non-attritors sample biases the results. Table 8 shows the results using the original data at baseline, and both imputed data sets. The coefficients without interactions represent the coefficients for the sample of non-attritors. The first three columns show the results for the entire sample. The coefficients for the non-attritors are similar between the original data and both the imputed data sets. However, the significance and magnitude of the interacted terms are different between samples. Specifically, the data set imputed with the basic model only shows a significance effect for the interaction between attrition and high school incomplete. Moreover, the estimate of the interaction between the indicator of attrition and the gender variable not only loses its significance when using the imputed data but the coefficient has the opposite sign.

[10] A test for the difference in the coefficients between samples was made and only the rural variable is significantly different at the 5% significance level between the original data and the basic imputed data set for the sample of males; and, the variable of some college or more between the original data and the extended imputed data set for the urban sample at the 10% significance level.

In this case the extended model, which incorporates migration expectation questions, shows more accurate results but still does not replicate perfectly the original data for the sample of attritors. This could indicate that including variables potentially unobserved in most of the surveys captures better the heterogeneity between tracked and non-tracked individuals at the baseline, but not perfectly. When stratifying the sample by gender and urban-rural, and focusing on the coefficients of the interaction with the attrition indicator, the results show that the extended model is in every case the most accurate. We can also infer that the extended model may work better because it captures the unobserved heterogeneity between tracked and non-tracked individuals.

However, for the sample of males and the sample of rural localities, the extended imputation model does not replicate the estimates of the original data. For both samples, both imputation models, do not properly capture the relationship between education and labor market success for the sample of attritors. Although the results in Table 7 indicate that a multiple imputation strategy is reliable for a model of earnings, the results in Table 8 show that the replication is not perfect particularly when focusing on the returns to education of the sample who attrited from the survey.

To sum up, the results in Tables 6 to 8, show that ignoring the sample of attritors would bias the estimations on a model of earnings. Both imputation models are successful in replicating on average the estimations of the model that uses the original data; however, the estimations for the sample of attritors is not perfectly matched,

particularly for males and rural areas. In the next sub-section we perform an imputation analysis for the earnings in 2005 of individuals who were loss in follow-up and measure the bias caused by ignoring the sample of attritors.

2.5.2 Imputation of Earnings in MxFLS2

Table 9 presents results for the model of earnings in 2005 using the version of the extended model for the multiple imputation. The first column shows the results of the earnings model for the sample that was tracked in 2005, the second column shows the results for the same sample that additionally includes multiple imputation estimates of earnings for the individuals between 21 and 65 that were working in 2002 and were not tracked in 2005, and the third column shows the significance of the difference of the coefficients between the two samples. If we rely on our imputation model, the results in Table 9 show that without a post-survey adjustment strategy, and relying only on the tracked sample the returns of education would be overestimated for females and in urban places. However, in previous results we show evidence of the potential underestimation in the returns of education for the attritor's sample in urban places when using a multiple imputation strategy. Without information on the attritors' sample in future follow-ups it is not possible to test the assumptions of the different techniques to handle missing data. In the next subsection we exploit information on individuals who attrited from the MxFLS in 2005 but were found in 2009. We make a contribution to the understanding of the potential bias of post-survey adjustments that rely solely on

observed characteristics at baseline by comparing the estimates from a model of earnings including information of attritors in the 2005 wave who were interviewed in 2009 with the estimates from a model that imputes missing data for these individuals.

2.5.3 Imputation of Earnings in MxFLS3

In this section we compare the distribution of the earnings in 2002 and in 2009 between two groups, the tracked individuals in both 2005 and 2009, and the non-tracked individuals in 2005 but found in 2009.[11] Figure 1 shows the distribution of the quartic-root of earnings, of the entire sample, for these two groups in 2002. A Kolmogorov-Smirnov t-test for equality of the distribution functions fails to reject equality of the distribution functions. When we compare the distributions of the quartic-root of earnings in 2009 between both groups (Figure 2), the t-test for equality of distributions is rejected. While we cannot reject that the distribution of earnings was equal in 2002 for both groups, they no longer appear to have similar earnings in 2009, and the distribution of the attritors in MxFLS2 is skewed to the right. This might be consistent with the previous results that show that more educated individuals are more likely to attrit in MxFLS2. These results persist when stratifying the sample by gender. Figures 3 and 4 show the results for males and Figures 5 and 6 for females.

[11] At this time, March 2014, for the sample of our interest in this chapter (individuals age 15 and older at baseline and interviewed in Mexico in both follow-ups),1,025 out of 2,461 individuals not tracked in 2005 have been tracked in 2009 (41.64%).

Table 10 shows the results of the estimates for a model of earnings for individuals age 15 and older at baseline who were active in the labor force at the time of the interview in the third wave. For each sub-sample the first column reports the estimates using the data collected in the third wave with no imputations; the second column, reports the estimates replacing the data of the 2005 attritors with data from a multiple imputation strategy; and, the third column reports the significance of the difference between the estimators using these two different samples. These results are thought to be conservative estimates because, by limiting the imputation process to attritors whom we know are actively working in 2009, we are providing the imputation model with additional information that the researcher would not actually have. In a real world situation, the researcher would not know if the attritors are working in 2009, and thus, they would have to either impute the probability that each attritor is participating in the labor market or assume that only those working in 2002 remained in the labor force in 2009.

The results suggest that a multiple imputation strategy adjust for attrition for the sample of females and individuals living in rural localities; but, the imputed sample underestimates significantly the returns to education in urban places. Even when including a rich set of baseline characteristics, including measures of education, in the imputation process, it is not enough to replicate the estimates for individuals who live in urban places. This is most likely due to the fact that attrition in urban places is

particularly high, and is related to unobserved characteristics or changes that cannot be captured by simply relying on data collected at baseline. Thus, we find an adjustment based solely on baseline characteristics is not able to compensate for attrition for this group of individuals.

Moreover, even though, the results in Table 10 suggest that a multiple imputation strategy might adjust for attrition for females and in rural localities, if the additional 60 percent of the 2005 attritors are significantly different from those who were interviewed in 2009, the attrition bias still persist and it is impossible to test its selection on unobservables. In order to make a contribution to this important discussion, in 2012 we started a follow-up survey within the third wave on a random sample of attritors. Combining information from the respondents tracked before intensive tracking with the intensive-tracking random sub-sample, appropriately re-weighted, will yield a uniquely rich analytical sample that fully represents the Mexican population who were living in Mexico in 2002. These data will be used to rigorously test whether attrition is ignorable by assessing whether attrition is explained by characteristics at baseline and, importantly, changes in the lives of respondents that occurred since the baseline interview. Standard post survey adjustments that are intended to take attrition into account will be rigorously evaluated. These include reweighting and multiple imputation methods.

2.6 Conclusions

The consequences of non-response are a relevant issue in any study that relies on household and individual level data. Longitudinal surveys may suffer from attrition if the targeted respondents are not reached in subsequent waves. Ignoring attrition or performing corrections based solely on observed characteristics may bias the estimations when attrition is selected on unobserved attributes that are correlated with the outcome of interest.

In the first wave of the MxFLS the attrition rate was nearly 11%. In terms of characteristics measured at baseline, attritors are significantly different from those who are interviewed in the first follow-up. These differences extend to, for example, expectations about the future and thoughts about migrating in the future; which are not included in most models of attrition. This evidence suggests that attrition may not be ignorable in Rubin's sense. Experiments with models of earnings functions estimated on the baseline sample indicate that coefficient estimates are biased if those who attrit are excluded from the baseline sample. Moreover, although on average a multiple imputation strategy seems to be a successful strategy to adjust for this attrition bias in the context of models of earnings, when focusing on the estimates of the sample of attritors, the multiple imputation underestimates the returns to education particularly for males and in urban places. The limitations of a post-survey adjustment that relies

only on observed characteristics at baseline is also evidenced when exploiting information of the 2005 attritors that were found and interviewed in the third wave.

Future work using the results of an innovative strategy that conducts and intensive tracking on a random sample of attritors will identify cost-effective ways to find and interview the hardest to find respondents as well as provide rigorous evidence on the most effective empirical methods for reducing the costs of attrition in behavioral models.

2.7 Tables and Figures

Table 1: Attrition Rates

Panel A.

	All		Age >=15	
TRACKING 2002-2005	#	%	#	%
Eligible for survey (total sample 2002)	35,677		23,803	
Died between Waves	543		525	
Eligible to be tracked of whom	35,134		23,278	
Total Found	31,338	89.20	20,652	88.72
Found in original HH	*29,293*	*83.38*	*19,165*	*82.33*
Found in new HH	*1,271*	*3.62*	*929*	*3.99*
Found in US	*774*	*2.20*	*558*	*2.40*
Not Found	3,796	10.80	2,615	11.23
Individual not found	*688*	*1.96*	*515*	*2.21*
Original household not tracked	*3,108*	*8.85*	*2,100*	*9.02*
New entrants	4,529		4,529	
Total Sample 2005	39,663	100	27,807	100

Panel B.

	All		Age >=15	
TRACKING 2002-2005	#	%	#	%
Eligible for survey (total sample 2002)	35,677		23,803	
Died between Waves	543		525	
Eligible to be tracked of whom	35,134		23,278	
Found with interview	30,050	85.53	19,780	*84.97*
Non proxy	*27,497*	*78.26*	*17,514*	*75.24*
Proxy	*2,553*	*7.27*	*2,266*	*9.73*
Found without interview	1,288	3.67	870	3.74
Not Found	3,796	10.80	2,615	11.23
New entrants	4,529		4,529	
Total Sample 2005	39,663	100	27,807	100

Table 2: Characteristics Lost vs. Found Age 15+

Variables measured in 2002	Lost	Found	Difference (Lost-Found)	
Basic demographics				
Age	33.53	38.54	-.501	**
% female	51.63	52.77	-1.14	
Years of Education				
Respondent	8.66	6.70	1.97	**
Father	5.79	3.43	2.36	**
Mother	4.94	3.01	1.93	**
Height	160.33	158.53	1.80	**
Cognitive Score	6.63	5.87	0.76	**
% married	63.28	69.03	-5.75	**
Household resources			0.00	
% HH owns farm business	7.61	23.48	-15.86	**
% HH own no farm business	15.54	15.69	-0.15	
% HH owns house	49.98	70.38	-20.40	**
Per capita expenditure[1]	1,941	1,415	525	**
Wealth[1]	1,317,446	667,707	649,740	**
Household size	4.51	4.96	-0.45	**
Household composition			0.00	
Number of coresident children	1.14	1.43	-0.30	**
% at least one coresident parent	32.06	33.34	-1.28	
% both parents dead	13.85	22.35	-8.50	**
Health			0.00	
% relative health status: better	36.60	33.95	2.65	*
% relative health status: same	59.01	58.98	0.04	
% relative health status: worse	4.39	7.07	-2.69	**
% ever visited hospital/clinic	16.13	16.78	-0.64	
% hypertension	34.82	38.98	-4.16	**
BMI	26.50	26.89	-0.39	**
Labor			0.00	
% worked last week	60.72	54.67	6.04	**
% worked for payment	57.47	51.09	6.37	**
% peasant in own plot	0.92	3.88	-2.96	**
% unpaid	3.06	3.45	-0.40	
% non agricultural employee	41.27	29.73	11.54	**
% agricultural employee	2.72	5.50	-2.79	**
% employer	2.23	2.63	-0.40	
% self-employed	10.33	9.35	0.98	

Continue on next page

1. Earnings in pesos of 2002 (9.56 pesos/US dollars)

Variables measured in 2002	Lost	Found	Difference (Lost-Found)	
Earnings Last 12 Months[1]	22,771	14,658	8,113	**
Earnings Last 12 Months - Self Employed	43,223	26,083	17,140	**
Earnings Last 12 Months -Employee	45,208	31,562	13,647	**
Earnings Last 12 Months - Excluding Unpaid	26,103	15,684	10,419	**
# hours worked regular week	45.58	43.76	1.81	**
Locality characteristics				
Rural	16.37	42.91	-26.55	**
Integration to community			0.00	
% know person/institution to borrow	37.43	33.78	3.65	**
% received income from Progresa	5.35	17.46	-12.11	**
% received income from Procampo	2.61	11.27	-8.66	**
% activities outside HH	20.90	14.36	6.55	**
% take care of elder or sick people	32.89	29.74	3.15	**
Migration			0.00	
% moved by age 12	27.78	23.46	4.32	**
% thought about migrating	25.74	16.22	9.52	**
Quality of life/Expectations			0.00	
% life has improved	39.88	31.86	8.02	**
% life has remained the same	52.20	59.56	-7.36	**
% life has worsen	7.92	8.58	-0.65	
% a lot of fear of being assualted during day time	8.31	6.28	2.02	**
% a lot of fear of being assualted during night time	12.32	10.05	2.27	**
% positive probability of being victim of assualt	24.48	18.78	5.70	**
% has been a victim of assaults	17.33	11.07	6.26	**
% in neighborhood: gangs	29.19	26.03	3.15	**
% in neighborhood: cooperative neighbors	10.79	10.42	0.37	
% in neighborhood: paramilitaries	5.44	4.78	0.66	
% in neighborhood: militaries	5.57	5.80	-0.24	
% in neighborhood: neighbors with guns	3.44	5.40	-1.96	**
% in neighborhood: robbery	21.53	18.24	3.29	**
% in neighborhood: insecurity	9.85	7.15	2.70	**
Emotional Status	6.53	7.11	-0.58	**
Observations	2,615	20,652		

** p<0.01, * p<0.05, + p<0.1

1. Earnings in pesos of 2002 (9.56 pesos/US dollars)

Table 3: Characteristics Lost vs. Found - By Gender Age 15+

	Female			Male		
Variables measured in 2002	Lost	Found	Difference (Lost-Found)	Lost	Found	Difference (Lost-Found)
Basic demographics						
Age	33.54	38.48	-4.94 **	33.53	38.61	-5.08 **
% female						
Years of Education	8.38	6.75	1.63 **	8.93	7.10	1.83 **
Respondent	8.30	6.46	1.84 **	9.05	6.96	2.09 **
Father	5.79	3.39	2.40 **	5.79	3.47	2.33 **
Mother	4.86	2.95	1.91 **	5.03	3.08	1.95 **
Height	155.00	152.93	2.07 **	166.86	165.51	1.35 **
Cognitive Score	6.47	5.70	0.77 **	6.81	6.08	0.73 **
% married	63.62	71.26	-7.64 **	62.92	66.54	-3.62 *
Household resources						
% HH owns farm business	6.95	22.63	-15.68 **	8.31	24.42	-16.11 **
% HH own no farm business	50.48	70.30	-19.82 **	49.45	70.46	-21.02 **
% HH owns house	15.01	15.69	-0.68	16.10	15.70	0.40
Per capita expenditure[1]	1,935	1,408	526 **	1,947	1,423	524 **
Wealth[1]	1,440,725	695,909	744,816 *	1,189,086	636,180	552,906 +
Household size	4.54	4.94	-0.39 **	4.48	4.99	-0.51 **
Household composition						
Number of coresident children	1.19	1.51	-0.32 **	1.08	1.35	-0.27 **
% at least one coresident parent	29.73	30.56	-0.82	34.56	36.54	-1.98
% both parents dead	14.50	22.29	-7.80 **	13.15	22.42	-9.27 **
Health						
% relative health status: better	35.67	32.59	3.07 *	37.74	35.65	2.09
% relative health status: same	59.11	58.77	0.35	58.89	59.24	-0.35
% relative health status: worse	5.22	8.64	-3.42 **	3.37	5.11	-1.74 *
% ever visited hospital/clinic	20.69	21.15	-0.46	10.58	11.30	-0.72
% hypertension	26.38	31.29	-4.91 **	45.55	48.81	-3.25
BMI	26.80	27.47	-0.67 **	26.13	26.16	-0.04
Labor						
% worked last week	43.56	33.41	10.16 **	79.29	79.08	0.22
% worked for payment	40.21	30.15	10.05 **	76.16	75.11	1.05
% peasant in own plot	0.09	0.41	-0.32	1.82	7.86	-6.04 **
% unpaid	3.26	3.21	0.06	2.83	3.73	-0.91
% non agricultural employee	29.38	19.64	9.74 **	54.14	41.29	12.85 **
% agricultural employee	0.93	1.09	-0.16	4.65	10.56	-5.92 **
% employer	1.59	1.90	-0.31	2.93	3.47	-0.54
% self-employed	8.21	7.11	1.10	12.63	11.92	0.70

Continue on next page

1. Earnings in pesos of 2002 (9.56 pesos/US dollars)

	Female			Male		
Variables measured in 2002	Lost	Found	Difference (Lost-Found)	Lost	Found	Difference (Lost-Found)
Earnings Last 12 Months[1]	10,428	6,952	3,476 **	37,800	23,970	13,829 **
Earnings Last 12 Months - Self Employed	24,307	16,263	8,045 +	55,060	30,604	24,456 **
Earnings Last 12 Months -Employee	32,444	27,828	4,616	52,544	33,305	19,240 **
Earnings Last 12 Months - Excluding Unpaid	12,512	7,488	5,024 **	41,195	25,447	15,748 **
# hours worked regular week	40.41	37.86	2.55 *	48.66	46.69	1.97 **
Locality characteristics						
Rural	16.30	42.30	-26.00 **	16.51	43.39	-26.88 **
Integration to community						
% know person/institution to borrow	32.22	29.34	2.87 +	43.80	39.33	4.47 *
% received income from Progresa	5.76	17.69	-11.93 **	4.92	17.19	-12.27 **
% received income from Procampo	2.80	10.98	-8.18 **	2.41	11.60	-9.18 **
% activities outside HH	14.23	9.60	4.62 **	29.15	20.31	8.84 **
% take care of elder or sick people	44.98	42.15	2.83 +	17.96	14.21	3.75 **
Migration						
% moved by age 12	27.26	23.17	4.09 **	28.41	23.82	4.59 **
% thought about migrating	23.58	15.48	8.10 **	28.41	17.14	11.27 **
Quality of life/Expectations						
% life has improved	38.91	32.93	5.97 **	41.08	30.52	10.56 **
% life has remained the same	53.13	58.13	-4.99 **	51.05	61.36	-10.32 **
% life has worsen	7.96	8.94	-0.98	7.87	8.12	-0.25
% a lot of fear of being assualted during day tim	11.54	9.20	2.34 *	4.31	2.63	1.67 **
% a lot of fear of being assualted during night tir	16.62	14.34	2.28 +	7.01	4.67	2.34 **
% positive probability of being victim of assualt	24.98	18.86	6.12 **	23.86	18.69	5.18 **
% has been a victim of assaults	13.63	7.86	5.77 **	21.89	15.09	6.80 **
% in neighborhood: gangs	30.42	26.49	3.93 **	27.88	25.52	2.36 +
% in neighborhood: cooperative neighbors	10.08	10.25	-0.16	11.53	10.61	0.91
% in neighborhood: paramilitaries	5.59	4.75	0.84	5.27	4.81	0.47
% in neighborhood: militaries	5.25	5.71	-0.45	5.90	5.91	-0.01
% in neighborhood: neighbors with guns	3.14	5.56	-2.42 **	3.75	5.21	-1.46 *
% in neighborhood: robbery	21.95	18.54	3.41 **	21.09	17.90	3.19 **
% in neighborhood: insecurity	10.41	7.15	3.26 **	9.25	7.15	2.10 *
Emotional Status	8.12	8.86	-0.74 *	4.74	5.06	-0.32
Observations	1,353	10,942		1,262	9,710	

** p<0.01, * p<0.05, + p<0.1

1. Earnings in pesos of 2002 (9.56 pesos/US dollars)

Table 4: Probability of Attrition - Basic Logit Model Age15+

	Basic Model				
Variables measured in 2002	All	Female	Male	Rural	Urban
Basic demographics					
(1) Age>15	0.990+	0.984*	0.996	0.986	0.99
	[0.006]	[0.008]	[0.007]	[0.012]	[0.007]
(1) Age>36	0.99	0.998	0.978+	0.976	0.993
	[0.011]	[0.015]	[0.012]	[0.023]	[0.012]
(1) Age>58	0.999	0.992	1.005	0.999	1.004
	[0.013]	[0.018]	[0.020]	[0.039]	[0.015]
(1) Age>80	1.108**	1.099*	1.130*	1.117	1.110**
	[0.036]	[0.045]	[0.069]	[0.089]	[0.045]
(1) Female	1.143*			1.252	1.113+
	[0.062]			[0.199]	[0.066]
Respondent's Education					
(1) Primary complete[1]	0.955	0.985	0.909	0.799	0.962
	[0.090]	[0.127]	[0.114]	[0.123]	[0.109]
(1) High school incomplete	1.039	0.982	1.072	0.772	1.094
	[0.099]	[0.148]	[0.118]	[0.124]	[0.124]
(1) High school complete	1.275+	1.008	1.540**	0.951	1.329*
	[0.164]	[0.165]	[0.244]	[0.279]	[0.188]
(1) Some college or more	1.344**	1.176	1.456*	1.152	1.395**
	[0.148]	[0.187]	[0.236]	[0.347]	[0.172]
Father's Education					
(1) Primary complete[1]	1.086	0.984	1.218+	1.306	1.051
	[0.101]	[0.123]	[0.134]	[0.220]	[0.120]
(1) High school incomplete	1.320**	1.258+	1.391*	0.963	1.401**
	[0.133]	[0.170]	[0.190]	[0.290]	[0.156]
(1) High school complete	1.587**	1.334	1.846**	3.365**	1.510*
	[0.283]	[0.329]	[0.363]	[1.166]	[0.293]
(1) Some college or more	1.703**	1.919**	1.478+	1.214	1.767**
	[0.242]	[0.317]	[0.309]	[0.565]	[0.267]
Zscore Raven's test	0.957	0.998	0.913+	0.914	0.966
	[0.034]	[0.045]	[0.045]	[0.071]	[0.039]
Height (cm)	1.008*	1.010+	1.001	1.012	1.007
	[0.004]	[0.006]	[0.006]	[0.010]	[0.004]
(1) Married	1.049	0.995	1.2	1.611*	0.917
	[0.096]	[0.110]	[0.159]	[0.337]	[0.094]
Household characteristics					
(1) Own farm business	0.811	0.723+	0.921	0.666+	0.999
	[0.139]	[0.132]	[0.187]	[0.149]	[0.246]
(1) Own house	0.442**	0.432**	0.453**	0.745+	0.382**
	[0.039]	[0.043]	[0.045]	[0.114]	[0.039]
(1) Own non-farm business	0.853	0.862	0.834	0.903	0.847
	[0.097]	[0.108]	[0.101]	[0.213]	[0.107]
(1) Quartile 2 PCE[2]	0.815	0.831	0.806	0.756	0.848
	[0.117]	[0.141]	[0.115]	[0.152]	[0.169]
(1) Quartile 3 PCE	0.874	0.899	0.859	0.534*	0.999
	[0.114]	[0.137]	[0.119]	[0.142]	[0.168]
(1) Quartile 4 PCE	1.202	1.205	1.207	0.735	1.36
	[0.189]	[0.230]	[0.197]	[0.230]	[0.269]

Continue on next page

	Basic Model				
Variables measured in 2002	All	Female	Male	Rural	Urban
(1) Quartile 2 pc wealth[2]	0.697**	0.687**	0.698**	0.946	0.638**
	[0.086]	[0.098]	[0.090]	[0.199]	[0.099]
(1) Quartile 3 pc wealth	0.578**	0.560**	0.587**	0.696	0.561**
	[0.079]	[0.091]	[0.081]	[0.176]	[0.087]
(1) Quartile 4 pc wealth	0.726*	0.745*	0.693**	0.736	0.738*
	[0.094]	[0.111]	[0.095]	[0.225]	[0.114]
Household size	0.991	1.02	0.962	0.959	1.006
	[0.026]	[0.030]	[0.027]	[0.041]	[0.031]
Household composition					
Number of children in the same dwelling	0.885**	0.869**	0.898*	0.715**	0.938
	[0.034]	[0.037]	[0.042]	[0.051]	[0.041]
(1) At least one parent coresident	0.707**	0.606**	0.795*	0.783	0.677**
	[0.055]	[0.070]	[0.092]	[0.145]	[0.062]
(1) Both parents dead	0.967	0.938	0.991	1.097	0.963
	[0.085]	[0.129]	[0.122]	[0.244]	[0.091]
Health					
(1) Relative GHS better	0.942	0.95	0.929	0.93	0.944
	[0.058]	[0.081]	[0.074]	[0.123]	[0.064]
(1) Relative GHS worse	0.845	0.759+	1.032	0.621	0.915
	[0.098]	[0.111]	[0.231]	[0.183]	[0.116]
(1) Hypertension	0.956	0.944	0.941	1.025	0.953
	[0.065]	[0.086]	[0.108]	[0.146]	[0.077]
Labor market					
(1) Work for pay	1.221	1.002	1.208	1.505	1.176
	[0.199]	[0.301]	[0.274]	[0.546]	[0.219]
(1) Quartile 2 earnings last month[2]	1.559	0.698	2.574*	1.463	1.578
	[0.489]	[0.379]	[1.110]	[0.789]	[0.601]
(1) Quartile 3 earnings last month	0.759+	1.137	0.583*	0.717	0.78
	[0.119]	[0.310]	[0.128]	[0.183]	[0.149]
(1) Quartile 4 earnings last month	0.852	1.273	0.667+	0.96	0.82
	[0.140]	[0.355]	[0.146]	[0.288]	[0.163]
Hours worked/week	1.004+	1.004	1.004+	1.007	1.003
	[0.002]	[0.003]	[0.003]	[0.004]	[0.003]
Locality Characteristics					
Rural	0.426**	0.325**	0.545*		
	[0.075]	[0.040]	[0.148]		
Constant	0.032**	0.046**	0.070*	0.005**	0.022**
	[0.021]	[0.037]	[0.077]	[0.008]	[0.014]
Number of observations	22,829	11,825	10,361	8,866	13,963
ll	-6,722	-3,449	-3,199	-1,460	-5,192
R-squared	0.171	0.177	0.166	0.147	0.141

1. Omitted category: Primay incomplete

2. Omitted category: Quartile 1

Municipality Fixed Effects

Standard errors in brackets (adjusted for clusters at municipality level)

** p<0.01, * p<0.05, + p<0.1

Table 5: Probability of Attrition - Extended Logit Model Age15+

	Extended Model				
Variables measured in 2002	All	Female	Male	Rural	Urban
Basic demographics					
(1) Age>15	0.990+	0.984*	0.996	0.987	0.99
	[0.006]	[0.008]	[0.007]	[0.012]	[0.007]
(1) Age>36	0.99	0.998	0.979+	0.975	0.993
	[0.011]	[0.015]	[0.012]	[0.023]	[0.012]
(1) Age>58	0.999	0.991	1.006	1.001	1.004
	[0.014]	[0.018]	[0.020]	[0.038]	[0.015]
(1) Age>80	1.107**	1.103*	1.124+	1.109	1.110*
	[0.036]	[0.045]	[0.069]	[0.088]	[0.045]
(1) Female	1.161**			1.289	1.125+
	[0.067]			[0.212]	[0.071]
Respondent's Education					
(1) Primary complete[1]	0.949	0.977	0.907	0.801	0.958
	[0.089]	[0.125]	[0.111]	[0.124]	[0.108]
(1) High school incomplete	1.035	0.98	1.067	0.77	1.1
	[0.098]	[0.147]	[0.119]	[0.125]	[0.124]
(1) High school complete	1.269+	1.003	1.551**	0.946	1.333*
	[0.163]	[0.166]	[0.250]	[0.289]	[0.189]
(1) Some college or more	1.307*	1.147	1.411*	1.125	1.363*
	[0.147]	[0.186]	[0.232]	[0.346]	[0.172]
Father's Education					
(1) Primary complete[1]	1.079	0.976	1.215+	1.297	1.044
	[0.099]	[0.123]	[0.133]	[0.216]	[0.117]
(1) High school incomplete	1.303**	1.254	1.385*	0.945	1.389**
	[0.131]	[0.175]	[0.186]	[0.296]	[0.154]
(1) High school complete	1.569*	1.308	1.853**	3.276**	1.489*
	[0.281]	[0.322]	[0.361]	[1.164]	[0.292]
(1) Some college or more	1.659**	1.883**	1.444+	1.138	1.719**
	[0.236]	[0.319]	[0.303]	[0.563]	[0.259]
Zscore Raven's test	0.953	0.993	0.911+	0.909	0.966
	[0.034]	[0.045]	[0.044]	[0.072]	[0.040]
Height (cm)	1.008+	1.010+	1.002	1.012	1.007
	[0.004]	[0.006]	[0.006]	[0.010]	[0.004]
(1) Married	1.059	1.005	1.207	1.613*	0.93
	[0.098]	[0.110]	[0.161]	[0.339]	[0.096]
Household characteristics					
(1) Own farm business	0.809	0.724+	0.917	0.658+	0.998
	[0.139]	[0.131]	[0.187]	[0.146]	[0.244]
(1) Own house	0.443**	0.434**	0.454**	0.757+	0.382**
	[0.039]	[0.043]	[0.045]	[0.118]	[0.038]
(1) Own non-farm business	0.862	0.87	0.842	0.892	0.86
	[0.098]	[0.109]	[0.102]	[0.208]	[0.108]
(1) Quartile 2 PCE[2]	0.817	0.837	0.811	0.745	0.855
	[0.118]	[0.143]	[0.116]	[0.151]	[0.170]
(1) Quartile 3 PCE	0.879	0.912	0.861	0.524*	1.018
	[0.114]	[0.141]	[0.118]	[0.139]	[0.170]
(1) Quartile 4 PCE	1.205	1.216	1.211	0.724	1.381
	[0.190]	[0.233]	[0.197]	[0.227]	[0.272]
(1) Quartile 2 pc wealth[2]	0.700**	0.686**	0.703**	0.939	0.646**
	[0.085]	[0.097]	[0.089]	[0.201]	[0.099]
(1) Quartile 3 pc wealth	0.583**	0.562**	0.594**	0.697	0.567**
	[0.079]	[0.090]	[0.081]	[0.177]	[0.087]
(1) Quartile 4 pc wealth	0.726*	0.739*	0.696**	0.724	0.737*
	[0.093]	[0.110]	[0.094]	[0.220]	[0.113]
Household size	0.992	1.021	0.963	0.959	1.008
	[0.026]	[0.030]	[0.028]	[0.041]	[0.031]

Continue on next page

	Extended Model				
Variables measured in 2002	All	Female	Male	Rural	Urban
Household composition					
Number of children in the same dwelling	0.886**	0.872**	0.899*	0.712**	0.94
	[0.034]	[0.037]	[0.043]	[0.052]	[0.041]
(1) At least one parent coresident	0.700**	0.602**	0.785*	0.781	0.666**
	[0.054]	[0.070]	[0.092]	[0.142]	[0.061]
(1) Both parents dead	0.965	0.928	1.001	1.104	0.957
	[0.086]	[0.130]	[0.123]	[0.247]	[0.092]
Health					
(1) Relative GHS better	0.94	0.948	0.933	0.925	0.945
	[0.057]	[0.078]	[0.073]	[0.121]	[0.062]
(1) Relative GHS worse	0.856	0.763+	1.041	0.602+	0.938
	[0.096]	[0.110]	[0.226]	[0.182]	[0.113]
(1) Hypertension	0.959	0.949	0.943	1.034	0.958
	[0.065]	[0.086]	[0.108]	[0.147]	[0.076]
Labor market					
(1) Work for pay	1.212	0.983	1.217	1.497	1.17
	[0.198]	[0.297]	[0.276]	[0.553]	[0.218]
(1) Quartile 2 earnings last month	1.516	0.668	2.574*	1.487	1.516
	[0.477]	[0.360]	[1.118]	[0.823]	[0.576]
(1) Quartile 3 earnings last month	0.771+	1.158	0.586*	0.71	0.8
	[0.121]	[0.313]	[0.129]	[0.187]	[0.151]
(1) Quartile 4 earnings last month	0.862	1.305	0.664+	0.952	0.832
	[0.142]	[0.362]	[0.145]	[0.289]	[0.165]
Hours worked/week	1.004+	1.004	1.004+	1.006	1.003
	[0.002]	[0.003]	[0.003]	[0.004]	[0.003]
Migration					
(1) Moved before age 12	1.07	1.091	1.054	1.094	1.047
	[0.077]	[0.087]	[0.110]	[0.163]	[0.087]
(1) Thought about moving in the future	1.277**	1.264*	1.304**	1.084	1.337**
	[0.117]	[0.148]	[0.129]	[0.179]	[0.145]
Integration to community					
(1) Know person/institution to borrow	0.966		0.902	1.464	0.886
	[0.060]		[0.111]	[0.359]	[0.084]
(1) Activities outside the household	1.066	1.029		1.187	
	[0.073]	[0.118]		[0.217]	
Subjective quality of life					
(1) Feel unsafe in neighborhood	0.988	1.083	0.9	0.944	0.964
	[0.137]	[0.166]	[0.142]	[0.427]	[0.141]
(1) Quartile 2 emotional status	1.001	0.986	1.031	1.109	0.972
	[0.072]	[0.098]	[0.097]	[0.181]	[0.081]
(1) Quartile 3 emotional status	0.929	0.938	0.919	0.863	0.95
	[0.077]	[0.123]	[0.096]	[0.138]	[0.097]
(1) Quartile 4 emotional status	0.974	0.951	1.07	1.037	0.936
	[0.085]	[0.109]	[0.146]	[0.203]	[0.097]
Locality Characteristics					
(1) Rural	0.423**	0.319**	0.544*		
	[0.075]	[0.039]	[0.151]		
Constant	0.032**	0.045**	0.064*	0.004**	0.024**
	[0.021]	[0.036]	[0.073]	[0.007]	[0.015]
Number of observations	22829	11825	10361	8866	13963
Chi-Squared	-6705	-3436	-3190	-1455	-5174
R-squared	0.173	0.18	0.168	0.15	0.144

1. Omitted category: Primay incomplete

2. Omitted category: Quartile 1

Municipality Fixed Effects

Standard errors in brackets (adjusted for clusters at municipality level)

** p<0.01, * p<0.05, + p<0.1

Table 6: Quartic-root of Earnings in 2002 with Interactions - Age [21-65]

	All		Females		Males		Rural		Urban	
Variables measured in 2002	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Age>21	0.033**	0.032**	0.045**	0.041*	0.027*	0.027*	0.026	0.027	0.037**	0.036**
	[0.008]	[0.009]	[0.015]	[0.016]	[0.010]	[0.011]	[0.016]	[0.017]	[0.010]	[0.010]
Age>32	-0.018	-0.018	-0.009	0.003	-0.016	-0.021	-0.026	-0.027	-0.012	-0.009
	[0.014]	[0.015]	[0.026]	[0.027]	[0.015]	[0.017]	[0.024]	[0.026]	[0.016]	[0.018]
Age>43	-0.035*	-0.038*	-0.052+	-0.069*	-0.030+	-0.026	-0.018	-0.015	-0.041*	-0.049**
	[0.015]	[0.016]	[0.027]	[0.028]	[0.018]	[0.019]	[0.029]	[0.030]	[0.016]	[0.018]
Age>54	-0.048*	-0.044+	-0.034	-0.025	-0.056*	-0.053+	-0.033	-0.037	-0.067**	-0.058+
	[0.022]	[0.024]	[0.046]	[0.050]	[0.026]	[0.029]	[0.040]	[0.040]	[0.025]	[0.031]
Female	-0.929**	-0.959**					-0.865**	-0.890**	-0.970**	-1.010**
	[0.058]	[0.061]					[0.097]	[0.100]	[0.072]	[0.077]
Education										
(1) Primary complete [OG (Primary incomplete)]	-0.201	-0.214	0.195	0.186	-0.229	-0.225	0.07	0.112	-0.342+	-0.378*
	[0.179]	[0.173]	[0.131]	[0.131]	[0.143]	[0.137]	[0.106]	[0.114]	[0.196]	[0.179]
(1) High school incomplete	0.111	0.079	0.685**	0.707**	0.046	0.013	0.360*	0.362*	0.045	0.013
	[0.201]	[0.202]	[0.122]	[0.133]	[0.162]	[0.163]	[0.141]	[0.148]	[0.201]	[0.196]
(1) High school complete	0.209	0.189	0.943**	1.002**	0.108	0.081	0.563*	0.706**	0.177	0.139
	[0.267]	[0.248]	[0.151]	[0.171]	[0.244]	[0.212]	[0.229]	[0.240]	[0.244]	[0.206]
(1) Some college or more	1.216**	1.181**	2.050**	2.086**	1.001**	0.960**	1.461**	1.473**	1.171**	1.138**
	[0.230]	[0.237]	[0.138]	[0.153]	[0.190]	[0.194]	[0.242]	[0.246]	[0.222]	[0.223]
Married	0.092	0.097	-0.333**	-0.322**	0.385**	0.398**	0.166	0.146	0.058	0.072
	[0.067]	[0.067]	[0.114]	[0.118]	[0.079]	[0.081]	[0.113]	[0.115]	[0.077]	[0.075]
Rural	-0.955**	-0.953**	-0.841*	-0.824*	-0.997**	-1.004**				
	[0.269]	[0.267]	[0.349]	[0.360]	[0.260]	[0.258]				
Lost		-0.884				-0.763				-1.135
		[0.809]				[0.861]				[0.910]
Age>21*Lost		0.008		0.039		-0.003		-0.006		0.009
		[0.025]		[0.042]		[0.029]		[0.061]		[0.027]
Age>32*Lost		0		-0.094		0.038		0.011		-0.01
		[0.040]		[0.080]		[0.051]		[0.102]		[0.046]
Age>43*Lost		0.032		0.101		-0.017		-0.065		0.06
		[0.054]		[0.083]		[0.074]		[0.147]		[0.061]
Age>54*Lost		-0.05		0.159		-0.043		0.104		-0.062
		[0.083]		[0.182]		[0.101]		[0.222]		[0.092]
Female*Lost		**0.268+**		-0.436				0.471		**0.294+**
		[0.145]		[1.244]				[0.399]		[0.154]
Education*Lost										
(1) Primary complete*Lost [OG (Primary incomplete)]		0.476		0.058		0.46		**-0.875+**		**1.010****
		[0.323]		[0.482]		[0.305]		[0.463]		[0.356]
(1) High school incomplete*Lost		**0.660***		-0.221		**0.822***		-0.09		**0.953***
		[0.332]		[0.445]		[0.332]		[0.405]		[0.366]
(1) High school complete*Lost		0.621		-0.507		**0.904***		**-1.946****		**1.051****
		[0.384]		[0.591]		[0.404]		[0.702]		[0.366]
(1) Some college or more*Lost		**0.643+**		-0.391		**0.875***		-0.173		**0.923***
		[0.361]		[0.379]		[0.384]		[0.564]		[0.383]
Married*Lost		-0.04		-0.16		-0.08		0.355		-0.116
		[0.190]		[0.246]		[0.238]		[0.423]		[0.209]
Rural*Lost		0.194		0.033		0.284		0.328		
		[0.206]		[0.377]		[0.232]		[1.512]		
Constant	6.258**	6.310**	3.961**	4.010**	5.670**	5.681**	4.352**	4.325**	6.190**	6.215**
	[0.294]	[0.317]	[0.531]	[0.564]	[0.389]	[0.411]	[0.446]	[0.482]	[0.304]	[0.331]
Observations	9,328	9,328	3,011	3,011	6,317	6,317	3,411	3,411	5,917	5,917
R-squared	24.47	24.47	27.7	27.7	24.39	24.39	20.16	20.16	19.97	19.97
F-statistic	60.07	33.18	30.59	21.06	30.00	18.49	13.49	10.58	58.92	34.24

Municipality Fixed Effects

Standard errors in brackets (adjusted for clusters at municipality level)

** p<0.01, * p<0.05, + p<0.1

Table 7: Original and Imputed Models - Quartic-root of Earnings in 2002- Age [21-65]

	All			Females			Males			Rural			Urban		
Variables measured in 2002	Original	Basic	Extended	Original	Basic	Extended	Original	Basic	Extended	Original	Basic	Extended	Original	Basic	Extended
Age>21	0.033**	0.033**	0.033**	0.045**	0.046**	0.046**	0.027*	0.027*	0.027*	0.026	0.028+	0.028+	0.037**	0.037**	0.037**
	[0.008]	[0.009]	[0.008]	[0.015]	[0.016]	[0.016]	[0.010]	[0.011]	[0.011]	[0.016]	[0.016]	[0.016]	[0.010]	[0.010]	[0.010]
Age>32	-0.018	-0.023	-0.023+	-0.009	-0.013	-0.011	-0.016	-0.023	-0.023	-0.026	-0.028	-0.027	-0.012	-0.019	-0.018
	[0.014]	[0.014]	[0.014]	[0.026]	[0.025]	[0.025]	[0.015]	[0.018]	[0.017]	[0.024]	[0.026]	[0.026]	[0.016]	[0.018]	[0.017]
Age>43	-0.035*	-0.032*	-0.033*	-0.052+	-0.053+	-0.057+	-0.030+	-0.023	-0.023	-0.018	-0.014	-0.014	-0.041*	-0.039*	-0.042*
	[0.015]	[0.015]	[0.015]	[0.027]	[0.031]	[0.033]	[0.018]	[0.019]	[0.018]	[0.029]	[0.029]	[0.029]	[0.016]	[0.018]	[0.019]
Age>54	-0.048*	-0.044*	-0.042+	-0.034	-0.034	-0.031	-0.056*	-0.051+	-0.050+	-0.033	-0.041	-0.042	-0.067**	-0.053+	-0.049
	[0.022]	[0.022]	[0.023]	[0.046]	[0.051]	[0.053]	[0.026]	[0.027]	[0.026]	[0.040]	[0.038]	[0.038]	[0.025]	[0.030]	[0.033]
Female	-0.929**	-0.963**	-0.964**							-0.865**	-0.918**	-0.915**	-0.970**	-0.994**	-0.997**
	[0.058]	[0.058]	[0.059]							[0.097]	[0.100]	[0.100]	[0.072]	[0.070]	[0.072]
Education															
(1) Primary complete [OG (Primary incompleted)]	-0.201	-0.207	-0.209	0.195	0.189	0.193	-0.229	-0.228	-0.229	0.07	0.103	0.103	-0.342+	-0.358+	-0.359+
	[0.179]	[0.176]	[0.177]	[0.131]	[0.123]	[0.127]	[0.143]	[0.144]	[0.145]	[0.106]	[0.117]	[0.117]	[0.196]	[0.186]	[0.187]
(1) High school incomplete	0.111	0.092	0.089	0.685**	0.700**	0.701**	0.046	0.019	0.016	0.360*	0.360*	0.362*	0.045	0.026	0.021
	[0.201]	[0.207]	[0.209]	[0.122]	[0.127]	[0.129]	[0.162]	[0.173]	[0.173]	[0.141]	[0.147]	[0.147]	[0.201]	[0.207]	[0.209]
(1) High school complete	0.209	0.228	0.241	0.943**	0.959**	0.982**	0.108	0.124	0.135	0.563*	0.628**	0.646**	0.177	0.19	0.202
	[0.267]	[0.248]	[0.249]	[0.151]	[0.161]	[0.170]	[0.244]	[0.219]	[0.218]	[0.229]	[0.234]	[0.237]	[0.244]	[0.218]	[0.218]
(1) Some college or more	1.216**	1.185**	1.187**	2.050**	2.005**	2.021**	1.001**	0.983**	0.980**	1.461**	1.465**	1.458**	1.171**	1.139**	1.142**
	[0.230]	[0.237]	[0.236]	[0.138]	[0.162]	[0.160]	[0.190]	[0.202]	[0.201]	[0.242]	[0.244]	[0.244]	[0.222]	[0.229]	[0.227]
Married	0.092	0.113	0.104	-0.333**	-0.264*	-0.278*	0.385**	0.376**	0.371**	0.166	0.101	0.098	0.058	0.125	0.113
	[0.067]	[0.078]	[0.072]	[0.114]	[0.128]	[0.127]	[0.079]	[0.089]	[0.082]	[0.113]	[0.110]	[0.108]	[0.077]	[0.090]	[0.083]
Rural	-0.955**	-0.977**	-0.969**	-0.841*	-0.981**	-0.961*	-0.997**	-0.966**	-0.964**						
	[0.269]	[0.254]	[0.254]	[0.349]	[0.372]	[0.369]	[0.260]	[0.233]	[0.231]						
Constant	6.258**	5.637**	5.639**	3.961**	3.985**	3.980**	5.670**	5.651**	5.657**	4.352**	4.733**	4.742**	6.190**	6.112**	6.178**
	[0.294]	[0.375]	[0.365]	[0.531]	[0.554]	[0.551]	[0.389]	[0.406]	[0.397]	[0.446]	[0.444]	[0.440]	[0.304]	[0.474]	[0.469]
Observations	9,328	9,328	9328	3,011	3,011	3011	6,317	6,317	6317	3,411	3,411	3411	5,917	5,917	5917
R-squared	24.47	25.22	25.43	27.70	28.07	28.42	24.39	24.74	24.93	20.16	20.67	20.69	19.97	20.89	21.20
F-statistic	60.07	56.5	57.4	30.59	28.07	29.37	30.00	20.92	23.36	13.49	13.98	13.9	58.92	58.66	60.23

Municipality Fixed Effects

Standard errors in brackets (adjusted for clusters at municipality level)

** p<0.01, * p<0.05, + p<0.1

F-statistic for hypothesis that all coefficients except the intercept are equal to zero

Table 8: Original and Imputed Models with Interactions - Quartic-root of Earnings in 2002- Age [21-65]

	All			Females			Males			Rural			Urban		
Variables measured in 2002	Original	Basic	Extended	Original	Basic	Extended	Original	Basic	Extended	Original	Basic	Extended	Original	Basic	Extended
Age>21	0.032**	0.031**	0.031**	0.041*	0.043*	0.042*	0.027*	0.025*	0.026*	0.027	0.025	0.025	0.036**	0.035**	0.035**
	[0.009]	[0.009]	[0.009]	[0.016]	[0.017]	[0.017]	[0.011]	[0.011]	[0.011]	[0.017]	[0.017]	[0.017]	[0.010]	[0.010]	[0.010]
Age>32	-0.018	-0.017	-0.018	0.003	-0.003	-0.002	-0.021	-0.019	-0.02	-0.027	-0.025	-0.025	-0.009	-0.01	-0.01
	[0.015]	[0.015]	[0.015]	[0.027]	[0.027]	[0.027]	[0.017]	[0.017]	[0.017]	[0.026]	[0.027]	[0.027]	[0.018]	[0.018]	[0.018]
Age>43	-0.038*	-0.037*	-0.038*	-0.069*	-0.062*	-0.063*	-0.026	-0.027	-0.027	-0.015	-0.015	-0.015	-0.049**	-0.048**	-0.049**
	[0.016]	[0.016]	[0.016]	[0.028]	[0.028]	[0.028]	[0.019]	[0.018]	[0.018]	[0.030]	[0.030]	[0.030]	[0.018]	[0.018]	[0.018]
Age>54	-0.044+	-0.043+	-0.042+	-0.025	-0.036	-0.035	-0.053+	-0.048+	-0.048+	-0.037	-0.039	-0.039	-0.058+	-0.052+	-0.051
	[0.024]	[0.024]	[0.024]	[0.050]	[0.048]	[0.048]	[0.029]	[0.028]	[0.028]	[0.040]	[0.040]	[0.040]	[0.031]	[0.031]	[0.031]
Female	-0.959**	-0.956**	-0.956**							-0.890**	-0.899**	-0.900**	-1.010**	-0.998**	-0.997**
	[0.061]	[0.058]	[0.058]							[0.100]	[0.100]	[0.100]	[0.077]	[0.073]	[0.074]
Education															
(1) Primary complete OG (Primary incompleted)	-0.214	-0.219	-0.218	0.186	0.168	0.167	-0.225	-0.224	-0.223	0.112	0.099	0.1	-0.378*	-0.381*	-0.380*
	[0.173]	[0.169]	[0.169]	[0.131]	[0.128]	[0.129]	[0.137]	[0.136]	[0.136]	[0.114]	[0.114]	[0.113]	[0.179]	[0.176]	[0.176]
(1) High school incomplete	0.079	0.082	0.083	0.707**	0.691**	0.689**	0.013	0.022	0.023	0.362*	0.351*	0.351*	0.013	0.019	0.021
	[0.202]	[0.199]	[0.199]	[0.133]	[0.131]	[0.131]	[0.163]	[0.164]	[0.164]	[0.148]	[0.151]	[0.151]	[0.196]	[0.196]	[0.196]
(1) High school complete	0.189	0.171	0.171	1.002**	0.939**	0.938**	0.081	0.066	0.065	0.706**	0.618*	0.617*	0.139	0.135	0.134
	[0.248]	[0.241]	[0.241]	[0.171]	[0.169]	[0.169]	[0.212]	[0.210]	[0.210]	[0.240]	[0.244]	[0.244]	[0.206]	[0.203]	[0.203]
(1) Some college or more	1.181**	1.166**	1.164**	2.086**	2.037**	2.029**	0.960**	0.953**	0.951**	1.473**	1.454**	1.455**	1.138**	1.123**	1.119**
	[0.237]	[0.233]	[0.232]	[0.153]	[0.151]	[0.151]	[0.194]	[0.193]	[0.193]	[0.246]	[0.250]	[0.250]	[0.223]	[0.220]	[0.220]
Married	0.097	0.092	0.091	-0.322**	-0.319**	-0.319**	0.398**	0.385**	0.383**	0.146	0.109	0.105	0.072	0.088	0.087
	[0.067]	[0.062]	[0.062]	[0.118]	[0.117]	[0.117]	[0.081]	[0.074]	[0.075]	[0.115]	[0.109]	[0.108]	[0.075]	[0.074]	[0.075]
Rural	-0.953**	-0.978**	-0.970**	-0.824*	-0.972*	-0.959*	-1.004**	-0.977**	-0.972**						
	[0.267]	[0.254]	[0.254]	[0.360]	[0.377]	[0.373]	[0.258]	[0.237]	[0.236]						

Continue on next page

	All			Females			Males			Rural			Urban		
Variables measured in 2002	Original	Basic	Extended	Original	Basic	Extended	Original	Basic	Extended	Original	Basic	Extended	Original	Basic	Extended
Lost	-0.884	-1.116	-1.06				-0.763	-1	-0.892		-0.941	-0.833	-1.135	-1.015	-0.977
	[0.809]	[1.084]	[0.850]				[0.861]	[1.156]	[0.972]		[2.125]	[2.156]	[0.910]	[1.126]	[0.860]
Age>21*Lost	0.008	0.021	0.02	0.039	0.027	0.028	-0.003	0.02	0.018	-0.006	0.036	0.031	0.009	0.014	0.013
	[0.025]	[0.039]	[0.033]	[0.042]	[0.064]	[0.059]	[0.029]	[0.043]	[0.039]	[0.061]	[0.088]	[0.086]	[0.027]	[0.038]	[0.032]
Age>32*Lost	0	-0.042	-0.036	-0.094	-0.076	-0.069	0.038	-0.034	-0.028	0.011	-0.032	-0.021	-0.01	-0.048	-0.042
	[0.040]	[0.064]	[0.050]	[0.080]	[0.092]	[0.089]	[0.051]	[0.077]	[0.064]	[0.102]	[0.148]	[0.140]	[0.046]	[0.063]	[0.051]
Age>43*Lost	0.032	0.048	0.037	0.101	0.053	0.037	-0.017	0.044	0.037	-0.065	0.017	0.013	0.06	0.059	0.045
	[0.054]	[0.061]	[0.061]	[0.083]	[0.145]	[0.166]	[0.074]	[0.082]	[0.064]	[0.147]	[0.160]	[0.147]	[0.061]	[0.065]	[0.074]
Age>54*Lost	-0.05	-0.014	-0.003	0.159	0.104	0.119	-0.043	-0.033	-0.027	0.104	-0.024	-0.043	-0.062	0.001	0.019
	[0.083]	[0.094]	[0.095]	[0.182]	[0.350]	[0.385]	[0.101]	[0.127]	[0.106]	[0.222]	[0.339]	[0.307]	[0.092]	[0.095]	[0.124]
Female*Lost	**0.268+**	-0.036	-0.042	-0.436	-0.867	-0.937				0.471	-0.302	-0.23	**0.294+**	0.044	0.025
	[0.145]	[0.207]	[0.216]	[1.244]	[1.501]	[1.333]				[0.399]	[0.468]	[0.463]	[0.154]	[0.226]	[0.237]
Education*Lost															
(1) Primary complete*Lost OG (Primary incompleted)	0.476	0.47	0.496	0.058	0.218	0.27	0.46	0.402	0.41	**-0.875+**	0.126	0.123	**1.010****	**0.689+**	**0.730***
	[0.323]	[0.338]	[0.348]	[0.482]	[0.619]	[0.751]	[0.305]	[0.553]	[0.494]	[0.463]	[0.641]	[0.666]	[0.356]	[0.372]	[0.351]
(1) High school incomplete*Lost	**0.660***	**0.454+**	**0.463+**	-0.221	0.003	0.049	**0.822***	0.454	0.44	-0.09	0.173	0.21	**0.953***	**0.563+**	**0.574+**
	[0.332]	[0.265]	[0.275]	[0.445]	[0.627]	[0.685]	[0.332]	[0.356]	[0.322]	[0.405]	[0.500]	[0.513]	[0.366]	[0.313]	[0.317]
(1) High school complete*Lost	0.621	0.79	**0.919+**	-0.507	0.071	0.276	**0.904***	0.891	0.975	**-1.946****	0.182	0.441	**1.051****	**0.893***	**1.018***
	[0.384]	[0.479]	[0.471]	[0.591]	[0.710]	[0.823]	[0.404]	[0.708]	[0.640]	[0.702]	[1.020]	[1.049]	[0.366]	[0.400]	[0.397]
(1) Some college or more*Lost	**0.643+**	0.532	**0.596+**	-0.391	-0.211	-0.041	**0.875***	0.68	**0.686+**	-0.173	0.12	0.048	**0.923***	0.626	**0.709+**
	[0.361]	[0.327]	[0.316]	[0.379]	[0.659]	[0.720]	[0.384]	[0.416]	[0.380]	[0.564]	[0.752]	[0.713]	[0.383]	[0.415]	[0.376]
Married*Lost	-0.04	0.152	0.092	-0.16	0.379	0.265	-0.08	-0.071	-0.094	0.355	-0.074	-0.074	-0.116	0.203	0.133
	[0.190]	[0.408]	[0.340]	[0.246]	[0.474]	[0.468]	[0.238]	[0.508]	[0.419]	[0.423]	[0.701]	[0.652]	[0.209]	[0.419]	[0.349]
Rural*Lost	0.194	0.198	0.219	0.033	-0.178	-0.065	0.284	0.36	0.34	0.328					
	[0.206]	[0.284]	[0.329]	[0.377]	[0.457]	[0.508]	[0.232]	[0.342]	[0.367]	[1.512]					
Constant	6.310**	5.731**	5.718**	4.010**	4.132**	4.128**	5.681**	5.713**	5.700**	4.325**	4.809**	4.806**	6.215**	6.137**	6.180**
	[0.317]	[0.367]	[0.369]	[0.564]	[0.566]	[0.564]	[0.411]	[0.402]	[0.401]	[0.482]	[0.473]	[0.472]	[0.331]	[0.402]	[0.461]
Observations	9328	9328	9328	3011	3011	3011	6317	6317	6317	3411	3411	3411	5917	5917	5917
R-squared	24.47	25.22	25.43	27.70	28.07	28.42	24.39	24.74	24.93	20.16	20.67	20.69	19.97	20.89	21.20
F-statistic	33.18	32.55	32.73	21.06	16.12	16.28	18.49	13.82	14.47	10.58	7.43	7.54	34.24	32.18	32.61

Municipality Fixed Effects

Standard errors in brackets (adjusted for clusters at municipality level)

** p<0.01, * p<0.05, + p<0.1

Table 9: Original and Imputed Models - Quartic-root of Earnings in 2005- Age [21-65]

	All			Female			Male			Rural			Urban		
	Tracked	Imputed	Diff	Tracked	Imputed	Diff	Tracked	Imputed	Diff	Tracked	Imputed	Diff	Tracked	Imputed	Diff
Age>21	0.004	0.007		0.011	0.012		0	0.003		-0.031*	-0.026+		0.026*	0.023*	
	[0.009]	[0.009]		[0.015]	[0.015]		[0.012]	[0.011]		[0.014]	[0.014]		[0.012]	[0.011]	
Age>32	0	-0.004		0.004	0		0	-0.004		0.042+	0.034		-0.026	-0.025	
	[0.016]	[0.016]		[0.029]	[0.028]		[0.021]	[0.019]		[0.024]	[0.023]		[0.021]	[0.020]	
Age>43	-0.061**	-0.062**		-0.062+	-0.062+		-0.062**	-0.062**		-0.068**	-0.062*		-0.057**	-0.062**	
	[0.016]	[0.015]		[0.033]	[0.031]		[0.020]	[0.018]		[0.024]	[0.025]		[0.021]	[0.019]	
Age>54	0.035	0.034		0.041	0.048		0.037	0.034		-0.001	-0.006		0.078*	0.070*	
	[0.024]	[0.023]		[0.047]	[0.049]		[0.030]	[0.029]		[0.034]	[0.035]		[0.035]	[0.033]	
(1) Female	-0.819**	-0.788**								-1.002**	-0.969**		-0.742**	-0.722**	
	[0.060]	[0.059]								[0.111]	[0.106]		[0.066]	[0.065]	
Respondent's Education															
(1) Primary complete [OG (Primary incomplete)]	0.201**	0.139*	*	0.145	0.095		0.205**	0.138+	*	0.144	0.13		0.222**	0.123	*
	[0.061]	[0.061]		[0.137]	[0.139]		[0.069]	[0.071]		[0.090]	[0.092]		[0.078]	[0.075]	
(1) High school incomplete	0.589**	0.511**		0.747**	0.648**	*	0.510**	0.443**	+	0.606**	0.568**	+	0.575**	0.466**	*
	[0.065]	[0.065]		[0.134]	[0.128]		[0.083]	[0.083]		[0.089]	[0.090]		[0.092]	[0.088]	
(1) High school complete	0.980**	0.855**	+	1.298**	1.104**	*	0.782**	0.703**		0.981**	0.910**		0.966**	0.807**	*
	[0.119]	[0.117]		[0.198]	[0.189]		[0.137]	[0.128]		[0.200]	[0.232]		[0.142]	[0.130]	
(1) Some college or more	1.768**	1.692**	+	2.199**	2.041**	+	1.523**	1.489**		1.762**	1.732**		1.771**	1.660**	+
	[0.109]	[0.102]		[0.157]	[0.156]		[0.133]	[0.129]		[0.181]	[0.190]		[0.135]	[0.123]	
(1) Married	0.094+	0.06		-0.352**	-0.305**		0.417**	0.341**		0.179*	0.163+		0.047	0.015	
	[0.056]	[0.054]		[0.124]	[0.109]		[0.072]	[0.069]		[0.081]	[0.084]		[0.075]	[0.069]	
(1) Rural	-0.348+	-0.375*		-0.063	-0.141		-0.428**	-0.446**							
	[0.181]	[0.185]		[0.448]	[0.425]		[0.134]	[0.147]							
Constant	7.281**	7.294**		4.606**	4.696**		6.831**	3.805**		6.714**	7.825**		5.916**	5.493**	
	[0.323]	[0.317]		[0.579]	[0.582]		[0.310]	[0.372]		[0.384]	[0.372]		[0.327]	[0.319]	
Number of observations	7,921	9,264		2,491	2,977		5,430	6,287		3,010	3,192		4,911	6,072	
R-squared	13.75	13.82		18.02	17.37		12.53	12.72		15.99	16.17		9.72	9.94	
F-statistic	45.7	48.45		37.45	32.63		31.07	33.2		32.57	30.04		35.63	38.64	

Sstandard errors in brackets (adjusted for clusters at municipality level)

** p<0.01, * p<0.05, + p<0.1

F. F-statistic for hypothesis that all coefficients are equal to zero

Table 10: Original and Imputed Models - Quartic-root of Earnings in 2009- Age 15+

	All			Female			Male			Rural			Urban		
	Tracked	Imputed	Diff	Tracked	Imputed	Diff	Tracked	Imputed	Diff	Tracked	Imputed	Diff	Tracked	Imputed	Diff
Age>21	-0.018*	-0.016+		-0.034*	-0.032*		-0.012	-0.009		-0.030*	-0.031*		-0.008	-0.003	**
	[0.008]	[0.008]		[0.014]	[0.014]		[0.011]	[0.012]		[0.013]	[0.012]		[0.012]	[0.012]	
Age>32	0.018	0.015		0.056*	0.055*		0	-0.006		0.026	0.029		0.011	0.003	**
	[0.015]	[0.015]		[0.023]	[0.023]		[0.018]	[0.019]		[0.020]	[0.020]		[0.021]	[0.022]	
Age>43	-0.060**	-0.062**		-0.099**	-0.104**	*	-0.040*	-0.040+		-0.061*	-0.065*		-0.056+	-0.057*	
	[0.019]	[0.019]		[0.032]	[0.033]		[0.020]	[0.020]		[0.023]	[0.025]		[0.028]	[0.028]	
Age>54	0.014	0.02		-0.015	-0.008		0.012	0.016		0.013	0.02		0.02	0.026	
	[0.043]	[0.043]		[0.086]	[0.089]		[0.045]	[0.044]		[0.055]	[0.057]		[0.057]	[0.056]	
(1) Female	-1.062**	-1.057**								-1.232**	-1.220**		-0.958**	-0.959**	
	[0.072]	[0.074]								[0.117]	[0.115]		[0.095]	[0.097]	
Respondent's Education															
(1) Primary complete OG (Primary incomplete)	0.313**	0.281**		0.433*	0.420*		0.221*	0.179+		0.257+	0.266+		0.321*	0.242+	**
	[0.097]	[0.095]		[0.181]	[0.181]		[0.104]	[0.104]		[0.134]	[0.134]		[0.130]	[0.126]	
(1) High school incomplete	0.697**	0.643**		0.684**	0.667**		0.669**	0.604**		0.719**	0.694**		0.658**	0.566**	*
	[0.111]	[0.116]		[0.185]	[0.180]		[0.112]	[0.122]		[0.136]	[0.136]		[0.148]	[0.153]	
(1) High school complete	1.072**	0.974**	+	1.299**	1.240**		0.875**	0.755**	*	0.793**	0.836**		1.135**	0.960**	**
	[0.150]	[0.160]		[0.210]	[0.202]		[0.198]	[0.214]		[0.286]	[0.296]		[0.176]	[0.187]	
(1) Some college or more	2.072**	2.014**		2.499**	2.446**		1.731**	1.676**		2.246**	2.221**		2.001**	1.902**	**
	[0.134]	[0.138]		[0.214]	[0.219]		[0.171]	[0.172]		[0.206]	[0.204]		[0.169]	[0.174]	
(1) Married	0.067	0.074		-0.344*	-0.288*		0.369**	0.346**		0.097	0.128		0.035	0.025	+
	[0.078]	[0.081]		[0.138]	[0.140]		[0.101]	[0.105]		[0.123]	[0.126]		[0.103]	[0.104]	
(1) Rural	-0.446*	-0.464*		-0.377	-0.351		-0.498+	-0.540+							
	[0.217]	[0.227]		[0.269]	[0.283]		[0.286]	[0.289]							
Constant	6.128**	6.173**		6.546**	6.426**		3.570**	3.584**		7.265**	7.286**		5.589**	5.445**	
	[0.264]	[0.278]		[0.413]	[0.416]		[0.317]	[0.328]		[0.300]	[0.300]		[0.248]	[0.317]	
Number of observations	8,856	8,856		3,139	3,139		5,717	5,717		3,666	3,666		5,190	5,190	
R-squared	21.71	20.02		26.68	22.35		19.56	18.29		23.72	23.10		15.76	14.99	
F-statistic	70.62	66.58		51.67	44.54		32.51	29.65		41.02	40.49		58.79	53.07	

Sstandard errors in brackets (adjusted for clusters at municipality level)

** p<0.01, * p<0.05, + p<0.1

F. F-statistic for hypothesis that all coefficients are equal to zero

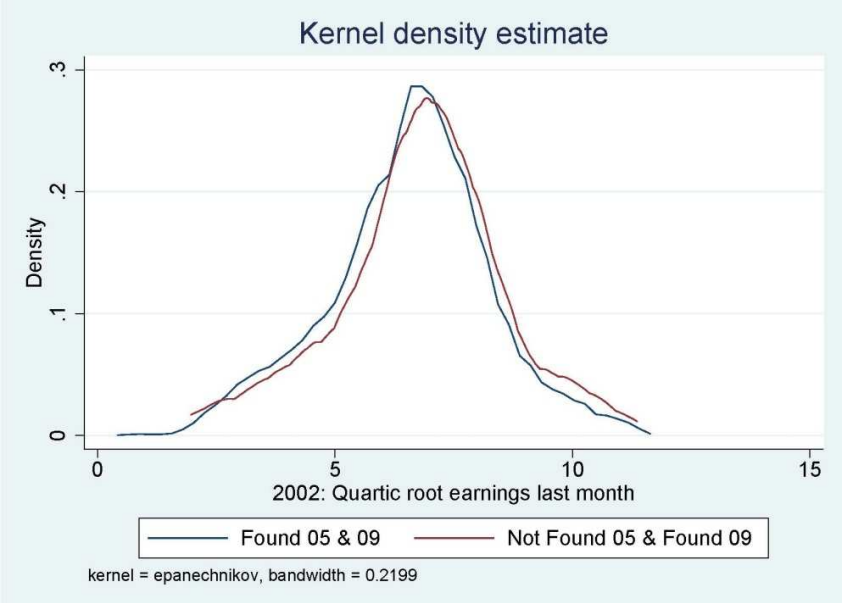

Figure 1: Distribution of the Quartic-root of Earnings in 2002

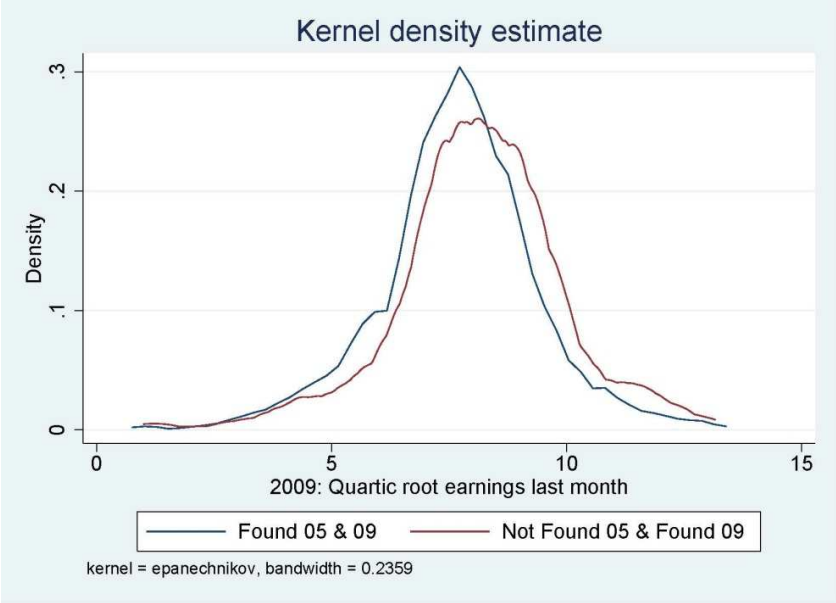

Figure 2: Distribution of the Quartic-root of Earnings in 2009

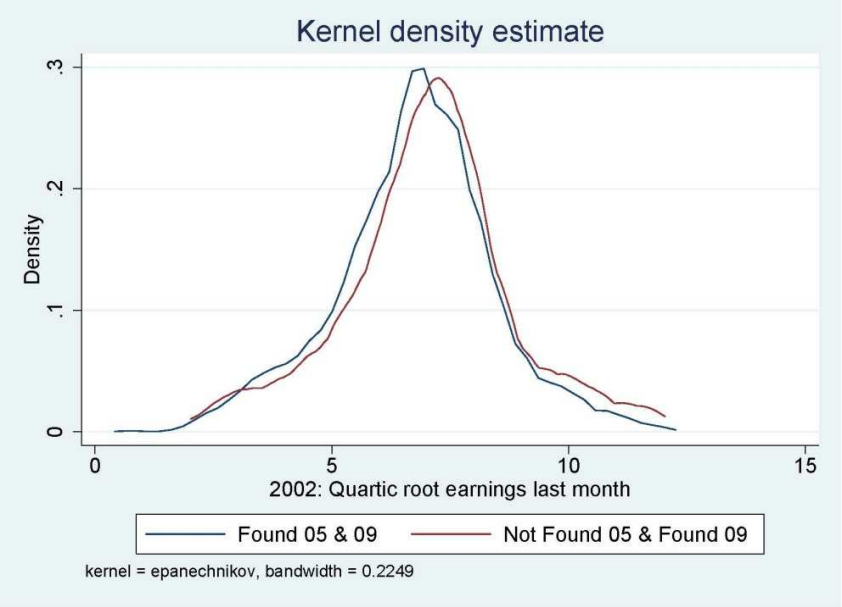

Figure 3: Distribution of the Quartic-root of Earnings in 2002 – Males

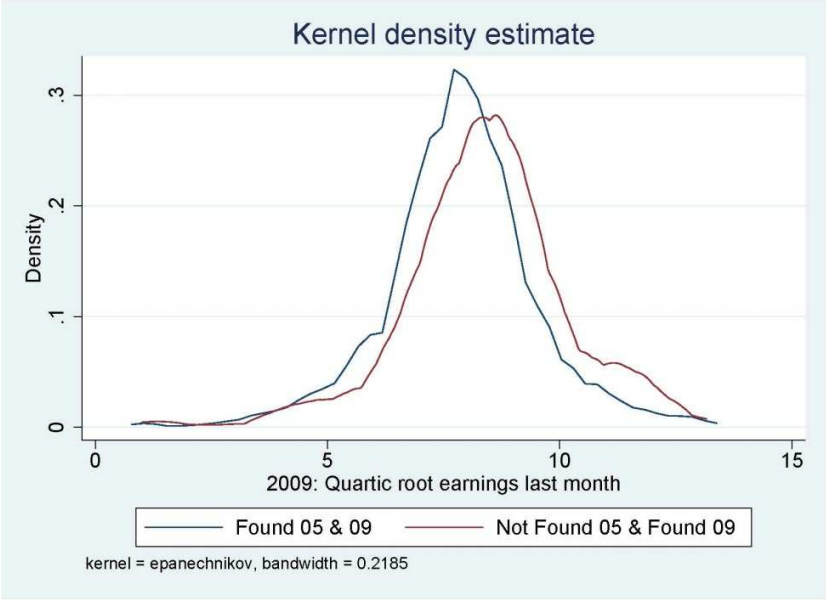

Figure 4: Distribution of the Quartic-root of Earnings in 2009 – Males

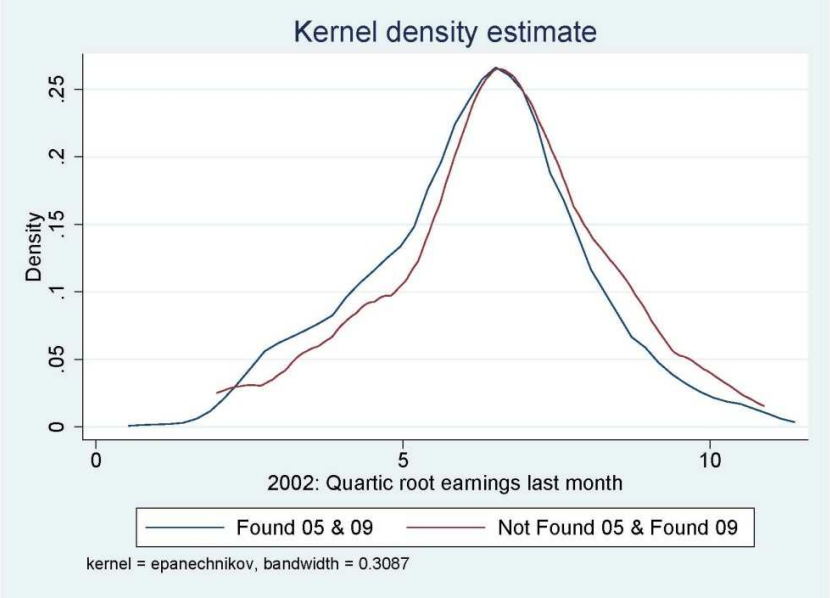

Figure 5: Distribution of the Quartic-root of Earnings in 2002 – Females

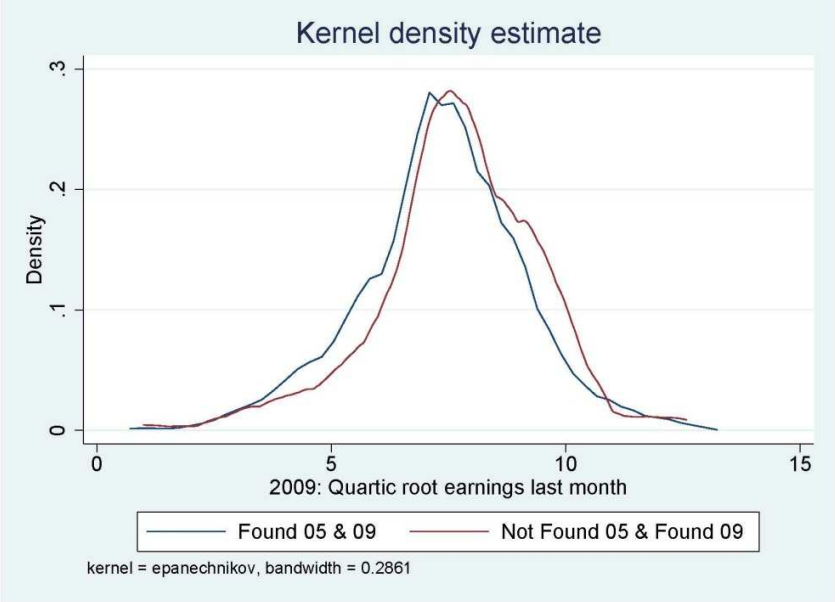

Figure 6: Distribution of the Quartic-root of Earnings in 2009 – Females

3. Selection and Assimilation of Mexican Migrants to the U.S.

3.1 Introduction

Mexican migration to the United States and the return of Mexican-born migrants to their country of origin are of substantial interest from both a policy and scientific point of view. Mexican-origin migrants are the largest Hispanic population in the U.S., accounting for nearly two-thirds of all Hispanic migrants. Moreover, Mexican migrants have traditionally followed two distinct patterns of migration; one fraction migrates to settle permanently in the U.S. while others are cyclical migrants moving frequently between the two countries. Recent evidence suggests that these patterns may be changing as migration from Mexico to the U.S. has declined sharply since the onset of the Great Recession and many of the migrants who were living in the U.S. have returned to Mexico. It is estimated that in the last few years, net migration from Mexico to the U.S. has fallen to zero.

This chapter uses novel data to provide new evidence on the characteristics that predict which Mexicans have chosen to migrate to the U.S. over the last decade and, among those migrants, the attributes that predict whether they settle in the U.S. for the longer term or return to Mexico. Finally, evidence is presented that sheds light on which characteristics at baseline are markers of successful assimilation in the U.S.

A large and active literature has examined the process of migration into the U.S. by Mexican citizens (see, for example, Donato, et al., 1992; Durand and Massey, 1992;

Durand, et al., 1996; Durand, et al., 2001; Fernández-Huertas Moraga, 2011; Hanson, 2006; Hoefer, et al., 2006; Ibarraran and Lubotsky, 2007; Massey et al., 1990; Massey and Singer, 1995; McKenzie and Rapoport, 2004; and Rendall, et al., 2011). Despite this extensive literature, evidence on recent changes in the selectivity of migrants, as well as, analysis of the characteristics that determine which migrants stay in the U.S. rather than return to Mexico is less well documented. This project sheds new light on these important subjects.

To provide scientific evidence on the selectivity of migrants, it is necessary to compare characteristics of migrants to those of non-migrants before the migration takes place. However, the absence of pre-migration information has been a challenge for research in this field. One line of inquiry uses data collected by the U.S. government, such as the American Community Survey in combination with census data from the Mexican government. This necessarily involves a comparison of the characteristics of non-movers to the attributes of movers after the migration has already occurred. Using data collected in this way, it is difficult to draw firm conclusions about the roles of time-varying characteristics in migration decisions. Moreover, studies based on data collected by the U.S. government are limited by the fact that such surveys are known to undercount the undocumented and most mobile migrants including those who cycle often between the U.S. and Mexico.

A second line of inquiry has relied only on Mexican census or survey data collected in Mexico on individuals that have already migrated. An important limitation of these data sources is that they only include migrants that have returned to Mexico or have at least one household member still living in Mexico and information about migrants is obtained by proxy. By design, these surveys exclude complete households that have migrated to the U.S. who make up an increasing and substantial fraction of migrants to the U.S. from Mexico. Both the selection and assimilation process of complete households that decided to migrate to the U.S. and never return to Mexico are likely to be quite different from the rest of the migrants and so studies based on these data sources are prone to biases due to the selection of those included in the study. An important, related source of data is the Mexican Migration Project (MMP) which is based on a sample of respondents in Mexico who have family in the U.S. While these data are supplemented by U.S. based snowball samples, the samples are also selected on having at least one family member remaining in Mexico and so these samples are also at risk of underrepresenting longer-term movers. Moreover, both types of surveys only contain information on movers after the migration event. The ideal source of data for a study of migrant selectivity would be a sample that is representative of the Mexican population prior to migration and proceeds to follow all respondents that migrate to the U.S., including those who stay for a short period and those who remain in the U.S. long-term.

We designed and implemented an approach to study migrant selectivity using this methodology.

The Mexican Family Life Survey (MxFLS) is an ongoing longitudinal data set that is representative of the Mexican population at baseline (in 2002). The first follow up was completed in 2005 and the second follow-up will be completed in mid-2013. In both follow-ups, movers to the U.S. have been tracked and interviewed in the U.S. The baseline respondents who are thought to have moved to the U.S. were found and interviewed at a rate of 90% and 85% in the first and second follow-up surveys, respectively. So far, we have re-interviewed over 1,000 baseline respondents who moved to the U.S. after 2002 and are still in the U.S., as well as, 570 respondents who migrated to the U.S. after 2002 and subsequently returned to Mexico.

The combination of successfully tracking and interviewing movers, including international movers, with detailed information on their labor market and migration experiences, families and resources of each respondent yields an extremely rich set of data for investigating the nature of selectivity of migrants to the U.S. and the selectivity of those who remain in the U.S. over the longer haul. With these data, we will draw comparisons between those people who have migrated to the U.S. since 2002, and stayed, those who have migrated to the U.S. since 2002 and returned to Mexico and those who have only migrated within Mexico since 2002.

3.2 Data

The MxFLS is an on-going large-scale population-representative longitudinal survey of Mexicans who were living in Mexico in 2002 when the baseline was conducted. The baseline survey, MxFLS1, collected detailed information on 35,677 individuals living in 8,440 households in 150 communities spread across 16 states in Mexico.

The second wave, MxFLS2, was conducted in 2005-2006. All baseline respondents and their biological children born after the 2002 baseline are eligible to be tracked in the follow-up surveys. They are our "panel respondents." Over 89% of the panel respondents were re-interviewed in MxFLS2.

A novel feature of MxFLS, which is key for this research, is that we decided to not only follow panel members that had moved within Mexico but also to follow respondents that had migrated to the U.S. Following movers is not straightforward. In the Mexican context, it poses special challenges because a significant number of people move to the U.S. Moreover, Mexican migrants are generally very mobile and the great majority is undocumented, adding additional challenges to the tracking process. In the context of the MxFLS, following migrants to the U.S. is important not only because it allows us to have a representative sample of recent migrants to the U.S. but also because it is crucial to maintain the representativeness of the baseline sample. If migrants to the U.S. are not followed, not only would attrition rates be higher than otherwise, but

attrition would also be selected on characteristics associated with migration to the U.S. Inferences about the evolution of many indicators of well-being of the Mexican population over the last decade would potentially be contaminated if domestic and international migrants were not followed.

As such, we designed and implemented an approach that allows us to have a representative sample of recent migrants to the U.S. for whom we have a rich set of characteristics measured at baseline, prior to migration. To achieve this, we follow all panel respondents that remain in Mexico, as well as, track respondents who move to the U.S. The MxFLS is the first population-representative large-scale longitudinal study that has attempted to follow migrants across an international border. Aware of the additional challenges this poses, we put substantial effort into developing and testing procedures to facilitate successfully interviewing migrants in the U.S. Those efforts have allowed us to maintain a very high re-contact rate: as shown in panel A of Table 11, in the second wave, 89.2% of the baseline respondents were re-interviewed. Moreover, we interviewed 91% of the 854 respondents believed to be in the U.S. (This includes anyone who was reported by an informant to have moved to the U.S. and was not interviewed in Mexico.) Furthermore, of the U.S moving panel respondents who were at least 15 years old at the time of the baseline, 90.4 percent were re-interviewed in the U.S. In the rest of the analysis we will focus on this age group. This choice is made because children are likely

to move because of a migration decision made by their parents and it is their parents' characteristics that are most likely driving the selection.

The third wave of the survey, MxFLS3, is currently in the final stages of fieldwork. We anticipate achieving the same re-contact rate as in MxFLS2. At March 2013, over 84% of the U.S. migrants have been interviewed in MxFLS3. Panel B of Table 11 shows current recontact rates of MxFLS3. The results for the entire sample show that over 85 percent of baseline respondents have been re-interviewed. From those baseline respondents, 1,876 have moved to the U.S., and at this time 85.2 percent has been re-interviewed. Due to the transient part of the Mexican migrant population, it is important to not only follow migrants that moved from Mexico to the U.S. but also to track those that return to Mexico after having lived in U.S. at some point between the waves. Out of the total U.S. migrants successfully re-interviewed, near 65% were still in the U.S. and the other 35% were found and interviewed back in Mexico. The recontact rates for the entire sample and the sample of respondents older than 15 at baseline are very similar.

In order to examine the selection and assimilation of Mexican migrants we will first group each respondent into one of several migration categories. These categories are based on the respondent's migration history and place of re-interview. We will describe in detail the migration history component and the migration categories used in our analysis.

3.2.1 Migration and Characteristics at Baseline

The MxFLS collects a rich array of information about each respondent, this include a component about the migration histories of all respondents age 15 or older at the time of the survey. The 2002 migration history component of the MxFLS includes all long-term movements (more than one year) and temporal movements (between one and 12 months) that occurred after age 12. This allows us to determine who has moved to the U.S. and who has moved within Mexico prior to 2002. In MxFLS2 the migration component follows the same structure and updates the migration history by assessing both permanent and temporal migration movements between the two waves. This feature, in addition to the successful tracking of migrants in the U.S. in the second wave, allows us to determine the migration trends of the Mexican population between 2002 and 2005. In a similar way, MxFLS3 updates the migratory movements between the second and third wave. In addition, we tracked and interviewed respondents thought to be in the U.S. as well as those confirmed to be back in Mexico. This unique feature allows us to describe the recent migratory trends of the Mexican population both within Mexico and to the U.S. and allows us to compare different groups of the population on a rich set of characteristics measured at baseline according to their migration experiences through the third wave.

Our analytical sample includes all panel respondents age 15 or older in 2002. In order to explore the migration status (between baseline and the third wave) of each

respondent we use data from the three waves of the MxFLS. Exploiting both the migration histories and the place of residence at the moment of each survey we can classify the migration status of each respondent in each wave. We classify the respondents as "non-movers", if they never moved from their locality of origin (the locality in which they were interviewed in 2002) for a period longer than a year; "movers within Mexico", if they moved out of their locality of origin but did not move to the U.S for a period longer than a year; "moved to U.S. and returned", if they moved to the U.S. for a period longer than a year but in the third wave were found and interviewed in Mexico; and "moved to the U.S. and stayed", if they were found and interviewed in the U.S. in the third wave. Table 12 shows the sample sizes in each group by gender.

Along with the migration component, the MxFLS contains information about the economic, social and health status of each member of a surveyed household. The questionnaire for adults includes sections on education, labor supply and earnings, marriage and fertility history, health status, and use of health care. In addition, one member is interviewed about information at the household level. This questionnaire includes a complete household roster including basic socio-demographic characteristics of each household member, and information of household expenditure, and asset ownership.

Another useful section of the MxFLS is the assessment of the presence of relatives in the U.S. for all baseline respondents. An important variable for predicting migration is the presence of networks in the destination place. Specifically, the presence of networks in the U.S. could affect the decision to migrate through several different channels. For example, networks in the place of destination may reduce the initial costs of migration if the relatives help with living expenses. In addition, they can offer valuable information about available jobs or connect the recent migrant to job networks. Our measure of direct networks in the U.S. prior to migration will allow us to explore these hypotheses.

Table 13 provides descriptive statistics for the variables that measure the presence of U.S. familial networks. Panel A provides evidence that conditional on being a migrant, those that moved to the United States have more relatives in America at baseline than those that only moved within Mexico. Moreover, disaggregated by family relationships, U.S. movers are significantly more likely to have relatives of each relationship type in the U.S. with the exception of extended family members. These results suggests than networks prior to migration have an important role in the subsequent decision to migrate.

In addition to influencing the initial decision to migrate, the presence of networks in the U.S. could affect the decision on the length of the migrants stay. Table 14 shows the same variables for U.S. migrants, distinguishing between those who

subsequently returned to Mexico and those who stayed in the U.S. The results show that the difference in the level of networks in the U.S is less significant when looking only at U.S. migrants. Even though current U.S migrant males are more likely to have relatives in the U.S. the presence of separate types of relatives is not significantly different between returners and stayers, with exception of siblings: current migrants are more likely to have siblings in US prior to migration. The results are similar for females with the one exception that female stayers are less likely to have extended family in the U.S. at baseline.

This preliminary evidence suggests that networks might be an important determinant for the migration decision but is not a strong predictor of the decision to stay in the U.S. In a later section, we will explore these relationships more rigorously in a regression framework that allows us to control for a broader group of characteristics measured at baseline both at the individual and household level.

3.2.2 Assimilation Outcomes

A second goal of this chapter is to determine the characteristics, prior to migration that have predictive power of the migrant's ability to assimilate in the U.S. For this analysis we will explore the sample of U.S. migrants interviewed in the U.S. in MxFLS2 and MxFLS3. In MxFLS2 we will use hourly wages to explore the baseline characteristics associated with successful assimilation. Additionally, we will compare this for migrants that later stayed in the U.S. and for those that later returned to Mexico

to see if baseline attributes had varying importance towards assimilation for these two groups.

Using MxFLS3 we can explore a broader group of markers of assimilation including: knowledge of English, whether the migrant's spouse is in U.S. conditional on being married, whether his/her children are in the U.S. conditional on having children and whether the migrants has sent remittances to Mexico in the last year.

For these analyses our sample is formed by all panel members age 15 or older at the time of the interview in MxFLS3. Table 15 shows basic statistics of the assimilation markers for our sample of respondents interviewed in U.S in the third wave.

3.3 Migration to the U.S.: Who Migrates, Who Stays, and Who Returns to Mexico

The first goal of this chapter is to explore the selectivity of both migration to the U.S. and return migration to Mexico. Four features of the data are key for this analyses.

First, we have detailed information about the lives of all the movers – and those who do not move – prior to the index international move (which occurred after 2002). Because of the design of MxFLS, these analyses are not contaminated by undercounts of the most mobile migrants from Mexico in U.S. surveys or by the loss of complete households that move. The latter concern is an increasingly common phenomenon among Mexican-origin migrants and is clearly documented in MxFLS (Farfan et al, 2012.)

Second, we follow respondents who return to Mexico and have detailed information about their experiences in Mexico prior to moving, their experiences while in the U.S., and their experiences in Mexico once they return.

Third, detailed information about migration experiences, labor market outcomes, and human capital are recorded in every wave of the MxFLS. It is, therefore, possible to provide a rich description of the nature of selection of migrants into the U.S. relative to those who stayed in Mexico. Similarly, focusing on those respondents who moved to the U.S. during the hiatus between the baseline and first re-survey, we will describe the characteristics that distinguish those who subsequently return to Mexico with those who stay in the United States.

Fourth, whereas much of the information described above is recorded in surveys that have been used for analyses of selectivity of migrants, MxFLS contains a far richer array of information on the lives of respondents than has been used in prior analyses. This included information at baseline, before the migration event, not typically found in surveys used to analyze migration such as questions about own wealth and the wealth of household and family members, living arrangements, and the presence of networks in locations other than the baseline community.

Table 16 and 17 present preliminary results for males and females, respectively, age 15 and older at baseline in 2002 using a linear probability model that explores which baseline characteristics are predictors of subsequent migration. We analyze the

migration decision for three different samples. The model in columns 1 and 2 predict the probability of any long-term migration, thus both within Mexico and to the U.S., for the entire sample of respondents. Columns 3 and 4, predicts migration to the U.S. conditional on being a migrant, and columns 5 and 6 predicts permanence in the U.S. conditional on being an U.S. migrant. For the purpose of this chapter we will focus on the results that predict migration to the U.S. (Columns 3 and 4) and staying in the U.S (Columns 5 and 6). For each of these samples we estimate two models that differ only on the measures used for the presence of networks in the U.S. The first model uses the number of relatives a respondent has in the U.S. at baseline as the measure of network presence while, in the second model, we disaggregate this variable by the relationship of the relative living in U.S. with our panel respondent. To highlight the role of individual and family factors in the selection process, these models include state fixed effects. Conceptually, we are comparing, for example in columns 3 and 4, those who move to the U.S. (both stayers and returners) with those who are from the same states and only move within Mexico.

Table 16 shows the results for males. There are four main results. First, young Mexicans are the most likely to move either within Mexico or to the U.S. and given that a respondent is a mover the youngest ones are the most likely to move to the U.S. However, once migration to the U.S. takes place the age of the migrant is not a predictor

of whether he remains in the U.S over the longer term. These patterns are unaffected by the choice of measure for networks in the U.S.

Second, human capital predicts of migration both within Mexico and to the U.S., and the effects are different depending on the migrant's destination. Males who have not completed high school are about 3 percentage points more likely to move than those with no education and those who have some college are over 4 percentage points more likely to move. Conditional on being a migrant, those with only primary education are the most likely to move to the United States and those with some college are by far the least likely to move. This inverted U shape has been established in many studies of migrants from Mexico to the United States. A result that is not as well established is that among those Mexican males who move to the United States, those with some college education are also the least likely to remain in the United States for the longer term. These results are qualitatively equivalent when we do not control for height and the cognitive score.

Two other dimensions of human capital are included in the models: a non-verbal cognitive assessment (the Ravens Progressive Colored Matrices test) and height. Height is not a significant predictor of migration, after controlling for education, but, even after controlling education, it is the migrants with lower cognitive scores that move to the U.S. Amongst the group that migrates to the United States, though, the cognitive score is not a significant predictor of their remaining in the United States for the longer term.

These results suggest that those in the lowest end of the distribution of education are the least likely to move to the United States, but having at least complete primary is enough to increase the likelihood of migrating to the U.S. However, the effect of education is non-lineal and higher levels of education do not increase the probability of migrating and in fact being in the highest level of the education distribution (college or more) significantly diminishes the probability of migrating to the U.S. These results hold when we estimate a specification that does not include height and the Ravens score. If the expected income of individuals with higher levels of education is greater in Mexico than in the U.S., it is reasonable to find that better educated individuals are more likely to stay in their country. Expected income at home and abroad plays an important role in the migration decision, but does not explain the whole picture. The migration of an individual can be understood as the decision of the whole household to economically support the migration costs. In addition, networks at the destination place can decrease the expected costs of migration (initial living expenses, information costs) and therefore, play a crucial role in the migration process. This will be explored in our next set of results.

An innovative feature of the MxFLS is that it collects information about networks in U.S. prior to migration, this will provide our third results of interest. Table 16 establishes that the presence of networks in the U.S. is a powerful predictor of migration. Results in columns 3 to 6 show that male migrants are more likely to move to the U.S.,

the higher the number of relatives they have in the U.S. and having only one relative is enough to increase the probability of staying in the U.S. Moreover, the presence of parents, son/daughters, or siblings in the U.S increases the likelihood of moving to the U.S. but only the presence of the son/daughters or siblings increases the probability of staying in the U.S. These results make a clear statement that the presence of networks in the destination place is important for making the decision to move to the U.S. and then, for staying more permanently.

The fourth main result is that living in a rural place at baseline has a different effect depending on whether the migrant is choosing to move or not versus the choice to move within Mexico versus to the U.S. While rural respondents are less likely to move, conditional on migrating, panel members living in a rural place at baseline, are 13 percent more likely to migrate to the U.S. Moreover, while not statistically significant, there is also suggestive evidence that rural respondents are more likely to stay in the United States given that they've migrated into the U.S. The skills of rural farmers might be useful in agricultural and seasonal jobs in the U.S., which could make U.S. migration more appealing, profitable, and sustainable for these individuals.

In addition to the covariates already discussed, the models control for a set of household characteristics that include household composition and household assets. These factors are relatively modest predictors of migration, with the exception of business ownership by the household at baseline and household wealth per capita at

baseline. Many studies have suggested that those with more assets in Mexico are more likely to keep ties with their home country and eventually return to Mexico. Ownership of a business while seeming to provide some of the recourses necessary to migrate to the U.S, does not appear to be a predictor of returning to Mexico, and in fact the sign of the coefficient suggests the opposite effect. Finally, conditional on moving to the U.S., being in the wealthiest quartile increases the likelihood of staying in the U.S.

Table 17 shows results of the same models for females. First, while age is a significant predictor of migration to the U.S. for males, it is not for women. Only women older than 50 at baseline have a lower probability of moving to the U.S. but in the younger groups there are not significant differences.

Second, education is a significant predictor of overall female migration but not of U.S. migration. Contrary to males, conditional on education, height is a significant predictor of overall migration and U.S. migration for females: taller women are more likely to migrate and more likely to migrate to the U.S. On the other hand, the cognitive raven's score does not have predictive power for migration to the U.S. for women.

Third, as for males, the presence of networks in the U.S. also plays an important role in the migration process for females. Having relatives in the U.S. increases the probability of moving in general and conditional on moving it increases the likelihood that the move will be to the U.S. Interestingly the effect sizes are considerably larger on

these characteristics for women than for men. However, relatives in U.S. do not play a role in the decision to stay in the U.S. for females, while it does for males.

Fourth, as before, coming from a rural locality increases the likelihood of migration to the U.S. yet the effect is half the size (6 percent) for females compared to males.

These results provide evidence that selection into migration and, then, into laying down roots for the long haul are dissimilar processes determined by different characteristics and that longer term-migrants are not the same as those who migrate to the U.S. for the short term. These results raise questions about what characteristics, if any, are predictive of success in the labor market in the U.S. The next section address this question.

3.4 Assimilation

The second goal of this research is to provide evidence about predictors of markers for successful assimilation in the U.S. For this analysis we will exploit information collected in U.S. during the second and third waves of MxFLS. First, information collected in the second wave allows us to compare return migrants to those that stayed in the U.S. to determine whether characteristics at baseline predicted a more successful assimilation for either group. In this analysis, assimilation is measured as the individual level of earnings in the U.S. A second analysis we have conducted uses data

from the third wave of MxFLS in which we measure assimilation outcomes for stayers whom we find and interview in the U.S. Assimilation is measured with four different markers: the earnings of the migrant, the per capita expenditure in the U.S., whether the migrant has sent transfers to Mexico, knowledge of English, whether the migrants' spouse (conditional on being married) lives in U.S. and whether his/her children live in the U.S. (conditional on having children alive).

For each of these outcomes, we assess whether socio-economic and demographic characteristics measured at baseline are predictive of the extent of assimilation for the select group of migrants who have stayed in the U.S. By drawing on the same models that are used in the analyses of selectivity of migrants, we provide a comprehensive picture of those characteristics that are predictive of both selection into migration and success in the new destination. Further, comparing the extent of assimilation in these dimensions of those who continue to stay in the U.S. with those who return to Mexico provides insights into the likely mechanisms that underlie decisions to set down roots for the longer haul.

Table 18 shows the results of assimilation using the log of hourly U.S. earnings measured in MxFLS2 for individuals age 15 or older at the time of the interview in U.S. as the outcome. In the first two columns we show the results for the entire sample of migrants interviewed in US in MxFLS2 who report positive earnings, the following two columns show the same estimations for the subset of migrants who subsequently

returned to Mexico and the last two columns show the results for the group of migrants who stayed in the U.S. For each sample, we show two different estimation models as in the previous tables: the first one includes networks in U.S. measured as the number of relatives and the last one measures networks as the relationship of the migrants to their connections in the U.S. We focus the analysis on the results for education and networks in the U.S.

The results in Table 18 show that female migrants earn on average 40 percent less than males, and the coefficient is very similar for migrants who stayed in the U.S. or return to Mexico. Education levels achieved at baseline are significant only for the sample of stayers: migrants whom had some years of high school or completed college earn higher hourly earnings than those in the lowest category of education. However, education attained at baseline does not seem to have any effect on the level of earnings of migrants that subsequently returned back to Mexico.

The results for the variables that measure networks in U.S. suggest that, even though their presence in the U.S. is an important predictor of migration, is not evident that they will determine a more successful assimilation in the U.S. Having extended family in the U.S. prior to migrate has a negative effect on the level of earnings and this effect holds only for return migrants.

Table 19 shows the results on what characteristics predict assimilation for the sample of U.S migrants who were found and interviewed in the third wave in the U.S.,

our sample of stayers. For this analysis we use the sample of individuals age 15 or older at the time of the interview in U.S. in MxFLS3. Table 19 shows the results for outcomes that measure economic assimilation: hourly earnings (columns 1 and 2), per capita expenditure (columns 3 and 4) and whether the individual sent transfers to Mexico (columns 5 and 6). Columns 1 and 2 show the results for the log of hourly earnings, and as in previous tables we estimate two models that differ by the measure of networks in U.S.

These results suggest that the gender wage gap that existed between migrants in the previous wave is still present in the most recent wave of the data. The measures of human capital show that education is an important determinant of higher earnings; but, its effect seems to be very linear.

Looking together at the results for the selection models and assimilation for stayers in Table 18, the results in Table 19 for networks in U.S. suggest that they not only predict migration to the U.S. but also affect how well the migrant does in the U.S. The results in Table 18 and Table 19 would together suggest that, the presence of relatives in the U.S. might have a positive effect on the level of earnings but, as suggested by Table 18, the effect is not immediate. The results using MxFLS3 show that the presence of two relatives in the U.S. increases earnings by 30 percent and in particular, siblings and the extended family are having a positive impact on the level of earnings.

Following the models for selection we control the assimilation estimates for household characteristics and, as in the previous models, these variables have a modest effect on the outcome of interest. However, a surprising result is the negative effect that being in the third quartile of the per capita wealth distribution in 2002 has on earnings. In future work we will explore non-parametric relationships to better understand these effects.

The results using the log of per capita expenditure PCE (columns 3 and 4) as the measure of assimilation confirm some of the results for earnings. In this analysis we keep only one observation per household and we keep the observation of the household head. A female household has a negative effect on the level of PCE and if the household head was married at baseline their current PCE is lower. A possible explanation put forth for this relationship is that these married household heads are spending less in the U.S. in order to send remittances to their families in Mexico; however, the results for transfers do not suggest that married individuals at baseline are more likely to send transfers. The education of the household head is an important determinant of the level of expenditure of the household. Household heads in the top of the distribution of education have on average a higher PCE.

Networks once again is an interesting part of the assimilation story. While individuals with more relatives in the U.S. do better in terms of earnings they spend less in the U.S. As with married household heads, one potential explanation is that migrants

with a larger network in the U.S may also have a larger network in Mexico and thus sends more transfers. However, the results for the likelihood of sending transfers do not show higher probabilities of sending transfers for migrants with a larger network in the U.S. In future stages of this work we will complement this analysis by looking at the savings of the household in U.S. in order to understand whether households with lower PCE are those with higher savings.

The last outcome of interest is whether the migrant sent transfers to Mexico in the last 12 months. An individual with deeper connections at home may have a more difficult time assimilating than an individual that does not leave as deep of roots in Mexico. Individuals that send transfers to Mexico might do it to keep savings back in Mexico or to invest in a business suggesting that the migrant has a plan to return to Mexico. Moreover, sending transfers to relatives as a financial help suggests that the migrant keeps deep roots in Mexico, which would affect their plans to stay in the U.S. permanently. When looking at transfers as an outcome we find very interesting results.

We find that individuals that are attending school in the U.S. are less likely to send remittances to Mexico by 30 percent. This may indicate that individuals that invest in their education in the U.S. have lower expectations of going back to Mexico; therefore, investing in maintaining relationships at home or investing at home is less necessary. Moreover, individuals with parents that are in the U.S. prior to migration are less likely to send remittances. Finally, the more recently the migrant has arrived to the U.S., the

more likely he/she is to send remittances to Mexico. This result may suggest that the first few years after migration are the most important in terms of maintaining connections and thus a potential safety net back in Mexico.

Another important variable for measuring assimilation in the U.S. is the level of English that the migrant speaks. The higher the level of English the migrant possesses the easier it is for him/her to build a network outside the Mexican circle. Moreover, speaking English well can help the migrant attain better and more permanent employment. Table 20 shows the results for assimilation measured by how well the migrant speaks English. The measure for how well the respondent speaks English is self-reported and he/she must assess whether he/she speaks Very Good, Good, Fair, Bad, very Bad or does not speak English. We build an index variable equal to one if the migrant self-reports that he/she speaks Fair, Good or very Good English. The results in Table 20 show that human capital measures are an important determinant of the level of English that the migrant speaks. The higher the education attained by the time of MxFLS3 the better the English spoken by the migrant. Further, this effect is non-linear, the higher the education attained the higher its impact on the respondent's English. Moreover, if the migrant attended school in the U.S. they, as expected, speak better English. Another measure of human capital is also an important determinant of this measure of assimilation as a higher cognitive score has a positive relationship to English proficiency. In addition, the earlier the migrant arrived to the U.S. the better his/her

knowledge of the foreign language. Interestingly, the presence of networks in U.S. does not seem to have a positive or negative effect on the respondent's level of English assimilation.

Finally, we analyze living arrangements in the U.S. as a proxy for assimilation (whether the migrant's spouse or children live in the U.S.). Table 21 shows the results for whether the migrant's spouse lives in the U.S. conditional on him/her being married and for whether the migrant's children live in the U.S. conditional on him/her having children alive. We show these results disaggregated by gender.

Interesting results are found for the presence of networks in the US on this measure of assimilation. For instance, while the number of relatives in the US at baseline are predictive of whether a males' children are in the U.S. in the third wave, this does not matter for women. Although the network size is not significant for females, whether her parents or siblings were in the U.S. prior to migration increases the likelihood of her children being with her in the U.S. at the time of the MxFLS3 survey. In terms of having one's spouse in the U.S. in the third wave we find that baseline characteristics do not have predictive power for females. For males, on the other hand, there are some noteworthy results.

First, owning a farm business at baseline increases the likelihood of having a spouse in U.S. in the third wave, while being in the second and third quartile of the per capita wealth distribution decreases this probability. Second, the more co-resident

children in the baseline home decreases the likelihood of the spouse being in U.S. at follow-up. Having several children in Mexico may make it more difficult for the spouse to migrate since each child increases the initial migration cost, through the additional expense of bringing them as well or the expense of setting up care for them once the spouse leaves. Third, the later the arrival date of the migrant the less likely that his spouse is with him at follow-up, which could suggest that a male migrant arrives alone first with a possible reunion with his spouse once a secure foundation has been established by the migrant.

3.5 Discussion and Future Work

In summary, the analyses presented in this research provide new insights into the mechanisms that underlie the selectivity of migrants within Mexico, how they differ from migrants who move from Mexico to the United States and how those who return differ from those migrants who remain in the United States over the longer haul. By estimating parallel models of multiple markers of assimilation in the United States, we can draw conclusions about the predictors of both selection into migration and the predictors of success in the destination among those who move and stay.

Human capital (as measured by education and cognitive skills) are predictive of migration within Mexico and to the United States. Those who move to the United States are not likely to be drawn from the bottom or top of the education distribution. There is

little evidence that the decision to stay in the United States is highly correlated with education of the movers except that those with some college education are far more likely to return to Mexico than any other migrant. Moreover, conditional on moving to the United States, there is little evidence that education carries a premium in terms of earnings in the labor market. This is in sharp contrast with results for natives in the United States. The results suggest that, relatively speaking, migrants to the United States from the bottom of the education distribution are doing better in the United States than those who are drawn from the top of the distribution.

In contrast, having relatives in the United States is not only a powerful predictor of migration to the United States but it is also predictive of success in the labor market. Specifically, males are more likely to move to the United States if their spouse, a parent, child or sibling was living in the United States. The presence of a child or sibling is predictive of staying in the United States (at least for males). However, it is only the presence of a spouse that is predictive of elevated earnings and having a child or extended family members is predictive of lower earnings.

3.6 Tables and Figures

Table 11: Sample Sizes and Recontact Rates in MxFLS

Panel A. Recontact rates in MxFLS2

	All			Age in 2002>=15		
	Eligible for survey	Interviewed	% Interviewed	Eligible for survey	Interviewed	% Interviewed
Total	35,134	31,338	89.20	23,222	20,612	88.76
In Mexico	34,280	30,564	89.16	22,606	20,055	88.72
In US	854	774	90.63	616	557	90.42

Source: MxFLS

Note - Excluded panel respondents who died between waves

Panel B. Recontact rates in MxFLS3

	All			Age in 2002>=15		
	Eligible for survey	Interviewed	% Interviewed	Eligible for survey	Interviewed	% Interviewed
Total	34,225	29,238	85.43	22,357	18,845	84.29
In Mexico	32,349	27,640	85.44	21,123	17,791	84.23
In US	1,876	1,598	85.18	1,234	1,054	85.41
US sample ivw in MX		570			430	
US sample ivw in US		1,028			624	

Source: MxFLS

Note - Excluded panel respondents who died between waves

Table 12: Migration Between Baseline and MxFLS3

	MALE	FEMALE	Total
0.Non movers	7,371	8,615	15,986
since 2002	*83.69*	*85.75*	*84.78*
1.Movers within Mexico	739	1,023	1,762
since 2002	*8.39*	*10.18*	*9.35*
2.Moved to U.S.	345	129	474
and returned	*3.92*	*1.28*	*2.51*
3.Moved to U.S.	353	280	633
and stayed	*4.01*	*2.79*	*3.36*
Total	8,808	10,047	18,855
col%	*100*	*100*	*100*

Table 13: U.S. Networks Reported at Baseline by MX/U.S. Migration Status

Panel A

	MALE				
	MX Mover		US Mover		P-value
Variables measured at baseline	mean	sd	mean	sd	Diff
% Has relatives in US	48.54	50.02	70.53	45.63	0.00
# of relatives in US	0.82	1.10	1.41	1.30	0.00
% Spouse in US	0.00	0.00	0.43	6.55	0.07
% Parents in US	1.20	10.91	5.19	22.20	0.00
% Daughter/son in US	2.05	14.17	5.36	22.55	0.01
% Siblings in US	13.18	33.87	27.97	44.93	0.00
% Extended family in US	17.95	38.42	20.50	40.41	0.32

Panel B

	FEMALE				
	MX Mover		US Mover		P-value
Variables measured at baseline	mean	sd	mean	sd	Diff
% Has relatives in US	51.93	49.99	81.86	38.58	0.00
# of relatives in US	0.91	1.14	1.82	1.37	0.00
% Spouse in US	1.58	12.46	10.78	31.06	0.00
% Parents in US	3.21	17.64	9.41	29.23	0.00
% Daughter/son in US	2.13	14.44	9.40	29.23	0.00
% Siblings in US	16.02	36.70	40.74	49.21	0.00
% Extended family in US	16.90	37.50	18.80	39.13	0.43

Table 14: U.S. Networks Reported at Baseline by U.S. Migration Status

Panel A

	MALE				
	US Returner		US Stayer		P-value
Variables measured at baseline	mean	sd	mean	sd	Diff
% Has relatives in US	66.57	47.25	74.35	43.73	0.03
# of relatives in US	1.32	1.29	1.49	1.32	0.09
% Spouse in US	0.29	5.38	0.57	7.52	0.58
% Parents in US	4.29	20.29	6.04	23.86	0.34
% Daughter/son in US	3.97	19.56	6.67	24.99	0.17
% Siblings in US	22.62	41.92	32.96	47.10	0.01
% Extended family in US	19.44	39.66	21.48	41.15	0.57

Panel B

	FEMALE				
	US Returner		US Stayer		P-value
Variables measured at baseline	mean	sd	mean	sd	Diff
% Has relatives in US	80.99	39.40	82.25	38.28	0.77
# of relatives in US	1.84	1.41	1.81	1.35	0.83
% Spouse in US	7.75	26.85	12.19	32.77	0.18
% Parents in US	6.03	23.92	10.94	31.27	0.13
% Daughter/son in US	11.93	32.56	8.26	27.59	0.28
% Siblings in US	28.44	45.32	46.28	49.96	0.00
% Extended family in US	25.69	43.89	15.70	36.46	0.03

Table 15: Characteristics U.S. Migrants - Panel Members Age15+

Variables measured in MxFLS3	mean	sd
Age in MxFLS3	30.58	13.19
% Female	41.11	49.23
Years of education	8.45	3.38
Height (cm)	161.56	9.46
Household size	3.15	1.89
% Married - conditional on age>=12	56.33	49.63
% Spouse in HH - conditional on married	82.60	37.96
% Has children alive	60.47	63.23
% Has children in US - conditional on having ch alive	81.90	38.53
% Worked last week	75.03	43.31
Ln(Hourly earnings)	1.99	0.72
% Speaks English Fair - Good or Very Good	43.46	49.60
% Has sent transfers to Mexico	67.37	46.91
Number Obs	934	

Table 16: Baseline Characteristics that Predict Migration Since 2002 - Males

	Moved since 2002 = 100		Moved to US since 2002 =100		Stayed in US = 100	
Variables measured at baseline	(1)	(2)	(3)	(4)	(5)	(6)
Basic demographics						
(1) Age: 20- 24 Omitted 15-19	-1.84	-1.28	-5.67	-4.25	-8.62	-8.33
	[1.688]	[1.690]	[3.686]	[3.733]	[5.963]	[6.136]
(1) Age: 25-34	-6.53***	-6.94***	-11.60***	-10.73**	-8.06	-8.96
	[1.613]	[1.615]	[4.196]	[4.220]	[7.065]	[7.134]
(1) Age: 35-49	-12.64***	-13.17***	-12.63**	-13.10**	-1.6	-2.61
	[1.738]	[1.741]	[5.076]	[5.122]	[9.043]	[9.197]
(1) Age> 50	-18.63***	-17.19***	-25.07***	-26.93***	-6.44	-12.35
	[1.906]	[1.906]	[5.779]	[5.809]	[11.987]	[12.537]
(1) Married	1.82	1.35	-4.68	-5.25	-3.37	-2.03
	[1.372]	[1.395]	[4.035]	[4.136]	[7.248]	[7.382]
Human Capital						
(1) Primary complete Omitted Primay incomplete	1.43	0.76	6.99*	7.60**	3.02	2.61
	[1.086]	[1.082]	[3.577]	[3.516]	[6.500]	[6.566]
(1) High school incomplete	3.02**	2.37**	2.27	3.28	3.54	3.49
	[1.194]	[1.184]	[3.530]	[3.529]	[6.387]	[6.415]
(1) High scool complete	1.89	1.57	4.52	4.5	1.93	1.74
	[1.850]	[1.837]	[5.741]	[5.751]	[9.582]	[9.472]
(1) Some college or more	4.82***	4.31***	-15.06***	-13.26***	-18.55*	-17.2
	[1.581]	[1.568]	[4.961]	[4.934]	[11.016]	[11.088]
Height (cm)	0	-0.02	0.16	0.15	-0.22	-0.23
	[0.059]	[0.059]	[0.198]	[0.199]	[0.359]	[0.365]
Z score Raven's Score	0.17	0.36	-3.79**	-3.48**	-0.21	-0.61
	[0.469]	[0.471]	[1.472]	[1.477]	[2.403]	[2.399]
Networks in the U.S.						
(1) One relative in US Omitted No relatives in US	3.47***		6.44**		10.03*	
	[0.926]		[2.830]		[5.133]	
(1) Two relatives in US	2.55**		16.06***		1.13	
	[1.274]		[3.621]		[6.102]	
(1) Three or more relatives in US	7.21***		15.19***		6.13	
	[1.420]		[3.573]		[5.970]	
(1) Spouse in US		1.95***		5.61***		8.72
		[0.677]		[1.258]		[30.896]
(1) Any parent in US		7.27**		17.37***		5.03
		[3.519]		[5.260]		[9.429]
(1) Daughter/Son in US		0.63		30.15***		21.94**
		[1.541]		[6.554]		[10.284]
(1) Siblings in US		5.12***		11.25***		11.57**
		[1.227]		[3.264]		[5.256]
(1) Extended family in US		0.34		2.16		7.71
		[1.155]		[3.598]		[6.037]

Continued on next page

	Moved since 2002 = 100		Moved to US since 2002 =100		Stayed in US = 100	
Variables measured at baseline	(1)	(2)	(3)	(4)	(5)	(6)
HH characteristics						
(1) HH own farm business	-0.63	-0.13	6.31**	5.65*	1.84	0.8
	[0.995]	[0.984]	[3.019]	[3.033]	[5.142]	[5.111]
(1) HH own a house	-3.41***	-3.37***	-1.68	-1.76	-0.53	-0.64
	[0.871]	[0.868]	[2.489]	[2.497]	[4.317]	[4.313]
(1) HH own a non-farm business	-0.05	0.23	5.17	5.50*	5.47	4.27
	[1.052]	[1.048]	[3.278]	[3.259]	[5.130]	[5.162]
(1) Quartile 2 wealth per capita	0.82	1.36	-0.04	-0.04	1.59	0.83
	[1.102]	[1.095]	[3.377]	[3.374]	[5.902]	[5.960]
(1) Quartile 3 wealth per capita	-1.13	-0.33	3.98	3.46	1.46	0.54
	[1.139]	[1.126]	[3.419]	[3.465]	[6.190]	[6.215]
(1) Quartile 4 wealth per capita	-2.44**	-1.65	4.79	4.94	22.01***	20.26***
	[1.219]	[1.206]	[3.900]	[3.863]	[6.361]	[6.437]
HH size	0.69***	0.94***	0.5	0.84	-1.01	-1.03
	[0.233]	[0.233]	[0.561]	[0.559]	[0.937]	[0.939]
# coresident children	-0.46	-1.02***	1.61	1.48	-0.36	-0.54
	[0.356]	[0.355]	[1.085]	[1.070]	[2.001]	[2.011]
(1) Coresident parents	1.9	1.08	4.33	5.14	0.57	1.52
	[1.277]	[1.283]	[3.512]	[3.580]	[6.433]	[6.456]
Locality characteristics						
(1) Rural	-1.72*	-1.54*	12.69***	12.62***	7.46	7.7
	[0.887]	[0.880]	[2.640]	[2.649]	[4.682]	[4.735]
Constant	18.18*	21.80**	18.81	19.18	68.33	70.9
	[10.047]	[9.980]	[33.446]	[33.509]	[59.875]	[60.840]
Observations	**8,808**	**8,808**	**1,437**	**1,437**	**698**	**698**
R- squared	0.134	0.148	0.359	0.362	0.09	0.09
Mean dependent variable	16.31	16.31	48.57	48.57	50.57	50.57

Robust standard errors in brackets

*** p<0.01, ** p<0.05, * p<0.1

Note: Includes State of origin Fixed Effects

Table 17: Baseline Characteristics that Predict Migration Since 2002 - Females

	Moved since 2002 = 100		Moved to US since 2002 =100		Stayed in US = 100	
Variables measured at baseline	(1)	(2)	(3)	(4)	(5)	(6)
Basic demographics						
(1) Age: 20- 24 Omitted 15-19	2.52	2.19	2.44	2.18	9.63	8.68
	[1.646]	[1.647]	[3.551]	[3.579]	[6.840]	[6.794]
(1) Age: 25-34	-5.50***	-6.41***	-0.47	-3.12	9.12	6.48
	[1.524]	[1.526]	[4.042]	[3.983]	[8.513]	[8.889]
(1) Age: 35-49	-10.35***	-11.35***	1.69	-3.33	-10.43	-14.14
	[1.580]	[1.581]	[4.543]	[4.555]	[10.503]	[11.163]
(1) Age: >50	-11.40***	-11.82***	-10.76**	-16.75***	-17.18	-19.3
	[1.693]	[1.694]	[5.046]	[5.240]	[14.027]	[15.428]
(1) Married	-0.42	0.65	-5.14	-5.22	-0.76	-0.39
	[1.363]	[1.364]	[3.358]	[3.280]	[8.230]	[8.168]
Human Capital						
(1) Primary complete Omitted Primay incomplete	2.17**	2.06**	-0.37	0.16	6.76	5.04
	[0.933]	[0.925]	[3.543]	[3.515]	[8.448]	[8.642]
(1) High school incomplete	2.49**	2.59**	2.47	3.33	7.28	5.78
	[1.078]	[1.075]	[3.788]	[3.775]	[8.671]	[8.937]
(1) High scool complete	2.94*	3.29*	6.94	5.97	15	13.88
	[1.738]	[1.719]	[5.496]	[5.511]	[10.653]	[10.784]
(1) Some college or more	3.80**	4.10***	-7.38	-4.04	-1.52	-1.61
	[1.555]	[1.555]	[5.082]	[5.025]	[13.273]	[13.633]
Height (cm)	0.11*	0.10*	0.28	0.31*	-0.56	-0.5
	[0.058]	[0.058]	[0.175]	[0.173]	[0.437]	[0.444]
Z score Raven's Score	0.96**	0.97**	-0.85	-1.11	-4.77	-4.35
	[0.412]	[0.412]	[1.352]	[1.342]	[2.962]	[2.926]
Networks in the U.S.						
(1) One relative in US Omitted No relatives in US	1.58**		12.56***		0.81	
	[0.801]		[2.807]		[7.675]	
(1) Two relatives in US	3.94***		18.27***		-2.72	
	[1.168]		[3.881]		[8.390]	
(1) Three or more relatives in US	7.01***		29.56***		-7.32	
	[1.263]		[3.830]		[8.253]	
(1) Spouse in US		1.01		2.75**		0.85
		[0.675]		[1.359]		[3.694]
(1) Any parent in US		11.78***		12.41**		3.37
		[3.085]		[6.113]		[7.900]
(1) Daughter/Son in US		1.86		42.92***		7.94
		[1.203]		[7.112]		[12.086]
(1) Siblings in US		4.80***		21.92***		7.71
		[1.071]		[3.379]		[5.747]
(1) Extended family in US		1.47		6.37*		-5.01
		[1.109]		[3.438]		[7.637]

Continued on next page

	Moved since 2002 = 100		Moved to US since 2002 =100		Stayed in US = 100	
Variables measured at baseline	(1)	(2)	(3)	(4)	(5)	(6)
HH characteristics						
(1) HH own farm business	-1.31	-1	12.25***	11.18***	9.77	7.19
	[0.875]	[0.864]	[3.538]	[3.566]	[6.321]	[6.307]
(1) HH own a house	-4.20***	-3.84***	8.29***	7.39***	0.6	0.4
	[0.791]	[0.790]	[2.455]	[2.459]	[5.546]	[5.564]
(1) HH own a non-farm business	0.05	-0.02	1.5	2.04	10.83*	10.15
	[0.949]	[0.936]	[3.247]	[3.239]	[6.277]	[6.296]
(1) Quartile 2 wealth per capita	-0.74	-0.45	-2.68	-2.4	5.01	4.22
	[0.983]	[0.977]	[3.273]	[3.246]	[7.427]	[7.459]
(1) Quartile 3 wealth per capita	-1.45	-1.34	2.04	2	9.9	7.77
	[1.039]	[1.032]	[3.421]	[3.367]	[7.811]	[7.762]
(1) Quartile 4 wealth per capita	-0.81	-0.64	1.88	2.71	21.60**	20.03**
	[1.116]	[1.111]	[3.866]	[3.782]	[8.694]	[8.461]
HH size	0.22	0.35*	-0.71	0	-1.32	-1.05
	[0.210]	[0.208]	[0.610]	[0.607]	[1.266]	[1.274]
# coresident children	-0.38	-0.63**	-0.79	-0.66	-1.47	-1.5
	[0.316]	[0.311]	[1.136]	[1.113]	[2.829]	[2.945]
(1) Coresident parents	3.15**	2.72**	-1.48	-1.44	-7.09	-7.44
	[1.236]	[1.227]	[3.385]	[3.361]	[7.046]	[7.218]
Locality characteristics						
(1) Rural	-1.57**	-1.65**	6.74**	6.48**	7.36	6.1
	[0.790]	[0.780]	[2.673]	[2.666]	[6.303]	[6.511]
Constant	1.07	1.28	-28.76	-33.95	142.23**	131.51*
	[9.107]	[9.034]	[26.882]	[26.459]	[71.503]	[72.836]
Observations	**10,046**	**10,046**	**1,432**	**1,432**	**409**	**409**
R- squared	0.102	0.118	0.236	0.25	0.17	0.18
Mean dependent variable	14.25	14.25	28.56	28.56	68.46	68.46

Robust standard errors in brackets

*** p<0.01, ** p<0.05, * p<0.1

Note: Includes State of origin Fixed Effects

Table 18: Baseline Characteristics Predictors of Assimilation in U.S. - MxFLS2

Dependent variable US Ln(Hourly Earnings) measured in MxFLS2	All		Returners		Stayers	
Variables at baseline	(1)	(2)	(3)	(4)	(5)	(6)
Basic demographics						
(1) Age: 15-19 Omiited Age<15	0.21*	0.17	0.18	0.08	0.2	0.1
	[0.107]	[0.137]	[0.169]	[0.224]	[0.146]	[0.196]
(1) Age: 20-24	0.27**	0.23	0.33	0.27	0.24	0.14
	[0.124]	[0.146]	[0.222]	[0.272]	[0.159]	[0.195]
(1) Age: 25-34	0.33**	0.28*	0.31	0.23	0.3	0.2
	[0.140]	[0.165]	[0.215]	[0.289]	[0.193]	[0.220]
(1) Age: 35-49	0.27	0.29	0.17	0.24	0.32	0.29
	[0.167]	[0.187]	[0.269]	[0.310]	[0.232]	[0.262]
(1) Age: >50	-0.21	-0.12	-1.11	-0.96	0.66*	0.59*
	[0.405]	[0.363]	[0.714]	[0.646]	[0.348]	[0.337]
(1) Female	-0.30***	-0.33***	-0.34**	-0.38**	-0.30***	-0.32***
	[0.082]	[0.088]	[0.152]	[0.168]	[0.095]	[0.102]
(1) Married	0.01	-0.02	0.09	-0.02	0.04	0.02
	[0.085]	[0.087]	[0.165]	[0.183]	[0.092]	[0.095]
Human capital						
Mother's years of education	0.02**	0.02**	0.05**	0.05**	0.01	0.01
	[0.009]	[0.009]	[0.019]	[0.020]	[0.013]	[0.012]
Father's years of education	-0.01	-0.01	-0.03*	-0.03*	0	0
	[0.009]	[0.008]	[0.015]	[0.015]	[0.013]	[0.012]
(1) Primary complete Omitted Primay incomplete	-0.03	-0.04	-0.12	-0.04	0.06	0.04
	[0.083]	[0.079]	[0.143]	[0.133]	[0.100]	[0.103]
(1) High school incomplete	0.08	0.08	-0.03	0.03	0.20**	0.21**
	[0.079]	[0.079]	[0.150]	[0.150]	[0.094]	[0.093]
(1) High scool complete	-0.05	-0.06	-0.22	-0.23	0.1	0.07
	[0.108]	[0.109]	[0.232]	[0.221]	[0.140]	[0.145]
(1) Some college or more	0.18	0.17	-0.01	-0.02	0.39*	0.40**
	[0.141]	[0.134]	[0.204]	[0.207]	[0.201]	[0.190]
Height (cm)	0	0	-0.01	-0.01	0	0
	[0.004]	[0.004]	[0.007]	[0.007]	[0.006]	[0.006]
Z score Raven's Score	0.02	0.01	0.12*	0.10*	-0.04	-0.05
	[0.030]	[0.030]	[0.061]	[0.058]	[0.039]	[0.038]
Networks in US						
(1) One relative in US Omitted No relatives in US	-0.02		-0.03		0.1	
	[0.068]		[0.130]		[0.080]	
(1) Two relatives in US	-0.08		-0.19		0	
	[0.086]		[0.148]		[0.111]	
(1) Three or more relatives in US	-0.05		-0.12		0.04	
	[0.083]		[0.158]		[0.098]	
(1) Spouse in US		0.30**		0.04		0.37**
		[0.146]		[0.259]		[0.187]
(1) Any parent in US		-0.06		0.01		-0.15
		[0.083]		[0.168]		[0.102]
(1) Daughter/Son in US		-0.39*		-0.43		-0.31*
		[0.224]		[0.487]		[0.189]

Continued on next page

Dependent variable US Ln(Hourly Earnings) measured in MxFLS2						
	All		**Returners**		**Stayers**	
Variables at baseline	(1)	(2)	(3)	(4)	(5)	(6)
(1) Siblings in US		-0.01		-0.16		0.04
		[0.067]		[0.136]		[0.084]
(1) Extended family in US		-0.20**		-0.39**		-0.11
		[0.103]		[0.180]		[0.116]
Year of arrival to US	0.01	0.01	0	0	0.02**	0.02**
	[0.009]	[0.008]	[0.015]	[0.014]	[0.008]	[0.008]
HH characteristics						
(1) Quartile 2 wealth per capita	-0.08	-0.09	-0.14	-0.15	-0.11	-0.09
	[0.058]	[0.060]	[0.113]	[0.113]	[0.079]	[0.082]
(1) Quartile 3 wealth per capita	-0.14*	-0.15*	-0.43***	-0.46***	0	0.01
	[0.082]	[0.083]	[0.145]	[0.147]	[0.107]	[0.105]
(1) Quartile 4 wealth per capita	-0.05	-0.08	-0.01	-0.09	-0.1	-0.12
	[0.074]	[0.076]	[0.156]	[0.168]	[0.096]	[0.096]
HH size	0	0	-0.01	-0.01	0.01	0
	[0.010]	[0.011]	[0.018]	[0.018]	[0.014]	[0.015]
Locality characteristics						
(1) Rural	-11.44	-12.42	12.43	12.69	-32.44*	-36.58**
	[18.526]	[17.157]	[31.001]	[28.703]	[16.705]	[15.799]
Constant	2.05	0.82	22.7	22.84	-18.11	-21.73
	[16.138]	[14.603]	[28.931]	[23.125]	[15.028]	[13.976]
Observations	**485**	**485**	**195**	**195**	**290**	**290**
R-squared	0.18	0.206	0.371	0.407	0.223	0.249
Mean dependent variable	1.89	1.89	1.86	1.86	1.92	1.92

Robust standard errors in brackets

*** p<0.01, ** p<0.05, * p<0.1

Note: Includes State of origin Fixed Effects

Table 19: Assimilation of Current U.S. Migrants - Economic Variables

	Ln(Hourly Earnings)		Ln(PCE)		Sent Transfers to MX=100	
Variables measured at baseline	(1)	(2)	(3)	(4)	(5)	(6)
Basic demographics						
(1) Age: 15-19 Omiited Age<15	0.01	0.06	-0.1	0.11	-5.31	2.17
	[0.203]	[0.208]	[0.262]	[0.288]	[10.421]	[11.201]
(1) Age: 20-24	0.08	0.08	0.1	0.28	-0.3	5.15
	[0.215]	[0.215]	[0.276]	[0.296]	[11.450]	[12.083]
(1) Age: 25-34	0.25	0.27	-0.1	0.06	4.54	9.79
	[0.212]	[0.210]	[0.252]	[0.270]	[11.628]	[12.079]
(1) Age: 35-49	0.2	0.18	0.17	0.35	13.83	19.04
	[0.252]	[0.253]	[0.275]	[0.289]	[13.102]	[13.558]
(1) Age: >50	0.13	0.04	-0.11	0.11	-19.91	-12.29
	[0.286]	[0.297]	[0.313]	[0.325]	[14.546]	[15.793]
(1) Female	-0.23**	-0.21**	-0.36***	-0.38***	-15.51***	-14.76***
	[0.100]	[0.100]	[0.077]	[0.078]	[3.845]	[3.910]
(1) Married	0.07	0.1	-0.27***	-0.31***	-8.44	-8.42
	[0.126]	[0.136]	[0.102]	[0.108]	[6.493]	[6.722]
Human capital						
Mother's years of education	0	0	0	0	-1.57***	-1.54***
	[0.013]	[0.013]	[0.011]	[0.011]	[0.572]	[0.580]
Father's years of education	0.01	0.01	0.01	0.01	1.09*	1.15**
	[0.012]	[0.012]	[0.011]	[0.011]	[0.565]	[0.576]
(1) *Primary complete Omitted Primay incomplete	0.25**	0.29**	0.09	0.06	-2.1	-2.46
	[0.113]	[0.116]	[0.107]	[0.106]	[6.004]	[6.159]
(1) High school incomplete	0.26**	0.28**	0.11	0.09	6.77	6.64
	[0.120]	[0.121]	[0.105]	[0.102]	[5.971]	[6.182]
(1) High scool complete	0.24*	0.28*	0.28**	0.27**	4.72	3.81
	[0.144]	[0.143]	[0.128]	[0.127]	[7.763]	[7.915]
(1) Some college or more	0.25	0.26*	0.29*	0.25*	8.03	6.27
	[0.156]	[0.159]	[0.148]	[0.147]	[8.179]	[8.365]
(1) Attended school in US	0.04	0.07	0.02	0.04	-32.02***	-28.92***
	[0.142]	[0.140]	[0.126]	[0.130]	[6.481]	[6.544]
Attained height (cm)	0	0	0	0	-0.08	-0.12
	[0.005]	[0.005]	[0.004]	[0.004]	[0.239]	[0.241]
Z score Raven's Score	0.04	0.04	0.01	0.01	0.98	0.76
	[0.035]	[0.034]	[0.034]	[0.033]	[1.765]	[1.753]
Networks in US						
(1) One relative in US Omitted No relatives in US	0.02		-0.21***		-6.5	
	[0.083]		[0.076]		[4.385]	
(1) Two relatives in US	0.26**		-0.08		-6.37	
	[0.108]		[0.083]		[5.022]	
(1) Three or more relatives in US	0.13		-0.20**		-2.07	
	[0.089]		[0.084]		[4.792]	
(1) Spouse in US		-0.19		0.11		-4.86
		[0.203]		[0.144]		[9.429]
(1) Any parent in Us		-0.05		-0.21*		-16.02*
		[0.118]		[0.110]		[8.690]
(1) Daughter/Son in US		0.25		-0.17		-6.06
		[0.180]		[0.187]		[9.699]

Continued on next page

	Ln(Hourly Earnings)		Ln(PCE)		Sent Transfers to MX=100	
Variables measured at baseline	(1)	(2)	(3)	(4)	(5)	(6)
(1) Siblings in US		0.16*		-0.05		1.44
		[0.080]		[0.071]		[4.522]
(1) Extended family in US		0.33***		-0.11		-3.26
		[0.111]		[0.094]		[5.663]
Year of arrival to US	0.01	0.01	0.01	0	0.49*	0.50*
	[0.007]	[0.007]	[0.005]	[0.005]	[0.265]	[0.264]
HH characteristics						
(1) HH own farm business	-0.07	-0.07	-0.09	-0.07	-5.38	-4.94
	[0.090]	[0.089]	[0.068]	[0.068]	[3.909]	[3.897]
(1) HH own a house	0.02	0.04	-0.02	-0.02	-5.17	-5.03
	[0.071]	[0.072]	[0.057]	[0.058]	[3.314]	[3.305]
(1) HH own a non-farm business	0	-0.01	0.01	0.01	1.38	1.91
	[0.082]	[0.083]	[0.071]	[0.072]	[4.350]	[4.366]
(1) Quartile 2 wealth per capita	-0.09	-0.09	0	0	-0.18	0.65
	[0.090]	[0.089]	[0.079]	[0.079]	[4.252]	[4.207]
(1) Quartile 3 wealth per capita	-0.28***	-0.28***	0.14*	0.16*	-0.53	-0.64
	[0.103]	[0.105]	[0.084]	[0.086]	[4.788]	[4.754]
(1) Quartile 4 wealth per capita	-0.07	-0.07	0.09	0.08	-4.7	-4.68
	[0.110]	[0.110]	[0.092]	[0.092]	[5.658]	[5.577]
HH size	0.02	0.01	-0.02	-0.02	1.27	0.85
	[0.014]	[0.015]	[0.014]	[0.014]	[0.807]	[0.813]
# coresident children	-0.04	-0.03	0.03	0.03	-0.68	-0.55
	[0.036]	[0.037]	[0.031]	[0.031]	[1.798]	[1.861]
(1) Coresident parents	0.13	0.12	0.05	0	-1.45	-2.14
	[0.111]	[0.116]	[0.094]	[0.099]	[5.992]	[6.105]
Locality characteristics						
(1) Rural	0.08	0.07	0.05	0.03	2.94	2.77
	[0.089]	[0.088]	[0.073]	[0.072]	[4.014]	[3.970]
Constant	-12.75	-12.35	-5.6	-3.53	-873.84	-889.32*
	[13.717]	[13.850]	[10.594]	[10.797]	[532.342]	[530.545]
Observations	**569**	**569**	**568**	**568**	**828**	**828**
R-squared	0.125	0.131	0.234	0.235	0.238	0.247

Robust standard errors in brackets

*** p<0.01, ** p<0.05, * p<0.1

Note: Includes State of origin Fixed Effects

* Note2: Education is attained education in MxFLS3

Table 20: Assimilation of Current U.S. Migrants - Use of English

Dept variable = 100 if speaks English Fair, Good or Very Good		
Variables measured at baseline	(1)	(2)
Basic demographics		
(1) Age: 15-19 Omiited Age<15	6.07	1.13
	[10.504]	[10.971]
(1) Age: 20-24	-3.57	-6.86
	[11.533]	[11.915]
(1) Age: 25-34	6.13	2.62
	[11.572]	[11.851]
(1) Age: 35-49	-7.12	-11.88
	[12.549]	[12.912]
(1) Age: >50	-24.87*	-31.47**
	[12.674]	[13.753]
(1) Female	-8.70**	-8.57**
	[4.013]	[4.005]
(1) Married	4.67	5.13
	[6.765]	[6.991]
Human capital		
Mother's years of education	1.66***	1.67***
	[0.613]	[0.619]
Father's years of education	0.2	0.32
	[0.639]	[0.651]
(1) *Primary complete Omitted Primay incomplete	2.11	2.23
	[5.613]	[5.820]
(1) High school incomplete	10.68*	11.03*
	[5.710]	[5.902]
(1) High scool complete	13.99*	14.83**
	[7.255]	[7.388]
(1) Some college or more	29.45***	31.62***
	[7.926]	[8.206]
(1) Attended school in US	44.15***	41.57***
	[5.288]	[5.354]
Attained height (cm)	0.16	0.19
	[0.237]	[0.238]
Z score Raven's Score	3.53*	3.65*
	[1.948]	[1.947]
Networks in US		
(1) One relative in US Omitted No relatives in US	-1.12	
	[4.774]	
(1) Two relatives in US	2.65	
	[5.307]	
(1) Three or more relatives in US	1.25	
	[5.148]	
(1) Spouse in US		-3.19
		[7.664]
(1) Any parent in Us		11.44
		[8.369]
(1) Daughter/Son in US		6.4
		[8.500]

Continued on next page

Dept variable = 100 if speaks English Fair, Good or Very Good		
Variables measured at baseline	(1)	(2)
(1) Siblings in US		-0.95
		[4.832]
(1) Extended family in US		-7.23
		[6.118]
Year of arrival to US	-0.61***	-0.63***
	[0.236]	[0.235]
HH characteristics		
(1) HH own farm business	-5.88	-5.34
	[4.206]	[4.177]
(1) HH own a house	-0.52	-1.4
	[3.483]	[3.502]
(1) HH own a non-farm business	-8.33**	-9.02**
	[4.087]	[4.032]
(1) Quartile 2 wealth per capita	-4.97	-5.14
	[4.722]	[4.712]
(1) Quartile 3 wealth per capita	-0.56	0.22
	[4.897]	[4.859]
(1) Quartile 4 wealth per capita	-2.6	-1.91
	[5.586]	[5.456]
HH size	-0.95	-0.55
	[0.847]	[0.866]
# coresident children	-2.79	-2.92*
	[1.755]	[1.695]
(1) Coresident parents	3.16	3.81
	[5.803]	[5.977]
Locality characteristics		
(1) Rural	-1.73	-2.11
	[4.071]	[4.098]
Constant	1,237***	1,263***
	[475]	[472]
Observations	**829**	**829**
R-squared	0.325	0.331

Robust standard errors in brackets

*** p<0.01, ** p<0.05, * p<0.1

Note: Includes State of origin Fixed Effects

* Note2: Education is attained education in MxFLS3

Table 21: Assimilation of Current U.S. Migrants - Relatives in U.S.

	MALES				FEMALES			
	IN HH Spouse		Children in US		IN HH Spouse		Children in US	
Variables measured at baseline	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Basic demographics								
(1) Age: 15-19 Omiited Age<15	6.01	-2.32	9.26	-28.71	-1.67	-0.51	0.14	-0.59
	[30.579]	[35.966]	[25.207]	[30.218]	[3.590]	[2.832]	[6.587]	[6.134]
(1) Age: 20-24	4.46	-3.74	13.96	-25.62	-4.16	-3.64	-7.44	-7.34
	[33.153]	[38.043]	[28.071]	[32.684]	[3.590]	[3.302]	[7.680]	[6.944]
(1) Age: 25-34	2.25	-5.12	15.67	-25.92	-4.22	-3.65	6.38	5.06
	[33.593]	[37.364]	[29.076]	[32.753]	[6.643]	[6.303]	[9.365]	[7.833]
(1) Age: 35-49	-17.12	-22.89	11.63	-27.07	-5.46	-4.03	-4.86	-0.92
	[34.341]	[38.719]	[29.387]	[33.615]	[8.268]	[6.469]	[9.944]	[9.017]
(1) Age: >50	2.45	-1.85	39.32	0.7	4.12	9.44	-2.7	1.14
	[33.185]	[37.381]	[28.339]	[32.081]	[6.348]	[7.710]	[10.591]	[10.501]
(1) Married	-13.3	-14.31	-15.61	-11.23	-2.85	-2.98	2.11	1.5
	[12.224]	[12.243]	[13.185]	[14.555]	[3.546]	[3.886]	[6.325]	[5.805]
Human capital								
Mother's years of education	-0.19	-0.16	0.53	0.27	-0.21	-0.25	0.47	0.58
	[1.465]	[1.511]	[1.142]	[1.178]	[0.215]	[0.245]	[0.573]	[0.540]
Father's years of education	-0.48	-0.31	-0.46	0.07	0.02	0.11	-0.73	-0.64
	[1.181]	[1.238]	[1.380]	[1.423]	[0.256]	[0.264]	[0.698]	[0.638]
(1) *Primary complete Omitted Primay incomplete	-0.04	1.43	16.81	19.82	2.19	1.35	-4.76	-5.31
	[13.710]	[13.931]	[12.500]	[12.554]	[4.915]	[4.280]	[5.012]	[4.603]
(1) High school incomplete	2.94	5.56	11.89	18.52	2.7	1.48	-8.46	-7.86
	[13.521]	[13.470]	[12.238]	[12.122]	[4.679]	[3.631]	[5.868]	[5.511]
(1) High scool complete	-14.28	-11.61	1.67	8.87	4.58	2.82	-8.1	-7.55
	[18.747]	[18.937]	[16.151]	[16.359]	[4.851]	[4.042]	[6.660]	[6.324]
(1) Some college or more	-16.2	-15.54	7.06	11.88	4.46	2.78	-24.27**	-22.78**
	[18.829]	[18.856]	[16.065]	[16.667]	[6.351]	[4.428]	[10.092]	[9.923]
(1) Attended school in US	-14.96	-13.27	7.32	4.16	0.59	1.43	0.73	0.58
	[17.056]	[17.586]	[13.944]	[14.630]	[2.993]	[3.158]	[4.729]	[4.299]
Attained height (cm)	0.96	0.92	0.1	0.08	-0.21	-0.16	-0.31	-0.28
	[0.630]	[0.649]	[0.640]	[0.646]	[0.231]	[0.204]	[0.226]	[0.197]
Z score Raven's Score	3.89	3.97	1	2.91	0	-0.08	1.32	1.41
	[4.697]	[4.606]	[3.899]	[4.101]	[0.708]	[0.674]	[1.724]	[1.651]
Networks in US								
(1) One relative in US Omitted No relatives in US	-0.04		21.99**		0.02		0.12	
	[8.669]		[9.300]		[2.624]		[4.325]	
(1) Two relatives in US	5.31		25.83**		-1.76		-4.12	
	[11.177]		[11.911]		[2.800]		[5.523]	
(1) Three or more relatives in US	10.38		29.11**		-1.71		0.57	
	[10.465]		[11.248]		[1.885]		[4.730]	
(1) Spouse in US				20.53				2.43
				[18.871]				[3.400]
(1) Any parent in Us		-1.12		-0.96		-0.39		1.48***
		[1.448]		[1.389]		[0.385]		[0.549]
(1) Daughter/Son in US		-5.54				-4.12		
		[10.780]				[4.990]		

Continued on next page

	MALES				FEMALES			
	IN HH Spouse		Children in US		IN HH Spouse		Children in US	
Variables measured at baseline	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
(1) Siblings in US		8.5		6.04		4.26		7.17**
		[8.129]		[6.506]		[2.632]		[3.201]
(1) Extended family in US		-2.08		-2.79		-0.52		-6.81**
		[10.413]		[6.308]		[3.536]		[3.431]
Year of arrival to US	-1.25**	-1.10**	-0.67	-0.5	0.19	0.26	-0.39	-0.4
	[0.534]	[0.537]	[0.439]	[0.445]	[0.282]	[0.311]	[0.258]	[0.268]
HH characteristics								
(1) HH own farm business	17.71*	18.71*	18.22**	17.81**	-2.83	-3.98	4.47	3.07
	[10.273]	[10.320]	[8.134]	[8.307]	[3.492]	[3.061]	[3.901]	[3.634]
(1) HH own a house	11.98	10.51	5.48	5.33	0.79	0.14	1.38	0.99
	[8.941]	[9.597]	[7.706]	[7.595]	[2.981]	[2.673]	[3.204]	[3.086]
(1) HH own a non-farm business	-4.58	-4.92	-1.56	-2.81	1.67	1	1.12	-0.12
	[9.174]	[9.700]	[9.320]	[9.602]	[1.488]	[1.369]	[4.197]	[3.908]
(1) Quartile 2 wealth per capita	-26.15***	-25.19**	-18.64*	-19.30**	4.55	4.87	5.3	5.29
	[9.924]	[9.918]	[9.676]	[9.579]	[3.967]	[4.010]	[4.043]	[3.780]
(1) Quartile 3 wealth per capita	-22.71**	-21.22**	-8.09	-9.68	6.09	5.95	4.12	6.09*
	[10.647]	[10.697]	[10.264]	[10.446]	[4.114]	[3.874]	[3.595]	[3.622]
(1) Quartile 4 wealth per capita	8.97	9.97	10.16	7.89	7.09	7.24	-0.26	-0.04
	[12.746]	[12.887]	[11.593]	[12.047]	[4.541]	[4.422]	[4.710]	[4.661]
HH size	2.86	2.99	2.35	2.67	0.15	0.18	-0.42	0.05
	[2.363]	[2.291]	[2.142]	[2.122]	[0.640]	[0.618]	[0.760]	[0.701]
# coresident children	-8.13**	-7.73**	-1.9	-1.49	1.25	1.15	-2.68	-3.05
	[3.428]	[3.585]	[3.291]	[3.169]	[0.948]	[0.929]	[2.668]	[2.232]
(1) Coresident parents	-3.34	-1.88	-8.67	-8.21	2.83	1.71	7.03*	6.69
	[11.893]	[12.113]	[10.112]	[10.535]	[2.669]	[1.972]	[4.033]	[4.293]
Locality characteristics								
(1) Rural	-3.55	-3.04	-6.56	-1.88	2.14	2.14	-4.48	-3.31
	[8.458]	[8.471]	[8.919]	[9.206]	[3.367]	[3.512]	[4.111]	[3.826]
Constant	2,422.77**	2,139.94*	1361.56	1077.23	-260.27	-407.66	933.48*	936.89*
	[1,076.599]	[1,087.328]	[865.599]	[885.383]	[538.094]	[599.133]	[522.708]	[543.083]
Observations	**189**	**189**	**226**	**226**	**218**	**218**	**259**	**259**
R-squared	0.436	0.437	0.312	0.294	0.144	0.174	0.314	0.38

Robust standard errors in brackets

*** p<0.01, ** p<0.05, * p<0.1

Note: Includes State of origin Fixed Effects

* Note2: Education is attained education in MxFLS3

4. The Economic Burden of Crime: Evidence from Mexico

4.1 Introduction

Since 2007, drug-related crime in Mexico has increased at an unprecedented pace and intensity. Between 2007 and 2011 the official homicide rate reported by the National Institute of Statistics and Geography (INEGI, its Spanish acronym) exploded from 8.5 to 24.4 per 100,000 inhabitants (Figure 7 provides the monthly homicide rate in Mexico from 2002 to 2011). Moreover, not only has the homicide rate increased in the last few years, but the characteristics of the violence have also dramatically changed. The amplified effort against Organized Crime Groups (OCGs) taken by Felipe Calderón's government had unanticipated consequences for the intensification and geographical dispersion of violence in Mexico. One of the main unexpected effects of the new military strategy, for example, was the growth in the number of OCGs from 6 in 2007 to 16 in 2010 (Guerrero, 2012a).

The loss in leadership within the drug cartels due to the death or incarceration of the previous bosses created power vacuums that have spawned the emergence of new leaders and have increased the struggle for dominance over critical drug running corridors.[1] As these competitions intensified, the nature of crimes has also rapidly

[1] The term "drug cartel" is a colloquial term to refer to organized crime organizations but it does not imply any collusion to set prices. I will use the term Organized Crime Groups (OCGs), traffickers' organizations and "cartels" indistinctively to refer to organized crime organizations involved in the drug-trafficking business.

changed. For example, the strategic placement of dead corpses has become one useful tool to spread fear and send messages to rival traffickers, authorities, and citizens (Rios, 2012). Moreover, in addition to changing the violent crime environment, the fracturing of the cartels has reduced the profits each individual OCG makes from drug trafficking and has pushed these groups to more aggressively pursue criminal enterprises that target non-combatants.[2]

Research suggests that an environment of elevated crime, such as the current Mexican environment, severely alters the context in which people operate and drastically impacts individual and household behaviors.[3] In particular, it can have an effect on labor markets, both on the demand and the supply side (Bozzoli et al., 2011; Calderón et al., 2011; Deininger, 2003; Fernández et al., 2011; Kondylis, 2007; Shemyakina, 2011). On the supply side, labor may diminish when violence increases if the fear of being victimized increases the cost of labor participation or if the outside option of drug-cartel membership reduces the relative utility of participation in the legal labor force. On the other hand, it may increase the labor supply if high levels of violence

[2] There is evidence that increasing conflict in Mexico has led to an expansion of crimes targeting citizens, such as, extortions, kidnapping, as well as, car and business thefts, with extortions increasing from 3 per 100,000 inhabitants in 2006 to 5.88 in 2009, business thefts from 54.42 in 2006 to 63.32 in 2009, (72.3 in 2011) (Robles et al., 2013) and executions by 60% between 2009 to 2010 (from 9,604 to 15,263) (Guerrero, 2012a).

[3] A rich strand of the literature has shown evidence of the effects of high levels of crime on health outcomes (Akresh et al, 2012; Baez, 2011; Bundervoet et al. 2009), human capital investment (Barrera and Ibáñez, 2004; Leon, 2012; Rodriguez and Sanchez, 2009; Shemyakina, 2010) and asset depletion and consumption (Justino and Verwimp, 2006; Ibáñez and Moya, 2010).

diminish the earnings of the main bread-winner and other members of the household have to join the labor force - this is known as the added-worker effect in the literature (Lundberg, 1986; Cunningham, 2001; Calderon et al. 2011). Labor demand is also likely to be impacted by violence, and as with supply, the direction of the effect is unclear. On the negative side, investment at the firm level may contract due to the added costs imposed by crime (Collier and Duponchel, 2013; Pshisva and Suarez, 2010; Camacho and Rodriguez, 2013), which would have a negative effect on availability of formal employment opportunities as some businesses close, reduce their size, or choose not to enter the market. Alternatively, increased presence of drug-cartel members, and thus their disposable income, may increase economic activity, leading to an increase on labor demand.

The purpose of this chapter is to quantify the impact of the increasing wave of crime in Mexico on individual labor market outcomes and household consumption. Due to the multidimensional impact of violence on labor market outcomes, the effect may differ significantly depending on the employment sector of the individual. For example, a reduction in hours worked may be a successful strategy to decrease the probability of being victimized. However, this is a strategy that is more easily implemented by self-employed individuals as they have more flexibility and control of their schedules. To shed light on the potential heterogeneity of the effect depending on type of employment,

I will identify the impact of violence separately for self-employed and wage employees, as well as occupation groups within these classes.

In addition, this chapter makes a contribution to the literature on conflict and gender, by examining specifically at the effects of violence on labor market outcomes of women and per capita expenditure in households with a female household head. Looking at women separately from men is imperative when examining the consequences of violent crime, as it is very likely that they face a different level and type of victimization than men, especially with regard to sexual assault (USAID, 2007). Moreover, in the specific case of the recent conflict in Mexico, evidence suggests that there has been an increase in violence specifically targeted towards women (Pantaleo, 2010). Additionally, since it is likely that the violence has heterogeneous effects on different job types, and since even within the same employment sector, men and women specialize in different types of occupations, separating the analysis by gender within occupation categories provides a richer understanding of the impact of violence on the labor market.

High levels of violence can create long-lasting impacts on the well-being of civilians. In particular, when labor markets are affected, households that do not have the capacity to smooth negative income shocks can end up in poverty traps, with the next generation's future welfare being affected by the lower household's income, lower consumption, and lower human capital investment. Advancing the understanding of the

effect of high levels of violence and crime is a first order question for governments that need to provide the adequate assistance to the victims of these circumstances.

However, assessing the effect of violence on economic outcomes imposes several empirical challenges. The major contribution of this work is the rigorous inclusion of and accounting for systematic migration, the use of respondent fixed effects, and the exploration of the heterogeneous impact of conflict on individual labor market outcomes and household per capita expenditure. An individual fixed effect strategy combined with an intent-to-treat approach controls for unobserved time-invariant heterogeneity that affects exposure to crime and individual outcomes while shielding the estimates from potential systematic migration as a behavioral response to crime. Moreover, unlike the previous literature on this subject I am able to explicitly analyze migration as a behavioral response to violence. In addition, by assessing the effect of violence on a broad range of outcomes, rather than just employment status and total earnings, the analysis provides a more comprehensive exploration of the labor market and welfare effects of crime. This chapter's analysis of labor market outcomes utilizes measures of participation in the labor market, total earnings, hours worked, hourly earnings and is complemented by a detailed analysis of per capita expenditure. The results confirm the importance of these advancements to this literature, as this chapter finds evidence of systematic migration as a response to violence, significant changes in hours worked,

hourly earnings, and per capita expenditure, as well as heterogeneous labor market effects by gender and occupation type.

To conduct these analyses I exploit available information on homicide rates from the INEGI at the municipality level from 2005-06 (a period of low levels of violence) to 2009-12 (a period of high levels of violence). In order to analyze the impact of violence on the outcomes of interest, the INEGI homicide data is matched, at the level of the municipality, to the Mexican Family Life Survey (MxFLS), a nationally representative longitudinal survey.

The MxFLS is ideally suited to address the questions of this chapter. One important feature of the survey is that the first follow-up was conducted between 2005 and 2006, during a time of relatively stable levels of conflict, and the second follow-up was performed from 2009 to 2013, when violence in Mexico rose with unprecedented intensity.[4] This feature of the timing, paired with the panel nature of the survey, allows the comparison of outcomes of the same individual in periods of low and high violence. Additionally, the detailed survey information contained in the MxFLS provides a rich set of controls for time-varying individual and household characteristics.

The results suggest that a higher homicide rate negatively affects labor market participation and the number of hours worked by self-employed women; and, there is a

[4] 94% of the sample of panel respondents interviewed in Mexico was interviewed between 2009 and 2010.

great deal of heterogeneity on the labor income of the women who stay in the labor force depending on their occupation and the level of violence to which they are exposed. On the other hand, crime does not affect self-employed males' labor market participation, but does negatively impact their hourly and total earnings. The adverse effect of crime is also evident for measures of per capita expenditure, in particular for households with a self-employed household head, as the results suggest that exposure to conflict has negatively affected consumption levels of these households.

The next section introduces the homicide data available in Mexico, provides background on the Mexican experience of violent crime over the last decade, and describes the INEGI homicide data. Section 3.3 reviews the literature on the effect of crime on labor market outcomes and wealth. Section 4.4 describes the Mexican Family Life Survey (MxFLS) and shows descriptive statistics of the analytical sample. Section 4.5 provides a description of the conceptual framework that motivates the empirical specification that is presented in section 4.6. Section 4.7 shows and discusses the main results. In section 4.8 robustness checks are discussed and section 4.9 concludes.

4.2 Background

The rapid increase in violent homicides has led to a debate about its causes.[5] One hypothesis is that the rapid increase in homicides and the degree of violence surrounding them is a byproduct of the military strategy of increased confrontation against OCGs that took place since late 2006 when Felipe Calderón became president (Molzahn, et al., 2012; Guerrero, 2011a).

Calderón's military strategy against the main "capos" of the cartels destabilized the old oligopolistic equilibrium where the OCGs operated while maintaining relatively low levels of violence. Guerrero (2012b) suggests that by confronting the main leader of a trafficking group two types of violence are created: first, an internal conflict arises for the leadership of the organization; and second, it increases the likelihood of confrontations from enemy organizations that seek to wrest territorial control by taking advantage of the weakened enemy. Rios (2013) describes this new equilibrium as a "self-reinforcing violent equilibrium" where the violent confrontations between traffickers increase the likelihood of government interventions which promote subsequent confrontations between the remaining traffickers to control the market left by their predecessors. Related to this debate, Dell (2011) compares municipalities where a mayor from Calderon's party (Partido Acción Nacional –PAN) won the election by a margin of

[5] Castillo et al. (2013); Dell (2011); Guerrero (2011a); Guerrero (2012b); Molzahn et al. (2012); Rios and Shirk (2011); Rios (2013); Shirk (2011).

5 percent or less to municipalities in which the PAN barely lost by the same margin. Comparing these municipalities 6 months before the election and 6 months after the election, the author finds a significant increase of drug related homicides in the municipalities where Calderón's party won, suggesting that PAN related policies may have triggered the massive increase in homicides.

This rapid and intense increase of violence is seen in the official homicide rate reported by the INEGI.[6] According to this data source the annual homicide rate has increased in Mexico from 8.5 homicides per 100,000 inhabitants in 2007 to 24.4 in 2011, an increase of almost 200% (Figure 7, red solid line). Comparing the numbers reported by the INEGI with those from the National Public Security System (SNSP, its Spanish acronym), which were collected under the direction of President Felipe Calderón (Figure 7, green dashed line), and compiles homicides exclusively related to organized crime, it is evident that most of the increasing trend of homicides reports by the INEGI is explained by drug-related violence.[7]

---

[6] In an effort to understand the new dynamics of crime in Mexico, in addition to the official numbers of intentional homicides reported by the INEGI, other data sources were assembled to shed light on the puzzling increase of crime in Mexico. For an analysis and comparison of the different data available to measure homicides see Rios and Shirk (2011) and Molzahn, Rios and Shirk (2012).

[7] The data that measures homicides related to organized crime was released by the Mexican government in December of 2010, and provides information of number of homicides "allegedly linked to organized crime" from December 2006 to September of 2011. The numbers from this data set show a rise in the annual homicide rate of more than 550% between 2007 and 2011; from 2.67 per 100,000 inhabitants in 2007 to 12.6 in 2010 and 18 in 2011 (green dashed line in Figure 1 shows monthly homicide rates).

Since 2006, not only has the intensity of executions increased but the geographical concentration has changed as well: in 2010, 168 municipalities reported 12 or more executions, whereas in 2007 only 48 municipalities reported 12 or more executions.[8] Although the INEGI dataset captures all intentional homicide rates and not only drug-related violence, it also shows the geographical dispersion found in the organized crime dataset. Figures 9 and 10 show the homicide rates in 2002 and 2005 before Calderón took office. At that point only a handful of municipalities were at the top of the distribution of homicides and they were highly concentrated in states with strong presence of OCG's. Although the presence of traffickers could increase the levels of violence in general the figures show that before 2007 the levels were relatively low and concentrated in only a few places.

Figures 11 to 13 show the geographical dispersion of the homicide rate from 2007 to 2010. The first year of Calderón's term, 2007, was a relatively stable year in terms of violence, even though his military strategy was already being implemented. The first dramatic increase of violence is observed in 2008 (Figure 8). Figures 12 and 13 show the almost epidemic dispersion of crime: there is a massive increase in violence in previously unaffected municipalities in Chihuahua, Durango, Sinaloa, Michoacán and Guerrero, among others.

[8] The mean of executions in 2012 was 11.76 using the dataset from the SNSP.

Rios (2013) documents the dynamics of war within and between traffickers' organizations after a leader of an OCG is captured. In 2008, for example, the Sinaloa cartel leader was captured which caused that cartel to split and as a consequence created a combative relationship between the Sinaloa Cartel and La Familia from Michoacán. One of the consequences of this was an intensification of violence in Guerrero where La Familia and the Sinaloa Cartel had operated together in previous years. The situation only got worse in 2009 and 2010 (Figures 12 and 13). In 2009 the Sinaloa Cartel split once again and an area that in 2007 had only one cartel operating became the territory of four competing cartels (Rios, 2013; Guerrero, 2011b). The fracturing of cartels intensified the violence in the competing territories and spread the violence as new groups attempted to increase their territorial control. In the meantime, Nuevo León and Tamaulipas suffered the consequences of a turf war between Los Zetas and their former employer the Gulf Cartel. The maps using the INEGI official homicides data show both the intensification of the violence and its geographical dispersion.[9]

In addition to the increased number of confrontations between groups, the rising competition between OCGs has led to a diversification of their financial sources. While drug trafficking activities still account for most of the drug cartels' economic resources,

[9] It is important that the INEGI data accurately reflects the relative levels of violence through Mexico, because unlike the data from other sources, the homicides reported in the INEGI are available from 1990 to 2011, which allows the analyses to include both the pre-violence and high violence periods and thus fully exploit the panel nature of the MxFLS.

in order to increase profits and fund their fight against the military and other OCGs, they have been relying more on criminal activities that directly affect the civil population, like kidnappings, extortions and car thefts. Moreover, the visibility of crimes have changed the dynamics of the violence in Mexico. "Narco-messages" to spread fear in the communities have become a very popular method to signal territorial presence and to spread fear across other OCG's but also toward authorities, journalists and any citizen that would not support their actions.

Moreover, the feeling of fear is exacerbated by the public's lack of trust in the State's institutions and the high levels of corruption and abuses from the police (Guerrero, 2011a; Díaz-Cayeros et al., 2011). Díaz-Cayeros et al. (2011) measure the strategies that the OCGs use against the civil society and measure how embedded they have become in the society. Their results show that both OCGs and police take advantage of citizens, particularly preying on the poor and less educated. By looking at the effects of violence for different occupations and type of employment, this chapter will provide empirical evidence on the heterogeneous effects of crime on labor markets.

In addition, although the main driver of the surge on violence has been the drug-related business, women have also been victims of the intensity of crime. Not only the number of female homicides increased 120% from 2007 to 2012 (INEGI) but also, the violence used in female homicides is more violent than the one used against men. While more than 40% of the cases of homicides of men are deaths caused by a fire arm, female

deaths are more violent - strangulation and the use of a sharp object are the most common incidences of homicide. Moreover, sexual violence against women in Mexico has also increased in the last years, and a study from the United Nations shows that Mexico is the country with the highest percentage of women who have suffered this kind of violence (Echarri Canovas, 2011).

The fear of being victimized can directly affect the decision to participate in the labor market. The effect might be different for men and women if the new dynamics of crime have had a differential effect on the decision to work of women. In particular, in cases where women are not the main bread-winner their opportunity cost of leaving the labor market might be lower. By measuring the effect of crime, separately for men and women, this chapter will provide evidence on these issues.

In order to measure the effect of crime on individual and household outcomes I will use the homicide rates reported by the INEGI. Although homicide rates have been extensively used in the literature as a measure of crime and violence, potential measurement error in the INEGI homicide rate when used as proxy for the measure of violence could bias the empirical estimation.[10] On the one hand, one concern is that since the INEGI data only captures registered homicides this source to measure

[10] Because the final act of a homicide is the violent death of a person, and because of the difficulty related with hiding a body, homicides are less likely to be subject to underestimation or misinterpretation in comparison to crime on property or physical violence; moreover, the classification of an homicide is homogeneous across regional boundaries which diminishes potential measurement error (Shrader, 2001).

homicide rates may reflect a lower bound on the actual level of violence. However, this does not seem to be a limitation, as the figures and maps show, the INEGI data captures the same increasing trend as the data on homicides related to organized crime. On the other hand, the increase in homicides might reflect a combination of actual increases in conflict as well as simply a shift in the ability/motivation to more accurately report homicides over this time period. In that case the effect of homicide rates on the outcomes of interest would be overestimated. Although shifts in reporting patterns could be an important source of measurement error in other scenarios, the increasing violence has not only been evident in the official numbers (INEGI and the SNSP datasets), but also in alternative data sources collected by academics, the national press, and NGOs, all of which have verified the findings of the public records (Molzahn, et al., 2012).[11]

4.3 The Economic Effects of Crime and Violence

Although a number of studies analyze the cause of the increasing violence in Mexico, relatively few studies measure the impact of crime on individual and household outcomes. Dell (2011) makes an important contribution to the literature and exploits a network model of drug-trafficking routes and information from the National Survey of

[11] An additional concern is related with systematic measurement error. This would be the case if, for example, there is larger underreporting in the most violent places, which would be related with non-classical measurement error. However, the close correlation of the INEGI dataset with other datasets that rely on alternative sources, does not seem to suggest non-classical measurement error on the INEGI homicide rates.

Occupation and Employment (ENOE) to measure the economic spillover effects of drug-related violence. Dell's results suggest no effects on male labor participation or wages in the formal sector but in contrast, negative effects on male wages in the informal sector and on female labor force participation.

Two recent working chapters have also measured the effect of crime on labor outcomes using data from the ENOE. The ENOE is a rotating panel where a household is followed during 5 trimesters and is then replaced by a new household. The limitation of this dataset to analyze the effects of violence is that it does not allow the researcher to control for migration as a behavioral response to crime. If, for example, the highest skilled individuals migrate when violence increases, the effects of crime on labor income will be overestimated. By using the MxFLS, a longitudinal survey designed to follow all migrants, I am able to measure migration as a response to crime and to account for it in my empirical specification.

Two working chapters have used the ENOE to examine the effects on labor outcomes. First, the empirical strategy of Robles et al. (2013) relies on an instrumental variable approach, using the variation in cocaine seizures in Colombia interacted with the distance of each municipality to the principal point of entry to the U.S., instrument from Castillo et al. (2013), as an instrument for homicides and find a dampening effect of violence on an individual's labor market participation, and on the proportion of business owners in the population but no effect on self-employment or overall economic

activity (measured as domestic electricity consumption). Second, BenYishai and Pearlman (2013) relies on individual fixed effects and an instrumental variable approach. The IV in this case is the number of kilometers of federal toll highways in each state. Their results suggest that the marginal effect of violence has reduced the average hours worked of both salaried and self-employed males, and for the latter the impact has been larger for those males who work from home. The analysis, though, does not provide evidence on whether the lower labor supply has had any effect on the labor income of Mexican citizens. Moreover, in order to deal with migratory behavior, individuals that were not born in the current state of residence are dropped from the study, which, if migration is selective, leads to a non-random analytical sample.

In addition to the chapters that study the recent violence in Mexico, there are important contributions in the literature that have measured the impact of crime on labor market outcomes and household's economic conditions in other settings (Bozzoli, et al., 2011; Calderón, et al., 2011; Deininger, 2003; Fernández, et al., 2011; Kondylis, 2007; Shemyakina, 2011). Most of these contributions are in settings were individuals have been forcibly displaced by violence and face considerable constraints in the labor markets of the destination cities.

Bozzoli, et al. (2011) study the effects of the Colombian conflict on the probability of being self-employed in rural settings and find a decreasing share of self-employment in municipalities with higher rates of conflict. Calderón, et al. (2011) estimate the effects

of displacement on the labor markets of destination places. The authors find that a larger supply of unskilled labor increases the likelihood of informality and reduces wages in this sector. Kondilys (2007) exploits data from a longitudinal study to explore the effects of displacement on labor market outcomes in post-war Bosnia and Herzegovina. The results show that displaced men experience higher unemployment levels, and displaced women are more likely to drop out of the labor force. Shemyakina (2011) measures the long-term impact of the 1992-1998 armed conflict in Tajikistan on education and labor market outcomes for both men and women. By performing a difference in difference regression the author finds that, for women, the conflict had a negative impact on educational attainment and a positive effect on labor market participation, while it had no effect for the outcomes of men.

Moreover, there is an extended literature analyzing the response of individuals and households to negative shocks. In particular, a number of studies examine the ability of households to smooth consumption in times of crisis (see for example, Townsend, 1995; Morduch, 1995; Frankenberg et al. 2003). If we consider exposure to violence as a negative shock to wealth, as high levels of violence may alter labor market outcomes and thus decrease the level of household income, there is potential for significant consumption smoothing. However, low socioeconomic status individuals may have very few insurance and credit mechanisms in order to smooth consumption. This suggests that the impact of violence on labor outcomes, per capita expenditure

(PCE), and wealth may have meaningful potential to create poverty traps for the most vulnerable population. A second pathway, through which violence may affect consumption behavior directly is through the fear of being targeted for conspicuous consumption. For example, it has been found that individuals in the United States decrease the consumption of visible goods to reduce the probability of victimization from property crime (Mejia and Restrepo, 2010). In this case there is a reallocation of consumption that does not affect total wealth, but can have a negative effect on utility functions at the household and individual levels.

This chapter makes a contribution to the understanding of the economic impact of Mexico's recent outbreak of violence and to the literature on the effects of crime in a number of ways. First, it exploits a robust identification strategy by estimating an individual fixed effect strategy combined with an intent-to-treat approach. The individual fixed effect strategy controls for time-invariant unobserved heterogeneity that affects exposure to crime and the outcomes of interest, and the intent-to-treat approach controls for potential selective migration as a behavioral response to crime. Moreover, I explicitly test the exogeneity of the increasing trend of crime by adding controls of economic performance at the state and municipality level, I estimate a model of the predictability of violence, and I provide evidence from a placebo test regarding the exogeneity of the increasing trend of crime. Second, I directly analyze migration as a response to violence and I show evidence of selective migration as a behavioral response

to high levels of violence. If we had not followed migrants and if I had not controlled for it in the empirical specification the estimates of the effects of crime on the outcomes of interest would have been biased. Third, by measuring the effect of crime on a broad range of individual and household outcomes I provide a more comprehensive analysis of the labor market and welfare effects of crime.

4.4 Data: Mexican Family Life Survey

In order to study the impact of crime on economic outcomes at the individual and household level the INEGI homicide data is matched with the Mexican Family Life Survey (MxFLS). The MxFLS is an ongoing longitudinal, nationally representative survey of individuals and households who were living in Mexico in 2002 when the baseline was conducted. It includes information on approximately 8,440 households and 35,600 individuals spread among 150 communities and 16 states throughout Mexico. The MxFLS is designed to follow all baseline respondents and their children born after 2002, independently of whether they have moved within Mexico or to the U.S.

The longitudinal nature of the MxFLS and its rich content make it an ideal source to study the dynamics of individuals and households living in Mexico in 2002 and, specifically for this research, the impact of crime on different individual and household behaviors and outcomes. However, the contribution of using longitudinal data depends on the extent to which the original sample is successfully re-interviewed in each

following wave. With regards to this study, it is important that attrition is unrelated both to observed and unobserved characteristics, and also to the increasing trend of violence.

Overall, the MxFLS has achieved low levels of attrition. In the second wave, conducted between 2005 and 2006, over 89% of the panel respondents were re-contacted (Panel A. of Table 22 shows the re-contact rates in MxFLS2). The third wave of the survey, MxFLS3, has recontacted approximately 87% of all panel respondents.[12] Panel B of Table 22 shows current recontact rates for MxFLS3.

Even though the recontact-rates provide evidence of a successful follow-up, if attrition is correlated with homicide rates the sample could lose its representativeness. In section 4.8 I present the results of a model that test for selected attrition on measures of violence. I estimate the model separately for men and women and in an alternative model I allow for non-linearities in the measures of violence, and in each analysis I find no evidence of selected attrition as a response to violence.

The MxFLS provides a number of advantageous features for studying the impact of the Mexican drug war on the lives of Mexican citizens. First, as mentioned previously, the MxFLS2 was conducted in 2005/06 (during a period of "normal" levels of homicide rates) and the second follow-up (MxFLS3) was conducted between 2009 and 2013

[12] These recontact rates are preliminary. Final recontact rates are currently being estimated.

(during a period of high levels of violence). As indicated in Figure 8, the timing of the MxFLS allows the comparison of outcomes of the same individual in periods of low and high levels of violence and the longitudinal nature of the data allows us to control for unobserved time-invariant characteristics.[13]

Second, in the MxFLS there has been a concerted effort made to follow migrants within Mexico and to the U.S. This is particularly important for this study because migration may be a behavioral response to crime. For the sample of interest in this chapter, 8.9 percent of respondents were found in a different municipality between the two waves and 7.7% were found in the same municipality but report a long-term migration (more than one year) outside their municipality of residence. If migration is due to unobserved characteristics correlated with labor market outcomes and related to violence, it could bias the coefficients of interest.

By using a panel dataset, I am able to limit this potential bias, as the empirical estimation controls for any time-invariant unobserved characteristics of the respondent. Moreover, since it is possible that individuals may be migrating as a reaction to or in anticipation of violence, the empirical specification relies on an intent-to-treat styled model where an individual will get assigned the homicide rate over the 12 months prior

[13] Although the information at baseline provides another measure of pre-violence status, this study focuses on the first and second follow-up. One of the reasons to focus only on the last two waves of the MxFLS is to avoid excluding from the sample young individuals just entering the labor market in 2005 and for whom labor market information is not available in 2002.

to the MxFLS interview from his/her municipality of residence in 2005 to remove the possibility for local violence based systematic migration. Moreover, a model will be estimated to determine if potential changes in local violence predict migration behavior. The results, which will be discussed in more detail in the empirical section, suggest that violence predicts migration of single women and women living in rural places, and of self-employed men. These results highlight the importance of following movers in longitudinal surveys, and in particular for the purpose of this chapter.

Third, the MxFLS has a rich set of characteristics about its participants, including information about the economic, social and health status of each member of a surveyed household. Of particular interest for this study, the MxFLS contains a great deal of content about a respondent's labor market participation. For instance, employment is defined to include formal and informal jobs, we collect information on occupation and sector of employment, and complete information on earnings, weeks, and hours worked is collected for the two main jobs. The information about the earnings and hours worked during the last 12 months allows analysis of the impact of crime not only on net participation in labor markets but also on total earnings and productivity, measured by hourly earnings.

4.4.1 Labor Force Participation

Before describing the conceptual and empirical framework, I describe in this section the changes on labor outcomes observed between 2005 and 2009, for the

analytical sample of this chapter, i.e. MxFLS respondents interviewed in Mexico in MxFLS2 and MxFLS3 age 18 and older at the time of the MxFLS2 interview. Table 23 reports employment transition matrices exploiting the longitudinal nature of the MxFLS. Employment rates are calculated as the fraction of the population age 18 and older at the time of the MxFLS2 interview who report being active in the labor force.[14]

The evidence showed in Panel A suggests that 81.1% of this cohort of males were active participants in the labor force in 2005, and their net participation in the labor force only increased by 0.6 percentage points by 2009. Panel B of Table 23 shows employment transition in and out of work, and between wage employment and self-employment. The results in the first columns show that, from the 9.2% of males leaving the labor force, 56% were employees in the pre-violence period and 33% were self-employed. In addition, the results of Panel B for males suggest that, even though male labor participation was pretty much constant between the two periods, there is some mobility across the labor sectors. Out of the 72% of males who work in both periods, 52% work as employee in both waves, 15% remain self-employed, and around 26% move between

[14] In MxFLS, employment is defined as having worked for at least one hour during the week prior to the interview for a salary, wage, or other payment in cash or in kind. It also includes individuals that did not work the week prior to the survey if the reason for missing work is a temporary cause such as a vacation or a short-term leave.

sectors.[15] The purpose of this chapter is to identify whether the increase of crime explains any of these changes.

Female participation, on the other hand, increased by almost 4 percentage points but more than half of women remain outside the labor market in both waves. The composition of women leaving the labor market is very similar to that of males: 53% were employees and 32% were self-employed. Out of the women working in both periods, 54% work as employee both periods, 14% are self-employed and do not switch and 24% move between wage employment and self-employment.

This evidence suggests that between 2005 and 2009 both males and females have seen changes in their labor force participation. The next sections of this chapter show tests for whether increases in crime explain some of these changes and whether these changes in labor participation have an effect on individual earnings and on household per capita expenditure.

4.5 Conceptual Framework

The conceptual framework that underlies the empirical analysis is based on an individual maximization problem.[16] Individuals' make consumption (*C*) and labor

[15] A very small percentage are unpaid workers in both waves. And, in total for 5.4% of males we have information on labor force participation but no information on their labor sector, and for females only 2.3% do not report employment sector.
[16] See Becker, 1965; Becker, 1968.

supply decisions into the wage labor market (E) or self-employment (S), which determines their leisure (L), to maximize their utility function:

$$\begin{aligned} & Max_{E,S,C}(C, L) \\ & s.t.\ w_E E + \pi_S S - PC \\ & \bar{T} = E + S + L \end{aligned} \tag{4.1}$$

Where w_E, is the wage in the labor market and π_S, is the profit function in the self-employment sector. The individuals' labor supply decision, in either sector, is a function of the prices of the goods, P, wages in the labor market, w_E, profits in the self-employment sector, π_S, community characteristics, X_c, the level of violence, V, and observed and unobserved idiosyncratic individual and household characteristics, ν_i, and ν_h.

$$E = E(P, w_E, \pi_S, X_c, V, \nu_i, \nu_h)\ and\ S = S(P, w_E, \pi_S, X_c, V, \nu_i, \nu_h) \tag{4.2}$$

As discussed in the introduction, the effect on the labor supply is ambiguous. On the one hand, given the characteristics of the conflict in Mexico, the effect of violence on the labor supply is expected to be negative if the fear of being victimized creates an additional cost on the labor force participation. As mentioned before, Organized Crime Groups have not only increased in number, but they have also started to use extortions of civilians as a financial resource and to increase the sense of fear in the community. An increasing sense of insecurity, lack of confidence in the police, and an increasing probability of being a direct victim of OCGs can induce, for example, business owners to close early, and street vendors to reduce their hours of exposure on the streets. The fear

factor, can also have a negative effect on the decision to participate in the wage labor market, though, it is important to note that self-employed individuals have more flexibility to reduce their hours or temporarily leave the labor market. Second, the cost imposed by the high levels of violence may cause some firms and businesses to close completely or reduce their size. In addition, the economic activity of the community may be impacted by violence, which negatively affects the profits of businesses and the self-employed and therefore may further reduce labor supply.

On the other hand, if for example the presence of drug-cartels incentivize the economic activity in the community, the effect of violence on the labor supply could be positive. An empirical test for the effect of violence on the labor force participation is a test of: $\frac{\partial E}{\partial V} < 0$ and $\frac{\partial S}{\partial V} < 0$. Thus, the total effect on the labor force participation will reflect a combination of the impact on labor supply and labor demand. In addition, I can test whether violence has had an effect on labor transition between the wage labor market and self-employment sectors.

This analysis also test the effect of high levels of violence on wages in the labor market, w_E, and profits in the self-employment sector, π_S:

$$w_E = w_E(P, X_c, V, v_i, v_h) \; and \; \pi_S = \pi_S(P, X_c, A, V, v_i, v_h) \tag{4.3}$$

Where A, are productive assets. The total effect on the wages depends on the extent to which violence changes labor market demand.[17] If firms close or reduce their size, the final effect on wages (for the workers that stay in the labor market) will depend on the negative magnitude of the shift in the labor demand and supply; but, unambiguously, the expected impact on labor participation would be negative. Evidence from Mexican newspapers as well as reports from respondents during the fieldwork of the third wave MxFLS suggest that both small businesses (self-employed) and bigger firms (in particular those related with tourism and manufacture) have been affected by the violence. By disaggregating each employment category by occupation this analysis is able to differentiate the impact of violence in different sectors of the labor market.

The effect of violence on the individual labor force participation might depend on idiosyncratic characteristics at the individual and household level. For example, the presence of young children in the household may incentivize one of the parent's to drop out of the labor force to take care of the children in order to avoid their direct exposure to violence. A rich set of characteristics measured before the increasing trend of violence will allow for exploration of this heterogeneity.

[17] The negative effect of conflict on various economic outcomes has previously been established in the literature (Collier 1999; Hoeffler and Reynal-Querol, 2003; Abadie and Gardeazabal, 2003; Deininger, 2003) and in particular on investment at the firm-level (Collier and Duponchel, 2013; Pshisva and Suarez, 2010; Camacho and Rodriguez, 2013).

4.6 Empirical Strategy

Drawing from a longitudinal survey where individuals were interviewed in 2005-06 (period of low levels of violence) and in 2009-13 (period of high levels of violence), the empirical specification relies on an individual fixed effect model, controlling for a rich set of individual and household time-variant characteristics, to estimate the effect of a plausibly exogenous shock of crime in Mexico on individual and household outcomes.

One of the challenges to the empirical estimation of this relationship is systematic behavioral response to crime. Specifically, non-random migration as a response to elevated conflict would hinder identification of the true impact of violence on labor market outcomes. For example, if individuals with higher earnings are more likely to migrate when violence increases, the impact of violence on wages would be overestimated. Table 24 shows the results of a linear probability model that predicts migration between municipalities, for the sample of interest in this chapter, as a function of the change in the homicide rate between the second and third waves of the MxFLS, using the following specification:

$$m_i = \delta_0 + \delta_1 Hom_{\Delta 2009-05} + \delta_2 x_{i2005} * Hom_{\Delta 2009-05} + \delta_3 x_{i2005} + \varepsilon_i \quad (4.4)$$

The measure of migration is a binary outcome equal to 1 if the respondent was interviewed in a different municipality in MxFLS2 and MxFLS3 (8.9%) or if in the migration history the respondent reported a long-term migration (one year or more)

away from the municipality of residence in MxFLS2 (7.7%). $Hom_{\Delta 2009-05}$ is the difference in homicide rates between 2005 and 2009 in the municipality of residence in MxFLS2; and, x_{i2005} are household and individual characteristics measured at the time of the MxFLS2 survey and include: age, years of education, marital status, cognitive score, household size and household composition, employment status and log of earnings, a dummy for whether the household has relatives in the U.S. and for whether the locality of residence is rural.

Columns 1 and 2 of Table 24 show the results for males. The results suggest that on average the change in the homicide rate does not predict migration of men. However, the model that adds interactions with individual and household characteristics shows that self-employed men are more likely to migrate when violence increases. If we had not followed migrants, we would have systematically lost self-employed men in places with high levels of changes in homicide rates. This result is consistent with anecdotal evidence that describe how business owners have migrated and closed their business to avoid being victimized by organized crime organizations. The results for women, in columns 3 and 4, show that single women and women living in rural places are more likely to migrate when violence increases. It is reasonable to think that the cost to migrate for married women might be higher since their migration possibly implies the migration of their entire family. Therefore, even if violence increases, married women are less likely to migrate than single women. The results for women living in rural

places is consistent with the fact that most of migration in Mexico is from rural to urban places. For women living in rural places higher level of violence decreases the relative cost of migration and therefore increases their likelihood of migration.

In order to control for migration as a behavioral response to crime, the empirical specification follows an intent-to-treat approach, where **t**he municipality of residence in the second wave of the MxFLS (MxFLS2) determines an individual's exposure to violence. By using an individual's location during the period of low and constant levels of violence, I am able to eliminate potential bias due to violence related endogenous migration.

In addition to behavioral responses, an additional challenge when estimating the effect of violence on economic outcomes is omitted variable bias. The difficulty in estimating the relationship between violence and economic outcomes could emerge from the fact that homicide rates have not increased in a random fashion over time and might not be orthogonal to unobserved factors that affect economic performance in the municipality or at the individual level.

To control for this unobserved heterogeneity the empirical specification will exploit an individual fixed effect model. The identification strategy exploits the variation over time of homicides rates between 2005 and 2009-12. For each of the outcomes of

interest an individual fixed effect model[18] is estimated, thus by definition the empirical strategy compares the same individual across time periods which captures the unobserved, time invariant factors that affect the dependent variable. This is particularly useful if we believe that there are time-invariant characteristics of individuals, such as ability or risk preferences, that are correlated with both labor outcomes and the violence level of the municipality in which the individual chooses to live. Moreover, if homicides are reported with error, the individual fixed effect strategy differences out error that is constant over time.

The empirical strategy can be generalized in the following regression framework:

$$y_{ijt} = \delta I(0 < HomRate_{jt} \leq 5) + \gamma I(5 < HomRate_{jt} \leq 15) + \rho I(15 < HomRate_{jt}) + \\ + X'_{ijt}\varphi + \theta_i + \alpha_{kt} + \beta_t + u_{ijt} \quad (4.5)$$

Where *y* is the outcome of interest of individual *i* living in municipality *j* at time *t*. I start the analysis by estimating a general model of labor force participation stratifying by sector of employment in 2005 (wage employment and self-employment) and by gender for respondents between the ages of 18 and 75 in 2005. Then, I estimate the model using hours worked and labor income as dependent variables on the sample of individuals working in both waves. The measure of earnings in the empirical section is the quartic root of hourly earnings and the quartic root of total earnings in the last 12

[18] With two periods an individual fixed effect model is similar to a first difference model. The results of a first difference model are qualitatively the same.

months.[19] Finally, I estimate a model that predicts the effect of violent crime on household expenditure (this model follows the same specification of equation (4.5) but estimates household fixed effects).

I estimate the effect of crime on the outcomes of interest exploiting a non-linear model. Although the levels of violence at the national level have more than tripled in only four years, there is a great deal of heterogeneity in the local exposure with many municipalities seeing no change in violence and others suffering increases of more than 500%.[20] A linear model might underestimate the effects of violence and, in addition, a non-linear specification allow for exploration of a threshold of violence after which the effects of violence are particularly significant.

In equation (4.5) $\left(0 < HomRate_{jt} \leq 5\right),\ I\left(5 < HomRate_{jt} \leq 15\right)\ and\ I\left(15 < HomRate_{jt}\right)$ are indicators that denote different levels of the homicide rate at the municipality level at time *t;* and, therefore, δ, γ and ρ are the coefficients of interest.[21] In an individual fixed effect framework, these variables will identify the effect of violence

[19] Earnings in the last month might be a noisy measure of labor income, particularly for self-employed individuals whose labor income can significantly vary within the year. For this reason I use earnings in the last 12 months. However, some individuals report non-positive earnings and using a logarithmic transformation for earnings would drop a number of observations. The quartic root behaves similarly to a logarithmic transformation for positive numbers (Thomas et al. 2006).

[20] In the state of Durango and Nuevo Leon, for example, there is a significant number of municipalities where homicide rates multiplied in only 4 years and went from being at the bottom of the national homicide rate distribution to the highest end of the distribution.

[21] The omitted category are municipalities with homicide rates equal to zero.

in municipalities whose homicide rate switched to each of the different groups between 2005 and 2009.

The rest of the specification continues as follows: X is a vector of individual and household time-varying characteristics (marital status, whether the respondent lives with his/her parents, household size, number of kids in the household, whether the place of residence is rural or urban, expectations of future migration, a measure of risk aversion and patience, a measure of emotional well-being and presence of relatives in the US); θ_i denotes the individual fixed effect; α_{kt} includes GDP at the state level; and β_t includes controls for year and quarter of interview.[22]

An additional challenge of the specification is the possible endogeneity of crime. In order to produce consistent estimates, the idiosyncratic error at each observation has to be uncorrelated with the variable that measures crime in both periods (Wooldridge, 2002). It would be reasonable to think that u_{ijt} is correlated with $HomRate_{jt}$ because crime is not allocated in a random way and it might be higher in municipalities with better economic performance, so the expected profit of the extortions to the civilians is larger. Second, we might think that u_{ijt} is correlated with $HomRate_{jt+1}$ if unobserved variables in 2005 affect both labor outcome variables in 2005 and the level of crime in 2009.

[22] Specification with time trends were estimated and the results do not change.

There are reasons to believe that increasing crime rates might be correlated with the error term in 2005: first, crime might be more likely to happen in areas with better economic growth because it is more profitable to extort civilians in these places, or it might be more likely in places with worse economic activity if that is a reflection of bad institutions and low state presence. While the individual fixed effect strategy controls for all time-invariant characteristics, if economic trends are changing differently in municipalities that have suffered a higher change of violence, the individual fixed effect estimates will be biased. In order to control for this I add controls for changes in GDP at the state level; moreover, in additional specifications I add unemployment rates at the municipality level. This is not my preferred specification since unemployment rates at the municipality level is clearly endogenous to the individual labor market measures. However, even adding this characteristic of the labor markets, the results of the main specification hold.

In order to examine the relationship between a municipality's change in violence to demographic and economic characteristics of that municipality, in section 4.8, I show the results of a model where municipality characteristics measured in 2002 and 2005 predict the levels of violence observed in 2009. Using pre-high violence data from the Census and from the MxFLS I do not find evidence of municipality characteristics being a predictor of the high levels of violence. In addition I estimate a placebo model to test the exogeneity of the surge in crime and the results give support to the assumption that

changes in crime levels was unrelated to other underlying economic trends (the results are shown and discussed in section 4.8).[23]

4.7 Results

4.7.1 Labor Market Outcomes

Tables 25 to 30 provide results for the analysis of the participation in the labor market, hours worked, the quartic root of hourly earnings, and total earnings stratified by gender and by occupation.

4.7.1.1 Females

Table 25 provides the results for women who were self-employed in 2005 in Panel A and those who were wage employees in 2005 in Panel B. Column 1 in Panel A shows the effect of violence on labor force participation (either as self-employed or wage employee) conditional on being self-employed in 2005.[24] These results show that exposure to violence has an adverse effect on the labor force participation for self-

[23] A final challenge for the empirical specification is that the main fieldwork was conducted during 2009 and 2010, and it would be reasonable to think that those interviewed in 2010 would have responded to the increasing trend of crime in a different way than those for whom the impact of violence was unanticipated. I compared the two groups of respondents and as expected those interviewed in 2010 have the characteristics of a group that in the field would be more difficult to track: younger, more educated, more likely to work, etc. I estimated the main specification with and without the individuals interviewed in 2009 and after 2009 and the results of the main specification hold.

[24] Transitions from self-employment to formal employment and vice versa are not considered a change in labor force participation in this version of the model. Models of transitions between self-employment and formal employment were estimated, but homicide rates are not a significant predictor of the change in occupation for any of these samples or models.

employed women. The results suggest that women living in a municipality that experienced any level of homicides above 0 were approximately 20% less likely to stay active in the labor market. These results are robust to changes in GDP and unemployment rates at the municipality level.

If the cost of participating in the labor market is rising due to increased crime, it is intuitive to think that women may re-allocate their time to other activities like household production. Column 2 shows whether the intensity of participation in the labor force, for women who stay active in the labor market, has been affected by the increase of violence. The results in column 2 of Panel A show a U-shaped effect on hours worked. Women reduce their hours worked if the municipality where they were leaving suffered a relatively small increase from zero to a positive homicide rate smaller than 5 or if their municipality's crime jumped to a rate larger than 15. Women living in these municipalities decreased their hours worked by around 70%. This means that from an average of 37 hours per week (7.4 hours per day in a work week of 5 days) self-employed women affected by violence, basically changed their status from being full-time workers to part-time workers.

The results in columns 3 and 4 show the effects on labor income. The results in column 3 and 4 show a great deal of heterogeneity depending on the degree of the change of homicide rates. Although, the results are imprecisely estimated for hourly earnings, the estimates suggest that, the women living in municipalities that switched to

the highest level of homicide rates that stayed in the labor market are being rewarded with an increase in their hourly earnings. On the other hand, women that stay in the labor market in municipalities that suffered the smallest increase in violence are not experiencing an increase in earnings. The dual impact of fewer hours worked and lower hourly earnings translate into significantly lower total earnings for women living in these municipalities.

Panel B of Table 25 shows the effects for women who were wage employees before the intensification of violence. Employee females compared to the self-employed do not suffer as adverse an impact from the increasing violence in Mexico. Columns 1 and 2 show negative effects of violence on labor force participation and in hours worked but the effects are not significantly different from zero. Moreover, the effect on the hourly earnings of employee women seems to be positive but the effects are not significant and the magnitude is smaller than the effect for self-employed. This positive effect seems to translate into overall positive earnings but the coefficients are imprecisely estimated. These findings reinforce the idea that wage employees who were likely to be under contract before the onset of violence are employed in jobs where wages are more sticky and individuals less able to adjust their labor force intensity.

Individual Heterogeneity

Individual characteristics may play an important role in the total effect of violence on labor outcomes. As discussed in section 4.6, households with young children

or varying levels of wealth may react differently to higher levels of crime. To test whether this is the case in the Mexican context, I estimate equation 4.5 adding interactions of the violence variables with individual characteristics measured in the pre-violence period.

Table 26 shows the results of this model for self-employed women. The same models were estimated for wage employee women but, similar to the main specification, no significant results were found for this sub-sample.

Columns 1 to 4 show the results adding an interaction of the homicide rate with an indicator equal to 1 if the woman has children living in her household. The results in column 1 show that women with children are significantly more likely to leave the labor market and the effect does not seem to vary with the level of violence suffered in the municipality of residence. In Mexico, the need of the new drug cartels to establish their territorial presence has created new relations between them and juvenile local gangs (Guerrero, 2012a). This has created financially profitable opportunities for young individuals and increased their opportunity cost of attending school. The combination of wanting to both protect their children from exposure to victimization and to keep them from joining the local gangs may cause women with young individuals in their household to react to high violence by leaving the labor force to better monitor their activity.

The results in column 3 show that the positive effect observed for hourly earnings in municipalities with the highest change in crime in the main specification becomes significant for women with no children. The opportunity cost of staying in the labor force might be lower for this group of women and the lower supply of labor might be having a positive effect on the hourly earnings of those who stay in the labor force.

Finally columns 5 to 8 show the results adding an interaction of the variables of violence and a dummy equal to one if the women's household per capita expenditure is in the top 50th percentile of the distribution. The results show than women in this end of the PCE distribution cut significantly their hours worked significantly more than less wealthy women when facing the highest level of crime.

The results in this section suggest that the labor market outcomes of self-employed women have been affected by violence. Violence has adversely affected the labor force participation and labor intensity of self-employed women. In addition the response in hourly income to increased homicide rates, for the women who stay in the labor force, is dependent on the level of violence exposure. These results suggest that a nontrivial number of self-employed females are leaving the labor market or reducing their hours worked, most likely to avoid being directly affected by violence or to take care of their children, but those who stay in the labor market may be receiving a wage premium. On the other hand, labor outcomes of women who were employees before the escalation of violence do not seem to be reactive to changes in the homicide rate.

Analysis by occupation

Given the fact that the fracturing of the drug cartels has led to an increase in non-violent crimes that are directed at non-combatants like extortions and kidnappings, it is reasonable to think that occupations where individuals are more exposed to potential victimization, such as business owners and street vendors, might be more adversely affected by the recent change in the conflict environment.

Table 27 shows the results of the main specification for self-employed women, in columns 1 to 4, disaggregating the sample by the three most popular occupations in self-employment.[25] The three most popular occupation categories for self-employed women are jobs in retail and commerce (51.47%), manufacturing (women in this sector are mainly artisans, tailors and food makers, 23.25% of self-employed women are in this category) and domestic employees and women working in personal services (16%).

Examining the results for self-employed women across the different occupations it is clear that women in retail/commerce and manufacturing are the most likely to leave the labor force. The results in column 2, suggest that for women that remain employed work intensity is decreasing similarly for all three occupation types (although the results are not significant). Interestingly, only the women working in personal service and

[25] The MxFLS provides information on the two-digit Mexican classification of occupations. I aggregate those in 1. Professionals, technicians; 2. Agricultural, cattle activities, foresting, hunting and fishing workers; 3. Manufacture activities; 4. Administrative jobs; 5. Retail and commerce; 6. Personal services, domestic employees; and 7. Other.

domestic employment are seeing their work incentivized by increases in hourly wages. This may be due to the fact that since these types of service providers have such a great deal of access to the employer's property/personal life, employers in a time of increased fear of victimization are providing an incentive for the women they are already comfortable and familiar with to stay in the labor market through increased wages. Finally, I find that women working as artisans, tailors, and food makers are suffering the largest deficit in their total earnings. This likely a result of reduced foot traffic and street sales brought on by fear-induced reductions in economic activity in areas with increased crime.

4.7.1.2 Males

Table 28 provides the results for male labor force participation, following the same structure of Table 25 for females. The results in Table 28 suggest that changes in the homicide rate over the last few years have not had a significant effect on labor force participation, for both self-employed and employee men.

Column 2 of Table 28 displays the results of the analysis using the log of hours worked in the last 12 months as the dependent variable. The estimates in Panel A (for the sample of self-employed males) and Panel B (for the sample of employees), provide evidence that the labor force intensity of males was also not affected by the surge in violence in Mexico. Taken together the results from columns 1 and 2 strongly suggest

that regardless of the conflict environment men do not change their level of engagement in the labor market.

The results for the impact of violent crime on the hourly earnings and total earnings in the last 12 months for men are found in columns 3 and 4. The coefficients in columns 3 and 4 of Panel A show an increasing negative effect of violence on both hourly and total earnings of self-employed men at the highest intensity of violence exposure. The results suggest that males living in municipalities where the homicide rates reached levels higher than 15 homicides per 100.000 saw a significant reduction in their hourly earnings and total earnings. Panel B shows the results for employees and while there is some evidence of reduced hourly earnings for these men, contrary to what was found for self-employed men, total labor income is not adversely affected by crime.

The negative effect on total earnings of self-employed may lend support to the anecdotal evidence in Mexico that suggests business owners have been particularly affected by violence. It has been suggested that businesses in places heavily exposed to violence tend to close earlier to avoid being directly victimized. Moreover, it is reasonable to think that individuals in those municipalities are less likely to engage in commerce in the evening due to increased insecurity, harming the profits of independent businesses. In a later subsection I check the hypothesis that the earnings of business owners may be more susceptible to increasing violence by looking at the disaggregated results by occupation.

Individual Heterogeneity

The results in Table 29 show the models with interactions of homicide exposure and having young children in the household, as well as, SES for the sample of self-employed men. As was the case with the female analysis, the sub-sample of employees did not show significant heterogeneity.

The results in column 1 to 4 suggest that, for men, the presence of children in the same household, unlike for women, does not differentially change their decision to participate in the labor force under conditions of increasing violence. In terms of a heterogeneous impact of violence by socio-economic status, columns 5 to 8 show that men in the highest percentile of the per capita expenditure are the most adversely affected by higher levels of violence, as they are more likely to reductions in the hourly income, as well as, their total income even after they increase their total hours worked. These results provide further evidence that is consistent with business owners being the most negatively affected by violent crime, as this the most common profession for high SES self-employed men. As mentioned previously, in order to test this directly I disaggregate the sample by occupation in the next set of results.

Analysis by occupation

Table 30 provides the results of the impact of violence on labor market outcomes for men disaggregated by occupation. Forty percent of self-employed men work in

agricultural, cattle activities, foresting, hunting and fishing activities (Panel A), 30% in manufacturing (Panel B) and 20% in retail and commerce (Panel C).

The results in columns 1 and 2 suggest that, independently of the intensity of violence and occupation, self-employed men do not leave the labor market or reduce the hours worked. Columns 3 and 4, on the other hand, support the previous hypothesis that self-employed men in the retail and commerce industry have been the most affected by violence and the effect is particularly strong in municipalities with the highest increase in the homicide rate. Moreover, the effect for the labor income of men in this industry is negative and significant. These findings strongly suggest that high incidence of crime through diminished business activity has a significant negative effect on the total earnings of men working in retail/commerce and these negative effects seem to be driven by lower hourly earnings.

Overall the results of the analyses in section 4.7.1 highlight the significant heterogeneity of the effects of violence in Mexico on labor market outcomes. Specifically, the findings suggest that self-employed individuals are the most adversely effected by increasing local violence. This relatively higher level of vulnerability to negative shocks may be a result of having less stable income sources. Moreover, the results suggest there is a significant difference in the labor market environment for self-employed males that faced increased violence relative to self-employed females exposed to more violence. Males' labor force participation is not being affected by the increasing violence

regardless of the level intensity, while self-employed women are highly reactive to any increase in the level of conflict. Moreover, while self-employed women that remain in the labor force in areas with the most violence see no change, or perhaps a positive change, in their total and hourly earnings, the comparable self-employed men are suffering significant losses.

The lower labor income some of these individuals are earning when exposed to violence may translate into lower consumption in these households and thus affect long-term investment that spills over into the well-being of the next generation. In order to assess whether increased exposure to violence has translated into lower consumption I analyze this outcome in the next section.

4.7.2 Per Capita Expenditure (PCE)

There is a growing literature on household responses to negative shocks (Townsend, 1995; Morduch, 1995; Beegle et al 2001; Frankenberg et al., 2003; Thomas et al. 2003; Thomas et al, 2007). Intuitively, the negative shock of increasing levels of violence in Mexico could have affected households' wealth in several ways. First, if labor income is negatively impacted by crime, households without mechanisms to smooth consumption will be affected by this unanticipated loss in resources. Second, in order to

minimize the probability of victimization, conspicuous consumption and consumption that increases the likelihood of being a target of exploitation, might decrease.[26]

This section of the chapter assesses whether the negative shocks in the labor market have translated into reductions in other measures of the household's wealth such as per capita expenditure. The MxFLS has a rich component of household expenditure measures that allow analysis on PCE to be disaggregated by different consumption categories in order to assess which components have been more or less affected by violence. This component records detailed information on household expenditures on food (including self-production) and non-food goods and services, personal care, household items and durable goods. I estimate equation (4.5) for four different categories of per capita expenditure in the last month: total per capita expenditure, expenditure on food, a category of arguably "conspicuous" consumption[27], and education. Panel A of Table 31 shows the results for all the households, Panel B for households with a female household head and Panel C for those with a male household head.

[26] It is also intuitive to think that expenses on security devices will increase. However, there is not a clear way to disentangle these expenses from the consumption section.

[27] I include in this category more visible expenditures or expenditures that could potentially increase the probability of being victimized like: meals out, recreation, communication, clothes, domestic appliances, transportation, furniture, ceremonies and vacations and semi-durable goods.

Column 1 shows the results for the log of total per capita expenditure for the three samples. Looking at the sample in Panel A that includes all households, the results suggest that the total PCE for households living in municipalities in which violence increased to a positive homicide rate smaller than 15 homicides per 100.000 decreased between 9 and 12%. However, the effect is significant only at the 10% confidence level. Disaggregating the sample by the gender of the household head there is not strong evidence that the effect is particularly damaging to the per capita expenditure of households with a certain gendered household head. Since the previous results suggested that it is self-employed men that are suffering the largest reductions in total earnings, I next tested if households led by these individuals are more susceptible to decreases in expenditure. To do this I estimate a model adding interactions between the variables of violence and a dummy for self-employment. The results (not reported in this version of the chapter) show that, as expected, it is self-employed male-headed households that face the largest reductions in per capita expenditure, while the self-employed female-headed households are not differentially reducing spending. This suggests that, the negative effects found for labor income is translating into consumption decisions and the most affected households may not be able to smooth the negative shock through the use of savings or other mechanisms.

Column 2 and 3 shows the results for expenditure on food and "conspicuous" consumption and these analyses provide little evidence of significant decreases in these

types of consumption. Column 4, on the other hand, shows negative and significant effects on education expenditure for households with female household heads living in municipalities where the violence increased the least. The results of previous section suggest that some self-employed women are leaving the labor market, and earnings of self-employed women working in commerce activities decrease with violence. Moreover, when the impact of expenditure on education is stratified by the employment type of the female household head, it is those household's led by a self-employed woman that are facing the largest decreases in education expenditure. The negative impact on education expenditure for female-headed households may be a result of the direct reduction in household income or due to a decrease in the woman's bargaining power in the household, as it had been shown that women are more likely to invest in a child's education. If this reduction in educational expenditure and loss in family income leads young kids of school age to be more likely to drop out of school and join the labor market it may lead to longer-term and persistent deficits in human capital accumulation for the next generation. With this in mind, Brown and Velásquez (2013) assesses the extent to which a violent environment may alter the educational attainment, cognitive scores, time allocation, and employment behavior of children and young adults.

4.8 Robustness Checks

4.8.1 Attrition

In order to test for potential selected attrition on measures of violence I estimate a model that predicts the probability of attrition from MxFLS3. The independent variables are individual and household characteristics measured in MxFLS2, the difference of homicide rates between 2005 and 2009 in the municipality of residence in MxFLS2 and the interactions of the measure of violence and individual characteristics. These interactions provide evidence of whether individuals with certain characteristics were more likely to attrit from the survey when violence increased. The sample includes all panel respondents age 18 and older at baseline and therefore at risk of being in the analytical sample of this chapter. The model is estimated as a linear probability model based on the following empirical specification:

$$A_i = \delta_0 + \delta_1 Hom_{\Delta 2009-05} + \delta_2 x_{i2005} * Hom_{\Delta 2009-05} + \delta_3 x_{i2005} + \varepsilon_i \quad (4.6)$$

Where A_i is a binary outcome equal to 1 if the panel respondent attrited from the survey in MxFLS3 conditional on being found in MxFLS2 and equal to zero if not. The model for attrition follows the same empirical specification as the one for migration in equation (4.4). Therefore, attrition from MxFLS3 depends on the change of violence between 2009 and 2005, individual and household characteristics measured in 2005 and the interaction of the measure of violence and own characteristics.

The results are reported in Table 35 in section 4.11 for the sample of males and Table 36 for the sample of females. The results for males in Column 1 suggest that on average the homicide rate in 2009 do not predict attrition; and, the results in Column 2 suggest that the change of homicide rates between 2005 and 2009 do not predict attrition in a different manner for males with different characteristics measured in 2005. The results for females in Table 36 show similar results. The change on homicide rates do not predict attrition and it does not predict attrition differently for females with different characteristics. In addition to this model I estimated a model where I allow for a non-linear effect of violence on attrition and the results do not change.

4.8.2 Prediction of Homicide Rates

In order to test the exogeneity of the surge in crime observed in Mexico since 2007 I estimate a model where the first, characteristics of the municipality in 2005, and second, trends in these characteristics between 2002 and 2005 predict the change of the municipality homicide rate between 2005 and 2009. The covariates are averages at the municipality level of MxFLS variables in one specification and Census variables in the other. Table 37 shows the results. Columns 1 and 2 show the results using variables measured in 2005 as covariates. The results suggest that municipalities with, on average higher labor participation of women had lower changes on their homicide rates, but those with lower illiteracy, and with higher electricity saw a higher change in violence. Places with a higher percentage of individuals with less than primary but also with

higher percentage of high school graduates had higher changes in their homicides levels. The results do not suggest a clear pattern by which pre-violence characteristics are related to the change in the homicide rate between 2005 and 2009, other than that suggest that violence increased more in more urban municipalities (those with better supply of public goods). Moreover, if the characteristics of the places where violence increase are constant over time, these fixed characteristics will controlled for in the fixed effect model. If however, some trends at the municipality level are correlated with the trends on violence and with labor outcomes, the specification in (5) could be biased.

The results from Columns 3 and 4 show that pre-violence trends in a rich set of economic and demographic characteristics are unrelated to future changes in violence. This provides suggestive evidence that the change in a municipality's level of violence is not simply a reflection of underlying trends in other characteristics of that municipality and that the change in violence levels in a municipality is not caused by pre-existing trends in the labor market environment.

4.8.3 Placebo Test

Measuring the impact of crime on economic outcomes imposes important challenges. One first order issue is that crime might be endogenous to economic activity, which makes it difficult to identify a causal relationship between crime and, for example, labor market outcomes. The increasing trend of crime in Mexico is evident from 2007, but from 2002 to 2005 the homicide rate was very stable. As a robustness

check, I estimate the same model discussed in the main specification, using individual and household observations measured in 2005 (MxFLS2) and in 2002 (MxFLS1), while continuing to utilize the same measures of violence (MxFLS2 and MxFLS3) as in the main specification. If it is the case that the level of change in crime in Mexican municipalities was not a result of underlying economic trends, no significant effects should be observed for the variable of violence in this specification, as future violence should not predict economic outcomes between 2002 and 2005.

Tables 32 and 33 shows the results for labor outcomes for women and men respectively. The estimates for women in Table 32 do not show any significant results and, moreover, the sign and magnitude of the coefficients do not follow the same pattern as those from the main specification.

Table 33 provides the results for males. The results in the main specification suggest that men do not change their labor participation but the total earnings diminish for men living in municipalities that switched to the highest percentile of the homicide rate distribution, but the effect is larger for self-employed men. The results in Table 22 show no evidence of these effects when looking at labor outcomes measured before the increasing wave of violence.

The results in Table 34 show the results of the placebo test for the PCE outcomes. These results suggest that municipalities where crime increased may have been on a positive trajectory of per capita expenditure for households with a female household

head and the results of the mains specification may actually be an underestimation for this sub-sample.

4.9 Conclusions

In order to test for potential selected attrition on measures of violence I estimate a model that High levels of crime may substantially alter the context in which individuals operate, affecting behaviors at the individual and household level (Verwimp et al., 2009). In particular, crime may have an effect on labor markets, as fear of victimization may increase the cost of participating in the labor market or reduced economic activity leads to decreases in investment and diminished job opportunities.

Measuring the impact of crime on labor market outcomes in Mexico is critical given the high incidence of crime observed since 2007. In addition OCGs have diversified their financial sources, and begun to rely increasingly on criminal activities that directly affect the civil population, like extortions, kidnappings, and car thefts.

Exploiting information from the MxFLS, which was collected in periods of both low and high levels of violent crime, this study estimates an individual fixed effect model that controls for time-invariant unobserved heterogeneity that could affect both exposure to violence and labor market outcomes. Moreover, combining this strategy with an intent-to-treat approach, where the municipality of residence in the pre-violence

period determines an individual's exposure to violence, this analysis is able to shield the estimates from potential bias due to violence-related endogenous migration.

The results of this study show that the increasing violence in Mexico has had a strong and significant negative effect on the labor market participation and intensity of self-employed women. The analysis further provides evidence that women who worked in personal services and as domestic employees and remain in the labor market are earning higher hourly wages, which leads to gains in total earnings. A possible explanation for this finding is that in times of increased feelings of vulnerability employers have a strong preference for domestic employees they know and trust as these service providers have a great deal of access to an employer's property and personal space. As such these women are being encouraged not to leave the labor market through high remuneration.

Similarly, for males, the negative impact is also stronger for self-employed individuals. A major difference between the genders though is that for males, labor participation and intensity is not significantly reduced and there is no evidence of a positive impact on earnings for any occupational subgroups. Moreover, earnings of self-employed males particularly in the commerce industry are negatively affected by an increased exposure to violent crime, suggesting these business owners' profits are being adversely affected by reductions in economic activity. The negative effects found for these men's labor income has been translated into decreases in measures of households'

well-being, as well. Suggesting the households are not able to perfectly smooth consumption when facing this shock of violence.

This chapter provides evidence that the increasing violence in Mexico not only affects the direct participants in the Mexican Drug War but also regular civilians. If this is the case, it may be in the best interest of the government to mediate the negative wealth shocks the rise in violence is imposing in order to reduce the long-term impact of this conflict. In on-going research I am exploring the impact of crime on measures of wealth at the household level, focusing specifically on whether households with more defined safety nets are less susceptible to changes in the violence environment.

4.10 Tables and Figures

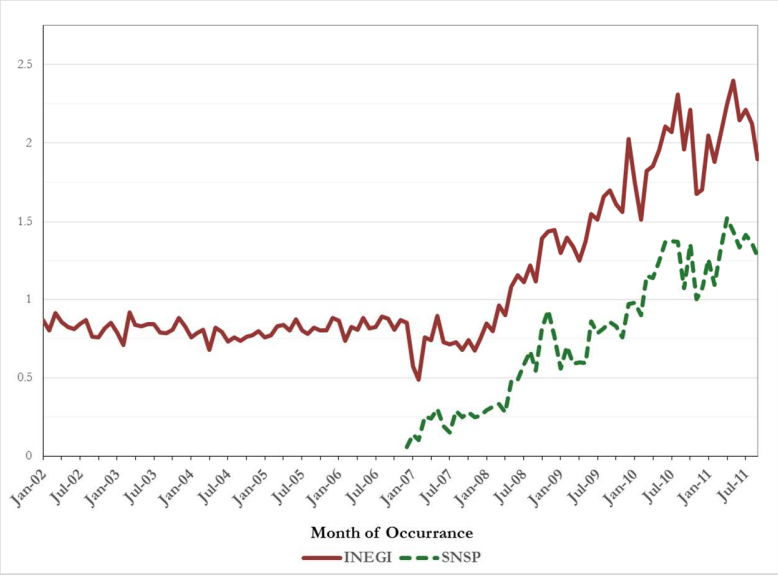

Figure 7: INEGI and SNSP - Monthly Homicide Rate (per 100,000)

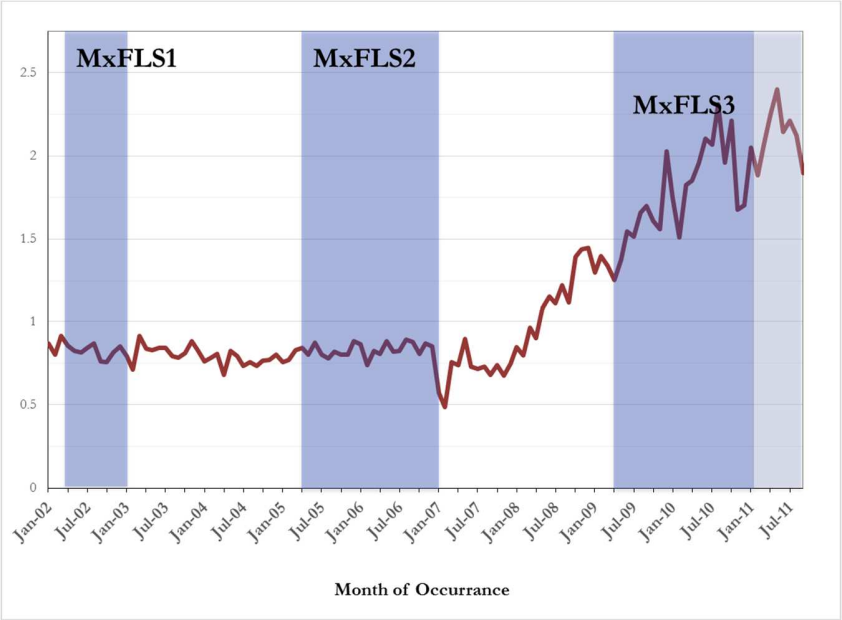

Figure 8: Timing of MxFLS and INEGI - Monthly Homicide Rate (per 1000,000)

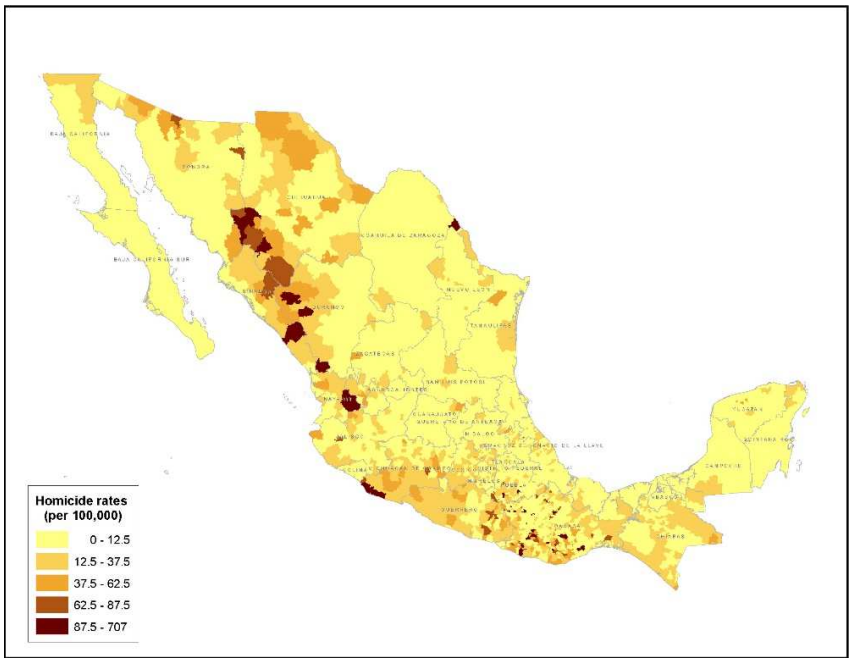

Figure 9: INEGI Annual Homicide Rate - 2002

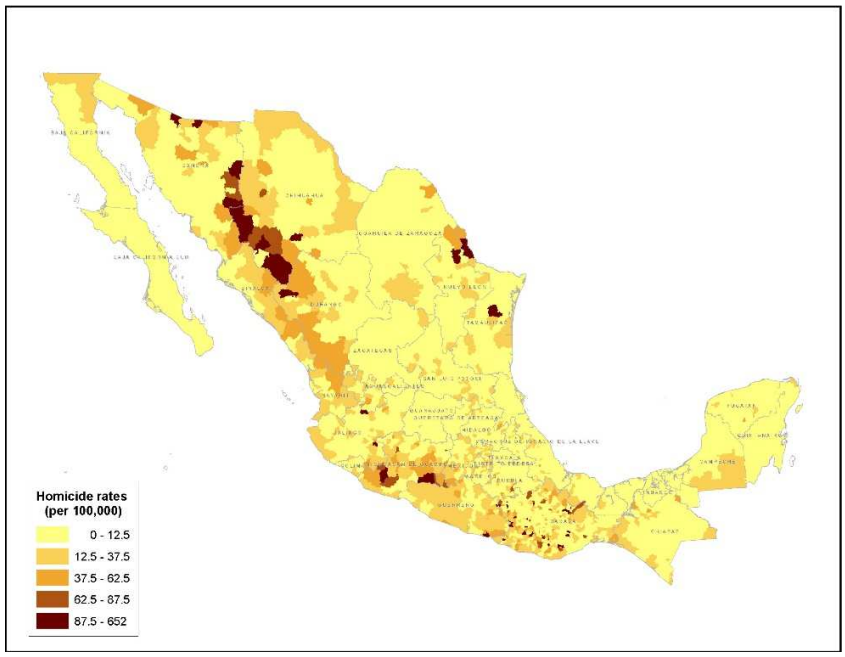

Figure 10: INEGI Annual Homicide Rate - 2005

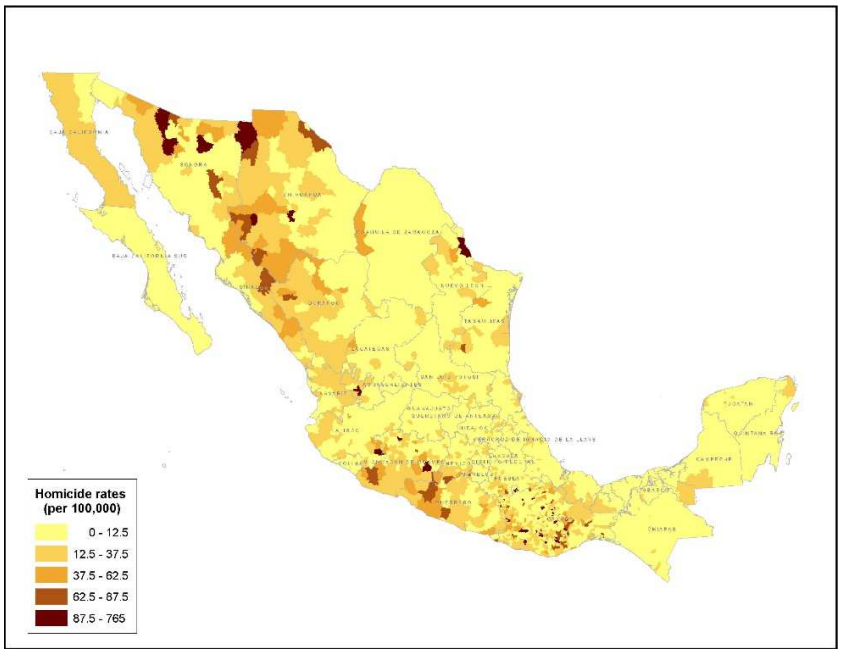

Figure 11: INEGI Annual Homicide Rate - 2007

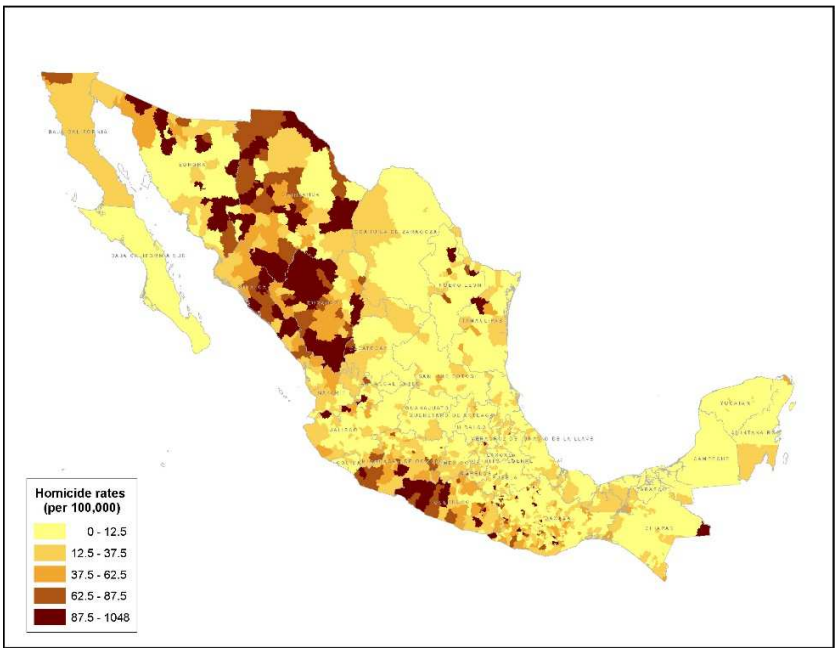

Figure 12: INEGI Annual Homicide Rate - 2009

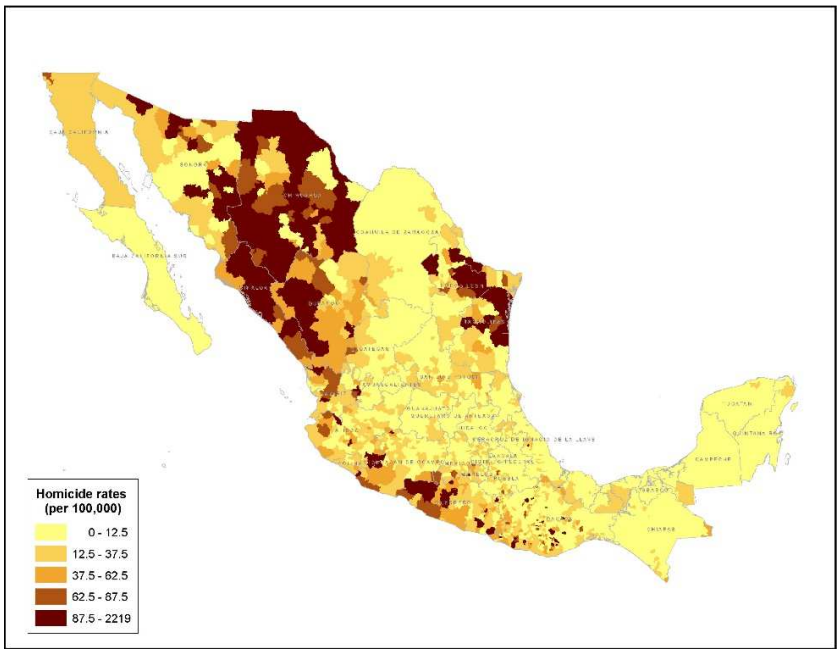

Figure 13: INEGI Annual Homicide Rate - 2010

Table 22: Sample Sizes and Recontact Rates in MxFLS

A. MxFLS2

	Eligible for survey	Recontact	% Recontact
Total	35,134	31,338	89.20
In Mexico	34,280	30,564	89.16
In US	854	774	90.63

B. MxFLS3

Total	34,360	29,798	86.72
In Mexico	32,551	28,206	86.65
In US	1,809	1,592	88.00
US sample ivw in MX		563	
US sample ivw in US		1,001	

Source: MxFLS

Note - Excluded panel respondents who died between waves

Table 23: Employment Transitions

A. By Gender

Males

2009	Not working	Working
2005		
Not working	9.2	9.7
Working	9.2	72.0

Females

2009	Not working	Working
2005		
Not working	51.2	15.5
Working	11.7	21.6

B. By Employment Status and Gender

Males

2009	Not working	Employee	Self-Employed	Unpaid	DK
2005					
Not Working	**9.2**	6.1	2.6	0.6	0.5
Employee	5.2	**37.7**	7.5	0.9	2.0
Self-employed	3.0	6.1	**10.9**	0.9	0.5
Unpaid	0.6	1.8	1.4	**0.2**	0.1
DK employment category	0.4	1.2	0.6	0.0	**0.3**

Females

2009	Not working	Employee	Self-Employed	Unpaid	DK
2005					
Not Working	**51.2**	8.5	4.9	1.5	0.5
Employee	6.2	**11.7**	1.6	0.4	0.6
Self-employed	3.8	1.1	**3.1**	0.6	0.1
Unpaid	1.4	0.7	0.9	**0.2**	0.0
DK employment category	0.4	0.5	0.2	0.0	**0.1**

Source: MxFLS2 and MxFLS3

Table 24: Prediction of Migration

Respondents Age 18 and older in 2005	**MEN**		**WOMEN**	
	Migration=100		Migration=100	
Variables measured in 2005	(1)	(2)	(3)	(4)
Δ Hom Rate (2009-05)	0.38	5.89	2.73	-3.36
	[2.725]	[13.721]	[2.807]	[13.061]
Δ Hom Rate (2009-05) *interacted with*				
Age		-0.11		-0.05
		[0.131]		[0.143]
Education		0.05		0.82
		[0.642]		[0.714]
Married		-1.06		-9.68**
		[3.983]		[4.802]
Worked last week		-13.18+		12.8
		[7.263]		[11.239]
Log(Earnings last 12 months)		0.73		-0.85
		[0.637]		[0.902]
Self-employed		10.05**		-1.21
		[4.572]		[9.738]
Rural		5.37		14.34**
		[6.831]		[6.201]
Sample size	6,310	6,310	8,428	8,428
Mean dependent variable	11.43	11.43	10.94	10.94
R-squared	0.030	0.031	0.028	0.032
F-test jointly =0 - P-value	0.338	0.338	0.006	0.006

Standard errors clustered at municipality level

*** p<0.01, ** p<0.05, + p<0.1

Notes: Includes controls for baseline characteristics (age, years of education, cognitive score, marital status, household characteristics, labor characteristics, whether respondent has relatives in US and place of residence rural/urban characteristic)

Table 25: Labor Outcomes of Women Age [18-75]

Panel A. Self-Employed in 2005

	(1)	(2)	(3)	(4)
	Worked last week (1)	Log(Hours worked last 12 months)	$\sqrt[4]{}$(Hourly Earnings)	$\sqrt[4]{}$(Earnings last 12 months)
(1) Homicide Rate: (0-5] [1]	-0.18***	-0.70**	-0.22	-2.61**
	[0.059]	[0.313]	[0.278]	[1.115]
(1) Homicide Rate: (5-15]	-0.19***	-0.31	0.24	0.51
	[0.052]	[0.280]	[0.244]	[1.057]
(1) Homicide Rate: (15-+]	-0.22***	-0.73**	0.44	0.83
	[0.071]	[0.314]	[0.276]	[1.123]
Constant	1.83***	8.29***	0.34	13.78**
	[0.383]	[1.488]	[1.342]	[6.773]
Sample size	780	327	327	327
R-squared	0.47	0.13	0.09	0.08

Panel B. Employee in 2005

	(1)	(2)	(3)	(4)
	Worked last week (1)	Log(Hours worked last 12 months)	$\sqrt[4]{}$(Hourly Earnings)	$\sqrt[4]{}$(Earnings last 12 months)
(1) Homicide Rate: (0-5] [1]	-0.03	-0.23	0.13	0.23
	[0.049]	[0.151]	[0.117]	[0.462]
(1) Homicide Rate: (5-15]	-0.05	-0.11	0.13	0.6
	[0.045]	[0.148]	[0.116]	[0.416]
(1) Homicide Rate: (15-+]	-0.03	-0.01	0.04	0.41
	[0.048]	[0.168]	[0.125]	[0.482]
Constant	1.20***	7.25***	1.86***	12.42***
	[0.124]	[0.324]	[0.190]	[1.174]
Sample size	1,985	1,160	1,160	1,160
R-squared	0.36	0.06	0.02	0.03

Standard errors clustered at municipality level

*** p<0.01, ** p<0.05, + p<0.1

1. Omitted variable is homicide rate = 0

Note: All models include marital status, household composition, rural/urban, migration expectations, preferences, emotional status, presence of relatives in US, year and quarter of interview and state GDP

Table 26: Labor Outcomes of Self-Employed Women Age [18-75]

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Worked last week (1)	Log(Hours worked last 12 months)	$\sqrt[4]{}$(Hourly Earnings)	$\sqrt[4]{}$(Earnings last 12 months)	Worked last week (1)	Log(Hours worked last 12 months)	$\sqrt[4]{}$(Hourly Earnings)	$\sqrt[4]{}$(Earnings last 12 months)
(1) Homicide Rate: (0-5] [1]	-0.12+	-0.69**	-0.21	-2.81**	-0.13	-0.96	0.25	-2.6
	[0.067]	[0.328]	[0.280]	[1.194]	[0.117]	[0.941]	[0.485]	[2.284]
(1) Homicide Rate: (5-15]	-0.13**	-0.37	0.3	0.7	-0.14+	-0.12	0.32	0.82
	[0.054]	[0.295]	[0.246]	[1.060]	[0.071]	[0.394]	[0.398]	[1.694]
(1) Homicide Rate: (15-+]	-0.17**	-0.80**	0.55+	1.15	-0.24***	-0.19	0.28	0.49
	[0.075]	[0.339]	[0.290]	[1.202]	[0.090]	[0.415]	[0.418]	[1.707]
(1) Homicide Rate: (0-5] [1] * Has children	-0.29**	0.33	-0.47	-0.19				
	[0.139]	[1.112]	[0.800]	[4.302]				
(1) Homicide Rate: (5-15] * Has children	-0.29***	0.73	-0.75	-2.48				
	[0.097]	[0.984]	[0.701]	[3.821]				
(1) Homicide Rate: (15-+] * Has children	-0.26**	0.71	-0.91	-2.74				
	[0.122]	[1.021]	[0.712]	[3.717]				
(1) Homicide Rate: (0-5] [1] * Upper 50th pctile PCE					-0.07	0.15	-0.59	0.01
					[0.143]	[1.089]	[0.586]	[3.182]
(1) Homicide Rate: (5-15] * Upper 50th pctile PCE					-0.08	-0.47	-0.08	-0.35
					[0.066]	[0.369]	[0.447]	[2.036]
(1) Homicide Rate: (15-+] * Upper 50th pctile PCE					0.03	-1.00**	0.29	0.66
					[0.089]	[0.470]	[0.496]	[2.201]
Constant	1.78***	8.39***	0.27	13.44**	1.83***	8.12***	0.36	13.99**
	[0.385]	[1.466]	[1.330]	[6.751]	[0.378]	[1.517]	[1.348]	[6.783]
Sample size	780	327	327	327	780	327	327	327
R-squared	0.47	0.13	0.09	0.09	0.47	0.15	0.09	0.09

Standard errors clustered at municipality level; *** p<0.01, ** p<0.05, + p<0.1

1. Omitted variable is homicide rate = 0

Note: All models include marital status, household composition, rural/urban, migration expectations, preferences, emotional status, presence of relatives in US, year and quarter of interview and state GDP

Table 27: Self-Employed Women by Occupation Age [18-75]

Panel A	**Retail/Commerce**			
	(1)	(2)	(3)	(4)
	Worked last week (1)	Log(Hours worked last 12 months)	$\sqrt[4]{}$(Hourly Earnings)	$\sqrt[4]{}$(Earnings last 12 months)
(1) Homicide Rate: (0-5] [1]	-0.19**	-0.53	-0.1	-2.44
	[0.077]	[0.424]	[0.506]	[2.361]
(1) Homicide Rate: (5-15]	-0.20***	-0.29	0.23	0.34
	[0.061]	[0.288]	[0.441]	[2.317]
(1) Homicide Rate: (15-+]	-0.20**	-0.57	0.46	0.63
	[0.076]	[0.347]	[0.479]	[2.372]
Constant	2.05***	6.74***	1.76	16.37
	[0.436]	[1.703]	[2.425]	[14.960]
Sample size	392	152	152	152
R-squared	0.489	0.162	0.0504	0.0816

Panel B	**Manufacturing**			
	(1)	(2)	(3)	(4)
	Worked last week (1)	Log(Hours worked last 12 months)	$\sqrt[4]{}$(Hourly Earnings)	$\sqrt[4]{}$(Earnings last 12 months)
(1) Homicide Rate: (0-5] [1]	-0.11	-0.83	-1.39	-7.51**
	[0.131]	[1.010]	[0.829]	[3.095]
(1) Homicide Rate: (5-15]	-0.17+	-0.15	-0.33	-1.41
	[0.094]	[0.551]	[0.417]	[1.670]
(1) Homicide Rate: (15-+]	-0.17	-0.8	0.03	-1.03
	[0.129]	[0.532]	[0.492]	[1.980]
Constant	2.48***	6.49	4.27	23.21
	[0.590]	[4.640]	[3.594]	[14.536]
Sample size	173	75	75	75
R-squared	0.559	0.312	0.335	0.306

Panel C	**Personal Services/Domestic Employee**			
	(1)	(2)	(3)	(4)
	Worked last week (1)	Log(Hours worked last 12 months)	$\sqrt[4]{}$(Hourly Earnings)	$\sqrt[4]{}$(Earnings last 12 months)
(1) Homicide Rate: (0-5] [1]	-0.05	-1.4	0.41	-1.31
	[0.163]	[0.842]	[0.625]	[2.824]
(1) Homicide Rate: (5-15]	-0.08	-0.45	0.83+	2.97
	[0.153]	[0.611]	[0.488]	[2.005]
(1) Homicide Rate: (15-+]	0.14	-0.63	0.90+	3.19
	[0.187]	[0.662]	[0.507]	[2.192]
Constant	-1.16	8.80**	0.8	13.47
	[0.902]	[3.845]	[2.921]	[15.193]
Sample size	110	62	62	62
R-squared	0.483	0.28	0.192	0.14

Standard errors clustered at municipality level; *** p<0.01, ** p<0.05, + p<0.1

1. Omitted variable is homicide rate = 0

Note: All models include marital status, household composition, rural/urban, migration expectations, preferences, emotional status, presence of relatives in US, year, quarter of interview and state GDP

Table 28: Labor Outcomes of Men Age [18-75]

Panel A. Self-Employed in 2005

	(1)	(2)	(3)	(4)
	Worked last week (1)	Log(Hours worked last 12 months)	$\sqrt[4]{}$(Hourly Earnings)	$\sqrt[4]{}$(Earnings last 12 months)
(1) Homicide Rate: (0-5] [1]	0.03	-0.03	-0.06	-0.54
	[0.028]	[0.130]	[0.160]	[0.824]
(1) Homicide Rate: (5-15]	0.01	0.06	-0.11	-0.45
	[0.022]	[0.096]	[0.114]	[0.553]
(1) Homicide Rate: (15-+]	-0.01	0.09	-0.29**	-1.55**
	[0.024]	[0.126]	[0.128]	[0.598]
Constant	0.90***	7.99***	1.84***	15.84***
	[0.066]	[0.276]	[0.371]	[1.971]
Sample size	1,658	1,069	1,069	1,073
R-squared	0.14	0.03	0.04	0.04

Panel B. Employee in 2005

	(1)	(2)	(3)	(4)
	Worked last week (1)	Log(Hours worked last 12 months)	$\sqrt[4]{}$(Hourly Earnings)	$\sqrt[4]{}$(Earnings last 12 months)
(1) Homicide Rate: (0-5] [1]	-0.01	-0.02	-0.05	-0.29
	[0.015]	[0.086]	[0.052]	[0.300]
(1) Homicide Rate: (5-15]	-0.01	0.11	-0.08+	-0.02
	[0.012]	[0.074]	[0.046]	[0.270]
(1) Homicide Rate: (15-+]	-0.02	0.11	-0.08	-0.14
	[0.014]	[0.086]	[0.050]	[0.315]
Constant	0.90***	7.80***	1.82***	12.62***
	[0.033]	[0.208]	[0.130]	[0.593]
Sample size	4,380	3,327	3,327	3,336
R-squared	0.11	0.02	0.01	0.01

Standard errors clustered at municipality level

*** p<0.01, ** p<0.05, + p<0.1

1. Omitted variable is homicide rate = 0

Note: All models include marital status, household composition, rural/urban, migration expectations, preferences, emotional status, presence of relatives in US, year and quarter of interview and state GDP

Table 29: Labor Outcomes of Self-Employed Men Age [18-75]

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Worked last week (1)	Log(Hours worked last 12 months)	$\sqrt[4]{}$(Hourly Earnings)	$\sqrt[4]{}$(Earnings last 12 months)	Worked last week (1)	Log(Hours worked last 12 months)	$\sqrt[4]{}$(Hourly Earnings)	$\sqrt[4]{}$(Earnings last 12 months)
(1) Homicide Rate: (0-5] [1]	0.02	-0.09	-0.09	-0.76	-0.01	-0.28	0.09	0.15
	[0.032]	[0.139]	[0.174]	[0.888]	[0.030]	[0.190]	[0.777]	[0.141]
(1) Homicide Rate: (5-15]	0.01	0.01	-0.1	-0.52	0	-0.07	0.51	0.13
	[0.023]	[0.096]	[0.119]	[0.583]	[0.028]	[0.129]	[0.553]	[0.110]
(1) Homicide Rate: (15-+]	-0.01	0.06	-0.31**	-1.68***	-0.03	-0.09	-0.86	-0.05
	[0.025]	[0.125]	[0.125]	[0.633]	[0.030]	[0.164]	[0.578]	[0.115]
(1) Homicide Rate: (0-5] [1] * Has children	0.05	0.34	0.11	1.13				
	[0.072]	[0.333]	[0.442]	[2.437]				
(1) Homicide Rate: (5-15] * Has children	0.01	0.29	-0.14	0.26				
	[0.050]	[0.303]	[0.270]	[1.146]				
(1) Homicide Rate: (15-+] * Has children	0.02	0.15	0.11	0.64				
	[0.051]	[0.305]	[0.251]	[0.972]				
(1) Homicide Rate: (0-5] [1] * Upper 50th pctile PCE					0.08	0.54**	-1.26	-0.43+
					[0.051]	[0.253]	[1.376]	[0.246]
(1) Homicide Rate: (5-15] * Upper 50th pctile PCE					0.02	0.27	-1.99**	-0.50***
					[0.034]	[0.182]	[0.913]	[0.164]
(1) Homicide Rate: (15-+] * Upper 50th pctile PCE					0.03	0.39+	-1.46	-0.51***
					[0.035]	[0.204]	[1.008]	[0.184]
Constant	0.90***	7.99***	1.84***	15.81***	0.89***	7.99***	15.73***	1.81***
	[0.066]	[0.271]	[0.374]	[2.000]	[0.067]	[0.271]	[1.982]	[0.364]
Sample size	1,658	1,069	1,069	1,073	1,658	1,069	1,073	1,069
R-squared	0.14	0.03	0.04	0.04	0.14	0.03	0.04	0.04

Standard errors clustered at municipality level; *** p<0.01, ** p<0.05, + p<0.1

1. Omitted variable is homicide rate = 0

Note: All models include marital status, household composition, rural/urban, migration expectations, preferences, emotional status, presence of relatives in US, year and quarter of interview and state GDP

Table 30: Self-Employed Men by Occupation - Age [18-75]

	SELF-EMPLOYED			
Panel A	**Agr/Hunt/Fish wrker**			
	(1)	(2)	(3)	(4)
	Worked last week (1)	Log(Hours worked last 12 months)	$\sqrt[4]{}$(Hourly Earnings)	$\sqrt[4]{}$(Earnings last 12 months)
(1) Homicide Rate: (0-5] [1]	0.01	0.19	0.04	0.84
	[0.044]	[0.229]	[0.188]	[0.955]
(1) Homicide Rate: (5-15]	-0.01	0.06	-0.14	-0.23
	[0.035]	[0.181]	[0.151]	[0.793]
(1) Homicide Rate: (15-+]	0.01	0.03	-0.21	-1.11
	[0.038]	[0.179]	[0.173]	[0.756]
Constant	1.02***	7.39***	1.77***	12.15***
	[0.087]	[0.403]	[0.407]	[1.980]
Sample size	627	384	384	384
R-squared	0.175	0.0506	0.0423	0.0536

Panel B	**Manufacturing**			
	(1)	(2)	(3)	(4)
	Worked last week (1)	Log(Hours worked last 12 months)	$\sqrt[4]{}$(Hourly Earnings)	$\sqrt[4]{}$(Earnings last 12 months)
(1) Homicide Rate: (0-5] [1]	0.08	0.1	-0.31	-1.62
	[0.061]	[0.268]	[0.334]	[1.692]
(1) Homicide Rate: (5-15]	0.01	0.23	-0.3	-1.17
	[0.037]	[0.191]	[0.251]	[1.276]
(1) Homicide Rate: (15-+]	-0.04	0.15	-0.3	-1.71
	[0.046]	[0.273]	[0.310]	[1.515]
Constant	1.07***	7.80***	-0.47	6.75
	[0.196]	[1.581]	[1.867]	[7.572]
Sample size	425	322	322	323
R-squared	0.124	0.0356	0.0897	0.0995

Panel C	**Retail/Commerce**			
	(1)	(2)	(3)	(4)
	Worked last week (1)	Log(Hours worked last 12 months)	$\sqrt[4]{}$(Hourly Earnings)	$\sqrt[4]{}$(Earnings last 12 months)
(1) Homicide Rate: (0-5] [1]	0.05	-0.31	-0.07	-1.7
	[0.085]	[0.333]	[0.570]	[3.234]
(1) Homicide Rate: (5-15]	-0.02	-0.32	-0.27	-3.57
	[0.066]	[0.292]	[0.486]	[2.717]
(1) Homicide Rate: (15-+]	-0.01	0.06	-0.68	-4.98+
	[0.072]	[0.292]	[0.500]	[2.778]
Constant	0.74**	6.91***	4.71***	25.30***
	[0.333]	[1.521]	[1.489]	[7.529]
Sample size	278	182	182	183
R-squared	0.209	0.0719	0.1	0.0797

Standard errors clustered at municipality level; *** p<0.01, ** p<0.05, + p<0.1

1. Omitted variable is homicide rate = 0

Note: All models include marital status, household composition, rural/urban, migration expectations, preferences, emotional status, presence of relatives in US, year, quarter of interview and state GDP

Table 31: Log(Per Capita Expenditure)

Panel A. All Households				
	(1)	(2)	(3)	(4)
	Log(Total PCE)	Log(PCE Food)	Log("Conspicuous" PCE)	Log(Education)
(1) Homicide Rate: (0-5] [1]	-0.12+	-0.06	-0.13	-0.09
	[0.059]	[0.058]	[0.106]	[0.095]
(1) Homicide Rate: (5-15]	-0.09+	-0.05	-0.1	0.05
	[0.053]	[0.052]	[0.094]	[0.072]
(1) Homicide Rate: (15-+]	-0.05	-0.04	-0.04	0
	[0.055]	[0.056]	[0.100]	[0.085]
Constant	7.15***	6.76***	5.34***	3.21***
	[0.061]	[0.062]	[0.101]	[0.119]
Sample size	8,350	8,350	8,037	5,752
R-squared	0.148	0.166	0.0406	0.0683

Panel B. Households with Female Househol Head				
	(1)	(2)	(3)	(4)
	Log(Total PCE)	Log(PCE Food)	Log("Conspicuous" PCE)	Log(Education)
(1) Homicide Rate: (0-5] [1]	-0.11	-0.08	-0.01	-0.40**
	[0.072]	[0.070]	[0.168]	[0.201]
(1) Homicide Rate: (5-15]	-0.13**	-0.11+	-0.13	-0.13
	[0.065]	[0.063]	[0.162]	[0.174]
(1) Homicide Rate: (15-+]	-0.06	-0.06	0.02	-0.12
	[0.063]	[0.063]	[0.158]	[0.177]
Constant	6.97***	6.61***	5.57***	3.62***
	[0.069]	[0.067]	[0.168]	[0.230]
Sample size	2,867	2,867	2,747	1,899
R-squared	0.19	0.207	0.0585	0.0515

Panel C. Households with Male Househol Head				
	(1)	(2)	(3)	(4)
	Log(Total PCE)	Log(PCE Food)	Log("Conspicuous" PCE)	Log(Education)
(1) Homicide Rate: (0-5] [1]	-0.13+	-0.06	-0.19	0.01
	[0.071]	[0.062]	[0.120]	[0.121]
(1) Homicide Rate: (5-15]	-0.07	-0.02	-0.08	0.11
	[0.063]	[0.057]	[0.102]	[0.096]
(1) Homicide Rate: (15-+]	-0.05	-0.02	-0.07	0.04
	[0.069]	[0.063]	[0.107]	[0.109]
Constant	6.79***	6.31***	5.57***	3.26***
	[0.070]	[0.068]	[0.134]	[0.156]
Sample size	5,483	5,483	5,290	3,853
R-squared	0.126	0.147	0.0363	0.0846

Standard errors clustered at municipality level; *** p<0.01, ** p<0.05, + p<0.1

1. Omitted variable is homicide rate = 0

Note: All models include marital status, household composition, rural/urban, migration expectations, preferences, emotional status, presence of relatives in US, year and quarter of interview and state GDP

Table 32: Placebo Test - Labor Outcomes of Women Age [18-75]

Panel A. Self-Employed in 2002

	(1)	(2)	(3)	(4)
	Worked last week (1)	Log(Hours worked last 12 months)	$\sqrt[4]{}$(Hourly Earnings)	$\sqrt[4]{}$(Earnings last 12 months)
(1) Homicide Rate: (0-5] [1]	0.02	-0.03	0.26	0.07
	[0.078]	[0.275]	[0.331]	[1.707]
(1) Homicide Rate: (5-15]	0.04	-0.04	0.2	0.49
	[0.066]	[0.268]	[0.237]	[1.374]
(1) Homicide Rate: (15-+]	-0.03	0.11	0.12	0.04
	[0.086]	[0.305]	[0.285]	[1.470]
Constant	0.73***	6.23***	1.62***	7.39***
	[0.113]	[0.364]	[0.324]	[2.457]
Sample size	820	357	358	367
R-squared	0.12	0.04	0.07	0.04

Panel B. Employee in 2002

	(1)	(2)	(3)	(4)
	Worked last week (1)	Log(Hours worked last 12 months)	$\sqrt[4]{}$(Hourly Earnings)	$\sqrt[4]{}$(Earnings last 12 months)
(1) Homicide Rate: (0-5] [1]	0.05	-0.13	-0.11	-0.81
	[0.054]	[0.123]	[0.136]	[0.962]
(1) Homicide Rate: (5-15]	0.07	-0.11	-0.04	-0.28
	[0.047]	[0.115]	[0.136]	[0.861]
(1) Homicide Rate: (15-+]	0.07	-0.21+	0.12	0.42
	[0.051]	[0.120]	[0.160]	[0.981]
Constant	0.77***	7.35***	2.44***	13.18***
	[0.068]	[0.251]	[0.191]	[1.127]
Sample size	1,641	1,106	1,106	1,112
R-squared	0.17	0.03	0.11	0.02

Standard errors clustered at municipality level

*** p<0.01, ** p<0.05, + p<0.1

1. Omitted variable is homicide rate = 0

Note: All models include marital status, household composition, rural/urban, migration expectations, preferences, emotional status, presence of relatives in US, year and quarter of interview and state GDP

Table 33: Placebo Test - Labor Outcomes of Men Age [18-75]

Panel A. Self-Employed in 2002

	(1)	(2)	(3)	(4)
	Worked last week (1)	Log(Hours worked last 12 months)	$\sqrt[4]{}$(Hourly Earnings)	$\sqrt[4]{}$(Earnings last 12 months)
(1) Homicide Rate: (0-5] [1]	-0.03	-0.1	0.67	0.08
	[0.039]	[0.151]	[0.756]	[0.149]
(1) Homicide Rate: (5-15]	-0.03	-0.01	0.64	0.04
	[0.029]	[0.109]	[0.621]	[0.098]
(1) Homicide Rate: (15-+]	0.01	-0.09	0.93+	0.13
	[0.034]	[0.108]	[0.558]	[0.102]
Constant	0.75***	7.31***	5.91***	1.35***
	[0.049]	[0.232]	[1.336]	[0.216]
Sample size	1,686	1,257	1,271	1,258
R-squared	0.04	0.03	0.02	0.02

Panel B. Employee in 2002

	(1)	(2)	(3)	(4)
	Worked last week (1)	Log(Hours worked last 12 months)	$\sqrt[4]{}$(Hourly Earnings)	$\sqrt[4]{}$(Earnings last 12 months)
(1) Homicide Rate: (0-5] [1]	0.02	-0.07	-0.25	-0.01
	[0.017]	[0.069]	[0.501]	[0.065]
(1) Homicide Rate: (5-15]	0.02	-0.10+	0.22	0.07
	[0.017]	[0.050]	[0.501]	[0.061]
(1) Homicide Rate: (15-+]	0.03	-0.09	0.17	0.06
	[0.018]	[0.068]	[0.601]	[0.075]
Constant	0.89***	7.52***	13.06***	2.15***
	[0.036]	[0.127]	[0.718]	[0.104]
Sample size	3,625	3,130	3,152	3,133
R-squared	0.03	0.01	0.03	0.06

Standard errors clustered at municipality level

*** p<0.01, ** p<0.05, + p<0.1

1. Omitted variable is homicide rate = 0

Note: All models include marital status, household composition, rural/urban, migration expectations, preferences, emotional status, presence of relatives in US, year and quarter of interview and state GDP

Table 34: Placebo Test - Log(Per Capita Expenditure)

Panel A. All Households				
	(1)	(2)	(3)	(4)
	Log(Total PCE)	Log(PCE Food)	Log("Conspicuous" PCE)	Log(Education)
(1) Homicide Rate: (0-5] [1]	0.08	0.06	0.12	0.01
	[0.058]	[0.051]	[0.090]	[0.101]
(1) Homicide Rate: (5-15]	0.07	0.03	0.13	-0.01
	[0.054]	[0.052]	[0.081]	[0.086]
(1) Homicide Rate: (15-+]	0	-0.03	0.08	-0.08
	[0.060]	[0.057]	[0.101]	[0.090]
Constant	5.95***	6.51***	5.21***	3.28***
	[0.093]	[0.071]	[0.151]	[0.190]
Sample size	7,447	7,447	7,208	5,170
R-squared	0.159	0.116	0.0256	0.102
Panel B. Households with Female Househol Head				
	(1)	(2)	(3)	(4)
	Log(Total PCE)	Log(PCE Food)	Log("Conspicuous" PCE)	Log(Education)
(1) Homicide Rate: (0-5] [1]	0.21**	0.19***	0.02	-0.2
	[0.089]	[0.069]	[0.133]	[0.153]
(1) Homicide Rate: (5-15]	0.16**	0.12+	0.19	-0.08
	[0.079]	[0.068]	[0.124]	[0.154]
(1) Homicide Rate: (15-+]	0.08	0.07	0.03	-0.18
	[0.081]	[0.067]	[0.148]	[0.164]
Constant	6.10***	6.62***	4.78***	3.56***
	[0.136]	[0.117]	[0.237]	[0.309]
Sample size	2,252	2,252	2,159	1,475
R-squared	0.257	0.167	0.0634	0.164
Panel C. Households with Male Househol Head				
	(1)	(2)	(3)	(4)
	Log(Total PCE)	Log(PCE Food)	Log("Conspicuous" PCE)	Log(Education)
(1) Homicide Rate: (0-5] [1]	0.03	0.01	0.17	0.08
	[0.064]	[0.058]	[0.102]	[0.121]
(1) Homicide Rate: (5-15]	0.03	-0.01	0.1	0.02
	[0.059]	[0.059]	[0.088]	[0.098]
(1) Homicide Rate: (15-+]	-0.04	-0.08	0.09	-0.03
	[0.071]	[0.069]	[0.107]	[0.102]
Constant	5.81***	6.51***	5.03***	3.03***
	[0.122]	[0.091]	[0.208]	[0.207]
Sample size	5,195	5,195	5,049	3,695
R-squared	0.129	0.101	0.0235	0.0928

Standard errors clustered at municipality level

*** p<0.01, ** p<0.05, + p<0.1

1. Omitted variable is homicide rate = 0

Note: All models include marital status, household composition, rural/urban, migration expectations, preferences, emotional status, presence of relatives in US, year and quarter of interview and state GDP

4.11 Supplementary Tables

Table 35: Prediction of Attrition from MxFLS3 - Men Age 18+

	Attrition=100	
Variables measured in 2005	(1)	(2)
Δ Hom Rate (2009-05)	0.68	-2.57
	[1.585]	[10.510]
Δ Hom Rate (2009-05) *interacted with*		
Age		0.15
		[0.131]
Education		0.83
		[0.532]
Married		-0.44
		[3.218]
Worked last week		-5.45
		[5.755]
Log(Earnings last 12 months)		-0.11
		[0.411]
Self-employed		-3.74
		[2.609]
Rural		-2.84
		[3.571]
Sample size	9,516	9,516
Mean dependent variable	15.46	15.46
R-squared	0.073	0.074
F-test jointly =0 - P-value	0.435	0.435

Standard errors clustered at municipality level

*** p<0.01, ** p<0.05, + p<0.1

Notes: Controls for baseline characteristics (age, years of education, cognitive score, marital status, household characteristics, labor characteristics, whether respondent has relatives in US and place of residence rural/urban characteristic)

Table 36: Prediction of Attrition from MxFLS3 - Women Age 18+

	Attrition=100	
Variables measured in 2005	(1)	(2)
Δ Hom Rate (2009-05)	0.45	0.34
	[1.669]	[10.327]
Δ Hom Rate (2009-05) *interacted with*		
Age		0.03
		[0.126]
Education		0.08
		[0.634]
Married		-1.51
		[3.678]
Worked last week		4.95
		[5.648]
Log(Earnings last 12 months)		0.07
		[0.662]
Self-employed		1.36
		[5.440]
Rural		-3.87
		[3.863]
Sample size	10,888	10,888
Mean dependent variable	13.43	13.43
R-squared	0.079	0.079
F-test jointly =0 - P-value	0.176	0.176

Standard errors clustered at municipality level

*** p<0.01, ** p<0.05, + p<0.1

Notes: Controls for baseline characteristics (age, years of education, cognitive score, marital status, household characteristics, labor characteristics, whether respondent has relatives in US and place of residence rural/urban characteristic)

Table 37: Prediction of Changes on Homicide Rate between 2005 and 2009

	(1)	(2)		(3)	(4)
Variables measured in 2005	MxFLS Individual Variables	Census Variables	Changes between 2002 and 2005	MxFLS Individual Variables	Census Variables
Years of Education	-0.26		Years of Education	0.99	
	[1.166]			[5.414]	
Married	-6.29		Married	-11.51	
	[18.604]			[52.449]	
Household Size	-3.05		Household Size	-3.39	
	[3.714]			[6.112]	
Worked last week - Female (1)	-41.59+		Worked last week - Female (1)	27.68	
	[21.187]			[49.251]	
Worked last weekd - Male (1)	18.69		Worked last weekd - Male (1)	-13.74	
	[30.330]			[57.241]	
Self-employed - Female (1)	16.79		Self-employed - Female (1)	-50.55	
	[14.665]			[42.449]	
Self-employed - Male (1)	-2.38		Self-employed - Male (1)	58.77	
	[18.124]			[45.018]	
Log Hourly Earnings - Female	-0.5		Log Hourly Earnings - Female	-5.62	
	[5.903]			[5.878]	
Log Hourly Earnings - Male	2.7		Log Hourly Earnings - Male	4.63	
	[5.729]			[6.911]	
Rural (1)	-1.51		Rural (1)	8.04	
	[5.657]			[7.030]	
Has relatives in US (1)	6.99		Has relatives in US (1)	-30.03+	
	[9.347]			[16.258]	
Log PCE	8.28		Log PCE	10.01	
	[9.667]			[13.459]	
Thoughts about future migration (1)	-8.27		Thoughts about future migration (1)	-27.98	
	[12.054]			[27.547]	
Fear in the day (1)	6.89		Fear in the day (1)	8.86	
	[39.344]			[57.344]	
Fear in the night (1)	-40.32		Fear in the night (1)	-48.2	
	[37.014]			[59.703]	
Rate of analfabetism		-371.37***	Rate of analfabetism		88.84
		[116.270]			[242.705]
Share dwellings with water		-27.58	Share dwellings with water		71.08
		[20.269]			[46.582]
Share dwelling with phone		-19.29	Share dwelling with phone		-10.56
		[53.626]			[49.221]
Share dwellings with sewage		-21.74	Share dwellings with sewage		-40.16
		[22.958]			[38.232]
Share dwellings with electricity		216.86**	Share dwellings with electricity		51.36
		[98.405]			[69.679]
Percentage with less than primary		171.58**	Percentage with less than primary		-110.62+
		[66.934]			[63.379]
Percentage with High School		70.23***	Percentage with High School		-152.52
		[26.475]			[102.922]
Percentage population younger than 18		123.49	Percentage population younger than 18		9.81
		[79.159]			[116.490]
Percentage population older than 65		-53.71	Percentage population older than 65		-69.31
		[107.284]			[193.190]
Constant	-31.35	-212.79***	Constant	6.36	10.54+
	[58.814]	[71.801]		[6.895]	[5.492]
Sample size	151	193	Sample size	134	193
R squared	0.0819	0.227	R squared	0.102	0.103
Standard errors clustered at municipality level			Standard errors clustered at municipality level		
*** p<0.01, ** p<0.05, + p<0.1			*** p<0.01, ** p<0.05, + p<0.1		

References

Abadie, A. and J. Gardeazabal, J. (2003), "The Economic Costs of Conflict: A Case-Control Study for the Basque Country," *American Economic Review*, 93(1), 113-133.

Abraham, K, Maitland, A. and Bianchi, S. (2006), "Nonresponse in the American Time Use Survey," *Public Opinion Quarterly*, 70(5), 676-703.

Akresh, R., Lucchetti, L. and Thirumurthy, H. (2012), "Wars and Child Health: Evidence from the Eritrean–Ethiopian Conflict," *Journal of Development Economics*, 99(2), 330-340.

Alderman, H., Behrman, J., Kohler, H.P., Maluccio, J.A. and Watkins, S. (2001), "Attrition in Longitudinal Household Survey Data: Some Tests from Three Developing Countries," *Demographic Research*, 5(4), 79-124.

Baez, J. E. (2011), "Civil Wars Beyond their Borders: The Human Capital and Health Consequences of Hosting Refugees," *Journal of Development Economics*, 96(2), 391–408.

Barrera, F. and Ibáñez, A. (2004), "Does Violence Reduce Investment in Education?: A Theoretical and Empirical Approach," Documento CEDE 2004-27, Universidad de los Andes, Bogota, Colombia.

Becker, G. (1965), "A Theory of the Allocation of Time," *Economic Journal*, 75(299), 496-517.

Becker, G. (1968), "Crime and Punishment: An Economic Approach," *Journal of Political Economy*, 76(2), 169-217.

Becketti, S., Gould W., Lillard, L., and Welch, F. (1988), "The Panel Study of Income Dynamics after Fourteen Years: An Evaluation," *Journal of Labor Economics*, 6(4), 472-92.

Beegle, K. (2005), "Labor Effects of Adult Mortality in Tanzanian Households," *Economic Development and Cultural Change*, 53(3), 655-684.

BenYishai, A. and Pearlman, S. (2013), "Homicide and Work: The Impact of Mexico's Drug War on Labor Force Participation," Working Paper.

Bozzoli, C., Brück, T. and Wald, N. (2013), "Self-Employment and Conflict in Colombia," *Journal of Conflict Resolution*, 57(1), 117-142.

Brown, R. and Velásquez, A. (2013), "The Effect of Violent Conflict in Mexico on the Human Capital Accumulation of Children and Young Adults," Duke University Working Paper.

Bundervoet, T., Verwimp, P. and Akresh, R. (2009), "Health and Civil War in Rural Burundi," *Journal of Human Resources*, 44(2), 536-563.

Calderón, V., Gáfaro, M. and Ibáñez, A.M. (2001), "Forced Migration, Female Labor Force Participation, and Intra-household Bargaining: Does Conflict Empower Women?," MICROCON Research Working Paper 56.

Camacho, A. and Rodriguez, C. (2013). "Firm Exit and Armed Conflict in Colombia" *Journal of Conflict Resolution*, 57(1), 89-116.

Castillo, J.C., Mejia, D. and Restrepo, P. (2013), "Illegal Drug Markets and Violence in Mexico: The Causes beyond Calderon," Universidad de los Andes, Working Paper.

Collier, P. (1999), "On the Economic Consequences of Civil War," *Oxford Economic Papers*, 51(1), 168–183.

Collier, P. and Duponchel, M. (2013). ''The Economic Legacy of Civil War: Firm Level Evidence from Sierra Leone,'' *Journal of Conflict Resolution*, 57(1), 65-88.

Cunningham, W. (2001), "Breadwinner or Caregiver? How Household Role Affects Labor Choices In Mexico" World Bank Policy Research Working Paper WPS2743.

Deininger, K. (2003), "Causes and Consequences of Civil Strife - Micro-Level Evidence from Uganda," World Bank Policy Research Working Paper Series WPS3045.

Dell, M. (2011), "Trafficking Networks and the Mexican Drug War," Forthcoming AER.

Díaz-Cayeros, A., Magaloni, B., Matanock, A. and Romero, V. (2011), "Living in Fear: Mapping the Social Embeddedness of Drug Gangs and Violence in Mexico," Working Paper.

Donato, K. M., Durand, J. and Massey, D. (1992), "Stemming the Tide? Assessing the Deterrent Effects of the Immigration Reform and Control Act," *Demography*, 29(2), 139–157.

Durand, J. and Massey, D. (1992), "Mexican Migration to the United States: A Critical Review," *Latin American Research Review,* 27(2), 3-43.

Durand, J., Kandel, W., Parrado, E. and Massey, D. (1996), "International Migration and Development in Mexican Communities," *Demography*, 33(2), 249-64.

Durand, J., Massey, D. and Zenteno, R. (2001), "Mexican Migration to the United States: Continuities and Changes," *Latin American Research Review,* 36(1), 107-27.

Echarri Canovas, C. (2011), "Feminicidio en México. Aproximación, Tendencias y Cambios, 1985-2009" ONU Mujeres, Entidad de las Naciones Unidas para la Igualdad de Género y el Empoderamiento de las Mujeres, Mexico.

Farfan, M., M. Genoni, L. Rubalcava, G. Teruel, D. Thomas and A. Velasquez (2012), "Mexicans in America," Duke University Working Paper.

Fernández-Huertas Moraga, J. (2011), "New Evidence on Emigrant Selection," *The Review of Economics and Statistics*, 93(1), 72-96.

Fernandez, M., Ibanez, A.M. and Peña, X. (2011), "Adjusting the Labor Supply to Mitigate Violent Shocks: Evidence from Rural Colombia," World Bank Policy Research Working Paper Series WPS5684.

Fitzgerald, J., Gottschalk, P. and Moffitt, R. (1998), "An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics," *Journal of Human Resources*, 33(2), 251-299.

Frankenberg, E., Smith, J. and Thomas, D. (2003), "Economic Shocks, Wealth and Welfare," *Journal of Human Resources*. 38(2): 280-321

Groves, R. M. and Couper, M. P. (1998), *Nonresponse in Household Interview Surveys*. New York: Wiley.

Guerrero-Gutiérrez, E. (2011a), "Security, Drugs, and Violence in Mexico: A Survey," 7th North American Forum, Washington D.C.

Guerrero-Gutiérrez, E. (2011b), "La Raíz de la Violencia," Revista Nexos, June, Available online: http://www.nexos.com.mx/?p=14318.

Guerrero-Gutiérrez, E. (2012a), "2011: La Dispersión de la Violencia," Revista Nexos, February, Available online: http://www.nexos.com.mx/?p=14705.

Guerrero-Gutiérrez, E. (2012b), "Epidemias de Violencia" Nexos, July, Available online: http://www.nexos.com.mx/?p=14884.

Hanson, G.H., (2006), "Illegal Migration from Mexico to the United States," *Journal of Economic Literature*, 44(4), 869-924.

Hoefer, M., Rytina, N. and Campbell, C. (2006) "Estimates of the Unauthorized Immigrant Population Residing in the United States," Washington, DC: Office of Immigration Statistics, Policy Directorate, U.S. Department of Homeland Security.

Hoeffler, A. and Reynal-Querol, M. (2003), "Measuring the Costs of Conflict," Oxford Working Paper.

Ibáñez, A.M. and Moya, A. (2010), "Vulnerability of Victims of Civil Conflict: Empirical Evidence for the Displaced Population in Colombia," *World Development,* 38(4), 647-663.

Ibarraran, P. and Lubotsky, D. (2007), "Mexican Immigration and Self-Selection: New Evidence from the 2000 Mexican census," in G. J. Borjas (ed.) *Mexican Immigration to the United States.* Chicago: University of Chicago Press.

Justino, P. and Verwimp, P. (2006), "Poverty Dynamics, Violent Conflict and Convergence in Rwanda" HiCN Working Paper 16.

Kondylis, F. (2007), "Conflict Displacement and Labour Market Outcomes in Post-War Bosnia and Hersegovina," *Journal of Development Economics,* 93(2), 235-248.

Leon, G. (2012), "Civil Conflict and Human Capital Accumulation: The Long-term Effects of Political Violence in Perú," *Journal of Human Resources*, 47(4), 991-1022.

Lundberg, S. (1985). "The Added Worker Effect," *Journal of Labor Economics*, 3(1), 11-37.

Maluccio, J. A., (2000). "Attrition in the Kwazulu Natal Income Dynamics Study, 1993-1998," FCND briefs 95, International Food Policy Research Institute (IFPRI).

Massey, D. S., Alarcon, R., Durand, J. and Gonzalez, H. (1990). *Return to Aztlan: The Social Process of International Migration from Western Mexico*, Berkeley: University of California Press.

Massey, D. and Singer, A. (1995), "New Estimates of Undocumented Mexican Migration and the Probability of Apprehension," *Demography*, 32(2), 203-13.

McKenzie, D and Rapoport, H. (2004), "Network Effects and the Dynamics of Migration and Inequality: Theory and Evidence from Mexico," Working Paper.

Mejia, D. and Restrepo, P. (2010), "Crime and Conspicuous Consumption," Documentos CEDE 007716, Universidad de los Andes

Molzahn, C., Rios, V. and Shirk, D. (2012), "Drug Violence in Mexico: Data and Analysis through 2011," Trans-Border Institute Joan B. Kroc School of Peace Studies University of San Diego.

Morduch, J. (1995), "Income Smoothing and Consumption Smoothing," *Journal of Economic Perspectives*, 9(3), 103-114.

Olsen, R.J. (2005), "The Problem of Respondent Attrition: Survey Methodology is Key," *Monthly Labor Review,* 128(2), 63-70.

Pantaleo, K. (2010), "Gendered Violence: An Analysis of the Maquiladora Murders," *International Criminal Justice Review,* 20(4), 349-365.

Pashiva, R. and Suarez, G. (2010), ''Capital Crimes: Kidnappings and Corporate Investment in Colombia'' in R. Di Tella, S. Edwards and E. Schargrodky (eds.) *The Economics of Crime: Lessons for and from Latin America*. Chicago: University of Chicago Press.

Rendall, M., Brownell, P. and Kups, S. (2011), "Declining Return Migration from the United States to Mexico in the Late 2000s Recession," *Demography*, 48(3), 1049-58.

Rios, V. and Shirk, D. (2011), "Drug Violence in Mexico: Data and Analysis through 2010," Trans-Border Institute Joan B. Kroc School of Peace Studies University of San Diego.

Rios, V. (2013), "Why Dis Mexico Become So Violent? A self-reinforcing violent equilibrium caused by competition and enforcement," *Trend in Organized Crime*, 16(2), 138-155.

Rodriguez, C. and Sanchez, F. (2009), "Armed Conflict Exposure, Human Capital Investments and Child Labor: Evidence from Colombia," HICN Working Paper 68.

Robles, G., Magaloni, B., and Calderon G. (2013), "The Economic Costs of Drug-Trafficking Violence in Mexico," Working Paper.

Rosenzweig, M. R. (1986), "Labor Markets in Low Income Countries," in H. Chenery and T. N. Srinivasan (eds.) *Handbook of Development Economics*. Amsterdam: Elsevier- North Holland.

Rubalcava, L. and Teruel, G. (2006), "User's Guide for the Mexican Family Life Survey First Wave", www.mxfls.uia.mx.

Rubin, D. (1987), *Multiple imputation for nonresponse in surveys*. New York: Wiley.

Rubin, D. and Zanutto, E. (2002), "Using Matched Substitutes to Adjust for Nonignorable Nonresponse through Multiple Imputations" in R. Groves, D. Dillman, J. Eltinge and R. Little (eds.) *Survey Nonresponse*. New York: Wiley.

Shemyakina, O. (2010), "The Effect of Armed Conflict on Accumulation of Schooling: Results from Tajikistan," *Journal of Development Economics,* 95(2), 186-200.

Shemyakina, O. (2011), "The Labor Market, Education and Armed Conflict in Tajikistan," World Bank Policy Research Working Paper Series WPS5738.

Shirk, D. (2011), "Drug Violence and State Responses in Mexico," Working Paper University of San Diego.

Shrader, E. (2001), "Methodologies to Measure the Gender Dimensions of Crime and Violence," World Bank Policy Research Working Paper Series.

Smith, J.P., and Thomas, D. (1998), "On the Road: Marriage and Mobility in Malaysia," *Journal of Human Resources* 33(4):805-32.

Sudman, S. and Bradbum, N. (1974), *Response Effects in Surveys*. Chicago: Aldine.

Thomas, D., Frankenberg, E. and Smith, J. P. (2001), "Lost but Not Forgotten: Attrition and Follow-up in the Indonesia Family Life Survey," *Journal of Human Resources*, 36(3), 556-592.

Thomas, D., Beegle, K. and Frankenberg, E. (2003) "Labor Market Transitions of Men and Women during an Economic Crisis: Evidence from Indonesia," in B. Garcia, R. Anker, and A. Pinnelli (eds.) *Women in the Labour Market in Changing Economies: Demographic Issues*. Oxford University Press.

Thomas, D., Frankenberg, E., Friedman, J., Habicht, J.P., Hakimi, M., Ingwersen, N., Jaswadi, Jones, N., McKelvey, K., Pelto, G., Sikoki, B., Seeman, T., Smith, J.P., Sumantri, C., Suriastini, W. and Wilopo, S. (2006), "Causal Effect of Health on Labor Market Outcomes: Experimental Evidence," Working Paper.

Thomas, D. and Frankenberg, E. (2007), "Household Responses to the Financial Crisis in Indonesia: Longitudinal Evidence on Poverty, Resources, and Well-being," in A. Harrison (ed.) *Globalization and Poverty*. Chicago: University of Chicago Press.

Thomas, D., Witoelar, F., Frankenberg, E., Sikoki, B., Strauss, J., Sumantri, C., and Suriastin, W. (2012), "Cutting the costs of attrition: Results from the Indonesia Family Life Survey," *Journal of Development Economics*, 98(1), 108-123.

Townsend, R.M. (1995), "Consumption Insurance: An Evaluation of Risk-Bearing Systems in Low-Income Economies," *Journal of Economic Perspectives,* 9(3), 83-102.

USAID (2007), "Women and Conflict: An Introductory Guide to Programming".

Verwimp, P., Justino, P. and Brück, T. (2009), "The Analysis of Conflict: A Micro Level Perspective," *Journal of Peace Research*, 46(3), 307-314.

Zabel, J. (1998), "An Analysis of Attrition in the Panel Study of Income Dynamics and the Survey of Income and Program Participation with an Application to a Model of Labor Market Behavior," *Journal of Human Resources*, 33(2), 479-506.

Biography

Andrea Velásquez was born January 22nd, 1982 in Bogotá, Colombia. Her research interests lie at the intersection of development, labor and population economics, with a special interest in the topics of data quality, conflict, and migration. She received an M.A. and Ph.D. in Economics from Duke University in 2011 and 2014, respectively. Before pursuing her graduate studies at Duke she received a B.A. in Economics in 2005 and an M.A. in Economics in 2007 from the Universidad de los Andes, Bogotá, Colombia. While completing the M.A. in Colombia, she worked on topics exploring the causes and economic consequences of forced displacement in her home country.

At Duke she was a member of the team that designed and collected the U.S. component of the third wave of the Mexican Family Life Survey. Using this innovative and extremely rich longitudinal data her current work rigorously examines the consequences of attrition in longitudinal surveys, Mexican migration to the United States, and the impact that unanticipated shocks in the socioeconomic environment, such as the recent surge in violent crime in Mexico, have on the labor market, migration, and time allocation of individuals and households. Her active involvement in the fieldwork for the third wave of the MxFLS motivated her to keep thinking about attrition in longitudinal surveys. In 2012, under the guidance of her advisor Professor Duncan Thomas she was awarded the National Science Foundation Doctoral

Dissertation Research Improvement Grant, to conduct post-survey intensive tracking in the United States and Mexico for the MxFLS.

Andrea will begin a position as an Assistant Professor of Economics at the University of Colorado, Denver in August of 2014.