Roads to Development: Experimental Evidence from Urban Road Pavement*

Marco Gonzalez-Navarro
UC Berkeley †

Climent Quintana-Domeque
Universitat d'Alacant ‡

This Draft

February 23, 2010

Abstract

Urban peripheries in many developing country cities lack basic local public goods like pavement, water, sewerage and electricity. We estimate the impacts of slum infrastructure upgrading using an experiment in urban road pavement provision in Mexico. Our findings show that homes in streets that were paved increased their value between 15 and 17%. Households living in streets that were paved obtained more credit, had higher per capita expenditures, increased motor vehicle ownership and were more likely to have made home improvements. The rate of return to road pavement is estimated to be 2% without considering externalities, but raises to 55% once externalities are accounted for. We also present a model to understand the experimental estimates. Increases in consumption are more strongly correlated with increases in housing value than reductions in transport costs, suggesting that the wealth effect generated by the road pavement was a stronger driver of consumption than the reduction in transport costs.

JEL: C92, C93, H41, O15, O12

---

*We would like to thank Anne Case, Angus Deaton, Frederico Finan, Paul Gertler, Alan Krueger, David Lee, Adriana Lleras-Muney, Jeremy Magruder, Edward Miguel, Christina Paxson, Elisabeth Sadoulet, Cecilia Rouse and Sam Schulhofer-Wohl for their comments and advice in this research project. The collaboration of the Acayucan 2005-2007 and 2008-2010 Municipal administrations is gratefully acknowledged. We also recognize the effort of José Luis Palma, Luz Uribe and Monica González at INSAD who were in charge of the survey, and José Luis Reyes who provided us with the house value assessments. Financial aid and support from Princeton University, Princeton Woodrow Wilson Scholars, Princeton Industrial Relations Section, Princeton Research Program in Development Studies, Robert Wood Johnson Scholars in Health Policy Research Program, Berkeley Economics, Universitat d'Alacant, the Lincoln Institute of Land Policy and the Spanish Ministry of Science and Innovation (ECO 2008-05721/ECON) is gratefully acknowledged. Any errors contained in the paper are our own.

†50 University Hall, MC 7360 Berkeley, CA 94720. Email: marcog@berkeley.edu

‡Departament de Fonaments de l'Anàlisi Econòmica, Universitat d'Alacant, Sant Vicent del Raspeig, 03690 Alacant, Spain. Email: climent@ua.es

1 Introduction

Developing countries are urbanizing at a much more rapid pace than was experienced by currently developed economies (Henderson, 2002; UN-Habitat, 2003). Rapid population growth in cities has generated a widespread lack of urban infrastructure, especially in the outskirts of cities. Because it is precisely in the outskirts of urban areas where welfare indicators are worse (Napier, (2009)), it is important to understand what the effects of urban infrastructure are. In particular, it is crucial for policy makers to know if investments in urban infrastructure can be an effective tool in poverty alleviation, in a context in which public funds must compete with cash transfer programs - such as *Progresa*.

Roads have been proposed for a long time as poverty reduction tools (Jalan and Ravallion, 2002). However, there is little convincing evidence that road paving affects social outcomes (Van de Walle, 2002). The main endogeneity challenge with any study focusing on impacts of infrastructure is that a simple comparison of places with and without infrastructure in observational data can be misleading (Duflo and Pande, 2007). Paraphrasing Van de Walle (2002) "The general point here is that unless road placement is truly random - which seems most unlikely - simple comparisons of outcome indicators in places with roads versus without them can be very deceptive." Our work is unique in that it is the first to solve the selection bias inherent in road infrastructure placement by using random assignment. When treatment is randomly assigned, the treatment is independent of other sources of variation, and any bias is balanced across treatment and controls.

We study the effects of an experiment in urban infrastructure provision in Mexico. The experimental design consisted of randomly selecting from a pre-approved set of street projects (defined as contiguous sets of unpaved city blocks connecting to the city's pavement grid) a subset to be treated with road pavement. Randomization of urban infrastructure provision is assessed through a household baseline survey and business census (pre-intervention) and the evaluation of the effects of urban infrastructure provision is done by means of a household follow-up survey and a business census (post-intervention).

At the household level, we find that road pavement increased property values by around 15% according to professional appraisals, and by around 24% according to homeowners. While collateral-based credit from the private sector more than doubled in terms of number of loans and size, we find no response in other forms of credit, such as non-collateral-based, or from family and friends. The provision of road pavement appears to have incentivized households to make home improvements and buy materials for home improvements. Moreover, the household head was substantially more likely to use motorized transport to go to work as a result of the paving of the road, and households in general increased by 50% their vehicle ownership. Plans to outmigrate for work reasons were reduced as a result of the infrastructure. Lastly, monthly per capita expenditure is estimated to have increased by around 10%, and the treated households increased the number of durable goods they possess.

At the neighborhood level, the provision of road pavement did not affect either immigration or out-migration flows. Further, immigrants to and out-migrants from paved and unpaved streets were not different in their observable characteristics, such as consumption, labor income, home ownership status or durable goods. The business census evidence suggests that the road pavement had no impact on business opportunities. Number of business units were unchanged, number of employees and firm profits did not vary with access to pavement.

Interestingly, we find experimental effects at the household but not at the business level. Although in a different context, these findings are consistent with Haughwout (2002), who concludes that the principal beneficiaries of infrastructure investment are property owners, not firms.[1]

In an approach reminiscent of Jacoby (2000), in which properties along paved roads are understood as assets whose price equals the net present value of the benefits they provide,

[1] His findings for the USA show that one-standard deviation increase in a city's infrastructure stock raises the value of an acre of city land by between \$ 11,000 and \$ 22,000 (an elasticity of 0.11 and 0.23). In contrast, the elasticity of productivity with respect to infrastructure is 0.038.

we provide a cost benefit analysis of road pavement by comparing the costs of the roads to the increase in property values they generated. The sum of values of properties along paved roads increased by slightly more than the cost of the pavement, generating a positive return of 2%. However, when externalities on unpaved road properties are taken into account, the estimated return to road pavement increases to 55%.

There is a large non-experimental macro literature on returns to infrastructure, mostly based on cross-regional and cross-country data (Antle, 1983; Aschauer, 1989; Holtz-Eakin, 1992; Kneller et al., (1999); Canning and Bennathan, (2000); Esfahani and Ramirez, 2003; Briceño, Estache and Shafik, (2004); Fan and Chan-Kang, 2005), which estimates returns to public infrastructure between 0 and 200%.[2] However, as already pointed out by Gramlich (1994), it is very difficult to obtain either reliable economic measures of rates of return to public infrastructure or meaningful econometric estimates of productivity impacts. Our study contributes to this macro-growth literature on infrastructure by providing an experimental estimate of the rate of return to road pavement in a developing country. This is particularly important given that public spending on infrastructure in developing countries averaged 9% of government spending (World Bank, 1994) and 15% to 20% of the World Bank's lending portfolio is targeted towards transport investments (Khandker, (2009)).

Despite the widespread belief that infrastructure is integral to development, evidence on how investment in physical infrastructure affects individual wellbeing remains limited, as pointed out by the World Bank (1994), Jimenez (1995) and Dinkelman (2008). By focusing on how treated households respond to urban road pavement, we can better understand the impact of public infrastructure on individual wellbeing. Not only this, we are also able to investigate some potential mechanisms by which urban road pavement affects household behavior. Hence, we also contribute to the infrastructure literature from a microeconomic point of view, and to the growing literature studying the impact of public infrastructure in

[2] In his review of the literature, Holtz-Eakin (1992) concludes that the evidence from state-level and region-level US data indicates a 0 elasticity of private output or productivity with respect to state or local government capital. However, Kneller et al. (1999) finds that across 22 OECD countries productive government spending has a social rate of return of between 10 and 20%.

the context of a developing country (Lokshin and Yemtsov, 2005; Van de Walle and Mu, 2007; Michaels, 2008).

The structure of the paper is as follows. Section 2 describes the experimental design. Section 3 offers a description of the experimental streets in 2006, before the intervention. Section 4 discusses the identification strategy and provides some testable implications. In Section 5 we present evidence that the randomization produced a balanced sample between treatment and control groups in terms of observable characteristics, and present our experimental estimates. Section 6 discusses some potential mechanisms to understand the experimental estimates and provides some evidence on them. In Section 7 we explore the role of externalities and present a cost-benefit analysis. Finally, Section 8 concludes.

2 Experimental Design

2.1 Institutional Context

Acayucan is a municipality in the southern part of the state of Veracruz, in eastern Mexico. In Table 1 we present some descriptive statistics from the entire city, column (1), and from the experimental streets, columns (2) and (3). According to the 2005 short Census (Conteo), the municipality has a population of 79,459, with the city accounting for about 50,000. The average altitude is 100 meters above sea level, with tropical climate. The sex ratio is 0.89 males for every female.[3] Of those aged 15 and more, 9% are illiterate. School enrollment is 94% among adolescents aged 12-14.

Regarding household level variables, electricity is enjoyed by almost everyone with 98% of homes having electricity in their property. Tap water is less common: 16% of private inhabited dwellings report not having access to piped water in their lot or home. In terms of assets, 81% of homes have a refrigerator, 55% have a washing machine, and 14% have

[3] Grech et al. (2003) have documented a falling male to female ratio in all of Mexico, but well above one. The only explanation we have encountered in the literature for low male to female sex ratios such as the one in Acayucan is the existence of male migrant labor (Bean, King, and Passel, 1980).

computers.

Interestingly, the descriptive statistics from the 2006 baseline survey are close to those of the 2005 short Census, with the exception of the fraction of households having access to piped water in their lot or home and the number of rooms in the house, which are less for inhabitants from our survey. Although census tracts with streets that are part of the experiment have worse indicators than census tracts in the downtown area of the city, there are many areas that were not part of the experiment with even worse socioeconomic indicators. This highlights the fact that although the experiment took place in parts of the city that are relatively poor, they did not contain the poorest households, which tend to live in scarcely populated areas with many vacant lots, where the municipal government was not yet interested in providing urban road infrastructure. This aspect must be taken into account when assessing the external validity of our estimates.

Municipal governments in Mexico have as their main responsibilities garbage collection, paying for public street illumination, providing local public safety, regulating businesses, tending to public gardens, and providing and maintaining public infrastructure including sewerage, road pavement, and sidewalks. Each three-year administration has freedom to choose what it will focus its budget on.

Mexico's government obtains its funds mainly from a national VAT, a national income tax, and oil proceeds from the state-run oil company. These funds are shared by the three orders of government: Federal, State and Local. Hence, funding of the municipal government comes mainly from transfers from the Federal and State Government. A significant portion of these transfers is conditional on being spent on things like infrastructure. Local sources of revenue (mainly the urban property tax) account for less than 10% of the total municipality budget.

2.2 The Experiment

The 2005-2007 Acayucan administration put forth as its priority providing pavement in city areas lacking these services. However, the infrastructure needs of the city were much larger than what could be provided for with the municipality budget. Under these circumstances, we proposed a randomized evaluation of their urban road pavement infrastructure investments.

Throughout the city, there are many streets without pavement. The administration was interested in upgrading those with higher population densities, and left for the future areas that were not yet heavily populated. The mayor and the public works personnel provided us with a set of 56 "street projects" they were interested in upgrading throughout the city. The administration was responsible for selecting and defining those projects. The street projects consisted of sets of contiguous city blocks that connected to the existing city pavement grid. One condition for being part of the experiment was for the street not to be paved. Once it became part of the experiment, the city determined if the tap water or sewerage lines would be replaced or upgraded.

Given that the administration would not be able to provide infrastructure to the 56 "street projects", council members and the mayor voted to let us use random assignment to choose which roads to pave within the set of interest to them. The municipality accepted the randomization requirement because they were interested in having a third party evaluating their public works program, and they understood it as a fair and transparent way to provide urban road pavement. We randomly assigned 28 out of the 56 "street projects" to the group of streets to be paved. The randomization provides a credible strategy to identify the benefits of such a policy because it manages to overcome the selection bias, a major concern in the infrastructure literature.

2.2.1 Treatment

One important challenge for the randomization was the many sources of uncertainty the Municipal government would face during the course of the infrastructure construction. These included volatile government income and input cost fluctuations. The main factor that could slow down construction was unforeseen weather: construction crews could not perform some important tasks on rainy days.

Given that the municipal government is free to choose its infrastructure program, the municipality decided there was no need to announce to the population the existence of this study. Moreover, the questionnaire did not mention that its objective was to measure the effects of infrastructure and field workers were trained to not mention this to respondents. Hence, changes in behavior among the treatment group (Hawthorne effects) and among the control group (John Henry effects) were minimized.

By March 2009, 17 of the streets in the treatment group had been completely treated, four were in process but unfinished, and seven had not been pursued. The municipal government argues that the weather and some technical difficulties did not allow them to provide road pavement to these eleven streets. Figure 1 shows the location of experimental areas throughout the city: ITT (streets assigned to the treatment group) and control (streets assigned to the control group). Table 2 lists all the projects assigned to the treatment group and the date in which this was completed.

The administration did fulfill the requirement of not paving the projects assigned to the control group. This is important because under one-sided non-compliance, the Bloom (1984) result tells us how to use the IV formula to estimate the average effect of the treatment on the treated.

Finally, it is worth mentioning that the administration did not agree to paving streets in a random order, mainly because there are efficiencies when paving streets that are near one another: by moving machinery around from one street project to another and traveling a very short distance; establishing a common point close to various street projects to distribute construction material; and having constant supervision of workers.

2.3 Sources of Information

In our experiment, the unit of randomization is the street project, though the main unit of analysis is the household within each street. Information on households was collected through the Acayucan Standards of Living Survey (ASLS, Encuesta de Condiciones de Vida y Opciones Económicas en Acayucan). The survey contains two rounds: a baseline (2006) and a follow-up (2009) conducted during the second half of February and the beginning of March. Both rounds were conducted during the same months with the intention of minimizing seasonality effects.

We decided to conduct a baseline survey because of three main reasons. First, a baseline survey provides an opportunity to check that randomization was conducted appropriately. This is particularly important in our context because when randomization is done at the cluster level instead of individually, there is a non-negligible probability that the randomization produces groups with different average characteristics (Bloom, 2006). Second, a baseline survey provides information on lagged outcomes that may generate more precise estimates of the effects of the treatment on outcomes by including them in a regression of the final outcome on the treatment variable and a constant (E. Duflo and Kremer, 2006; Kling, Liebman, and Katz, 2007).

We also obtained information on business units through a very short business census in 2006 and 2009 and home value assessments made by a real estate agent.

2.3.1 Household Surveys

The survey firm created a sampling frame from all inhabited residential dwellings found in experiment streets in early 2006. As recognized by Deaton (1997), the use of out-dated or otherwise inaccurate sampling frames is an important source of error in survey estimates. The sampling procedure was clustered sampling: From the list of dwellings in each cluster we randomly chose a pre-specified fraction to be interviewed.

Given the uncertainty about the total number of projects the municipality would be able to conclude by the time of the follow-up survey, and since we did not have any prior about having different sample variances of the outcome means in the treatment and control groups, we decided to sample with a higher intensity in the intent-to-treat (ITT) group (70% of dwellings) than in the control group (50% of dwellings).[4]

Notice that some dwellings would contain more than one household (defined as a group of one or more people who live in the same house and share food expenditures). The procedure in case of multiple households per dwelling was to interview all of them. It is worth noting that neither quota sampling nor substitution of non-responding households or individuals (whether refusals or non-contacts) were permitted at any stage.

The household questionnaire collects detailed information for each individual in the household (*ie.* age, sex, etc.) and characteristics at the household level (*ie.* wall material, electricity availability, etc.). Both household and individual questions were answered by a reference person who was targeted to be either the household head or the spouse/partner of the head. In over 95% of cases the respondent was the household head or the spouse as intended. If the household head or the spouse were not going to be available in a second visit, but a knowledgeable adult was willing to participate, the interview took place.

Survey weights (or expansions factors) represent the inverse of the probability that a dwelling or household is included in the sample. They are constructed taking into account the proportion of households that we attempted to interview in each cluster and cluster specific non-response. In the construction of the weights non-response is assumed to be random: it simply inflates the weight given to households in a project that were successfully interviewed. The response rate in the baseline survey was 94%.

In order to compare nominal variables between the follow-up (2009) and the baseline (2006) survey, the cumulative inflation between 2006-2009, 14.68%, is taken into account.

[4] Duflo et al. (2005) and List et al. (2009) offer a detailed discussion on optimal size arrangements.

2.3.2 Business Census

A very short business census was applied to all business units with their main entrance on the street project in 2006 and 2009. Mobile business units were excluded from analysis (for example, a seller on a motorcycle, or a water distributor going around on a truck). The supervisors of each survey team were in charge of locating all business units on their street project and administering the questionnaire. The questions included the type of business, years of operation, employees, total sales, expenditures, changes in profits, and whether the business unit is located in a house, a special purpose commercial locale, or on the street.

2.3.3 Housing Value Assessments

Finally, the other source of data in the study was the housing value assessments produced by a real estate agent. As pointed out by Gonzalez-Navarro and Quintana-Domeque(2009), having only one real estate agent performing all the assessments has the positive feature that heterogeneity of assessment practices, which require a lot of subjective decision-making, is minimized. The real estate agent was asked to visit one out of every two successfully interviewed homes and to assess the market value of the house in 2006 and 2009. In each year, the assessments were performed once the household survey fieldwork was completed.

3 Baseline Descriptive Statistics

3.1 Baseline Household Characteristics

This section offers a detailed picture of the experimental streets in 2006. Table 3 provides descriptive statistics. The survey obtained information from 1,231 households, with an average and median of 4 members, suggesting a nuclear family rather than an extensive household with multiple generations. Only 3% of individuals speak an indigenous language. Individuals aged 15 and over have a median of 8 years of schooling (mean of 7.5 years) and 88% of them have ever attended school. Lack of schooling is mostly concentrated among the elderly. 88% of individuals aged 5 and over declare to be literate - defined as being able to read and write a note in Spanish.

Home ownership is relatively high: 84% if households declare to be owners of the property they live on. Only 71% of homeowners have a title of property. Indeed, some homeowners even declared to have acquired the property by invading it. The survey asked for an estimate of property value, and although non-response for this question is quite high, Gonzalez-Navarro and Quintana-Domeque (2009) have shown that the probability of non-response is uncorrelated with professionally appraised values of these same properties. The median house value estimate is 114,680 in 2009 Mexican pesos (13,500 PPP-adjusted 2009 US dollars). Houses in the sample are relatively simple, with a median of 2 rooms (mean of 2.3 rooms), 93% of the homes with cement floor (or hard-floor) and 92% with cement walls. Cement roof is not the norm, given that only 37% of homes have it. Asbestos or metal sheets were by far the most common form of roofing. 41% of homes have the bathroom located outside the house. 25% of households use wood or charcoal as cooking fuel. In households that cook with wood the kitchen is typically under a roof outside the main structure.

In terms of labor, 51% of the 4,099 individuals aged 8 and over worked the previous week. A person is defined as working if he or she engaged in any income generating activity or worked without pay in the family business or farm. Among those who worked, the median number of days worked was 6 and the average was 5.5 days. Work is usually 8 hours per day. Multiplying *days worked* by the number of *hours worked per day* provides a measure of weekly hours worked. With a median of 48 and an average of 46.5 hours worked last week, part-time employment does not seem to play an important role. Average monthly labor income is 3,374 pesos, with a median of 2,408 pesos (around 280 PPP adjusted 2009 US dollars). Our estimated rate of unemployment (excluding students, housewives, the elderly and anyone not looking for work) is around 6%, which compares to the 3.5% unemployment rate for Mexico as a whole in the first quarter of 2006. This suggests that the inhabitants of the city's outer neighborhoods do not seem to experience the degree of high joblessness encountered in other urban contexts (Magruder, 2009).

In the consumption panel of Table 3 we can see that household expenditure is on average 3,748 pesos, with a median of 3,211 pesos. Dividing monthly household expenditure by the number of family members provides a measure of per capita expenditure, which is on average 1,067 pesos with a median of 860 pesos (around 100 PPP-adjusted 2009 US dollars per month). The median per capita expenditure is slightly higher than the 2 dollar a day[5] poverty line (Banerjee and Duflo, 2007). In terms of durable goods, 12% of households have an automobile, and 8% have a pick up truck. The median household does not have either. In terms of other durable goods, 79% of households have a refrigerator, 51% have a washing machine, 38% have a video player, 20% have a microwave oven, 10% have a personal computer, and 6% have air conditioning. Television and radio ownership were not asked because other surveys have showed that they have been almost universally adopted in Mexican households. Participation in government welfare programs, such as Progresa-Oportunidades, DIF food aid, is positive in 7% of households.

The panel on public services shows that the distance to the nearest paved street is on average 1.4 blocks. Street blocks in Acayucan vary in size but are roughly 200 meters long. 78% of households have tap water in their lot. During the course of the fieldwork, we learned that some households do have a tap water line to their property but do not use it because they have not opened an account with the water company. These families either fetch their water from a neighbor, or use a water well. 87% of homes have a sewerage line connected to the city sewerage system. Electricity was available in practically all homes (98%). Regarding garbage collection, although the service is free and supposedly universal, only 58% of homes declared to have refuse collection services. Those without collection service either burn their refuse or take it to a street where the garbage is actually collected. 22% of homes experienced a flooding of their home in the past year. In terms of public safety, 11% of homes experienced a burglary in the past 12 months, and 62% declare to feel safe walking in their street at night.

[5]The 2 dollar a day poverty line was for 1985, which is 2.92 dollars in 2005.

The credit panel shows that 17% of households had a bank account, and 10% had a credit card. Use of collateral-based credit, such as mortgages and private bank loans was positive for 3% of individuals aged 18 and over, whereas uncollateral-based credit was positive for 0.4% of individuals (Electronic and furniture store credit, credit card, automobile loans, etc.) Credit from informal sources, such as friends and family, was relatively uncommon, as only 0.4 and 0.1% of individuals reported using credit from these sources in the past year.

In terms of schooling, 87% of children aged 5-17 were reported to be able to read and write a note in Spanish. 91% of children aged 5-17 were enrolled in school, and 21% reported missing at least 1 school day in the past month.

The Table also inquires about expectations. 45% of individuals in the sample had plans to out migrate for work reasons. Regarding the satisfaction of living in Acayucan, the average and the median on a 4 point increasing scale was 3: satisfied. When asked about 13 different kinds of home improvements performed in the past 6 months, the average number of improvements was 0.45 and the median was 0. Only 5% of households declared to have opened a business in the past year. 64% of those new businesses were located at home.

3.2 Baseline Business Characteristics

In the baseline round, 250 business units were located and successfully interviewed in the experimental streets. We present summary statistics from the business census for employment (firm size), sales, expenditures, and estimated profits. We trim the sample by 5% from above and below in terms of profit rank to eliminate the influence of extreme outliers. The median firm in the study areas has 1 worker (who is usually the owner) and an average of 1.66 workers. The largest firm had 10 workers. Median sales were 2,300 pesos, median expenditures were 1,200 pesos and median profits were 917 pesos per month (around 108 PPP-adjusted 2009 US dollars). The averages are larger but a similar picture emerges. All measures are suggestive of very small businesses. The most common business is a small shop (*tiendita*) selling goods like milk, tortillas, beer, sodas, potato chips, canned goods, candy and snacks (50% of units). In 84% of cases, the business is located in the house of the owner. The business census revealed that less than 15% of adults work in business units located in the neighborhoods under study. These neighborhoods are thus primarily for housing people whose work is located either in the downtown areas of the city or in ranches/farms located around the city. The little commerce that exists is mainly providing basic foods to neighbors of these areas.

4 Identification and Testable Implications

The identification framework contained in this section draws on Duflo et al. (2006) and Angrist and Pischke (2009). See Imbens and Angrist for a seminal discussion on identification of average treatment effects.

For expositional purposes, assume for now that the unit of randomization and the unit of analysis are the same. The important thing to keep in mind is that the standard errors of our estimates must be clustered at the unit of randomization level to account for intra-cluster correlation. We also use household weights.

4.1 ATET: Average Treatment Effect on the Treated

In our analysis, the treatment is defined at the street level (being paved or not) and it is described by a binary random variable, $D_i = \{0, 1\}$. The outcome of interest is denoted by Y_i. The question is whether Y_i is affected by the treatment. To address this question, we use the potential outcomes framework (Rubin, 1974). Hence, for any individual or household living in a street there are two potential outcome variables: Y_i^0 is the outcome of an individual or household had his street not been paved, irrespective of whether it actually was, and Y_i^1 is the individual's or household's outcome if his street is paved. We would like to know the difference between Y_i^1 and Y_i^0, which can be said to be the causal effect of paving the street for an individual or household i. The observed outcome, Y_i^0, can be written in terms of potential outcomes as

$$Y_i = Y_i^0 + (Y_i^1 - Y_i^0) \cdot D_i$$

where $Y_i^1 - Y_i^0$ is the causal effect of pavement for an individual or household. Because we never see both potential outcomes for any one individual or household, we must learn about the effects of pavement by comparing the average outcome of those whose streets were and were not paved. The comparison of average outcome conditional on treatment status is formally linked to the average causal effect by the equation

$$\underbrace{E[Y_i|D_i = 1] - E[Y_i|D_i = 0]}_{Comparison} = \underbrace{E[Y_i^1|D_i = 1] - E[Y_i^0|D_i = 1]}_{ATET} + \underbrace{E[Y_i^0|D_i = 1] - E[Y_i^0|D_i = 0]}_{SB}$$

The term $E[Y_i^1|D_i = 1] - E[Y_i^0|D_i = 1] = E[Y_i^1 - Y_i^0|D_i = 1]$ is the average causal effect of treatment on those who were treated (ATET). The term $E[Y_i^0|D_i = 1] - E[Y_i^0|D_i = 0]$ is called selection bias (SB). This term is the difference in average Y_i^0 between those who were and those who were not treated. Random assignment of D_i solves the selection problem because random assignment makes D_i mean independent of potential outcomes:

$$E[Y_i^0|D_i = 1] - E[Y_i^0|D_i = 0] = E[Y_i^0] - E[Y_i^0] = 0$$

Hence, under random assignment of D_i:

$$\underbrace{E[Y_i|D_i = 1] - E[Y_i|D_i = 0]}_{Comparison} = \underbrace{E[Y_i^1 - Y_i^0|D_i = 1]}_{ATET}$$

Let Z_i be the random assignment to treatment versus no-treatment. In our experiment, $D_i \neq Z_i$. Although non-compliance can be in two-directions, in many randomized trials, such as job training programs (e.g., JTPA), only one-sided non-compliance occurs. On the one hand, participation is voluntarily among those randomly assigned to receive treatment, $D_i = \{0, 1\}$ if $Z_i = 1$. On the other hand, no one in the control group has access to the experimental intervention, $D_i = 0$ if $Z_i = 0$. Since the group that receives (i.e. complies with) the assigned treatment is a self-selected subset of those offered treatment, comparison between those actually treated and the control group is misleading. The selection bias in this case is almost always positive; those who take advantage of randomly assigned economic interventions such as training programs tend to earn more anyway (Angrist and Pischke, 2009).

In general, using IV in a randomized trial with one-sided non-compliance allows us to estimate the ATET, (Bloom, 1984). The IV estimate is obtained by regressing the outcome of interest on the treatment, where the latter is instrumented by assignment status[6].

$$E[Y_i^1 - Y_i^0 | D_i = 1] = \frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[D_i|Z_i = 1]}$$

In the first part of our experimental analysis we need to show that we can actually estimate the ATET. In other words, we need to offer evidence that randomization successfully balanced subjects' characteristics across the intent-to-treat (ITT) and control groups. To do that, we compare pretreatment (observable) characteristics X_i across groups. If we do not find systematic differences in mean (observable) characteristics between the ITT and control groups before the intervention, the assignment to the ITT group is random, and hence we have a valid instrument to identify the ATET[7]. Hence, our first testable implication is the following:

Testable Implication 4.1 (ATET Identification: based on baseline characteristics) *If the ITT and control groups have the same mean pre-treatment characteristics, the groups are balanced, and we have a valid instrument to identify the ATET. The ATET is identified if*

[6] Frölich and Blaise (2008) show that if additional control variables are included in the model, treated and compliers are not identical, and ATET $\neq$ LATE. They discuss several reasons for doing so. First, when the treatment is randomly assigned but the assignment probability differs between individuals. Second, non-response and attrition are universal problems of most randomized trials, particularly when one is interested in medium to long-term effects of a treatment. Third, when including additional covariates to separate direct from indirect effects. Finally, when the instrumental variable has not been randomly assigned and therefore might be confounded, unless we condition on several background characteristics. None of these scenarios apply to our case.

[7] The assumption being that if there are no mean differences in observable characteristics, there will be no mean differences in unobservable characteristics.

we cannot reject H_0:

$$H_0 : E[X_i|Z_i = 1] = E[X_i|Z_i = 0]$$

$$H_1 : E[X_i|Z_i = 1] \neq E[X_i|Z_i = 0]$$

We provide evidence supporting the identification of the ATET in Table 5 (Section 5).[8]

4.2 ATE: Average Treatment Effect

As discussed above, random assignment of D_i solves the selection problem. Further, ATE=ATET:

$$\underbrace{E[Y_i|D_i = 1] - E[Y_i|D_i = 0]}_{Comparison} = \underbrace{E[Y_i^1 - Y_i^0|D_i = 1]}_{ATET} = \underbrace{E[Y_i^1 - Y_i^0]}_{ATE}$$

In words, under random assignment (and perfect compliance, $D_i = Z_i$), the average effect of treatment on the treated is the same as the average effect of the treatment on a random chosen individual-household.

With one-sided non-compliance, comparing OLS with IV estimates should give us the magnitude of the selection bias:

$$\underbrace{E[Y_i^0|D_i = 1] - E[Y_i^0|D_i = 0]}_{SB} = \underbrace{E[Y_i|D_i = 1] - E[Y_i|D_i = 0]}_{OLS} - \underbrace{E[Y_i^1|D_i = 1] - E[Y_i^0|D_i = 1]}_{IV}$$

In other words, since the experimental protocol is violated, we should expect to identify at most the ATET. However, in our randomized control trial clusters of individuals (streets) rather than independent individuals are randomly allocated to intervention groups: the outcome of interest occurs at the individual level whereas the randomization occurs at the cluster (street) level. Hence, in our case, one-sided non-compliance does not come from the fact that some individuals decided whether to participate in the program or not, but because

[8] The intention-to-treat effect (ITTE) is immediately identified by regressing the observed outcome of interest Y on a constant and Z. However, in our case, this effect gives us the average causal effect of being randomly selected to be paved. Given that people did not know about their assignment status, the ITTE does not seem to provide meaningful estimates.

the government could not comply in providing the randomly assigned treatment by the time we ran the follow-up survey or had not started. Hence, unless government non-compliance is related to both socio-economic characteristics of the places that could not be paved and those of the families living there, which may accidentally occur due to the non-random order in paving streets, selection bias is much less likely to be a concern.

There are three testable implications to check whether selection bias (endogeneity) is a concern:

Testable Implication 4.2 (ATE Identification 1: based on baseline characteristics) *If the ITT-treated and the ITT-untreated groups have the same mean pre-treatment characteristics, the groups are balanced, and there is no selection on pre-treatment characteristics. There is no selection on pre-treatment characteristics if we cannot reject H_0:*

$$H_0 : E[X_i|D_i = 1, Z_i = 1] = E[X_i|D_i = 0, Z_i = 1]$$

$$H_1 : E[X_i|D_i = 1, Z_i = 1] \neq E[X_i|D_i = 0, Z_i = 1]$$

Moreover, if there is no selection into the treatment within the ITT group, and given balance between ITT and control groups, we should expect to find no pre-treatment differences between paved (treated) and unpaved (control and ITT-untreated) streets.

Testable Implication 4.3 (ATE Identification 2: based on baseline characteristics) *If the treated and the untreated groups have the same mean pre-treatment characteristics, the groups are balanced, and there is no selection on pre-treatment characteristics. There is no selection on pre-treatment characteristics if we cannot reject H_0:*

$$H_0 : E[X_i|D_i = 1] = E[X_i|D_i = 0]$$

$$H_1 : E[X_i|D_i = 1] \neq E[X_i|D_i = 0]$$

Further, if we find that there is no selection on pre-treatment characteristics, we should not find statistically differences between OLS (ATE) and IV (ATET) estimates. This suggests testing the following implication:

Testable Implication 4.4 (ATE Identification 3: based on follow-up estimates) *Let us write the following outcome and first-stage equations:*

$$Y_i = \alpha + \beta \cdot D_i + u_i$$

$$D_i = \gamma + \pi \cdot Z_i + e_i$$

where D_i is the potentially endogenous treatment and Z_i is a valid instrument, e.g., the random assignment to treatment. Then a test of the null hypothesis that D_i is not correlated with u_i is equivalent to a test of the hypothesis that ρ equals zero in the following auxiliar regression

$$Y_i = \alpha + \beta \cdot D_i + \rho \cdot \widehat{e_i} + u_i$$

where $\widehat{e_i}$ represents the fitted residual from the first stage regression, i.e., $\widehat{e_i} = D_i - \widehat{\gamma_{OLS}} - \widehat{\pi_{OLS}} \cdot Z_i$. The term $\widehat{e_i}$ comprises the potential endogenous component of D_i that is related to Y_i, all exogenous influences being captured in D_i. Thus, the test of exogeneity would be the test of $H_0 : \rho = 0$. If we cannot reject H_0, exogeneity of D_i cannot be rejected.

Testable Implication 4.4 uses the Durbin-Wu-Hausman (DWH) test, a regression-based form of the Hausman test for the presence of systematic differences between OLS and IV estimates (Wooldridge, (2002); Cameron and Trivedi, (2009)), that under independent homoskedastic standard errors turns out to be asymptotically equivalent to the original form of the Hausman test (Hausman 1978, 1983). The DWH test produces a robust test statistic (Davidson, 2000), even under heteorskedastic errors[9]. In general, this test lacks of power due to the low correlation of the instrument with the potentially endogeneous variable. How-

[9] The Hausman test is based on the assumption that $\widehat{Var}(\widehat{\beta_{IV}} - \widehat{\beta_{OLS}}) = \widehat{Var}(\widehat{\beta_{IV}}) - \widehat{Var}(\widehat{\beta_{OLS}})$, which is correct only if $\widehat{\beta_{OLS}}$ is the fully efficient estimator under the null hypothesis of exogeneity, an assumption that is valid only under the very strong assumption that model errors are independent and homoskedastic.

ever, this is not a problem in our case, since random assignment to treatment with one-sided non-compliance satisfies the relevance condition by construction.

5 Experimental Analysis

This section is divided into three subsections: migration, baseline balance and experimental estimates. The first subsection shows that neither the intensity of out-migration or immigration, nor the characteristics of out-migrants or immigrants were affected by urban road pavement. The second and third subsections focus only on households that were surveyed in 2006 and again in 2009, stayers.

Households interviewed in 2006 can be partitioned into two groups: those who stayed in the experimental areas and those that moved out between 2006 and 2009. Due to budgetary reasons, the survey only followed up on non-mover households. Although mover households were not contacted, we do have information about them from the 2006 ASLS that allow us to understand along which dimensions they were different from those who stayed.

Field workers had maps with locations of dwellings interviewed in 2006. They approached the dwelling in 2009 with questionnaires that had a pre-filled section with identifying information about the household that was interviewed in 2006. If the head of the household was the same as in 2006, the households were matched.[10] If there were additional households in the dwelling (new households or subdivisions), they were all interviewed, but were coded as new households (with no household counterpart in 2006). Matched households allow an analysis of changes within households over time.

In the second round of the ASLS we were able to measure how the infrastructure provision changed the composition of the neighbors that inhabit the road pavement projects. The characteristics of new neighbors were assessed in two ways. First, by interviewing new households living in dwellings vacated by mover households: if a dwelling that was surveyed

[10] The exception to this rule occurs if family members are the same, but now another member is declared to be the household head.

in 2006 had a different household in 2009, the new household was interviewed (but had no household counterpart in 2006). The purpose of these surveys is to measure if neighbors attracted by the infrastructure have different characteristics than those who were already living there. However, not all new households come to live in pre-existing houses, some families build new houses upon arrival. To include them in our measurement, households living in all residential constructions built between 2006 and 2009 were also interviewed.

The matched data set thus contains three types of households: 1- Those interviewed in 2006 and 2009, 2- those interviewed in 2006 only that could not be followed because they out migrated, and 3- new households with information from the 2009 round only (which can be further subdivided into: new households replacing those that out migrated, households inhabiting new constructions, subdivisions of households from 2006, and new households that neither substituted mover households nor were part of the household in 2006).

Table 4 presents a summary of the interview results for the 2006 and 2009 matched data set. The baseline sampling frame was all inhabited residential structures with main entrance facing the proposed road pavement projects found in early 2006. Out of 1,275 inhabited residential structures selected for interview, completed interviews were obtained from 1,193 dwellings. The response rate was thus a very high 94%. In those 1,193 dwellings, 1,231 household interviews were obtained, because in a few cases, there were 2 (and even 3) households living in the same property.

The 2009 survey was intended to follow up on the 1,231 households successfully interviewed in 2006. 900 follow-up households were successfully located and interviewed. In 56 cases the family was located but refused to participate in the survey, and in 271 cases the household had moved. The household was categorized as having moved if neighbors or new dwelling inhabitants had information that the previous family had moved out. This means the 2009 ASLS survey had a recontact rate of 73% of households interviewed in 2006. The main reason for the low recontact rate was household out-migration.

In 2009, 183 new households were interviewed. 120 of the newly interviewed households were families living in the dwellings left by mover households (labeled "Substitution" in the Table). 27 families were interviewed in new inhabited residential constructions. 22 cases were family subdivisions; typically one of the sons got married, had a child, and created a new household in the same plot because food expenditure was not shared with the parents anymore. 14 cases were simply defined as new households and occurred whenever the 2006 household was contacted, but now there was an additional family in the household. For example, if a room in the property was now rented out to another family. The 2009 ASLS round obtained a total of 1,083 completed surveys.

5.1 Migration

We have just seen that by the time of the follow-up survey in 2009, 271 baseline households (originally in our sample) moved to other places, while 183 new households (originally out of our sample) arrived to the experimental streets. The out-migrant baseline households create attrition into our panel, and we need to examine whether this attrition is random or not. If attrition was random, our experimental estimates based on stayers are going to be unbiased or consistent but imprecise. However, if it was non-random, our estimates are going to be biased or inconsistent.

In order to understand the nature of attrition, Table 5 addresses two important questions:

1. Did pavement induce more outmigration?

2. Are outmigrants from paved streets different than those from unpaved streets?

To answer the first question, the top-panel of Table 5 reports OLS and IV estimates based on the following regressions:

$$OLS: \; I_{outmigrated} = \alpha_0 + \alpha_1 \cdot Pavement_i + \epsilon_{1i}$$

$$IV: \; I_{outmigrated} = \alpha_2 + \alpha_3 \cdot \widehat{Pavement}_i + \epsilon_{2i}$$

where $I_{outmigrated} = 1$ if the household out-migrated and, in the IV regression, $Pavement{=}D_i$ is instrumented with random assignment (Z_i). The bottom-panel answers the second question by reporting OLS and IV estimates based on the following regressions:

$$OLS: \; Y_i^{2006} = \beta_0 + \beta_1 \cdot Pavement_i + e_{1i}$$

$$IV: \; Y_i^{2006} = \beta_2 + \beta_3 \cdot \widehat{Pavement}_i + e_{2i}$$

where $Y_i^{2006} \in \{log(PCE), log(LaborIncome), Homeowner, SumofDurableGoods\}$

The results in the top-panel show that the probability of out-migration does not depend on the street being paved. The rate of attrition is around 24%, but it is unrelated to the pavement status of the street. Further, the bottom-panel shows that those who out-migrated from paved streets were not different than those who out-migrated from unpaved streets in terms of per capita expenditure, labor income, home ownership status and durable goods. All in all, these results suggest that, if anything, attrition due to out-migration was random.

We also inquire about the role of pavement in attracting new households to the neighborhood. In this regard, Table 6 answers two important questions:

1. Did pavement induce more immigration?

2. Are immigrants to paved streets different than those to unpaved streets?

To answer the first question, the top-panel of Table 6 reports OLS and IV estimates based on the following regressions:

$$OLS: \; I_{immigrated} = \alpha_4 + \alpha_5 \cdot Pavement_i + \epsilon_{3i}$$

$$IV: \; I_{immigrated} = \alpha_6 + \alpha_7 \cdot \widehat{Pavement}_i + \epsilon_{4i}$$

where $I_{immigrated} = 1$ if the household immigrated and, in the IV regression, $Pavement{=}D_i$ is instrumented with random assignment (Z_i). The bottom-panel answers the second question by reporting OLS and IV estimates based on the following regressions:

$$OLS: \ Y_i^{2009} = \beta_4 + \beta_5 \cdot Pavement_i + e_{3i}$$

$$IV: \ Y_i^{2009} = \beta_6 + \beta_7 \cdot \widehat{Pavement}_i + e_{4i}$$

where $Y_i^{2009} \in \{log(PCE), log(LaborIncome), Homeowner, SumofDurableGoods\}$

Interestingly, the results show that the probability of immigration does not depend on the street being paved. The rate of immigration is around 17% and is unrelated to the pavement status of the street.[11] Moreover, immigrant households to paved streets were not different than those who migrated to unpaved streets in terms of per capita expenditure, labor income, home ownership status and durable goods (except by a statistically significant coefficient in the OLS for durable goods). Overall, we find evidence that neither pavement attract a higher fraction of immigrants, nor immigrants were different between paved and unpaved streets.

To sum up, we can focus our experimental evaluation on the stayers, without expecting any bias due to either out-migrant-based-attrition or masked effects by differential immigration flows to paved and unpaved streets.

5.2 Baseline Balance

In order to test if ATET and ATE are identified, Table 7 presents average baseline characteristics for three different groups: ITT versus Control, Treated (ITT & treated) versus ITT & untreated, and Treated versus Untreated. Standard errors are calculated using the survey weights and clustering at the road pavement project level.[12]

[11] The difference between out-migration and migration rates must be taken with caution, since for the surveyors it was much easier to identify out-migrant than immigrant households.

[12] Following Deaton (2009), an alternative test of equality of means is a two sample t-test with unequal variances between groups using Welch's (1947) approximation. This alternative provides a solution to the Fisher-Behrens problem of testing the significance of the difference between the means of two normal populations with different variances. The standard errors using this alternative test were very similar.

The Table reports baseline characteristics by treatment status for 56 indicators of demographic characteristics, housing quality, credit, labor, consumption, public services, schooling of children, and business units characteristics. We find evidence of balanced characteristics across ITT and control groups before the intervention. Only in 1 out of 56 cases (1.8%), the differences are statistically significant: labor income in the ITT group appears to be 18% higher than in the control group at the 10% significance level. Hence, we cannot reject H_0 of testable implication 4.1: Z_i is a valid instrument and ATET is identified.

The comparison of average characteristics for the second group (ITT treated versus ITT untreated) show that only 7 out of 56 mean differences are statistically significant: 5 with $p{-}value < 0.1$ and 2 with $p{-}value < 0.05$. Roughly speaking, this means that we get around 10% statistically significant differences. This suggests that we cannot reject H_0 of testable implication 4.2: it does not seem to be selection into treatment based on pre-treatment characteristics of the ITT group and ATE appears to be identified.

Given that there is balance in pre-treatment characteristics between ITT and control groups and between ITT treated and ITT untreated groups, we should expect to find evidence of balanced characteristics across treated and untreated groups before the intervention. Indeed, only in 2 out of 56 cases (3.6%), the differences are statistically significant: dwellings in the treated group are almost 11% more likely to have tap water connection in lot (p-value<0.1) and gas delivery service appears to be 7.4% more common in treated streets (p-value<0.01). This suggests that we cannot reject H_0 of testable implication 4.3: ATE appears to be identified.

Overall, our baseline balance findings suggest that: (i) we have a valid instrument to identify the ATET, (ii) ATET=ATE, and (iii) both OLS and IV estimates should provide similar estimated effects. In the next subsection we report our experimental estimates and we provide a test for detecting systematic differences between OLS and IV estimates (ATE Identification 3).

5.3 Experimental Estimates

5.3.1 Household Survey Estimates

We present our main experimental estimates for different outcomes in Table 8. In the first two columns, we report OLS estimates without and with the lagged outcome variable as a regressor. Adding the lagged outcome variable as a control variable is standard in the impact evaluation literature (Imbens and Angrist, 1994;Duflo et al., 2006; Kling, Liebman, and Katz, 2007) in order to reduce the standard error on the coefficient of interest. Columns (3) and (4) report the corresponding IV estimates, where the treatment variable is instrumented with the treatment status assignment. Finally, the last column provides the mean of the outcome variable for the control group in 2009. All regressions include a constant term, use the survey weights and standard errors are clustered at the street project level to account for intra-street correlation.

The top panel in Table 8 focuses on housing indicators. The first thing to note is that the distance to the nearest paved street in terms of street blocks decreased by around 0.7. Home ownership was not affected by the treatment. The log home owner estimate of housing value did go up because of treatment. People estimate the properties on paved roads to be worth about 24% more than without pavement, while the increase according to the real estate agent's home valuation is around 15%.

As in many developing country contexts, Acayucan households improve and expand their house over time. Hence, home characteristics at any point in time provide a measure of cumulative investments in the house. We asses differences in housing quality by treatment group using a set of house quality indicators.[13] However, in the short run, we find no evidence of changes in the overall housing stock characteristics, as measured by quality of flooring, walls and roofing, or number of rooms. Similarly, having a bathroom inside the house - a good measure of housing quality in this context - is unchanged by treatment status in

[13]Following Kling et al. (2007), for questions with multiple related outcomes, such as construction materials in the house, durable goods in the household and improvements to the house, we use a summary index of outcomes by adding up dummy variables.

the short run. Nevertheless, we do find differences in the number of home improvements made in the last 6 months: households in paved streets appear to be involved in more home improvement, such as floor improvement, plumbing, electrical, toilets, remodeling, and air conditioning, than households in control streets. Also, they are more likely to have bought material for home improvement in the last 6 months.

Table 8 also shows that collateralized credit composed of mortgages and private bank loans increases with pavement status. In particular, individuals in paved street projects are more likely to have collateral-based credit than individuals from unpaved streets. Not only are more individuals using collateralized credit, the average credit size is around three-four times as large in paved than in unpaved streets. Indicators for the household having a credit card or a bank account do not show any effect of the intervention. Access to non-collateral credit (and its amount) does not seem to respond to the pavement intervention either. Also, credit from family and friends is unaffected by pavement (see Table 13 in the Appendix).

Looking at labor variables, there is some weak evidence of a labor supply increase due to the pavement intervention. More interestingly, are the results on labor market expectations and motor transport to work: first, households in paved streets are 7-11% less likely to have a member planning to migrate for work reasons than those in unpaved ones; second, household heads in paved streets are more likely to use a motor transport to go to work. Results on satisfaction living in Acayucan was unchanged by paving the street (see Table 13 in the Appendix).

In terms of consumption, our results suggest that paving the street was reflected in higher household per capita expenditure (PCE). The estimated differences in columns (2) and (4) are 8% (p-value<0.1) and 10% (p-value=0.103), respectively. These magnitudes are in line with Khandker et al. (2009), who find an increase in household per capita consumption of 8-10% due to rural road improvement. Notice that the difference in PCE is not explained by higher household participation in government welfare programs.

There is strong evidence that durable goods increased among households in paved streets.

Out of 7 durable goods, control households had an average of 2.4 goods. Treated households had around 0.21 more durable goods according to column (2) and 0.26 more goods according to column (4). Also, car-truck indicator is higher for households in paved streets.

Finally, pavement of the street did not make burglaries more likely in treated households. Actually, members of households in paved streets were 10 percentage points more likely to feel safe while walking in their street at night than control households where only 62% felt safe walking in their street at night (see Table 13 in the Appendix). Notice also that the urban road pavement generated no changes in school enrollment or school absences among children aged 5-17.

5.3.2 Business Census Estimates

Business unit results are reported in Table 9. The top panel, labeled "intensive margin", presents regression results at the firm level. The results show that the average behavior of firms in the study area did not vary according to treatment. Neither number of employees, log sales, log expenditures, nor log profits varied by pavement status, either in the OLS or IV regressions. Although unreported in the Table, we obtained the same results for type of locale the business was in (formal independent, formal inside a residential lot, inside a house, or on the street). To determine if positive results were being masked by a temporary negative effect in streets recently paved (due to street blockages during construction) we performed tests of differences in sales, expenditures and profits according to an indicator for pavement taking place within the past 6 months and more than 6 months. We found no differences in outcomes for firms along this dimension.

The bottom panel in Table 9 reports tabulations for the sum of business units both in 2006 and 2009, to determine aggregate changes in economic outcomes by treatment status. Although the number of business units in ITT projects increased more than in control projects, both in absolute and in percentage terms, these differences were not borne out in terms of total employment. Similar results hold in a comparison of paved to unpaved areas. The business unit results suggest are somewhat unsurprising. Given the peripheral location of the street projects, pavement provision did not result in increased traffic (at least in the short run) and business units seemed to be serving the same clientele as before treatment.

To sum up, our experimental evaluation suggests that urban road pavement affected households but not business (or firms). Interestingly, this is consistent with the findings by Haughtwout (2002), who shows that in the USA the principal beneficiaries of infrastructure are property owners, not firms. Nevertheless, we should emphasize that our business estimates suffer from lack of power.

6 Understanding the Experimental Estimates

The two first order effects of pavement were an increase in home value and a reduction in the distance to the nearest paved street. In order to understand how these first order effects translated into household outcomes, such as consumption, this section explores the expected effects of road pavement from an economic perspective.

We first perform a comparative statics exercise on the changes the road pavement is expected to bring about from the perspective of a standard household utility maximization model with housing wealth and transport costs, where households are credit constraint. In the second part, we test our model predictions by means of a regression analysis.

6.1 A Simple Model

We present a simple model that captures the two direct effects of road pavement, namely, the previously documented increase in the value of properties located along paved roads, and the expected reduction in transport costs. The household's utility u depends on consumption and leisure. The utility maximization problem of the representative household is:

$$\max_{c,l,k} u(c,l)$$

$$\begin{aligned}\text{subject to:} \quad (1-\tau)f(k,h)+rH &= c+\phi k \\ h+l &= 1 \\ k &\leq \alpha H\end{aligned}$$

where c denotes consumption, l leisure, h hours of work, f is the production function of the household, k is capital, H is housing wealth, r is the return to wealth, ϕ is the rental cost of capital, and α is the fraction of housing wealth that can be used as collateral. The representative household has an endowment of time, normalized to 1. Transport costs are introduced as reducing output by a factor τ. The utility function satisfies the following properties: $u_c > 0$ and $u_l > 0$; $u_{cc} \leq 0$, $u_{ll} \leq 0$, and $u_{cl} \geq 0$. The production function satisfies the following properties: $f_k > 0$ and $f_h > 0$; $f_{kk} \leq 0$, $f_{hh} \leq 0$, and $f_{kh} \geq 0$.

Under no credit constraints, the solution to this problem is well known. Here, we focus on the credit constrained case, i.e., $k = \alpha H$. From the first order conditions:

$$u_l = (1-\tau)f_h u_c \tag{1}$$

$$\lambda[(1-\tau)f_k - \phi] = \mu \tag{2}$$

$$(1-\tau)f(k,1-l) + rH = c + \phi k \tag{3}$$

$$k = \alpha H \tag{4}$$

where $\lambda > 0$ and $\mu > 0$. Taking the total differential from equations (1), (3), and (4), and using (2), we otain:

$$dc = \beta_0 dH + \beta_1 d\tau \tag{5}$$

where

$$\beta_0 = \left(\frac{\alpha(1-\tau)^2 f_{hk} f_h u_c + [(1-\tau)(u_{cl} f_h - u_c f_{hh}) - u_{ll}][\alpha\frac{\mu}{\lambda} + r]}{(1-\tau) f_h [u_{lc} - u_{cc}(1-\tau) f_h] + (1-\tau)(u_{cl} f_h - u_c f_{hh}) - u_{ll}} \right)$$

$$\beta_1 = -\left(\frac{f_h u_c + [(1-\tau)(u_{cl} f_h - u_c f_{hh}) - u_{ll}] f}{(1-\tau) f_h [u_{lc} - u_{cc}(1-\tau) f_h] + (1-\tau)(u_{cl} f_h - u_c f_{hh}) - u_{ll}} \right)$$

and

$$dk = \alpha dH \tag{6}$$

Given the utility and production function properties, and that we are focusing on the case where all constraints are binding, i.e., $\lambda > 0$ and $\mu > 0$, then

$$\left.\frac{dc}{dH}\right|_{d\tau=0} = \beta_0 > 0$$

That is, an increase in home value increases household consumption through two mechanisms: first by a wealth effect, and second by relaxing the household credit constraint and allowing the use of more capital.

$$\left.\frac{dc}{d\tau}\right|_{dH=0} = \beta_1 < 0$$

An increase in transport costs decreases consumption for two reasons: first by a substitution effect, making labor less productive and thus raising the relative price of consumption; secondly by an income effect, less income leads to a reduction of consumption.

As mentioned above, for credit constrained households, an increase in home value leads to an increase in capital:

$$\left.\frac{dk}{dH}\right|_{d\tau=0} = \alpha > 0$$

Analogously, we can derive the implications for household labor supply. However, the predictions for labor supply are ambiguous. A reduction in household transport costs increases the relative price of leisure, hence households will increase their labor supply. However, it also makes households wealthier, reducing their labor supply. Similarly, an increase in home valuation has two different effects. On the one hand, it makes households wealthier, decreasing their labor supply. On the other hand, it relaxes household credit constraints, which may lead to an increase in labor supply via a substitution effect.

6.2 Empirical Specification and Results

The measurement of transport costs poses a major difficulty for empirically investigating the role of transportation in a wide range of economic activities (see McFadden (2007) for complexity in consumer transport decisions). Although we do not observe household transport costs τ directly, we observe the household distance to the nearest paved street d. In our context, we have reasons to think of household transport costs as depending positively on the distance to the nearest paved street.

The relationship between household transport costs and distance to the nearest paved street is postulated to be

$$\tau = \theta d + \epsilon$$

where $\theta > 0$ and ϵ contains any other determinant of transport costs. In general, the problem of using distance as a proxy for transport costs is that distance is related to other determinants that we do not observe, $cov(d, \epsilon) \neq 0$. In such a case, the effect of distance cannot be disentangled from other potential factors affecting transport costs. However, the road pavement experiment changed the distance to the nearest paved street. Hence, we have:

$$\Delta\tau = \theta\Delta d + \Delta\epsilon \tag{7}$$

Assuming that the pavement experiment only affected transport costs through the reduction in distance, $cov(\Delta d, \Delta\epsilon) = 0$.

Equations (5), (6) and (7) suggest estimating the following empirical model:

$$\Delta c = \beta_0 \Delta H + \gamma_1 \Delta d + \Delta u$$

$$\Delta k = \alpha \Delta H$$

where $\gamma_1 = \beta_1 \theta$ and $\Delta u = \beta_1 \Delta \epsilon$. If we know the sign of θ and assuming that $cov(\Delta \tau, \Delta \epsilon) = 0$, we can identify the sign of β_1. However, the fact that we do not observe directly the change in transport costs increases the residual variance. Hence, although we can identify the sign, we must be aware of the fact that, the higher is the increase in the residual variance, the lower will be the power to statistically detect the effect of the change in transport costs on the change in consumption.

Including constants in the empirical model above we have the following final specification:

$$\Delta c = \kappa_0 + \beta_0 \Delta H + \gamma_1 \Delta d + \vartheta$$

$$\Delta k = \kappa_1 + \alpha \Delta H + \nu$$

where $\vartheta \equiv \Delta \epsilon + \eta$ and ν are error terms that capture any random variation not accounted for in our model, such as measurement error in c and k. c is measured as household PCE, k is measured as household total credit amount, d is the distance to the nearest paved street, and H is the home value appraisal.

We estimate the model in two different ways. First, using OLS equation by equation. Second, we also estimate the two equations simultaneously by SUR in order to exploit the potential correlation between ϑ and ν. More specifically, we perform SUR estimation in three different ways: (i) using weights, (ii) constraining the constants terms to be zero, and (iii) using cluster-bootstrapped standard errors (500 replications).

Table 10 presents the estimates. As expected from our empirical model, we cannot reject either $\kappa_0 = 0$ or $\kappa_1 = 0$. More interestingly, β_0 and γ_1 are estimated to be positive and negative, respectively, though only the former is statistically significant. Given the point estimate of β_0, an increase in home value of $22,000$ Pesos \$, which is the increase in the average home value due to pavement (15% of the average home value in the control group, $145,000$ Pesos \$), seems to translate into an increase of 44 Pesos \$ in monthly per capita consumption.

The fact that the estimated γ_1 is not statistically significant is not surprising. As discussed above, the change in distance only captures a small part of the change in transport costs. The point estimate of γ_1 is very imprecisely estimated and fluctuates between -6 and -17, i.e., a reduction in transport costs equivalent to a one$-$block decrease in the distance to the nearest paved street is associated with an increase of per capita consumption of 6-17 Pesos \$.

The table also shows that an increase in home valuation is positively associated with an increase in capital, proxied by the total amount of credit. However, we cannot reject that $\alpha = 0$. Indeed, the point estimate is imprecisely estimated. If households were credit constrained, we should expect to find evidence of $\alpha > 0$. Unfortunately, we cannot conclude that they are. It is possible that our credit variables are very noisy measures of capital, since using other credit measures as proxies for capital, we still do not find statistically significant results, and the sign of the relationship is unstable across different proxies for capital. This non-statistically significant and unstable-sign relationship across capital and home value suggests that we cannot measure capital very well in our context.

This non-experimental analysis shows that increases in consumption are more strongly correlated with increases in housing value than reductions in transport costs, suggesting that the wealth effect generated by the road pavement was a stronger driver of consumption than the reduction in transport costs.

7 Discussion

7.1 External Effects

Our experimental estimates do not take into account the presence of externalities, in other words, households in the control group may benefit from the pavement provided to the treated group. This is because pavement reduces the distance to the nearest paved not only for houses in the treatment group but also for those in the control group. Indeed, we should expect that this reduction in the distance to the nearest paved street increases home value for houses in the control group. If that is the case, our control group is contaminated, and our previous experimental estimate on the impact of pavement on home value is likely to be downward biased.

In order to measure this external effect, we ran a regression of the change in log home value on the change in the distance to the nearest paved street and a constant. Our finding shows that a reduction in distance by one block increased average home value in the control group by 3%, albeit the effect is estimated imprecisely (p-value = 0.119). Given the experimental estimate on the effect of pavement on home value, 15%, the actual increase in home value due to pavement is around 17%. This additional 2% comes from the average increase in home value due to distance reduction for the control group, 0.7 blocks, on average. Put it differently, our previous experimental estimate was downward biased by more than 10%.

7.2 Cost-Benefit Analysis

To evaluate the desirability of the road pavement experiment, we compare the costs of providing road pavement with its benefits. The costs are measured by looking at the municipality expenditure on road paving. Specifically, in constructing the road projects, the municipality expended $11,304,642$ Pesos (2009). More decision making is involved when measuring the benefits.

In our context, it is reasonable to assume that the main benefits of the road pavement experiment are captured by the increase in home values. We also need to consider two different scenarios. The first one does not consider external effects, focusing just on the effect of pavement on home values of the treatment group. The second scenario tries to account for external effects, accounting for the average increase in the value of homes located in the control group due to the average decrease in the distance to the nearest paved street.

Table 11 reports the results of our cost benefit analysis (CBA). The top-panel in the table reports the rate of return to the road pavement project absent from external effects. Given the experimental estimated effect on the value of homes located in the treatment group, around 15% of the average home value in 2006 for the control group, we obtain a rate of return of 2.1%. However, things are dramatically different when we account for external effects. In the bottom-panel, we consider this new scenario. As we already acknowledged, the average increase on the value of homes located in the control group due to the average decrease in the distance to the nearest paved street (0.7 blocks) is around 2%. This means that the effect of pavement is 17%, not 15%. The new rate of return is estimated to be around 55%.

Before concluding one important shortcoming of our cost benefit analysis (CBA) must be emphasized: there may have been negative external effects on the set of homes that were already paved. By increasing the supply of homes with road paving in the city, homes that were already paved could have suffered a reduction in their value as a result of the road pavement experiment. Hence, a note of caution in the interpretation of our CBA is warranted given these potential general equilibrium effects.

7.3 Multiple Outcomes and Multiple Testing

In this subsection, we argue that in our experiment, the large number of measured outcomes does not raise real concerns about multiple inferences.

In general, in an experimental evaluation, significant effects may emerge simply by chance. The larger the number of tests, the easier it is to make the mistake of thinking that there is an effect when there is none, i.e., "Type I" error. The problem is well known in the theoretical literature (Romano and Wolf, (2005)), and it has recently received some attention in the policy evaluation literature (Kling et al., 2007; Anderson, (2008)).

In our experimental evaluation we are not examining many outcomes for a given dimension.[14] Rather, we are testing for the existence of differences in outcomes associated to

[14] This is what would happen, for example, in an experimental educational program where evaluators are

different dimensions across treatment and control groups.

The most common approach to adjusting p values for multiple testing is to control the familywise error rate, and the simplest way to do this is by means of the Bonferroni correction. This correction consists in multiplying each p value by the number of tests performed in each dimension. The problem with this method is its lack of power (see Anderson (2008) for a more powerful alternative technique). However, given the low-dimensionality of the outcomes evaluated in our experiment, even if we adjust the p values for multiple testing using Bonferroni's technique, our main effects are still there: the decrease in the distance to the nearest paved street and the increase in home value.

8 Conclusion

This paper provides the first experimental analysis of the effects of public infrastructure. We study the impact of randomly assigned urban road pavement in Acayucan, Mexico. Using information from a baseline survey conducted in 2006 we confirm that randomization worked as intended, balancing control and treatment (or intent-to-treat) groups. We estimate the effects of urban road pavement by means of a follow-up survey conducted in 2009.

Our findings show that urban road pavement increased home value by around 15-17%. Households in paved streets had higher access to collateral-based credit, and this amount was higher for households in paved streets. Urban road pavement also made households respond by substantially increasing car and truck ownership and making home improvements. We also find that households in paved streets had higher labor income, per capita expenditure, and consumption of durable goods, and they reduced their plans to out-migrate for work reasons.

The two first order effects of pavement were an increase in home value and a reduction

testing for differences in several scholastic achievement measures across treatment and control groups. In that case, the multiple-inference problem should be addressed, either by adjusting the p-values for each test accounting for the number of hypotheses being tested, or summarizing the different measures into an index (see Anderson, 2008).

in the distance to the nearest paved street. In order to understand how these first order effects translated into household outcomes, such as consumption and capital, we present a simple household model with housing wealth and transport costs, where households are credit constrained. A regression analysis based on the model suggests that the wealth effect is a stronger predictor of increases in consumption than the reduction in transport costs.

Finally, our Cost-Benefit Analysis suggests that the rate of return to urban road pavement is between 2% (ignoring external effects) and 55% (accounting for externalities).

These results provide evidence that lack of urban public infrastructure such as paved roads in a city slum can reduce available credit and consumption among households inhabiting those neighborhoods. We take the evidence presented here as suggestive that the lack of infrastructure can be a bottleneck in the process of development for poor countries.

References

ANDERSON, M. L. (2008): "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects," *Journal of the American Statistical Association*, 103(484), 1481–1495.

ANGRIST, J. D., AND J.-S. PISCHKE (2009): *Mostly Harmless Econometrics. An Empiricist's Companion*. Princeton University Press.

ANTLE, J. M. (1983): "Infrastructure and Aggregate Agricultural Productivity: International Evidence," *Economic Development and Cultural Change*, 31(3), 609–619.

ASCHAUER, D. A. (1989): "Is Public Expenditure Productive?," *Journal of Monetary Economics*, 23(2), 177–200.

BANERJEE, A. V., AND E. DUFLO (2007): "The Economic Lives of the Poor," *Journal of Economic Perspectives*, 21(1), 141–167.

BEAN, F. D., A. G. KING, AND J. S. PASSEL (1983): "The Number of Illegal Migrants of Mexican Origin in the United States: Sex Ratio-based Estimates for 1980," *Demography*, 20(1), 99–109.

BLOOM, H. (2006): *Learning More From Social Experiments*. Russell Sage Foundation.

BLOOM, H. S. (1984): "Accounting for No-Shows in Experimental Evaluation Designs," *Evaluation Review*, 8(2), 225–246.

BRICEÑO GARMENDIA, C., A. ESTACHE, AND N. SHAFIK (2004): "Infrastructure Services in Developing Countries: Access, Quality, Costs, and Policy Reform," *SSRN eLibrary*.

CAMERON, C., AND P. TRIVEDI (2009): *Microeconometrics using Stata*. Stata Press, College Station, Texas.

Canning, D., and E. Bennathan (2000): "The Social Rate of Return on Infrastructure Investments," *SSRN eLibrary*.

Deaton, A. (1997): *The Analysis of Household Surveys. A Microeconometric Approach to Development Policy*. John Hopkins University Press.

Dinkelman, T. (2008): "The Effects of Rural Electrification on Employment: New Evidence from South Africa," *mimeo*, Princeton University.

Duflo, E., and R. Pande (2007): "Dams," *Quarterly Journal of Economics*, 122(2), 601–646.

E. Duflo, R. G., and M. Kremer (2005): "Randomization as a Tool for Development Economists," *MIT Mimeo*.

Esfahani, H. S., and M. T. Ramirez (2003): "Institutions, Infrastructure, and Economic Growth," *Journal of Development Economics*, 70(2), 443–477.

Fan, S., and C. Chan-Kang (2005): "Road Development, Economic Growth, and Poverty Reduction in China," *IFPRI Research Report*, 138.

Frölich, M., and M. Blaise (2008): "Identification of Treatment Effects on the Treated with One-Sided Non-Compliance," *IZA DP No. 3671*.

Gonzalez-Navarro, M., and C. Quintana-Domeque (2009): "The Reliability of Self-Reported Home Values in a Developing Country Context," *Journal of Housing Economics*, 18(4).

Gramlich, E. M. (1994): "Infrastructure Investment: A Review Essay," *Journal of Economic Literature*, 32(3), 1176–1196.

Grech, V., P. Vassallo-Agius, and C. Savona-Ventura (2003): "Secular Trends in Sex Ratios at Birth in North America and Europe over the Second Half of the 20th Century," *Journal of Epidemiology and Community Health*, 57, 612–615.

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

Jacoby, H. G. (2000): "Access to Markets and the Benefits of Rural Roads," *The Economic Journal*, 110(465), 713–737.

Jalan, J., and M. Ravallion (2002): "Geographic Poverty Traps? A Micro Model of Consumption Growth in Rural China," *Journal of Applied Econometrics*, 17(4), 329–346.

Jimenez, E. (1995): "Human and Physical Infrastructure: Public Investment and Pricing Policies in Developing Countries," vol. 3 of *Handbook of Development Economics*, chap. 43, pp. 2773–2843. Elsevier.

Khandker, S., Z. Bakht, and G. Koolwal (2009): "The Poverty Impact of Rural Roads: Evidence from Bangladesh," *Economic Development and Cultural Change*, 57, 685–722.

Kling, J. R., J. B. Liebman, and L. F. Katz (2007): "Experimental Analysis of Neighborhood Effects," *Econometrica*, 75(1), 83–119.

Kneller, R., M. F. Bleaney, and N. Gemmell (1999): "Fiscal policy and growth: evidence from OECD countries," *Journal of Public Economics*, 74(2), 171 – 190.

List, J. A., S. Sadoff, and M. Wagner (2009): "So you want to run an experiment, now what? Some Simple Rules of Thumb for Optimal Experimental Design," *Carlo Alberto Working Papers Series*, (125).

Lokshin, M., and R. Yemtsov (2005): "Has Rural Infrastructure Rehabilitation in Georgia Helped the Poor?," *World Bank Economic Review*, 19, 311–333.

Magruder, J. (2009): "High Unemployment Yet Few Small Firms: The Role of Centralized Bargaining in South Africa," *UC-Berkeley ARE Mimeo.*

McFadden, D. (2007): "The behavioral science of transportation," *Transport Policy*, 14(4), 269 – 274.

Michaels, G. (2008): "The Effect of Trade on the Demand for Skill: Evidence from the Interstate Highway System," *Review of Economics and Statistics*, 90(4), 683–701.

Napier, M. (2009): *Making Urban Land Markets Work Better in South African Cities and Towns: Arguing the Basis for Access by the Poor*in Lall, Freire, Yuen, Rajack, Helluin: "Urban Land Markets: Improving Land Management for Successful Urbanization". Springer Netherlands.

Romano, J., and M. Wolf (2005): "Stepwise Multiple Testing as Formalized Data Snooping," *Econometrica*, 73, 1237–1282.

Van de Walle, D. (2002): "Choosing Rural Road Investments to Help Reduce Poverty," *World Development*, 30(4), 575–589.

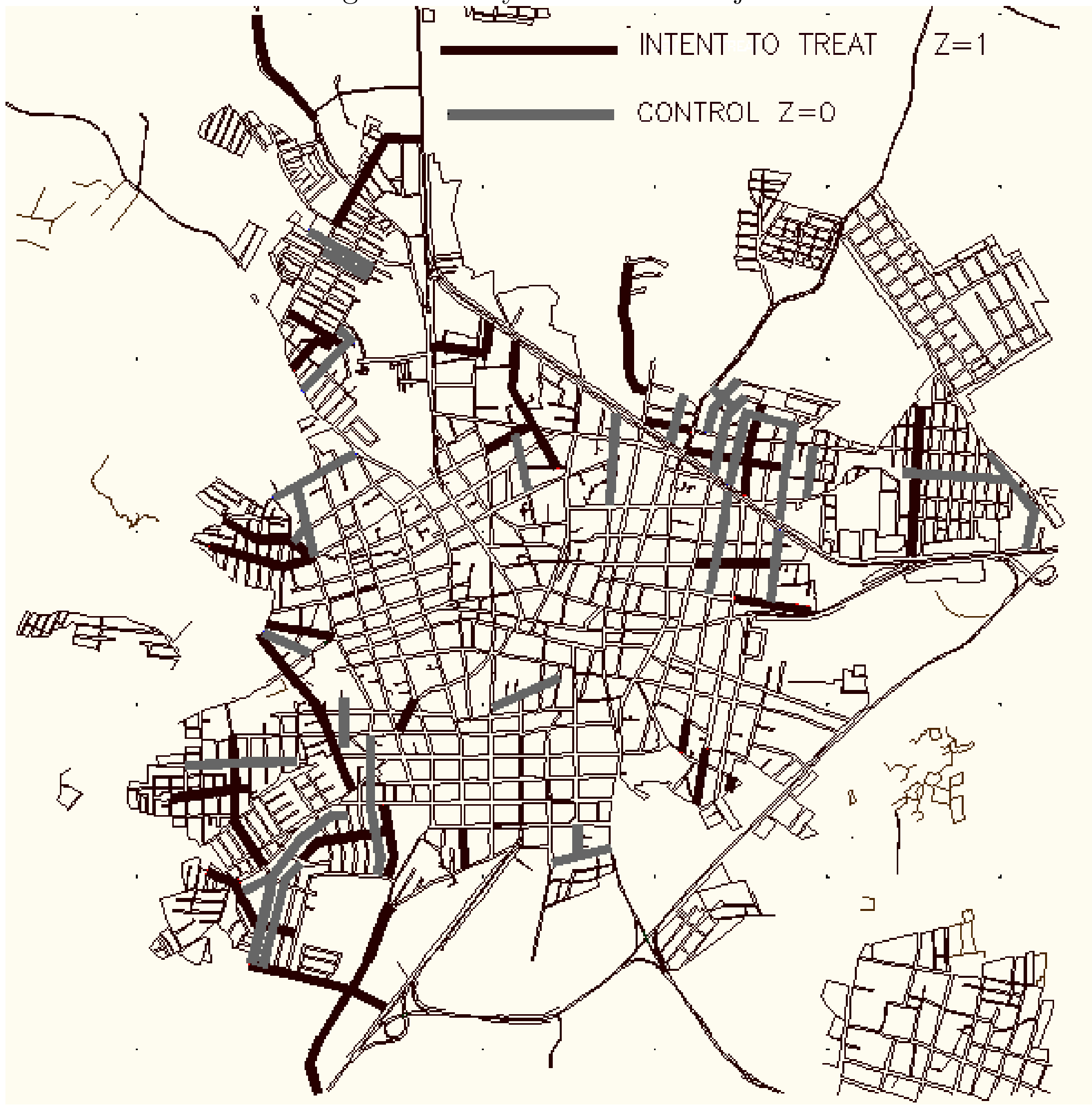
Van de Walle, D., and R. Mu (2007): "Fungibility and the Flypaper Effect of Project Aid: Micro-Evidence for Vietnam," *Journal of Development Econommics*, 84, 667–684.

Welch, B. L. (1947): "The Generalization of 'Student's' Problem when Several Different Population Variances are Involved," *Biometrika*, 34(1/2), 28–35.

Wooldridge, J. M. (2002): *Econometric Analysis of Cross Section and Panel Data*. MIT Press.

World Bank (1994): "World Development Report 1994 : Infrastructure for Development," *Washington D.C. World Bank*.

Figure 1: Acayucan Street Projects



Tables

Table 1: Comparison to City

Individual Level Variables	Acayucan (2005 Census)	Experimental Streets (ASLS 2006)	Experimental Streets (ASLS 2006, Weighted)
Population	49,945	4,943	9,088
Males/Females	89%	89%	89%
Share Aged 0-5	11%	11%	11%
Share Aged 65+	6%	5%	5%
15+ Illiterate	9%	11%	11%
6-14 Not Enrolled in School	4%	4%	4%
12-14 Not Enrolled in School	6%	7%	7%
15-24 Enrolled in School	48%	48%	48%
Household Level Variables			
Families	12,874	1,231	2,264
Dwellings	12,693	1,193	2,197
1 Room Dwelling	22%	27%	27%
2 Room Dwelling	17%	36%	36%
3+ Room Dwelling	60%	37%	37%
No Tap Water in Lot	16%	22%	21%
Electricity	98%	98%	98%
Fridge	81%	80%	80%
Washing Machine	55%	51%	52%
Computer	14%	10%	11%

First column data from locality census Iter 2005 (INEGI). Second and third column data from baseline Acayucan Standards of Living Survey 2006. Weights are the street-project inverse of sampling probability.

Table 2: ITT Road Pavement Projects Finish Date

Project Name	**Road Pavement Finish Date**
Heroes de Nacozari	Aug. 2007
Belisario Dominguez	Nov. 2007
Calabaza	Dec. 2007
Altamirano	Dec. 2007
Felipe Angeles	Dec. 2007
Salvador Allende	Dec. 2007
Ramon Corona	Dec. 2007
Porvenir	May. 2008
Guanajuato	May. 2008
Alacio Perez	May. 2008
Antonio Plaza-lado izq.	Oct. 2008
Las Arboledas	Dec. 2008
Lombardo Toledano	Feb. 2009
Antonio Plaza	Feb. 2009
David Davila y Bugambilias	Feb. 2009
Lopez Mateos	Feb. 2009
Prol. Murillo Vidal	Feb. 2009
Simon Bolivar	In process
Flores Magon	In process
Cartas Leandro Valle	In process
Gutierrez Zamora	In process
Del Arroyo y del Pantano	No progress
Ignacio Zaragoza	No progress
Prol. Atenogenes Perez y Soto	No progress
Juan de Dios Pesa-lado izq.	No progress
Veracruz	No progress
Cuahutemoc y Calle 6	No progress
Prol. Venustiano Carranza	No progress

Table 3: Baseline Descriptive Statistics (2006)

Variable	Obs.	Mean	Median	SD	Min	Max
Demographic Indicators						
Household members	1,231	4.01	4	1.80	1	16
Female (=1)	4,943	0.53	1	0.50	0	1
Age	4,939	28.0	24	19.6	0	96
Literate (=1)	4,401	0.88	1	0.33	0	1
Indigenous (=1)	4,401	0.03	0	0.15	0	1
Has ever attended school (age>15)	3,332	0.88	1	0.32	0	1
Years of schooling (age≥15)	3,289	7.52	8	4.7	0	20
Housing Quality						
Home owner (=1)	1,230	0.84	1	0.36	0	1
Property title (=1)	1,025	0.71	1	0.45	0	1
House value estimate	730	223,448	114,680	322,527	3,440	3,440,400
Cement floor (=1)	1,231	0.93	1	0.26	0	1
Cement roofing (=1)	1,231	0.37	0	0.48	0	1
Cement walls (=1)	1,229	0.92	1	0.27	0	1
Rooms (=1)	1,231	2.30	2	1.18	1	8
Bathroom inside house	1,231	0.59	1	0.49	0	1
Wood Fuel (=1)	1,221	0.25	0	0.44	0	1
Labor						
Worked last week	4,099	0.51	1	0.50	0	1
Days worked last week	2,018	5.54	6	1.55	1	7
Daily hours worked	2,031	8.10	8	3.18	0	16
Monthly labor income	1,735	3,374	2,408	3,592	0	62,500
Consumption						
Household expenditure	1,203	3,748	3,211	2,544	0	25,000
Per capita expenditure	1,203	1,067	860	846	0	9,174
Automobile	1,231	0.12	0	0.32	0	1
Pick-up truck	1,231	0.08	0	0.27	0	1
Motorcycle	1,231	0.02	0	0.15	0	1
Sum of durables	1,231	2.07	2	1.50	0	7
Refrigerator	1,231	0.79	1	0.40	0	1
Washing machine	1,231	0.51	1	0.50	0	1
Video player	1,231	0.38	0	0.48	0	1
Microwave oven	1,231	0.20	0	0.40	0	1
Computer	1,231	0.10	0	0.30	0	1
Air conditioning	1,231	0.06	0	0.24	0	1
Government program (=1)	1,231	0.07	0	0.25	0	1

Table 3: Baseline Descriptive Statistics (2006)

Variable	Obs.	Mean	Median	SD	Min	Max
Public services						
Distance to nearest paved street	1,231	1.40	1	1.31	0	9
Water in lot	1,229	0.78	1	0.42	0	1
Sewerage	1,227	0.87	1	0.34	0	1
Electricity	1,230	0.98	1	0.15	0	1
Garbage collection	1,230	0.58	1	0.49	0	1
Burn Garbage	1,230	0.15	0	0.37	0	1
Flooding (12 months)	1,211	0.22	0	0.42	0	1
Burglary (12 months)	1,224	0.11	0	0.30	0	1
Feel safe walking in your street at night	1,231	0.62	1	0.52	0	1
Credit						
Bank Account	1,221	0.17	0	0.37	0	1
Credit Card	1,223	0.10	0	0.30	0	1
Collateral based private credit	2,995	0.03	0	0.16	0	1
Uncollateralized private credit	2,995	0.04	0	0.20	0	1
Family or friends credit	2,995	0.01	0	0.08	0	1
Schooling (Ages 5-17)						
Literate	1,405	0.87	1	0.33	0	1
Enrolled in School	1,368	0.91	1	0.27	0	1
Absences Last Month (> 0)	1,243	0.21	0	0.40	0	1
Health						
Sick last month	4,851	0.46	0	0.50	0	1
Infection/parasite last year	4,851	0.15	0	0.35	0	1
Expectations, Investment						
Plans to out migrate for work	1,160	0.45	0	0.55	0	1
Home improvements (6 months)	1,231	0.45	0	1.08	0	10
New business (12 months)	1,231	0.05	0	0.21	0	1
New business at home	61	0.64	1	0.47	0	1
Satisfaction living in city	1,229	3.00	3	0.70	1	4
Business Units						
Number of employees	225	1.66	1	1.22	1	10
Sales	225	3,865	2,293	4,885	64	34,404
Expenditures	225	2,420	1,204	3,729	0	31,078
Profits	225	1,445	917	1,867	−1,950	8,715

Mean calculation takes survey weights into account, except in business units, which is a census.

Literate is being able to read and write a note in Spanish, asked for individuals 5 and older. *Indigenous* is speaking an indigenous language, asked for individuals 5 and older. *Has Ever Attended School* and *Years of Education* are for people aged 15 and older. *Property title* is not asked for renters. *House value estimate* in 2009 Mexican pesos. *Rooms* is the number of rooms in the house excluding kitchen, unless it is also used for sleeping. Labor questions are asked for people aged 8 and older. Labor statistics are calculated for the set of people who worked the previous week, except for *Worked last week*. *Hours Per Day* is coded as 0 when the person worked an average of less than 1 hour per day, and is top coded at 16 hours. *PCE* is per capita monthly expenditure in 2009 Mexican pesos at the household level. *Sum of durables* is a sum of indicators for: Refrigerator, washing machine, computer, video player, air conditioning, microwave oven, and motorcycle in the household. Government welfare programs include: Liconsa, Progresa-Oportunidades, DIF, etc. *Distance to nearest paved street* in terms of city blocks, each of around 200 meters. *Burn garbage* means the household commonly disposes of garbage by burning it. *Credit Card* and *Bank Account* are coded as 1 if anyone in the household has them. Other credit questions are asked for all adults 18 and older. Informal private credit sources are: Money lenders, merchants, and local pawn shops. Collateral based credit sources are private bank loans and mortgages. Uncollateralized credit sources are credit cards, furniture and appliance stores, automobile loans, and casas de crédito popular. Home improvements is a sum of indicators of improvements in: flooring, walls, roofing, sewerage connection, plumbing, toilets, electrical, room construction, remodeling, air conditioning installation, security measures, and improvements to house front.

Table 4: Non Response and Recontact

	2006 Dwellings		2009 Households
Eligible selected	1,275	Follow up	1,231
Completed	1,193	Completed at follow up	900
Response rate	94%	Household moved	271
		Non response	56
		Other	4
		Recontact rate	73%
		New households	183
		of which:	
		Subdivision	22
		Substitution	120
		New household	14
		New construction	27
		Completed in 2009	1,083

Eligible dwelling category excluded plots without a dwelling, unoccupied dwellings or temporary use dwellings.
2006 non response is in terms of dwellings selected from the frame, and the number of dwellings with completed household survey. 2009 recontact is in terms of households. There were 1,231 households in 1,193 dwellings in 2006, so that in some cases there is more than one household per dwelling.
Completed at follow up is defined as having recontacted at least one member of the household interviewed in 2006.
New households defined as not having been interviewed in 2006.
Subdivisions happen when one of the members in 2006 creates a new household living in the same plot, for example if the son gets married and lives in his parent's house but does not share food expenses.
Substitutions are new households found in 2009 that occupy the dwelling inhabited by an interviewed family in 2006, for example if the house is rented.
New household occurs when the interviewed family is still in the dwelling, but now there is an additional household, for example if a room in the house is now rented out.
New constructions are households interviewed in which the residential structure was not there in 2006 but is there in 2009.

Table 5: Outmigration

Outmigration Rate	$y = 1$ if household outmigrated	
	OLS	IV
Pavement	−0.010	0.013
	(0.026)	(0.044)
Constant	0.24***	0.23***
	(0.018)	(0.022)
Obs	1,171	1,171

Outmigrant Characteristics	log(PCE)		Labor Income		Homeowner (=1)		Sum of Durable Goods	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
Pavement	0.04	−0.04	0.03	−0.14	0.008	−0.05	0.29	−0.05
	(0.12)	(0.18)	(0.12)	(0.19)	(0.090)	(0.15)	(0.25)	(0.41)
Constant	6.84***	6.86***	7.88***	7.91***	0.50***	0.52***	1.86***	1.95***
	(0.05)	(0.07)	(0.06)	(0.07)	(0.05)	(0.07)	(0.13)	(0.15)
Obs	255	255	367	367	271	271	271	271

Weighted regressions, standard errors clustered at the road pavement project level.
In top panel dependent variable is a dummy for household having outmigrated by 2009, sample is households surveyed in 2006. Probit specification yields same results.
In lower panel specification in OLS columns is $y_i = \alpha_1 + \alpha_2 \cdot Pavement_i + \epsilon_i$. In IV columns, Pavement is instrumented with assignment to treatment.
PCE and *Labor income* in 2009 Mexican pesos.

Table 6: Immigration

Immigration Rate	$y = 1$ if household immigrated	
	OLS	IV
Pavement	−0.012	−0.012
	(0.023)	(0.039)
Constant	0.17***	0.17***
	(0.01)	(0.02)
Obs	1,083	1,083

Immigrant Characteristics	log(PCE)		Labor Income		Homeowner		Sum of Durable	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
Pavement	0.03	−0.06	0.15	0.12	0.06	−0.001	0.49*	0.49
	(0.12)	(0.16)	(0.15)	(0.23)	(0.12)	(0.16)	(0.26)	(0.37)
Constant	6.88***	6.90***	7.87***	7.88***	0.46***	0.47***	2.11***	2.11***
	(0.05)	(0.05)	(0.80)	(0.10)	(0.05)	(0.06)	(0.13)	(0.15)
Obs	181	181	249	249	183	183	183	183

Weighted regressions, standard errors clustered at the road pavement project level.
In top panel dependent variable is a dummy for household having immigrated by 2009, sample is households surveyed in 2009. Probit specification yields same results.
In lower panel specification in OLS columns is $y_i = \alpha_1 + \alpha_2 \cdot Pavement_i + \epsilon_i$. In IV columns, Pavement is instrumented with assignment to treatment. Sample is the 2009 round immigrants and correlates characteristics to treatment.
PCE and *Labor income* in 2009 Mexican pesos.

Table 7: Pre Intervention Balance in Means (Stayers)

	Group (1)			Group (2)			Group (3)		
Variable	ITT ($Z=1$)	Control ($Z=0$)	*Diff.*	ITT & Treated ($D=1, Z=1$)	ITT & Untreated ($D=0, Z=1$)	*Diff.*	Treated ($D=1$)	Untreated ($D=0$)	*Diff.*
Demographic Indicators									
Household members	4.09	4.13	−0.05	4.02	4.19	−0.16	4.02	4.15	−0.12
	(0.095)	(0.088)	(0.128)	(0.107)	(0.17)	(0.20)	(0.107)	(0.078)	(0.13)
	487	413	900	300	187	487	300	600	900
Female (=1)	0.52	0.54	−0.019	0.52	0.52	−0.0004	0.52	0.53	−0.015
	(0.007)	(0.011)	(0.013)	(0.009)	(0.011)	(0.014)	(0.009)	(0.009)	(0.013)
	1,997	1,716	3,713	1,212	785	1,997	1,212	2,501	3,713
Adult literate	0.88	0.86	0.018	0.88	0.88	−0.001	0.88	0.87	0.013
(=1)	(0.012)	(0.007)	(0.014)	(0.018)	(0.014)	(0.022)	(0.018)	(0.007)	(0.019)
	1,783	1,567	3,350	1,086	697	1,783	1,086	2,264	3,350
Adult schooling	7.35	7.12	0.23	7.63	6.93	0.69	7.63	7.07	0.55
	(0.36)	(0.28)	(0.45)	(0.54)	(0.39)	(0.65)	(0.54)	(0.235)	(0.57)
	1,194	1,080	2,274	722	472	1,194	722	1,552	2,274
Adult age	39.88	40.77	−0.89	40.31	39.22	1.09	40.31	40.42	−0.11
	(0.55)	(0.40)	(0.67)	(0.74)	(0.82)	(1.12)	(0.74)	(0.38)	(0.83)
	1,251	1,107	2,358	754	497	1,251	754	1,604	2,358
Home Characteristics									
Homeowner (=1)	0.93	0.94	−0.011	0.93	0.93	0.004	0.93	0.94	−0.006
	(0.018)	(0.012)	(0.022)	(0.018)	(0.04)	(0.042)	(0.018)	(0.012)	(0.022)
	487	412	899	300	187	487	300	599	899
Log owner estimate	11.71	11.80	−0.09	11.77	11.60	0.17	11.76	11.76	0.003
of house price	(0.11)	(0.09)	(0.14)	(0.14)	(0.16)	(0.21)	(0.14)	(0.08)	(0.16)
	301	290	591	198	103	301	198	393	591
Log professional	11.61	11.63	−0.015	11.67	11.52	0.14	11.67	11.60	0.064
appraisal of home	(0.07)	(0.06)	(0.091)	(0.10)	(0.10)	(0.13)	(0.10)	(0.05)	(0.11)
value	223	195	418	136	87	223	136	282	418
Number of rooms	2.34	2.37	−0.03	2.40	2.23	0.17	2.40	2.34	0.064
	(0.06)	(0.06)	(0.09)	(0.08)	(0.09)	(0.12)	(0.08)	(0.05)	(0.10)
	487	413	900	300	187	487	300	600	900
Cement roof+	2.15	2.20	−0.046	2.22	2.04	0.17	2.22	2.16	0.057
cement walls +	(0.07)	(0.05)	(0.087)	(0.088)	(0.108)	(0.14)	(0.088)	(0.048)	(0.098)
hard floor $[0-3]$	485	413	898	300	185	485	300	598	900
Bathroom inside	0.54	0.58	−0.04	0.57	0.48	0.085	0.57	0.56	0.014
house (=1)	(0.04)	(0.04)	(0.05)	(0.04)	(0.08)	(0.093)	(0.04)	(0.03)	(0.05)
	487	413	900	300	187	487	300	600	900
Water connection	0.41	0.47	−0.05	0.43	0.38	0.06	0.43	0.44	−0.011
inside house (=1)	(0.05)	(0.04)	(0.06)	(0.07)	(0.09)	(0.11)	(0.07)	(0.04)	(0.07)
	487	412	899	300	187	487	300	599	899
Distance to nearest	1.486	1.351	0.135	1.48	1.49	−0.009	1.48	1.38	0.099
paved road (blocks)	(0.16)	(0.147)	(0.216)	(0.20)	(0.26)	(0.33)	(0.20)	(0.128)	(0.233)
	487	413	900	300	187	487	300	600	900
Bought materials for	0.25	0.22	0.031	0.27	0.22	0.049	0.27	0.22	0.049
home improvement (=1)	(0.02)	(0.02)	(0.029)	(0.03)	(0.033)	(0.041)	(0.03)	(0.02)	(0.030)
	487	412	899	300	187	487	300	599	899
Number of home	0.55	0.46	0.088	0.549	0.55	−0.0007	0.55	0.48	0.067
improvements $[0-13]$	(0.05)	(0.05)	(0.072)	(0.060)	(0.097)	(0.117)	(0.06)	(0.05)	(0.076)
(6 months)	487	413	900	300	187	487	300	600	900
Tap water	0.78	0.78	−0.004	0.86	0.64	0.21*	0.86	0.75	0.109*
connection in lot (=1)	(0.05)	(0.05)	(0.067)	(0.05)	(0.10)	(0.11)	(0.05)	(0.04)	(0.063)
	487	412	899	300	187	487	300	599	900
Sewerage (=1)	0.84	0.88	−0.04	0.87	0.80	0.06	0.87	0.86	0.003
	(0.04)	(0.03)	(0.048)	(0.04)	(0.07)	(0.08)	(0.04)	(0.03)	(0.045)
	487	412	899	300	187	487	300	599	899

Table 7: Pre Intervention Balance in Means (Stayers)

Variable	Group (1) ITT ($Z=1$)	Group (1) Control ($Z=0$)	*Diff.*	Group (2) ITT & Treated ($D=1, Z=1$)	Group (2) ITT & Untreated ($D=0, Z=1$)	*Diff.*	Group (3) Treated ($D=1$)	Group (3) Untreated ($D=0$)	*Diff.*
Electricity (=1)	0.98	0.98	0.0014	0.98	0.98	−0.003	0.98	0.98	−0.0006
	(0.005)	(0.014)	(0.015)	(0.007)	(0.008)	(0.011)	(0.007)	(0.011)	(0.013)
	486	413	899	299	187	486	299	600	899
Property title (=1)	0.71	0.75	−0.043	0.71	0.70	0.006	0.71	0.74	−0.030
	(0.031)	(0.03)	(0.043)	(0.04)	(0.05)	(0.06)	(0.04)	(0.025)	(0.04)
	452	388	840	279	173	452	279	561	840
Garbage collection	0.52	0.59	−0.068	0.58	0.42	0.16	0.58	0.55	0.030
(=1)	(0.05)	(0.06)	(0.080)	(0.07)	(0.07)	(0.10)	(0.072)	(0.05)	(0.086)
	487	413	899	300	187	487	300	600	900
Gas delivery	0.955	0.914	0.040	0.989	0.903	0.08**	0.989	0.912	0.074***
service (=1)	(0.017)	(0.026)	(0.031)	(0.006)	(0.039)	(0.039)	(0.006)	(0.022)	(0.023)
	487	412	899	300	187	487	300	599	899
Cleanliness of	0.37	0.48	−0.10	0.40	0.34	0.05	0.40	0.45	−0.048
street $[1-5]$	(0.06)	(0.07)	(0.09)	(0.07)	(0.10)	(0.12)	(0.07)	(0.06)	(0.087)
	487	413	900	300	187	487	300	600	900
Cost of taxi to	29.54	30.47	−0.92	23.10	39.83	−16.73	23.10	32.63	−9.53
city center	(5.56)	(4.89)	(7.33)	(3.53)	(12.56)	(12.75)	(3.53)	(4.74)	(5.83)
	486	411	897	299	187	486	299	598	897
Credit									
Collateral-based	0.024	0.024	0.0006	0.025	0.023	0.002	0.025	0.023	0.002
credit (=1)	(0.004)	(0.006)	(0.007)	(0.007)	(0.006)	(0.009)	(0.007)	(0.004)	(0.008)
	1,215	1,089	2,304	734	481	1,215	734	1,570	2,304
Non collateral	0.044	0.032	0.012	0.046	0.041	0.005	0.046	0.034	0.012
based credit (=1)	(0.007)	(0.005)	(0.009)	(0.009)	(0.010)	(0.014)	(0.009)	(0.005)	(0.010)
	1,215	1,089	2,304	734	481	1,215	734	1,570	2,304
Collateral-based	535	361	173	544	522	21	543	397	146
credit amount	(200)	(127)	(234)	(283)	(279)	(393)	(283)	(117)	(301)
	1,215	1,090	2,305	734	481	1,215	734	1,571	2,305
Non-collateral	420	240	179	478	329	149	479	260	218
based credit amount	(110)	(76)	(132)	(163)	(119)	(197)	(163)	(64)	(171)
	1,215	1,090	2,305	734	481	1,215	734	1,571	2,305
Credit card (=1)	0.09	0.09	0.008	0.12	0.05	0.066*	0.12	0.08	0.041
	(0.02)	(0.01)	(0.024)	(0.034)	(0.015)	(0.04)	(0.03)	(0.009)	(0.035)
	484	410	894	298	186	568	298	596	894
Bank account (=1)	0.146	0.163	−0.017	0.18	0.094	0.083**	0.18	0.15	0.03
	(0.026)	(0.018)	(0.032)	(0.035)	(0.025)	(0.039)	(0.035)	(0.015)	(0.037)
	483	410	893	298	185	568	298	595	893
Credit from family	0.004	0.005	−0.0002	0.005	0.004	0.001	0.005	0.005	0.0004
and friends (=1)	(0.002)	(0.002)	(0.003)	(0.003)	(0.004)	(0.005)	(0.003)	(0.002)	(0.003)
	1,215	1,089	2,304	734	481	1,215	734	1,570	2,304
Informal private	0.003	0.006	−0.003	0.005	0.00	0.005	0.005	0.005	0.0006
credit (=1)	(0.001)	(0.002)	(0.003)	(0.002)	.	.	(0.002)	(0.002)	(0.003)
	1,215	1,089	2,304	734	481	1,215	734	1,570	2,304
Labor									
Work (=1)	0.603	0.596	0.006	0.611	0.589	0.022	0.611	0.595	0.017
	(0.014)	(0.017)	(0.022)	(0.021)	(0.017)	(0.027)	(0.021)	(0.014)	(0.025)
	1,127	1,001	2,128	694	433	1,127	694	1,434	2,128
Unemployed (=1)	0.048	0.072	−0.023	0.052	0.041	0.011	0.052	0.065	−0.013
	(0.010)	(0.016)	(0.019)	(0.014)	(0.016)	(0.022)	(0.014)	(0.013)	(0.019)
	614	548	1,162	383	231	614	383	779	1,162
Daily hours	8.39	8.19	0.201	8.297	8.534	−0.236	8.297	8.262	0.035
worked	(0.174)	(0.147)	(0.226)	(0.243)	(0.215)	(0.316)	(0.243)	(0.126)	(0.267)
	523	452	975	323	200	523	323	652	975

Table 7: Pre Intervention Balance in Means (Stayers)

	Group (1)			Group (2)			Group (3)		
Variable	ITT ($Z = 1$)	Control ($Z = 0$)	*Diff.*	ITT & Treated ($D = 1, Z = 1$)	ITT & Untreated ($D = 0, Z = 1$)	*Diff.*	Treated ($D = 1$)	Untreated ($D = 0$)	*Diff.*
Monthly log labor income	7.959	7.781	0.178*	8.010	7.882	0.127	8.010	7.803	0.207
	(0.083)	(0.049)	(0.096)	(0.125)	(0.083)	(0.146)	(0.125)	(0.043)	(0.128)
	420	390	810	254	166	420	254	556	810
Head motor transport to work (=1)	0.624	0.549	0.076	0.660	0.554	0.106	0.660	0.550	0.110
	(0.045)	(0.060)	(0.074)	(0.054)	(0.079)	(0.094)	(0.054)	(0.048)	(0.072)
	181	111	292	120	61	181	120	172	292
Plans out migration for work (=1)	0.42	0.42	−0.002	0.43	0.401	0.027	0.43	0.42	0.012
	(0.03)	(0.02)	(0.036)	(0.044)	(0.025)	(0.050)	(0.044)	(0.02)	(0.046)
	456	286	842	278	178	456	278	564	842
Business opening last year (=1)	0.053	0.036	0.017	0.058	0.047	0.011	0.058	0.039	0.019
	(0.012)	(0.009)	(0.015)	(0.018)	(0.015)	(0.023)	(0.018)	(0.007)	(0.019)
	487	413	900	300	187	487	300	600	900
Consumption									
Log per capita expenditure	6.75	6.67	0.080	6.77	6.70	0.071	6.77	6.67	0.099
	(0.066)	(0.048)	(0.080)	(0.099)	(0.069)	(0.117)	(0.10)	(0.040)	(0.104)
	470	408	878	292	178	470	292	586	878
Sum of durable goods in household $[0-7]$	2.10	2.06	0.042	2.24	1.88	0.35	2.24	2.02	0.218
	(0.14)	(0.078)	(0.161)	(0.219)	(0.128)	(0.25)	(0.219)	(0.066)	(0.226)
	497	413	900	300	187	487	300	600	900
Sum of car and truck $[0-2]$	0.173	0.202	−0.029	0.200	0.128	0.072	0.200	0.185	0.015
	(0.039)	(0.028)	(0.048)	(0.062)	(0.021)	(0.066)	(0.062)	(0.022)	(0.065)
	487	413	900	300	187	487	300	600	900
Government welfare program (=1)	0.071	0.084	−0.013	0.075	0.064	0.010	0.075	0.079	−0.004
	(0.018)	(0.015)	(0.023)	(0.025)	(0.025)	(0.035)	(0.025)	(0.013)	(0.027)
	487	413	900	300	187	487	300	600	900
Satisfaction living in city $[1-4]$	2.99	3.05	−0.061	2.95	3.05	−0.107	2.95	3.05	−0.103
	(0.045)	(0.05)	(0.067)	(0.060)	(0.068)	(0.090)	(0.060)	(0.04)	(0.072)
	485	413	898	300	185	485	300	598	898
Public Safety									
Burglary in past 12 months (=1)	0.11	0.11	−0.004	0.103	0.12	−0.016	0.103	0.115	−0.012
	(0.016)	(0.016)	(0.022)	(0.019)	(0.02)	(0.027)	(0.019)	(0.013)	(0.022)
	486	412	898	300	186	486	300	598	898
Feels safe walking in street at night (=1)	0.62	0.61	0.017	0.67	0.55	0.12*	0.67	0.59	0.075
	(0.03)	(0.03)	(0.041)	(0.026)	(0.06)	(0.062)	(0.026)	(0.025)	(0.036)
	487	413	900	300	187	487	300	600	900
Vehicle stolen or vandalized (=1) (12 months)	0.069	0.020	0.049	0.034	0.164	−0.13*	0.034	0.050	−0.016
	(0.036)	(0.019)	(0.040)	(0.029)	(0.070)	(0.072)	(0.029)	(0.023)	(0.036)
	65	46	111	47	18	65	47	64	111
Schooling of Children (Age 5-17)									
Age children (=1)	9.10	9.46	−0.36	9.20	8.94	0.26	9.20	9.33	−0.136
	(0.27)	(0.22)	(0.34)	(0.37)	(0.37)	(0.51)	(0.37)	(0.19)	(0.402)
	744	607	1,351	457	287	744	457	894	1,351
Literate (=1)	0.86	0.88	−0.011	0.85	0.89	−0.04	0.85	0.88	−0.028
	(0.02)	(0.01)	(0.023)	(0.03)	(0.02)	(0.03)	(0.03)	(0.01)	(0.030)
	568	478	1,046	352	216	568	352	694	1,046
Enrollment in school (=1)	0.94	0.93	0.008	0.94	0.95	−0.01	0.94	0.94	−0.0005
	(0.01)	(0.01)	(0.016)	(0.016)	(0.01)	(0.021)	(0.015)	(0.01)	(0.018)
	556	471	1,027	344	212	556	344	683	1,027
Absences>0 last month (=1)	0.19	0.18	0.002	0.20	0.17	0.029	0.20	0.18	0.017
	(0.01)	(0.02)	(0.026)	(0.02)	(0.02)	(0.030)	(0.02)	(0.02)	(0.026)
	522	432	954	322	200	522	322	632	954

Table 7: Pre Intervention Balance in Means (Stayers)

	Group (1)			Group (2)			Group (3)		
Variable	ITT ($Z = 1$)	Control ($Z = 0$)	*Diff.*	ITT & Treated ($D = 1$, $Z = 1$)	ITT & Untreated ($D = 0$, $Z = 1$)	*Diff.*	Treated ($D = 1$)	Untreated ($D = 0$)	*Diff.*
Health									
Sick last month	0.48	0.46	0.017	0.50	0.43	0.067*	0.50	0.45	0.048
	(0.02)	(0.02)	(0.029)	(0.025)	(0.03)	(0.036)	(0.025)	(0.017)	(0.030)
	1,950	1,690	3,640	1,184	766	1,950	1,184	2,456	3,640
Fungus, parasites	0.137	0.16	−0.023	0.148	0.119	0.028	0.148	0.152	−0.003
skin infections	(0.014)	(0.02)	(0.021)	(0.016)	(0.026)	(0.031)	(0.016)	(0.013)	(0.021)
	1,950	1,690	3,640	1,184	766	1,950	1,184	2,456	3,640
Business Unit Census									
Number of employees	1.78	1.56	0.22	1.83	1.68	0.16	1.84	1.59	0.25
	(0.13)	(0.10)	(0.16)	(0.19)	(0.14)	(0.25)	(0.20)	(0.08)	(0.22)
	102	123	225	64	38	108	64	161	225
Log sales	7.72	7.62	0.10	7.77	7.64	0.13	7.77	7.62	0.15
	(0.14)	(0.12)	(0.19)	(0.18)	(0.30)	(0.37)	(0.18)	(0.12)	(0.22)
	102	123	225	64	38	102	64	161	225
Log expenditures	7.19	7.00	0.18	7.14	7.29	−0.15	7.14	7.07	0.06
	(0.17)	(0.15)	(0.23)	(0.25)	(0.22)	(0.34)	(0.25)	(0.13)	(0.28)
	98	117	215	63	35	98	63	152	215
Log profits	6.89	6.89	0.005	6.92	6.85	0.075	6.92	6.88	0.04
	(0.13)	(0.13)	(0.18)	(0.15)	(0.31)	(0.36)	(0.15)	(0.12)	(0.20)
	94	107	201	60	34	94	60	141	201

Coefficients from OLS regressions using survey weights. Standard errors clustered at the road pavement project level
Coefficients from OLS regressions using survey weights. Standard errors clustered at the road pavement project level
Business units regressions use clustered standard errors at the road pavement project level. Business units analysis includes all firms with complete information from 2006 with a 5% trimming according to profit rank from above and below. Expenditures Sales and Profits in terms of 2009 Mexican pesos.
Literate is defined as being able to read and write a note in Spanish, and is asked for people aged 5 and older.
Adult is defined as being aged 18 and older.
PCE is per capita monthly expenditure in Mexican pesos at the household level.
Estimate of house value in 2006 Mexican Pesos.
Number of Rooms is the number of rooms in the house excluding kitchen, unless it is also used for sleeping.
Informal private credit sources are: Money lenders, merchants, and local pawn shops.
Collateral based credit sources are private bank loans and mortgages. Uncollateralized credit sources are credit cards, furniture and appliance stores automobile loans, and casas de crédito popular.
Credit Card and *Bank Account* are coded as 1 if anyone in the household has them. Other credit questions are asked for all adults 18 and older.
Durable goods in household is a sum of dummies for having: Refrigerator, washing machine, computer, video player, air conditioning, microwave oven, and motorcycle.
Government welfare programs include: Liconsa, Progresa-Oportunidades, DIF, etc.
Labor questions are asked for people aged 8 and older. Labor statistics are calculated for the set of people who worked the previous week, except for *Worked last week*. *Hours Per Day* is coded as 0 when the person worked an average of less than 1 hour per day, and is top coded at 16 hours. Weekly hours worked is a multiplication of *hours per day* and *days worked last week* for each individual that works.
Home improvements is a sum of indicators for improving: flooring, walls, roofing, sewerage connection, plumbing, toilets, electrical, room construction, remodeling, air conditioning, security measures, and house front.
Distance to nearest paved street in terms of city blocks, each of around 200 meters.
Satisfaction with Government on a 4 point scale where: 1 is very unsatisfied, 2 is unsatisfied, 3 is satisfied and 4 is very satisfied.

Table 8: Impacts on Stayers

Variable	OLS	OLS+LO	IV	IV+LO	Mean Control 2009
Home Characteristics					
Distance to nearest paved street	−0.623***	−0.651***	−0.636***	−0.709***	0.645
(in number of street blocks)	(0.068)	(0.076)	(0.153)	(0.124)	(0.069)
	893	893	893	893	407
Homeowner (vs renter) (=1)	−0.009	−0.001	−0.030	−0.019	0.954
	(0.022)	(0.009)	(0.036)	(0.015)	(0.014)
	897	897	897	897	411
Log owner estimate of house price	0.230	0.201*	0.189	0.241*	11.99
	(0.177)	(0.102)	(0.225)	(0.143)	(0.081)
	535	535	535	535	275
Log professional appraisal	0.174	0.133***	0.110	0.146***	11.57
of house price	(0.114)	(0.038)	(0.153)	(0.047)	(0.061)
	394	394	394	394	185
Bought material for home	0.053**	0.047*	0.090*	0.084*	0.146
improvement (=1) (6 months)	(0.026)	(0.026)	(0.046)	(0.046)	(0.021)
	894	894	894	894	409
Number of home improvements	0.215*	0.207*	0.435**	0.419**	0.400
$[0-13]$ (6 months)	(0.120)	(0.118)	(0.201)	(0.200)	(0.064)
	900	900	900	900	413
Cement roof+cement walls+	0.081	0.028	−0.053	−0.016	2.25
hard floor [0-3]	(0.094)	(0.039)	(0.130)	(0.058)	(0.047)
	894	894	894	894	411
Number of rooms	0.043	0.004	−0.039	−0.015	2.43
	(0.131)	(0.094)	(0.189)	(0.137)	(0.079)
	900	900	900	900	413
Credit					
Collateral based credit (=1)	0.018	0.018	0.028*	0.028*	0.018
	(0.011)	(0.011)	(0.014)	(0.014)	(0.004)
	1,984	1,984	1,984	1,984	937
Non-Collateral based credit (=1)	0.001	−0.001	0.002	−0.001	0.069
	(0.011)	(0.012)	(0.019)	(0.019)	(0.009)
	1,984	1,984	1,984	1,984	937
Collateral based credit amount	1,627*	1,613**	1,759**	1,740**	427.1
	(816)	(799)	(827)	(811)	(92.2)
	1,984	1,984	1,984	1,984	937
Non-collateral based credit amount	233	236	412	416	716
	(421)	(424)	(577)	(581)	(178)
	1,984	1,984	1,984	1,984	937
Credit card (=1)	0.056	0.047	0.058	0.055	0.155
	(0.037)	(0.036)	(0.053)	(0.052)	(0.021)
	890	890	890	890	410
Bank account (=1)	0.059*	0.048	0.065	0.070	0.138
	(0.035)	(0.032)	(0.049)	(0.044)	(0.020)
	891	891	891	891	410

Continued on next page

Table 8: Impacts on Stayers

Variable	OLS	OLS+LO	IV	IV+LO	Mean Control 2009
Labor					
Work (=1)	−0.021	−0.029	−0.027	−0.031	0.627
	(0.026)	(0.023)	(0.037)	(0.032)	(0.015)
	2,128	2,128	2,128	2,128	1,001
Unemployed (=1)	0.004	0.006	−0.009	−0.004	0.076
	(0.023)	(0.023)	(0.031)	(0.031)	(0.014)
	1,162	1,162	1,162	1,162	548
Daily work hours	0.306	0.292	0.682**	0.543*	8.24
	(0.284)	(0.226)	(0.396)	(0.310)	(0.166)
	975	975	975	975	452
Log labor income	0.244**	0.143***	0.195	0.050	7.82
	(0.100)	(0.051)	(0.130)	(0.081)	(0.046)
	810	810	810	810	390
Plans to migrate for work (=1)	−0.065*	−0.065*	−0.106*	−0.103*	0.474
	(0.038)	(0.035)	(0.058)	(0.055)	(0.027)
	801	801	801	801	370
Head motor transport	0.220***	0.160***	0.252***	0.292***	0.492
to work (=1)	(0.068)	(0.050)	(0.089)	(0.066)	(0.042)
	292	292	292	292	111
Consumption					
Log per capita expenditure	0.122*	0.0788*	0.158*	0.101	6.73
	(0.066)	(0.038)	(0.093)	(0.063)	(0.027)
	822	822	822	822	385
Sum of durable goods in	0.381	0.209*	0.337	0.261*	2.41
household [0-7]	(0.266)	(0.114)	(0.299)	(0.153)	(0.079)
	900	900	897	897	413
Sum of car and truck [0-2]	0.127*	0.106***	0.096	0.113**	0.202
	(0.072)	(0.038)	(0.083)	(0.051)	(0.025)
	900	900	900	900	413
Government welfare program (=1)	−0.001	−0.001	−0.004	−0.004	0.033
	(0.013)	(0.013)	(0.019)	(0.019)	(0.009)
	897	897	897	897	411

Continued on next page

Table 8: Impacts on Stayers

Variable	OLS	OLS+LO	IV	IV+LO	Mean Control 2009
Public safety					
Burglary (=1)	0.010	0.011	0.048	0.049	0.060
(12 months)	(0.020)	(0.020)	(0.034)	(0.033)	(0.012)
	893	893	893	893	410
Vehicle stolen/vandalized (=1)	−0.009	−0.011	0.001	0.007	0.094
(12 months)	(0.051)	(0.052)	(0.071)	(0.072)	(0.044)
	111	111	111	111	46
Health					
Sick last month (=1)	−0.024	−0.031	−0.005	−0.008	0.523
	(0.023)	(0.023)	(0.039)	(0.039)	(0.017)
	3,152	3,152	3,152	3,152	1,445
Parasites or fungus last year	−0.004	−0.003	0.003	0.010	0.167
(=1)	(0.023)	(0.022)	(0.037)	(0.036)	(0.017)
	3,145	3,145	3,145	3,145	1,444
Schooling					
School enrollment (=1)	−0.003	0.002	0.019	0.022	0.939
	(0.021)	(0.020)	(0.030)	(0.028)	(0.013)
	700	700	700	700	313
Absenteeism last month (=1)	0.054	0.051	0.048	0.045	0.140
	(0.045)	(0.043)	(0.058)	(0.057)	0.026
	645	645	645	645	280

IV uses intent to treat assignment as the instrumental variable for getting road pavement. *LO* stands for lagged outcome included as regressor. regressions use survey weights and standard errors clustered at the street project level.
Home value estimate, *Professional appraisal* in 2009 Mexican pesos.
Rooms is the number of rooms in the house excluding kitchen, unless it is also used for sleeping.
Collateral based credit is one for mortgages and bank loans. *Non collateral basedcredit* is one for store credit (appliances, furniture, etc.), automobile loan, credit card and casa de credito popular.
Labor questions are asked for individuals aged 18-59. *Work* is one if the person worked last week or has work but is on leave, 0 otherwise (0 includes students, housewives, etc.) *Employed* distinguishes employed from unemployed (Excluding students, housewives, etc.) *Daily hours* is top coded at 16 hours.
HHD motor transport to work is one if the head of the household uses a car, bus or taxi to go to work.
Per capita expenditure at the household level in 2009 Mexican pesos, 1% trimmed from above and below.
Government welfare programs include: Liconsa, Progresa-Oportunidades, DIF, etc.
Sum of durables is a sum of indicators for: Refrigerator, washing machine, computer, video player, air conditioning, microwave oven, and motorcycle in the household.
Sick Last Month = 1 if Vomit, diarrhea, bronchitis, stomach pain, flu, fever, coughing were present in the past month.
Infection/parasite Last Year = 1 if person presented or was diagnosed skin infection, fungus in feet or hands, or intestinal parasites in the past year.
Schooling outcomes are for children aged 5-17.

Table 9: Business Unit Results

Intensive Margin (Regressions)

	OLS	IV		OLS	IV
Dep Var: Number of Employees			Dep Var: Log Sales		
Paved	0.007	−0.047	Paved	0.17	−0.14
	(0.136)	(0.219)		(0.14)	(0.25)
Constant	1.65***	1.67***	Constant	7.61***	7.71***
	(0.084)	(0.096)		(0.105)	(0.12)
Obs.	248	248	Obs.	247	247
Dep Var: Log Expenditures			Dep Var: Log Profits		
Paved	0.11	0.15	Paved	0.11	0.14
	(0.15)	(0.23)		(0.15)	(0.23)
Constant	7.20***	7.19***	Constant	7.20***	7.19***
	(0.096)	(0.12)		(0.09)	(0.12)
Obs.	243	243	Obs.	243	243

Extensive Margin (Tabulations)

	All Business Units		Difference	All Employees		Difference
	2006	2009		2006	2009	
Paved	64	77	+13	118	128	+10
Unpaved	161	171	+10	256	283	+27
ITT= 1	102	123	+21	182	202	+20
ITT=0	123	125	+2	192	209	+17

Data from a short census of all business units in the study projects.
5% trimmed sample from above and below in terms of profit rank.
Sales, *Expenses* and *Profits* in Mexican Pesos per month.
Employees defined as people working in the business unit including the owner.
Profits is obtained by subtracting *Expenditures* from *Sales*.

Table 10: Mechanisms

	OLS		SUR		SUR		SUR	
	ΔC	ΔK	ΔC	ΔK	ΔC	ΔK	ΔC	ΔK
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ΔH	0.0021***	0.0011	0.0021***	0.0011	0.0020***	0.0007	0.0020***	0.0009
	(0.0006)	(0.0018)	(0.0007)	(0.0137)	(0.0007)	(0.0136)	(0.0008)	(0.0035)
Δd	−6.05		−8.27		−33.95		−16.96	
	(28.20)		(30.97)		(24.08)		(28.70)	
Constant	62.06	192.56	60.04	192.56			49.18	30.91
	(48.13)	(554.99)	(44.99)	(720.07)			(48.56)	(436.65)
R^2	0.03	0.00	0.03	0.00	0.03	0.00	0.03	0.00
Obs	325	325	325	325	325	325	325	325

(1) and (2) are OLS weighted estimates with standard errors clustered at the street project level. (3) and (4) are SUR weighted estimates. (5) and (6) are SUR estimates constraining the constant to be 0. (7) and (8) are SUR estimates with bootstrapped standard errors.

Table 11: Cost Benefit Analysis

	Houses	Estimated house value	Estimated impact	Benefit per dwelling	Total benefits	Total costs
Ignoring externalities						
Paved	576	133,595	0.15	20,039	11,542,589	11,304,642
Rate of return: 2.1%						
Accounting for externalities						
Paved	576	131,311	0.17	22,323	12,857,984	11,304,642
Unpaved	1,622	137,781	0.02	2,893	4,693,083	0
					17,551,067	11,304,642
Rate of return: 55.3%						

Estimated house value uses professionally appraised house values to obtain an estimate of average property price and then corrects for program impacts. In the top panel $\frac{\hat{y}_{paved}}{(1+0.15)}$. In the lower panel $\frac{\hat{y}_{paved}}{(1+0.17)}$ and $\frac{\hat{y}_{unpaved}}{(1+0.02)}$.

Appendix

Table 12: Regression-Based Hausman Tests

	p-values of equality of coefficients	
	OLS=IV	OLS+LO=IV+LO
Daily hours	0.07	0.21
Log labor income	0.51	0.08
Head transportation to work	0.44	0.54
Plans to migrate for work reasons	0.30	0.35
Collateral based credit (=1)	0.33	0.32
Collateral based credit amount	0.95	0.92
Log owner estimate of house value	0.78	0.69
Log appraised house value	0.48	0.65
Bought materials for home improvement	0.28	0.28
Number of home improvements	0.18	0.20
Log per capita expenditure	0.65	0.68
Sum of durable goods	0.76	0.64
Car and truck	0.35	0.80
Distance to nearest paved street	0.93	0.61
Garbage collection	0.03	0.10

Table 13: Impacts on Stayers (Additional outcomes)

Variable	OLS	OLS+LO	IV	IV+LO	Mean Control 2009
Bathroom inside house (=1)	0.020	0.005	−0.021	0.014	0.561
	(0.065)	(0.040)	(0.092)	(0.059)	(0.037)
	894	894	894	894	411
Water connection inside house (=1)	0.057	0.058	−0.027	0.024	0.522
	(0.069)	(0.038)	(0.101)	(0.056)	(0.038)
	898	898	898	898	412
Water in lot (=1)	0.132**	0.067**	0.011	0.024	0.793
	(0.063)	(0.031)	(0.099)	(0.047)	(0.035)
	898	898	898	898	412
Sewerage (=1)	0.041	0.039	−0.019	−0.007	0.930
	(0.028)	(0.025)	(0.051)	(0.042)	(0.022)
	898	898	898	898	412
Property title (=1)	−0.035	−0.024	−0.092	−0.063	0.731
	(0.044)	(0.037)	(0.070)	(0.058)	(0.033)
	831	831	831	831	385
Cleanliness of street (increasing scale $[1-5]$)	0.148***	0.149***	0.181***	0.185***	0.733
	(0.032)	(0.031)	(0.056)	(0.055)	(0.027)
	880	880	880	880	406
Family and friends credit (=1)	0.003	0.003	0.002	0.002	0.004
	(0.003)	(0.003)	(0.005)	(0.005)	(0.002)
	1,984	1,984	1,984	1,984	937
Informal private credit (=1)	−0.001	−0.001	0.001	0.001	0.002
	(0.002)	(0.002)	(0.003)	(0.003)	(0.002)
	1,984	1,984	1,984	1,984	937
Collateral-based credit amount $\mid$ CB-credit>0 in 06 or 09	6,000	5,613	10,485*	9,868*	7,274
	(4,990)	(4,533)	(5,935)	(5,542)	(1,494)
	329	1,984	1,984	1,984	143
Credit amount	1,533	1,297	2,360*	2,023*	1,103
	(1,099)	(889)	(1,299)	(1,081)	(231)
	1,984	1,984	1,984	1,984	937
Credit amount $\mid$ credit>0 in 06 or 09	6,000	5,613	10,485*	9,868*	7,274
	(4,990)	(4,533)	(5,935)	(5,542)	(1,494)
	329	329	329	329	143
Satisfaction living in city (increasing scale $[1-4]$)	0.007	0.008	0.005	0.006	3.14
	(0.043)	(0.043)	(0.056)	(0.056)	(0.023)
	897	897	897	897	412
Business opening last year (=1)	0.021	0.021	0.039	0.039	0.040
	(0.018)	(0.018)	(0.026)	(0.026)	(0.010)
	897	897	897	897	411
Garbage collection (=1)	0.134*	0.118**	−0.020	0.024	0.707
	(0.076)	(0.055)	(0.121)	(0.087)	(0.053)
	899	899	899	899	413
Gas delivery (=1)	0.088***	0.053	−0.025	−0.051	0.940
	(0.023)	(0.016)	(0.057)	(0.043)	(0.024)
	898	898	898	898	411
Cost of taxi to city center	−1.57	−0.985**	−0.198	−0.580	18.14
	(1.03)	(0.483)	(1.67)	(0.765)	(0.697)
	889	889	889	889	407
Feel safe walking in street (=1)	0.120**	0.103**	0.050	0.047	0.623
	(0.046)	(0.482)	(0.071)	(0.066)	(0.028)
	888	888	888	888	410

IV uses intent to treat assignment as the instrumental variable for getting road pavement. *LO* stands for lagged outcome included as regressor. Regressions use survey weights and standard errors clustered at the street project level.
$Water\ in\ lot = 1$ if property has running water service, but not necessarily inside the house.
$House\ flooding = 1$ if house has suffered from flooding in the past year.
$Cost\ of\ taxi$ in 2009 Mexican pesos.