UNIVERSITÀ COMMERCIALE LUIGI BOCCONI

HEALTH AND SOCIAL NETWORKS IN DEVELOPMENT ECONOMICS

Lucia Corno

Matricola 19851
Ph.D. in Economics, XX Cycle

Committee in charge:
Prof. Eliana La Ferrara (chair)
Prof. Martina Bjorkmäm
Prof. Edward Miguel
Prof. Guido Tabellini

To my parents

Contents

Preface **3**

Acknowledgements **5**

Introduction **6**

1 Social Networks and Crime: Evidence from homeless people in Italy **11**

1.1 Introduction . . . 13

1.2 Background . . . 15

1.2.1 Previous Works on social networks . . . 15

1.2.2 Homelessness . . . 16

1.3 The data . . . 19

1.3.1 Survey design . . . 19

1.3.2 Rainfall data . . . 23

1.4 Descriptive statistics . . . 24

1.5 Empirical Strategy . . . 28

1.5.1 Identification . . . 28

1.5.2 Size of the social networks . . . 29

1.5.3 Social networks in dyadic data . . . 30

1.5.4 Potential drawbacks . . . 33

1.6 Results . . . 34

1.6.1 Rainfall . . . 34

1.6.2 Social Networks size and prison . . . 35

1.6.3 Peer effects and prison 37
1.7 Conclusions 40
1.8 References 41

2 Learning (or not) in health seeking behavior: Evidence from rural Tanzania 47
2.1 Introduction 48
2.2 Health care system in Tanzania 51
2.3 Data and Descriptive statistics 53
2.4 Empirical Strategy 59
2.4.1 Estimating equations 59
2.4.2 Identification strategy 61
2.5 Econometric results 63
2.5.1 Switching to formal health care 63
2.5.2 Natural disasters and the probability of illness 65
2.6 Concluding remarks 67
2.7 References 68

3 Getting a job but loosing health:HIV and miners in Lesotho 73
3.1 Introduction 74
3.2 Miners and HIV/AIDS in Lesotho 77
3.3 Data and Descriptive statistics 79
3.4 Empirical Strategy 84
3.5 Results 86
3.5.1 HIV Status and HIV test 86
3.5.2 Condom Use 88
3.5.3 Abstinence during the last 12 months 89
3.5.4 Miners and distance to South Africa Borders 90
3.6 Robustness check 91
3.7 Conclusions 91
3.8 References 92

Preface

This work is an attempt to shed lights on some of the crucial issues affecting every day millions of poor people in developing and richer countries. Despite recent achievements and economic performances of some developing countries, the number and complexity of pending problems is amazing. On the other hand, the number of poor people is rapidly increasing also in richer economies, where basic needs, such as living in a house, are not satisfied for a considerable part of the population. Motivated by this evidence, I started to investigate more deeply the phenomena of homelessness. During Red Cross tours, I realized how many people in Milan are forced to sleep on the street and how this topic is totally excluded from the economic research. The first chapter of the thesis presents and discusses a new data collection among homeless people. I tried to understand the main causes of homelessness and the functioning of social interactions. More specifically, I show the importance of social networks in helping homeless to survive on the street without committing crime.

The second and the third chapters focus on health problems in two African countries, with the lowest life expectancy at birth in the world: Tanzania (51 years old) and Lesotho (40 year old). In Tanzania, I investigate the role of the informal health sectors and the demand for traditional health treatments. Witchcraft beliefs are strong in western Tanzania where a large proportion of the population follows traditional religions and has never adopted Christianity or Islam. A recent example for the relevance of the informal health sector in Tanzania has been the scandal of witch-doctors who use body parts from albinos in magic potions to bring people good luck or fortune. The demand for traditional medicine is still very strong in traditional societies in many other African countries and, this aspect is fostering the existence of inefficient and unqualified doctors, worsening the already precarious health conditions of the poorest.

The last chapter deals with one of the major concern in many developing countries: HIV/AIDS epidemic. The epidemic is especially acute in Lesotho, where roughly one quarter of the population is HIV/AIDS infected. This part of the thesis aims at sheding lights on who are those more at risk of contracting HIV infection. The idea came out during my staying in Lesotho, where I met many former miners who had been retrenched from South Africa mines after the Apartheid regime. The mines host a vibrant sex industry and once men are away from their families, they might be more likely to have multiple sexual partners. At the same time, women, who waits for their husbands to come back from the mines have been known to engage in sexual relationships with other partners as well. This has the potential to create a dangerous "network effect" in the transmission of HIV through multiple partners. Prevention efforts targeting more risky groups and new innovative approaches to induce safer sexual behavior have been desperately called for in Lesotho.

All these open issues should be considered a great challenge rather than discourage economists, institutions and NGOs working on development. Fundamental needs, such as health and housing could became a common practice everywhere. It is just necessary to be persistent.

Acknowledgements

My first acknowledgement must go to my advisor Eliana La Ferrara. Eliana provided me the inspiration and the encouragment to undertake this effort and she taught me to have a critical and original thinking towards development issues. A sincere thanks goes to Martina Bjorkman for injecting me enthusiasm for field works and to Guido Tabellini, always a source of precious advice. I am grateful to professor Edward Miguel for valuable supervision during my visiting at University of California, Berkeley.

I want to express my thanks to Michela Braga, who shared with me opinions, discussions, and troubles during these years. I am also grateful to all my friends, students, and colleagues who participated in the homeless project; a special thanks to Erica. Last but not least, I am indebted to Francesco for his encouragement when it was most required.

I remain solely responsible for all errors, omissions, and interpretations.

Introduction

The present doctoral dissertation investigates the link between health and social behaviors on one hand, and economic development on the other. The thesis is composed by three chapters:

-The first chapter investigates the role of social networks in predicting criminal activities among homeless people using new data I collected among homeless people in Milan;

-The second one is focused on the choice between formal and informal health care in Tanzania;

-The third one studies the determinants of HIV/AIDS in Lesotho, with a particular attention on the vulnerability of miners;

The first chapter analyzes the most extreme forms of poverty in industrialized countries: the homelessness. When I decided to study homelessness from an economic prospective, I realized that there were no data available on homeless people, neither in Italy nor in Europe and only few statistics were available in US. Therefore, with the help of more than 300 volunteers we counted and interviewed a representative sample of homeless people in the metropolitan area of Milan. Since typically homeless move from one street to another one in different nights, the probability to have double counts or double interviews of a single person is very high. To minimize this risk, we counted and interviewed homeless people in Milan in one single night. We found about 3860 homeless and the survey includes a total sample of 910 observations among those founded on the street, in shelters, and in slums in the periphery of Milan. The main aims of this chapter, "**Social network and crime: Evidence from homeless people in Italy**", is to investigate the link between social network and criminal activity. I study the effect of having a criminal friend on the likelihood to be in prison. The key difficulty in identifying network effects is separating the causal impact of the network from the role of unobservable characteristics shared by network members, especially as individuals potentially sort themselves in some network. To solve this endogeneity, economists have tried to instrument the group of people representing the social networks of an individual with some exogenous shocks not correlated with economics outcomes to study. The main novelty of the paper is to instrument bilateral links among individuals (instead of a group of individuals) to study the impact of social networks on criminal decisions with rainfall shock. More specifically, I use the fraction of rainy days in the overlapping period on the street (shelter) between two homeless, i and j. The basic idea behind the instrumental variable procedure is that a higher fraction of rainy days should positively affect social networks, by inducing a higher probability to meet more friends and to form more relationships. Since rainfall allocates randomly individuals across the streets, i and j have a higher probability to meet during their period as homeless because they live the same weather condition at the same time. The analysis points out some fascinating peer effects in the realm of criminality. As main results, I found that homeless who have bad friends are more likely to commit crime, but having more friends (non criminal) decreases the probability to commit crime.

This work has potentially relevant implications from a policy perspective. Given the role played by friends in decreasing criminality and by bad peers in increasing it, it would be inappropriate to not consider this effect when designing interventions for homeless people. If having a social networks on the street is negatively correlated with criminal activities, voluntary organizations should work to develop tools to stimulate social interactions and aggregation among homeless (i.e. meetings, free movies, dinners and so on). On the other hand, given the role played by bad friends, rehabilitation in prison and housing assistance would be an important intervention to avoid former inmates to end up living on the street. In general, this paper highlights the need for a new research agenda to create additional survey on homeless people.

The main contribution of the second chapter, "**Switching (or not) in health seeking behavior: evidence from rural Tanzania**", is to study individual's choice between formal and informal health care, as a function of the previous therapy. The idea of the paper is to understand whether the result of a therapy (sick or not sick) is a good incentive to switch to an alternative type of care. Understanding incentives/causes driving individuals towards more efficient health care is relevant to designing long term policy measures in the health systems. The key result is that agents do not switch from informal to formal health care, even when they remain sick after an informal treatment. The intuition under this result is straightforward: as long as the costs to test a new treatment are higher then the potential benefits coming from the new treatment, even a completely "rational" agent will choose not to experiment. This player gets stuck in a non-optimal equilibrium, simply because the cost of trying something else is higher compared with the potential benefits. In the paper, an agent has to compare the certain cost of formal health care (far away government-run clinics, long queues, etc.) with an uncertain benefit (the probability to heal).

The paper attempts to address three main questions. First, do agents switch to a formal caregiver, following only private information about their bad health status after the informal treatment? Second, what factors influence the demand for formal health care? Third, what are the main determinants of becoming ill?

Regarding the first question, I show that agents do not learn from their own past experience that it is better to switch to formal medical care after an ineffective informal therapy. Individuals choose once and for all at the beginning of the period and they do not update their believes, even if they have evidence of inefficient outcomes. The formal institutions I considered are hospitals, health centres, clinics and dispensaries and the informal ones are practitioners' homes, pharmacies, family homes and self-care. Ceteris paribus, to consult informal providers and to remain sick after the treatment decreases the probability to visit formal establishments in the next period by 11.5 percentage points.

A second result concerns the determinants involved in consulting formal or informal health care providers, conditional on reporting illnesses. The main determinant of the probability to seek formal care is the distance between the household and formal health establishments. As expected, more educated individuals tend to choose formal caregivers. Having access only to bad sources of drinking water, such as rain, lake and river water and living in an inadequate house, decreases the likelihood to seek formal medical care. Education, the quality of drinking water and living conditions also capture income and the ability to afford better services.

Finally, I show that disaster, such as drought, epidemic, insect and crop diseases, happened in a community six months prior to the survey increases the probability of illnesses. Years of education and bad living conditions are the main variables that determine one's possibility of becoming ill, affecting it respectively in a negative and in a positive way. Furthermore, women are more likely to report diseases compared to men.

The proposed analysis is tested using a 4 years household panel data from Kagera, Tanzania. The fact that in Tanzania, once individuals opted for a traditional healer never switch in the formal sector suggested that the prior belief about what is the best medical care dominates over the illness status of individuals. This feature of patient's demand fosters the existence of low quality and unqualified doctors and suggests implications from a policy perspective. The first step would be to allocate efforts and resource to enforce the demand for formal care, for example by disseminating informative campaigns to overcome cultural bias towards informal caregivers. Second, a more capillary distribution of government-run health services is required to increase the availability of formal health practitioners for the poorest.

The third paper, "**Getting a job but loosing health: HIV and miners in Lesotho**" (joint with D. de Walque), analysis the determinants of HIV/AIDS and related sexual behaviors. As main findings, we point out the most vulnerable groups to HIV infection in Lesotho: miner and current married women. Lesotho is the countries with the third highest HIV prevalence rate in the world, after Swaziland and Botswana. In 2005, the estimated adult HIV prevalence was 23.2 percent; nearly one over three Basotho adults aged 15-49 is currently infected by the virus. The paper explores whether historical past dynamics are correlated with the current spreading of HIV.

Lesotho is a small country located in the middle of South Africa and its economy is mainly dependent on this state. HIV was first detected in Lesotho in 1986, during the apartheid regime, when Basotho's men brought HIV home from South Africa, where roughly 60 percent of the workforce was employed in the mines. Working in South Africa mines means spending long period away from the household in Lesotho. Since then, the nation has experienced a dramatic escalation in the HIV/AIDS epidemic. Motivated by this anecdotal evidence the paper investigates whether the massive percentage of Lesotho labor force employed in South Africa for a long period in the past is the main cause of HIV/AIDS health emergency in Lesotho.

Since HIV can also effect the likelihood to work and to be a miner we found a variable that is correlated with being a miner in South Africa but not with being HIV infected. As proxy for being a miner, we use the distance between households in Lesotho and South Africa borders.

The results show that miners are more likely to be infected. Moreover, we found that education appears to have a protective effect: it is negatively associated with HIV infection and it strongly predicts preventive behaviors. The findings also show that married women who have extra-marital relationships are less likely to use a condom than non-married women. Our findings suggest policy recommendations to redirect prevention efforts towards more vulnerable groups in Lesotho, such as miners and women.

Chapter 1

Social Networks and Crime: Evidence from homeless people in Italy

Abstract[1]

The paper investigates the effect of peers and social networks size on criminal decision. I present a novel identification strategy to estimate endogenous social effects using a unique dataset on homeless people, I collected in Milan, in January 2008. By exploiting information on peers' names and surnames and by using a dyadic framework, the paper studies weather individual's criminal activities vary with the prevalence of criminal friends in his social networks. To proxy social networks, I exploit exogenous variation across dyads in rainfall shocks in the overlapping period on the street between a pair of homeless. Rainfall fosters homeless concentration in sheltered places (i.e. under bridges) and increase the probability of new meetings

[1] This project required the collaboration of many people. First of all I am deeply indebted with Michela Braga, my co-worker during all stages of the project. The data collection benefits from the help of Italian Red Cross's volunteers, Caritas and Opera Cardinal Ferrari. I wish to thank Stefano Dellavigna, Eliana La Ferrara and Edward Miguel for their invaluable comments and suggestions. I also benefit from the input by Martina Bjorkman, Enrico Moretti and Guido Tabellini. I thank ARPA and the Metherological Service of the Military Aereonautics for making rainfall data available to me. Financial support from the Empirical Research in Economics (ERE) and Fondazione Rodolfo De Benedetti (FRdB) is gratefully acknowledged. All errors are my own.

and friendship. Results show that the individual's probability of imprisonment increases with criminal peers, while decreases with network size.

> *"A closed door may feel safe, but an opened door may save your life. Let's open the doors!"*
> Ambrogio, Homeless in Milan

1.1 Introduction

Social interactions have been the objective of studying for their role in coping with economic shocks, especially among the poorest or in presence of market imperfections. Economists have been recently interested in how social interactions can foment bad behaviors, by describing, for example, network effects among criminals (Gleaser, Sacerdote and Sheinkman 1996, Case and Kats 2001, Kling and Kudwig 2007). This paper explores the implications of social interactions and peer effects on criminal behavior among poor people: could social interactions create a safety net for the poorest and decrease crime rates?

The motivations for this question go beyond the obvious policy implications for crime prevention. First, crime is a negative externality with social costs. If the existence of social networks reduces crime among the poorest throught its effect in coping with economic shock, forms of aggregation should be stimulate. The benefit of reducing crime might also have an indirect effect on poor people by increasing incentives to switch from criminal activities towards the formal labor market.

Second, the empirical evidence collected so far on the role of peer effects on crime decisions among the poor is very scarce. The scarcity of individual data public available on crime decision limited the empirical research on this topic. The data used for this paper come from a very innovative and detailed survey among homeless people, collected by the author in January 2008. The survey includes questions on imprisonment and names and surnames of the first five best friends of each homeless. I exploit this question by creating a dyadic framework with 600 links. The advantage of this framework is that we are able to precisely identify the social network of the respondents. Furthermore, the main difficult in estimating the effect of social interaction on criminal activity is that unobserved characteristics affecting social networks are likely to be correlated with unobservables influencing the decision to engage in crime. To solve this endogeneity, economists have tried to instrument the group of people representing social networks with some exogenous shock not correlated with economics outcomes (Munshi 2003). The main novelty of my paper is to instrument bilateral links among individuals (instead of a group of individuals) to study peer effects on criminal behavior.

Third, crime is over-represented among the poor and the reason behind this results is that the benefits from a crime outweigh the cost of potential punishment (Becker 1968). I investigates the role of social interactions among homeless people, the most extreme form of poverty in western economies. The role of the interactions among this group of individual can shed some light on long term policies to decrease homelessness on one hand, and crime on the other. The data used for the paper come from an innovative survey among homeless people, conducted by the author in one single night in January 2008 in Milan.

I start by analyzing the effect of network size, measured by the number of homeless friends, on the probability to go to prison during an homeless spell. To address endogeneity problem I instrument the size of the network with the fraction of rainy days in the individual's homeless spell. The idea is that homeless are more likely to be in sheltered places during rainy days (i.e. under bridges or porches) and they have a higher probability to meet more people and to create more links. Changes in rainfall strongly and positively predict the number of homeless friends. The well know reflection problem (Manski (1993)) has been addressed by including prior information on the structure of social network: the number of peers who have been in prison *before* homelessness. As first result, I show that social networks size allows to prevent criminal behavior. Homeless with more friends on the street have a greater probability to survive without committing crime. The likelihood to go to prison during an homeless spell decreases by 4.6 percent on average with an extra friend. By including exogenous peers characteristics, I also find that having at least one friend with criminal records before becoming homeless increases i's probability of incarceration by 13%, showing strong peer effects on criminal decisions.

I next test the effect of having a criminal friends on the likelihood of imprisonment by exploiting a dyadic framework. Similar to the previous identification, I instrument the probability of link between two homeless i and j with the fraction of rainy days in the overlapping period on the street between i and j. Instrumental variable estimates reveal a significant relationship between social interaction and incarceration: having a criminal friend increases the likelihood of imprisonment during homelessness, but having a good friend decreases it. Moreover, I find that social identity drives link formation: being of same nationality and same gender strongly increase the probability of links.

The remainder of the paper is organized as follows. Section 2 reviews the literature on social network and introduces some background on homelessness. Section 3 illustrates the research design for the data collection and section 4 describes data and provides some descriptive statistics. In section 5, I present the identification strategy and in section 6 the main results. Section 7 contains some concluding remarks and policy recommendations.

1.2 Background

1.2.1 Previous Works on social networks

The role of social networks in shaping economic outcomes has been widely documented. The empirical literature analyzes the relevant role of social interactions on risk sharing (Foster and Rosenzweig (2000), Fafchamps and Lund (2001), De Weerdt and Dercon (2006)), on credit transactions (La Ferrara (2003)), on welfare participation (Bertrand, Luttmer and Mullainathan (2000)), on educational choices (De Giorgi, Pellizzari and Redaelli (2007)) and its implication on labor market outcomes (Munshi (2003), Beaman (2008)). A part of this literature documented the role of social interactions on crime decisions. Case and Kats (1991), using data from the 1989 NBER survey of youth living in low income Boston neighbors, find that a 10 percent increase in the neighborhood juvenile crime rate increases the individual probability to become a delinquent by 2.3 percent. With an alternative approach, Gleaser, Sacerdote and Scheinkman (1996) provides an index of social interactions which suggests that the amount of social interactions is highest in petty crimes, moderate in more serious crimes and negligible in murder or rape. Kling and Ludwig (2007), by exploiting a source of exogenous variation in crime rate across neighborhoods in US, show little evidence of peer effects in crime. Recently, part of the literature focuses on peer effects on post released behavior. Drago and Galbiati (2008) study whether individual's post released behavior (i.e. recidivism) varies with the prevalence of recidivism of the peers who served the sentence in his same prison. They exploit the Italian Collective Clemency Bill to estimate recidivism of an inmate by using the residual sentence of his peers as a source of exogenous variation. The results shows that an increase in peers' average recidivism increases the individual probability of recidivism. Regarding the theoretical literature, Calvo-Armengol and Zenou (2003) provide a model in which delinquents compete with each other in criminal activities but may benefits from being friend with other criminal. They end up with different equilibrium outcomes in which identical individuals connected through social network can end up with being employed, criminal in networks or isolated criminals.

Despite this vast literature, identification of social interaction remain very problematic. The key difficulties in estimating peer effects are endogeneity - due to self selection in a group and common group effects - and "reflection" (Manski 1993) - endogenous effects (impact of the average peers' outcome on individual outcome), cannot be distinguished by exogenous effects (average characteristics of the group) because in a peer group everyone's behavior affects the others. Several authors have suggested that prior information on the structure of social network can help to resolve Manski's (1993) problem. For example, De Giorgi et al. (2007) instrument the mean outcome of i's group with the peers of peers who are not in one's own peer group.

This previous notion of social network, defined as a clearly delineated group of people who interact more or less frequently (e.g. family, kinship, village, student in the same class), is in contrast with the emerging literature on network formation for which a network is a bilateral links formed between agents. Theoretical works on network formation comes, among others, from Jackson and Wolinsky (1996), Bala and Goyal (2000) and Galeotti, Goyal, Jackson and Vega-Redondo (2008). This new approach for the formation of bilateral links has been translated in the empirical literature by using mainly as level of observation the dyad, or a pair of agents (Fafchamps and Gubert (2007), Conley and Udry (2008) De Weerdt and Fafchamps (2008) Fafchamps, Goyal and van der Leij (2008)). This paper deals with this last definition of network, applied among homeless people in Milan. A dyadic framework delineates in a more precise way one's social network (i.e. I can be part of the same class, but not be friends of my classmates), but, on the other hand, it is more difficult to find a valid instrument for bi-later links compared to a group of people. The main novelty of my paper is to instrument a link between two individual i and j by exploiting rainfall variation across dyads in the overlapping period on the street between a pair of agents.

1.2.2 Homelessness

The economic literature studying homelessness is really poor, mainly due to the lack of reliable data and comprehensive surveys on homeless people. The only institution that regularly carries out homeless counts is the U.S. Department of Housing and Urban Development (HUD). Since 1984, the HUD requires homeless counts every two years on a national sample of 80 geographically diverse communities. HUD's most recent estimates indicate 754,000 persons in US as living in emergency shelter, transitional housing, and on the streets on any given night, and a fairly stable homeless population at around 0.2 – 0.3 percent of the total population. California registers the highest number of homeless people, about 170,000, followed by New York, Florida, Texas and Georgia. HUD's assessment also measures homelessness on a longer time horizon showing that those who lives an homeless spell over the whole year are around the 1 percent of the total population. Although, the European Parliament (source) has underline the need of more effort towards regular homeless counts, European research on this phenomena is scarce and it is impossible to do cross country comparison for the lack of data. Besides the Milan homeless survey conducted by the author in January 2008, in Italy, the only attempt to count homeless people has been managed by the Survey Commission on social exclusion (Commissione di Indagine sull'esclusione sociale) joint with Zancan foundation in March 2000 on a sample of representative Italian municipalities. The count has been done by sampling shelters and soup kitchens and all places generally attended by homeless people. It indicates about 17.000 homeless people in Italy. At regional level, in Veneto, the University of Padua conducted a survey on a sample of about 80 homeless people to gather data on some socio-demographic characteristics.

Most of the empirical works analyze the causes of the substantial increase in the incidence of homelessness during 1980s in USA, by relying on homeless estimates provided by the HUD (O' Flaherty (1996)(check), Tucker (1989), Quigley (1990), Honing and Filer (1993), Quigley, Raphael and Smolenski (2001)) or by the US Census Bureau (Burt (1990)). A common result among these studies is that variations in homelessness arise from changed circumstances in the housing market and in the income distribution. Tuker (1989), shows that cities with rent control and lower vacancy rates have higher rate of homelessness. Subsequent papers have confirmed the robustness of these findings: Quigley et al. (2001) shows that one percentage point increase in the vacancy rate (from an average of 8.4%) combined with a decrease in average monthly rent-to-income ratios from 17.5% to 16.8% (median rent to median household income) could reduce the rate of homelessness by one-fourth. While previous works used intercity aggregate data, this is the first paper using micro-level data to study homeless from an economic perspective. Another strand of homeless literature investigates homeless duration based on hazard model and provide evidence that homeless population experiences temporary but recurrent spells of homelessness (Piliavin, Wright., Mare and Westerfelt (1994), Allgood, Moore and Warren (1997)). Homeless spells are longer for those with history both of drugs and alcohol abuse, while having received government benefits in the past decreases the average length of a homeless spell persons with a certain demographic characteristics and behavioral history (Allgood and Warren (2003)). Finally, a smaller strand of research studies homeless spatial distribution within a city, showing that they tend to settle in proximity to employment agencies (Marr (1997), Okamoto (2007)) or to commercial and service usage (Schor et al. 2003). Social networks among homeless people have not been the objective of extensive research. Different kind of mechanisms underline network formation among homeless people. For example, members of informal homeless communities share information about jobs, soup kitchens, shelters availability and rules and they benefit from peers who protect themselves against harassment from residents, police and other homeless. Iwata and Karato (2007) examine the effect of homeless networks on geographic concentration in Osaka City by using a spatial autoregressive model. They found that the number of homeless people in a census block is influenced by the number of homeless in neighboring census blocks, suggesting that networks determine homeless concentration. Different from the latter, the present paper relies on notion on network as a bilateral link between two or more individuals.

Social policies and institutions targeting homeless people change by town to town. In Milan there are mainly two type of social services. A permanent help centre financed by Milan municipality and located at the central station (Centro Aiuto Stazione Centrale) and the so called "cold emergency". The first one offers general information for all needy people on shelters availability, soup kitchen's location and it sorts homeless towards more specific voluntary organizations, depending on homeless's characteristics (i.e. immigrants, with or without the permit of stay, and so on.) The second is a temporary services and as the aim to increase shelters' host capacity during winter months. All other services are managed by private organizations with the goal to satisfy homeless basic needs, such as warm free meals, bed, clothes, blankets, medicines. Homeless have the opportunity to care their personal hygiene in public toilets, that generally costs 0,50 euros and they offer them showers with shampoo, bath foam, toothbrush, toothpaste and a towel. Sick homeless can seek medical cares in some health services in Milan, targeting specific groups of people: for example, immigrants, drugs or alcohol addicted.

Besides shelters, soup kitchen and health centres, additional "road's services" have the task to help people in emergency situation and especially during night, by providing them blankets and hot drinks. In Milan, there are mainly three organizations specialized in these type of services: the Red Cross, City Angels and S. Egidio Community. Some others organizations targeting homeless people adopted a different approach and they more oriented towards listening and comprehension of people excluded from the society (Caritas, SOS Exodus).

1.3 The data

1.3.1 Survey design

The data used in this paper come from a very innovative and representative survey of 910 homeless people, collected by the author on January 2008 in the city of Milan, Italy. The reference population includes people residing in (i) places not meant for human habitation, such as cars, parks, sidewalks, abandoned buildings (unsheltered homeless); (ii) emergency shelters (sheltered homeless); (iii) disused areas/shacks/slums.

The Milan Homeless Survey (MHS) includes two major phases: counting homeless people and conducting face to face interviews. The first step was to count all homeless people found in the streets, shelters and slums during the night of January 14^{th}, 2008. The count was essential in order to construct a homeless census from which randomly selecting the respondents. Interviews of unsheltered homeless were performed on the following night, January 15^{th}, while we surveyed people who were sleeping in shelters and in slums on January 16^{th} and 19^{th}, respectively. The whole data collection was completed in a single week to minimize sample attrition due to the high level of homeless' migration within the city. The count and the interviews were not conducted on the same day for two main reasons. First, it is not feasible to attempt to interview people during a one-night count. During the count, enumerators minutely checked the presence of homeless by walking along all streets in Milan and there would not be remaining time to also select and to interview them. Second, while it is optimal to conduct a late-night count (from around midnight to 3 a.m.) to maximize the probability to observe more visible people sleeping outside, the ideal time for interviews is around 9 p.m. when they are already settled down, but still awake and able to talk. Because it is easier to count people in shelters than on the street, conducting the count on a night when shelters are most full will likely lead to the most accurate count. In January shelters are at their peak capacity. Moreover, counting and interviewing people sleeping in open locations during the winter months may lead to a more realistic picture of chronically unsheltered homeless. Finally, in order to generate comparable numbers and to increase public awareness on homeless issues, we chose the same period of the year used by the US Department of Housing and Urban Development. To facilitate the identification of homeless people we picked a day of the week with less pedestrian traffic (Monday night).

The project involved about 300 volunteers in one single night, recruited among those who used to work with homeless people such as shelters workers, social workers in associations (i.e. Red Cross), students and private citizens. Enumerators were trained to produce an accurate count and a complete questionnaire, but also to avoid possible risks involved in approaching homeless during the nights (i.e. by always moving in team of more than two people).

On this respect, the paper add a methodological contribution in providing reliable estimates on the size of homeless people in Milan and in collective qualitative data. An accurate estimate of the unsheltered and sheltered homeless is essential for projections of service needs.

Homeless Count

Homeless people are both territorially mobile and likely to enter into and exit out of the homeless state depending on the time of day, season of the year, level of police harassment and other factors. The risk to count and to interview the same person twice is therefore very high among this population. A methodology to count homeless people by ensuring a minimal double counting is the so called *S - Night approach (Shelter and Street Night) or "Point in time survey"*,[2] meaning counting all homeless people (sheltered and unsheltered) contemporaneously in one reference night in the whole town. This approach also allows the enumerators to judge whether observed individuals fit the study's definition of homeless and it guarantees minimum variation in the criteria to identify them since they are counted by the same enumerators. The *S-Night* approach shows a snapshot of the homeless population, but if the count is repeated with regular intervals, it will give insights of the trend over time. Different methods have been

[2] *The S-Night approach has been proposed for the first time by the US Bureau of Census in 1990 in five US City (Chicago, Los Angeles, New Orleans, New York and Phoenix).*

proposed to count homeless population.[3] Among the more recent ones, is the capture-recapture method (CR)[4]. The CR method estimates the total homeless population from the sum of the population actually observed and an estimate of the unobserved population. The estimate of the unobserved is calculated with the number of people not caught in repeated observations (Cowan, Breakey and Fisher (1998), David and Snijkders (2002) Fisher, Turner, Pugh and Taylor (1994)). A limitation of the CR approach consists in estimating homeless during an entire year. Therefore, it assumes that all individuals identified as homeless remain homeless for the full year. Brent (2007) identifies street homeless through CR method by using a repeated daytime street observations to estimate the size of homeless people in Toronto. However, counting homeless based on a clear set of visual indicators undoubtedly misses those not engaged in public behaviors (e.g. people with shopping carts carrying a significant numbers of bags). Furthermore, the homeless population in Milan really care about their appearance: they are generally clean and well-dressed, thus "unrecognizable" during daytime.

“Point-in-time” counts could be criticized for missing homeless hidden from public view during late-night hours. We applied some efforts to overcome this criticism and to have the most reliable estimates. First, we divided Milan into 65 smaller areas, following the main roads, so that a team of 3-4 enumerators could reasonably cover it during the night of the count. Second, we distributed enlarged maps of each smaller section, defining the itinerary to be followed in detail with a complete list of all streets in each area. Third, we reported on each map the everyday places attended by homeless, identified in advance by consulting homeless services providers, so that enumerators could paid additional attention to the pointed out places. We assumed some criteria for the count: closed tends and closed paperboard were counted as one homeless, while in the abandoned cars/caravans enumerators tried to determine exactly how many homeless were sleeping there.

During this night, besides counting unsheltered homeless people, enumerators have two additional tasks. First, we asked them to describe homeless’ location as precise as possible, by reporting the name of the street, the closest civic number but also the sleeping place (i.e. Sarfatti

[3] To have a detailed descriptions of methodologies to count homeless people see Brent (2007).

[4] These methods were originally developed to estimate the size of a closed animal population and it is very common in the epidemological literaure. For further explanations on the capture recapture technique see Brent (2007).

street closed to number 25 on a bench in front of Bocconi University). They also recorded observable characteristics, such as ethinicity, sex and estimated age. Second, enumerators should inform homeless people about the next day interviews by leaving a flier next to their sleeping bag or paperboard. Reporting observable features of the individuals was useful to cross check this information with the ones collected through questionnaires. Volunteers joined these statistical activities with hot beverages and food distribution. At the same time, other teams of enumerators collected the lists of sheltered homeless people in each of the 19 emergency shelters in the city, with names, sex, age and nationality of each guest.

We met the enumerators about one hour before the count in five strategic headquarters in Milan, to distribute tools useful for the night (torches, food, beverages and block-notes) and to be sure all the teams would have started the count at the same time.

The procedure to count people living in slums was not straightforward. Slums in Milan are stable villages where people (mainly gypsy) are generally monitored by the municipal police. During the three months prior to the survey, we identified the typology of village (authorized/not authorized), the type of ethnic group and the number of people living in each area thought field visits. During these visits, we asked for the permission to interview people in the slums to the village's chief and we announced the date of the survey. In the reference night, enumerators checked dimension and location of pre-identified slums/disused areas.

The average length of the count was about 3 hours.

Homeless Interviews

Upon estimating the total size of homeless population, 75 enumerators interviewed a sample of them. We shirked the number of interviewers as much as possible to minimize answer bias and we recruited them for all the three interviews' nights to exploit the "learning by doing" effect.

Enumerators interviewed unsheltered homeless in the night following the count by going back to the locations reported. Two additional volunteers/assistants for each interviewer were distributing food and hot tea to unsheltered homeless people to make them more comfortable during the survey. Sleeping persons were not awakened for an interview, but when visible, enumerators reported some observable characteristics such as estimated age, race, and sex. The same team of volunteers were sent to different identified locations after finishing the first round of interviews to maximize the probability to recapture as many unsheltered homeless as possible. Information from street interviews may, in some cases be of questionable quality, since some homeless could be reluctant to participate, incoherent or fearful about giving out information. To deal with these caveats each enumerator filled out a module at the end of each questionnaire to evaluate interviews' quality and respondent's status, We dropped low quality questionnaire.

Sheltered homeless were randomly sampled from the population on the basis of shelter's dimension. We agreed with the responsible of each shelter the best time to run interviews. Among 25 shelters, four refused to participate and one had no guests at the time of the survey. Some interviews were directly conducted by shelter's managers.

Finally, slums were sampled through a stratified random sample method, based on geographic location, typology and dimension. We selected a total of 12 of the 56 slums. Within each selected area, we randomly extract respondents. Since the biggest slums had, on average, 100 people, we sent teams of 8-20 volunteers depending on slums' dimension. Volunteers also distributed napkins and kids' clothes to the households in the slums.

A potential drawback in doing the count and the interviews in two different days (even if very closed date) is the attrition rate, since people counted could have moved the day after. To control for the fact that homeless counted were the same of homeless interviewed, we include as first question in the survey "Did you sleep here last night?" and if no "Where did you sleep?". We cross checked this information with homeless locations recorded during the count. We did not consider the questionnaire if the respondent slept in places not identified in the count.

To avoid time-consuming interviews, enumerators distributed vouchers, that could be spent in restaurants, supermarkets, shops and pharmacies in Milan, to the respondents who fully completed the questionnaire. The questionnaire was translated into Romanian and English and the average length to fill it out was about 30 minutes each.

1.3.2 Rainfall data

This paper used rainfall data, as a proxy for homeless social networks. Rainfall data come from ARPA. (Regional Agency for the Environmental Protection) and from the Meteorological Department of the Military Aeronautics. Monthly rainfall data have been collected from 1960 to 2008. I use information on rainfall from six weather stations within Milan municipality, but some stations lack data for some months per year. I calculate the average number of rainy days (more than one millimeter of rainfall per day) among weather stations per months as a proxy of rainfall in Milan.

The rainfall shock variable used as instrument for social network has been constructed as the ratio between the number of rainy days in the overlapping period spent on the street/shelter between two individuals over the their total number of overlapping days. I exploit variation across pairs to estimate the social network effects on economic outcomes.

1.4 Descriptive statistics

The total homeless population in Milan is about 3860, split in 408 unsheltered homeless, 1152 sheltered homeless and about 2300 adults (older than 16 years old) in slums. Among unsheltered homeless we interviewed 34.6 percent, 12 percent refused to answer, 16.4 percent were already sleeping and 21 percent was not found during the night of the interviews. Due to time constraints, we did not send teams of enumerators for 16 percent of the homeless counted. In shelters, we sampled 525 homeless people over 1152 and we interviewed 420 respondents. While 6.7 percent of the sample were not in the shelter at the day of the interview, 11.3 percent of sheltered homeless were not interview for time constraints and about 2 percent refused to answer. In slums, we sample 525 adult homeless over 2300 and we collected 349 interviews, 33.5 percent of them were not found.[5] These percentages are reported in table A1 in the appendix.

[Insert figure 1]

Figure 1 shows the spatial distribution of sheltered and unsheltered homeless in the city. We found a high concentration of unsheltered homeless in the centre of the city, especially in proximity of the train stations (Cadorna Station and Central Station) and at Linate's airport where every night used to sleep about 15-20 people. In Milan there are 25 shelters and they are mainly located in the countryside.

[5] In slums, we did the interviews on Saturday afternoon, to capture also those who work during the week. We decided not to go there during the night for security reasons.

Unfortunately, information on names and surnames of respondents were not collected in slums, because homeless people where reluctant to reveal their own identity. Most are illegal immigrants and they were afraid to be identify by the police. Thus, the following statistics are restricted to unsheltered and sheltered homeless, ending up with a sample of 561 observations.

[Insert table 1]

How do homeless people survive on the street if they do not have a job? About 13 percent of the sample have no income source, while 21 percent did not know or did not answer the question regarding the first source of income. Among the remaining 66 percent, about 51 percent of Italian homeless receive government subsidies, while 9 percent get help from family and friends. This is in line with the economic literature focuses on family as source of insurance among its member especially against economic shocks. For example, Bentolila and Ichino (2006) study how countries with different family ties (namely Italy and Spain with strong family ties and US and UK with weaker ties) cope with unemployment shocks. They find that the consumption losses after the termination of a job are much lower in Mediterranean Europe, due to strong family ties. Most of the Italian homeless have lost all contacts with their original families and they do not want to get in touch with them anymore. As expected, among immigrants the percentage of those getting government subsidies is lower. To be eligible for a welfare checks from the municipality of Milan, individuals must be Italian and resident in Italy, while regular immigrants can benefit from pension, disability and unemployment insurance if they are eligible. About 24.2 percent of them receive help from friends/family and more than 60 percent declare wages as first source of income. The oldest mainly rely on government support. Only few people declare to gain from illegal activities and this percentage is higher among younger.

[Insert table 2]

However, as table 2 shows, the percentage of respondents who have been in prison is about 28 percent (39 percent of Italians and 23 percent of immigrants). Among those, 67 percent declare a period in prison after homelessness and roughly 28 percent before it, showing how extreme poor condition leaves people close to the edge of survival and when people have nothing left to lose, crime could become more frequent.[6]

[Insert table 3]

The identification strategy of the paper exploits the questions in the survey "Can you please tell me the name and surname (or alternatively, the first three letters of the surname) of the first five homeless friends on whom do you rely on in case of need?". Note that we asked about friends among the homeless people. Table 3 reports the distribution of friends including respondents from the street and from shelters. About the 36 percent of homeless people do not rely on any friends on the street and this percentage is higher for people who slept in shelter during the night of the count. Each individual has an average of 1.3 friend. The 20 percent have one friend and only the 5 percent report names and surnames of five friends. The 12 percent of the respondents did not answer and this percentage was higher among street homeless. To link friends' names with each questionnaire, we also asked the name and the surname of the interviewed. Names and surnames of sheltered homeless have been checked with administrative data provided by shelters' administration, while for unsheltered homeless names we consulted soup kitchens or social service centres registers. For example, to find missing surnames we crossed information on the name, age, nationality provided during the questionnaire with name, surname, age and nationality coming from administrative data.[7] The distribution of friends varies depending on incarceration. Table 3 also reports incarceration before and after homelessness, by number of friends. Among homeless with no friends, those who experienced imprisonment are 50 percent, compared to 32 percent without criminal records. Among respondents who provide the name of five friends, about 7 percent have never been to prison during homelessness, while only 0.94 have criminal records. The average number of friends is 0.87 for criminals compared with 1.47 friends among those with no criminal records after homelessness. These statistics provide some insight on the importance of having friends to decrease imprisonment. The table does not show substantial differences in the number of links on imprisonment before the homeless status.

[Insert table 4]

[6] There are no individuals who have been in jail both before and after their homelessness.

[7] In case the respondent cited only friend's name without surname, we associated him with the respective questionnaire only in the case the name does not appear twice in the list of names' respondents.

Table 4 shows the share of i's peers who have been in prison *before* homelessness. 70 percent of the homeless interviewed do not have friends who have been in prison, while 10 percent of the sample have at least one peer who have been in prison before his homeless spell. About 21 percent have friends who were not interviewed or who did not answer the question on imprisonment.

[Insert table 5]

The empirical analysis in this paper uses the dyad as level of observation, or a pair of agents ij. Dyadic regression analysis includes each pair of homeless twice: ij and ji pair. I construct a matrix value $L(i,j) = 1$ if there is a link between i and j, and $L(i,j) = 0$ indicates the absence of a link. In my sample the dyadic relationship could be directional, so that $L(i,j) = 1$ doesn't need to be equal to $L(j,i)$, meaning that i cited j as a friend, but j doesn't cited i. As described in table 6, in the sample, 657 links or network members are identified. Of these, 433 or 66% are links among members in the survey and 34% are links with homeless not covered by the survey. The percentage of links among homeless interviewed is higher because we did a census on the street. 62.4% of links are directional, while only the 37% are symmetric.

[Insert table 6]

Table 6 describes network using a dyad as level of observation. Following Fafchamps and Gubert (2006), I will focus on the link between sampled homeless, by dropping individuals who didn't answer or cited a name of a friend not in the sample and the observations in which friends had not been interviewed or didn't answer. The total number of dyadic observations is therefore N(N-1) minus those who dropped. The dependent variable used in the empirical analysis is equal to one if there is a link between two surveyed homeless people, and 0 if they are no friends. As expected, the percentage of links with surveyed homeless is very low and equal to 0.21%.

[Insert figure 3]

Social networks can be analyzed by graphical representations. Figure 3 depicts social networks among homeless based on their friendship. The graph is based on a spring embedding algorithm from UCINET. The idea is how a graphical representation can influence viewers' perceptions of the network structure. Each node represents an individual and the arrows are the links between them. Some forms of clustering emerged. First of all, it emerges a great proportion of men (blue) compared to women (pink). Second, the shapes of the objects depicting nationality are highly clustered. We see mainly three clustered groups: a high proportion of Italian homeless (circles), followed by Romanian (squared), Moroccans (box), and women from Russia, Moldova and Ukraine (up triangle). Italians tend to communicate and to create relationship among each other but they also have connections with Romanian, Morocco and Egyptians. Homeless from Morocco have also friends among other African countries but they do not communicate with Romanian. Eritrean and Ethiopian are the most isolated groups, followed by homeless from Southwest Asia (mainly China and India). The different colors of the line linking nodes describe whether the connection happens between or within shelter and street. The grey lines represent connection within the group, while the black ones describes links between the groups, showing a high level of clusterization among places where they usually sleep.

1.5 Empirical Strategy

1.5.1 Identification

The goal of the empirical section is to assess the effect of having criminal friends on one hand and of the number of friends on the other, on the individual's probability to be imprisoned. There is one main challenge in identifying network effects: endogeneity. Endogeneity arises because the propensity of an individual to behave in some way varies with the behavior of his peers' group. Therefore, a researcher who tries to infer whether the average behavior in the group influences the behavior of the individuals that comprise the group,.faces the well-known "reflection problem", pointed out by Manski (1993). In a peer group everyone's behavior affects the others, as in a mirror, and we cannot know if one's action is the cause or the effect of peers' actions. Manski (1993) and subsequent authors suggested that inference on endogenous effects is not possible unless the researcher has prior information specifying the composition of the reference groups. I begin the empirical section by testing the influence of peers and of network size on the likelihood of imprisonment. I adopt a structural framework providing social network identification presented in Bramoullè, Djebbary, Fortin (2007). I will next test peers effect by proposing a new identification strategy by exploiting a dyadic data structure.

1.5.2 Size of the social networks

Suppose we have a population of N individuals. For each individual i, let $g_i = \{j/g_{ij} = 1\}$, that is the set of individuals j to whom i is directly connected. This set g_i would be the individual's own peer group. Also, let N_i the total number of possible friends. I adopt the convention that for each $i, i \notin g_i$, so that $g_{ii} = 0$ for all i.

Let's formalize that problem using simple linear in means:

$$y_i^{After} = \alpha + \beta NS_i + \gamma E\left[y_j^{Before}|j \in g_i\right] + \delta E\left[x_j|j \in g_i\right] + zx_i + \epsilon_i \tag{1.1}$$

where y_i^{After} is a dummy variable equal to one if the individual i has been in prison *after* becoming homeless, NS_i is the social networks size going from a minimum of zero to a maximum of five friends, x_i is a vector of individual's traits and ε_i is the error term. $E\left[x_j|j \in g_i\right]$ contains the averages of the x in the peer group of the individual i, g_i.In a standard framework, $E\left[y_j|j \in g_i\right]$, the average of the dependent variable in the individual i peer group is included among the regressors to estimate peer effects and so γ would measure the endogenous effects and δ the exogenous effects. The main difficulties in estimating equation (1) are endogeneity in the estimate of β and reflection in the estimate of γ. To correctly identify β and γ authors seek instruments for $E\left[y_j|j \in g_i\right]$ and NS_i to the extent that unobservable shocks are correlated with group outcomes. Natural instrument for $E\left[y_j|j \in g_i\right]$ would be the average characteristic vectors of peers who are not in i's peer group (De Giorgi, Pellizzari, Redaelli (2007)).

My setting naturally offers an alternative solution to solve the reflection problem. Instead of estimate the endogenous effect of $E\left[y_j|j \in g_i\right]$, I will estimate the effect of predetermined characteristics of y_j : $\left[y_j^{Before}|j \in g_i\right]$ is the fraction of individual who have been in prison *before* becoming homeless in the peer group of individual i. Inmates who committed crimes before ending up on the street are those intrinsically criminal. They could affect the behavior of i once they become friends on the street, but their behavior is not influenced by i, since i and j_s became friends only during homelessness.

An additional source of endogeneity remains when we investigate the impact of social networks size. For example, a larger social network could depend on the fact that individuals meet up a many friends during the imprisonment. We have therefore to find something that is correlated with network size but not with the criminal activities. I apply an instrumental variable strategy to estimate the β coefficient by instrumenting network size with rainfall shock as follows:

$$NS_i = \alpha + \theta R_i + \gamma X_i + \varepsilon_i \tag{1.2}$$

where NS_i is the size of the network, represented by the number of friends (from 0 to 5), X_i is a variable capturing individual characteristics and ε_i is the error term. In this case we need a statistical instrument that determines changes in the size of the network but that is uncorrelated with the probability to commit a crime: R_i is the fraction of rainy days during i's period on the street. The idea is that homeless are more likely to be in sheltered places during rainy days, such as under bridges or porches, and they have a higher probability to meet more people and to create more links. Criminology literature argues a well established connection between weather and crime (Cohn 1996): at times of extremes in temperature and humidity the number of crimes drops, while in good weather the number rises. This would invalidate my instrumental variable strategy. I am fairly confident in my IV for two main reasons. First, as far as my knowledge, there is not a clear empirical evidence on the effect of rainfall on criminal rates. Second, in this special setting crime are generally committed to survive on the street: by stealing food, wallet and so on. Homeless people use to live on the street with whatever weather conditions and they act independently. The criminal activities is correlated with the satisfaction of their basic needs, reaching peaks when they do not have anything left.

1.5.3 Social networks in dyadic data

A second step in the analysis would be to identify the endogenous peer effects on the homeless's probability to go in prison by exploiting a dyadic data structure. In this framework the unit of observation is a dyad or a pair of individuals, ij. I begin by testing the determinants of network formation including among the regressors an exogenous shock that I will use, in the second stage equation, as an instrument for social networks. I estimate the following model:

$$L_{ij} = \alpha + \beta R_{ij} + \gamma_1(x_i - x_j) + \gamma_2(x_i + x_j) + \gamma_3 w_{ij} + \varepsilon_{ij} \tag{1.3}$$

where i and j are individuals, L_{ij} is an $N \times (N-1)$ matrix assuming value one if i reports j as a friend and 0 otherwise.[8] Individual attributes, such as age, education and duration on the street, are denoted x_i , x_j for the individual i and j respectively. Following the framework developed by Fafchamps and Gubert (2007) in studying dyadic equations, I include among the regressors the sum and the difference of x_i and x_j. This is necessary to preserve the symmetry requirement in estimating directional dyadic equations, for which the effect of (x_i, x_j) on L_{ij} must be the same as the effect of (x_j, x_i) on L_{ji}.[9] γ_1 measures the effect of differences in attributes on L_{ij}, while γ_2 captures the effect of the combined level of x_i and x_j on the dependent variable. To investigate the role of social distance in network formation I include w_{ij}, a dummy variable equal to one if i and j have the same nationality or the same gender. My solution to the endogeneity problem is to instrument for the link between two individuals. A valid instrument setting would determine the probability of a link, L_{ij}, while remaining uncorrelated with the economic outcome for which we would like to study the social network effects (i.e. the probability to go to prison or other determinants of crime). Rainfall shocks could be suitable for the purpose, because it allocates randomly homeless across streets. Specifically, my instrument, R_{ij} is the fraction of rainy days in the overlapping period on the street (shelter) between i and j. For example, let's say that homeless i slept for the first time on the street on November 2005, while homeless j arrived on the street on July 2007. The overlapping period on the street between i and j would be between July 2007 to January 2008, the data of the survey. In this period, I calculate the ratio between the total number of rainy days in Milan over the total number of overlapping days between an homeless pair. In an ideal setting, I would use data on the exit from homelessness spells to calculate the overlapping duration. Unfortunately, it was

[8] The total number of possible ij pairs is N^2, but I drop the N ij pairs on the diagonal.

[9] In case of unidirectional dyadic relationship, L_{ij} needs to be equal to L_{ji} and the symmetry requirement is satisfied by including the sum and the difference of the attributes in absolute values. For further details on the identification in dyadic estimation see Fafchamps and Gubert (2007).

impossible to collect data on the exit date for former homeless people. The basic idea behind the instrumental variable procedure is that a higher fraction of rainy days should positively affect social network formation, by inducing a higher probability to meet more friends and to form more relationships. Why two homeless people are more likely to become friends if they have a higher fraction of rainy days in their homeless spells? Since rainfall induces a higher concentration of homeless in sheltered places, i and j have a higher probability to meet during their homeless spell because they live the same weather condition at the same time. For instance, when it suddenly starts to rain, homeless people would stop to the closest sheltered place (i.e.: under a bridge, at the train station) and thus they would be more concentrated in indoors places. The probability to meet other homeless during rainy days is reasonable higher than during sunny days, in which homeless people used to stroll around in the street without a destination. Therefore, rainfall shocks constitute a good proxy for social networks. I exploit the variation across dyads to estimate the network effect in this paper.

The second point to deal with is the correlation of the residuals, ε_{ij}, among observations in dyadic data. This invalidates the classical assumption of no correlation among errors in the classical OLS framework. In dyadic equations $E\left[\varepsilon_{ij},\varepsilon_{ik}\right]\neq 0$, $E\left[\varepsilon_{ij},\varepsilon_{kj}\right]\neq 0$, $E\left[\varepsilon_{ij},\varepsilon_{jk}\right]\neq 0$ and $E\left[\varepsilon_{ij},\varepsilon_{kj}\right]\neq 0$. Following the previous literature (Conley (1999); Fafchamps and Gubert (2007), Comola (2008)), I apply the extension of White robust covariance matrix developed to deal with spatial correlation of errors. The formula for network corrected covariance matrix is:

$$AVar(\overset{\wedge}{\beta})=\frac{1}{N-K}(X\prime X)^{-1}\left(\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{l=1}^{N}\frac{m_{ijkl}}{2N}X_{ij}u_{ij}u_{kl}^{'}X_{kl}\right)(X\prime X)^{-1} \tag{1.4}$$

where $\overset{\wedge}{\beta}$ denotes the vector of coefficients, N is the number of dyadic observations, K is the number of regressors, X is the matrix of all regressors, X_{ij} is the vector of regressors for dyadic observation ij and $m_{ijkl}=1$ if $i=k, j=l$, $i=l$ or $j=k$ and 0 otherwise. The only structure imposed on the covariance structure is that $E\left[u_{ij},u_{ik}\right]\neq 0$, $E\left[u_{ij},u_{kj}\right]\neq 0$, $E\left[u_{ij},u_{jk}\right]\neq 0$ and $E\left[u_{ij},u_{kj}\right]\neq 0$ for all k but that $E\left[u_{ij},u_{km}\right]=0$. Formula (4) was initially developed for linear regressions but it applies to maximum likelihood estimation provided that x_{ij} is everywhere replaced by the corresponding score l_{ij}. The division of the inner term by two corrects for the double counting of the dyads ij and ji.

Having established the correlation between rainfall shock and friendship, I estimate and discuss the validity of using it as instrument in the 2SLS estimates as follows::

$$P_i^{After} = \alpha + \beta_1 L_{ij} + \beta_2 (L_{ij} * P_j^{Before}) + \beta_3 P_j^{Before} + \tag{1.5}$$

$$\gamma_1 (x_{it} - x_{jt}) + \gamma_2 (x_{it} + x_{jt}) + \gamma_3 w_{ijt} + u_{ijt}$$

where P_i^{After} is equal to one if the individual i has been in prison *after* becoming homeless and 0 otherwise, P_j^{Before} is equal to one if the individual j has been in prison *before* becoming homeless and L_{ij} is equal to one if i and j are friends. The interaction term $L_{ij} * P_j^{Before}$ measures the effect of peers on one's probability of incarceration. In other words, β_2 captures the effect of having friends intrinsically criminals. As we previously discussed we instrument the probability of links and the interaction term with rainfall shocks, to deal with the endogeneity problem. The reflection problem in this equation has been addressed by including pre-existing characteristics of the peers in the i's group: the probability that the individual j has been in prison *before* his homeless spell, and so before i and j would have met. The likelihood that the individual j has criminal records before ending up in the street is a proxy for the type of i's friends. Incarceration before homelessness would also capture the probability that a bad guys have more bad friends or a greater probability to go back to prison.

1.5.4 Potential drawbacks

There are essentially two data problems that we must deal with in this paper. First, a possible caveat in the estimation strategy is that rainfall is not allocated randomly across individuals. i and j could not be friends, not only because it had not rained a lot during their overlapping period, but also because homeless i could have relatives who might host him during rainy nights. So homeless i and homeless j did not meet because one of the two was not on the street during rainfall time. In general omission of individuals from the sample will bias the estimated rainfall effect on network downward, since we are attributing all the rainfall effect to a fraction of links. By assuming that none would spend nights by their relatives during rainy days, there will be more links due to rainfall shock and the impact of rainfall on the probability to have a link would increase. In the empirical analysis I will restrict the sample to only those who never move back and forth over time to obtain a relatively clean estimates of the network effect.

Second, crime is not directly observed. Even with a further data collection on incarceration from the Ministry of Justice we cannot measure crime directly. Following Moretti (2003), we assume that arrest is an increasing function of the amount of crime committed.

1.6 Results

In the paper social networks is seen to influence crime behavior in two ways. The first one is the effect of peers on the probability to have criminal records. The second one is the impact of the size and the quality of social networks on individual's crime behavior. I begin this section by investigating the statistical distribution of the instrument used in more detail. Then I will show the main findings of the impact of network size on crime using individual level data and the peer effects in criminal decisions by exploiting a dyadic framework.

1.6.1 Rainfall

In this paragraph I describe the distribution of rainfall in Milan. Figure 3 (a, b, c, d) depicts the number of rainy days per months from January 1960, when the "oldest" homeless arrived on the street to January 2008, the month of the survey. I split the time-series in year subgroups to better visualize rainfall trends. The graphs show that the most rainy months are, on average, October and November. Thus, homeless who, by chance, slept for the first time on the street in these rainy months would have a higher percentage of rainy days compared with those who arrived on the street, for example in July. We assume that the date of arrival on the street is random. In the empirical section, I adjust my estimates for seasonality, by including the number of Novembers since a pair has been on the street.

Second, the distribution of rainfall for "older pairs", those who arrived on the street at maximum 20 years and at minimum 5 months prior the survey could tend to the general distribution of rainfall in Milan and the variation in my instrument could be explained only by the "younger pairs", those who spent at maximum five months on street (for example, from August 2007 to January 2008). Figure 4 shows the distribution of the fraction of rainy days considering all pairs in the sample and by splitting the sample between older and younger pairs. As expected, the graph describing rainfall distribution for the younger pairs shows slightly higher mean and variance compared to rainfall distribution of the all sample. Looking at the histogram we note an higher fraction of rainy days for younger couples. However, the trend of rainfall distribution for the older and younger couples are very similar to the total distribution in the sample.

1.6.2 Social Networks size and prison

How peers and network size affect the individual's probability to go to prison? This section provides linear estimates of the likelihood that an homeless would go to prison during his homeless spell.

[Insert table 7]

In the first four columns of table 7 I report the reduced form equation: the fraction of rainy days in the i's homeless spell is clearly uncorrelated with the probability to go to prison after becoming homeless (column (2) and (1)). Table A2 in the appendix reports the means test on the fraction of rainy days in the period on the street for those who have been in prison after their homeless spell. The t-test do not reject the null hypothesis of the equality of mean in the fraction of rainy days for criminal and not criminal, both before and after homelessness. Note that the probability of imprisonment is positively and significantly correlated with the time spent on the street, while duration squared is negatively correlated with the dependent variable, suggesting that homeless have a higher probability to commit crime in the middle of their homeless spell. Women are less likely to go to prison. Columns (3) and (4) report the reduced form regressions by splitting the rainfall variable in four sub-categories depending on the percentage of rainy days on the total number of days on the street. For example, the first variable (the omitted variable) is equal to one if the individual was in the rain for less that 0.17 percent of his days on the street.

Columns (5)-(10) show the first stage estimates. The fraction of rainy days during i's period on the street strongly predicts network size: rainfall increases the probability to have more friends. In term of magnitude: a one standard deviation increase in the fraction of rainy days for homeless i increases the probability to have an additional friend by 22.6 percent (2.55*0.089). Rainfall represents a good instrument for the size of social networks: the rainfall coefficient is statistically significant at 1 percent and the F- statistics is above the critical value at the 5 percent level. We should thus be fairly confident about the consistency of the IV estimates. While in columns (5)-(6) I report the first stage by including the share of i's peers who have been in prison before homelessness, columns (7) and (8) consider as additional regressor a variable equal to one if at least one peer in i's group went in prison before homelessness. Columns (9) and (10) report first stage equation by splitting the rainfall variable: homeless with more than 30 percent of rainy days in their spell are more likely to have more friends, and this remains true also if one controls for the length or time spent on the street.

[Insert table 8]

Table 8 presents OLS and two least squared estimates of the impact of social network size and peers on the probability of incarceration after becoming homeless. The probability of imprisonment decreases with the network size and this effect is stable across specifications, suggesting that homeless with more friends have greater chances to survive on the street without committing crime. Having more friends means having a greater safety net on the street at which homeless can rely on in case of need. An alternative explanation could be that having more friends reduces the probability of arrest conditional of crime. This would be the case if, for example, homeless with larger social networks benefit by having more friends because they can protect him against imprisonment. While this reasoning could be true in some contexts (the Italian Mafia is a well-known example), it seems irrelevant among homeless people: they do not have credential to be trusted by the police and no incentive to be involved in a trial as witness to protect a homeless friend.

Columns (1) and (2) report OLS results. The likelihood to go to prison during homelessness decreases by 4.6% on average with an extra friend. Note that among the regressors I include the share of i's friends who have been in prison before homelessness. These estimates clearly indicate the presence of significant peer effects in the probability of committing criminal activities. In term of magnitude: having at least one friend with criminal records before becoming homeless increases i's probability of incarceration by 13%. The estimates show the effect of having bad friends on criminal behavior. These results hold also by including controls. Women are less likely to go to prison and duration on the street is positively correlated with the probability of incarceration: the longer these people are on the street, the more exposure they have to criminal activity. Theft of food and clothing, (what we might call "survival crimes") escalates into more serious crime. Duration squared is a decreasing function suggesting that homeless in the middle of their status have a higher probability to commit crime. OLS estimates give an inconsistent estimate of β given omitted unobserved characteristics of homeless i. Columns (3)-(8) report IV estimate. Once we instrument the number of friends with rainfall shocks, the network size effect is even greater: an extra friend is estimated to lead to a 11 percent points increase in the probability of going to prison. Looking at peer effects, the estimates indicate that one additional average peer spending a period in prison before his homelessness increase the probability that homeless i will go to prison by approximately 22.3 percentage points. In columns (5) and (6) we include an alternative definition of peer, including a dummy equal to one if at least on peer in i's group has been in prison before his homeless spell. In this specification the estimates indicate that having at least one criminal in the peer groups increase the dependent variable by 28 percentage points.

1.6.3 Peer effects and prison

I next show the results of peer effects on the likelihood of going to prison by exploiting a dyadic data structure.

[Insert table 9]

Table 9 reports linear first stage estimates on the probability of link between i and j. The fraction of rainy days in the overlapping period between i and j positively affects the probability of friendship also including additional controls (column 2). In term of magnitude, with a standard deviation of 0.121 for the rainfall variable, it turns out that a one standard deviation increase in rainfall fraction rises the probability of friendship by 0.082% (0.0068*0.121). This is a non-negligible effect: given that the mean level of having a link in the sample is 0.002, this constitutes an increase by 41%.(0.0008/0.002=0.41). Social identity drives link formation: being of same nationality and same gender strongly increase the probability of links. It is interesting to note that the difference in age between i and j is negative and significant, showing that younger respondents are more likely to report links with older individuals, capturing inter-generational altruism

In columns (3)-(10) the robustness of this result is tested against competing explanations.

From the estimates reported in columns (1) and (2) we could argue that these results are driven by the fact that more rainfall bring people to shelter and, in shelters, the probability to make friends is higher. So this would constitute a "shelter effect" rather than a "rainfall effect". Therefore, columns (3) and (4) report estimates by splitting the sample by place of interview. Rainfall remains positive and statistically significant also considering sheltered and unsheltered homeless separately. [10] A partial counter-intuitive results is the negative sign in the sum of i and j duration in column (3): a possible explanation is that people who have spent a long period of time on the street/shelter have lost their hope and are not interested anymore in building up relationship with other people. A possible caveat in the estimation strategy implemented so far could be that rainfall is not allocated randomly across individuals. i and j could not be friends, not only because it did not rain a lot during their overlapping period, but also because homeless i could have relatives who host him during rainy nights. So homeless i and homeless j did not meet because one of the two was not on the street during rainfall time. As additional control, in column (5) I restricted the sample to chronic homeless people, those who never went out from the street since they arrived. By restricted the sample, the size of the coefficient increases and it is still statistically significant at 1 percent level. Another interesting question is to verify if once homeless are randomly allocated across street they are more prone to match with similar individuals. To answer to this question I include among regressors the interaction between all the covariates and the fraction of rainy days between i and j (column (6)). The table reports only the coefficients statistically significant: once they are allocated randomly, homeless are more likely to match with homeless of the same gender.

The quantity of rain for each pair depends on the month of the year they arrived on the street. For example, those who end up on the street before November, they will be on the street

[10] The program used to correct dyadic standard errors requires a squared matrix among respondent and friends. To estimate the effect of rainfall shocks on social networks in restricted samples of unsheltered and sheltered homeless I drop friends who slept in shelter for those interviewed on the street and friends who slept on the street for those interviewed in shelter (Fafchamps and Gubert 2007).

in the most rainy months. I adjust the estimates for seasonality by including the number of Novembers each pair have lived on the street. Column (7) reports seasonally adjusted estimates and it shows the consistency in sign and significance of the β coefficient.

Finally, in the last column of table 10, I checked for an alternative measure of rainfall shock by splitting the rain fraction in three dummies variables capturing the level of rainfall, between 10 and 20, 20 and 30 and more than 30 and leaving as omitted category if it rained less that 10. We note a non linear effect of rainfall and the only significant coefficient is when it rained more that 30% of the overlapping days.

[Insert table 10]

I now proceed to directly verify the relationship between social network and incarceration. The dependent variable is equal to one if i registered criminal records *after* becoming homeless and 0 otherwise. Starting with a preliminary OLS, column 1 of table 10 shows how social network affects the probability of imprisonment: the likelihood of criminal records for the individual i negatively depends on the linkage between i and j, showing that having a friend on the street reduces the probability incarceration. Peer effects are capturing by the interaction between a dummy equal to one if i and j are friends and a binary variable assuming value one if j was a previous inmate. The results show strong peer effects in the realm of criminality: having a criminal friend increases the likelihood of imprisonment. To account for endogeneity, in column 2, I estimate a two stage least-squares model (2SLS) by doing two consecutive OLS regressions: a first stage OLS regression of L_{ij} on R_{ij} to get $\overset{\wedge}{L_{ij}}$ with dyadic robust standard errors followed by a second stage OLS of P_i on $\overset{\wedge}{L_{ij}}$ and on $\overset{\wedge}{L_{ij}} * P_j$. Having a friend reduces the probability of arrest, but having a "bad" friend increase it. In term of magnitude: one standard deviation increases in the having an extra criminal friends rises the likelihood to go to prison by 3% (3.079*0.03sd).[11] Columns (3) and (4) reports binary probit estimates. The sign of the coefficients are stable across specifications. However, the magnitude of the coefficients is very different across specifications. Further investigation on the more appropriate model to use in a dyadic data structure is required.

[11] The big size of the coefficient links to$Link_{ij}$ is due to the small percentage of those with a link (0.21%). I am currently working to obtain a more realist estimation.

1.7 Conclusions

This paper makes two contributions. First, it proposes and discusses the first data collection among homeless people in Europe. The economic research on homelessness is very scarce because of the lack of extensive surveys on this population. In January 2008, during one single night, we counted and collected data on a representative sample of homeless people in Milan, Italy. We found about 3896 homeless including those sleeping on the street, in shelters, in slums and we interviewed about 24% of them. The dataset will be available to the academic community to conduct further research on homelessness. By knowing the number and the socio-demographic characteristics of homeless, policy makers can target the right amount of resources and design specific policies to speed up the reintegration of homeless within a community and to increase welfare benefit programs targeting those who need them most.

The second contribution is an attempt to solve endogeneity in studying the effect of having a criminal friend on the likelihood to be in prison. The main challenge in identifying network effects is separating the causal impact of the network from the role of unobservable characteristics shared by network members, especially as individuals potentially sort themselves in some network. The proposed identification is divided in two parts. In the first one, the paper exploits the share of i's peers who have been in prison *before* homelessness to estimate the likelihood of individual's imprisonment *after* homelessness. Inmates *before* homelessness are those intrinsically criminals and can affect the criminal behavior of their friends. To estimate the impact of social networks size on the likelihood of imprisonment, we instrumented the number of friends with the rainfall shock during i's homeless spell. I find that the likelihood to go to prison during an homeless spell decreases by 4.6 percent on average with an extra friend. By including exogenous peers characteristics, I also find that having at least one friend with criminal records before becoming homeless increases i's probability of incarceration by 13%, showing strong peer effects on criminal decisions.

Second, the main novelty of the paper is to estimate peer effects on criminal activities by exploiting a dyadic data structure. I instrument the probability to have a friend by using exogenous variation across dyads in rainfall in the overlapping period on the street among a couple of individuals. The idea is that during rainy days homeless look for a sheltered place to sleep and the probability to meet other homeless people and to find new friends is higher during rainy days. The results show that those who have bad friends are more likely to commit crime, but having more friends (non criminal) decrease the probability to commit crime.

These results have potentially relevant implications from a policy perspective. Given the role played by friends in decreasing criminality and by bad peers in increasing it, it would be inappropriate to not consider this role when designing interventions for homeless and in the realm of criminality. If having a social networks on the street is negatively correlated with criminal activities (and so the cost for the society and increase security) , voluntary organizations should work to develop tools to stimulate social interactions and aggregation among homeless people (i.e. meetings, free movies, dinners and so on). On the other hand, given the role played by bad friends, rehabilitation in prison and housing assistance would be an important intervention to avoid former inmates to end up in the street.

In general, this paper highlights the need for a new research agenda to create additional survey among homeless and to better understand how social interactions affect economic outcomes. This aspect is not only relevant in Western economies. Homelessness is an increasing phenomena worldwide. In China, Thailand, Indonesia and the Philippines, despite their growing prosperity, homelessness is rampant and it is mainly due to migrant workers who have trouble in finding permanent homes. In developing countries such as India, Nigeria, and South Africa, millions of children living and working on the streets.

1.8 References

Allgood S., Moore M. L., Warren R. S., "The duration of sheltered homelessness in a small city", *Journal of Housing economics* 6, 60-80, 1997.

Allgood S., Warren R. S., "The duration of homelessness from a national survey", *Journal of Housing economics* 12, pp. 273-290, 2003.

Angelucci M., De Giorgi Giacomo, Rangel M., Rasul I., "Family Networks and school enrolment: Evidence from a Randomized Social Experiments", April 2007.

Bala V., Goyal S., "A Noncooperative Model of Network Formation", *Econometrica*, Vol. 68, No. 5, pp. 1181-1229, 2000.

Bandiera O., Rasul I., "Social Networks and Technology Adoption in Northern Mozambique", *Economic Journal* 116: 862-902, 2006.

Barret B., Santos P, "What do we learn about social networks when we only saple individuals? Not much", 2008.

Berry B., "A Repeated Observation Approach for Estimating the Street Homeless Population", *Evaluation Review*, 31; 166, 2007.

Beaman L., "Social Networks and the Dynamics of Labor Market Outcomes: Evidence from refugees", working paper, 2008.

Bertrand M., Luttmer E. F. P., Mullainathan S., "Network Effects and Welfare Cultures", *Quarterly Journal of Economics* 115(3), pp. 1019-55, 2000.

Braga M., Corno L. "Being an homeless in Milan: a descriptive analysis", 2009.

Bramoullé Y., Djebbari H., Fortin B., "Identification of peer effects through social networks", 2007.

Burt M., "Over the edge: The growth of homelessness in the 1980s", *Urban Institute Press*, Washington DC, 1992.Calvo-Armengol A., Zenou Y., Social Network and Crime Decisions: The role of Social Structure in facilitating delinquent behavior, 2003.

Cameron C., Golotvina N., "Estimation of Coutry Pair Data models controlling for clustered errors: with international trade applications", *working paper N. 06-13*, Department of Economic US Davis, 2005.

Case A., Katz L. F., The Company You Keep: The Effects of Family and Neighborhood on Disadvantaged Youths, NBER Working Paper No. W3705, May 1991.

Cohn, E. G., "The effect of weather and temporal variations on calls for police services", American Journal of Police, Vol. XV, No. 1 1996.

Conley T., "GMM Estimation with Cross Sectional Dependence" *Journal of Econometrics* Vol. 92 Issue 1, 1-45, 1999.

Conley T., Udry C. "Learning About a New Technology: Pineapple in Ghana", September 2008 Revision.

Conley T., Udry C., "Social Learning Through Networks: The Adoption of New Agricultural Technologies in Ghana", *American Journal of Agricultural Economics*, Vol. 83 Issue 3, 668-673, 2001.

Cowan C.D., Breakey W. R., Fischer P.J, "The methodology of counting the homeless: A Review", in *Homelessness, health and human needs* 169-82, Washington DC, Institute of Medicine, National Academy Press, 1988.

David B, Snijders T.A.B, "Estimating the size of the homeless population in Budapest", *Quality and Quantity* 36:291-303, 2002.

De Giorgi G., Pellizzari M., Redaelli S., "Be as careful of the books you read as of the company you keep: Evidence on Peer Effects in Educational Choices", *IZA working paper* No. 2833, 2007.

De Weerdt J., "Risk Sharing and endogenous network formation", 2002.

De Weerdt J, Dercon S., "Risk Sharing networks and insurance against illness", *Journal of Development Economics* 81, 337-356 , 2006.

De Weerdt J., Fafchamps M., "Social Identity and the Formation of Health Insurance Network", March 2008.

Drago F., Galbiati R., "Network effects in Prison: Evidence from the Italian Collective Clemency Bill", July 2008.

Early, D, "The role of subsidized housing in reducing homelessness: an empirical investigation using micro-data", *Jjournal of policy Analysis and Management,* VOl.17, No 4, pp.687-696, 1998.

Early, D., Olsen, E. O., "Subsidized Housing, emergency shelters, and homelessness; and empirical investigation using data from 1990 Census", *Jjournal of Economic Analysis & Policy,* Vol 2, Issue 1, 2002.

Early, D., "An empirical investigation of the determinants of street homelessness", *Journal of housing economics* 14, pp.27-47, 2005.

Fafchamps M., Gubert F., "The Formation of Risk sharing networks", *Journal of Development Economics,* 83(2):326-50, 2006.

Fafchamps M., Lund S., "Risk Sharing Networks in Rural Philippines", *Journal of Development Economics,* 71:261-87*,* 2003.

Fafchmaps M., Goyal S., Van der Leij M., "Matching and Network Effects", working paper, 2008.

Farrell S. J., Reissing E. D., "Picking up the challenge: Developing a methodology to enumerate and assess the needs of street homeless population*", Evaluation Review* 28;104, 2004.

Fisher N., Turner S.W., Pugh R., Taylor C., "Estimated numbers of homeless and homeless mentally ill people in north east Westminster by using capture-recapture analysis"*, British Medical Journal*, 308:27-30, 1994.

Foster A., Rosenzweig M., "Imperfect commitment, Altruism and the Family: Evidence from Transfer behavior in low income rural areas", *Review of Economics and Statistics*, 83(3):389-407, 2001.

Freeman R. B., Hall B. "Permanent homelessness in America*"*, *NBER working paper series* No. 2013, 1986.

Galeotti A., Goyal S., Kamphorst J., "Network formation with heterogeneous players", *Games and Economic Behavior*, 54, 2, 353-372, 2006.

Galeotti A.,Goyal S., Jackson M., Vega-Redondo F., Yariv L., "Network Games", (Revised and Resubmitted to The Review of Economic Studies), 2008.

Glied S., Hoven C. Moore R., Garrett B. A., "Medicaid and service use among homeless adults", *NBER Working paper* 5834, November 1996.

Glaeser E., Sacerdonte B., Scheinkman J., "Crime and Social Interactions", *Quarterly Journal of Economics*, Vol. 111, No 2, May 1996.

Hojman D., Szeidl A. "Core and Periphey in Networks", 2007.

Honig M., Filer R. K., "Causes of Intercity variation in Homelessness", *The American Economic Review*, Vol.83, No. 1., pp.248-255, 1993.

Iwata S., Karato K., "Homeless networks: Testing peer and homed networks against location choice", *CIRJE Discussion Paper F-522*, 2007.

Iwata S., Karato K., "Homeless networks and Geographic Concentration:Evidence from Osaka City", *CIRJE Discussion Paper F-527*, 2007.

Jackson M. O., Wolinsky A. "A Strategic model of social and economic networks", *Journal of Economic Theory* 71, 44-74, 1996.

Jecks, C., "The Homeless". *Cambridge University press, Harvard University press* 1994.

La Ferrara E., "Kin Groups and Reciprocity: A model of Credit Transactions in Ghana", *American Economic Review*, 93(5), 1730-1751, 2003.

Lochner L., Moretti E., "The effect of Education on Crime: Evidence from prison inmates, arrest and self-reports", *American Economic Review* 94(1), 2004.

Ludwig J., Kling J.R., "Is crime contagious?", *Journal of Law and Economics*, 50:3, 2007.

Manski, C. "Identification of Endogenous Social Effects: The Reflection problem", *Review of Economic Studies*, 60, 531-542, 1993.

Martin E., "Assessment of S-Night Street Enumeration in the 1990 Census", US Bureau of the Census, Washington DC, 1992.

Munshi K., "Networks in the modern economy: Mexican migrants in the US labor market". *Quarterly Journal of Economics* 118, 549-599, 2003.

Piliavin I., Wright, E. R. B., Mare, D. R., Westerfelt H.,A., "The dynamic of homelessness", *Institute for research on poverty, discussion paper,* No.1035, 1994.

Quigley J., "Does Rent control cause Homelessness? Taking the claim seriously", *Journal of policy analysis and Management*, pp.89-93, 1990.

Quingley J. M., Raphael S., Smoloensky E., "Homeless in America, Homeless in California", *Review of Economics and Statistics,* 83(1), 37-51, 2001.

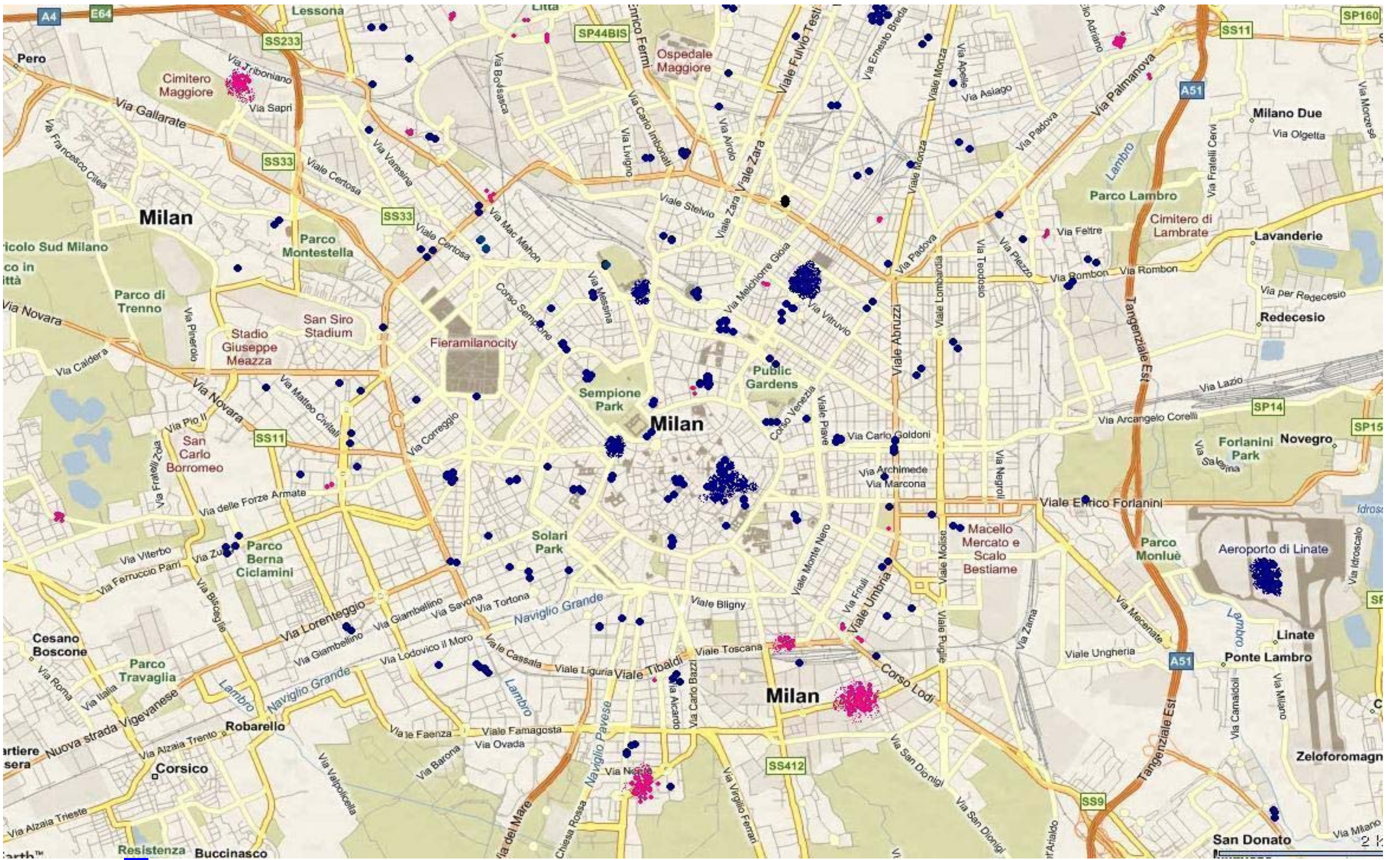
San Francisco Human Services Agency, "San Francisco 2007 Homeless Count Final Report", March 2007.

Tucker W., "America's Homeless: victims of rent control", *Backgrounder N. 685, The Heritage Foundation*, Washington DC, 1989.

White H., "A Heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity", *Econometrica*, Vol. 48, N. 4, 1980

Figures

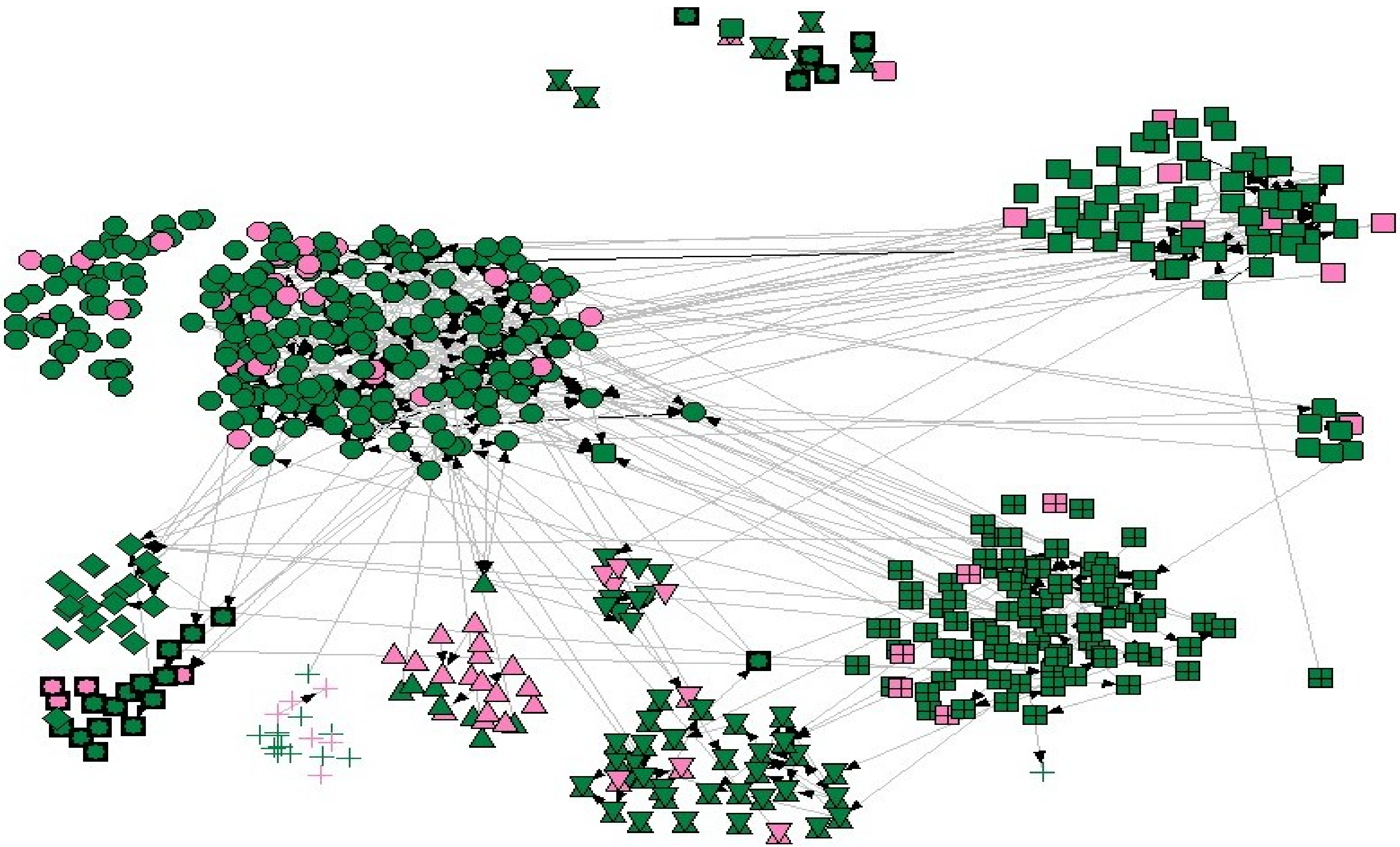
Figure 1: Spatial distribution of unsheltered homeless and localization of shelters in Milan



Legend: [blue] =Localization of unsheltered homeless, each dot=1 homeless
[magenta] =Localization of shelters, each dot=10 homeless

Source: MHS 2008

Figure 2: Social interactions among homeless



Legend:
Woman = []   Man=[]
Italians=circle, Romanian=square, Russia/Moldavia/Ukraine/Bulgaria=>up triangle, Morocco/Algeria/Tunisia=box, South America=>down triangle, Egypt=Diamond, Eritrea/Ethiopia=plus, Niger/Mali/Chad/Sudan=rounded squared, Senegal/Gambia/Ivory Coast/Togo=>thing, Asia=>circle in box

Figure 3a: Number of rainy days per month (1969-1970)

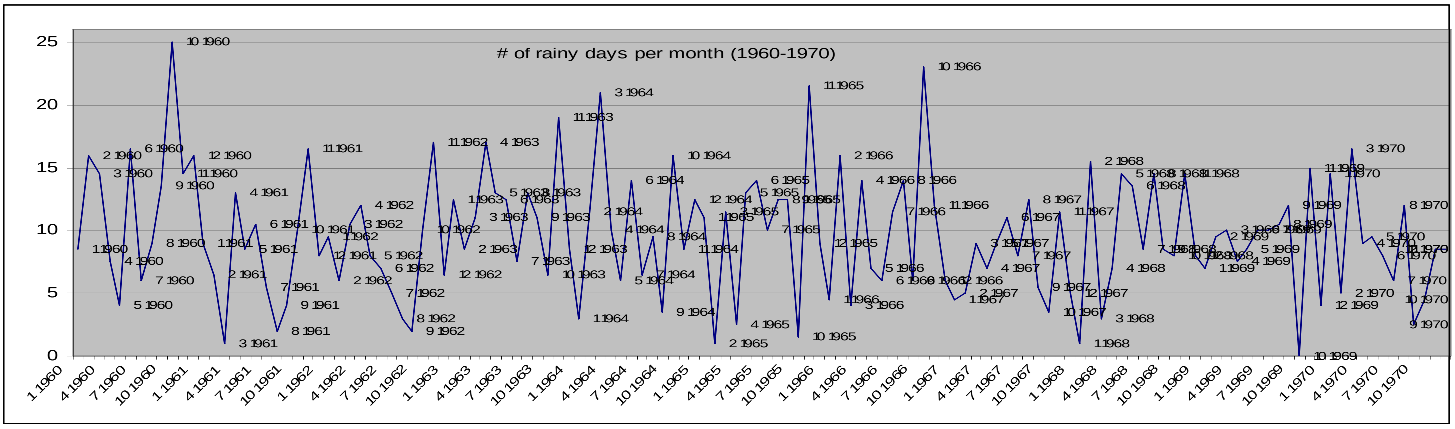


Figure 3b: Number of rainy days per month (1971-1980)

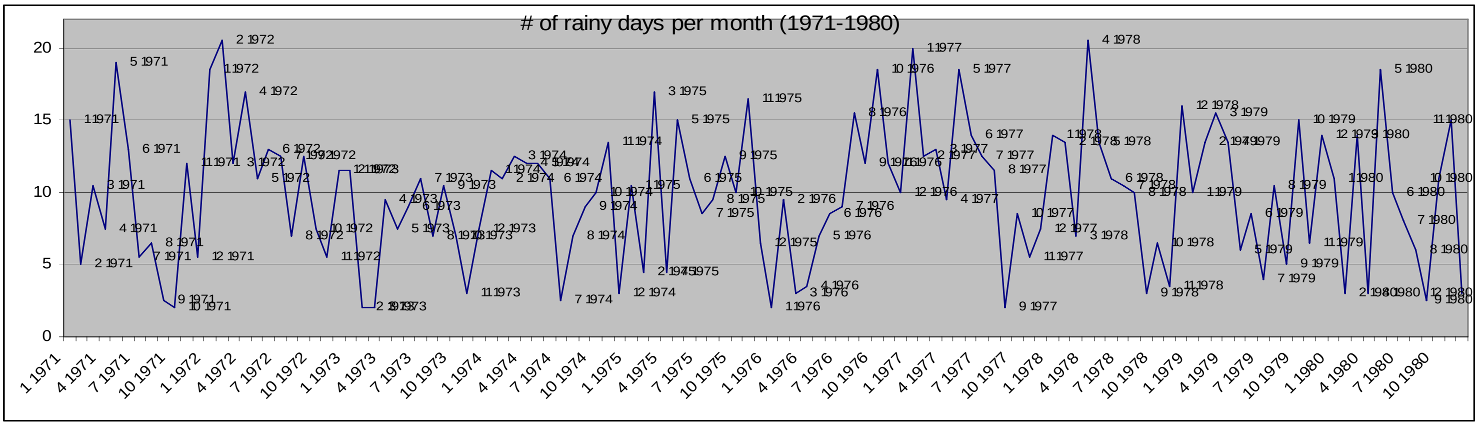


Figure 3c: Number of rainy days per month (1981-1995)

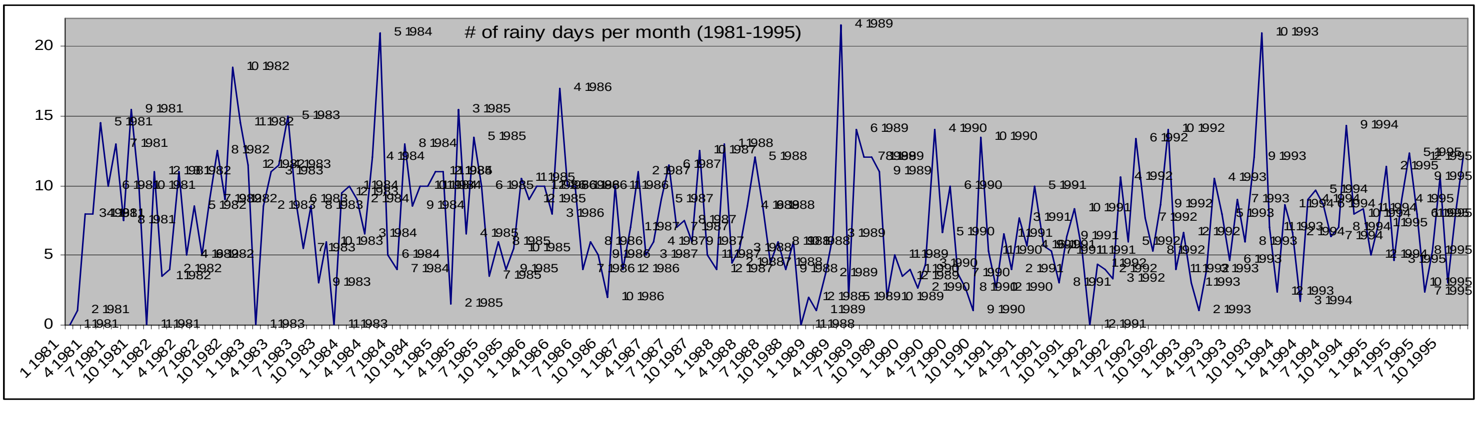


Figure 3d: Number of rainy days per month (1969.1970)

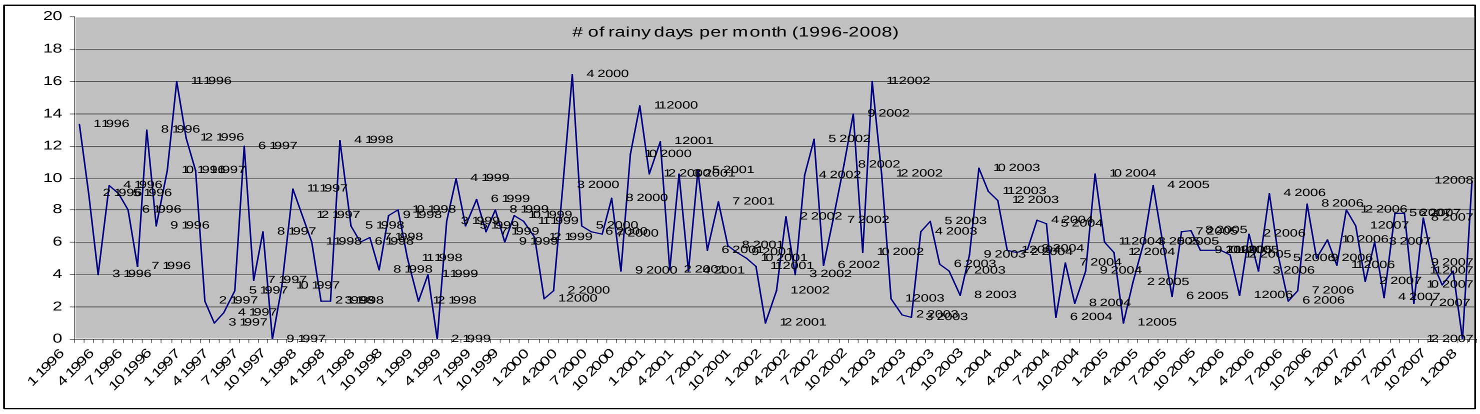
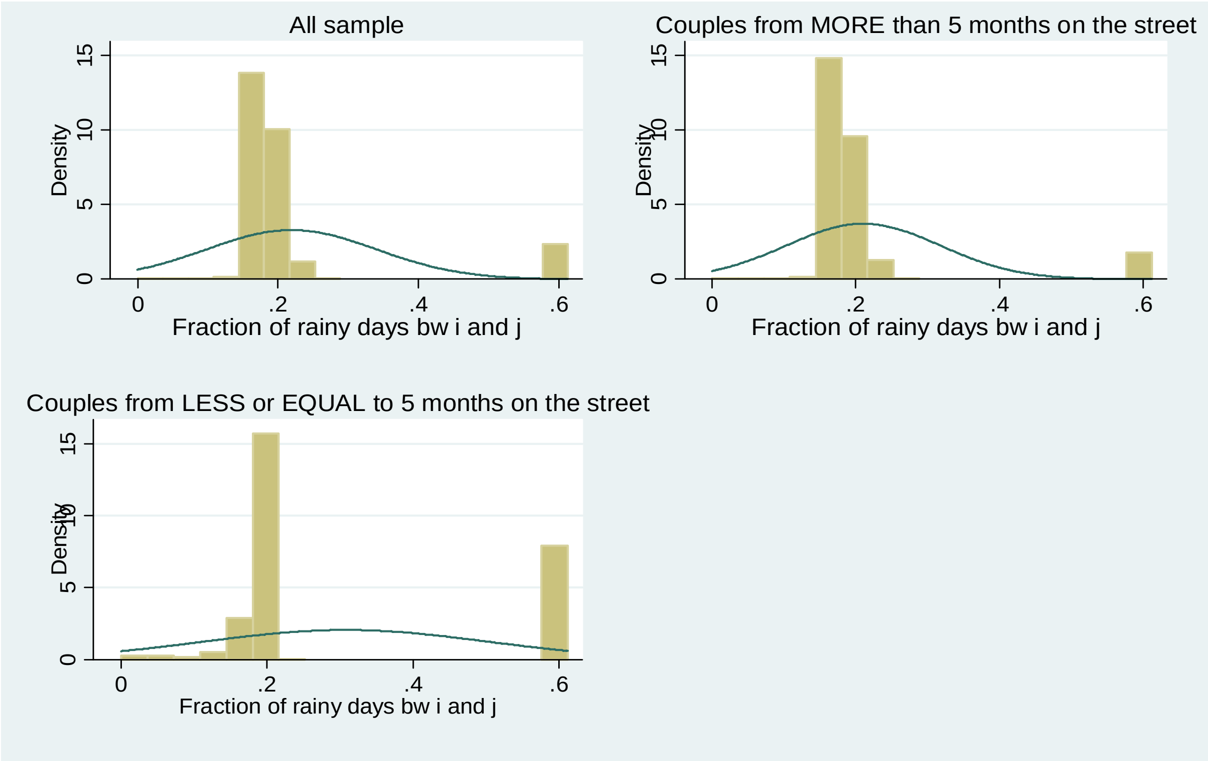


Figure 4: Rainfall distributions



Tables

Table 1: Source of income, among those who declared an income

Sources of income	*Italians*	*Immigrants*	*age>59*	*age<60*
Gov. subsidies	51.63	6.45	73.17	16.32
Family/Friends	9.24	24.19	4.88	20.14
Job (Perman/occas)/savings	34.78	63.44	17.07	58.33
Illegal activities	1.63	1.61	-	2.08
Shelter subs./church	2.72	4.3	4.88	3.13
Obs.	184	186	82	288

Note: [1] We include welfare checks, disability/unemployment insurance, pensions. To be eligible for welfare checks it is required to be Italian and to have residence in Italy. Regular immigrants can benefit from disability/unemployment insurance and pension if they are eligible.

Table 2: Homelessness and prison

	Obs.	*Percent*
Prison at least once	156	27.81
Never in prison	391	69.7
Don't answer	14	2.5
Total	561	100
Prison after homelessness	156	67.95
Prison before homelessness	156	32.05

Table 3: Distribution of friends, by place of interview and incarceration

Distribution of friends	*All Sample*	*Street %*	*Shelter*	*After Homelessness*		*Before Homelessness*	
				% not in prison	*% in prison*	*% not in prison*	*% in prison*
0 links	36.01	28.37	38.57	32.2	50	35.61	36
1 links	20.86	19.86	21.19	22.68	16.04	21.33	22
2 links	13.37	16.31	12.38	14.29	11.32	14.08	10
3 links	5.35	6.38	5	5.67	4.72	5.43	6
4 links	6.24	7.09	5.95	6.8	4.72	6.24	8
5 links	5.70	5.67	5.71	7.03	0.94	5.84	6
Don't know/Don't answer	12.48	16.31	11.19	11.34	12.26	11.47	12
Mean	1.34	1.53	1.28	1.47	0.87	1.35	1.39
Observations	561	141	420	441	106	497	50

Table 4: Share of *i*'s peers who have been in prison before homelessness

Share of i's peers who in prison before homelessness	*Obs.*	*Percent*
0	395	70.41
0.25	5	0.89
0.333	8	1.43
0.5	14	2.5
0.666	2	0.36
1	16	2.85
Peers not in the survey/don't answer	121	21.57
Total	561	100
At least one peer in prison before homelessness	45	10.23

Table 5: Number of links among homeless

	Total links	*Links with homeless in the survey*		*Links with homeless not in the survey*	
	No.	*No.*	*%*	*No.*	*%*
Total links	657	433	65.9	224	34.1
Not symmetric links	--	270	62.4	--	--
Symmetric links		163	37.6	--	--

Table 6: Network using dyadic data

Original sample	*Freq.*	*%*
No links	245967	78.29
Links with homeless interv.	433	0.14
Not answer	28560	9.09
Links with homeless not interv.	39200	12.48
Total	314160	100
Final sample		
No links	192,763	99.79
Links with homeless interv.	397	0.21
Total	193,160	100

Table 7: Reduced form and First stage regressions: rainfall and social network size

	Reduced Form				First Stage					
Dependent Var.:	*Prison after becoming homeless*				*Network Size=# of friends*					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Fraction of Rainy days	-0.0759	-0.0745			2.55***	2.43**	2.349***	2.25**		
	[0.1441]	[0.1609]			[0.84]	[0.83]	[0.81]	[0.83]		
Rain from 0.17-0.20			-0.0312	-0.0737					0.259	0.193
			[0.0574]	[0.0631]					[0.264]	[0.275]
Rainfall from 0.20-0.30			0.1040	-0.0338					0.102	-0.127
			[0.0641]	[0.0537]					[0.276]	[0.330]
Rainfall more than 0.30			-0.0931	-0.0939					1.186**	1.119**
			[0.0780]	[0.0991]					[0.424]	[0.430]
Share of *i*'s peers who have been in prison before homelessness					1.68***	1.6***				
					[0.34]	[0.35]				
At least one peer in prison before homelessness							1.63***	1.56***	1.55***	1.55***
							[0.23]	[0.24]	[0.24]	[0.24]
Duration on the street		0.0019***		0.0018***		-0.0007		-0.0007		-0.0013
		[0.0006]		[0.0007]		[0.0019]		[0.0019]		[0.0024]
Duration sq.[1]		-0.0024**		-0.003**		-0.0016		-0.0016		-0.0008
		[0.0015]		[0.0014]		[0.005]		[0.005]		[0.0054]
Female		- 0.0954***		-0.0964***		-0.3139		-0.3139		-0.282
		[0.0306]		[0.0290]		[0.2023]		[0.2023]		[0.22]
Observations	535	525	535	525	422	416	422	416	422	416
R-squared	0.0001	0.07	0.03	0.07	0.07	0.09	0.12	0.13	0.124	0.14
F-test	0.28	6.99	4.55	9.40	6.52	4.2	6.45	5	2.2	1.7

Notes: Linear estimates. * denotes significance at 10 percent level, ** at 5 percent level, *** at 1 percent level. Robust standard errors in parenthesis, adjusted for clustering at the place of the interview level. Columns (1)-(4): reduced form regressions, Columns (5) - (10): first stage regressions. Columns (2), (4), (6) include additional controls: age, age sq., education, place of the interview, nationality.
[1] coefficient and se multiplied by 1000.

Table 8: Second stage regressions: social network size and crime

	Dep. Var.: 1 if has been in prison after homelessness					
	OLS		*IV*			
	(1)	(2)	(3)	(4)	(5)	(6)
Network size	-0.0460***	-0.0432***	-0.1154***	-0.1251***	-0.1304***	-0.1403***
	[0.0098]	[0.0098]	[0.0414]	[0.0449]	[0.0459]	[0.0488]
At least one peer in prison before homelessness	0.1314***	0.1514***			0.2805***	0.3076***
	[0.0562]	[0.0562]			[0.0893]	[0.0852]
Share of *i*'s peers who have been in prison before homelessness			0.2231*	0.2613**		0.1956*
			[0.1161]	[0.1156]		[0.1027]
Age		0.0040		0.0027		0.0021
		[0.0072]		[0.0076]		[0.0076]
Age sq.		-0.0000		-0.0000		0.0000
		[0.0001]		[0.0001]		[0.0001]
Duration		0.0017***		0.0016***		0.0016***
		[0.0005]		[0.0005]		[0.0005]
Duration sq.		-0.0000**		-0.0000**		-0.0000**
		[0.0000]		[0.0000]		[0.0000]
Female		-0.0997*		-0.1292**		-0.1344**
		[0.0547]		[0.0586]		[0.0631]
Years of education		0.0023		0.0038		0.0041
		[0.0058]		[0.0069]		[0.0070]
Place of interview		0.0569		0.0227		0.0288
		[0.0389]		[0.0435]		[0.0445]
Nationality		0.0040		0.0021		0.0024
		[0.0035]		[0.0033]		[0.0035]
Constant	0.244***	-0.125***	0.3296***	0.0960	0.3348***	0.0919
	[0.23]	[0.195]	[0.0553]	[0.1530]	[0.0594]	[0.1520]
Observations	433	416	422	416	422	416
Shea Partial R-squared	--	--	0.0217	0.0196	0.0194	0.0177
First stage F-test	--	--	6.52	4.2	6.45	4.71
R-squared	0.03	0.10	--	--	--	--

Notes: * denotes significance at 10 percent level, ** at 5 percent level, *** at 1 percent level.

Table reports linear estimates; robust standard errors in parenthesis, adjusted for clustering at the place of the interview level.

Table 9: Probability of link - First stage

Dependent variable=1 if link between i and j	*All Sample*		*Shelter*	*Street*	*Chronic homeless*	*All Sample*		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Fraction of rainy days$_{ij}$	0.0087***	0.0068***	0.0059***	0.0830***	0.0057***	0.0068***	0.0070***	
	[0.0005]	[0.0012]	[0.0016]	[0.0208]	[0.0012]	[0.0012]	[0.0012]	
Rain from 0.10-0.20								0.0002
								[0.0006]
Rainfall from 0.20-0.30[1]								0.0003
								[0.0006]
Rainfall more than 0.30								0.0034***
								[0.0008]
Same gender$_{ij}$		0.0015***	0.0022***	0.0042**	0.0011***	0.0014***	0.0015***	0.0016***
		[0.0002]	[0.0002]	[0.0020]	[0.0002]	[0.0002]	[0.0002]	[0.0002]
Same nationality$_{ij}$		0.0047***	0.0063***	0.0153***	0.0041***	0.0047***	0.0046***	0.0046***
		[0.0004]	[0.0006]	[0.0024]	[0.0005]	[0.0004]	[0.0004]	[0.0004]
Sum of i and j's age		-0.013***	-0.0000***	-0.0002***	-0.015***	-0.0000***	-0.0000***	-0.0000**
		[0.0029]	[0.0000]	[0.0000]	[0.0005]	[0.0000]	[0.0000]	[0.0000]
Diff in in i and j's age		-0.010***	-0.0000**	-0.0001*	-0.0009**	-0.0000***	-0.0000***	-0.0000***
		[0.0029]	[0.0000]	[0.0000]	[0.0004]	[0.0000]	[0.0000]	[0.0000]
Sum in i and j years of educ.[1]		0.025	0.008	0.2	0.036**	0.006	0.0000	0.0000**
		[0.010]	[0.026]	[0.2]	[0.0018]	[0.01]	[0.0000]	[0.0000]
Diff in i and j years of educ.[1]		-0.002	0.01	-0.0002	-0.0002	-0.0036	-0.0000	-0.0000
		[0.010]	[0.023]	[0.0002]	[0.0002]	[0.017]	[0.0000]	[0.0000]
Sum in i and j's duration[1]		-0.001	-0.0025*	-0.0038	-0.0044***	-0.00008	-0.0000**	-0.0000
		[0.002]	[0.0014]	[0.007]	[0.0008]	[0.00006]	[0.0000]	[0.0000]
Diff in i and j's duration[1]		0.002	0.0011	0.007	0.002	-0.00008	0.0000	0.0000
		[0.001]	[0.0012]	[0.007]	[0.001]	[0.00006]	[0.0000]	[0.0000]
Same gender*rain[1]						0.0054*		
						[0.0031]		
# of Novembers							-0.0919*	
							[0.0480]	
Place id	yes	yes	no	no	yes	yes	yes	yes
Observations	193160	193160	11990	109230	122850	193160	193160	193160

Dyadic robust standard errors in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%

Note: [1] coefficient and se multiplied by 1000. Linear Estimates.

Table 10: Social Networks and Arrests: Second Stage

Dependent Variable=1 if i has been in prison after homelessness	OLS	2SLS	BINARY PROBIT	
			Coeff.	*Marg.*
	(1)	(2)	(3)	(4)
Link bw i and j	-0.1021***	-29.3555***	-2.492***	-0.00006
	[0.0213]	[5.7157]	(0.120)	[0.0017]
Prison before j*Link ij	0.1279**	3.0794**	2.616***	0.950
	[0.0516]	[1.2760]	(0.191)	[.00022]
Prison before $homelessness_j$	-0.0032**	-0.0123***	-0.024***	-0.00086
	[0.0016]	[0.0045]	(0.008)	[.0946]
Same Gender	0.0380*	0.0843***	0.146	0.00015
	[0.0224]	[0.0250]	(0.094)	[.7538]
Same Nationality	0.0113	0.1450***	0.040	0.001
	[0.0195]	[0.0343]	(0.071)	[.2532]
Sum in i and j age	0.0012	0.0004	0.005	0.00003
	[0.0007]	[0.0006]	(0.003)	[107.51]
Diff in i and j age	0.0000	0.0000	0.005	-0.00003
	[0.0000]	[0.0000]	(0.009)	[19.58]
Sum in i and j years of educ.	0.0008	0.0016	0.003	0.0015
	[0.0023]	[0.0022]	(0.009)	[32.80]
Diff in i and j years of educ.	0.0002	0.0003	0.003**	-0.0014
	[0.0023]	[0.0022]	(0.001)	[14.762]
Sum in i and j duration[1]	0.0003**	0.0002**	0.001***	-0.00004
	[0.0001]	[0.0001]	(0.000)	[150.55]
Diff in i and j duration[1]	0.0004***	0.0005***	0.0001***	0.00003
	[0.0001]	[0.0001]	[0.0000]	[23.365]
# of Novembers since homelessness	0.0047*	0.0072***	0.007	.00014
	[0.0026]	[0.0027]	(0.019)	[.0369]
Place id	yes	yes	yes	--
First Stage F-test	-	30.04	30	
Observations	172640	172640	164906	164906

Notes: * denotes significance at 10 percent level, ** at 5 percent level, *** at 1 percent level.

Table reports linear estimates; robust standard errors in parenthesis, adjusted for clustering at individual level. Restricted the sample, by dropping friends who didn't answer.

Appendix

Table A1: Descriptive statistics on the homeless count

	Street	*Shelter*	*Slums*
Homeless found	408	1152	2300
Sampled	408	500	525
Interviewed	34.6	84	66.5
Refused	12	2	0
Not found	21	6.7	33.5
Not interviewed for time constraints	0	7.3	0
Sleeping	16.4	0	0
Not sent a team	16	0	0
Total obs	141	420	349

Source: MHS 2008

Table A2: Means test on the fraction of rainy days for those criminals and not criminals

	Obs.	Not Prison	Prison	P-value	t-stat
			After homelessness		
Fraction of rainy days for individual *i*	535	0.2088	0.2048	0.685	0.405
			Before homelessness		
Fraction of rainy days for individual *i*	535	0.209	0.1966	0.345	0.945

Table A3: Descriptive statistics

Dyadic regressors					
Variable	*Obs.*	*Mean*	*Std. Dev.*	*Min*	*Max*
Link $_{ij}$	193160	0.002	0.045	0	1
Rainy days in the overlapping period$_{ij}$	175228	0.220	0.121	0	0.61
Prison	190087	0.296	0.456	0	1
Prison ^{After} homelessness$_{i}$	190087	0.196	0.397	0	1
Prison ^{Before} homelessness$_{j}$	190087	0.099	0.299	0	1
Prison$^{Before}{}_{j}$*Link$_{ij}$	190087	0.000	0.016	0	1
Same Nationality	193160	0.269	0.443	0	1
Same Gender	193160	0.757	0.429	0	1
Same Street	193160	0.117	0.321	0	1
Sum of Age	189660	91.573	18.605	40	158
Diff in Age	189660	0.000	18.648	-62	62
Diff in Duration	182756	0.00	132.55	-576.47	576.47
Sum in Duration	182756	126.67	132.24	0.10	1117.00
Sum in education	188790	18.041	5.784	0	40
Diff in education	188790	0.000	5.797	-20	20
Not Dyadic variables					
Fraction of rainy days$_{i}$	544	0.21	0.09	0.16	0.61
Network Size = # of friends	491	1.34	1.54	0	5
Prison	547	0.29	0.45	0	1
Prison ^{After} homelessness$_{i}$	547	0.193	0.395	0	1
Prison ^{Before} homelessness$_{i}$	547	0.091	0.288	0	1
Share of i's peer in prison before homelessness	491	0.285	0.202	0	1
At least 1 peer in prison before homelessness	491	0.0916	0.29	0	1
Female	561	0.10	0.30	1	2
Age	522	44.68	13.35	19	82
Years of educ.	522	9.12	4.08	0	20
Duration in months	544	61.30	93.26	0.03	577

Source: MHS 2008

APPENDIX A4: HOMELESS - STREET SURVEY

Good evening, my name is **______________** and I am volunteering with a Milan University. We are carrying out interviews in order to understand cause and consequences of homelessness in Milan. We could use **your help** to better understand the needs of the city's homeless and find ways to improve social support services within the community. Any answers or information you provide will remain totally **confidential and not be shown to other persons**. During the interview, you do not have to answer any of the questions you are very uncomfortable with. To thank you for your time in answering the questionnaire, you will receive a voucher to spend in many Milan supermarkets, restaurants, pharmacies or bars. At this time, do you want to ask me anything about the survey? Do you agree to do the interview? Would you like a warm drink as we do the survey? Can I sit here? Would you like a copy of the survey?

INTERVIEW START TIME: I___I___I___I___I

SECTION 1: GENERAL INFORMATION

Q1	Did you sleep **here** last night?

Yes	1	=>Q3
No	2	
Don't know (DO NOT READ)	9	
No Answer (DO NOT READ)	99	

Q2	Where did you sleep?

Street (STREET NAME)	1 (______________________________)
Shelter (FACILITY NAME/STREET)	2 (______________________________)
Other (Specify)	3 (______________________________)
Don't know (DO NOT READ)	9
No Answer (DO NOT READ)	99

NOTES

Q3	Do you usually sleep in the same place every night?

Yes	1
No	2
Don’t know (DO NOT READ)	9
No Answer (DO NOT READ)	99

NOTES

Q4	Gender (DO NOT ASK, ENTER THE ANSWER DIRECTLY)

Male	1
Female	2

Q5	Can you tell your Year of Birth ?

Year of birth	I__I__ I__I__I
Don’t know (DO NOT READ)	9
No Answer (DO NOT READ)	99

NOTES

Q6	What is your **first nationality**? (country of origin)

Italian	1	
Romanian	2	=>Q8
Russian/Moldavian/Ukrainian	3	=>Q8
Moroccan/Algerian	4	=>Q8
South American	5	=>Q8
Other (Specify)	6 (______________________)	=>Q8
Don’t know (DO NOT READ)	9	
No Answer (DO NOT READ)	99	

Q7	What **Province of Italy** were you born in? (ENTER THE NAME OF THE BIRTH PROVINCE)

Province of		=> Q9
Don’t know (DO NOT READ)	9	=> Q9
No Answer (DO NOT READ)	99	=> Q9

NOTES

Q8	What **Year** did you come to Italy for the **first time**?

MONTH (*in letters*)	YEAR
I__I__I__I__I__I__I__I__I__I	I___I___I___I___I
Don’t know (DO NOT READ)	9
Don’t remember (DO NOT READ)	99

NOTES

Q9	Are you religious/practicing? If Yes what faith do you practice?

No	0
Catholic	1
Protestant	2
Christian Orthodox	3
Jewish	4
Muslim	5
Hindu	6
Buddhist	7
Other (Specify)	8 (__________________________)
Don’t know (DO NOT READ)	9
No Answer (DO NOT READ)	99

SECTION 2: CURRENT CONDITIONS AND EXPECTATIONS

Q1	Can you tell me **the main reasons** that you started sleeping on the street? (*Enter the number in the order they are mentioned; read out the options **only** if they do not answer*)

ANSWER	ORDER NUMBER
Free Choice	I__I
Family Relationships (separation/abuse)	I__I
Unemployment	I__I
Immigration	I__I
Drug/Alcohol Dependency	I__I
Disability/Illness	I__I
Previous Convictions	I__I
Gambling	I__I
Other (Specify)	I__I (________________________________)
Don’t know (DO NOT READ)	9
No Answer (DO NOT READ)	99

NOTES

Q2	When was the first time you ever slept on the “street”?

Enter the respondent’s exact answer	

Don’t know (DO NOT READ)	9
No Answer (DO NOT READ)	99

NOTES

Q3	Have you slept on the street ever since?

Yes	1	=> Q6
No	2	
Don’t know (DO NOT READ)	9	
No Answer (DO NOT READ)	99	

Q4	Do you remember where you slept when you left the street and for how long? (*Do not read the options out loud)*

		PERIOD (*Specify days/months/years)*
My home	1	
Shelter (emergency/longterm)	2	
Church/Parish	3	
Friends/relatives’ home	4	
Disused Area/Shacks	5	
Hospital	6	
Other (Specify)	7 (____________________)	
Don’t know (DO NOT READ)	9	
No Answer (DO NOT READ)	99	

NOTES

Q5	When did you end up back on the street?

Enter the respondent’s exact answer	

Don’t know (DO NOT READ)	9

No Answer (DO NOT READ)	99

NOTES

Q6	**How much longer do you expect** to sleep on the street? (*Do not read out options)*

Less than 1 month	1	
1 to 3 months	2	
3 to 6 months	3	
6 months to 1 year	4	
Over 1 year	5	
Forever	6	=> Q8
Don't know (DO NOT READ)	9	
No Answer (DO NOT READ)	99	

Q7	How do you plan to get out of your situation? *(Only 1 possible answer)*

Will go back to my home country (of origin)	1
Will ask for community housing	2
Will ask friends/family to host me	3
Will go back to my own home	4
Will enter a drug/alcohol rehab centre	5
Other (Specify)	6 (_____________________)
Don't know (DO NOT READ)	
No Answer (DO NOT READ)	

NOTES

Q8	How long did you **expect to stay** on the street for when you first arrived?

Less than 1 month	1
1 to 3 months	2
3 to 6 months	3
6 months to 1 year	4
Over 1 year	5
Forever	6
Don’t know (DO NOT READ)	9
No Answer (DO NOT READ)	99

NOTES

SECTION 3 : FAMILY

Q1	At the moment _________________is/are: not important, indifferent, important?

	Not important	Indifferent	Important
Family	1	2	3
Friends	1	2	3
Religion	1	2	3
Work	1	2	3
Politics	1	2	3

Q2	On a scale of 1 to 3, where 1 means disagree, 2 is neutral and 3 means agree, state **how much you agree with the following statements**:

	Disagree	**Neutral**	**Agree**
We must always love and respect our parents, regardless of their faults or merits	1	2	3
Parents are responsible for doing what is best for their children, even at the cost of their own happiness and wellbeing	1	2	3
A child needs a father and a mother to grow up happily	1	2	3
The institution of marriage is out of	1	2	3

fashion			

Q3	Are you currently:

Widowed	1	
Married	2	=> Q5
Separated/Divorced	3	=> Q5
Single	4	=> Q5
Other (Specify)	5 (________________)	=> Q5
No Answer (DO NOT READ)		

Q4	Was your wife/husband/partner still alive when you ended up on the street?

Yes	1
No	2
Don’t know (DO NOT READ)	9
No Answer (DO NOT READ)	99

Q5	Do you have any children? If yes how many?

No	0	=>Q7
Yes *enter number of children*	I__I __I	
Don’t know (DO NOT READ)	9	
No Answer (DO NOT READ)	99	

NOTES

Q6	Are all your children still alive?

Yes, all are still alive	0
No, none	1
No, 1 child passed away	2
No, 2 children passed away	3
No, 3 children passed away	4
No, 4 children passed away	5
No, over 4 children passed away	6
Don’t know (DO NOT READ)	9

No Answer (DO NOT READ)	99

Q7	Is your mother still alive?

Yes	1	=>Q9
No	2	
Don’t know (DO NOT READ)	9	
No Answer (DO NOT READ)	99	

Q8	Was your mother still alive when you started living on the street?

Yes	1	
No	4	=>Q10
Don’t know (DO NOT READ)	9	
No Answer (DO NOT READ)	99	

Q9	Was your mother living with you when you started living on the street?

Yes	1
No	2
Don’t know (DO NOT READ)	9
No Answer (DO NOT READ)	99

Q10	Is your father still alive?

Yes	1	=>Q12
No	2	
Don’t know (DO NOT READ)	9	
No Answer (DO NOT READ)	99	

Q11	Was your father still alive when you started living on the street?

Yes	1	
No	2	=>Q13
Don’t know (DO NOT READ)	9	
No Answer (DO NOT READ)	99	

Q12	Was your father living with you when you started living on the street?

Si,	1
No	2
Don’t know (DO NOT READ)	9
No Answer (DO NOT READ)	99

Q13	Have you spoken to any relatives in the **last three months**? If yes who did you speak with?

No, haven’t spoken with anyone	0
Father/Mother	1
Son/Daughter	2
Brother/Sister	3
1st grade (cousins, aunts/uncles...)	4
Other (Specify)	5 (______________________)
Don’t know (DO NOT READ)	9
No Answer (DO NOT READ)	99

Q14	Have you spoken with any relatives in the **last year**? If yes with whom?

No, haven’t spoken to anyone	0
Father/Mother	1
Son/Daughter	2
Brother/Sister	3
1st grade (cousins, aunts/uncles...)	4
Other (Specify)	5 (______________________)
Don’t know (DO NOT READ)	9
No Answer (DO NOT READ)	99

NOTES

SECTION 4: WORK AND INCOME

Q1	Did you have a job before ending up on the street (in the shelter) ?

Yes	1	
No	2	=>Q5
Don't know (DO NOT READ)	99	
No Answer (DO NOT READ)		

Q2	What was your occupation? *(Read answers only if they have trouble responding)*

Factory worker	1
Administration	2
Teacher	3
Cook/Waitress/Waitor	4
Domestic Worker/Nanny/Cleaner	5
Gardener	6
Artisan	7
Self-employed	8
Truck-driver	9
Artist	10
Other (Specify)	11 (____________________)
No answer	99

NOTES

Q3	What was your **monthly** income? *(Read options only if respondent has difficulty responding)*

Enter monthly income	I__I __I __I __I EURO

less than 300 euro	1
from 300 to 500 euro	2
from 500 to 1000 euro	3
from 1000 to 2000 euro	4
over 2000 euro	5
Don’t know (DO NOT READ)	9
No Answer (DO NOT READ)	99

Q4	Was this the last job you had?

Yes	1	=>Q16
No	2	
Don’t know (DO NOT READ)	9	
No Answer (DO NOT READ)	99	

NOTES

Q5	Are you currently working?

Yes	1	
No	2	=>Q9
No Answer (DO NOT READ)	99	

Q6	What is your current job? *(Read answers only if they have trouble responding)*

Factory worker	1
Administration	2
Teacher	3
Cook/Waitress/Waitor	4
Domestic Worker/Nanny/Cleaner	5
Gardener	6
Artisan	7
Self-employed	8
Truck-driver	9

Artist	10
Other (Specify)	11 (____________________)
No answer	99

NOTES

Q7	What is your **monthly** income? *(Read options only if respondent has difficulty responding)*

Enter monthly income	I__I __I __I __I EURO

less than 300 euro	1
from 300 to 500 euro	2
from 500 to 1000 euro	3
from 1000 to 2000 euro	4
over 2000 euro	5
Don’t know (DO NOT READ)	9
No Answer (DO NOT READ)	99

Q8	What type of contract do you have?

Permanent Contract	1	=>Q20
Non permament Contract	2	=>Q20
Don’t have a contract/paid under the table	3	=>Q20
Don’t know (DO NOT READ)	9	=>Q20
No Answer (DO NOT READ)	99	=>Q20

Q9	Did you work last month? (*Read options)*

Yes, always	1	
Yes, occasionally	2	
No	3	=>Q13
No Answer (DO NOT READ)	99	

Q10	What job did you do? *(Read answers only if they have trouble responding)*

Factory worker	1
Administration	2
Teacher	3
Cook/Waitress/Waitor	4
Domestic Worker/Nanny/Cleaner	5
Gardener	6
Artisan	7
Self-employed	8
Truck-driver	9
Artist	10
Other (Specify)	11 (__________________)
No answer	99

NOTES

Q11	What was your **last month's** income? *(Read options only if respondent has difficulty responding)*

Enter monthly income	I__I __I __I __I EURO

less than 300 euro	1
from 300 to 500 euro	2
from 500 to 1000 euro	3
from 1000 to 2000 euro	4
over 2000 euro	5
Don't know (DO NOT READ)	9
No Answer (DO NOT READ)	99

Q12	What type of contract did you have **last month**?

Permanent Contract	1	=>Q16
Non permament Contract	2	=>Q16
Don't have a contract/paid under the table	3	=>Q16
Don't know (DO NOT READ)	9	=>Q16
No Answer (DO NOT READ)	99	=>Q16

NOTES

Q13	What was the last job you held? *(Read answers only if they have trouble responding)*

Factory worker	1
Administration	2
Teacher	3
Cook/Waitress/Waitor	4
Domestic Worker/Nanny/Cleaner	5
Gardener	6
Artisan	7
Self-employed	8
Truck-driver	9
Artist	10
Other (Specify)	11 (__________________)
No answer	99

NOTES

Q14	What was your **monthly** income? *(Read options only if respondent has difficulty responding)*

Enter monthly income	I__I __I __I __I EURO

less than 300 euro	1
from 300 to 500 euro	2
from 500 to 1000 euro	3
from 1000 to 2000 euro	4
over 2000 euro	5
Don’t know (DO NOT READ)	9
No Answer (DO NOT READ)	99

Q15	What type of contract did you have in your very last job?

Permanent Contract	1
Non permament Contract	2
Don’t have a contract/paid under the table	3
Don’t know (DO NOT READ)	9
No Answer (DO NOT READ)	99

NOTES

Q16	Are you looking for a job?

Yes	1	
No	2	=> Q18

NOTES

Q17	How are you looking for a job ? (through what channels)

Friends/Relatives	1
Work placement office (Municipality)	2
Temporary work agency	3
Voluntary associations (Red Cross, Caritas..)	4
Other (Specify)	6 (_____________________)
Don’t know (DO NOT READ)	9
No Answer (DO NOT READ)	99

Q18	If a job were available in the next two weeks, what would be the **minimum monthly** income you would accept to start working?

Enter monthly income	I__I __I __I __I EURO
I wouldn't accept a job in the next two weeks	1
Don't know (DO NOT READ)	9
No Answer (DO NOT READ)	99

Q19	What are your main sources of income *(where do they get money from)*? *(More than 1 answer is possible but do not read the answers out loud)*

	TYPE OF INCOME	TOTAL (Euro)	WEEKLY / MONTHLY	
Welfare cheque	1	I__I__I__I__I	S	M
Unemployment Insurance	2	I__I__I__I__I	S	M
Disability Insurance/Pension	3	I__I__I__I__I	S	M
Stable work	4	I__I__I__I__I	S	M
Occasional work	5	I__I__I__I__I	S	M
Family/Relatives	6	I__I__I__I__I	S	M
Friends	7	I__I__I__I__I	S	M
Pension	8	I__I__I__I__I	S	M
Begging	9	I__I__I__I__I	S	M
Other (Specify)	10 (______________________)	I__I__I__I__I	S	M
Don't know (DO NOT READ)	9			
No Answer (DO NOT READ)	99			

Q20	Are you able to save any money from your earnings?

Yes	1
No	2
Don't know (DO NOT READ)	9
No Answer (DO NOT READ)	99

Q21	Have you ever received financial help from **close family members** (parents, husband/wife, children) in the last year?

Yes	1
No	2
Don’t know (DO NOT READ)	9
No Answer (DO NOT READ)	99

Q22	Do you receive any **non-financial** help (eg. food, clothes, other objects …)?

Yes	1	
No	2	=>Q25
Don’t know (DO NOT READ)	9	
No Answer (DO NOT READ)	99	

Q23	What type of help and how often? *(List the alternatives and write an “X” in the corresponding cell)*

Type	Daily	2 or more times a week	Weekly	Monthly	A few times a month	A few times a year
Food						
Clothes						
Sleeping bag/blanket/tent						
Medicines						
Don’t know (DO NOT READ)						
No Answer (DO NOT READ)						

NOTES

Q24	Who gave you help in kind?

TYPE OF HELP	WHO
Food	
Clothes	
Sleeping bag/blanket/tent	
Other:	
Don’t know (DO NOT READ)	
No Answer (DO NOT READ)	

NOTES

Q25	In the last week have you bought/spent money on ___________? How much did you spend?

ITEM	ENTER “X”	AMOUNT (EURO)
Water/beverages		
Food		
Clothes		
Sleeping bag/blanket/tent		
Cigarettes		
Wine/Spirits		
Prepaid phone cards/Telephone expenses		
Mobile phone		
Medicines		
Entertainment (movies, hanging out ...)		
Travel		
Places to sleep (hotels, hostels …)		
Gambling (horses, lotteries, betting)		
Hand bag/luggage/carrying case		
Other (Specify)		
Don’t know (DO NOT READ)		
No Answer (DO NOT READ)		

NOTES

SECTION 5: EDUCATION

Q1	Do you know how to read (in your language)?

Yes	1
No	2
Don’t know (DO NOT READ)	9
No Answer (DO NOT READ)	99

Q2	Do you know how to write (in your language)?

Yes	1
No	2
Don’t know (DO NOT READ)	9
No Answer (DO NOT READ)	99

Q3	Have you ever gone to school? If Yes, what is the highest level of education you received?

Never went to school	0
Elementary School	1
Middle School	2
Professional Diploma	3
Highschool Diploma	4
University/Bachelor Degree	5
Master/PhD	6
Don’t know, don’t remember (DO NOT READ)	9
No Answer (DO NOT READ)	99

NOTES

Q4	Did your mother go to school? If Yes, what is the highest level of education she received?

She never went to school	0
Elementary School	1
Middle School	2
Professional Diploma	3
Highschool Diploma	4
University/Bachelor Degree	5
Master/PhD	6
Don’t know (DO NOT READ)	9
No Answer (DO NOT READ)	99

NOTES

Q5	Did your father go to school? If Yes, what is the highest level of education he received?

He never went to school	0
Elementary School	1
Middle School	2
Professional Diploma	3
Highschool Diploma	4
University/Bachelor Degree	5
Master/PhD	6
Don’t know (DO NOT READ)	9
No Answer (DO NOT READ)	99

NOTES

SECTION 6: CONTACTS AND TRUST

Q1	Do you know other people who sleep on the street? If yes, how many?

No, don't know anyone	0	=>Q3
from 1 to 5	1	
from 5 to 10	2	
from 10 to 20	3	
over 20	4	
Don't know (DO NOT READ)	9	
No Answer (DO NOT READ)	99	

Q2	Of these, can you please tell me the name of the first 5 people that you turn to for help? And how long you know these people?

NAME	SURNAME *(first two letters of the last name)*	TIME KNOW EACH OTHER *(Specify days, months, years)*
Don't know (DO NOT READ)	9	
No Answer (DO NOT READ)	99	

Q3	Do you **usually** pass the night with a group (1 or more) of people? If yes, how many people?

Enter the respondent's exact answer	

Don't know (DO NOT READ)	9
No Answer (DO NOT READ)	99

Q4	Have you asked anyone for help in the **last year**?

Yes	1	
No, I haven't asked anyone for help	2	=>Q6
No, I don't need help	3	=>Q6
Don't know (DO NOT READ)	9	
No Answer (DO NOT READ)	99	

Q5	Who did you ask for help and how often? (*Over the past year)*

		NUMBER OF TIMES
No-one	0	
Father/Mother	1	
Son/Daughter	2	
Brother/Sister	3	
1st grade relatives (cousins, aunts/uncles...)	4	
Friends	5	
Church/Parish	6	
Voluntary Associations	7	
Other (Specify)	8 (________________________)	
Don't know (DO NOT READ)	9	
No Answer (DO NOT READ)	99	

Q6	On a scale of 1 to 3, where 1 means no trust at all and 3 means lots of trust, how much do you trust in: _________ ? (*Read the options out loud*)

	No trust at all	Indifferent	Lots of trust
Family	1	2	3
People living in the same situation	1	2	3
People of other nationalities	1	2	3
Churches	1	2	3
Hospitals	1	2	3
Voluntary Associations (Red Cross, Saint Francis Friars, OCF, City Angels..)	1	2	3
Police	1	2	3
Government	1	2	3

Q7	Consider two **imaginary situations**. In the first scenario you are given **800 euro** right away. In the second you are asked to **toss a coin**: if the coin lands on HEADS you won't win a thing; if it lands on TAILS you will get **8.500 euro**. What option would you choose? (*Ensure that the respondent clearly understands the question)*

800 EURO right away	1
Would toss the coin and see what happens	2
Don't know (DO NOT READ)	9
Doesn't answer (DO NOT READ)	99

Q8	Consider another two **imaginary situations**. If you could be given a house right away or 1000 euro every month. What option would you choose? (*Ensure that the respondent clearly understands the question)*

House	1
1000 euro a month	2
Don't know (DO NOT READ)	9
Non rispondere (NON LEGGERE)	99

SECTION 7: AWARENESS

Q1	Can you tell me **today's date**? If yes, what is the date today?

		Day	Month	Year
Yes	1	I__I__I	I__I__I__I__I__I__I__I__I__I	I__I__I__I__I
No, Don't know (DO NOT READ)	2			
No Answer (DO NOT READ)	99			

NOTES

Q2	Can you tell me what day of the week this is?

Monday	1
Tuesday	2
Wednesday	3
Thursday	4
Friday	5
Saturday	6
Sunday	7
Don’t know (DO NOT READ)	9
No Answer (DO NOT READ)	99

Q3	When was the last time you read a newspaper (daily)?

Today	1
1 week ago	2
1 month ago	3
6 months ago	4
1 year ago	5
Over 1 year ago	6
Never read a newspaper	7
Don’t know (DO NOT READ)	9
No Answer (DO NOT READ)	99

Q4	When was the last time you saw a television newscast or listened to a radio newscast?

Today	1
1 week ago	2
1 month ago	3
6 months ago	4
1 year ago	5
Over 1 year ago	6
Never seen/heard a newscast	7
Don’t know (DO NOT READ)	9
No Answer (DO NOT READ)	99

Q5	Who is the Prime Minister of Italy? *(Read the options)*

Silvio Berlusconi	1
Romano Prodi	2
Giorgio Napolitano	3
Adriano Celentano	4
Gianni Agnelli	5
Don't know (DO NOT READ)	6
No Answer (DO NOT READ)	6

NOTES

SECTION 8: HEALTH AND GENERAL INFORMATION

Q1	Have you ever gotten sick in the last month? For example have you had coughs, colds, the flu, diarrhea or lesions/wounds?

No	1	=> Q4
Yes	2	
Don't know (DO NOT READ)	9	
No Answer (DO NOT READ)	99	

Q2	What sickness did you have?

Enter exact response	

Q3	Did you go to anyone for a check-up? Where?

No, I didn't go for a check-up	0
Hospital/Emergency	1
Shelter Doctor	2
Clinic	3
NAGA (medical/social assistance association)	4
Opera San Francesco (St. Francis Friars)	5
Other (Specify)	6 (__________________)
No Answer (DO NOT READ)	99

NOTES

Q4	Serious disabilities (DO NOT ASK THIS QUESTION: OBSERVE THE RESPONDENT AND ENTER THE RESPONSE DIRECTLY)

Yes	1
No	2

NOTES

Q5	What is the reason for your current living conditions?

Bad luck	1
Lack of opportunities	2
Lack of worth	3
Don’t know (DO NOT READ)	9
No Answer (DO NOT READ)	99

Q6	Do you have a permit of stay?

Yes	1
No	2
No Answer (DO NOT READ)	99

Q7	Have you ever been in jail?

Yes	1	
No	2	=>Q9
No Answer (DO NOT READ)	99	

NOTES

Q8	Have you been in jail before or after becoming homeless?

Before homeless	1
After homeless	2
Don’t know	9
No Answer (DO NOT READ)	99

NOTES

Q9	In the **last year** do you believe that your life has ___________________?

Greatly improved	1
Slightly improved	2
Remained the same	3
Slightly worsened	4
Greatly worsened	5
Don’t know (DO NOT READ)	9
No Answer (DO NOT READ)	99

NOTES

Q10	In the **next year** do you believe that your life will ___________________?

Greatly improve	1
Improve	2
Remain the same	3
Worsen	4
Greatly worsen	5
Don’t know (DO NOT READ)	9
No Answer (DO NOT READ)	99

NOTES

INTERVIEW FINISH TIME I___I__I__I__I

****

TO BE FILLED IN BY THE INTERVIEWER AT THE END OF THE INVERVIEW

INTERVIEWER FIRST NAME	I__I__I__I__I__I__I__I__I__I__I__I__I__I__I__I__I__I__I

INTERVIEWER SURNAME	I__I__I__I__I__I__I__I__I__I__I__I__I__I__I__I__I__I__I

MILAN ZONE WHERE THE INTERVIEW TOOK PLACE (01-66)	I__I__I

ENTER NAME OF STREET/ SQUARE	I__I__I__I__I__I__I__I__I__I__I__I__I__I__I__I__I__I__I__I__I__I__I

STATE OF MIND OF THE RESPONDENT

Lucid	0
Aware	1
Problematic	2

INTERVIEW LANGUAGE

Italian	0
English	1
Spanish	2
Romanian	3
Other (Specify)	4 (___________________________)

FINAL NOTES

Chapter 2

Learning (or not) in health seeking behavior: Evidence from rural Tanzania

Abstract[1]

The aim of this paper is to understand the functioning of individuals' health seeking behavior. It studies empirically how and if economic agents adjust their health seeking behavior over time, after they have gained experience about the quality of the previous caregiver consulted. I find that agents seek medical care repeatedly from the same type of health provider, even if the treatments are ineffective. Specifically, they do not switch to formal health sector, even though the informal one has failed to treat their illness in previous years. The paper also investigates the determinants of illnesses, showing the relevant role of education in avoiding diseases. I apply the Wooldridge's (1995) procedure to deal with the sample selection bias problem, by exploiting past natural disasters as exclusion restriction. These effects are tested using a 4 years panel data from a household survey in Tanzania.

[1]I wish to thank Eliana La Ferrara for her guidance and support; Martina Bjorkman, Domenico Depalo, Andrea Gamba and the participants at the CSAE Conference in Oxford and at the EUDN Conference at Paris School of Economics for their very useful comments. I am also indebted to Kathleen Beegle for suggestions on the datasets used and to Paola Porta for sharing information on Kagera's diseases. All errors are my own.

2.1 Introduction

Illness is one of the most serious problems that can affect a household in African communities, where a considerable part of the population seeks medical care in the informal health sector, e.g. traditional healers and witch-doctors, or it doesn't receive any health treatments. Economics research (Björkman and Svensson 2006), international organizations (USAID 2006, WHO 2005, The World Bank 2004) and governments' efforts are mainly focused on the supply-side of the health market: how to improve health systems and how to increase the number and the quality of accessible medical services. In recent years, the growing attention in studying health care demand (Lindelow, 2002, Leonard 2007) has started to give some insights to better understand failures in the health systems of many developing countries. A limitation of this literature, however, is that it doesn't analyze the pattern of health seeking behavior over time. The main contribution of the present paper is to study individual's health seeking behaviors, conditional on the outcome of the previous therapy. It explores whether agents switch to a better quality health provider, after the failure in treating an illness on the part of the first caregiver consulted. The idea is to understand whether remaining sick after a treatment is a good incentive to switch to an alternative care. Understanding incentives/causes driving individuals towards more efficient health care is relevant to designing long term policy measures in the health systems. As main results, the paper shows that individuals, who sought care from informal caregivers don't switch to formal health cares even if they are still suffering from the same disease after the informal treatment. It emerges that the prior belief about what is the best medical care dominates over the illness status of individuals. The intuition under this result is straightforward: as long as the costs to test a new treatment are higher then the potential benefits coming from the new treatment, even a completely "rational" agent will choose not to experiment. This player gets stuck in a non-optimal equilibrium, simply because the cost of trying something else is higher compared with the potential benefits. In the paper, an agent has to compare the certain cost of formal health care (far away government-run clinics, long queues, etc.) with an uncertain benefit (the probability to heal). The confirmation of health seeking choices independently of their outcomes fosters the existence of unqualified doctors and health services. In addition, this work proposes a unique procedure to solve the selection bias problem, by exploiting past disasters as exclusion restriction, correlated with the probability of illness but not with the probability to look for formal health care.

The paper attempts to address three main questions. First, do agents switch to a formal caregiver, following only private information about their bad health status after the informal treatment? Second, what factors influence the demand for formal health care? Third, what are the main determinants of becoming ill?

Regarding the first question, I show that agents do not learn from their own past experience that it is better to switch to formal medical care after an ineffective informal therapy. Individuals choose once and for all at the beginning of the period and they do not update their believes, even if they have evidence of inefficient outcomes. The formal institutions I considered are hospitals, health centres, clinics and dispensaries and the informal ones are practitioners' homes, pharmacies, family homes and self-care. Ceteris paribus, to consult informal providers and to remain sick after the treatment decreases the probability to visit formal establishments in the next period by 11.5 percentage points.

A second result concerns the determinants involved in consulting formal or informal health care providers, conditional on reporting illnesses. The main determinant of the probability to seek formal care is the distance between the household and formal health establishments. As expected, more educated individuals tend to choose formal caregivers. Having access only to bad sources of drinking water, such as rain, lake and river water and living in an inadequate house, decreases the likelihood to seek formal medical care. Education, the quality of drinking water and living conditions also capture income and the ability to afford better services.

Finally, I show that disaster, such as drought, epidemic, insect and crop diseases, happened in a community six months prior to the survey increases the probability of illnesses. Years of education and bad living conditions are the main variables that determine one's possibility of becoming ill, affecting it respectively in a negative and in a positive way. Furthermore, women are more likely to report diseases compared to men.

The proposed analysis is tested using household panel data from Kagera, Tanzania. The Kagera Health Developing Survey has several features that make it particularly appropriate for studying individual's health seeking behavior. First, it contains detailed information on ill individuals as well as the type of illness reported, the decision to seek care, the type of health provider chosen and if respondents are still suffering from the same disease after a health visit.

Second, more than one half of respondents reported an illness or an injury in the four weeks proceeding the survey. This high percentage provides a significant sample of ill individuals to analyze. Finally, the KHDS includes a community questionnaire. Hence, I exploit information on community shocks as an instrument to study the probability of illness and to solve the sample selection bias problem.

The paper is related to three main strands of the literature. First, it relies on studies on the determinants of health seeking behavior, such user fees (Gertler and Van Der Gaag 1990, Dasgupta and Gupta 2002), travel distance (Acton 1975), individual and household characteristics (Lindelow 2005). Case, Menendez and Ardington (2005) examine patterns of health seeking behavior prior to death among individuals in a South Africa district, finding that all adults who were ill prior to death sought treatment from a Western medical provider and, the fifty percent of them, from a traditional healer, suggesting that traditional medicine is seen as a complement to, rather than a substitute for, Western care. While previous works are mainly based on cross-section datasets and they adopt a static approach, this paper studies the health seeking behavior over time in a panel dataset, by conditioning the choice of health provider to the results of the previous therapy.

Second, the paper fits, for some aspects, into the more recent research highlights the role of social learning played in health seeking behavior. Das and Hammer (2004) evaluate how provider quality affects both the demand for health services and health outcomes, analyzing the competence of caregivers through vignettes and clinical observations in Delhi. Luke and Munshi (2007) assess the role of social affiliation, measured by caste in India, on household health care decisions and they propose a network-based explanation for why investments in health may differ across castes. Within this literature, the paper most related to my work is Leonard (2007). Using two separate datasets from Tanzania, he shows that households, by gathering information and by communicating with nearby households, learn about the best quality clinicians. Different from the latter, the idea of the present paper is to study the capacity to modify/switch behavior whenever new individual experience about provider's quality is acquired.[2] Psychological research indicates that some people have a cognitive bias that leads them to misinterpret new information as supporting previously held hypotheses.

[2] In the literature this is better known as the concept of "learning by doing" (Foster and Rosenzweig, 1995).

Indeed, the results of the paper can be thirdly absorbed into the literature studies confirmation bias. Rabin and Schrag (1999) show that such confirmatory bias induces overconfidence: the agent may come to believe a false hypothesis despite receiving an infinite amount of information. The paper provide empirical evidence that agents confirm the choice of the first health provider consulted, no matter if past experience and new information have been acquired.

The reminder of the paper is organized as follows. Section 2 introduces some notions on the health system in Tanzania. Section 3 describes the data and illustrates the main pattern of health seeking decision. Section 4 shows the empirical strategy used and in section 5 I expose the main results. Section 6 concludes.

2.2 Health care system in Tanzania

Looking at other eastern African countries, Tanzania's health care system is relatively well-established. Health indicators are still slightly above the average for sub-Saharan Africa and the health sector is doing relatively well. Infant mortality rate fell from 99 deaths per 1,000 live births in 1999 to 68 per 1,000 in 2005. Children's malnutrition status also improved. Between 1999 and 2005, the incidence of stunting decreased from 44 percent to 38 percent, wasting from 5 to 3 percent and underweight from 29 to 22 percent. In contrast, maternal mortality remains high at 578 deaths per 100,000 live births in 2005, life expectancy at births is 51 years old and according to the 2003-04 Tanzania HIV Indicator Survey (THIS) the 7% of adults where infected with HIV/AIDS. Malaria is the leading cause of death in Tanzania and the major public health concern, especially among pregnant women and children under five years (NBS and ORC Macro 2006).

The health sector in Tanzania improved after the independence from the United Kingdom in 1964, especially during the Arusha Declaration[3], when there was an emphasis for a more equal and efficient access to social services; before they were mainly basic and concentrated only in more developed urban centres. Free medical services were introduced and facilities were redirected towards rural and poorer areas. In 1991, the Government revised its approach towards private initiatives and it officially recognized the private activities of medical practitioners and

[3] A declaration outlining Tanzania's policy on socialism and self-reliance, by Julius Nyerere.

dentists. Today, even if the government remains the main source of financing for the health sector, private organizations and qualified persons can actively participate in the development and management of health care services in Tanzania, with the approval from the Ministry of Health.

The health system has a pyramidal structure from services at district level to medical care at national level.[4] The lowest level of health care delivery in each district of the country is the village health service or health post. It essentially provides preventive services which can be offered in patients' homes, such as advice on personal habits and domestic hygiene (water, food), as well as the distribution of basic drugs. Usually, each health post has two workers chosen by the community who take short training courses before starting to provide services. The second stage of health services at district level is the dispensary. The dispensary provides care approximately for 6,000 to 10,000 people and supervises all the village health posts in its ward. Dispensaries in Tanzania are generally located in rural centres, serving population spread over a wide area. They are staffed by a rural medical aid (RMA) with one or two helpers. The RMA receives a 3 years course of training in anatomy, physiology and hygiene with good grounding in diagnostic methods and treatment of common diseases, but they still offer limited types of surgery. Some dispensaries are managed by religious groups and, as a whole, they tend to be better equipped than those in the state sector, with higher standard of cleanliness and hygiene. The health centre provides a slightly higher level of medical care. It is expected to cater to 50,000 people, which is approximately the population of one administrative division. Most of health centres have a room for minor surgery and they provide 20-30 beds for inpatients including maternity cases. Each district is supposed to have a hospital. The district hospital is the base for staffing and supplying all rural units and it is to the district hospital that any difficult of serious case is referred. The regional hospital offers similar assistance like that agreed at the district level, but with specialists in various fields and additional services. Finally, at the top there are four referral hospitals which serve the whole country. Referral hospitals are managed by the Ministry of Health, while local governments are responsible for dispensaries, health centres and the district/regional hospitals. In this official classification clinics do not

[4] Ministry of Health and Social Welfare, The United Republic of Tanzania, National website, www.tanzania.go.tz., 2007.

appear. They provide similar services to health centres and they are generally managed by NGOs or private organizations.

In addition to this formal structure of the health care delivery system, the sector includes individuals not recognized by the Ministry of Health such as practitioners working from their homes, traditional healers, quacks and witch-doctors. They serve as primary health care providers for a large part of the population. Witch-doctors have no formal training and acquired all of their practices from their parents who in turn learned from their grandparents. Their only diagnostic tools are patients' symptoms and for each ailment they prescribe a combination of songs, prayers, particular kinds of soil, and tree roots. In slums and rural area, sick poor people tend to consult them, since formal health establishments are far away and with long waiting lists; while informal caregivers generally visit people at home and so they are more accessible or easier to reach by all patients.

2.3 Data and Descriptive statistics

The data used for this study come from a research project conducted by the World Bank and the University of Dar es Salaam in the Kagera region, located in the North-western corner of Tanzania and bordering Uganda to the North and Rwanda and Burundi to the West. The prime objective of the survey was to measure the economic impact of fatal illness in the region and to propose cost-effective strategies to help survivors. Therefore, it includes a large set of questions on health care and health behaviors. The Kagera Health and Development Survey (KHDS) surveyed 816 households four times from 1991-1994, with an average interval between surveys of 6-7 months. Households were selected in the six districts of Kagera region (Karagwe, Bukoba Urban, Bukoba Rural, Muleba, Biharamulo and Ngara), mainly in rural villages (Figure 1).[5]

[Insert table 1]

The sample includes 19,009 individuals interviewed up to four times in 51 clusters. The analysis focuses on two dependent variables: the probability to seek formal health care in the

[5] In 2004 another round data was collected (KHDS 2004) but the paper presents evidence only on the first four years due to the lack of a key variable for my identification strategy in the last wave. The KHDS 2004 doesn't include the question on the individual health outcome (still sick or not) after a treatment.

four weeks prior the survey and the probability of illness. Table 1 provides information on the percentage of ill individuals by gender and age. In the sample, more than the half of individuals reported being ill or injured in the four weeks proceeding the survey. As expected, illness reporting was particularly high for the elderly (46 years or older) and for infants between 0-2 years old. The lowest percentage of ill individuals belonged to the class of children between 6-15 years old. Women were slightly more prone to report illness than men in all age categories, except for male children under 2 years old.[6]

[Insert table 2]

Evaluating patients' choice towards a specific caregiver requires data on the disease suffered as it allows to account for the tendency to seek formal care only for more serious illnesses. Given the impossibility of controlling directly for the disease for lack of data, I exploit all the information I have regarding type of symptoms. First, I grouped the reported symptoms in two broader categories, depending on how serious they are: "routine" and "not routine" symptoms (Luke and Munshi, 2007). The "routine" ones are the most common and most widespread such as acute diarrhea, chronic diarrhea, weight loss, weakness, vomiting, acute fever, chills, cough, sore throat, severe headache, skin rush and wound. The "not routine" are recurring fever, fainting, difficult breathing, productive cough, coughing blood, abdominal pain, pain on passing urine, genital sores, burn, fracture, child birth and mental disorder.[7] Table 2 describes the symptoms suffered by the sick agents in the sample. Almost 90% of individuals reported less severe symptoms. Among those, the prevalent is acute fever (23.55%), one of the main symptoms of malaria. Roughly 13% of the sample reports cough and about 10.89% is suffering from severe headache, followed by chills, wound and acute diarrhea. Approximately 21% of individuals reported to have other types of symptoms, different from the listed ones. Finally, among the not routine symptoms, the share of individuals having abdominal pain is around 6.5%. Second, I divided the symptoms reported in two other categories: cyclic and not cyclic,

[6] In what follows, I consider illness and injury as the same variable.

[7] The classification between routine and not routine symptoms has been costructed by an anonymous doctor working in Kagera region.

depending on the fact that they appear or not in regular intervals of time.[8] This distinction is relevant to understand the efficacy of a treatment. Cyclic diseases show up independently from the medical care chosen. In this case, the health status (sick or not sick) doesn't depend on the therapy but only on the illness's course. In the empirical part of the paper, I restrict the analysis to the sub-samples of routine, not routine and not cyclic symptoms.

[Insert table 3]

Third, I infer a patient's illness by combining the first two symptoms reported by each sick respondent with the most spread illnesses in Kagera region. This additional check is relevant since for example one or more routine symptoms could even generate a serious disease. The first row of table 3 shows the most widespread diseases in Kagera, while the first column describes all the possible symptoms listed in the survey. Each illness is linked to the corresponding symptoms. In the table, 1 and 2 are, respectively, the first and the second symptom for that specific disease. The X represent all additional symptoms that could appear if the agent is suffering from the disease reported in the first row of the table. I infer an agent's predicted illness by considering the first two symptoms suffered by individuals. For example, if an individual reported acute fever as first symptom and chills as the second one, he might probably suffer from malaria.[9] In the empirical section, the estimates include predicted diseases as additional controls.

[Insert table 4]

The survey proceeds as follows: individuals who reported an illness in the four weeks prior to the survey were asked if anybody had been consulted for treating this illness and, consequently, which was the first place where the patient sought care. The treatment choices available in Kagera region are hospitals, health centres, dispensaries, clinics, pharmacies, practitioner's homes, and patient's homes. As table 4 (panel A) shows, the greater part of the sample (68.3%) doesn't consult anyone for their disease and this percentage is slightly higher for women. In slums and rural area, sick poor people don't seek health care because clinics are too far away

[8] The symptoms reported as not cyclic are: Acute Diarrhea, Weight Loss,Acute Fever, Skin Rash, Weakness, Chills, Productive Cough, Blood Cough, Urine Pain, Genitals, Sore Throat, Burn, Fracture, Wound, Child Birth (Harrison 2006).

[9] Table 3 has been costructed with the help of two anonymous doctors, one of whom, work in Kagera region.

and the queues too long. In many rural areas there are no government-run clinics or dispensaries, so the only solutions are often to give up a treatment or to opt for a quack. Splitting the sample by age groups, it emerges that the highest percentage of individuals receiving medical care are infants (40%). To analyze individual health seeking behavior, I aggregate the health establishment choices in two broader categories: informal and formal health care. Hospitals, health centres, clinics and dispensaries are considered formal cares.[10] Informal health establishments include pharmacies, practitioner's homes, their own home and "other".[11] Table 4, panel B reports the percentages of sick agents who decide to consult someone for treating their illness: more than 13% of these choose informal health care. The reaction to illnesses could vary with individual characteristics, such as gender and age. More men than women tend to seek informal care and the highest percent of individuals who report an informal consultation is represented by elderly over 45 years. Among formal health care, roughly the 23% of infants have been visited in hospitals compared to the 30% of adults between 31 and 45 years old. Dispensaries are the most widespread formal establishments where individuals decided to seek treatments, conditional on illness and consultation, and they are particularly widespread among kids aged 3-5 years old. Approximately 10% of agents sought care in a health centre, while clinics are less attended. In the informal sector, individuals are mainly treated in traditional practitioner's homes and this percentage increases for the elderly. The survey also investigates if sick people look for health care for the same illness in any other establishments after the first visit. The 88% of individuals don't consult any additional caregivers for the same disease. This is a crucial information since to empirically study the respondent's health seeking behavior, we have to be sure that the health status after a health care visit (formal or informal) is a consequence of that specific treatment and not of many treatments within the same month.

[Insert figure 2]

The crucial point of the paper is to investigate health behaviors in the wave after a formal or informal health treatment. The formal or informal therapy didn't work whether individuals sick in the four week proceeding the survey are still ill from the same disease during the survey

[10] The formal health centres are all the institutions officially recognised by the State in Tanzania (Tanzania National Website, 2008: www.tanzania.go.tz/health.htlm).

[11] Partages, shops or private laboratories.

after a medical consultation. Figure 2 describes the timing of the questions asked in the survey. The questions are repeated for four waves.

[Insert table 5]

Table 5 is a transition matrix showing agent's behavior from wave t-1 to wave t for those who were still suffering from the same disease during the survey in t-1. Only the 20 percent of respondents, who have visited an informal caregivers (or haven't received any care) and have remained sick afterwards, have switched towards formal health providers in the following wave. This shows that the 80 percent of the still sick informal seekers have confirmed their health provider's choice. While the 58 percent of the "formal care visitors" in t-1, for which the treatment didn't work, has shifted towards informal cares (or has opted for no treatment) in wave t. From table 5, I construct the main independent variable in the empirical section: INFORMAL/NO CARE DIDN'T WORK is a dummy taking value 1 if the respondent has consulted an informal providers in t-1 and he is still sick during the survey in t-1 (1013 observations) and equal 0 if he remained sick in t-1 although he visited a formal caregivers (384 observations).

[Insert table 6]

Table 6 shows a cost-benefit analysis as a tool to evaluate advantages and disadvantages of the formal health sector in Kagera. The table provides descriptive evidence that searching care in formal institutions increases the probability to heal after the treatment. The 47 percent of patients, who received care for the first time by an informal provider, is still suffering from the same disease compared to 40 percent of those remained sick after therapy in hospitals or clinics. The test for the equality of the two coefficients rejects the null hypothesis with a p-value of 0.01. The t-test is valid also considering the sub-group of routine symptoms, while the difference between formal and informal care is not statistically significant when considering individuals suffering from more serious diseases, suggesting that some chronic illnesses cannot be benefit neither from the informal, nor from the formal health sector. By comparing agents who seek formal caregivers with those who haven't consulted any health providers, sick individuals are 40% and 42% respectively, but the t-test is not statistically significant anymore. Considering only those suffering from not routine symptoms, I found a greater percentage of sick agents after a formal therapy compared to those who didn't opt for medical care. The latter result can be explained by a higher proportion of death among those affected by a serious disease who had not sought health care.

A second slot of results describes the fraction of ill individuals treated by a doctor by type of health care. The percent of patients who have been visited by a doctor is 19 percent for those seeking formal care and 0.4 percentage for the ones consulting informal health providers. So, why should rational agents look for medical care in the informal health sector? The last row of table 6 shows the main cost associated with formal health care: the reported distance between the household and the medical establishment individuals have consulted. The distance from the household and the informal health facilities visited by each member is 2 km, much lower than the distance necessary to reach formal establishments. In the empirical section, I compute the average distance, among household members, between the household and the formal/informal caregivers they visited. On average, the distance between a household and a formal establishment is equal to 5.6 percent km, while the distance to visit an informal practitioner is lower than 1 kilometer.

Table A1 in the appendix shows summary statistics of the sample. As explanatory variables I used individual and household characteristics. The individual features include age, gender, years of education, religion and health status. The best measure of income in the KHDS is the value of physical assets held by the household (including the value of land, business, equipment, livestock, and dwelling). In the empirical part, I measure wealth as the log value of durable assets per household. Further, I control for the source of drinking water for the family and the living conditions. The measure of unsafe water includes water from rivers, lakes, wells without pump and rain. 80% of individuals had access only to unsafe drinking water and the 25% lived in a dwelling composed of walls of mud or bamboo, with soil as the main flooring and by roofs made of grass or mud. Finally, I exploit a variable from the community questionnaire as exclusion restriction in the empirical strategy: disasters that happened in each community in the past six months, such as flood, drought, epidemic and crop diseases.

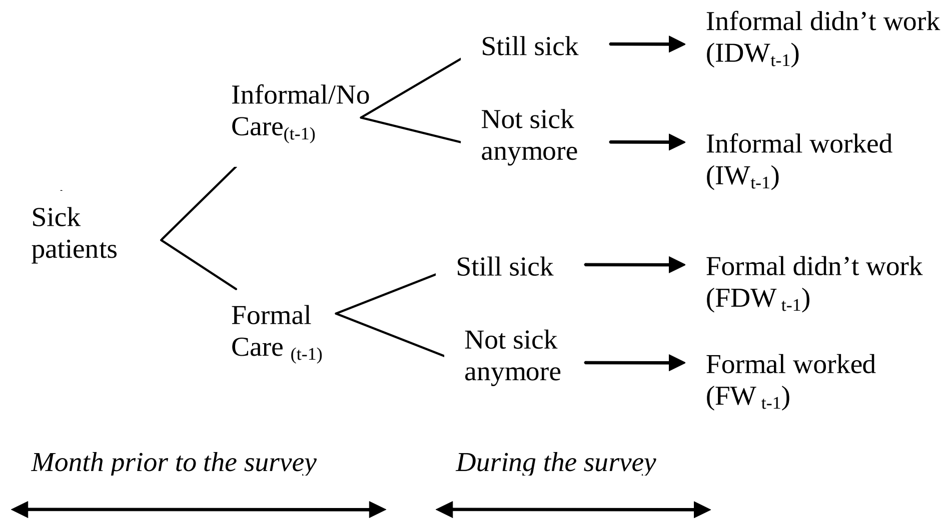
2.4 Empirical Strategy

2.4.1 Estimating equations

The main set of equations has the purpose to assess whether a patient's decision between formal and informal health care changes over time, depending on an agent's health status after an informal treatment. A patient would switch towards a formal health provider in the following wave when the informal one hasn't been effective in caring his/her disease. I adopt the following estimation strategy:

$$\begin{aligned} Y_{it} &= \alpha + \beta_1 IDW_{t-1} + \beta_2 IW_{t-1} + \beta_3 FW_{t-1} + \beta_4 Distance_{it} + \\ &\quad \beta_5 X_{it} + C_{it} + d_t + d_d + \varepsilon_{it} \end{aligned} \tag{2.1}$$

where Y_{it} is a binary variable equal one if the agent i seeks formal medical care and equal zero if he gets informal treatments or no treatments at time t. This variable is observed only if the individual is ill at time t, because, by construction, an healthy individual doesn't consult neither formal nor informal health providers. The selection bias problem involved in estimating equation 1 has been discussed and solved in paragraph 4.2. The KHDS investigates patient's health status after the formal/informal treatment, by asking whether he/she is still suffering from the same disease today (at the time of the interview). The following graph shows the four combinations of possible events and how the independent variables in equation 1 have been constructed.



The main variable of interest in equation 1 is INFORMAL/NO CARE DIDN'T WORK in t-1 (IWD_{t-1}), a dummy assuming value one if the agent i sought informal care or hadn't medical care in a month prior the survey in wave $t-1$, and he is still suffering from the same disease during the survey in the same wave. IW_{t-1} (Informal/no care worked) and FW_{t-1} (formal worked) are dummies equal 1 if agent visited an informal/formal provider and he is not sick anymore after the visit. In equation 1, the omitted category is represented by all the individuals who consulted formal doctors at time $t-1$ and are still sick during the survey at time $t-1$ (FDW_{t-1}). The paper aims to understand whether remaining sick after a treatment represents a good incentive to shift type of therapy. According to equation (1), the probability to look for formal treatment increases when $\beta_1 > 0$, suggesting that agents will switch from informal medical care to formal treatments, after a negative outcome (still sick) coming from the informal sector. In this case, a patient has consulted an informal provider in $t-1$, he has ascertained that it failed to treat his disease and therefore he updates his believes regarding the best quality caregiver. This updating will be translated in a different health provider's choice in the next wave. Agents confirm their health seeking choice whether $\beta_1 < 0$: they don't switch caregiver, even after the failure to treat diseases by the previous health provider consulted. X_{it} is a vector of individual and household characteristics, including age, sex, years of education, head of household, head's religion, household size and wealth, proxy as physical stock held by the family, living conditions (type of housing and access to safe water) and the dependency ratio (number of elderly and kids over household size). Notice that among the controls is the average distance between the household and the caregivers (formal or informal).[12] C_{it} is the children dummy equal 1 if the respondent is a kid younger than 12 years old, d_t and d_d are respectively time and district dummies and ε_{it} is the error term. Fixed effects at district level capture time-invariant features of the health market supply, such as the number of health facilities per district. Time t is the wave in which the survey has been done with an average interval between waves of 6-7 months.

The main concern with this empirical strategy is that the health status after the treatment could not necessarily be the result of success or failure of the previous treatment received. This could emerge for two reasons. First, patients might potentially visit many health care providers during their illnesses. The survey investigates the number of providers consulted since the beginning of the illness before checking respondent's health status at the time of the survey. Therefore, I estimate equation 1 also on the sample of those who only had received one medical advice for the same illness episode. Second, as explained in section 3, the health status could also depend on the type of symptoms and diseases suffered. In the empirical section, I restrict the estimates to non cyclic disease, routine and not routine symptoms and to predicted disease. Equation 1 will be estimated using a probit model.

2.4.2 Identification strategy

The main problem in identifying the probability of choosing a formal health provider is related to sample selection bias. The Kagera Health Development Survey reported information related to health care decisions only conditional to previous illnesses, hence equation 1 can be estimated only for sick agents. There is a problem of selection whether some unobservable variables influence both the probability of illness and the utilization of health services (i.e. health endowment, health habits, etc.).

In cross-section datasets there have been attempts to solve this issue with a two-steps

[12] The distance is self-reported. I calculate the average distance among household members who sought informal care and among those who sought formal care.

method such as the one proposed by Heckman (1979). The econometric literature does not yet provide a single solution to deal with sample selection bias in panel datasets. The first approach proposed is to control for unobserved time-invariant individual characteristics. Heckman and Macurdy (1980) argued that a fixed effect tobit model could be a solution for sample selection bias. Nijman and Verbeek (1992) consider a random effect model under the assumption of normality and serial independence of the idiosyncratic errors in both the selection and the main equation. Vella (1998) provides a review and additional references.

I control for sample selection bias following the procedure introduced by Wooldridge (1995, 2002). This method requires a standard probit or tobit model in the selection equation for each time period followed by a multivariate linear regression in the second stage equation. The Mills ratio coming from the first stage estimates are included as controls in the second stage linear estimate. I estimate the first stage equation with a probit model, for each time period t (four waves) as follows:

$$\Pr(ill)_{it} = 1[\beta_0 + \beta_1 Disaster_{ht} + \beta_2 X_{it} + \beta_3 \bar{X_{it}} + C_{it} + d_t + \varepsilon_{it} > 0] \tag{2.2}$$

where $\Pr(ill)_{it}$ is the binary selection indicator equals to1 for agents reporting an illness or an injury in the month prior the survey for individual i at time t. $Disaster_{ht}$ is a dummy variable equal 1 if there was a shock in the community h in the six months prior to the survey, such as drought, epidemic, insect and crop diseases.[13] This variable is the exclusion restriction in the first stage equation correlated with the probability of illness, but not with the probability to seek formal care (equation 1). Due to data limitations, we don't have an exclusion restriction at individual level: $Disaster$ is constant within each cluster in the sample. Anyway, there are many reasons to think at it as a good exclusion restriction. In Tanzania, climatic conditions and unfavorable natural shocks, not only have a huge impact on agriculture (which accounts for almost half of GDP and employs 80% of the work force[14]) limiting cultivated crops, but they also influence people's health status, by worsening the already precarious living conditions. X_{it} represents individual and household controls as specified in equation 1, $\bar{X_i}$ is the average of

[13] The potential type of disasters mentioned in the questionnaire were: flood, drought, epidemic, insect, war and crop diseases. I didn't consider flood and war because of their potential endogeneity in supply side of health market (i.e. impact on the health facilities).

[14] Central Intelligence Agency 2007.

each independent variables across time for the individual i[15],C_{it} is the children dummy equal 1 if the respondent is a kid younger than 12 years old, d_t represents four years time dummies and ε_{it} is the error term. The errors in the selection equation are assumed to be normally distributed, but to display arbitrary serial correlation and unconditional heteroskedasticity. I found reasonable not to include in $\bar{X}_i$ the average of the binary variables, because they present a good stability over time, and therefore, to decrease the risk of multicollinearity between the average and the single binary regressor.

The resulting estimates of the first stage equation are used to obtain the inverse Mills ratios, $\hat{\lambda}_{it}$ for all t and i.[16] Following Wooldridge (1995), I estimate the probability of seeking formal medical care Y_{it} by pooled OLS on X_{it}, $\bar{X}_i$, $\hat{\lambda}_{it}$ and $d_t{*}\hat{\lambda}_{it}$ for those observations for which $\Pr(ill)_{it} = 1$. Finally, the asymptotic variance of the estimated coefficients has been corrected by bootstrapping the standard errors.

To my knowledge this is the first attempt to solve the selection bias problem in panel data by empirically applying the Wooldridge's (1995) method.

2.5 Econometric results

2.5.1 Switching to formal health care

Table 7 reports estimated and marginal probit coefficients for a model where the dependent variable is equal 1 if the individual opts for formal establishments when sick.

[Insert table 7]

Columns 1 and 2 test the determinants of the probability to seek formal health care (hospitals, clinics, dispensaries, health centres), conditional on being ill in the four weeks prior to the survey. The distance between the household and the formal establishments shows a negative and significant coefficient: ceteris paribus one more kilometer of distance decreases the probability to seek formal care by 0.2 percentage points. On the contrary, the dependent

[15] This is best viewed as the Mundlak approach (1978). A Chamberlain approach (1980) would replace $(X_{it}, \bar{X}_i)$ with X_i.

[16] The Mills ratio is $\hat{\lambda} = \phi(x\delta)/\Phi(x\delta)$, where $x\delta$ are the residuals of the equation 1.

variable increases with the number of years of education and the coefficient is significant at 5 percent level, suggesting the relevant role played by education in demanding higher standards of health care. Poorer households have a lower probability to be treated in hospitals or clinics: the dependent variable decreases with inadequate houses and with access only to unsafe water. The magnitude of the latter effect is substantial: conditional on illness, ceteris paribus, having access only to unsafe water (rain, river and lake) as the main source of drinking water reduces the likelihood to consult formal caregivers by 9 percentage points. As expected, household size decreases the dependent variable. Bad water, bad house and household size also capture a wealth effect, showing that, on average, wealthier households can afford better health services.

Columns 3 to 14 report probit estimates by conditioning the choice of a formal or informal medical care in time t to the outcome coming from the previous caregivers consulted. I test whether respondents switch health seeking behavior over time. The table reports only the coefficients of interest. In columns 3 and 4 I consider the full sample of individuals. Remember that the variable INFORMAL/NO CARE DIDN'T WORK$_{t-1}$ is equal 1 if the patient was ill four week prior to the survey, he consulted an informal health care provider (or he didn't receive a therapy) and he is still sick at the time of the interview. INFORMAL/NO CARE DIDN'T WORK$_{t-1}$ shows a negative and significant coefficient at 1 percent level. Looking at the magnitude of the coefficient we note that, ceteris paribus, having visited informal providers and being still sick reduces the probability to have formal care in the following year by 11.4 percentage points. Patients don't switch from informal to formal provider, even after the failure in treating their disease by the previous informal caregiver consulted. This result suggests lack of learning from own past experience in health seeking behavior. Moreover, the dependent variable decreases by 13.1 percentage points also when a patient has consulted an informal provider and he is not sick anymore (Informal worked $t-1$). These results confirm that remaining sick after a medical consultation isn't a good incentive to try alternative practitioners. The health status after a therapy doesn't count as a crucial factor in the health seeking choice in the next years. To isolate the effect of informal care from the one of not having care in t-1 without loosing observations, I include in all the estimates, a dummy variable equal 1 if the individual didn't consult anyone for his disease in t-1. The probability to seek formal care at time t decreases with the average distance between the household and the formal establishment (controlling also for the distance to informal practitioners) and the coefficient is statistically significant at 5 percent level. Compared to the previous specification, the head of household dummy is positive and significant, showing a higher probability to afford better services for the head of the family.

An individual's choice among different caregivers' options depends on the seriousness and on the type of symptom he is suffering from. Columns 5 to 10 report the findings by splitting the sample between routine (less serious), not routine (more serious) and not cyclic symptoms.[17] With all three restrictions the sign of the coefficient linked to INFORMAL/NO CARE DIDN'T WORK_{t-1} is still negative and significant at 1% level, confirming the lack of learning. Columns 7 and 8 show an interesting result: for serious diseases, distance doesn't influence the probability to consult formal providers, while it is still statistically significant at 1 percent level for routine and not cyclic symptoms.

In columns 11 and 12 I include predicted diseases as additional controls. The dependent variable decreases with INFORMAL/NO CARE DIDN'T WORK_{t-1} and patients suffering from malaria have a higher likelihood to look for formal health care, suggesting greater propensity to consult formal provider for serious disease. Holding other controls at the sample mean, the probability to get a formal care increases by 8.9 percentage points for agents at whom malaria has been inferred.

An explanation for the results so far could be driven by the fact that patients have consulted informal practitioners as first choice and then, they have sought other medical treatments within the same month. In this case, the concern would be that the outcome from the therapy is not the consequence of the first health care sought. To deal with this problem, I further restrict the sample to patients who only had one consultation for the same disease. Columns 13 and 14 show the estimates: once again β_1 remains negative and significant at 5 percent level.

2.5.2 Natural disasters and the probability of illness

To test the robustness of the results I control for sample selection bias, using as exclusion restriction a dummy variable equal 1 if a disaster happened in the community in the past six months, such as, drought, epidemic, insects and crop disease. By solving the selection bias

[17] Note that in this case it is relevant to consider patients suffering from not cyclic disease in time t-1, to capture the type of symptom after the therapy. Cyclic symptoms appear from time to time and the agent's health status would not be the consequence of a treatment, but rather by the type of disease.

problem, I demonstrate that the results so far are not driven by the selected sample of ill individuals.

[Insert table 8]

Columns 1 and 2 of table 8 report the first stage regression, using as a dependent variable the probability of illness in the past month. Individuals who live in a community hit by natural disasters have a higher probability to be sick in the following months. The coefficient is statistically significant at 1% level, suggesting that disaster could represent a good exclusion restriction in this context. Women have a higher probability of reporting illnesses in the four week proceeding the survey compared with men: being a woman increases the likelihood of illness by 4.4 percentage points. Education decreases the probability of illness with a coefficient statistically significant at 5 percent level. The head of household dummy has instead a positive effect on the likelihood to report an illness in the month prior to the survey. The head of a household is generally the one who has more opportunities to travel away from the family, increasing the risk of contracting diseases. As proxy of income I use an alternative wealth indicator: the amount of physical stocks held by the household.[18] Even controlling for district fixed effects, I show that the dependent variable decreases with the amount of physical assets held by the family, showing that richer individuals have a less probability to contract a disease. Finally, I investigate the relationship between diseases and household's characteristics. Extreme living conditions, such as a dwelling composed by wall of mud or bamboo, by earth as main floor and by roof made by grass or mud are associated with a higher probability of illness. To assess the magnitude of this coefficient, note that, ceteris paribus, having an inadequate house and drinking dirty water increase the probability of illness by 2.6 and 1.9 percentage points, respectively.

Columns 3 to 8 provide linear estimates on the probability to consult a formal caregivers controlling for sample selection bias and bootstrapping standard errors with 200 replications. With all the types of sample restrictions (routine, not routine, not cyclic symptoms and only one health visit) getting an informal health consultation and remaining sick has a negative

[18] The physical stock considered are land, business, equipment, livestock, and dwelling.

and significant effect on the probability to seek formal health care in the next period. In the baseline specification, ceteris paribus the likelihood to switch health provider decreases by 11 percent with INFORMAL/NO CARE DIDN'T WORK$_{t-1}$. Patients don't update their believes on health seeking behavior after a negative outcome, once again revealing lack of learning. By controlling estimate for sample selection bias, I confirm the results reported in table 7, with a relevant role of distance, household size and access only to dirty water in the reducing the likelihood to seek formal health care.

As a robustness test for the pertinence of the exclusion restriction, I estimate equation 1 by including "disaster" as additional regressor. Table A2 in the appendix shows that a disaster happened in the community doesn't have any impact on the choice of formal health care.

2.6 Concluding remarks

The main contribution of the paper is to add a dynamic perspective in analyzing individual's health seeking behavior. Overall, the empirical evidence suggests that agents are biased towards one type of health care and they don't switch caregivers even if the treatment has failed to heal them. Patients behave without taking into account the private information on their health status. The paper also investigates how the choice between a formal health care provider (hospitals, health centres, dispensaries, clinics) and an informal one (pharmacy, practitioner's home, family homes, self-care or no-care) changes with individual and household characteristics. The main cost associated to the formal health sector is the distance between the household and the facility, while education positively affects the likelihood to have formal therapies. All the estimates are controlled for types of symptoms and diseases. By applying Wooldridge's procedure (1995) and by exploiting disasters which happened in the community as exclusion restriction (correlated with the probability of illness but not with the likelihood to have formal treatments), I attempt to solve the sample selection bias problem. The findings of the paper are not driven by the selected sample of sick individuals.

These results shed light on a relevant problem in Tanzania and they have important implications from a policy prospective. The individuals' health conditions are not only driven by the generally inefficient supply side of the health market, but even from the interesting structure of patient's demand. The success of a therapy is not an important factor involved in the choice of a specific caregiver. This feature fosters the existence of low quality and not qualified doctors and health services. The first step to enforce the demand for formal care is to promote education and to disseminate informative campaigns to overcome cultural bias towards informal caregivers. Second, a more capillary distribution of government-run health services is necessary. This is a very costly and long term solution, anyway. An alternative response could be the promotion of groups of official doctors in charge of visiting sick poor households in rural area from time to time.

In general this paper underlines the need for a better understanding of the demand for health services. Future research should explore more deeply the world of informal health institutions, by gathering data on estimated number of traditional healers and user fees.

2.7 References

Acton, J. P., *Nonmonetary Factors in the Demand for Medical Services: some empirical evidence*, Journal of Political Economy V. 83, n. 3, pp. 595-641, June 1975.

Alderman, H., Gertler, P., *Family Resources and Gender Differences in Human Capital Investments: The demand for Children's Medical Care in Pakistan*, Intrahousehold resource allocation in developing countries: Models, methods, and policy, pp. 231-48, Johns Hopkins University Press for the International Food Policy Research Institute, 1997.

Althaus, F. A., *Work in progress: The expansion of Access to Abortion services in South Africa following Legalization*, International Family Planning Perspectives, Vol.26, No. 2, pp.84-86, June 2000.

Banks, J. S., Sundaram, K. R., *Switching cost and the Gittins index,* Econometrica, Vol. 62, No. 3, pp. 687-694, May 1994.

Beegle K., De Weerdt J., Dercon S., *Kagera Health and Development Survey (KHDS), 2004, Basic Information Document*, mimeo, The World Bank, 2006b.

Björkman M., Svensson J. *Power to the People: Evidence from a Randomized Experiment of a Community Based Monitoring Project in Uganda*, CEPR Discussion Paper No.6344, 2007

Börgers T., *On the relevance of evolution and learning to economic theory*, Economic Journal 106, 1274-1385, 1996

Case A., Menendez A., Ardington C., *Health Seeking Behavior in Northern KwaZulu-Natal*, Princeton University working paper n.146, 2005.

Central Intelligence Agency (CIA), *The World Fact Book*, Tanzania, 2007, https://www.cia.gov/library/pub world-factbook/index.html

Conley, G. T., Udry C. R., *Learning About a New Technology: Pineapple in Ghana*, Yale Economic Growth Center Discussion Paper N. 817, July 2000.

Cretin, S., Keeler, E. B., *Uses of Cost -benefit Analysis: Editorial*, Journal of Health Economics v6, n3, pp. 275-278, September 1987.

Das, J., Hammer J., *Strained Mercy: The Quality of medical care in Delhi*, Policy Research Working Paper Series: 3228, The World Bank, 2004.

Das, J., Hammer, J., *Which Doctor? Combining Vignettes and Item-Response to Measure Doctor Quality*, Journal of Development Economics, 78, 348-383, 2005

Dasgupta, P., Gupta, I., *Demand for curative health care in rural India: Choosing between Private, Public and No care*, Working paper series N. 82, National Council of Applied Economic Research, January 2002.

Dupas, P. *The impact of Conditional In-Kind Subsidies on Preventive Health Behaviors: Evidence from Western Kenya*, EHESS-PSE, Paris, July 11, 2005.

Economic Development Initiative, *Kagera Rural Core Welfare Indicator Questionnaire (CWIQ), Baseline survey on poverty, Welfare and Services in Kagera Rural District,* April 2004.

Foster, A. D., Rosenzweig, M. R., *Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture*, Journal of Political Economy v103, n6 pp. 1176-1209, December 1995.

Fauci A.S, Braunwald E., Kasper D. L., Hauser L. S., Longo L. D., Jameson L., Loscalzo J. *Harrison's Principles of Internal Medicine*, 17th Edition, McGraw-Hill, June 2008.

Gertler, P., Van Der Gaag, J., *The willingness to Pay for Medical Care*, John Hopkins University Press for the World Bank, 1990.

Gittins, J. C., Jones, M. D., *A dynamic allocation index for the discounted multiarmed bandit problem.* Biometrika 66, pp. 561-565. 430, 1976.

Green, M., *Public reform and the privatization of poverty: some institutional determinants of health seeking behavior in southern Tanzania*, Culture, Medicine and Psychiatry 24, pp. 403-430, 2000.

Greene, W., *Estimating Econometric Models with Fixed Effects*, New York University, Leonard N. Stern School of Business, 2001.

Heckman, J., *Sample Selection Bias as a Specification Error*, Econometrica Vol.4, No. 1., pp. 153-162, Jan., 1979.

Heckman, J., Macurdy, T. E., *A Life Cycle Model of Female Labour Supply*, The Review of Economic Studies, Vol. 47, No 1, Ecometrics Issue, pp. 47-74, Jan, 1980.

Kasper, D., Braunwald E., Hauser, S., Longo, D., Jameson, J. L., Fauci, A., Harrison's Principles of Internal Medicine, The McGraw Hill, 16th Edition, 2006.

Kyriazidou, E., *Estimation of a Panel Data Sample Selection Model*, Econometrica, Vol. 65, No 6, pp 1335-1364, November 1997.

Leonard K., *Learing in health care: evidence of learning about clinician quality in Tanzania,* Economic Development and Cultural Change, 55 (3) pp.531-555, 2007.

Leonard K., Mliga G. M., Mariam D. H, *Bypassing Health Centres in Tanzania: Revealed Preferences for Observables and Unobservable Quality*, Journal of African Economies v11, n4, pp. 441-71, December 2002.

Leonard K., Masatu M. C., Vialou, A., *Getting Doctors to do their best: Ability, Altruism and Incetives*, Journal of Human Resources Vol.42 (3) 2007, pp. 682-700, 2005.

Lindelow, M., *The Utilization of Curative Healthcare in Mozambique: Does Income Matter?*, Journal of African Economics, Volume 14, Number 3, pp. 432-482, 2005.

Luke, N., Munshi, K., *Social Affiliation and the Demand for Health Services: Caste and Child Health in South India*, Journal of Development Economics 83(2):256-279, 2007.

Ministry of Health and Social Welfare, The United Republic of Tanzania, National website, www.tanzania.go.tz., 2007.

Munshi, K., *Social Learning in a heterogenous population: technology diffusion in the Indian Green Revolution*, Journal of Development Economics 73, 185-213, 2004.

National Bureau of Statistics and Macro International, *Tanzania Service Provision Assessment Survey 2006*, Dar es Salaam, Tanzania 2006.

Nijman, T, Verbeek, M., *Testing for Selectivity Bias in Panel Data Models*, International Economic Review, Vol. 33, No. 3, pp. 681-703, August 1992.

Rabin M., Schrag J. L., *First Impressions Matter: A Model of Confirmatory Bias,* The Quarterly Journal of Economics, Vol. 114, No. 1, pp. 37-82, 1999

Sahn D. E., Youger, S. D., Genicot, G., *The Demand for Health Care Services in Rural Tanzania*, Oxford Bulletin of Economics and Statistics, 65, 2, 0305-9049, 2003.

Semykina A., Wooldridge J., *Estimating panel data models in the presence of endogeneity and selection: theory and application*, working paper, June 2005.

Tanzania Government and ORC Macro, "Tanzania HIV/AIDS Indicator Survey 2003-04", Tanzania Commission for AIDS and National Bureau of Statistics, Dar es Salaam, Tanzania, 2005.

Tanzania Government and ORC Macro, "Tanzania Demographic and Health Survey 2004-2005", National Bureau of Statistics, Dar es Salaam, Tanzania, 2005.

The World Bank, "User's Guide to the Kagera Health and Development Survey Datasets", mimeo, 2004.

USAID, Report to Congress: *Health-Related research and development activities at USAID*, June 2006.

Vella, F., *Estimating Models with Sample Selection Bias: A survey*, The Journal of Human Resources XXXIII, 1998.

WHO and European Observatory on Health Systems and Policies Series, *Purchasing to improve health system performance*, Edited by Josep Figueras, Ray Robinson, Elke Jakubowski, Open University Press, 2005.

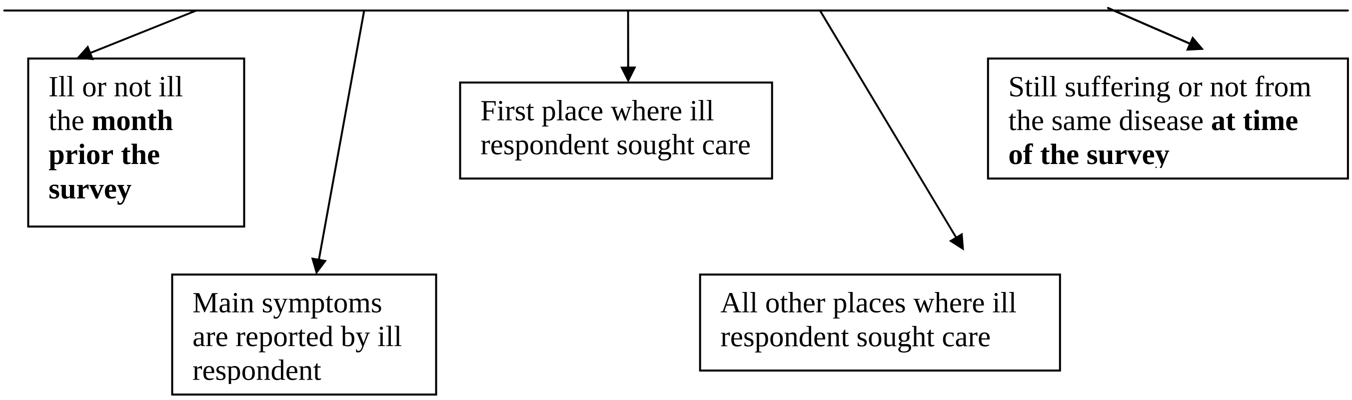
Wooldridge, J. M., *Selection corrections for panel data models under conditional mean independence assumptions*, Journal of Econometrics 68 pp.115-132, 1995.

Wooldridge, J. M., *Econometric analysis of cross section and panel data*, Massachusetts Institute of Technology, 2002.

Figure 1: Location of the KHDS clusters in Kagera Region, Tanzania

Source: The World Bank, 2004

Figure 2: Timing of the questions in the KHDS 1994-1997



Source: The World Bank, 2004

Tables

Table 1: Percent of individual reporting illness and injury in the four weeks prior to survey, by age and gender

Age	Total	Male	Female
0-2	63.63	64.02	63.17
3-5	48.48	47.96	48.99
6-15	40.59	40.52	40.67
16-30	45.87	42.87	48.61
31-45	55.73	51.27	58.92
>45	68.88	66.12	71.14
Total	51.24	49.77	52.60

Source: author's calculation on KHDS

Table 2: Symptoms of those reporting illness or injury in month prior to survey

Symptom	*%*	*Symptom*	*%*
Routine	88.21	**Not Routine**	11.79
Acute Diarrhea	3.7	Recurring fever	2.25
Chronic Diarrhea	0.11	Productive Cought	1.84
Weight Loss	0.05	Coughing Blood	0.08
Vomiting	0.85	Difficulty breathing	0.38
Acute Fever	23.55	Abdominal pain	6.51
Sever Headache	10.89	Pain on passing urine	0.02
Chills	8.21	Genital Sores	0.02
Cough	12.95	Burn	0.21
Sore throat	0.15	Fracture	0.19
Wound	4.06	Child birth	0.17
Weakness	1.68	Mental disorder	0.03
Skin Rash	1.22	Fainting	0.09
Other	20.79		

Source: author's calculation on KHDS.

Table 3: Symptoms reported associated with most widespread diseases in Kagera

	Diseases													
Symptoms	Malaria	Acute Respiratory infection	Helmintiasis	Skin Disease	AIDS	TB	Trauma	Veneral Disease	Child Birth	Diarroheal diseases	Mental illnesses	Urinary Tract Infection	Migraine	Burn
Acute Diarrhea			2			x				1				
Chronic Diarrhea					2									
Weight Loss					1	2	x					2		
Acute Fever	1	1				x								
Recurr. Fever					x		x	2						
Skin Rash				1	x									
Weakness	x					x	x						x	
Headache													1	
Fainting							x							
Chills	2					x								
Vomiting										2			2	
Cough		2												
Prod. Cough		x												
Coughing blood						1								
Urine Pain												1		
Genitals sores								1						
Mental Disorders	x										1			
Abdominal pain			1		x	x								
Sore Throat		x												
Breathing		x												
Burn														1
Fracture							2							
Wound							1							
Child Birth									1					

Note: 1=Main symptom in the disease, 2=second important symptom in the disease, X=additional symptoms eventually present in the disease.

Source: Fauci et al. 2008.

Table 4 : Choice of health care provider on sample of ill individuals

Place	*Total*	*By gender* Male	Female	*By age* 0-2	3-5	6-15	16-30	31-45	>45
Panel A									
Look for care	31.7	33.2	30.4	40.5	27.2	28.8	32.3	32.3	32
No Care	68.3	66.8	69.6	59.5	72.8	71.2	67.7	67.7	68.0
No.Obs	8770	4119	4651	991	828	2408	1951	988	1604
Panel B									
Informal care	*13.3*	*14.1*	*12.6*	*11.9*	*13.8*	*11.4*	*13.3*	*14.7*	*15.8*
Pharmacy	1.3	1.2	1.3	1	0.4	1.3	1.0	2.5	1.4
Practitioner	6.7	7.7	5.7	6.7	8.0	4.8	6.4	7.8	8.6
Own home	4.3	4.4	4.2	3.5	4.4	3.9	4.8	3.5	5.5
Other	1	0.7	1.3	0.7	0.9	1.4	1.3	0.9	0.4
No.Obs	370	178	192	48	31	79	84	47	81
Formal Care	*86.7*	*85.9*	*87.4*	*88.0*	*86.2*	*88.6*	*86.7*	*85.3*	*84.2*
Hospital	25.2	22.9	27.4	23.2	19.6	21.9	28.6	29.8	26.6
Health Centre	10.4	11.1	9.7	12.5	7.6	10.4	9.4	11.6	10.6
Dispensary	50	51.1	49	49.4	57.8	55.2	48.5	43.0	46.5
Clinic	1.1	0.8	1.3	3.0	1.3	1.2	0.2	0.9	0.6
No. Obs.	2410	1236	1174	353	194	615	545	272	431

Source: author's calculation on KHDS

Table 5: Switching between formal care and informal/no care from wave t-1 to wave t for those still ill after a therapy

Wave t	*Wave t-1* Informal/No Care		Formal Care		Total
	Obs.	%	Obs.	%	Obs
Informal/No Care	815	80	224	58	1039
Formal Care	198	20	160	42	358
Total	1013	100	384	100	1397

Source: author's calculation on KHDS

Table 6: Benefits and Costs of formal health care

Sample	*Formal* μ_F	*Informal* μ_I	*P-value* μ_F-μ_I=0	*Obs.*	*No care* μ_{NC}	*P-value* μ_F-μ_{NC}=0	*Obs.*
Fraction of ill individuals in the months prior the survey, who are still ill during the survey							
Pooled Sample	0.40	0.47	0.01	2781	0.42	0.13	8401
Routine Symp.	0.39	0.45	0.03	2349	0.42	0.01	7431
Not Routine Symp.	0.47	0.54	0.24	432	0.40	0.04	970
Fraction of ill individuals, treated by a doctor							
Pooled Sample	0.19	0.04	0.00	2661	-	-	-
Distance (Km) between the facility and the household [(a)]							
Pooled Sample	4.9	2	0.00	2782	-	-	-

Source: author's calculation on KHDS

[(a)] The distance from health facility to household is equal 0 if agents seek care in their own homes.

Table 7: Probability of care by formal providers if the informal sector didn't work

Dependent variable=1 if an individual seek formal care

	Baseline				*Type of symptoms*						*Type of diseases*		*Single visit (t-1)*	
					Routine		*Not routine*		*Not cyclic symp. (t-1)*					
	Coeff.	*Marg.* [(a)]	*Coeff.*	*Marg.*	*Coeff.*	*Marg.*	*Coeff.*	*Marg.*	*Coeff.*	*Marg.*	*Coeff.*	*Marg.*	*Coeff.*	*Marg.*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Infomal/no care didn't work $_{(t-1)}$			-0.352**	-0.115**	-0.306**	-0.099**	-0.877*	-0.297**	-0.334**	-0.108**	-0.338*	-0.121*	-0.470**	-0.165**
			(0.147)	(0.045)	(0.156)	(0.048)	(0.454)	(0.134)	(0.168)	(0.052)	(0.196)	(0.067)	(0.211)	(0.065)
Formal worked $_{(t-1)}$			-0.011	-0.004	-0.021	-0.007	0.070	0.026	-0.050	-0.017	-0.041	-0.015	0.006	0.002
			(0.085)	(0.029)	(0.091)	(0.030)	(0.267)	(0.100)	(0.099)	(0.033)	(0.110)	(0.040)	(0.094)	(0.036)
Informal/No Care worked $_{(t-1)}$			-0.400***	-0.132***	-0.375**	-0.122**	-0.743	-0.257*	-0.400**	-0.130**	-0.344*	-0.124*	-0.150	-0.056
			(0.147)	(0.046)	(0.155)	(0.048)	(0.473)	(0.149)	(0.168)	(0.052)	(0.195)	(0.068)	(0.199)	(0.073)
Distance to Formal	-0.007***	-0.002***	-0.019***	-0.006***	-0.019**	-0.007***	-0.006	-0.002	-0.023***	-0.008***	-0.014*	-0.005*	-0.018**	-0.007**
	(0.003)	(0.001)	(0.005)	(0.002)	(0.006)	(0.002)	(0.020)	(0.007)	(0.007)	(0.002)	(0.007)	(0.003)	(0.009)	(0.003)
Education	0.022**	0.008**	0.006	0.002	0.007	0.002	-0.004	-0.001	0.017	0.006	0.014	0.005	0.027	0.010
	(0.009)	(0.003)	(0.013)	(0.004)	(0.013)	(0.005)	(0.041)	(0.015)	(0.015)	(0.005)	(0.016)	(0.006)	(0.021)	(0.008)
HH Head	0.113	0.039	0.176*	0.062*	0.127	0.044	0.485*	0.186*	0.241**	0.084**	0.205*	0.077*	0.299*	0.117*
	(0.069)	(0.025)	(0.091)	(0.033)	(0.097)	(0.034)	(0.279)	(0.108)	(0.108)	(0.039)	(0.116)	(0.044)	(0.158)	(0.062)
Bad water	-0.252***	-0.090***	-0.271***	-0.096***	-0.252**	-0.088***	-0.517**	-0.199**	-0.256***	-0.090***	-0.361***	-0.138***	-0.399**	-0.156***
	(0.047)	(0.017)	(0.064)	(0.024)	(0.068)	(0.025)	(0.223)	(0.087)	(0.076)	(0.028)	(0.086)	(0.034)	(0.102)	(0.040)
Bad house	-0.095*	-0.032**	-0.102	-0.035	-0.092	-0.031	-0.169	-0.062	-0.135*	-0.045*	-0.126	-0.046	-0.098	-0.037
	(0.049)	(0.016)	(0.067)	(0.022)	(0.072)	(0.023)	(0.199)	(0.072)	(0.078)	(0.025)	(0.086)	(0.031)	(0.116)	(0.044)
HH size	-0.010*	-0.004*	-0.004	-0.001	-0.003	-0.001	-0.008	-0.003	-0.010	-0.004	0.005	0.002	-0.014	-0.005
	(0.005)	(0.002)	(0.007)	(0.003)	(0.008)	(0.003)	(0.023)	(0.009)	(0.009)	(0.003)	(0.010)	(0.004)	(0.013)	(0.005)
Predicted Malaria											0.230*	0.087*		
											(0.118)	(0.046)		
Observations	7384		2795		2477		318		1476		1570		955	
Wald chi2	277.87		164.85		153.1		40.13		121.7		119.2		59.54	
Prob > chi sq.	0.00		0.00		0.00		0.07		0.00		0.00		0.00	

Notes: * denotes significance at the 10 percent level; ** at the 5 percent level; *** at 1 percent level. Probit Estimates. Standard errors in parenthesis and corrected for heteroskedasticity and clustering of the residuals at the household level.

(a) Marginal probit coefficients are calculated at the mean. For dummies, marginal effect is calculated for discrete change from 0 to 1.

Estimates controlled for not having care in t-1, distance to infomal care, age, age sq., gender, head's religion, log of physical stock, household size, dependency ratio, wave dummies and district dummies.

.

Table 8: Probability of illness and of formal care, controlling for sample selection

Dependent variable=	1 if ill		1 if Formal Care					
			Baseline	*Type of symptom*			*Type of disease*	*Single visit $_{(t-1)}$*
	Coeff.	*Marg.* [(a)]		*Routine*	*Not Routine*	*Not cyclic $_{(t-1)}$*		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Disaster [(b)]	0.072***	0.029***						
	(0.026)	(0.010)						
Informal/no care didn't work $_{(t-1)}$			-0.118**	-0.102**	-0.287*	-0.115**	-0.122*	-0.146**
			(0.048)	(0.051)	(0.157)	(0.056)	(0.070)	(0.069)
Formal worked $_{(t-1)}$			-0.011	-0.012	-0.028	-0.022	-0.027	-0.002
			(0.033)	(0.035)	(0.095)	(0.041)	(0.042)	(0.038)
Informal/No Care worked $_{(t-1)}$			-0.136***	-0.122**	-0.264	-0.136**	-0.129*	-0.049
			(0.047)	(0.052)	(0.176)	(0.058)	(0.070)	(0.073)
Distance to Formal			-0.004***	-0.004***	-0.001	-0.005***	-0.004**	-0.004*
			(0.001)	(0.001)	(0.006)	(0.001)	(0.002)	(0.002)
Female	0.111***	0.044***	-0.052*	-0.053*	0.025	-0.039	-0.075*	-0.041
	(0.026)	(0.010)	(0.027)	(0.032)	(0.098)	(0.037)	(0.045)	(0.056)
Education	-0.027**	-0.011**	0.019	0.019	0.005	0.042	0.041	0.024
	(0.012)	(0.005)	(0.018)	(0.019)	(0.071)	(0.022)	(0.026)	(0.037)
HH Head	0.175***	0.069***	0.082	0.083	-0.043	0.074	0.056	0.240
	(0.045)	(0.018)	(0.072)	(0.083)	(0.253)	(0.090)	(0.108)	(0.160)
HH Size	-0.003	-0.001	-0.122***	-0.140***	0.034	-0.010	-0.004	-0.023
	(0.010)	(0.004)	(0.037)	(0.041)	(0.050)	(0.017)	(0.023)	(0.030)
Bad water	0.028	0.011	-0.142***	-0.129***	-0.213**	-0.133***	-0.195***	-0.230***
	(0.031)	(0.012)	(0.028)	(0.031)	(0.084)	(0.034)	(0.038)	(0.049)
Bad house	0.055*	0.022*	-0.029	-0.021	-0.044	-0.048	-0.050	-0.003
	(0.030)	(0.012)	(0.029)	(0.029)	(0.104)	(0.031)	(0.038)	(0.063)
Predicted Malaria							0.103**	
							(0.047)	
Observations	18340		2687	2380	307	1943	1505	900
R sq.			0.0877	0.0883	0.3983	0.0938	0.15	0.174
Pseudo R sq.	0.0449							

Note: standard errors in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Columns 1 and 2 show probit estimates pooling the four waves together, altough to costruct the Mills ratios I separately estimate the probability of illness in each wave. Columns 3 to 8 report linear estimates.

Estimates controlled for not having care in t-1, distance to infomal care, age, age sq., log of physical stock, head's religion, dependency ratio, wave and district dummies

(a) Marginal probit coefficients are calculated at the mean. For dummies, marginal effect is calculated for discrete change from 0 to 1.

(b) Disaster includes drought, epidemic, insect and crop diseases.

Appendix Tables

Table A1: Definition of variables and descriptive statistics

Variable	Definition	Obs	Mean	S.D.
Informal/No care didn't work $_{(t-1)}$	Equal 1 if an ill individual in a month prior the survey consulted an informal provider at wave t-1 and he is still ill during the survey at wave t-1	2795	0.293	0.46
Formal didn't work $_{(t-1)}$	Equal 1 if an ill individual in a month prior the survey consulted an informal provider at wave t-1 and he is still ill during the survey at wave t-1	2795	0.211	0.41
Informal worked $_{(t-1)}$	Equal 1 if an ill individual in a month prior the survey consulted an informal provider at wave t-1 and he isn't ill during the survey at wave t-1	2795	0.359	0.48
No care $_{(t-1)}$	Equal 1 if an individual didn't have health care at wave t-1	2797	0.61	0.49
Formal$_{(t)}$	Equal 1 if formal provider (hospital, health centre dispensary clinic), 0 if informal (pharmacy, patient's home, other, self care) at wave t	7384	0.324	0.47
Distance to formal care	Average distance (km) between the household and the formal health establishments visited by HH's members	7384	5.509	7.72
Distance to informal care	Average distance (km) between the household and the informal health establishments visited by HH's members	7384	0.12	0.7
Ill	Equal to 1 if individual reported ill or injury and 0 if not, in a month prior to survey	19009	0.512	0.5
Female	Equal 1 if female, 0 otherwise	19009	0.521	0.5
Age	Age in years	19009	22.47	20.2
Education	Years of education	19009	2.714	3.11
HH head education	Years of education of the HH head	19009	4.952	3.22
HH head	Equal 1 if head of household, 0 otherwise	19009	0.173	0.38
HH Catholic	Equal to 1 if the HH Head is catholic and 0 if he is muslim, protestant, other christian, traditional and other religions	19009	0.594	0.49
HH size	Number of member in each hh	19009	8.091	4.11
Log Phys. Stock	Log value of physical asset	19009	13.1	1.59
Bad water	Equal 1 if the source of drinking water for household river/lake, well without pump, rain 0 otherwise	19009	0.8	0.4
Bad house	Equal 1 if inadequate house (bad wall bad floor bad roof), 0 otherwise	19009	0.253	0.43
Dependecy Ratio	No. of elderry and kids in the household over HH size	19009	0.479	0.2
Children	Equal 1 if children younger than 12 years old	19009	0.396	0.49
Disaster	Equal 1 if there was a disaster in the past 6 months in the community, such as: drought, epidemic, insect, crop diseases.	18340	0.56	0.5

Source: author's calculation on KHDS

Table A2: Impact of disaster on the probability to look for formal care

Dependent variable=1 if an individual seek formal care	
Disaster [b]	0.041
	(0.076)
Informal/no care didn't work $_{(t-1)}$	-0.118
	(0.048)**
Formal worked $_{(t-1)}$	-0.010
	(0.032)
Informal/No Care worked $_{(t-1)}$	-0.135
	(0.047)***
Distance to Formal	-0.004
	(0.001)***
Observations	1159
R squared	0.0877

Note: standard errors in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%. Linear Estimates.
As additional controls I included age, age sq., gender, educatioon, distance to informal care household head, hh size, dep. ratio, wave and district dummies plus the average of all covariates, except for the dummies (Wooldridge 1995).

(b) Disaster includes drought, epidemic, insect and crop diseases.

Chapter 3

Getting a job but loosing health:HIV and miners in Lesotho

Abstract[1]

This paper analyzes the socioeconomic determinants of HIV infection and related sexual behaviors using the 2004 Lesotho Demographic and Health Survey. The paper points out the more vulnerable groups to HIV infection in Lesotho: miner and current married women. Being a miners increases the probability to be HIV positive by 11 percentage points. Working in mines in South Africa means spending long period away form the household in Lesotho. This creates dangerous "network effects" in extra-merital partners. Education appears to have a protective role: it is negatively associated with HIV infection and it strongly predicts preventive behaviors. The findings also show that married women who have extra-marital relationships are less likely to use a condom than non-married women. This is another important source of vulnerability that should be addressed in prevention efforts.

[1] We thank John Nkonyana, Noosi Nkone, Mahlapane Lekometsa, Mahlape Ramoseme at the Ministry of Health and Social Welfare in Lesotho as well as Charles Annor Frampong Preeti Arora, Elena Bardasi, Louise Fox, T. Mpoy-Kamulayi, Rachel Kline, Oscar Picazo, Elizabeth Sherwood and Ritva Reinikka at the World Bank for useful discussions and suggestions.

3.1 Introduction

The HIV/AIDS pandemic is the greatest challenge Lesotho is facing. Lesotho is the country with the third highest HIV prevalence rate in the world, after Swaziland and Botswana. In 2005, the estimated adult HIV prevalence was 23.2 percent, equal to 270000 HIV infected (UNAIDS, 2006). Nearly one over three Basotho adults aged 15-49 is currently infected by the virus. [2] Although, Sub-Saharan Africa has been heavily affected, HIV do not hit this region, equally. For example, HIV prevalence rate in Congo is 5.3 percent and in Kenya is 6.1 percent. The present paper explores whether historical past dynamics are correlated with the current spreading of HIV and it investigates the socioeconomic determinants of HIV/AIDS and related sexual behaviors in Lesotho.

Lesotho is a small country located in the middle of South Africa and its economy is mainly dependent on this state. HIV was first detected in Lesotho in 1986, during the apartheid regime, when Basotho's men brought HIV home from South Africa, where roughly 60 percent of the workforce was employed in the mines. Working in South Africa mines means spending long period away from the household in Lesotho. Since then, the nation has experienced a dramatic escalation in the HIV/AIDS epidemic in common with neighboring countries in southern Africa. Motivated by this anecdotal evidence the present paper investigates whether the massive percentage of Lesotho labor force employed in South Africa for a long period in the past is the main cause of HIV/AIDS health emergency in Lesotho More precisely, the paper explores if the probability of being HIV positive is higher for miners. Moreover, it also analyzes the socioeconomic determinants of HIV/AIDS and related sexual behaviors at the individual level. Understanding the determinants of HIV infection and of sexual related habits is crucial for targeting prevention.

As main finding, we show that being a miner increases the likelihood to be HIV positive by 11 percentage points and a miner currently married is about 14 percentage points more likely to be HIV positive compared with other currently married men. From a policy perspective, we shed lights on the key categories in the society that policy makers should target in Lesotho in order to prevent HIV/AIDS.

[2] People from Lesotho.

The present paper mainly uses data coming from the first Demographic and Health Survey conducted in Lesotho in 2004. The LDHS is particularly appropriate to study the determinants of HIV/AIDS for different reasons. First, because eligible women and men were asked if they consented to be tested for HIV. Further, the HIV results obtained can be merged with the socio-demographic data collected in the individual questionnaires. Compared to previous surveys including information on HIV/AIDS, the DHS has the great advantage to include a bio-marker and not to rely exclusively on self-reported behavior. Second, it includes GPS data for all households interviewed with latitude, longitude and altitude for each cluster in the sample: we exploit these data to proxy the cost of migration from Lesotho to South Africa. Finally, the data are nationally representative and the variables are defined similarly across countries, so that it is possible to perform useful international comparisons. To investigates more deeply the magnitude of the miners phenomena, we additionally collected data on the number of miners over time from TEBA a non-governmental organization and the location of South Africa mines from the South Africa Centre For Advanced Satellite and Mineral Exploration and from the South Africa Department of Minerals and Energy.

The socioeconomic profile of the HIV/AIDS epidemic in Africa has been analyzed in the epidemiological literature and, to a lesser extent, in the economics literature. Few of these studies have used nationally representative samples. Datasets which include the results of individual HIV tests are generally drawn from surveillance data taken from pregnant women attending ante-natal care clinics (Fylkesnes and others, 1997; Kilian and others, 1999) or from high risk groups (Nagot and others, 2002). A further limitation of this literature is that most of these datasets have only a limited number of socio-demographic variables and most of them cannot claim to be representative. More recently, Clark and Vencatachellum (2003) use a nationally representative sample from South Africa. Fylkesnes and others (2001) compare results from surveillance data among pregnant women with surveys based on the whole population. De Walque (2006) presents a comparison of the determinants of HIV infection and associated sexual behaviors using data coming from five Demographic and Health Surveys (Burkina Faso, Cameroon, Ghana, Kenya, Tanzania). For the whole set of countries, the author finds evidence on riskier group: people who have been in successive marriages are more likely to be infected and married women who engage in extra-marital sex are less likely to use condoms. In a following paper, he investigates HIV infection at the level of the cohabiting couple (De Walque 2007), showing room for prevention strategy among couple where only one of the partner is HIV positive. Among the literature explaining the main factors for the spread of AIDS pandemic in Africa, a part of this argues that migrants in Africa are more likely to be infected by HIV virus, such as trucker and migrants (Lurie et al, 2003). To test the hypothesis for which migrants have higher rates of risky behaviors, Oster (2007) estimates the relationship between exports and HIV infections from early 1980s to the late 1990s in nine countries of Sub-Saharan Africa. She argues a positive relationship among HIV and exports: a doubling exports leads to approximately a quadrupling of new infections. This paper also fits in the literature related to prevention campaigns through randomized evaluation, even if we use non-experimental data. Duflo, Dupas, Kremer and Sinei (2006) conducted a randomized evaluation comparing three different school-based HIV/AIDS interventions in Kenya: training teachers, encouraging students to debate the role of condoms and write essays on how to protect themselves against HIV/AIDS, and reducing the cost of education. After two years, in schools where teachers had been trained girls were more likely to be married in the event of a pregnancy and the condom debates increased practical knowledge and self-reported use of condoms without increasing self-reported sexual activity. Dupas (2006) carried on a campaign that provided Kenyan teenagers with the information that HIV prevalence was much higher among adult men than among teenage boys because of their longer exposure to the HIV virus. As a result of this campaign, she found a 65 percent decrease in the incidence of pregnancies by adult partners among teenage girls.

Finally, the paper can be included in the broad economic theory explaining individual migration decision. Starting from traditional neoclassical models (Todaro (1969), Harris and Todaro (1970)) migration is modelled as the result of a cost - benefit analysis in which before moving individuals compare the expected income differentials between the home and the receiving country.

The remainder of the paper is organized as follows. Section 2 gives a background of migrant miners' phenomena from Lesotho to South Africa. Section 3 describes the data and summary statistics. The empirical analysis and the identification strategy is discussed in section 4. Section 5 shows the results and section 6 describes the robustness analysis. Section 7 concludes.

3.2 Miners and HIV/AIDS in Lesotho

Lesotho is a small, land-locked country, located in the southern part of Africa and completely surrounded by the Republic of South Africa. The country is divided into ten administrative districts: Butha-Buthe, Leribe, Berea, Maseru, Mafeteng, Mohale's Hoek, Quthing, Qacha's Nek, Mokhotlong, and Thaba-Tseka. The population is estimated at 2.1 million with a population growth rate of 0.144% (Central Intelligence Agency, 2007). Lesotho is extremely mountainous, and only 10 percent of the land is arable; the rest is suitable only for grazing of livestock. Due to the geographic location, there is a long history of close interaction between Lesotho and South Africa. Indeed, the most important source of income for most households is wage labor in South Africa, where Basotho are employed as migrant laborers. The Lesotho Labour Force Survey conducted in1999 found that 37 percent of the people interviewed reported a family member working in South Africa, 26 percent reported a family member permanently settled there, 21 percent had sought medical care there, and 18 percent admitted to possessing South African identification documents. Further, the people of Lesotho are almost wholly Basotho, but the numbers of ethnic Basotho in South Africa are even higher.

In the past decade, a very high proportion of Basotho labor force was temporarily employed outside its borders. For most of the 20th century, as much as half of the adult male population worked on a temporary basis in South Africa, predominantly in the gold mines but also in most other economic sectors. To work in the mines for at least a year or two was a normal part of becoming an adult. In contrast, women have been prevented from legal migration for work through controls exerted by their spouses, chiefs and the colonial administration. South African sources suggest that total employment in mining peaked in 1987, at about 126000 (Sweetman 1999). At that time, these miners outnumbered the ranks of people with formal wage-earning jobs in Lesotho, a situation that persisted until the late 1990s.

With the heightened opposition to South Africa's apartheid regime and the coming of majority rule in South Africa, Lesotho's migration patterns have changed substantially. The retrenchment of Basotho men from the mines began in 1987. The numbers of Basotho miners have since declined dramatically, also as a result of mechanization and relative stagnation in gold mining, the decrease of the price of gold and the preference for South African labor. The average number of employed migrant workers was 101,262 people in 1996 and it was 52, 450 persons in 2005 (National employment Services). However, some provinces (mainly the Free State, North West , Gauteng, Mpumalanga and Limpopo) still employ, in large scale, numbers of Basotho nationals, shifting from gold mines to platinum mines. After the retrenchment there was a very limited re-employment of men. Among miners retrenched since 1991, only 33 percent were working by mid-1993 (Sweetman 1999). The generally low level of formal education among miners limited the chances of ex-miners to find a job, suggesting a predictable correlation between unemployment and the former-miners.

The mines host a vibrant sex industry and anecdotal evidence suggests that, once men are away from their families, they might be more likely to have multiple sexual partners. Once they have contracted HIV, they might infect their partners when they return to their families in Lesotho. At the same time, women, who waits for their husbands to come back from the mines have been known to engage in sexual relationships with other partners as well. This has the potential to create a dangerous "network effect" in the transmission of HIV through multiple partners.

The poor sociocultural status of women might be another reason for the spread of the disease, making more complicated the implementation of effective policies to prevent HIV/AIDS in Lesotho. According to law, a Basotho woman must obtain her husband's approval to have surgery, take contraceptives, take out a loan, run for public office, and until recently she had no power to refuse any sexual relations or insist on condom use. Women cannot own property, leaving them socioeconomically vulnerable. Many engage in other unsafe means of survival, including prostitution, early marriage, or sexual favours for older men. The country has recently passed a bill providing equal status to married women.[3] The enactment of this bill is considered a first step to remove barriers for the access to HIV prevention, treatment, care and support services for women and girls (UNAIDS 2006).

A number of institutions are currently implementing HIV/AIDS prevention programs in Lesotho. One of these is the Life Skill Curriculum, proposed by the Ministry of Education, in which HIV/AIDS education has been integrated into the primary school curriculum. It includes information on the biology of HIV/AIDS, its transmission channels and the consequences of the epidemic. It emphasizes abstinence until marriage and the importance of "Saying NO" to

[3] The Sexual Offernce Act passed in 2003.

sex before marriage. Unfortunately, the curriculum provides only a limited scope for discussing contraception or safer sex. Discussions about condoms and safe sex are not encouraged through the formal education channel.

Another initiative proposed by the Ministry of Health is the "Know Your Status" campaign, with the aim of making every person in Lesotho aware of their HIV status. Around 3,600 community health workers will be trained to do a simple HIV test that involves pricking a finger to get a drop of blood for testing. This is a very ambitious undertaking. Its full impact still needs to be assessed.

Finally, the country has been exposed to a huge distribution of condoms, especially through the Population Services International (PSI), a nonprofit organization that addresses health problem in low income countries. PSI markets three male condom brands and it also distributes low cost, high quality workplace condoms, both in rural and urban areas. Since the launch of this program, called Condom Social Marketing (CSM) in 2001, PSI has tripled the number of condoms that are sold and the number of shops where condoms can be purchased.

3.3 Data and Descriptive statistics

The data used for this study come mainly from the Lesotho Demographic and Health Survey (LDHS) 2004, collected by the Ministry of Health and Social Welfare (MOHSW) in collaboration with the Bureau of Statistics (BOS) in Lesotho and ORC Macro International. In addition, we assembled data on the number of mines in South Africa provided by the South Africa Centre For Advanced Satellite & Mineral Exploration and by the South Africa Department of Minerals and Energy. Finally, we gathered data on individuals employed in the mine sector at national level from TEBA Development.[4]

The LDHS was conducted using a representative sample of women and men of reproductive age living in 8,592 households in the ten districts of Lesotho (Butha-Buthe, Leribe, Berea, Maseru, Mafeteng, Mohale's Hoek, Quthing, Qacha's Nek, Mokhotlong, Thaba-Tseka.), both in urban and rural areas. The survey utilized a two-stage sample design. In the first stage, 405

[4] TEBA is a Non-Governmental Organization, located in many southern Africa states. The purpose of TEBA Development is to play a leading role in a collective endeavour to improve living conditions and livelihoods of communities that have provided labour to the mining industry for decades.

clusters (109 in the urban and 296 in the rural areas) were selected from a list of enumeration areas from the 1996 Population Census frame. In the second stage, a complete listing of households was carried out in each selected cluster. Households were then systematically selected for participation in the survey. All women aged 15-49 years who were either permanent household residents in the 2004 LDHS sample or visitors present in the household on the night before the survey were eligible to be interviewed. In addition, in every second household selected for the survey, all men aged 15-59 years were eligible to be interviewed if they were permanent residents or visitors present in the household on the night before the survey. The final sample includes 7,095 women between 15-49 years of age and 2,797 men between 15-59 years old.

[Insert figure 1]

In Lesotho, the 26 percent of women is HIV positive, compared to the 18 percent of men. In many other sub-Saharan African countries the prevalence of HIV is considerably higher among women then men, reflecting that the biological probability of transmission from male to female is substantially higher than from female to male. Figure 1 reports the age profile of HIV prevalence for males and females in Lesotho. The percentage of woman HIV positive is greater compared to infected men in almost all the age groups, except for men from 40 years and older. The graph shows that the prevalence of HIV is considerably higher among young women than among young men. Recent studies in Africa (Dupas, 2006) suggest an explanation for this pattern: unprotected sex between teenage girls and adult men. Men involved in sexual relationships with younger women are more likely to be HIV infected compared to younger boys, because they have been sexually active for longer. Moreover, older men have generally a higher income and, therefore, they have more power to negotiate unprotected sex.

[Insert table 1]

The LDHS includes specific variables describing the coverage of HIV testing for 7140 individuals randomly selected to be tested. Among these eligible individuals, HIV tests were conducted for about 75 percent of women and men. Among the sample of all individuals eligible to be tested, about 11 percent were not interviewed in the DHS. Table 1 reports the coverage rate for HIV testing among eligible people selected for the HIV test as well as the test coverage among interviewed individuals, both for women and men. Women are more likely to have been tested than men: 86.2 percent of interviewed women and 81.4 percent of interviewed men took the HIV test respectively. Almost 15 percent of eligible individuals refused testing when asked for consent and the refusal rate is higher among men, showing a lower propensity of men to know their HIV status. About 0.37 percent of the respondents were not at home at the time the test was conducted in the household, and others were missing test results for other reasons (e.g. technical problems prevented taking blood sample).

[Insert table 2]

Table 2 shows HIV prevalence rates, by employment status. The first column describes the percentage of individuals in the sample, by the job they reported having in the last year.[5] Note that more than 41 percent of the sample reported to be unemployed. This statistic might reflect the big retrenchment of Basotho miners from South Africa mines after the end of the apartheid (see section 3.2 for more details). Thus, it is very likely that most of the unemployed are former miners. Indeed, only about 4 percent of the respondents are currently working in the mines, while about the 24 percent is employed in the agricultural sector. The second column of table 2 shows the percentage of HIV infections by respondent's occupation. Miners have the highest HIV prevalence rate (40%), followed by transport equipment operators (39%) and bricklayers (32%). The last two columns report shows descriptive statistics on partner's occupation for women. Similar to the previous statistics, the highest HIV prevalence rate is among transport equipment operators (40%), followed by miners (35%).

A drawback in this analysis is that we cannot construct the retrospective employment history of each respondent for lack of data. Therefore, we can not recover the information on the percentage of unemployed that were actually former miners before the 2004.

[Insert figure 2]

In which South African provinces Basotho miners are more likely to migrate? The literature on migration suggests that people tend to migrate to the closest region which offers labor

[5] We report only occupations with more than 10 observation each.

opportunity and where there is a consistent social network of individuals of the same nationality. Figure 2 shows in which South Africa provinces Basotho people are mainly clustered. As expected, there is a higher rate of migrating in the South Africa provinces where gold mines are mainly located: Free State, North West and Gauteng (South Africa Census 2001).

[Insert table 3]

Table 3 shows the percentage of miners and of women who answered to have a miner as partner, by district. The districts with the greater prevalence of miners are those that border with the Free State, in the North-western part of Lesotho: Mohale's Hoek, Mafeteng, Butha-Buthe, Leribe, Berea, Maseru. Workers who live close to South Africa are more likely to migrate. In the empirical part of the paper, we will exploit the distance between the household in Lesotho and South Africa borders as a proxy for migration towards mines in South Africa.

[Figure 3]

Figure 3 shows the cross border points between Lesotho and South Africa. There are 13 cross borders points between the two countries (Maseru Bridge, Van Rooyen, Sepapo Gate, Makhaleng, Tele bridge, Qacha's Nek, Ramalesito, Sanipass, Monatsa pas, Caledonspoort, Fricksburg, Peka) and we selected six of those for the identification strategy. The selection of the cross border points is based on the prevalence of mines and on percentage of Basotho miners in South Africa provinces. As we report in figure 2, Basotho mainly migrate in South Africa provinces located in the northern side of Lesotho (Free State, North West and Gauteng), therefore we consider only customs bordering with these provinces. Maseru and Ficksburg bridges are the two most popular border crossing points between Lesotho and the Free State. DHS data provide information on geographical location (city, municipality and district) for each households interviewed and GPS location, including latitude, longitude and altitude for each cluster. We exploit variation in the distance between Lesotho clusters and South Africa borders to instrument the probability of working as miner. The distance variable has been constructed by computing the shortest distance between the selected cross borders and the clusters' location of each respondent, by using the great-circle formula.[6] The idea is that Lesotho migrants have to

[6] Having geographic coordinates of two points A and B on the earth surface (latitudeA, lati-

face higher costs of migration as far they are from South Africa. Working in the mines means spending long time away from their households with higher risk to engage in extra-marital relationship.[7]

Table A1 in the appendix presents summary statistics by gender for the variables used in the empirical analysis. All the variables are weighted with the sample weights given by the data provider. In addition to HIV infection and HIV test, we used behaviors and attitudes related to HIV prevention as dependent variables. 11 percent of females and 19 percent of males tend to use condoms during extramarital intercourse, while only 0.7 percent of women and 0.5 percent of men use this preventive method within the marriage. There is a limited discrepancy across gender regarding abstinence in the last 12 months. The analysis of the independent variables shows that a higher percentage of women than men live in urban areas, respectively 27 and 24 percent. An interesting feature of Lesotho emerges from the education statistics. Women reach highest education grade, both at the primary and secondary levels. This is peculiar to Lesotho since in most other African countries educational achievement is higher for boys. In developing countries, parents have to consider the relative substitution of education against the likelihood of a child finding employment and, in most contexts, this means educating boys. In case of Lesotho, boys have not needs education to become breadwinners, since during the past, they had a guaranteed job in the mines in South Africa. Younger boys are generally in charge of herding cattle (the so called "herdboys") and this benefits girls who are more likely to go to school. The measure of wealth is a set of durable goods held by the household: television, radio, refrigerator, motorbike, car, bicycle, electricity. We grouped the religious affiliation in three categories: Catholic (Roman Catholic Church), Protestant (Evangelical, Methodist, Anglican, Adventist, Pentecostal, other Christian) and no religion.

---

tudeB, longitudeA, longitudeB) in order to compute the shortest distance between them the formula is d=3963.0*arccos[sin(lat1/57.2958)*sin(lat2/57.2958) + cos(lat1/57.2958) *cos(lat2/57.2958) *cos(lon2/57.2958 -lon1/57.2958)].

[7] In the north of Lesotho there is one mining site, but we did not consider it for two reasons. First, the proportion of miners working in Lesotho is very small compared with those working in South Africa. Second, those who works Lesotho mines generally do not have to spend long period away from their households.

3.4 Empirical Strategy

This section analyzes HIV prevalence and sexual behaviors using a multivariate analysis. We run two separate sets of regressions. The first one uses HIV status and the probability of being tested for HIV as dependent variables. The second set of equations uses as dependent variables sexual behaviors and other attitudes related to HIV/AIDS epidemic, such as condom use, extra marital sex and abstinence. Specifically, we tested attitudes related to the so-called "ABC" prevention campaign ("Abstain, Be Faithful and use a Condom"), the most widespread information campaign on HIV prevention in Lesotho and in many other African countries. We then investigate if miners have a higher risk to contract AIDS pandemic compared with agents having other types of jobs. Understanding the dynamic of sexual behaviors is crucial to designing efficient prevention policies. Specifically, we estimate the following equation:

$$Y_{ij} = \beta_0 + \beta_1 M_{ij} + \beta_2 X_{ij} + \delta_j + \epsilon_i \tag{3.1}$$

$$for\ i = 1, 2...n$$

where Y_{ij} is the outcome of interest: a dummy variable equal to one if the agent i in district j is HIV positive or if she adopts risky behaviors. M_{ij} is a dummy taking value one if man i is a miner or if woman i has a miner as partner and 0 otherwise. X_{ij} represents a set of individual characteristics including age and education and household features, such as durable goods and urban status and ϵ_{ij} is the error term. Note that by adding district fixed effects δ_j we control for unobserved characteristics that might affect HIV, such as the number of hospitals, VTC centres or HIV preventive campaigns at the district level. Our identifying assumption is that, conditional on area and controls X_{ij}, the probability of having HIV/AIDS is positively correlated with the the miner status We run separate regressions for men and women.

Two problems can potentially exist in estimating equation (3.1). The first one is linked to an omitted variables bias that explain both the probability of having AIDS and the probability of being a miner. For example, miners could be individuals who really like to spend time away from their households and this is correlated with an higher likelihood to have multiple sexual partnership and with HIV infection. We address this question by including in the estimates two additional regressors capturing whether individuals have a higher propensity to be away from the household for some periods. We exploit two questions in the survey: "In the last 12 months how many separate occasions have traveled away from this household and slept away?" and "In the last 12 month have you been away from your home community for more than 1 month at a time?". These questions have been asked only to male respondents. The second one is a problem of endogeneity. HIV reduces the participation in the labor force throught two channels. The first one is that miners cannot work because once they are HIV positive, they become sicker and weaker and, consequently, they can not carry on job characterized by physical strength. A second concern is that HIV positive individuals have more difficulties in finding a job due to stigmatization. Stigmatization in Lesotho, due to the lack of information on HIV/AIDS transmission, is still a relevant problem. Many firms are reluctant to hire HIV infected people. For example, 50 percent of men in the sample answered that they would not buy vegetables from an infected vendor (DHS 2004).

In a second set of results, we exploit the distance to South Africa borders to instrument the probability of being a miner. The identification strategy is based on the economic theory modelling migration as the result of a cost - benefit analysis in which before moving individuals compare the expected income differentials between the home and the receiving country. Our hypothesis is that being a miner is correlated with the distance to the nearest borders to South Africa: men living very close to the borders have a higher probability to migrate in South Africa and to become miners.

We estimate the following instrumental variable model:

$$Y_{ij} = \beta_0 + \beta_1 M_{ij} + +\beta_2 X_{ij} + \delta_j + \epsilon_i \tag{3.2}$$

$$M_{ij} = \beta_0 + \beta_2 X_{ij} + \beta_3 Distance_{ij} + \delta_j + \epsilon_i \tag{3.3}$$

Y_{ij} is equal to one if the agent i in district j is HIV positive. M_{ij} is a dummy taking value 1 if man i is a miner. X_{ij} is a set of individual characteristics including age, education and some household features, such as durable goods and urban status and ϵ_{ij} is the error term. In the first stage, $Distance_{ij}$ is the distance between the cluster where the individual lives and the nearest cross border point. The distance to the border is used to instrument the probability of being a miners. To control for district unobservable characteristics we include districts dummies, δ_j, in both the regressions. To check that our results are not bias by the functional form of the equations estimated, we estimate equation (3.1) with a linear model, consistent but not efficient and with a probit model.

3.5 Results

3.5.1 HIV Status and HIV test

We begin by examining the probability of being HIV positive.

[Insert table 4]

Table 4 presents the marginal effects of a probit model, where the dependent variable is a dummy equal to one if the individual is HIV positive, and zero otherwise. Controls include demographic characteristics, such as age, age squared, education and religion and household features, such as durable goods and if the household lives in urban or rural area. We control the regressions with district dummies. In all estimates, standard errors are corrected for clustering of the residual at village level.

According to the estimates of column 1, ceteris paribus a miner is 11 percentage points more likely to be HIV positive. Living in urban area increases the probability of being infected by 9 percentage points and the coefficient is statistically significant at 1 percent level. Age squared is negatively correlated with the dependent variable: as expected, working age individuals have an higher probability of infection. Ownership of durable goods, used as proxy for the level of wealth, is negatively correlated with HIV, but it is not statistically significant. Surprisingly, education is not significant and these results are consistent with an explanation that working in mines is the main determinants of getting the disease, whatever is the education level of the individual. In columns 2 we restrict the sample to currently married men. Ceteris paribus, a miner currently married is about 14 percentage points more likely to be HIV positive compared with other currently married men. This result suggests that, not only being a miner is a "risky" job because of the sexual industry developed in mines, but also because currently married miners leave their wife alone for long period of time. During these periods wives also incurred in multiple sexual partnership and they increase the probability to infect their husband when he come back from the mines. Note that being catholic increases the probability of getting the virus for currently married men. The fact that catholic are more conservative could potentially reduce the probability of using condoms.

A potential drawback in the interpretation of the results so far is that it could exist some omitted variables that explains both the probability of having AIDS and the probability of being a miner. For example, miners could be individuals who like to spend time away from their households, and this is correlated with an higher likelihood to have multiple sexual partnership and with HIV. As explained in section 3.3, we address this caveat by including in the estimates two regressors capturing the propensity to be away from the household for some period: sleep away in the last 12 months and being away from your home community for more than 1 month. In columns 3-4 we include a dummy equal to one if the individual have slept away from the household in the last 12 months at least once. Ceteris paribus, a miner is 10 percentage points more likely to be HIV positive, while the probability of sleep away doesn't affect the dependent variable. Age positively affected the dependent variable. The results hold also considering the sub-sample of currently married man and the magnitude of the coefficient increased by 3 percentage points. In columns 5-6 we include among the regressors a variable equal to one if the respondent spent more than one month away from the home community in the last 12 months. While this coefficients doesn't influence the probability to get AIDS, the dependent variable is positively and statistically significant correlated with miners. The magnitude of the coefficient is substantial: being a miner increase the probability of contracting HIV virus by 17 percentage points.

The last two columns of table 4 investigate the probability to being HIV positive for women who have a partner who is a miner. Ceteris paribus, having a partner working in the mines increases the probability to be a HIV positive woman by 6.1 percentage points. This result confirms the previous explanation: women with a miners as partner have a higher probability of infection, both because they have a higher likelihood to have extramarital sex and also because they are at risk to be infected by they husbands coming back from the mines.

[Insert table 5]

Some individuals who were randomly selected to be tested in the survey do not have a test result either because they refused to be tested or because they were absent. Table 5 reports the marginal effect of probit estimates of the probability of being tested for HIV during the survey. Being a miner does not affect the dependent variable, while urban status is strongly negatively associated with the probability of being tested. The coefficient is significant at one percent level, both for males and females. One potential reason might be that people in urban areas are less likely to be present at the time of the test. Compared to individuals with no education, men who have been to primary school are more likely to be tested, but women who have been to secondary school are less likely to be tested. The latter result could suggest that educated women are more likely to be tested before the survey and they think they do not need to be tested again or they may have a greater probability of working outside the household and therefore were absent during the test. Durable goods in the household decrease the probability of testing. The fact that some socioeconomic factors have an impact on the probability of being tested indicates that there is a potential risk of bias in the interpretation of the results in table 5. The sign and the magnitude of the bias would depend on whether we expect individuals who were not tested even though they were selected to be more or less likely to be HIV positive than the average individual.

3.5.2 Condom Use

We next move to the analysis of the determinants of reported behaviors. It is important to note that all those behaviors are self-reported and therefore more subjective than the result of an HIV test. The first set of regressions focuses on the probability of using a condom during the last intercourse. This aspect is particularly interesting for Lesotho, where condoms use is still relatively limited. The Demographic Health Survey asks whether the last intercourse occurred with a spouse or with another partner, thus we can compare this behavior inside and outside the marriage. The use of a condom is recommended in both situations, but the absence of a condom during extra-marital sex is considered even more risky. Unfortunately, in the survey, precise questions about condom use at the last intercourse were only asked to females, so that when we are considering males the dependent variable is the last contraceptive method used.

[Insert table 6]

Table 6 reports the results of the analysis. Being a miner or having a partner as miner is not associated with condom use in the last intercourse in the all sample, while it turns out to be statistically significant by restricting the sample to condom use with spouse or not with spouse. Having a partner as miner decreases the probability to use a condom for women within the marriage, while being a miner has a negative effect on dependent variable when we consider extra-marital intercourse. Ceteris paribus working as a miner reduces the probability to use condom as last contraceptive method by 8.9 percentage points. This is an important source of vulnerability that should be taken into consideration for prevention efforts. Urban status is positively and strongly correlated with the likelihood of using a condom, both inside and outside marriage. People living in urban districts are generally more exposed to information campaigns promoting the use of condoms and there is higher availability in urban areas. The probability of using a condom inside and outside the marriage increases with educational achievement. The relationship is robust and statistically significant for men both with primary and secondary education and for women with secondary education only, suggesting that women bargaining power for having safe sex increase with the level of her education. As expected, wealth positively affects condom use, both for sexual intercourse with the spouse and with other partners.

3.5.3 Abstinence during the last 12 months

Abstinence is alternative strategy to prevent AIDS, in addition to condom use. The LDHS investigates if the respondent did not have sexual intercourse in the last year.

[Insert table 7]

In table 7, we consider the probability of not having had sex in the last 12 months as the dependent variable. As expected, the probability of abstinence decreases for miners and the magnitude of the coefficient is substantial: working as miner reduce the probability to abstain in the last year by 12 percentage points. Having a partner working in the mines increases the probability of abstinence for women (*Restrict the sample to currently married*) Abstinence is also negatively associated with urban status for males, but positively for females. Men with primary education are less likely to have abstained during the last year. The likelihood of not having had sex decreases with wealth for women. Finally, being a Catholic man tends to decrease the probability of abstinence, compared with being Protestant.

3.5.4 Miners and distance to South Africa Borders

The last step in the analysis is to estimate equation 3.2 by addressing the endogeneity problem. More specifically, we attempt to instrument the probability of working in South Africa mines with the distance between the household in Lesotho and South Africa borders.

[Insert table 8]

Table 8 contains the results for the first stage equation. In columns 1-2 we report the linear estimates, while in columns 3-4 we show the marginal coefficient of a probit model. Controlling for district fixed effects and by clustering the standard errors at the village level, we note that higher distance to South Africa borders is negatively correlated with the probability of being miner, suggesting that people living closer to the South Africa have a higher probability to become miners. One standard deviation increase in the distance reduce the probability of working in South Africa mines by 3 percentage point. Considering the sample of women, distance reduces the probability to have a partner working in the miners. The sign of the coefficient is statistically significant at 5 percent level also by considering only women. Note that among the regressors we include a dummy variable equal to one if the individual have lived continuously in the same household to take into account those who moved after getting a job in the mines. The results are robust also by considering a probit model. The F-statistic is above the critical value of 5, showing that distance could be a good instrument in this setting.

[Insert table 9]

Table 9 describes the effect of being a miner on the probability of HIV infection. Working as a miner or having a miner as partner increase the probability of having HIV/AIDS and the coefficient is statistically significant at 5 percent level. The results still hold by using a probit model. This is a first attempt to control the estimates for endogeneity and omitted variable bias and the results show so far give some insight on the vulnerability of miner or former miner for HIV infection. However, the size of the coefficient links to the miner let us suspicions on the validity of the IV estimates: being a miner increases the probability of infection by 78 percentage points from a baseline HIV prevalence of 23 for non-miners, bringing to 100%. This implied 100% HIV prevalence for miners which seems unrealistic. Therefore, we think that further data on the road distance between household and South Africa Borders are required to complete the analysis.

3.6 Robustness check

As robustness check we investigate the probability of being HIV positive for respondents who have other types of jobs.

[Insert table A2]

Table A2 reports linear estimates on the probability of having HIV. The first four columns of table A2 study the probability of infection for teacher and farmers. The coefficients related to these jobs do not have any impact on the dependent variable. Columns 5-8, include type of jobs more similar to miners: carpenters/bricklayer and transport equipment operators. Especially, the latter includes the possibility of spending some periods of time away from the households in South Africa. The probability of having HIV doesn't increase with these type of jobs.

The robustness check analysis confirms the evidence that being employed as miner increases the probability to get HIV/AIDS in Lesotho.

3.7 Conclusions

This paper analyzes the socioeconomic determinants of HIV infection and related sexual behaviors using the 2004 Lesotho Demographic and Health Survey. This is the first nationally representative survey in Lesotho to include an HIV test. The paper focuses in understanding the most vulnerable groups to HIV infection in Lesotho. As main result we show that miners working in South Africa are more likely to be HIV infected. Mines host a vibrant sex industry and getting a job in mines means to face a higher risk of contracting the virus through extramarital partners. Moreover, the paper shows the protective role of education: it is negatively associated with HIV infection and it strongly predicts preventive behaviors like condom use, voluntary counseling and testing and knowledge about AIDS. This adds to the many reasons for supporting the current efforts by the Government of Lesotho in expanding access to primary and secondary education, especially among men. Further steps must be taken to address the endogeneity issues and to find a more valid instrument to proxy the likelihood to be a miner. However, some of our findings directly suggest policy recommendations for increased or more focused prevention efforts, such as preventive and informative HIV/AIDS in mining sites. The battle against AIDS is far from won, but there are signs that companies in high prevalence areas are being innovative in dealing with disease. Recently, an increasing number of mine companies, such as Anglo Platinum, have policies offering education, testing, counselling and access to treatment.

3.8 References

De Walque D., “Who Gets AIDS and How? The determinants of HIV infection and sexual behaviors in Burkina Faso, Cameroon, Ghana, Kenya and Tanzania”. World Bank Policy Research Working Paper # 3844, 2006a.

De Walque D., Sero-Discordant Couples in Five African Countries: Implication for Prevention Strategies. in Population and Development Review 33(3): 501-523. 2007a.

Duflo, E.,Dupas, P. Kremer, M., Sinei, S. "Education and HIV/AIDS prevention : evidence from a randomized evaluation in Western Kenya", 2006.

DHS, “Lesotho Final Report, 2004”, DHS 2004.

Ferguson J., "The Anti-Politics Machine: Development, Depoliticization and Bureaucratic Power in Lesotho", University of Minnesota Press, 1994.

Fylkesnes K., Musonda R. M., Kasumba K., Ndhlovu Z., Mluanda F., Kaetano L., Chipaila C., ”The HIV epidemic in Zambia: socio-demographic prevalence patterns and indications of trends among childbearing women”, AIDS, 11, 339-45, 1997.

Gersovits M., “The HIV epidemic in four African countries seen throught the Demographic and Health Surveys”, The Journal of African Economies, 14: 191-246, 2005.

Lucas R. E. B., "Emigration to South Africa's Mines", The American Economic Review, Vol.77, No.3. pp.313-330, June 1987.

Lurie, MN., Williams BG., Zuma K. Mkaya-Mwambury et al. "The impact of migration on HIV-1 Transmission in South Africa: A study of migrants and non-migrants men and their partners" Sexually transmitted Diseases 2003, 30(2), 139-156.

Oster E. "HIV and Sexual Behavior Change: Why not Africa?", April 2007.

Oster E. "Routes of Infection: Exports and HIV Incidence in Sub-Saharan Africa", NBER Working Paper Number 13049 November 2007.

Sweetman C., "Sitting on a rock: the implications of socioeconomic change in Lesotho for gender, men and development",1999.

Figures and tables

Figure 1: HIV Prevalence by Age Group and Sex

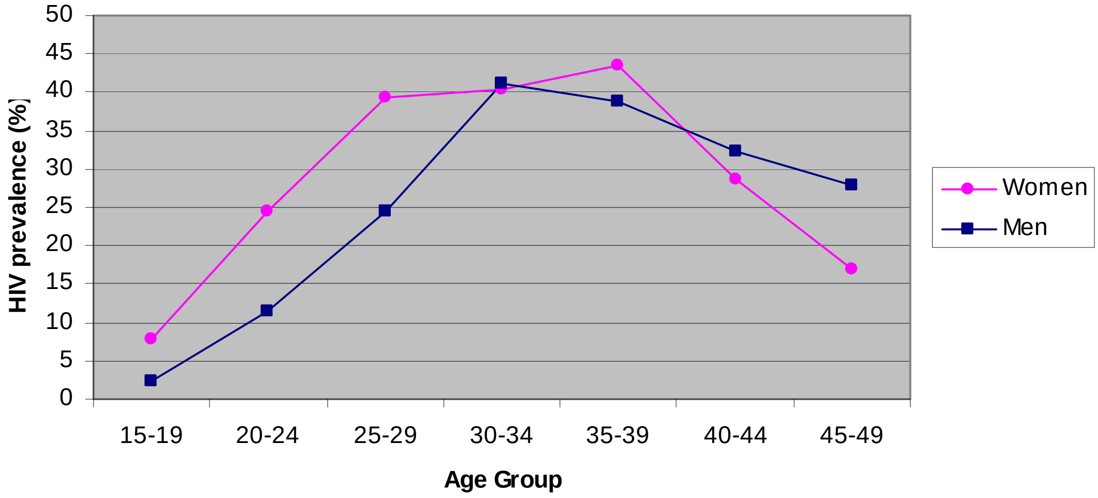


Figure 2: Lesotho Migrants in South Africa Provinces

Legend:
=43982 Basotho
=31167 Basotho
=23376 Basotho

Source: South Africa Census 2001

Figure 3: GPS clusters locations and borders cross points

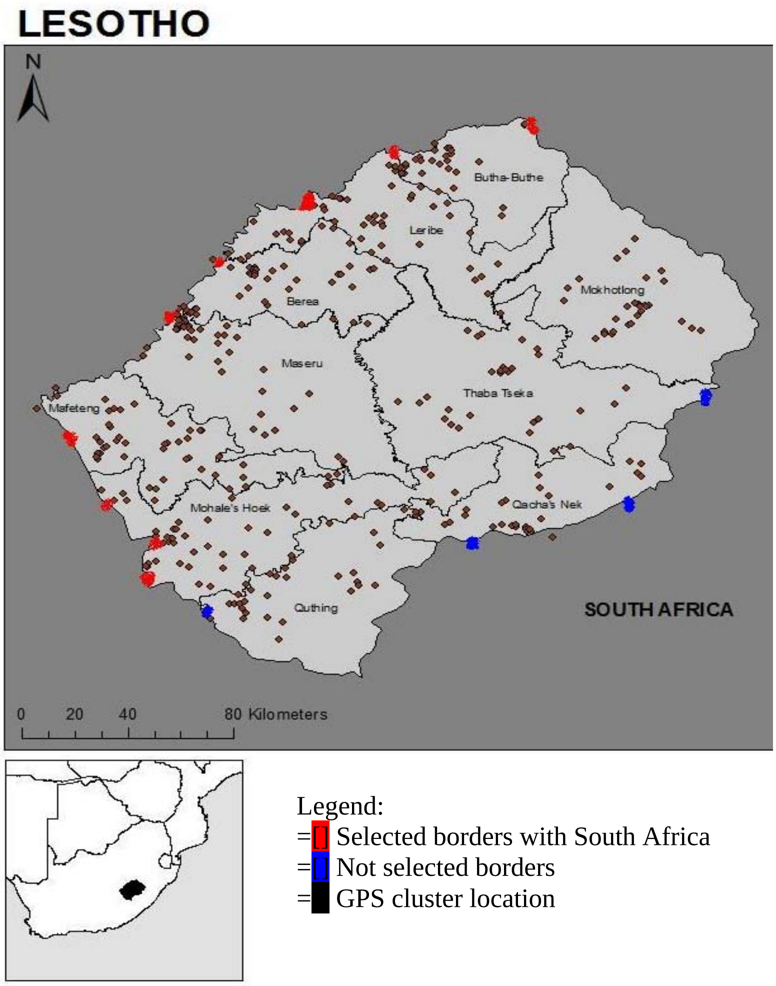


Table 1: Testing Status

		Total		Women		Men	
		% of eligible people selected for HIV test	*Coverage of HIV testing among interviewed individuals*	*% of eligible people selected for HIV test*	*Coverage of HIV testing among interviewed individuals*	*% of eligible people selected for HIV test*	*Coverage of HIV testing among interviewed individuals*
TEST	INTERVIEW	Percent	Percent	Percent	Percent	Percent	Percent
Tested	Interviewed	75.04	84.1	80.10	86.23	69.2	81.42
	Not interviewed	0.34		0.31		0.38	
Refused	Interviewed	12.97	14.61	11.85	12.85	14.26	16.83
	Not interviewed	2.4		1.42		3.54	
Absent	Interviewed	0.33	0.37	0.29	0.32	0.37	0.44
	Not interviewed	6.59		4.38		9.14	
Technical problems	Interviewed	0.05	0.06	0.02	0.02	0.09	0.11
	Not interviewed	0.08		0.53		0.17	
Other	Interviewed	0.75	0.85	1.11	0.58	1	1.2
	Not interviewed	1.45				1.85	
	Total	7,140	6,294	3,836	3,513	3,304	2,781

Note: Data have been weighted with the sample weights recommended by the data provider.
Source: LDHS 2004

Table 2: HIV prevalence by type of job

	Men		*Women*	
	Respondent's occupation		*Partner's occupation*	
	% of the respondents	*% of HIV Positive*	*% of the respondents*	*% of HIV Positive*
Unemployed	41.47	20.23	7.35	28.08
Teachers	1.5	22.22	1.39	31.57
Salesmen, shop assts, and related	2.74	30.00	3.21	36.73
Helpers and related housekeeping svc	0.31	42.80	0.36	22.22
Protective service workers	3.32	26.00	3.86	33.33
Farmers (herd boys)	12.42	20.18	18.47	22.79
Agricultural and animal husbandry worker	11.18	13.36	6.38	15.64
Miners, quarrymen, well drillers, and related	3.93	0.40	31.58	35.27
Spinners, weavers, knitters, dryers	1.46	13.63	2.01	26.31
Bricklayers, carpenters, and construction	7.87	32.67	9.27	35.26
Transport eqmt operators	1.77	39.28	3.69	40.57
Laborers not elsewhere classified	4.29	21.25	4.01	34.48
Obs.	2,262		4,737	

Note: Table reports only the occupations with more than 10 observations each.

Source: DHS 2004

Table 3: Miners and women having a miner as partner, by district

District	*# of miners*	*%*	*# of women having a miner as partner*	*%*
Butha-Buthe	10	11.24	152	10.16
Zeribe	8	8.99	166	11.10
Berea	8	8.99	151	10.09
Maseru	5	5.62	176	11.76
Mafeteng	15	16.85	236	15.78
Mohale's Hoek	23	25.84	217	14.51
Quthing	2	2.25	139	9.29
Qacha's Nek	8	8.99	98	6.55
Mokhotlong	7	7.87	95	6.35
Thaba-Tseka	3	3.37	66	4.41
Total	89	100.00	1496	100.00

Source: DHS 2004

Table 4 : Probability of HIV for miners/partner miner

Dependent variable=1 if HIV positive	*All Men*	*Men Currently married*	*All Men*	*Men Currently married*	*All Men*	*Men Currently married*	*All Women*	*Women Currently married*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Miner	0.1177**	0.1389**	0.1082*	0.1406**	0.1771*	0.1935*		
	[0.0574]	[0.0653]	[0.0589]	[0.0673]	[0.0999]	[0.1117]		
Partner is a miner							0.0611**	0.0435
							[0.0246]	[0.0274]
Sleep away last 12 months			0.0200	0.0243				
			[0.0209]	[0.0325]				
Away more than 1 month					-0.0026	-0.0137		
					[0.0314]	[0.0452]		
Age	0.0666***	0.0465***	0.0663***	0.0458***	0.0796***	0.0560***	0.0849***	0.0757***
	[0.0058]	[0.0126]	[0.0059]	[0.0128]	[0.0097]	[0.0177]	[0.0095]	[0.0103]
Age squared	-0.0009***	-0.0007***	-0.0009***	-0.0006***	-0.0010***	-0.0008***	-0.0013***	-0.0012***
	[0.0001]	[0.0002]	[0.0001]	[0.0002]	[0.0001]	[0.0002]	[0.0001]	[0.0002]
Urban	0.0926***	0.0937**	0.0894***	0.0859**	0.1214***	0.1527**	0.0996***	0.0906***
	[0.0306]	[0.0429]	[0.0302]	[0.0424]	[0.0438]	[0.0646]	[0.0269]	[0.0311]
Primary Educ.	-0.0088	-0.0257	-0.0051	-0.0208	-0.0111	-0.0029	-0.0086	-0.0040
	[0.0248]	[0.0367]	[0.0249]	[0.0375]	[0.0395]	[0.0584]	[0.0528]	[0.0581]
Secondary Educ.	-0.0128	-0.0317	-0.0130	-0.0238	-0.0683	-0.1046	0.0169	0.0216
	[0.0314]	[0.0491]	[0.0320]	[0.0506]	[0.0435]	[0.0660]	[0.0581]	[0.0622]
Durable goods	-0.0121	-0.0105	-0.0133	-0.0156	-0.0186	-0.0029	-0.0106	-0.0073
	[0.0103]	[0.0165]	[0.0105]	[0.0168]	[0.0166]	[0.0245]	[0.0113]	[0.0118]
Catholic	0.0266	0.0652**	0.0278	0.0687**	0.0358	0.0480	0.0137	0.0219
	[0.0201]	[0.0305]	[0.0205]	[0.0311]	[0.0315]	[0.0467]	[0.0202]	[0.0224]
No Religion	-0.0157	0.0061	-0.0121	0.0138	-0.0579	-0.1458*	0.1921	0.2557*
	[0.0385]	[0.0645]	[0.0388]	[0.0649]	[0.0548]	[0.0785]	[0.1236]	[0.1498]
District dummies	yes	yes	yes	yes	yes	yes	yes	yes
No. Obs.	1781	932	1749	915	761	425	2018	1594

Notes: * denotes significance at 10 percent level, ** at 5 percent level, *** at 1 percent level. Table reports marginal probit coefficient; standard errors in parenthesis, adjusted for clustering of the residuals at the village level

Estimates weighted with the sample weights provide by the data providers

Table 5: Determinants of being tested for HIV

Dependent variable=1 if HIV tested

	(1) Males	(2) Females
Miner	-0.0392	
	[0.0530]	
Partner miner		0.0029
		[0.0140]
Age	-0.0030	-0.0046
	[0.0053]	[0.0054]
Age squared	0.0000	0.0000
	[0.0001]	[0.0001]
Urban	-0.0928***	-0.0716***
	[0.0347]	[0.0216]
Primary Educ.	0.0423*	-0.0498
	[0.0242]	[0.0383]
Secondary Educ.	-0.0111	-0.0997*
	[0.0330]	[0.0549]
Durable goods	-0.0175*	-0.0112*
	[0.0093]	[0.0059]
Catholic	0.0009	-0.0020
	[0.0221]	[0.0127]
No Religion	-0.0033	-0.2613***
	[0.0385]	[0.0985]
Unemployed	0.0348	
	[0.0223]	
Unemployed Partner		0.0126
		[0.0232]
District dummies	yes	yes
Observations	2232	2325

Notes: * denotes significance at 10 percent level, ** at 5 percent level, *** at 1 percent level
Table reports marginal probit coefficient; standard errors in parenthesis, adjusted for clustering of the residuals at the village level

Estimated weighted with the sample weights provide by the data providers

Table 6: Determinants of using a condom

Dependent variable=1 if condom used	*All Sample*		*With spouse*		*Not with Spouse*	
	Male	*Female*	*Male*	*Female*	*Male*	*Female*
	(1)	(2)	(3)	(4)	(5)	(6)
Miner	-0.0485		0.0041		-0.0895***	
	[0.0506]		[0.0437]		[0.0245]	
Partner miner		-0.0175		-0.0356***		0.0014
		[0.0134]		[0.0120]		[0.0018]
Age	0.0041	0.0190***	-0.0017	0.0157***	-0.0194***	0.0009
	[0.0050]	[0.0053]	[0.0068]	[0.0048]	[0.0056]	[0.0006]
Age squared	-0.0002**	-0.0003***	-0.0000	-0.0003***	0.0001*	-0.0000
	[0.0001]	[0.0001]	[0.0001]	[0.0001]	[0.0001]	[0.0000]
Urban	0.0936***	0.0906***	0.0409	0.0507***	0.0553*	0.0089**
	[0.0308]	[0.0206]	[0.0265]	[0.0191]	[0.0315]	[0.0036]
Primari Educ.	0.0895***	0.0405	0.0023	0.0470	0.0926***	0.1045
	[0.0269]	[0.0429]	[0.0236]	[0.0498]	[0.0264]	[0.0944]
Secondary Educ.	0.2482***	0.1201**	0.1034**	0.1282*	0.2099***	0.4732
	[0.0416]	[0.0581]	[0.0466]	[0.0753]	[0.0484]	[0.3455]
Durable goods	0.0148	0.0128***	0.0091	0.0138***	0.0128	-0.0004
	[0.0111]	[0.0049]	[0.0088]	[0.0047]	[0.0094]	[0.0005]
Catholic	0.0121	-0.0125	0.0153	-0.0215*	-0.0130	0.0010
	[0.0212]	[0.0131]	[0.0178]	[0.0117]	[0.0187]	[0.0014]
No Religion	0.0255	-0.0340	0.0480	0.0219	0.0160	
	[0.0413]	[0.0475]	[0.0495]	[0.0698]	[0.0375]	
Unemployed	0.0049		0.0080		0.0118	
	[0.0184]		[0.0217]		[0.0173]	
Unemployed Partner		-0.0194		-0.0168		0.0036
		[0.0216]		[0.0199]		[0.0045]
District dummies	yes	yes	yes	yes	yes	yes
Observations	2243	4114	1132	3448	1825	3420

Notes: * denotes significance at 10 percent level, ** at 5 percent level, *** at 1 percent level
Table reports marginal probit coefficient; standard errors in parenthesis, adjusted for clustering of the residuals at the village level

Estimated weighted with the sample weights provide by the data providers

For women condom used reffered to the last intercourse , while for men is the most recent contraceptive method

Table 7: Abstinence during the last 12 months

Dependent variable=1 if no sexual intercourse in the last 12 months	*Males*	*Females*
	(1)	(2)
Miner	-0.1278***	
	[0.0281]	
Partner miner		0.0346***
		[0.0128]
Age	-0.0473***	-0.0082
	[0.0046]	[0.0051]
Age squared	0.0006***	0.0002**
	[0.0001]	[0.0001]
Urban	-0.0518**	0.0362**
	[0.0243]	[0.0153]
Primary Educ.	-0.0501**	0.0204
	[0.0230]	[0.0282]
Secondary Educ.	-0.0371	0.0022
	[0.0285]	[0.0312]
Durable goods	-0.0084	-0.0305***
	[0.0106]	[0.0062]
Catholic	-0.0499***	-0.0022
	[0.0186]	[0.0110]
No Religion	0.0161	-0.0813***
	[0.0320]	[0.0175]
Unemployed	0.0416*	
	[0.0217]	
Unemployed Partner		-0.0111
		[0.0184]
District dummies	0.0602	0.0071
Observations	2243	4681

Notes: * denotes significance at 10 percent level, ** at 5 percent level, *** at 1 percent level
Table reports marginal probit coefficient; standard errors in parenthesis, adjusted for clustering of the residuals at the village level

Estimated weighted with the sample weights provide by the data providers

Table 8 : Probability of being a miner (or having a partner as miner) instrumented with distance to South Africa borders – First Stage

Dependent variable=1 if miner or women have a miner as partner

	OLS		*Marginal probit coeff.*	
	Men	*Women*	*Men*	*Women*
	(1)	(2)	(3)	(4)
Distance to SA borders	-0.0008**	-0.0020***	-0.0001***	-0.0023***
	[0.0003]	[0.0004]	[0.0002]	[0.0005]
Age	0.0023	0.0329***	0.0021***	0.0407***
	[0.0024]	[0.0057]	[0.0023]	[0.0083]
Age sq.	0.001	-0.0002***	-0.000012	-0.0004***
	[0.003]	[0.0001]	[0.000032]	[0.0001]
Primary Educ.	0.0029	0.0368	0.0028	0.0422
	[0.0112]	[0.0356]	[0.011]	[0.053]
Secondary Educ.	-0.0112	0.0276	-0.01371	0.0445
	[0.014]	[0.0374]	[0.0146]	[0.053]
Durable goods	0.008	0.0012	-0.00679	0.029
	[0.004]	[0.0063]	[0.0047]	[0.0099]
Catholic	0.0027	-0.0115	0.0027	-0.0167
	[0.009]	[0.0132]	[0.0095]	[0.02011]
No Religion	-0.0046	-0.0810	-0.004	-0.0720
	[0.0173]	[0.0645]	[0.0172]	[0.1004]
Unemployed	-0.0723***		-0.0705***	
	[0.009]		[0.009]	
Living continuously in the same HH	0.0246**		0.0298**	
	[0.011]		[0.011]	
Unemployed Partner		-0.2985***		
		[0.0252]		
District dummies	yes	yes	yes	yes
Observations	1677	1900	1677	1900
F-stat	7.72	6.73	6.74	7.75
R-squared	0.09	0.19	0.20	0.21

Robust standard errors in parenthesis, * significant at 10%; ** significant at 5%; *** significant at 1% adjusted for clustering of the residuals at the village level

Table 9 : Probability of HIV/AIDS for miner - Second Stage

Dependent variable=1 if HIV positive	*OLS*		*Marginal probit coeff.*	
	Men (1)	*Women* (2)	*Men* (3)	*Women* (4)
Miner	1.4859*		0.786***	
	[0.8683]		[0.036]	
Partner as miner		0.5609*		0.517***
		[0.2926]		[0.3247]
Age	0.0548***	0.0575***	0.055***	0.057**
	[0.0066]	[0.0149]	[32.4]	[32.4]
Age sq.	-0.0008***	-0.0010***	-0.001***	-0.001***
	[0.0001]	[0.0002]	[1198]	[1198]
Primary Educ.	-0.0078	-0.0345	-0.006	-0.028
	[0.0313]	[0.0540]	[0.55]	[0.65]
Secondary Educ.	0.0114	0.0006	0.009	0.002
	[0.0393]	[0.0607]	[0.192]	[0.3068]
Durable goods	-0.0206	-0.0064	-0.026*	-0.006
	[0.0143]	[0.0127]	[0.819]	[0.89]
Catholic	0.0230	0.0312	0.002	0.029
	[0.0245]	[0.0230]	[0.434]	[0.434]
No Religion	-0.0056	0.2188*	-0.002	0.208*
	[0.0427]	[0.1230]	[0.787]	[0.009]
Unemployed	0.0785		0.095*	
	[0.0609]		[0.435]	
Living continuously in the same HH	-0.0522		-0.045	
	[0.0377]		[0.793]	
Unemployed Partner		0.1622		0.157
		[0.1007]		[0.0652]
District dummies	yes	yes	yes	yes
Observations	1677	1900	1677	1900

Robust standard errors in parenthesis, * significant at 10%; ** significant at 5%; *** significant at 1% adjusted for clustering of the residuals at the village level

Appendix

Table A1: Summary Statistics

	Definition	*Gender*	*Obs.*	*Mean*	*Std. Dev.*
	Dependent variables				
HIV positive	Equal 1 if the individual is HIV positive, 0 otherwise[1]	Female	3020	0.26	0.439
		Male	2234	0.18	0.389
HIV test	Equal 1 if the individual has been tested for HIV in the survey, 0 otherwise	Female	3513	0.87	0.329
		Male	2781	0.82	0.378
Used condom (not with spouse)	Equal 1 if used condom during the last sexual intercourse, if not with spouse	Female	4967	0.11	0.31
		Male	2026	0.199	0.40
Used condom (with spouse)	Equal 1 if used condom during the last sexual intercourse with spouse	Female	4967	0.07	0.255
		Male	2026	0.055	0.228
No sex last 12 months	Equal 1 if abstinent in the last 12 months, 0 otherwise	Female	7090	0.29	0.457
		Male	2785	0.28	0.446
	Independent Variables				
Miner	Equal to one if the respondent is a miner	Male	2262	0.039	0.194
Partner miner	Last partner's occupation for currently married people and last partner's occupation for formerly married people	Female	4737	0.31	0.46
Distance	Distance between clusters and closeses South Africa borders in the Lesotho Northern part		38252	48.16	37.13
Durables	# of durable goods held by the hh: TV, radio, refrigerator, motorbike, car, bicycle, electricity		40255	0.87	1.119
Urban	Equal 1 if the individual is in urban area, 0 otherwise	Female	7095	0.274	0.446
		Male	2797	0.24	0.432
Primary education	Equal 1 if achieved at least some primary education	Female	7082	0.60	0.487
		Male	2791	0.54	0.49
Secondary/Tertiary educ.	Equal 1 if achieved at least some secondary education or more	Female	7082	0.36	0.48
		Male	2791	0.25	0.43
Catholic	Equal 1 if roman catholic church, 0 otherwise	Female	7083	0.45	0.49
		Male	2797	0.46	0.49
No religion	Equal 1 if no religion	Female	7083	0.09	0.94
		Male	2797	0.067	0.25
Protestant	Equal 1 if evangelical, methodist, anglican, adventist, pentecostal and other christian, 0 otherwise	Female	7083	0.54	0.497
		Male	2797	0.48	0.499

[1] In these percentages are not included the indeterminate test results, equal to 0.65% of tested individuals.

Notes: Standard errors in brackets. N.a: not applicable, variable not included. The data are from the 2004 DHS and have been weighted with the sample weights recommended by the data provider. Standard errors take into account the clustering at the enumeration area level.

Table A2 : Probability of HIV for other type of job

Dependent variable=1 if HIV positive	*All men*	*Currently married men*	*All men*	*Currently married men*	*All men*	*Currently married men*	*All men*	*Currently married men*
	Marg.Coeff. (1)	*Marg.Coeff.* (2)	*Marg.Coeff.* (3)	*Marg.Coeff.* (4)	*Marg.Coeff.* (5)	*Marg.Coeff.* (6)	*Marg.Coeff.* (7)	*Marg.Coeff.* (8)
Teacher	-0.0447	-0.0597						
	[0.0795]	[0.1166]						
Carpenter/Bricklayer							0.0356	-0.0275
							[0.0349]	[0.0440]
Transport eqmt operator					0.1118	0.1335		
					[0.0887]	[0.1138]		
Farmer			-0.0181	-0.0177				
			[0.0299]	[0.0446]				
Age	0.0672***	0.0483***	0.0671***	0.0480***	0.0667***	0.0475***	0.0663***	0.0485***
	[0.0058]	[0.0126]	[0.0058]	[0.0126]	[0.0058]	[0.0126]	[0.0058]	[0.0126]
Age squared	-0.0009***	-0.0007***	-0.0009***	-0.0007***	-0.0009***	-0.0007***	-0.0009***	-0.0007***
	[0.0001]	[0.0002]	[0.0001]	[0.0002]	[0.0001]	[0.0002]	[0.0001]	[0.0002]
Urban	0.0915***	0.0927**	0.0907***	0.0918**	0.0902***	0.0925**	0.0897***	0.0959**
	[0.0306]	[0.0431]	[0.0305]	[0.0429]	[0.0307]	[0.0428]	[0.0305]	[0.0433]
Primary Educ.	-0.0077	-0.0233	-0.0087	-0.0242	-0.0098	-0.0266	-0.0088	-0.0219
	[0.0247]	[0.0365]	[0.0250]	[0.0368]	[0.0248]	[0.0366]	[0.0247]	[0.0364]
Secondary Educ.	-0.0107	-0.0294	-0.0150	-0.0347	-0.0146	-0.0351	-0.0125	-0.0331
	[0.0320]	[0.0497]	[0.0319]	[0.0495]	[0.0312]	[0.0488]	[0.0315]	[0.0492]
Durable goods	-0.0107	-0.0077	-0.0114	-0.0092	-0.0121	-0.0105	-0.0103	-0.0094
	[0.0103]	[0.0167]	[0.0103]	[0.0165]	[0.0103]	[0.0167]	[0.0103]	[0.0165]
Catholic	0.0263	0.0638**	0.0263	0.0642**	0.0256	0.0622**	0.0265	0.0640**
	[0.0200]	[0.0304]	[0.0200]	[0.0304]	[0.0200]	[0.0304]	[0.0200]	[0.0304]
No Religion	-0.0186	0.0005	-0.0191	0.0009	-0.0215	-0.0079	-0.0162	0.0002
	[0.0381]	[0.0641]	[0.0381]	[0.0642]	[0.0376]	[0.0622]	[0.0385]	[0.0642]
District dummies	Yes	yes	yes	yes	yes	yes	yes	yes
No. Obs.	1781	932	1781	932	1781	932	1781	932

Notes: * denotes significance at 10 percent level, ** at 5 percent level, *** at 1 percent level
Table reports marginal probit coefficient; standard errors in parenthesis, adjusted for clustering of the residuals at the village level