Title
Essays in Development and Health Economics

Author
Gracner, Tadeja

Publication Date
2015

Essays in Development and Health Economics

by

Tadeja Gračner

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Economics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Frederico S. Finan, Chair
Professor Paul J. Gertler
Professor Edward A. Miguel

Spring 2015

Essays in Development and Health Economics

Abstract

Essays in Development and Health Economics

by

Tadeja Gračner

Doctor of Philosophy in Economics

University of California, Berkeley

Professor Frederico S. Finan, Chair

This dissertation combines three projects, each analyzing how markets or individuals respond to different policy or environmental changes in low and middle income countries. Each study theoretically motivates the hypotheses of interest and combines novel data from a wide range of sources to test them.

The first chapter examines the role of prices of foods, rich in different nutrients in the obesity and diet-related disease epidemic in Mexico. In response to the growing epidemic of obesity and diet-related chronic diseases, a number of governments are proposing taxes designed to reduce the consumption of unhealthy foods and thereby improve health outcomes. In this paper, I provide the first estimates of the effects of price changes of foods rich in sugar on the prevalence of obesity and diet-related chronic diseases, such as diabetes and hypertension. The analysis is made possible by rich longitudinal and nationally representative micro data on food prices and objective measures of health outcomes in Mexico for the period 1996-2010. I employ a unique bar-coded level price dataset with product-specific nutritional information combined with two datasets on health outcomes: (1) a state-level administrative dataset and (2) an individual panel dataset. Exploiting plausibly exogenous within-state variation in prices over time, I show that a decrease in the price of sugar-rich foods significantly increases the prevalence of abdominal obesity, type 2 diabetes, and hypertension. In addition, the least healthy and most impatient individuals seem to be more responsive to price changes, suggesting that time preferences are an important mechanism driving the results. Overall, the effect of sugar prices on the incidence of chronic diseases is large. Since the signing of NAFTA, I estimate that the reduction in prices of sugar-rich foods explains 20 percent of the increase in diabetes.

The second chapter, joint work with Paul J. Gertler, identifies how junk food and soda taxes are passed through to consumer prices. I use a unique bar-code level price dataset with product-specific nutritional information and the introduction of junk food and soda taxes in Mexico in January 2014 to assess whether and how the excess tax burden is split between consumers and producers. Preliminary results suggests that pass-through is the strongest

for sodas, followed by snacks, candy and cakes. For these products a full pass-through materializes within six months after the introduction of the tax. On average, the tax is not fully passed through to consumer prices in the case of cereal and cookies. However, the pass through is increasing over time for the latter, reaching a full pass-through by May 2014. We also observe large variation in pass-through across cities - from over to under shifting; hence observing decreases in consumer prices as well. Using price elasticities of health from Gracner (2015), a one time introduction of junk food and soda taxes in Mexico would imply approximately one and a half percentage point decrease in abdominal obesity and between a quarter to one half of a percentage point decrease in type II diabetes prevalence, under the assumption of a symmetric response to a price change.

The third chapter, based on joint work with Paul J. Gertler, Marco Gonzalez-Navarro, and Alex D. Rothenberg, provides evidence of the effects of road quality on local economic activity using temporal variation generated by maintenance investments in roads. A long panel of firms and households allows us to shed light on the effects of road quality for pre-existing households and firms. Methodologically, we propose a new road quality instrument using a nationwide panel dataset of road surface roughness to predict road quality from temporal variation in budgets exogenously allocated to different road maintenance authorities. We first show that higher road network quality improves household consumption and income. We then show that this is partly due to job creation in the manufacturing sector. Third, we show evidence of an occupational shift from agriculture into manufacturing and higher profits for those who stay in agriculture. The gap in average income between agriculture and manufacturing employment is reduced with road quality but not eliminated. Because wages in the manufacturing sector do not change with road quality the results are consistent with dual labor markets in Indonesia.

To my parents, Frida and Jože,
and to my sister Maja.

Contents

List of Figures

List of Tables

Acknowledgments

This dissertation would not have been written as it is without tremendous support and advice of my advisors, colleagues, and especially my friends and family.

I would like to thank Frederico Finan for his support and guidance over the course of my graduate studies. Your generous, constructive and direct feedback has strengthened each chapter of this dissertation and was crucial for my professional and personal development. Your constant motivation to search for a topic I truly am passionate about is a big part of the reason of why this dissertation is what it is. Thanks to Ted Miguel for eagerly supporting and standing behind the projects I pursued, for being an inspiring teacher, and for continuous enthusiasm and encouragement. I would also like to thank Paul Gertler, who played an integral role in my becoming an economist. You have not only taught me how to pursue a research project from the beginning to the end, but also how to always keep a positive and optimistic mindset while carrying it out. You were able to see and bring the best out of me and I will be forever grateful for your guidance and infectious enthusiasm for relevant work that you do and I aspire to do myself.

Many other faculty members have been generous with their time, wisdom and kindness. Among them, I would like to thank Hillary Hoynes and Elizabeth Sadoulet in particular. I also want to thank Marco Gonzalez Navarro and Alex Rothenberg for a great and joyful learning experience while working on the third chapter of this dissertation.

I greatly appreciate the help of the National Institute of Public Health, and National Institute of Statistics and Geography in Mexico, who provided invaluable support with the data. This project would not have been possible without Etienne Gagnon and his incredible generosity with sharing his dataset with me. Thanks to Mauricio Varela for kindly sharing his data as well. I thank UC Mexus and Ad Futura for their generous funds, and Bhavna Challa, Kristy Kwak, and Cesar Augusto Lopez for their excellent research assistance.

Thanks to Helena Schweiger for your honest advice and support from well before I knew I will end up here - I will do my best to pay your favors forward. Thanks to Paolo Carmine de Salvo, my partner in crime when applying to graduate schools - who knew we will share the same city again.

I have also gained tremendously from talking to my fellow graduate students, who I share fond memories with. I would like to especially thank Aisling, Attila, David, Eric C., Fenella, Katalin, Kaushik, Marc, Marquise, Maryam, Matt, Michel, Mitar, Paolo, Santiago, Simon, and Waldo.

I would like to say special thanks to Aniko, truly an amazing friend, colleague and person. Your strength, confidence and unconditional support, from Berkeley or afar, have made me a better friend and researcher. I will forever keep fond memories of us working together in your living room (aka Fertile Grounds), eating too much gelato across the street, or making plans just about anything (just because)! On that note, thanks to Misha for being away during May 2012, and being here for the rest of this journey. I would also like to thank

Hedvig for always willingly sharing your econometrics wisdom, to Willa for always having a word of advice, to Michele for her positive spirit, support and all the pizzas we ate, and to Liang for being one of the kindest and most thoughtful people I know. Jamie and Eva - we did it!

Thank you to my brilliant housemates, who helped making Berkeley my second home. Thanks to Trucy for many years of support, and to Monica for always cheering me up. Thanks to Tarso, who calls it as it is since our first day in Berkeley. Nina, Wen and Maude - you are a treasure! There are no words to describe how much I appreciate your support over the last couple of years.

Lastly, thank you to my family. Thanks to my grandparents, Bojana, Riko, Jože and Ivan for your cheering from afar. Pieter, thank you for loving and understanding me and my crazy, and being there every step of the way. A special thanks to my big sister Maja, for being my first and best teacher and friend, for caring and reminding me to forever believe in myself. Finally, I want to thank my parents, Frida and Jože, for your unconditional love and support, for all the sacrifices you made so I can be where I am today, and for teaching me independence, empathy, and appreciation. Thank you for being the strongest people I know. I learn from the best.

Chapter 1

Bittersweet: How Prices of Sugar-Rich Foods Contribute to the Diet-Related Disease Epidemic in Mexico

1.1 Introduction

Since 1980, worldwide obesity has almost tripled and today more than 1.5 billion adults are overweight (WHO, 2008). Over the same period of time, the prevalence of diabetes and hypertension has almost doubled. Today almost ten percent of adults are diabetic and more than one third are hypertensive, and these numbers are expected to increase another twofold by 2030 (IDF, 2011). While the obese are at the greatest risk for diabetes and hypertension, another 40 percent of adults at normal weight also manifest some form of "metabolic syndrome" (Basu et al., 2013).[1] These chronic diseases account for the greatest share of premature deaths and disabilities worldwide, and the total cost of these chronic diseases in low- and middle-income countries alone is forecast to surpass seven trillion US dollars by 2030 (UN, 2011).

One of the biggest contributors to obesity and related chronic diseases has been a significant shift to unhealthy diets. In fact, the rise of the obesity and chronic disease epidemic has been commensurate with a significant increase in the price differential between healthy and unhealthy foods. This has lead not only to a substantial increase in total caloric intake, but also a shift towards consuming more calories from sugar, refined carbohydrates and fat relative to a lower intake of fiber (Cutler et al., 2003; Drewnowski and Darmon, 2005; Popkin, 1994). These observations have led some academics and policymakers to advocate for

[1] Metabolic Syndrome is defined as the simultaneous presence of three of the following five risk factors: abdominal obesity, elevated blood pressure, decreased HDL (the "good") cholesterol, elevated triglycerides, or elevated fasting glucose (USDA).

taxing products that are rich in sugar or fats as a method of redress.[2]

The effectiveness of these taxes depends on how health is impacted by changes in the prices of foods that are rich in these supposedly unhealthy nutrients. While there is some evidence that changes in relative nutrient prices do significantly alter the composition of food consumption (Dubois et al., 2013; Harding and Lovenheim, 2014), there is little rigorous evidence on the extent to which changes in the price of sugar- or fat-rich foods alter dietary intake enough to translate into a decreased prevalence of obesity and diet-related chronic diseases. The existing evidence relating food prices to obesity is weak: much of it is based on correlation studies using small and mostly cross-sectional, or short longitudinal, data sets.[3] To the best of my knowledge, there are no studies thus far relating food prices and chronic diseases.[4]

It is not evident that changes in relative prices of foods would necessarily translate into better health. Specifically, the complex preference pattern of substitutability of food items makes it difficult to unambiguously predict the effects of a relative price change on health. For instance, recent evidence shows that while increased prices of items rich in sugar unambiguously reduce sugar and total caloric intake, price increases of fatty foods[5] that decrease consumption of fat also increase soda and sugary foods intake, suggesting that fat and sugar are substitutes (Harding and Lovenheim, 2014). Moreover, even if price elasticities of food item consumption are known, mapping from consumption to health depends on the nature of the productive relationship between nutrients on health and on how existing health mediates those relationships.[6]

In this paper, I provide the first rigorous estimates of the effects of changes in the price of sugar-rich foods on obesity, abdominal obesity, diabetes, and hypertension *directly*, using nationally representative data from Mexico from 1996 to 2010. In contrast to previous research, I combine detailed nationally representative price data with objective measures of obesity and chronic diseases. Previous research on health outcomes has not had access to representative price data that can be objectively aggregated by the nutritional content of food items. Studies have typically circumvented this issue by looking at food groups as a

[2]Healthier diet habits extend one's life-span by a mean of 1.9-3.4 years (WHO, 2002). If not applied, this implies around a trillion dollars in life-year lost annually in the US alone, valuing life-years at $100, 000 (Gruber and Koszegi, 2000). Mexico launched a soda and "junk food" tax in January 2014. Denmark introduced what was known as a fat tax on items containing more than 2.3 percent saturated fat in 2011, yet abolished it one year later.

[3]Most longitudinal studies focus on a specific group, such as children through fifth grade (Sturm and Datar, 2005; Datar et al., 2004) or older adolescents (Powell et al., 2007a). In developing countries, data is mostly focused on women of childbearing age and preschoolers (Popkin et al., 2012).

[4]BMI is the only health outcome to be examined so far, with the exception of Grossman et al. (2014) who use body fat alongside BMI as the obesity measure.

[5]I use the term "fatty" prices when refering to prices of foods rich in fat.

[6]Recent research suggests that the relative overconsumption of sugar - fructose in particular - has played a critical role in the chronic-disease epidemic through its effect on insulin resistance and lower satiety (Basu et al., 2013; Reaven, 1991; Teff et al., 2009; Bremer et al., 2011; Johnson et al., 2007). Even so, several scholars attribute this epidemic to the overconsumption of calories coming from dietary fats (Bray and Popkin, 1998).

whole, and have failed to disaggregate the prices beyond the somewhat subjective grouping of "healthy" (e.g., vegetables and fruits) versus "unhealthy" foods (e.g., fast foods and sweet beverages) (Auld and Powell, 2009; Beydoun et al., 2008; Kim and Kawachi, 2006; Sturm and Datar, 2005). I overcome this obstacle by assembling a unique dataset that tracks over 25,000 retail food prices annually along with the hand-collected detailed nutritional composition of these products over a 15-year period.[7] Using cluster analysis, I divide these products into nutritionally-similar food clusters, and then construct individual price indices for foods rich in sugar, protein, fat, and fiber.[8] Since food prices are tracked continuously at the store level across 46 Mexican cities, these "nutrient" prices are almost fully comparable over time.[9] Previous research has also not had access to high-quality longitudinal data on obesity and diet-related chronic diseases. I merge my longitudinal price information with 15 years of state-level administrative data on chronic disease incidence diagnosed through the health care system and a nationally representative, individual-level panel data on health outcomes, spanning the period 2002 to 2009. The nationally representative data provides stronger external validity of the results, whereas individual level data allows for exploring the heterogeneity in results. This combined data has enabled me to utilize the variation of prices within cities and states, conditional on location and year fixed effects as the main identification strategy.

Recent developments in Mexico constitute an ideal setting for my empirical analysis. From 1996 to 2010, there has been significant variation in food prices, spatially and over time.[10] After the signing of the North American Free Trade Agreement (NAFTA) in 1994, gradually expanding import quotas, reduced tariffs, and the removal of barriers to foreign direct investments resulted in an outward shift in the supply of processed foods that are particularly rich in sugar and fat, and a substantial decrease in their prices.[11] Since food expenditures in Mexico represent more than one-third of an average family's income, these price changes played an important role in a significant shift from a traditional diet to a "Western" diet over this same period (Clark et al., 2012).[12] Simultaneously, Mexico has experienced one of the most rapid epidemiological transitions. In the course of only two decades, obesity rates in Mexico soared from 30 percent to more than 70 percent. Today, nearly one out of every five Mexican adults is estimated to be diabetic, while one out of every

[7]The longest duration of price data combined with nutritional data thus far is the US Nielsen Homescan Data, which spans a period of seven years, relating them to consumption (Harding and Lovenheim, 2014).

[8]I use a k-mean clustering algorithm, similar to Harding and Lovenheim (2014).

[9]The prices used in this literature thus far, such as prices drawn from American Chamber of Commerce Researchers Association (ACCRA) in the US, are not recorded in the same cities over time and hence, not as comparable over time. Furthermore, they are collected only for a small number of food items (e.g. the prices of only seven fruits and vegetables were surveyed.)(Powell and Chaloupka, 2009).

[10]As a source of exogenous price variation, Fletcher et al. (2010a), Fletcher et al. (2010b) and Finkelstein et al. (2010) use the changes in states' soda taxes as natural experiments, observing small effects on weight.

[11]I provide some case studies of suggestive evidence on the supply driven variation in prices spatially and over time due to variation in transportation costs, supermarket entry, or tariff policies over the observed period.

[12]Western diet tend to be rich in refined carbohydrates, namely sugar, and fat.

two is estimated to be hypertensive. In addition, diabetes is considered the number one cause of death in the country, followed by hypertension and cardiovascular diseases. Considering that these diseases account for more than two-thirds of all chronic-disease health care costs in Mexico, understanding the cause of this burgeoning epidemic is crucial (See Figure 1.1) (de Salud, 2010).

I find that the decrease in the prices of sugar-rich foods significantly increases the type 2 diabetes and hypertension incidence rates, waistline measurements, and the probability of becoming obese and abdominally obese.[13] The effect is strongest in the first year following a price change and diminishes over a period of four years. I show that changes in the prices of foods rich in other nutrients are not significantly correlated with health outcomes. I also discern that low prices of foods rich in sugar have negative effects across the entire health distribution, measured at baseline, yet the price effect is strongest for those at the highest risk for developing chronic diseases. Simple calibrations based on these estimates suggest that the decrease in sugary prices explains approximately 20 percent of the increase in diabetes prevalence in Mexico since NAFTA was signed in 1994.[14]

To help interpret these results, I develop a theoretical model which demonstrates the role of prices and time preferences in the evolution of health over time. Consistent with this theory, I provide evidence that the heterogeneity in my results is partly attributable to differences in time preferences between individuals. Individuals defined as less patient weigh present consumption of food more, while internalizing future health costs less. This results in the accumulation of worse health over time and its significantly stronger response to changes in sugar-rich food prices. These findings complement a growing body of work that focuses on the role of time preferences in weight gain. [15]

These results are robust with regard to checks that address several important concerns. One of the main threats to identification is the strongly positive within-state trend of chronic disease, alongside negative trends in the real prices of food. However, results are robust to including state, year, region-year fixed effects which control for time-varying unobservable factors that are consistent within regions, to linear state trends, and to controlling for trends by individual baseline risk for diseases. In addition, future prices of sugary foods do not

[13] Sugar-rich food-price elasticities of BMI and waistline (between -0.02 and -0.05, respectively) are most comparable to the BMI elasticity to fast food restaurant food prices (Powell et al., 2007a; Chou et al., 2005).

[14] Chou et al. (2004) find that decreased food prices explain between ten to fifteen percent of the obesity increase the US. Currie et al. (2009) show that fast food restaurants entry explains below three percent of a 10-year increase in women and adolescents' weight.

[15] Courtemanche et al. (2014) provide evidence on the cheapest calories that lead to the largest weight gains among those who are the most impatient. Fuchs (1982), Smith et al. (2005) and Chabris et al. (2008) find positive associations between impatience and obesity, and also other health behavior, such as smoking. Despite existing evidence on an inverse/positive relationship between obesity and type 2 diabetes and socioeconomic status in developed/developing countries (Sturm and Datar, 2005; Drewnowski and Specter, 2004; Wardle et al., 2002; Baum II and Ruhm, 2009; Monteiro et al., 2004), and a stronger price sensitivity in health of the poor, (Monteiro et al., 2004), I observe no such robust relationships in my data. My findings, however, are consistent with Sturm and Datar (2005); Powell et al. (2007b), which show higher price sensitivity of health for those overweight/at a higher risk for obesity.

have a systematic relationship with health outcomes. This test also addresses the concern of reverse causality.

I address the reverse causality concern further by controlling for time variant, such as income, work status, and invariant individual characteristics (e.g., tastes), by inclusion of individual fixed effects. In addition, I test whether changes in the price of sugary foods are correlated with unhealthy behavior as proxied by using a measure of smoking behavior, predictive of obesity and chronic disease (Gruber and Frakes, 2006). I find that there is no systematic relationship between changes in smoking behavior and sugary food prices. I address the concern of the widespread availability of cheap calories and local demand shocks affecting health irrespective of prices by controlling for the number of local fast food restaurants and their advertising expenditures. Additionally, there is a possibility that areas where sugary food prices fell have witnessed larger expansions in disease diagnostics than areas where sugar calories became relatively more expensive, overestimating my results. I refute this concern by estimating a placebo test with type 1 diabetes and asthma, diseases orthogonal to food prices, yet of similar diagnostic needs as type 2 diabetes and hypertension. This placebo test reconfirms that, conditional on state fixed effects, changes in sugary prices are not correlated with state characteristics.

This paper makes a number of important contributions to the literature. It is the first to provide rigorous evidence on the relationship between economic incentives and chronic diseases, in addition to obesity, in the context of a middle-income country. In these countries, the related and existing literature so far has mainly looked at the role of income and socioeconomic status (Fernald, 2007; Fernald et al., 2008; Monteiro et al., 2007; Strauss and Thomas, 1998; Monteiro et al., 2004), gender (Case and Menendez, 2009), or urbanization in obesity prevalence. Moreover, this study is one of the first to focus on health deterioration as a consequence of calorie over- rather than under-consumption due to price changes in the developing world (Pitt and Rosenzweig, 1984; Dasgupta, 1997; Thomas and Strauss, 1992).

This project is one of the first to examine the effect of prices of nutritionally similar food clusters, as opposed to thus far considered cruder classifications of healthy and unhealthy foods, and their relationship to health. The empirical finding that mainly sugary food price changes alter health outcomes complements the growing medical literature pointing to the relative harmfulness of sugar as a nutrient (Lustig, 2013; Taubes, 2007). By contributing to the debate on the ability of price changes to influence behavior and health (Gruber and Mullainathan, 2005; Evans and Ringel, 1999; Adda and Cornaglia, 2006; Wasserman et al., 1991), this paper also relates to recent evidence on proposed chronic disease management solutions, such as obesity remediation through taxes (Powell and Chaloupka, 2009; Fletcher et al., 2010b), or diabetes and obesity management by disseminating information, either through medical diagnosis (Oster, 2014), nutritional labeling (Abaluck, 2011; Bollinger et al., 2010; Downs et al., 2009) or advertising (Ippolito and Mathios, 1995). This paper also has policy implications that apply to both developing countries, where there has been an influx of cheap sugar calories and a substantial decrease in prices due to globalization (Atkin et al., 2014; Hawkes, 2006), and developed countries, where these results could apply to less affluent households, who, incidentally, are at the highest risk for obesity and related diseases

(Drewnowski, 2009).

This paper proceeds as follows. Section 1.2 provides the theoretical framework that will assists in the interpretation of my empirical findings. Section 1.3 provides the context in which the proposed research questions are answered. Section 1.4 presents the data of my research, and Section 1.5 describes the main empirical strategy with the robustness checks. In Section 1.6, I discuss the results and policy implications. I conclude in Section 1.7.

1.2 Theoretical Framework

In this section, I present a simple theoretical framework drawing on Lakdawalla and Philipson (2002), Auld and Powell (2009), and Grossman (1972) to support some of my main empirical findings. The model theoretically demonstrates the role of prices and time preferences for the evolution of health over time. I identify under which conditions cheaper calories from foods rich in a particular nutrient, such as sugar, deteriorate the consumers' health. In addition, I show that the effect of prices is stronger for individuals with already worse health, i.e., for people who are at a higher risk for developing the disease.

Consider an individual in a discrete-time environment who in each period t chooses how to allocate consumption between two kinds of foods, one being rich in nutrient n and the other one being rich in some other nutrient o.[16] I assume that consumption of foods rich in n and o is measured in calories, hence total consumption of calories equals

$$C_t = n_t + o_t.$$

Consumption of those foods yields a consumer some positive period t utility

$$u(n_t, o_t) \quad = \quad n_t^{\alpha} \cdot o_t^{1-\alpha}$$

and at the same time affects the consumer's health negatively, and $u_t' > 0$ and $u_t'' < 0$ in food consumption.[17] In particular, following Grossman (1972), the stock of health H_{t+1} evolves according to

$$H_{t+1} = (1-d)H_t + I(n_t, o_t).$$

The main idea of this equation is that people receive an endowment of health capital at birth H_0, which depreciates with age but can be raised through investments. For simplicity, I assume throughout this section that everyone is given the same stock of health at birth. Hence, $I(n_t, o_t)$ is gross investment and d is the exogenous rate of depreciation during period

[16] I call foods rich in nutrient n simply as n and likewise for foods rich in the other nutrient (o) hereafter.

[17] $\overline{I}$ denotes any individual-specific health investments that are independent from consumption, such as exercise or medical habits. This does not mean I abstract away from exercise altogether, but I assume that the individual makes exercise and consumption decisions independently.

t.[18] Furthermore, I assume that the observed subject is an individual, who is overeating at any time t, so additional food consumption has an unequivocal negative effect on their future health. To make the model as parsimonious as possible, I assume a linear form for $I(n_t, o_t)$, allowing however for the possibility that n can be relatively more harmful to health than o. λ measures the relative harmful effect of foods rich in nutrient n. In particular, the parameter $\lambda > 1$. Net investment is, thus, given by

$$I(n_t, o_t) = \overline{I} - (\lambda n_t + o_t).$$

where $G_t \equiv \lambda n_t + o_t$ is reffered to as the "*effective*" consumption, and $\overline{I}$ incorporates other investments in health (such as exercise). For notational simplicity, I hereafter simply write i for foods rich in nutrient i, where $i \in \{n, o\}$.

Then, a consumer with discount factor $\delta \in (0, 1)$ solves the following optimization problem:

$$V(H_t) \equiv \max_{n_t, o_t} u(n_t, o_t) + H_t + \delta V(H_{t+1}) \quad \text{s.t.} \quad n_t \cdot p_t + o_t \leq w_t.$$

I assume a Cobb-Douglas utility function from food consumption today, $u_t(n, o)$, with parameter $\alpha \in (0, 1)$. I denote the price of foods rich in nutrient n at time t by p_t, normalizing the price of foods rich in other nutrients to 1; w denotes the consumer's food budget.

The budget constraint of the consumer must be binding. Hence, the optimal n_t^*, o_t^* must satisfy the following first order condition of the Bellman equation for V:

$$\underbrace{u(n_t^*, w_t - p_t n_t^*)\left(\frac{\alpha}{n_t^*} - \frac{(1-\alpha)p_t}{w_t - p_t n_t^*}\right) - D\lambda + Dp_t}_{F(n_t, o_t)} = 0$$

where $D = \sum_{i=1}^{\infty} \delta^i \cdot (1-d)^{i-1}$.

Using the Implicit Function Theorem I calculate the marginal effect on nutrient n as its price changes:

$$\frac{\partial n_t^*}{\partial p_t} = \frac{\overbrace{(1-\alpha)\frac{u(n_t^*, o_t^*)}{o_t^*}(\alpha - \frac{(1-\alpha)p_t n_t^*}{o_t^*}) - \frac{w_t - 2p_t n_t^*}{w_t - p_t n_t^*}}^{=-A} - D}{\underbrace{u(n_t^*, o_t^*)\left(\left(\frac{\alpha}{n_t^*} - \frac{(1-\alpha)p_t}{o_t^*}\right)^2 - \frac{\alpha}{n_t^{*2}} - \frac{(1-\alpha)p_t^2}{o_t^{*2}}\right)}_{=X<0}}$$

[18]One could assume, though, that rate of depreciation is endogenous and a negative function of the stock of health, discussed later. Applying this in the model would not change its predictions of interest.

Note that $X = \frac{\partial F}{\partial n}(n,p)\big|_{n=n^*}$ refers to denominator, and $A + D = \frac{\partial F}{\partial p}(n,p)\big|_{n=n^*}$. I use these abbreviations throughout the model solution.

Proposition 1 shows that an increase in the relative price p_t of nutrient n improves health if and only if the relative price for nutrient n is smaller than the relative harmfulness of nutrient n for health.

Proposition 1. *Increase/decrease in price p_t improves/deteriorates health if $p_t < \lambda$. The effect is increasing in λ.*

Proof. The net effect of price change p_t on health H_{t+1} equals to:

$$\frac{\partial H_{t+1}}{\partial p_t} = \frac{\partial H_{t+1}}{\partial G_t}\left(\lambda\frac{\partial n_t^*}{\partial p_t} + \frac{\partial o_t^*}{\partial p_t}\right) = -((\lambda - p_t)\frac{\partial n_t^*}{\partial p_t} - n_t^*)$$

Since $\frac{\partial H_{t+1}}{\partial G_t} < 0$ by definition, total calories consumed will decrease when $\lambda\frac{\partial n_t^*}{\partial p_t} + \frac{\partial o_t^*}{\partial p_t} < 0$, or $\lambda\frac{\partial n_t^*}{\partial p_t} < -\frac{\partial o_t^*}{\partial p_t}$. This is true when $p_t < \lambda$. Hence, this condition is less restrictive in the case when allowed for foods rich in different nutrients to have differentially harmful effect on health compared to the usually examined one with equally harmful food for health.[19] One should note, however, when this condition is not satisfied, the theoretical prediction regarding the health impact of a price change is ambiguous. Yet, since there exists vast empirical evidence, also supported by my data, that the foods, rich in supposedly harmful nutrients, such as carbohydrates and sugars, are relatively cheaper than its healthier alternatives, I will hereafter assume $p_t < \lambda$ (Drewnowski and Darmon, 2005).

In addition, since $\lambda \geq 1$, one can see that the effect of change in price p_n is stronger when food is relatively more harmful to one's health. □

There are many reasons as to why one's health responses to change in price might be heterogenous. In particular, time preferences and how forward-looking buyers are affects how much health will be accumulated by different individuals over time and how price changes in a given period affect future health outcomes. First, this model predicts that more impatient individuals will have accumulated less health at any time t compared to more patient (and otherwise identical) individuals who have faced the exact same price path. Second, the impact of a price change on impatient individuals is stronger than for patient individuals. This is summarized in the following proposition.

Proposition 2. *Individual's health is increasing in one's discount factor, that is, those more impatient have lower health H_t compared to the more patient ones: $\frac{\partial H_t}{\partial \delta} > 0$. Health response to change in p_t is decreasing in δ, hence $\frac{\partial H}{\partial p \partial \delta} < 0$.*

[19] Hence, a tax on a particular nutrient will only be effective if it is relatively more harmful than other nutrients and relative prices do not account for this negative externality on health. In other words, taxing the wrong nutrient (even if harmful for health) can decrease health outcomes if it leads to people substituting food consumption towards foods rich in a relatively more harmful nutrient.

Proof. By the Implicit Function Theorem: $\frac{\partial n^*_{t-1}}{\partial \delta} < 0$, so it follows that

$$\frac{\partial H_t}{\partial \delta} = -((\lambda - p_t)\frac{\partial n^*_{t-1}}{\partial \delta} - n^*_{t-1}) > 0.$$

□

Proof. See details in Appendix A.

$$\frac{\partial H_{t+1}}{\partial p_t \partial \delta} = -((\lambda - p_t)\frac{\partial^2 n^*_t}{\partial p_t \partial \delta} - \frac{\partial n^*_t}{\partial \delta}) < 0$$

□

This differential effects of health responses to price changes imply that at any given time t, the effect of a price change affects less healthy individuals more than the more healthy ones. I show that in the following Proposition.

Proposition 3. *Increase/decrease in price p_t improves/deteriorates health H_{t+1} more for those less healthy, that is, those with lower H_t:* $\frac{d}{dH_t}\frac{dH_{t+1}}{dp_t} < 0$.

Proof. See details in Appendix A.

$$\frac{\partial}{dH_t}\frac{\partial H_{t+1}}{\partial p_t} \qquad = \frac{\partial}{\partial \delta}\frac{\partial H_{t+1}}{\partial p_t} \cdot \frac{\partial \delta}{\partial H_t} = \frac{\frac{\partial^2 H_{t+1}}{\partial \delta \partial p_t}}{\frac{\partial H_t}{\partial \delta}} < 0$$

□

Note that in this model I assume the current health stock does not affect the marginal effect of food consumption today on health because $I(n, o)$ was imposed to be independent of H_t for technical simplicity. In reality, however, less healthy individuals might react more radically to a change in sugar consumption. For instance, even a small increase in sugar consumption can result in a full-blown diabetes or disfunctional pancreas for those already highly pre-diabetic (Stanhope et al., 2011). Hence, there might be an additional effect of health stock on the effectiveness of price changes. The role of impatience should, however, remain unaltered in such a generalized setup.

In summation, this simple model predicts that while an increase in price may very likely improve health, it does so only under certain conditions and is therefore to be tested empirically. In particular, it shows that depending on the relative harmfulness of nutrients and relative prices, the effect of price changes can be very different. This model also shows that health response to price changes is increasing in relative harmfulness of the nutrient, one's impatience and is decreasing in one's pre-existing health condition. I present a simple intuitive example to the graphical solution of the model in Figure 1.2. I then check whether data supports some of these theoretical predictions.

1.3 Context

In this section, I first discuss evidence of plausibly exogenous shocks to food prices, which help identify a causal relationship between them and health outcomes. I then shortly discuss the change in dietary patterns and present the evolution of obesity and diet-related chronic diseases in Mexico over the last two decades.

Food Price Dynamics

After the signing of the North American Free Trade Agreement (NAFTA) in 1994, gradually expanded import quotas, reduced tariffs, and removed barriers to foreign direct investments were associated with substantial downward adjustments in food prices that varied spatially and over time (see Figure 1.4, Panel A). Pass-through of liberalization on prices due to tariff changes varied spatially through differential transaction costs, increasing in distance from points of entry (e.g. ports). Nicita (2009) shows that prices of cereals were mostly affected closer to the US border, whereas tariff cuts had almost no effect on their prices in the south. The opposite was true for oils and vegetables, mostly brought to Mexico through southern ports. Figure 1.5, panel A, supports this evidence. Prices of sugar-rich processed foods varied differentially within and between states, changing most rapidly in the northern states (see Figure 1.5, panel B). An additional example on changes in prices being associated with supply-side trade shocks is related to a 20 percent tax on high fructose corn syrup (HFCS) sweetened beverages between 2002-2005, applied by Mexico on the US imports. This resulted in a large drop in HFCS imports (see Figure 1.5, panel A), and a substantial increase in sugar and sugary food prices (see Figure 1.5, panel B).

In addition, the number of foreign-owned supermarkets expanded from 204 centrally located to more than 1300 supermarkets throughout the country between 1995 and 2014, contributing to additional spatial variation in prices over time (Atkin et al., 2014). According to Atkin et al. (2014), foreign retailers, such as Walmart, on average charged 12 percent lower prices for identical barcode-level products of the same quality. Also, entry of a supermarket is shown to result in higher frequency of changes in local prices, especially those of energy dense and fresh foods (Basker, 2007; Basker and Noel, 2009). Using within state variation in supermarkets between 1996-2006, I find consistent evidence on a negative relationship between supermarket density and prices of foods rich in sugar. In my dataset, the number of supermarkets between 1996 and 2006 more than doubled - the number of states with less than five hypermarkets went from 14 in 1996 to barely 4 in 2006 (see Figure 1.4, Panel A). [20] At the same time, prices of foods rich in sugar on average followed a downward trend where supermarkets were expanding (see Figure 1.4, Panel B, C, D). Table 1.4 shows that prices of foods rich in sugar on average decrease by about two percent for every additional

[20]State level panel data was kindly provided by Mauricio Varela. Details on this dataset can be found in Varela (2013).

supermarket in the area within three years.[21] This provides suggestive evidence that price variation in foods rich in sugar over the observed period is associated to significant retail expansion.

Nutritional Transition

Parallel to these trends in food prices, Mexico's dietary intake shifted from a traditional to "western diet". Rich in fat and refined carbohydrates, namely sugars, and low in fiber, the purchase of fruits and vegetables decreased by almost 30 percent between 1988 and 1999. The purchase of refined carbohydrates and soda, both rich in sugar, increased by more than six and slightly less than 40 percent, respectively. Households' consumption of dairy, particularly ice cream and frozen desserts, more than trippled (Rivera et al., 2004). Compared to 69 liters per capita in 1991, at 172 liters per capita per year, Mexico is the largest consumer of soda today (ENSANUT, 2012). In addition, more than 30 percent of the Mexican population is at risk of excessive carbohydrate intake. The average national percentage of total food energy from fat increased as well, albeit less dramatically. Consumption of fat increased from 23 to more than 30 percent, with 12 percent of people being at risk for excessive fat intake (Clark et al., 2012). Hence, Mexicans' diet today is not only unhealthy in terms of total calories, but also in terms of its nutrient composition.

Epidemiological Transition

Mexico is a country that experienced one of the most rapid epidemiological transitions worldwide. Over only two decades, Mexico's disease profile has transformed from malnutrition, communicable infectious and parasitic diseases to a country dominated by obesity, diabetes, hypertension and other diet-related chronic diseases.

Prevalence of excess weight and obesity in adults in Mexico, based on the body mass index (BMI), has gone from less than 30 to more than 70 percent between 1988 and 2012, at an annual increase almost five times greater than the one experienced by the United States.[22] Similarly, the fraction of overweight children has risen from 9 to more than 23 percent in the same period. This worrisome trend is also reflected by the waist circumference of Mexican adults: more than 75 percent are considered to be abdominally obese.[23]

Obesity is considered a serious and chronic condition that increases risk for numerous preventable, behavior-induced, and mostly irreversible chronic diseases, such as type 2 diabetes and hypertension (Catenacci et al., 2009). Nevertheless, more than 20 percent of

[21] This result is robust to various controls and robustness checks and consistent with the finding from Atkin et al. (2014). They find that prices of domestic retailers fall by about two to three percent in two years after the opening of a foreign supermarket and remain stable thereafter.

[22] Someone is considered obese if their body mass index ($BMI = \frac{kg}{m^2}$) is larger than 30, whereas one is considered overweight if their BMI is larger than 25.

[23] Abdominal obesity is specified as a waist circumference over 80 cm for females and 90 cm for males Alberti et al. (2006).

Mexicans diagnosed with type 2 diabetes are of normal weight and more than 10 percent of non-obese are diabetics; similar results hold for hypertension. This underscores the importance of focusing not only on the increase in prevalence of obesity, but also of diet-related chronic diseases. The prevalence of type 2 diabetes in Mexico more than doubled between 1993 and 2012. Today, 9.5 percent of Mexican adult population is diagnosed with type 2 diabetes, and more than 30 percent is diagnosed with hypertension. However, due to many individuals going undiagnosed, some sources estimate type 2 diabetes to already affect almost every fifth Mexican adult and half of the country's adult population to be hypertensive (Barquera et al., 2013).[24]

Both of these diseases represent a high burden for both individuals and society. This includes both direct costs, such as health care expenditures, and indirect costs, such as productivity loss due to morbidity or early death, or costs of complications (e.g. retinopathy, nephropathy, other cardiovascular diseases). For instance, between 2000 and 2007 alone, the mortality rate due to type 2 diabetes increased from 77.9 to 89.2 per 100,000 people. Today, diabetes costs the lives of more than 80,000 Mexicans each year,[25] and is considered the number one cause of deaths in the country, followed by hypertension and cardiovascular diseases (Sánchez-Castillo et al., 2005; Sánchez-Barriga, 2010). Despite the tripled health costs due to chronic disease over the last decade, this burden is expected to increase even more in the coming years. As the Mexican population ages, additional complications driven by chronic conditions are expected to compound the effects of an aging population, which in itself is projected to double or triple healthcare consumption (McKinsey, 2012).

Descriptive evidence shows that states experiencing significant drops in real prices of foods rich in sugar over the last two decades also faced stark increases in diabetes and hypertension incidence. The negative relationship between prices and health is evident in states where prices of foods rich in sugar increased, too; even if prices increased only shortly, chronic disease incidence decreased as well (see Figure 1.6).

1.4 Data

In this section I describe primary data sources used to estimate the effect of price changes of foods rich in sugar or other nutrients on diet-related chronic diseases.

Price and Nutrition Data

The central dataset used for this empirical analysis is a novel dataset on annual time series of retail food prices grouped by main macronutrients[26] between 1996 and 2010. Specifically, I

[24]One is diagnosed as diabetic with a fasting (8-12 hours) plasma glucose of larger or equal to 126mg/dl. Hypertension is diagnosed when systolic or diastolic blood pressure exceeds or equals 140 mmHg or 90mmHG, respectively.

[25]Almost three times the number of homicides due to drug violence.

[26]Macronutrients refer to fats, protein, and carbohydrates, which further consists of sugar and fiber.

construct price indices for foods, rich in sugar, fats, protein or fiber.[27] I assemble this data by combining two different databases; first, a panel data of retail prices with barcode-equivalent food product's description and second, detailed nutritional information of those products, included on their nutritional label.[28]

My price data consists of 25000 food price quotes per year from a nationally representative sample of urban areas across 46 Mexican cities. Data is collected by Banco de México (Banxico) for the purpose of computing the Mexican CPI, and is therefore representative of more than two-thirds of Mexican consumers' expenditures. There are many reasons why this data is suitable for the purpose of my analysis. First, food prices are tracked for the same or a very similar product using a unique product identifier continously within stores over 15 years, which makes them comparable over time, and hence making it possible to exploit their time variation within regions. In addition, price data spanning over almost two decades allows me to observe a dynamic relationship between prices and health outcomes of interest as well.

Second, required by Articulo 20-Bis of the Codigo Fiscal de la Federacion, Central Bank publishes store price microdata together with precise item descriptions in the official gazette of the Mexican government, the Diario Oficial de la Federacion (see Figure 1.7).[29] Crucially for this project, products' price quotes are very narrowly defined. Definitions include product's name and brand, packaging type and weight, such as Kellog's Cereals, Zucaritas, box of 250 grams, sold in outlet 1100 in Mexico City.

Detailed item's description enables me to match each food product with its calorie content and exact nutritional composition of main macronutrients. In particular, I obtain information on amount of energy in kilocalories (kcal), grams of fats, protein, sodium, carbohydrates, of those grams in sugar and fiber per 100 grams.[30] The motivation for collecting detailed nutritional information per product is the following. Individual product prices are nested within 106 product categories, such as yoghurt, cereals, or snacks. To obtain price indices of foods rich in different nutrients one could take a somewhat subjective or ad-hoc approach and divide foods by macronutrients based on the average product category nutritional value (Miljkovic and Nganje, 2008). However, this approach masks a large between product differences in the nutrient content within each product category and does not take into account the within product correlation of nutrients (Griffith and O'Connell, 2009). Figure 1.8 shows an example for *''Galletas Popular"*, a product category consisting of both, salty and sugary snacks. One can see that nutritional composition varies substantially across products, making

[27]For simplicity, I will interchangeably use the term *''nutrient prices"*, or prices of sugar, fats, protein or fiber.

[28]The price quotes for 1996-2010 were kindly provided by Etienne Gagnon at the Federal Reserve Board in Washington D.C. Detailed description of his data can be found in (Gagnon, 2009).

[29]The National Institute of Statistics and Geography (INEGI) took over the collection of prices from 2011 onwards and publishes them on their website.

[30]Information on fiber is often missing or reported as smaller than 0, in which case I either record it as missing or assign value 0, respectively. Macronutrients are converted from grams to total calories per 100grams by multiplying grams of carbohydrates by 4, grams of proteins by 4, and grams of fats by 9 (USDA).

it difficult to disentangle the effect foods rich in one nutrient from another or their combination on health outcomes of interest.[31] To overcome this challenge, collection of detailed nutritional data and matching it to product characteristics is a crucial step in constructing the price indices of interest.

I collect nutritional information on products from several sources. I manually search nutritional information on product's manufacturer's websites, and websites such as Factual.com, Superama.com.mx, or Walmart.com. These websites' nutritional information is of reliable quality - for instance, nutritional database at Factual.com consists of 600,000 consumer packaged goods in a UPC centric US database, and Superama.com.mx and Walmart.com report nutritional information provided by manufacturers. In addition, a very important source of information on nutritional composition is Mexican Food Composition Table. Nutritional information was manually gathered from Fatsecre.com.mx, or Caloriecount.com as well. Matching nutritional information to each product followed a double blind entry method, where each product was cross-checked at least twice. In addition, each match was always compared to a ''generic" match in either Mexican Food Composition Table or USDA Food Composition Table. Whenever exact match between the product and its nutritional information cannot be found, product's nutritional composition is compared to the next most similar product found. If nutritional composition at the brand level either cannot be found or is incomplete (eg. information on sugar or other macronutrient is missing), nutritional composition assigned corresponds to a similar product of a different brand.[32] I pay special attention to product's fat and sugar content throughout the nutritional composition matching. For instance, I differentiate between skimmed and whole milk, plain or fruit yoghurt, and diet or regular soda. I assign average nutritional values at the higher food group level only in few cases, such as in the case of spices, or roasted coffee.

Third, using each item's unique identifier, consisting of a product number, store, city and food category, I can not only track product's price trajectory over time, but also assign it a constant nutritional content. Since Banxico reports changes in product's representation, brand, or type, I can assign an appropriate, updated nutritional composition to substitutions of existing or addition of new items.[33] Fourth, prices of food items are mostly conveniently expressed either per 1 kg or 1 liter, which makes the interpretation and scaling of the nutritional composition fairly straightforward. All food items for which prices are not reported either in kilograms or liters are excluded. Lastly, product division and unit of measure make it convenient to combine the store microdata with Household Expenditure Survey data (ENIGH), from which I obtain the weights, used for the price indices calculation (see

[31] For instance, whether the price of snacks is a proxy for price for sugary or fatty foods is unclear, since average values per product category are high in of both nutrients.

[32] For instance, Brand XY 2% low fat milk is assigned a nutritional composition of a generic or 2% low fat milk of another brand.

[33] Banco de Mexico published complete lists of item descriptions in March 1995 and July 2002, corresponding with major basket revisions. Therefore, items between 2002-2010 cannot be traced back to earlier years due to a change in their key identifier and hence separate nutritional matching had to be done.

Section 1.4).[34] Since ENIGH is collected bi-annually during the third quarter, I compute a three-month average price of individual items in each year's third quarter for the purpose of this empirical analysis.

Nutrition clustering

To fairly objectively construct price indices representing foods rich in each macronutrient individually, I use the k-mean clustering approach (Harding and Lovenheim, 2014).

First, I classify 106 food groups into 13 mutually exclusive categories that roughly correspond to major food areas of USDA categorization.[35] These are grains, snacks and candy, meat, condiments, oils, juices and syrups, sodas, warm beverages, fruits, vegetables, prepared or packaged meals, dairy and milk.[36] Second, I separate these categories using the k-mean clustering approach. This approach separates the initial 13 product categories into 29 product-nutritional clusters. Finally, based on nutritional composition of each cluster, I choose those primarily rich in sugar content and no other nutrient.[37] I use them to construct price indices of foods rich in different nutrients. Roughly, chosen clusters identifying foods rich mainly in sugar come from within the food category of sodas, juice and syrups, sweets and candies, and fruits food category.[38] Similarly, I identify groups of items rich in fats, items rich in fiber, and items rich in protein relative to other nutrients.[39]

K-means clustering method is an iterative learning algorithms that solves the clustering or grouping problem. The main idea behind this algorithm is to partition a set of objects into k distinct groups or clusters. K is a parameter that is initially set externally. Using a set of covariates and a measure of distance, the centroid of each cluster and the distance of each object to its cluster's centroid are calculated. The centroid for each cluster is the point to which the sum of Euclidian distances from all objects in that cluster is minimized.[40] The goal of k-means clustering is minimizing the distances within clusters (having similar objects within clusters) while maximizing the distance between the clusters (having different objects across clusters). Given a clustering outcome, each object has a silhouette value which measures how close each point in one cluster is to points in the neighboring clusters. It ranges from +1, indicating objects within the assigned cluster are well-separated from all other clusters in the object space, through 0, indicating objects that are not well distinguished

[34] Food categories in retail price data are representative of the ones in ENIGH, accounting for at least 0.02 percent of households' expenditures, which captures well above of the 95% of Mexican households' expenditures Gagnon (2009).

[35] USDA: `www.ars.usda/ba/bhnrc/fsrg`

[36] For details, see Appendix, Table 1.1.

[37] The largest share of other nutrients mostly does not account for more than 20 percent of food's serving.

[38] For instance, within sodas, I chose the cluster of regular, non-diet sodas. Within fruits, I choose canned fruits from the cluster mainly rich in sugar since much of their sugar content is due to added sugar and not fructose only. Results are not sensitive on either including or excluding this category (or any other, one by one).

[39] See Figure 1.9 for more detailed representation of clusters' nutritional composition.

[40] Some other distance measure may be chosen.

across clusters, to -1, which means objects are probably assigned to the wrong cluster. The average silhouette value provides a measure of success of the clustering method and can be used to determine which k is ideally used.

For each of the 13 food categories, I employ k-means clustering to determine food subgroups within these categories and choose the k that maximized the average silhouette value as described above.[41] The covariates used to determine the distance measures are the product's total calories, calories from fat, grams of protein, carbohydrates, sugar and sodium per 100 grams.[42] On average, food categories are divided into 2 or 3 clusters.[43] As an example, I plot the silhouette values for soda products at two partitions. Figure 1.10 shows that not only this methodology successfully separates products into different product-nutrient clusters, but also stresess the importance of de-grouping the products beyond the product category level. For instance, in diet soda, we observe 0 grams of sugar, yet an average regular soda contains more than 30 grams of sugar per can (12 fl).[44]

Prices

Based on k-mean clustering results, I construct the Laspeyres price index for foods rich in sugar, fats, protein and fiber for each of the 46 cities or 32 states. As weights, I use 2008 product category budgets shares at the urban state level from ENIGH.[45] Since there exists no information on consumption at the dissagreggated product level, I first calculate median price for each product category within clusters of choice and then assign it an appopriate weight.[46] Lastly, I obtain real prices by deflating Laspeyres index with the 2008 city level CPI. Figure 1.5 shows within state variation of real prices of foods rich in sugar between 1995 and 2010.

Health Data

Incidence Data

I motivate the relationship between prices of sugar and diet related chronic diseases by combining state average prices with data on state-year incidence rate of hypertension, and type 2 diabetes between 1996-2010.[47] Data is collected by the Mexican National Epidemiological Surveillance System (SINAVE). The SINAVE collects data on new cases of disease from more

[41] I set k=1,...,15.

[42] The reason for excluding information on fiber from k-mean clustering analysis is due to its miss- or under-reporting on the nutritional panel. However, k-mean clustering with fiber as an additional attribute gives very similar results.

[43] This suggests that increasing the number or partitions would not have changed my results.

[44] Harding and Lovenheim (2014) obtains very similar results in the division of sodas and clustering of other food categories as well.

[45] Urban areas are defined as those with more than 2500 inhabitants.

[46] For instance, the weight I used for cluster of regular soda refers to any soda in ENIGH, since budget shares for diet and regular soda separately is not available.

[47] I express incidence rate as per 100,000 population.

than 95 percent of all local health centers in Mexico.[48] They use the 9th or 10th Revision of International Statistical Classification of Diseases and Related Health Problems (ICD-10) coding system when reporting diseases on a standardized data collection form. More than 85 percent of health centers reports epidemiologic information on a weekly basis.[49] SINAVE calculates incidence rates per 100,000 population using 1990-2050 population projections from the appropriate Population Censuses (CONAPO). State-year panel data allows me to not only avoid the disease self-report bias due to adminstrative nature of the data, but also enabels me to look at the contemporanous and lagged relationships between prices and health outcomes over 15 years. See summary statistics in Table 1.2. [50]

MxFLS

Individual level data comes from the Mexican Family Life Survey (MxFLS). This is a nationally representative longitudinal survey, collected at the individual level in 2002. 2005 and 2009. With less than a 10 percent attrition rate, detailed information on health, personal traits, and socioeconomic data is collected and tracked for more than 35000 individuals (8400 households) in 150 urban and rural communities, 136 municipalities and 16 states.[51] The MxFLS contains detailed anthropometric module, including information on height, weight, or waist circumference, which allows me to calculate one's body mass index or abdominal obesity, respectively. All three values are measured by a nurse practicioner, avoiding the self-reporting bias (Thomas and Frankenberg, 2000). I use the information on hypertension and self-reported diabetes as well. Data contains many different demographic characteristics, such as age, gender, educational attainment, individual time allocation, employment status, or self-reported household level consumption expenditures and assets.

1.5 Empirical Strategy

In order to estimate the effect of prices of foods rich in sugar on type 2 diabetes and hypertension incidence rate, I first exploit the within state variation in prices to estimate the

[48] They are included in IMSS, ISSSTE, IMSS-Oportunidades, PEMEX, SEDENA, SEMAR, DIF or SALUD.

[49] 16,468 out of 16900 local health centers, 2428 municipalities, and 234 health jurisdictions in Mexico are included in this system. Among those that miss weekly reports, the main reasons include physicians on leave, vacations, or sickness and lack of transmission means (Tapia-Conyer et al., 2001).

[50] There exists no other data on disease incidence rates in Mexico. However, if one applies a simple excercise assuming difference in prevalence of disease between years, adjusted for mortality, equals incidence rate, results from Mexican National Nutrition and Health Surveys give comparable results to the data I observe here.

[51] Survey collects the data including for those who changed households, and migrated within Mexico or emigrated to the United States. Number of communities, municipalities and states increases over time due to migration.

following equation:

$$\mathrm{y}_{st} = \alpha_s + \alpha_t + \sum_{j=0}^{4} \beta_{t-j} \log\left(\mathrm{P}_{\mathrm{sugar}}\right)_{st-j} + \mathbf{z}'_{st}\theta + \varepsilon_{st} \tag{1.1}$$

where y_{st} is the dependent variable, either type 2 diabetes or hypertension incidence rate, observed for the state s at time t, expressed as age-adjusted incidence rate of disease per 100,000 population at risk.[52] α_s control for time invariant, state-specific unmeasured factors that are correlated with prices and health. Time fixed effects, α_t, control for common trends. Variable $\log\left(\mathrm{P}_{\mathrm{sugar}}\right)_{st}$ measures the log of average real calorie prices of foods rich in sugar in state s, at time t. To observe relationship between health and change in prices of foods rich in fat, protein or fiber, and to control for a general food cost at a city level, their one-year lags are included as well. [53] The vector $\mathbf{z}_{s(m)t}$ controls for time variant changes in food availability and income due to rainfall shocks, proxied with drought index at the state level.[54] State level GDP, which absorbs local macroeconomic variation, is included as well. To address the concern that widespread availability of cheap calories might affect health irrespective of prices (Currie et al., 2009; Anderson and Matsa, 2011), number of fast food restaurants per squared kilometer at the state level is added as a control. In addition, I control for local demand shocks that are potentially correlated with local prices and one's health, such as advertising, with fast food services advertising expenditures per capita at the state level (Chou et al., 2005; Saffer and Chaloupka, 2000).[55] To estimate the persistence of the price effect on health, prices with lag j are added. Unless stated otherwise, all parameters in this equation are estimated using state and year fixed effects ordinary least squares. To account for correlation of the residuals ε_{st} within state, I report standard errors clustered by state.[56] The key identification assumption of the equation (1.1) is that after conditioning on the vector $\mathbf{z}_{st}$, state and year fixed effects, changes in disease incidence rates are not systematically related with changes in prices of foods rich sugar or other nutrients.

Second baseline specification using individual level panel data from MxFLS is the following:

$$\mathrm{y}_{it} = \alpha_i + \alpha_t + \beta \log\left(\mathrm{P}_{\mathrm{sugar}}\right)_{c(i)t} + \mathbf{x}'_{it}\theta + \mathbf{z}'_{m(i)t}\delta + \varepsilon_{it} \tag{1.2}$$

where y_{it} is the dependent variable observed for an individual i at time t. Health outcomes of interest are log of BMI, an obesity indicator, log of waistline measure (in cm), an abdominal

[52]Diabetes is type 2 diabetes, unles stated otherwise; used interchangeably.

[53]Results stay nearly unchanged if other period prices of foods rich in other nutrients or price index of other foods are included.

[54]See Dell (2012) for details on how drought index is constructed. Rainfall data is obtained from the University of Delaware's center for Climatic Research.

[55]I obtain this data from the Mexican Population and Economic Census data from 1999, 2004 and 2009. Economic census reports the number of economic establishments per municipality, using North American Industrial Activity Classfication (SCIAN) classification. I linearly interpolate number of fast food restaurants and service establishments and their advertisement expenditures at the state level for missing years. SCIAN codes used to record fast food restaurants and services are 722211, 722212 and 722219.

[56]I also repeat the empirical exercise using the wild bootstrap with 1000 repetitions. Results remain nearly the same (see Appendix Table A7 and A8).

obesity indicator, an indicator variable for whether you were ever diagnosed with type 2 diabetes or hypertension by a doctor. I define someone as obese if BMI is greater or equal to 30. Someone is considered abdominally obese, if his(her) waistline is greater or equal to 90(80) cm. Type 2 is an indicator variable that equals one for non-insulin dependent individuals, diagnosed with diabetes by a doctor after age of 35. Diabetes diagnosis is self reported and does not distinguish between two types of diabetes (type 1 and type 2). Hence, I base my definition of type 2 diabetes following the WHO (2002) and Evans et al. (2000), who suggest that diabetes diagnosed after age of 35 is most likely of type 2, not of type 1, and is non-insulin dependent in most cases. Hypertension is defined as an indicator variable that equals one if an individual was ever diagnosed with hypertension.[57]

Common macroeconomic fluctuations are controlled for with the inclusion of year fixed effects, α_t. Controlling for individual fixed effects, α_i implies that results are not driven by any variable which differs across individuals, such as genetics, or tastes.[58] Since tastes and preferences for different food might vary locally, influencing local demand and/or supply of food and health differentially within the state, this allows me to relax the assumption of homogenous tastes within states from equation (1.1) while also addressing the aforementioned concern of reverse causality. Hence, the identifying assumption of the price effect on health outcomes is that changes in unobservable determinants of one's health are uncorrelated with changes in prices of sugar over time.

The vector $\mathbf{x}_{it}$ represents a set of individual and household level time-varying controls, such as socioeconomic status decile indicator, household size, house ownership status, individual's age and education, work status and log of annual labor income and distance to nearest city in kilometers, as well as controls for prices of foods rich in other nutrients or food price index.[59] I also want to control for ways people might spend their calories. One can burn calories through basal metabolism, which affects the rate of energy expenditure at rest, thermic effect, which burns calories through processing food, and physical activity (Cutler et al., 2003). All but physical activity might be controlled for with individual fixed effects, since one's daily routine might change over time. Thus, I control for sedentary lifestyle and physical activity by adding controls on weekly hours spent on exercise, watching tv, or using the internet. In addition, one might change the habit of cooking his own meals due to change in relative price of foods, potentially affecting his health through different composition of caloric intake, quantity or quality wise (eg. increasing food consumption away from home instead). Thus, I include the control on cooking at home.

Individuals might also be exposed to health awareness campaigns at various locations and times, so different spatial trends in health consciousness could bias my results. Hence,

[57]In addition, the variable equals one if one's measure of systolic and diastolic blood pressure is higher than 140/90 mmHG, respectively. In MxFLS, systolic and diastolic blood pressure are both measured twice. I calculate the average of two measures when defining the variable of interest.

[58]I assume preferences and tastes are time-invariant or change very slowly, especially since I am looking at the 2002-2009 period only (Atkin, 2013, 2010)

[59]SES deciles are obtained using principal component analysis on household income, size, assets, and house materials.

I construct a proxy for individual health awareness. Priot to taking measurments on height and weight, MxFLS asks individuals whether they know their measurments and if so, what they think they are. Health awareness proxy then equals the sum of indicator variables of whether you guessed your height or weight compared to the measured one close to 5cm or 3kg, respectively, of whether you exercised at least once a week and whether you smoke or not. The higher the value, ranging from 0 to 4, the more health concious you are. Lastly, to control for differential trends in access to health care and diagnostics between areas, I include an indicator whether an individual has medical insurance or not as an additional control variable. The vector $\mathbf{z}_{s(m)t}$ serves the same purpose as in 1.1, only that number of fast food services and their advertising expenditures are observed at the municipality level.

The variable $\log\left(\mathrm{P}_{\text{sugar}}\right)_{c(i)t}$ measures the log of real price of foods rich in sugar in i's nearest city $c(i)$, at time t. As in (1.1), I include prices of foods rich in other nutrients.[60] Using individual's municipality identifier, I link each individual's municipality's centroid to the nearest municipality centroid of 46 cities for which the price data is available.[61] Median distance of urban individuals to the nearest city is 26 kilometers, and more than 75 percent of people lives within 50 km radius of a city that they are assigned to (See Figure 1.14).[62] There are 39 cities being merged to urban individuals over all three periods, however 30 of them are being used for the analysis on average. Analysis is focused on those who remain assigned to the same city throughout the analysis to maintain a more balanced cluster size. Observing individual's geographical location, I relax the assumption on no cross-state migration, set in equation (1.1). Since prices are collected in cities, I focus my analysis on urban areas only.

1.6 Empirical Results

In this section I provide empirical results, then discuss identification concerns and describe the robustness checks I apply to rebut them.

Main Results

Table 1.6, Columns 1-2, shows my baseline empirical estimates of the effect of changes in prices of foods rich in sugar and other nutrients on type 2 diabetes and hypertension incidence rates per 100,000 population (see equation 1.1 above). Adding macroeconomic controls (Coloumn 3) or controlling for food environment (Column 4) does not seem to change the estimates. Results show that a relative decrease in real price of calories of foods rich in

[60] Even though β seems to measure contemporanous price effects, health data is collected over 1-2 years, hence, contemporanous effect in this regression can be comparable to a one year lag effect.

[61] INEGI provided me with a list of municipality codes from where store prices were collected - one city spans over more municipalities. In addition to municipality centroid matching, I also re-do the analysis on using the linear distance between inidividual's municipality's centroids and city's polygon border. Results remain unchanged.

[62] Results are not sensitive on limiting the sample to various distance cutoffs.

sugar significantly increases the incidence rate of type 2 diabetes and hypertension. Changes in real prices of foods rich in other nutrients, however, do not (Coloumn 5). Even though coefficients for prices of foods rich in fats and fiber are of expected sign, they are all non-significant at the conventional levels.[63] On average, a 10 percent decrease in prices of foods rich in sugar results in 9 new diagnosed cases of type 2 diabetes and 16 new diagnosed cases of hypertension per 100,000 people within one year.[64] Prices have a diminishing significant effect on the current incidence rate of diabetes for up to two to three years. A similar, yet more stable effect over time, is observed in the case of hypertension. The total effect of a 10 percent increase in prices of sugary items results in 17 new diabetic and 33 new hypertensic cases per 100,000 population over 3 or 4 years, respectively, which is equivalent to an approximately five percent decrease in disease incidence rates over the same period.

I obtain comparable results estimating the equation (1.2) using MxFLS. Table 1.11, Panel A, suggests that decrease in prices of foods rich in sugar significantly increase the probability of becoming diabetic. Specifically, a decrease in prices of items rich in sugar content by 10 percent on average increases the probability of becoming diabetic by 0.5 percentage points (Column 1), equivalent to an almost 5 percent increase from current 11 percent diabetes prevalence rate of urban adults. This translates into almost 300 thousand new diabetics within one year, counting urban areas and adults only. Since diabetes is still underdiagnosed, this is probably a conservative estimate. Results are sensitive neither to additional time-variant individual controls (Column 3) nor to local economic ones (Coloumn 4). Again, changes in prices of other nutrients do not change the main result, suggesting that only change in prices of sugar items significantly affects the probability of becoming diabetic.

I observe a negative, non-significant relationship between prices and probability of being diagnosed with hypertension (See Table 1.11, Panel B). One reason for an imprecise and non-significant effect might be due to the under-diagnosing. Lower statistics could be explained by unawareness of having the disease (Lloyd-Sherlock et al., 2014).

Price effects matter for adults' waistlines and their probability of becoming abdominally obese as well (Table 1.10, Panel A and B, respectively). Decreasing price of items rich in sugar by ten percent, on average, increases waistline by almost 0.5 percent (Panel A, Column 1-5). This translates into an almost half a centimeter larger waistline in one year of time. At the same time, probability of becoming abdominally obese increases by 1.5 percentage points (Panel B, Columns 1-5), where changes in prices of sugary items are the only ones significantly affecting this outcome of interest. Higher sugar price elasticities for abdominally obese compared with waistline results suggests that individuals at the right tail of the waistline distribution are more price elastic. Results remain robust to additional controls.

Changes in prices of sugary foods affect children's probability of becoming obese, too. Table 1.8 shows that a ten percent decrease in prices result in 0.3 percentage point increase

[63] Results remain unchanged if controlling for other combination of nutrients as well.

[64] I assume people's health response to prices is symmetric either to price increase or decrease.

in probability of becoming obese, which is equivalent to around three percent increase in children obesity.[65]

These estimates suggest that real prices of foods rich in sugar explain approximately 20 percent of the trend in type two diabetes prevalence in Mexico in the last two decades. This translates into about 1 million more people being diagnosed with diabetes between 1996-2010 due to cheaper sugary processed foods.[66] Taking into account the direct (US dollars 743) and indirect costs (US dollars 3,528) of diabetes per capita (Barcelo et al., 2003), additional diabetes due to decreasing costs of sugary foods sum up to around 4.5 billion US dollars over this period. [67] Hence, if a one-time ten percent tax on foods rich in sugar were applied, this would prevent almost half a million of people from being diagnosed with type two diabetes within one year. In addition, the tax would prevent around 1 million people from becoming abdominally obese.

Robustness checks

One concern with the estimates is that even after conditioning on year, state or individual fixed effects and time varying individual and local characteristics, the changes in prices of foods rich in sugar may still be associated with other determinants of chronic diseases that I cannot control for. In this section, I present robustness checks that account for those identification concerns.

One possible identifying concern is the strongly positive within-state trend of chronic disease, alongside a negative one in the real prices of food. Several points lend credibility to my results. Firstly, I include both year and region-year fixed effects, which control for any omitted variable that varies over time within region, linear state trends, and linear trends by baseline risk for disease using MxFLS. There is a risk for attenuation due to sweeping out variation and the inclusion of an excessive number of controls in the regression, however results remain very similar (Table 1.12, Columns 7 - 10). Secondly, the prices of fatty foods follow a decreasing trend within many states where disease incidence is increasing. If this trend is driving my results, I could expect a negative relationship between health changes and the price changes of foods that are rich in fat as well. However, I find no evidence of any such relationship. This also adds additional evidence that changes in prices of fatty foods do not matter for health (see Figure 1.13, panel A). Lastly, conditioning the regression for lead prices of sugary foods, I show no systematic relationship between these prices and health outcomes (see Table 1.7, Columns 1-2).

This last test also addresses the concern of reverse causality. Over the last three decades, Mexico experienced a dramatic increase in the import of processed foods and fast food restaurant supply (Clark et al., 2012). The identification concern goes that imports of those

[65]Results on log(BMI) and obesity indicator have the expected sign (See Appendix, Table 1.9), however are smaller in magnitude.

[66]Average decrease in real prices of sugar between 1996 and 2010 was around 20 percent.

[67]Assuming that costs per capita remain constant throughout. For this calculation, I used population projections by CONAPO.

foods and establishment of new fast food services did not locate randomly with respect to consumer's demand. In Table 1.7, Columns 1-2, I ask whether current incidence rate of type 2 diabetes and hypertension may be correlated with future prices of sugar in states that experienced some unobserved upwards trends in the demand of foods rich in sugar content. Namely, if future food prices predict contemporaneous health conditional on current food prices, individuals of particular health are likely to influence prices rather than the other way around. Hence, future prices of sugar should not affect health outcomes of interest. It is evident from Figure 1.11 that the relationship between lead prices and diabetes is not significant. I repeat this exercise using MxFLS data as well, using a one year price lead. Results remain unchanged (See Table 1.12, Column (1)-(3)). In addition, Figure 1.12 confirms that there is no underlying relationship between lead prices and health outcomes in the data. I further invalidate this concern by controlling for time variant, such as work and income, and time invariant individual characteristics, such as tastes and preferences, by including individual fixed effects. I also test whether changes in the price of sugary foods are correlated with unhealthy behavior, as proxied by using a measure of smoking behavior, predictive of obesity and chronic disease (Gruber and Frakes, 2006). This test addresses the concern that areas more prevalent in unhealthy behavior attracts investments offering relatively more processed foods than areas with relatively healthier behavior. I find that there is no systematic relationship between changes in smoking behavior and prices of sugary foods. Furthermore, by controlling for the number of local fast food restaurants and their advertising expenditures in most of regressions, I address the concern of widespread availability of cheap calories and local demand shocks that might affect health irrespective of prices.

Even though the statistical health system in Mexico is recognized as one of high quality, many individuals are still going undiagnosed with diseases, specifically, type 2 diabetes or hypertension. If disease under-reporting was constant over time or shared a common trend countrywide, this would not be a concern. Yet, over a past couple of decades, Mexico expended considerable effort into improving the national statistics of non-communicable disease tracking, diabetes in particular. One might be worried that areas with relatively cheaper sugar calories have a faster increasing trend of better disease diagnostics than those where sugar calories became relatively more expensive, thus overestimating my results. It could be that people in those areas are less likely to be insured and therefore go undiagnosed more often. The Mexican Health and Nutrition survey 2012 (ENSANUT) records that 9 out of 100 uninsured people tested positive on diabetes, yet among the insured only 2 out of 100 were newly diagnosed.[68] Mexico undertook significant measures to achieve universal health coverage especially after 2004, with an intentional aim of Seguro Popular to ensure universal access to preventative healthcare such as diabetes screening and treatment of chronic diseases. Hence, this could pose a valid concern regarding the bias of my estimates (Knaul et al., 2012).[69].

[68] Even though the ratio between uninsured and insured people in that sample was almost one to one.

[69] Seguro Popular is a national health insurance program, which started in 2004 and by 2012 expanded

If this is so, I should find a negative relationship between other diseases of similar diagnostic needs and changes in real prices of items rich in sugar as well. Hence, I estimate equation (1.1) using type I diabetes and asthma incidence rate per 100,000 population as new outcome variables. Diabetes mellitus type 1, also known as juvenile or insulin dependent diabetes, is an autoimmune disease in which a persons pancreas stops producing insulin. The causes of type 1 diabetes are not yet entirely understood, however scientists are certain that the onset of this disease has nothing to do with diet or lifestyle and cannot be prevented.[70] Type 1 diabetes is, just as type 2 diabetes, diagnosed through a blood test, followed by additional tests to distinguish it from type 2 diabetes. Similarly, asthma requires significant testing, either through physical examination, lung function test, or bronchoprovocation tests among others. Similar to symptoms of other diseases, it is not straightforward to diagnose. Hence, if the concerns outlined above are unjust and my identification is credible, type 1 diabetes and asthma incidence rates should not be correlated with prices of sugar or its lags (or prices of foods rich in any other nutrient). Indeed, I find no evidence that incidence of type 1 diabetes and asthma are correlated with prices of foods rich in sugar or any other nutrient (See Table 1.7, Columns 3-6). With this placebo test, I also show that, conditional on state fixed effects, changes in prices of sugary foods are not correlated with state characteristics. Moreover, controlling for number of medical units at the state or municipality level does not change results.[71] Together, these results add credibility to casual interpretation of my results.

Heterogeneous effects and Mechanisms

High Risk vs Low Risk

So far I have shown that economic incentives, such as falling real prices of sugar, on average, contribute to prevalence of diet-related chronic diseases and abdominal obesity. The substantially higher elasticity of abdominal obesity to prices of sugary foods (Table 1.10, Panel B) compared with the (log of) waistline measure (Table 1.10, Panel A) suggest that individuals at the higher end of the waistline distribution are more price elastic than the ones at the lower end one. In this section, I provide evidence that similar results hold true for individuals who are at a high risk for developing type 2 diabetes or hypertension.

We would expect for type 2 diabetes and hypertension to develop over a longer period of time. For instance, high blood sugar can precede the development of type 2 diabetes for as long as 10 years. Hence, nearly everyone who has type 2 diabetes was pre-diabetic first and is, to a certain degree, able to prevent pre-diabetes from becoming type 2 diabetes through making changes in weight, exercise, and especially diet (Ezzati et al., 2003).[72] On the other

access to health care for tens of millions of previously uninsured Mexicans (Knaul et al., 2012).

[70]See American Diabetic Association

[71]Similarly, controlling for the linearly interpolated share of population enrolled in Seguro Popular from 2005 and 2009 Population Census to other years at the state or municipality level does not affect the results.

[72]A pre-diabetic is someone whose blood glucose levels are higher than normal, but not high enough to be

hand, several studies show that drinking sugary beverages daily for only two weeks increases cholesterol and triglyceride levels by 20 percent, and daily consumption of sugary drinks for six month increases fat deposits in the liver by 150 percent, directly contributing to both, diabetes and heart disease (Stanhope et al., 2011; Maersk et al., 2012). A new report from the Centers for Disease Control and Prevention shows that less than 10 percent out of more than 75 million adults with pre-diabetes know they are pre-diabetic. In Mexico numbers are unknown, but probably even higher. This means that there is a substantial share of population whose increased sugar consumption even over a very short period of time might slide them into a chronic disease, such as type 2 diabetes or hypertension.[73] This could, first, explain the strong effects of changes in prices of sugary foods on health within a short period of time (see Section 1.6), and second, suggest that health most responsive to prices is of those at the highest risk for disease development.

I divide individuals into a moderate to high and low risk group for diabetes development based on The Type 2 Diabetes Risk Assessment Form (see Figure 1.15). The Type 2 Diabetes Risk Assessment Form is an example of an effective patient questionnaire with eight scored questions. The total test score provides a measure of the probability of type 2 diabetes development within the following ten years. I exclude the question on daily vegetable consumption due to its unavailability, and proxy for genetic predisposition of disease by assigning three points if at least one household members has diabetes of either type. Information on whether elevated glucose levels is available for pregnant individuals, as per question 5. Due to a lower total number of points, the cutoff point for being considered moderate to high risk is ten. A slightly elevated risk is considered for scores between six and nine. Scores below six are considered as low risk for developing type 2 diabetes.[74] Following the Center for Disease Control and Prevention risk factor guidelines, I construct a similar risk assessment questionnaire for hypertension. Each risk factor weighs one point (obesity and abdominally obesity, smoking and not exercising, experiencing sleeping problems and stress, being diabetic, and being older than 45 years old). People at high risk for hypertension are those scoring at least 4 points, those below are considered as low risk. Individuals' risk is assessed at the values of their initial survey year.

To check whether health elasticity to prices differs between people at different risks for disease development, I estimate the following equation:

$$\mathrm{y}_{it} = \alpha_i + \alpha_t + \beta_1 \log\left(\mathrm{P}_{\text{sugar}}\right)_{c(i)t} + \sum_{j=2}^{3} \beta_j \log\left(\mathrm{P}_{\text{sugar}}\right)_{c(i)t} \times \mathbb{1}_{j_i} + \mathbf{x}'_{it}\theta + \mathbf{z}'_{m(i)t}\delta + \varepsilon_{it} \quad (1.3)$$

classified as diabetic (that is, fasting blood glucose level is below 126mg/dcl). Early pre-diabetes treatment can return blood glucose levels back to normal range - one can lower risk for type 2 diabetes by almost 60 percent through losing 7 percent of body weight and exercising 30 minute per day.

[73]More than 1 in 4 pre-diabetics will develop type 2 diabetes within 3-5 years. Chen et al. (2004) observe cumulative and long-term effects of the yearly blood glucose level gain only during the winter holidays. Similarly, (Tobenna and Rahkovsky, 2014) find significant increase in glucose levels among diabetics already within only a 3-month window during relative increase of healthy food prices.

[74]This means that at least one in six individuals with a score more or equal to ten will develop type 2 diabetes within ten years; or at most 1 in 100 will develop the disease if their score is below six.

where $\mathbb{1}_{j_i}$ is an indicator variable that equals 1 if an individual i is considered j, where j is either high/slightly elevated or low risk, in the initial survey year.[75] I include controls for linear time trend by risk group. Table 1.13 shows that individuals who are at a high risk for developing type two diabetes or hypertension are substantially more price elastic than those at a lower risk. In particular, for every ten percent decrease in sugar prices, people at high risk for disease development are two percentage points more likely to develop the disease relative to those healthier individuals (Columns 1-2).[76] Moreover, I find suggestive evidence that even though the effects are strongest within one year of a price change for those at the highest risk for disease development, lower prices of sugary foods deteriorate health of those at a better health with a lag as well (see Table 1.14).

Impatience

Decreased prices of sugary foods alone cannot explain the heterogeneity in individuals' responses. So, why do some individuals, such as those at a higher risk for either developing the chronic disease or becoming abdominally obese, respond to changes in relative prices of sugar significantly more than healthier individuals?

Following my theoretical framework, differences in time preferences could help explain the heterogeneity of individuals in their price sensitivity of health outcomes. I expect that less patient individuals put more emphasis on present costs such as prices of foods rich in sugar, and internalize future costs such as worse health condition as a consequence to present (consumption) decisions less. Therefore, their health is more price sensitive than the health of the patient individuals. To test this hypothesis, I estimate the following equation:

$$\mathrm{y}_{it} = \alpha_i + \alpha_t + \beta_1 \log\left(\mathrm{P_{sugar}}\right)_{c(i)t} + \beta_2 \log\left(\mathrm{P_{sugar}}\right)_{c(i)t} \times \mathbb{1}_{impatient_i} + \mathbf{x}'_{it}\theta + \mathbf{z}'_{m(i)t}\delta + \varepsilon_{it} \quad (1.4)$$

where $\mathbb{1}_{impatient_i}$ is an indicator variable that equals 1 if someone is considered impatient and 0 otherwise. I measure time preference using a battery of six questions, available in the 2009 MxFLS survey wave: "Suppose you won a lottery of \$1000. Would you prefer to take the money now or get \$X in a month?". Varying the amount of money one can receive in one month time makes it possible to back out a range of implicit discount rates for each individual, assuming each response is coming from a discounted utility calculation (Samuelson, 1937). Given the question setup, I can divide individuals into 5 different groups (see Figure 1.16). I define someone as most impatient if their discount rate falls either within the highest and/or one of the two highest discount rate interval. Throughout, I assume that the discount rate is the same across different activities (e.g. one exhibits the same discount rate with respect to money and health), and is constant over time.[77] Since discount rate

[75] For hypertension, one indicator variable for risk assessement is constructed, and the equation is adjusted accordingly.

[76] In the Appendix, Table A2 confirms that individual's health conditional on initial risk for disease is mostly sensitive to changes in prices of sugary items.

[77] Even if I would allow for time preferences to vary over time, individual differences in impatience seem to be quite stable (Mischel et al., 1989).

might incorporate expected inflation and the uncertainty in the future cash flow in addition to preferences for current consumption, I include indicators for whether one expects inflation, economic crisis or improved economic wellbeing within one to three years.[78] In my sample of urban adults, around 50 percent of individuals are considered impatient and half of them as relatively more patient (See Figure 1.17). Other controls are the same as listed in equation (1.2).

The results of equation 1.4 are reported in Table 1.15 and Table 1.16, where most impatient individuals are labeled as Impatient(5) or (4+). Results for abdominal obesity (Table 1.15, Columns 1-3) suggest that the majority of price effect on increased waistline is driven by the most impatient people. The probability of becoming abdominally obese is on average almost twice as high for the impatient individuals relative to the patient individuals. However, results are only statistically significant at the ten percent level and lose their significance once I control prices of rich in other nutrients.[79]

Impatience and change in the prices of sugary foods combined seem to matter substantially in affecting the probability of becoming diabetic or becoming diagnosed with hypertension, especially for those at the highest risk for disease development (see Table 1.16). In particular, relatively impatient people who are at a high risk for developing type two diabetes are significantly more likely to become diabetic at the event of lower prices of sugary foods (Columns 3-5). For instance, a ten percent decrease in price of sugar increases the probability of becoming diabetic for the impatient individuals at the high risk for developing the disease by almost two percentage points more than for the patient ones. Similar holds for the case of hypertension (Columns 8 - 10). Changes of prices rich in sugar items affect hypertension for the impatient individuals regardless of initial risk for disease development (Columns 6 - 7). Results remain nearly the same when adding interaction and level terms for variables, potentially correlated with impatience, such as income, education, gender, expectation on inflation or future social status (see an example in Table A6.)

Changes in price of foods rich in other nutrients, no matter the individual's time preference or risk for disease, have no significant effect on health outcomes at conventional levels (Table 1.16, Columns 5 and 10). A decrease in prices of foods rich in fat, however, do suggest a decrease in the probability of becoming diabetic for the relatively more impatient individuals by around 0.1 percentage point more compared to the relatively more patient ones for each percent decrease in prices.[80] One of the possible explanations for this could be that a decrease in prices of fatty foods increases consumption of fatty foods at the expense of decreased consumption of sugar. This might decrease the probability of becoming diabetic if fat is relatively less harmful to health than sugar.[81] This intuition is consistent with the

[78]Since this information is available in 2005 and 2009 only, I generate a missing variable indicator variable and replace missing values with zeros.

[79]Results do not change if Impatient(4+) is interacted with prices of foods rich in other nutrients.

[80]The coefficient is not significant, yet suggestive of that direction.

[81]In my data, I find a significant and positive relationship between soda consumption and abdominal obesity. Also, given that 1 gram of fat contains double the calories of 1 gram of sugar, this could (only) intuitively suggest that where calories come from matter.

recent findings by Harding and Lovenheim (2014), where households substituted away from consuming relatively more expensive fats to increased soda consumption. Since impatient individuals are more responsive to current costs, a change in nutritional composition of their diet might therefore be larger and the shift towards healthier foods large enough to translate into better health (Crescenzo et al., 2014; Atkins, 2002).

I check whether results presented above are consistent with other measures that could serve as proxies for impatience. I use the question on whether an individual considers the future or not when making financial or savings decisions. Since these statements are about the attitude referring to trade-offs between the present and the future, they could be a good reflection of one's discount rate (Borghans and Golsteyn, 2006). Hence, I construct an indicator variable that equals 1 if people never think about the future when making financial or savings decisions and 0 otherwise. Almost 40 percent of people say that they never think about their future when making financial or savings decisions. Table 1.17 supports my results from Table 1.16 and Table 1.15. I observe that health, such as diabetes, waistline or abdominal obesity, of those relatively impatient individuals is not only significantly more responsive to prices of sugar, but is also directly correlated to one's health status. For instance, impatient high risk individuals are significantly more likely to be diagnosed as diabetic or have a larger waistline compared to relatively patient ones, irrespective of prices. Impatience does not matter for the onset of hypertension directly. In addition to the results above, I re-estimate equation 1.4 using another measure of impatience, proxied with the income management variable. I expect that respondents with higher discount rates are more inclined to spend money immediately. Individuals are asked the following question:"Now imagine that you have a rich relative who gives you $1,000 pesos today. In the next 30 days, would you spend all of it, save all of it, or spend a portion and save the other?" I define someone as very impatient or impatient if they spend more than 90 or 50 percent of the offered income, respectively. Results in Table A5 are very similar to those in Table 1.4, especially in the case of diabetes.

1.7 Conclusion

Despite the pervasive price-based policy discussions on curbing the obesity and the chronic disease epidemic, there exists no rigorous evidence on whether changes in prices of foods, rich in sugar or fat in particular, alter dietary intake enough to translate into decreased prevalence of obesity or diet-related chronic diseases. Evidence from developing countries, where significant variability in relative food prices is expected to have a particularly strong effect on health due to a higher share of food costs in family income, is non-existent. The plentiful substitutes of foods available, the productive relationship between nutrients and health and one's existing health mediating these relationships make it hard to unambiguously predict the effect of relative price changes on health.

Using a unique, detailed bar-coded price dataset with product-specific nutritional information, combined with the overlapping health panel data at the state and individual level, I estimate the effect of price changes of foods rich in sugar on obesity, abdominal obesity, diabetes, and hypertension in Mexico over the period 1996-2010. Reduced form results show that a decrease in prices of sugary foods increases abdominal obesity, type two diabetes and hypertension significantly, yet changes in prices of foods rich in other nutrients have no significant effect on health. While previous literature on this topic has mostly focused on contemporaneous or one year lagged effects of price changes on health, this empirical analysis shows that the effect of price change in foods rich in sugar is strongest within the first year and diminishes over a period of four years. A simple calibration based on these estimates suggests that decreases in prices of sugary foods in Mexico explain around one-fifth of the steep increase in diabetes prevalence over the past 15 years.

Since diabetes and hypertension are both chronic diseases, preventing those highly pre-diabetic or pre-hypertensive from becoming diagnosed with diabetes or hypertension is crucial. Empirical analysis suggests non-linear effects of prices of foods rich in sugar, depending on one's risk for developing chronic disease. In particular, I show that over time prices of foods rich in sugar have negative effects across the entire health distribution, yet the effect is strongest for those at the highest risk for developing chronic disease. By investigating the different possible mechanisms through which these effects materialize, I show that the most impatient individuals are those most responsive to changes in sugary food prices. This suggests that time preferences are an important mechanism driving these results.

Overall, my results suggest that prices of foods rich in sugar in particular play an important role in abdominal obesity and the chronic disease epidemic in a middle income country or among poor households in developed economies, who, incidentally, are at highest risk for obesity and diet-related chronic diseases. Importantly, however, whether these effects translate into individual's welfare remains an open question. While increased prices of sugary foods do translate to better health, the consumer welfare costs remain uncertain, and could potentially be larger for those less advantaged. In addition, despite these relationships suggesting that relatively cheaper sugary items lead to worse health, they do not explain whether consumption of sugary foods in particular harmed health the most. These relationships only suggest that prices of foods particularly rich in sugar stimulate consumption of nutritional bundles in such way so that they affected obesity, and incidence rates of diabetes or hypertension. Comprehensive overlapping dataset on consumption and health would be needed to fully assess these questions. Further research should also improve the understanding of the role of behavior (e.g. self-control) as the mechanism in health.

Table 1.1: Food Categories

FOOD CATEGORY	PRODUCT GROUPS
1 **Grains**	Tortillas, corn and wheat flour, corn, corn dough, toasted tortillas, bread, pastries, pasta, cereals, rice
2 **Snacks & Candy**	Cookies, candies and caramels, jams, honey, chokolate in powder and tablet, gelatin, sugar
3 **Meat**	Chicken, beef, pork, ham, sausages (eg. chorizo, salchichas, embutidos...), enchiladas, dried meat, bacon, seafood (eg. fish, squid, shrimps), canned tuna and sardines
4 **Spices**	Salt, pepper, cinnamon, cumin
5 **Oils**	Margarine (with and without salt), butter, oil (eg. olive, soy, canola oil), lard
6 **Cold beverages**	Juice, juice powders, syrups
7 **Warm beverages**	Tea, coffee and capuccino (in powder or liquid/made)
8 **Fruits**	Fresh and canned fruits
9 **Vegetables**	Fresh, packaged and canned vegetables, fresh and dried chiles
10 **Prepared food**	Instant soups, french fries, tomato and other cream sauces, pizza, roasted chicken, roasted peanuts, ketchup, mayonnaise, mustard, bouillon cube, fried meatballs, barbecued meat, prepared baby food
11 **Dairy**	Condensed milk, powdered milk, cheeses, yoghurt, cream, eggs, ice cream
12 **Sodas**	Diet and regular sodas
13 **Milk**	Milk

Note: Alcoholic beverages are excluded from the analysis. Manufacturers are not obliged to report their nutritional composition, therefore nutritional composition for most alcoholic items is missing.

Table 1.2: Summary statistics - Incidence Rates and MxFLS

Variable	Obs	Mean	Std. Dev.	Min	Max
Incidence Rate					
Annual Yearbook					
Diabetes II	495	357.636	118.461	114.69	816.3
Diabetes I	363	11.928	10.667	0	58.81
Hypertension	495	559.3	230.349	151.36	1402.68
Hospital discharges					
Diabetes	25895	41.303	66.319	0	990.099
Diabetes 1	25372	.87	4.682	0	99.602
Obesity	25902	.223	2.789	0	213.675
Cardiovascular diseases	25871	87.843	101.824	0	996.145
Hypertensive diseases	25902	16.592	39.811	0	1351.351
MxFLS					
Ind. Characteristics					
Age	22073	51.321	12.828	35	100
Sex	22073	.465	.499	0	1
School years	19177	7.682	3.614	1	13
Worked last week	18258	.545	.498	0	1
Cooked last week	17798	.616	.486	0	1
Hrs using internet pw	17792	.582	3.63	0	95
Hrs doing sports pw	17774	.572	2.532	0	50
Hrs watching tv pw	17771	10.864	9.625	0	90
Ind. Health Outcomes					
Overweight or/and obese	16235	.781	.414	0	1
Obese	16228	.354	.478	0	1
Abdominally obese	16315	.777	.416	0	1
Diabetes II	17055	.117	.321	0	1
Hypertension	17812	.209	.406	0	1
Child obesity	13341	.108	.31	0	1
HH Characteristics					
HH size	15407	4.305	2.083	1	21
Room N	15325	2.162	1.072	1	14
Telephone	15338	.521	.5	0	1
Electricity	15388	.977	.149	0	1
Own house	14759	.739	.439	0	1
Tap in dwelling	15404	.236	.425	0	1
Monthly soda consumption PAE (l)	10290	7.457	10.251	0	252.583

Note: Incidence rates are expressed per 100,000 population. Annual Yearbook Data records average state-year incidence rates, whereas incidence rates obtained from hospital discharge data are expressed at the municipality level for adults older than 35 years old, and younger than 15 years old for type 1 diabetes incidence rate. Summary statistics for individuals is reported for urban adults older than 35 years old. Child obesity is reported for children between 6 and 18 years old living in the urban areas.

Table 1.3: Summary statistics - Nutritional Composition of Prices

Group Of Products	Obs	Mean	Std. Dev.	Min	Max
Rich in Sugar Content					
Grams of **Sugar** per 100g	46	46.9	4.708	35.525	57.483
Grams of Protein per 100g	46	1.218	.365	.767	2.46
Grams of Fats per 100g	46	.867	.417	.309	1.543
Grams of Sodium per 100g	46	.06	.031	.031	.216
Grams of Fiber per 100g	46	.588	.403	.158	2.38
Rich in Fat Content					
Grams of Sugar per 100g	46	.003	.006	0	.017
Grams of Protein per 100g	46	6.846	1.641	3.743	10.396
Grams of **Fats** per 100g	46	72.953	1.747	67.66	75.395
Grams of Sodium per 100g	46	.497	.082	.365	.717
Grams of Fiber per 100g	46	.076	.105	0	.52
Rich in Protein and Fat Content					
Grams of Sugar per 100g	46	1.161	.41	.302	2.259
Grams of **Protein** per 100g	46	17.187	.933	15.162	20.084
Grams of **Fats** per 100g	46	22.601	.858	20.89	25.151
Grams of Sodium per 100g	46	.482	.051	.405	.602
Grams of Fiber per 100g	46	.03	.05	0	.298
Rich in Sugar and Fat Content					
Grams of **Sugar** per 100g	46	35.921	6.164	17.038	45.845
Grams of Protein per 100g	46	5.842	1.146	4.216	8.852
Grams of **Fats** per 100g	46	10.488	2.395	5.029	15.587
Grams of Sodium per 100g	46	.129	.043	.072	.256
Grams of Fiber per 100g	46	.644	.562	.069	1.951

Note: Nutritional content (per city per 'price') is reported for 2008 items, and is very similar for other years.

Table 1.4: Supermarkets and Prices

Dep.Var: $Log(Price_{sugar})$	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
$Supermarkets_t$	-0.004	0.002	0.002	0.000	0.003
	(0.004)	(0.006)	(0.006)	(0.005)	(0.006)
$Supermarkets_{t-1}$		-0.000	-0.000	-0.001	-0.000
		(0.005)	(0.005)	(0.005)	(0.006)
$Supermarkets_{t-2}$		0.001	0.001	0.001	0.001
		(0.010)	(0.010)	(0.010)	(0.011)
$Supermarkets_{t-3}$		-0.017	-0.022	-0.021	-0.022
		(0.008)**	(0.013)*	(0.012)*	(0.013)
$Supermarkets_{t-4}$			0.007	0.009	0.007
			(0.014)	(0.014)	(0.014)
N	224	224	224	224	224
Adj. R^2	0.02	0.03	0.03	0.03	0.02
Y Mean					4.60
N Clusters	32	32	32	32	32
State FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Controls	No	No	No	No	Yes
Region Trends	No	No	No	Yes	No

Supermarkets refer to stores such as Walmart supercenters and Carreffour. Controls are state level log GDP, and number of fast food restaurants. Robust standard errors in parentheses, clustered at the state level. */**/*** denotes significant at the 10% / 5% / 1% levels.

Table 1.5: Diabetes II Incidence Rates - Panel A

Panel A					
Dep. Var: Diabetes II	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
$Log(P_{sugar})_t$	-93.942	-23.702	-22.128	-22.945	-23.625
	(20.072)***	(20.550)	(19.583)	(19.705)	(19.874)
$Log(P_{sugar})_{t-1}$		-86.782	-83.978	-83.986	-82.625
		(21.111)***	(19.897)***	(20.084)***	(21.597)***
$Log(P_{sugar})_{t-2}$		-24.936	-31.116	-30.925	-31.447
		(17.067)	(14.783)**	(14.966)**	(16.212)*
$Log(P_{sugar})_{t-3}$		-29.770	-27.084	-27.117	-27.822
		(15.873)*	(16.756)	(16.671)	(16.378)*
$Log(P_{sugar})_{t-4}$		-0.879	-2.829	-2.134	-3.510
		(22.409)	(22.220)	(22.678)	(24.068)
$Log(P_{fats})_{t-1}$					-6.587
					(55.198)
$Log(P_{protein})_{t-1}$					-4.722
					(27.584)
$Log(P_{fiber})_{t-1}$					-10.476
					(30.502)
N	351	351	351	351	351
Adj. R^2	0.35	0.40	0.41	0.40	0.40
F-Stat	10.83	11.26	15.95	13.35	13.17
Y Mean					358
Y SD					116
N Clusters	32	32	32	32	32
Long-run effect					-169.029 (34.795)***
State FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Macro Controls	.	.	Yes	Yes	Yes
FastFood&Adv	.	.	.	Yes	Yes

Incidence rate is defined as per 100000 population. Controls 1 include state-level log of GDP and drought index. Controls 2 include state-year level log of number of fast food services per km^2 and their advertising expenditures per capita. Results are not sensitive on neither years, inclusion/exclusion of control variables one by one, nor on other nutrients price lags used. Total effect is calculated up to the fourth lag. Robust standard errors in parentheses, clustered at the state level. */**/*** denotes significant at the 10% / 5% / 1% levels.

Table 1.6: Hypertension Incidence Rates - Panel B

Panel B					
Dep. Var: Hypertension	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
$Log(P_{sugar})_t$	-166.373	-94.274	-89.421	-87.511	-89.801
	(48.829)***	(42.808)**	(40.949)**	(39.894)**	(40.407)**
$Log(P_{sugar})_{t-1}$		-65.648	-70.734	-73.398	-69.101
		(37.584)*	(37.419)*	(38.464)*	(39.744)*
$Log(P_{sugar})_{t-2}$		-50.707	-47.812	-49.660	-50.413
		(26.526)*	(28.720)	(28.545)*	(29.304)*
$Log(P_{sugar})_{t-3}$		-87.641	-88.228	-87.832	-89.356
		(35.838)**	(35.966)**	(36.412)**	(37.333)**
$Log(P_{sugar})_{t-4}$		-16.864	-19.464	-22.775	-26.903
		(46.669)	(46.043)	(47.933)	(48.541)
$Log(P_{fats})_{t-1}$					-11.813
					(87.579)
$Log(P_{protein})_{t-1}$					-18.374
					(36.543)
$Log(P_{fiber})_{t-1}$					-27.738
					(44.779)
N	351	351	351	351	351
Adj. R^2	0.65	0.68	0.69	0.69	0.68
F-Stat	47.43	43.43	51.84	51.40	100.31
Y Mean					548.55
Y SD					226.00
N Clusters	32	32	32	32	32
Long-run effect					-325.574 (82.020)***
State FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Macro Controls	.	.	Yes	Yes	Yes
FastFood&Adv	.	.	.	Yes	Yes

Incidence rate is defined as per 100000 population. Controls 1 include state-level log of GDP and drought index. Controls 2 include state-year level log of number of fast food services per km^2 and their advertising expenditures per capita. Results are not sensitive on neither years, inclusion/exclusion of control variables one by one, nor on other nutrients price lags used. Total effect is calculated up to the fourth lag. Robust standard errors in parentheses, clustered at the state level. */**/*** denotes significant at the 10% / 5% / 1% levels.

Table 1.7: Diabetes II and Hypertension - Robustness Checks

Check:	**Reverse Causality**		**Placebo**				**Region x Year**		**State Trend**	
	Diab II	**Hyper**	**Diab I**	**Diab I**	**Asthma**	**Asthma**	**Diab II**	**Hyper**	**Diab II**	**Hyper**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**
$Log(P_{sugar})_t$	-18.580	-90.239	-0.394	-0.570	-12.572	-18.176	-22.785	-93.856	-25.753	-64.131
	(27.993)	(41.669)**	(2.759)	(2.694)	(58.575)	(56.589)	(26.554)	(51.201)*	(20.160)	(40.328)
$Log(P_{sugar})_{t-1}$	-107.452	-113.239	1.537	1.754	22.945	26.283	-104.761	-84.407	-90.961	-82.109
	(30.285)***	(49.341)**	(3.795)	(3.815)	(35.300)	(34.312)	(20.408)***	(40.232)**	(26.014)***	(43.083)*
$Log(P_{sugar})_{t-2}$	-36.978	-72.343	-1.408	-1.098	-22.807	-22.667	-29.337	-59.174	-33.039	-42.048
	(14.807)**	(32.658)**	(1.686)	(1.576)	(31.281)	(30.854)	(18.556)	(32.436)*	(18.712)*	(30.190)
$Log(P_{sugar})_{t-3}$	7.351	13.285	3.519	3.604	-32.883	-31.556	-50.681	-105.188	-17.883	-62.308
	(23.058)	(44.088)	(3.490)	(3.524)	(27.938)	(27.366)	(22.349)**	(40.695)**	(19.282)	(35.761)*
$Log(P_{sugar})_{t-4}$	-27.855	-88.647	-4.616	-4.821	3.692	-3.500	9.623	-18.316	-8.463	-25.506
	(20.411)	(64.564)	(4.316)	(4.429)	(37.574)	(38.143)	(22.988)	(51.785)	(27.218)	(55.719)
$Log(P_{sugar})_{t+1}$	7.203	30.188								
	(21.277)	(37.496)								
$Log(P_{sugar})_{t+2}$	-7.421	-5.027								
	(20.482)	(53.785)								
$Log(P_{sugar})_{t+3}$	-9.700	-41.714								
	(33.284)	(49.746)								
$Log(P_{fats})_{t-1}$				1.254		11.985				
				(4.151)		(41.095)				
$Log(P_{protein})_{t-1}$				-3.185		-56.360				
				(3.386)		(39.717)				
$Log(P_{fiber})_{t-1}$				1.090		-17.391				
				(3.333)		(49.828)				
N	287	287	320	320	340	340	351	351	351	351
Adj. R^2	0.44	0.71	0.30	0.30	0.05	0.06	0.44	0.69	0.84	0.89
Y Mean	358.08	559.90	11.93	11.93	317.00	317.00	358.08	559.90	358.08	559.90
Y SD	120.04	232.94	10.80	10.80	196.18	196.18	120.04	232.94	120.04	232.94
N Clusters	32	32	32	32	31	31	32	32	32	32

All regressions include state, year fixed effects, controls 1 and controls 2 (see Table 1.6 for their detailed description). Robust standard errors in parentheses, clustered at the state level. */**/*** denotes significant at the 10% / 5% / 1% levels.

Table 1.8: Children Obesity

Dep.Var: Obese Child	**6≤All≤18**			**6≤Girls≤18**	**6≤Boys≤18**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**
$Log(P_{sugar})_t$	-0.035	-0.032	-0.032	-0.070	0.011
	(0.023)	(0.018)*	(0.018)*	(0.026)**	(0.039)
N	11958	11958	11958	6087	5871
Adj. R^2	0.00	0.01	0.01	0.01	0.01
F-Stat	2.47	16.00	23.95	6.92	20.05
Y Mean			0.11	0.11	0.11
Y SD			0.31	0.31	0.31
N Clusters	29	29	29	29	29
Individual FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Controls 1	.	Yes	Yes	Yes	Yes
Controls 2	.	.	Yes	Yes	Yes

Controls 1 include age, household size, SES decile indicator, and log of annual household's income. Controls 2 include state-level log(GDP), drought index, log of number of fast food services and their advertising expenditures at the municipality level. Robust standard errors in parentheses, clustered at the city level. */**/*** denotes significant at the 10% / 5% / 1% levels.

Table 1.9: Adult Obesity - BMI

Panel A				
DV: Log(BMI)	**(1)**	**(2)**	**(3)**	**(4)**
$Log(P_{sugar})_t$	-0.015	-0.016	-0.016	-0.016
	(0.010)	(0.009)*	(0.009)*	(0.010)
$Log(P_{fats})_t$				0.002
				(0.013)
$Log(P_{fiber})_t$				0.004
				(0.013)
N	21591	21591	21591	21591
Adj. R^2	0.06	0.06	0.06	0.06
F-Stat	158.98	157.66	320.24	401.69
Y Mean				27.80
Y SD				5.27
N Clusters	32	32	32	32

Panel B				
DV: Obese	**(1)**	**(2)**	**(3)**	**(4)**
$Log(P_{sugar})_t$	-0.032	-0.037	-0.036	-0.055
	(0.024)	(0.027)	(0.027)	(0.032)*
$Log(P_{fats})_t$				0.055
				(0.037)
$Log(P_{fiber})_t$				-0.019
				(0.039)
N	21589	21589	21589	21589
Adj. R^2	0.03	0.03	0.03	0.03
F-Stat	94.04	223.25	390.37	555.18
Y Mean				0.29
Y SD				0.45
N Clusters	32	32	32	32
Controls 1	.	Yes	Yes	Yes
Controls 2	.	.	Yes	Yes

Individual and year fixed effects are included in all regressions. Abdominal obesity is an indicator based on one's waistline, and equals 1 if waistline is larger than 90 and 80 for men and women, respectively. Controls 1 include Individual and household controls such as age, years of schooling, working status, household size, SES decile indicator, log of annual household's income, and distance to dearest city in km. Controls 2 include state-level log(GDP), and drought index. Additional controls are log of number of fast food restaurants and services and their total advertising expenditures at the municipality level. Results remain unchanged if weekly hours of doing sports, watching tv, using the internet, an indicator on whether you cooked last week, or proxy for health awareness variable are added as additional controls. Robust standard errors in parentheses, clustered at the city level. */**/*** denotes significant at the 10% / 5% / 1% levels.

Table 1.10: Adult Obesity - Waistline

Panel A				
DV: Log(waistline)	**(1)**	**(2)**	**(3)**	**(4)**
$Log(P_{sugar})_t$	-0.035	-0.042	-0.041	-0.038
	(0.025)	(0.020)**	(0.020)**	(0.021)*
$Log(P_{fats})_t$				-0.033
				(0.021)
$Log(P_{fiber})_t$				-0.031
				(0.035)
N	21753	21753	21753	21753
Adj. R^2	0.28	0.28	0.28	0.29
F-Stat	319.84	361.24	333.15	268.97
Y Mean				90.13
Y SD				13.36
N Clusters	32	32	32	32
Panel B				
DV: Abd. Obese	**(1)**	**(2)**	**(3)**	**(4)**
$Log(P_{sugar})_t$	-0.129	-0.145	-0.142	-0.148
	(0.066)*	(0.049)***	(0.049)***	(0.053)***
$Log(P_{fats})_t$				-0.040
				(0.051)
$Log(P_{fiber})_t$				-0.087
				(0.095)
N	21753	21753	21753	21753
Adj. R^2	0.12	0.12	0.12	0.12
Y Mean				0.66
Y SD				0.47
N Clusters	32	32	32	32
Long-run effect				-0.243
(over 5 years)				(0.099)**
Controls 1	.	Yes	Yes	Yes
Controls 2	.	.	Yes	Yes

Individual and year fixed effects are included in all regressions. Abdominal obesity is an indicator based on one's waistline, and equals 1 if waistline is larger than 90 and 80 for men and women, respectively. Controls 1 include Individual and household controls such as age, years of schooling, working status, household size, SES decile indicator, log of annual household's income, and distance to dearest city in km. Controls 2 include state-level log(GDP), and drought index. Additional controls are log of number of fast food restaurants and services and their total advertising expenditures at the municipality level. Results remain unchanged if weekly hours of doing sports, watching tv, or using the internet, an indicator on whether you cooked last week, or proxy for health awareness variable are added as additional controls. Robust standard errors in parentheses, clustered at the city level. */**/*** denotes significant at the 10% / 5% / 1% levels.

Table 1.11: Diabetes II and Hypertension

Panel A				
DV: Diabetes II	**(1)**	**(2)**	**(3)**	**(4)**
$Log(P_{sugar})_t$	-0.052	-0.048	-0.043	-0.055
	(0.015)***	(0.015)***	(0.015)***	(0.015)***
$Log(P_{fats})_t$				0.008
				(0.016)
$Log(P_{fiber})_t$				-0.027
				(0.017)
N	11485	11485	11485	11485
Adj. R^2	0.05	0.05	0.05	0.05
F-Stat	78.19	127.29	175.04	110.21
Y Mean				0.11
Y SD				0.31
N Clusters	29	29	29	29
Panel B				
DV: Hypertension	**(1)**	**(2)**	**(3)**	**(4)**
$Log(P_{sugar})_t$	-0.019	-0.014	-0.013	-0.029
	(0.036)	(0.034)	(0.032)	(0.036)
$Log(P_{fats})_t$				0.015
				(0.038)
$Log(P_{fiber})_t$				-0.056
				(0.055)
N	11758	11758	11758	11758
Adj. R^2	0.04	0.04	0.04	0.04
F-Stat	46.45	74.57	70.11	133.69
Y Mean				0.18
Y SD				0.38
N Clusters	29	29	29	29
Controls 1	.	Yes	Yes	Yes
Controls 2	.	.	Yes	Yes

Individual and year fixed effects are included in all regressions. Diabetes is an indicator that equals 1 if you were ever diagnosed with non-insulin dependent diabetes after age 35. Hypertension is an indicator that equals 1 if you were ever diagnosed with hypertension or, if missing, diagnosed with hypertension based on the measure (above 140mmHg systolic or 90mmHg diastolic blood pressure). Controls 1 include Individual and household controls such as age, years of schooling, working status, household size, SES decile indicator, log of annual household's income. Macro controls include state-level log(GDP), and drought index. Additional controls are log of number of fast food restaurants and services and their total advertising expenditures at the municipality level. Results remain unchanged if weekly hours of doing sports, watching tv, or using the internet, an indicator on whether you cooked last week, or proxy for health awareness variable are added as additional controls. Robust standard errors in parentheses, clustered at the city level. */**/*** denotes significant at the 10% / 5% / 1% levels.

Table 1.12: Obesity, Diabetes II and Hypertension - Robustness Checks

Check:	**Reverse Causality**			**Placebo**	**Region x Year FE**			**Risk for Disease Trend**
	Abd.Obese	**Diab II**	**Hyper**	**Smoking**	**Abd.Obese**	**Diab II**	**Hyper**	**Diab II**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
$Log(P_{sugar})_t$	-0.040	-0.051	-0.021	0.002	-0.041	-0.034	0.032	-0.050
	(0.023)*	(0.022)**	(0.046)	(0.028)	(0.029)	(0.025)	(0.052)	(0.016)***
$Log(P_{sugar})_{t+1}$	-0.001	0.011	0.009					
	(0.024)	(0.019)	(0.066)					
N	22367	11683	11959	26820	20968	11207	11442	11960
Adj. R^2	0.28	0.05	0.04	0.01	0.29	0.06	0.04	0.06
Y Mean	0.66	0.12	0.14	0.21	0.66	0.12	0.20	0.12
Y SD	0.47	0.32	0.34	0.41	0.47	0.32	0.40	0.32
N Clusters	32	29	29	32	29	28	28	29
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls 1	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls 2	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Smoking is an indicator for whether you smoke or not. Controls 1 include Individual and household controls such as age, years of schooling, working status, household size, SES decile indicator, log of annual household's income. Controls 2 include state-level log(GDP), and drought index. Additional controls are log of number of fast food restaurants and services and their total advertising expenditures at the municipality level. Results remain unchanged if weekly hours of doing sports, watching tv, or using the internet, an indicator on whether you cooked last week, or proxy for health awareness variable are added as additional controls. Robust standard errors in parentheses, clustered at the city level. */**/*** denotes significant at the 10% / 5% / 1% levels.

Table 1.13: Diabetes II and Hypertension - Risk for disease

DV	**Diabetes II**	**Hypertension**
	(1)	**(2)**
$Log(P_{sugar})_t$	0.028	0.041
	(0.039)	(0.057)
Med Risk Diab x $Log(P_{sugar})_t$	-0.033	
	(0.065)	
High Risk Diab x $Log(P_{sugar})_t$	-0.194	
	(0.084)**	
High Risk Hyper x $Log(P_{sugar})_t$		-0.209
		(0.089)**
N	10524	10727
Adj. R^2	0.06	0.03
Y Mean	0.12	0.20
Y SD	0.32	0.40
N Clusters	29	29
Individual FE	Yes	Yes
Year FE	Yes	Yes
Controls 1	Yes	Yes
Controls 2	Yes	Yes

People who scored at least 10, between or below 6 points on the Finnish Diabetes Risk Score, are considered as high, moderate or low risk group for type 2 diabetes, respectively (see Figure 1.15). Parental history is adjusted so that 3 points are given in the case of more than one household member having diabetes. The question on your daily vegetable consumption is excluded, and question 5 only related to high glucose level during pregnancy. People at high risk for hypertension are those scoring at least 4 points, each risk factor counting 1 point (initially obese, abdominally obese, smoking and not exercising, experiencing sleeping problems and stress, being diabetic, and being older than 45 years old). Since test scores only go from 1-8, individuals are divided into only two groups, high and low risk for hypertension. See Appendix, Table A2 for other nutrients. Results are not particularly sensitive on cut-offs or group of risk factors included. Robust standard errors in parentheses, clustered at the city level. */**/*** denotes significant at the 10% / 5% / 1% levels.

Table 1.14: Long Run Effect by Initial Risk for Disease

	Diabetes II			**Hypertension**		
	All	**High risk**	**Low risk**	**All**	**High risk**	**Low risk**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
$Log(P_{sugar})_t$	-0.037	-0.188	0.037	-0.041	-0.105	-0.005
	(0.030)	(0.083)**	(0.043)	(0.052)	(0.086)	(0.060)
$Log(P_{sugar})_{t-3}$	-0.018	-0.027	-0.045	-0.036	-0.029	-0.042
	(0.021)	(0.064)	(0.024)*	(0.048)	(0.073)	(0.050)
$Log(P_{sugar})_{t-5}$	0.016	0.033	-0.019	-0.100	-0.025	-0.135
	(0.034)	(0.126)	(0.037)	(0.058)*	(0.118)	(0.061)**
N	10437	2273	4308	10678	3511	7167
Adj. R^2	0.06	0.06	0.05	0.03	0.03	0.04
F-Stat	576.41	132.73	847.68	122.88	54.39	98.86
Y Mean	0.12	0.28	0.04	0.20	0.25	0.17
Y SD	0.32	0.45	0.20	0.40	0.43	0.38
N Clusters	29	29	29	29	29	29
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls 1	Yes	Yes	Yes	Yes	Yes	Yes
Controls 2	Yes	Yes	Yes	Yes	Yes	Yes

People who scored at least 10 or below 6 points on the Finnish Diabetes Risk Score, are considered as high and low risk group for type 2 diabetes, respectively.[a] People at high risk for hypertension are those scoring at least 4 points, each risk factor counting 1 point (initially obese, abdominally obese, smoking and not exercising, experiencing sleeping problems and stress, being diabetic, prehypertensive and older than 45 years old). Results are not particularly sensitive on cut-offs or group of risk factors included. Robust standard errors in parentheses, clustered at the city level. */**/*** denotes significant at the 10% / 5% / 1% levels. Note that these results for illustrative purpose of disease development over 5 years span only. Reason being, initial levels of risk are assesed in the individual's first year of survey, which could be more than five years after the first price shock has taken place. Miss-measurment of the risk and consequent grouping of people mostly attenuates my results. For instance, if I believe that a risk for disease stayed constant or increased over time, then high-risk group will consists of potentially miss-allocated initally low risk individuals, who will attenuate short term price elasticity. Low risk group will consists of people who had even lower risk for disease before, and would belong to this group either way. If risk remains constant over time, problem does not exist. On the contrary, if the risk is falling over time, then individuals in a high-risk group were at even greater risk for disease development before, so their assigment to the high risk group is a correct one. Individuals in the low-risk group, however, might have been at a greater risk for disease development as assigned, and therefore should respond to contemporanous prices more than their lags. Yet, if they were responsive to $t-5$ contemporanously, they bias the lagged effect for this group downwards, so caution when interpreting these results might be necessary. I estimate a similar regression focusing on sample 2005 and 2009 and 3-year lag only (see Appendix, Table A3). Coefficients are almost the same as in Table 1.14, which gives more confidence in the results.

Table 1.15: Impatience and Obesity

DV:	**Log(waistline)**			**Abdominally Obese**		
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
$Log(P_{sugar})_t$	-0.008	-0.019	-0.009	-0.073	-0.081	-0.121
	(0.019)	(0.018)	(0.023)	(0.057)	(0.051)	(0.067)*
$Log(P_{sugar})_t$ x Impatient(5)	-0.046		-0.047	-0.112		-0.056
	(0.016)***		(0.026)*	(0.061)*		(0.084)
$Log(P_{sugar})_t$ x Impatient(+4)		-0.024			-0.087	
		(0.016)			(0.053)	
$Log(P_{fats})_t$			-0.028			0.028
			(0.019)			(0.062)
$Log(P_{fiber})_t$			-0.031			-0.095
			(0.032)			(0.088)
$Log(P_{fat})_t$ x Impatient(5)			0.013			-0.056
			(0.021)			(0.062)
$Log(P_{fiber})_t$ x Impatient(5)			-0.009			-0.019
			(0.013)			(0.039)
N	16627	16627	16627	16627	16627	16627
Adj. R^2	0.30	0.30	0.30	0.13	0.13	0.13
Y Mean			90.46			0.67
Y SD			13.31			0.47
N Clusters	32	32	32	32	32	32
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls 1	Yes	Yes	Yes	Yes	Yes	Yes
Controls 2	Yes	Yes	Yes	Yes	Yes	Yes

I also control for expectations on inflation, economic crisis and on economic well-being from 2005 and 2009, where indicator variables for missing variables are added as well and a base value of 0 is assigned to missing variable values. Results remain unchanged. In addition, coefficients do not change significantly if regression is repeated including 2005 and 2009 year only; precision is compromised due to smaller sample size. Robust standard errors in parentheses, clustered at the city level. */**/*** denotes significant at the 10% / 5% / 1% levels.

Table 1.16: Impatience and Diseases

DV:	**Diabetes II**					**Hypertension**				
	All		**High Risk**			**All**		**High Risk**		
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**
$Log(P_{sugar})_t$	-0.047	-0.046	-0.045	-0.002	0.029	0.042	0.042	0.038	0.030	0.048
	(0.021)**	(0.021)**	(0.063)	(0.057)	(0.058)	(0.039)	(0.040)	(0.064)	(0.058)	(0.072)
$Log(P_{sugar})_t$ x Impatient(5)	0.012		-0.118			-0.069		-0.188		
	(0.025)		(0.069)			(0.039)*		(0.085)**		
$Log(P_{sugar})_t$ x Impatient(4+)		0.009		-0.177	-0.247		-0.062		-0.160	-0.166
		(0.029)		(0.073)**	(0.076)***		(0.037)		(0.076)**	(0.121)
$Log(P_{fats})_t$					-0.086					-0.090
					(0.098)					(0.069)
$Log(P_{fiber})_t$					0.002					0.043
					(0.087)					(0.076)
$Log(P_{fat})_t$ x Impatient(4+)					0.184					0.079
					(0.122)					(0.101)
$Log(P_{fiber})_t$ x Impatient(4+)					-0.092					-0.085
					(0.096)					(0.065)
N	9095	9095	1965	1965	1965	9276	9276	3026	3026	3026
Adj. R^2	0.06	0.06	0.06	0.07	0.07	0.04	0.04	0.06	0.06	0.06
Y Mean		0.11			0.27		0.19			0.26
Y SD		0.32			0.44		0.40			0.44
N Clusters	29	29	29	29	29	29	29	29	29	29
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls 1	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls 2	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

I also control for expectations on inflation, economic crisis and on economic well-being from 2005 and 2009, where indicator variables for missing variables are added as well and a base value of 0 is assigned to missing variable values. Results remain unchanged. In addition, coefficients do not change significantly if regression is repeated including 2005 and 2009 year only; precision is compromised due to smaller sample size. Robust standard errors in parentheses, clustered at the city level. */**/*** denotes significant at the 10% / 5% / 1% levels.

Table 1.17: Impatience proxied with thinking about future (financial decisions)

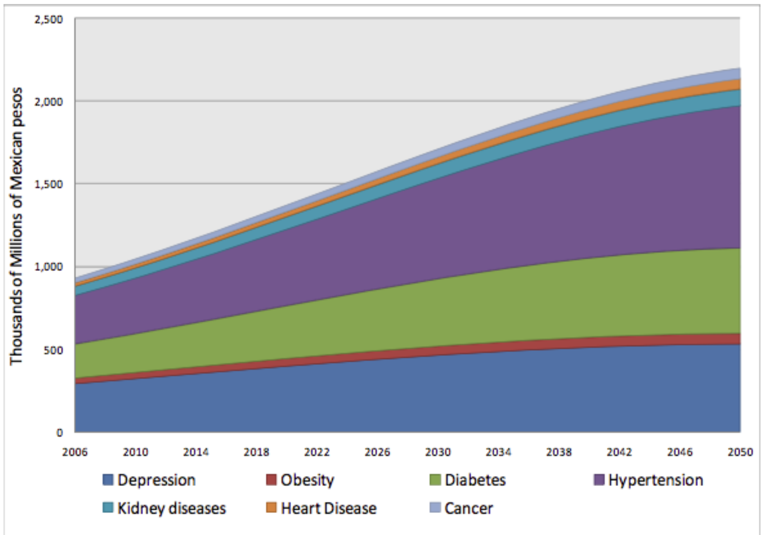
Panel A		
DV: Diabetes II	**All**	**High Risk**
	(1)	**(2)**
Log(P_sugar)	-0.042	-0.058
	(0.021)*	(0.054)
$Log(P_{sugar})_t$ x Never Future	-0.003	-0.129
	(0.037)	(0.066)*
Never Future	0.017	0.623
	(0.170)	(0.300)**
N	11187	2472
Adj. R^2	0.05	0.06
Y Mean	0.12	0.27
Y SD	0.32	0.45
N Clusters	29	29

Panel B		
DV: Hypertension	**All**	**High Risk**
	(1)	**(2)**
Log(P_sugar)	-0.008	-0.081
	(0.037)	(0.067)
$Log(P_{sugar})_t$ x Never Future	0.006	0.045
	(0.070)	(0.092)
Never Future	-0.023	-0.215
	(0.326)	(0.420)
N	11434	3790
Adj. R^2	0.05	0.05
Y Mean	0.20	0.25
Y SD	0.40	0.43
N Clusters	29	29

Panel C		
DV:	**Log(waistline)**	**Abd. Obese**
	(1)	**(2)**
Log(P_sugar)	-0.034	-0.117
	(0.020)	(0.058)*
$Log(P_{sugar})_t$ x Never Future	-0.020	-0.075
	(0.011)*	(0.060)
Never Future	0.094	0.349
	(0.051)*	(0.274)
N	20476	20476
Adj. R^2	0.34	0.14
Y Mean	90.31	0.66
Y SD	13.33	0.47
N Clusters	32	32

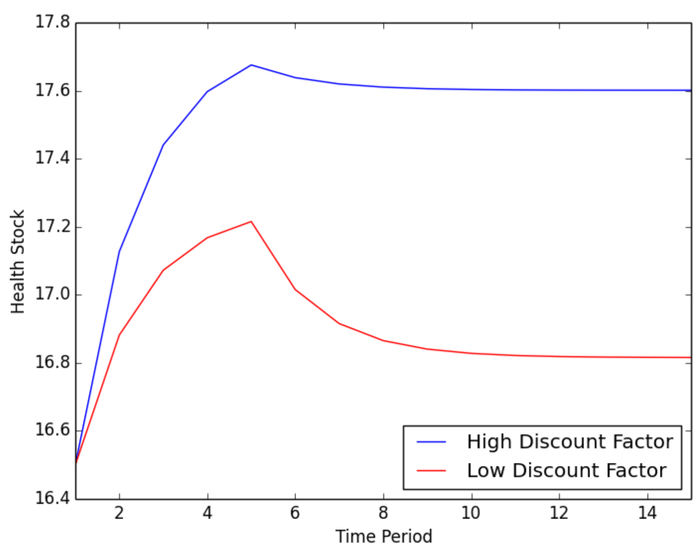
Individual and year fixed effects, controls 1 and controls 2 included in all regressions. Indicator variable Never Future equals 1 if one never thinks about future when making financial or savings decisions. Information comes from 2005 and 2009 survey and is imputed as their average for 2002. Individuals whose value was 0.5 in 2002 are assigned value 0, since I consider them relatively less impatient than those who kept their decision constant. Results do not change significantly if they are removed or assigned value 0. Control for missing/imputed variable is added as well. Controls 1 and Controls 2 are the same as in previous Tables. Robust standard errors in parentheses, clustered at the city level. */**/*** denotes significant at the 10% / 5% / 1% levels.

Figure 1.1: Projected Health Expenditure Trends by Disease in Mexico



Note: (de Salud, 2010).

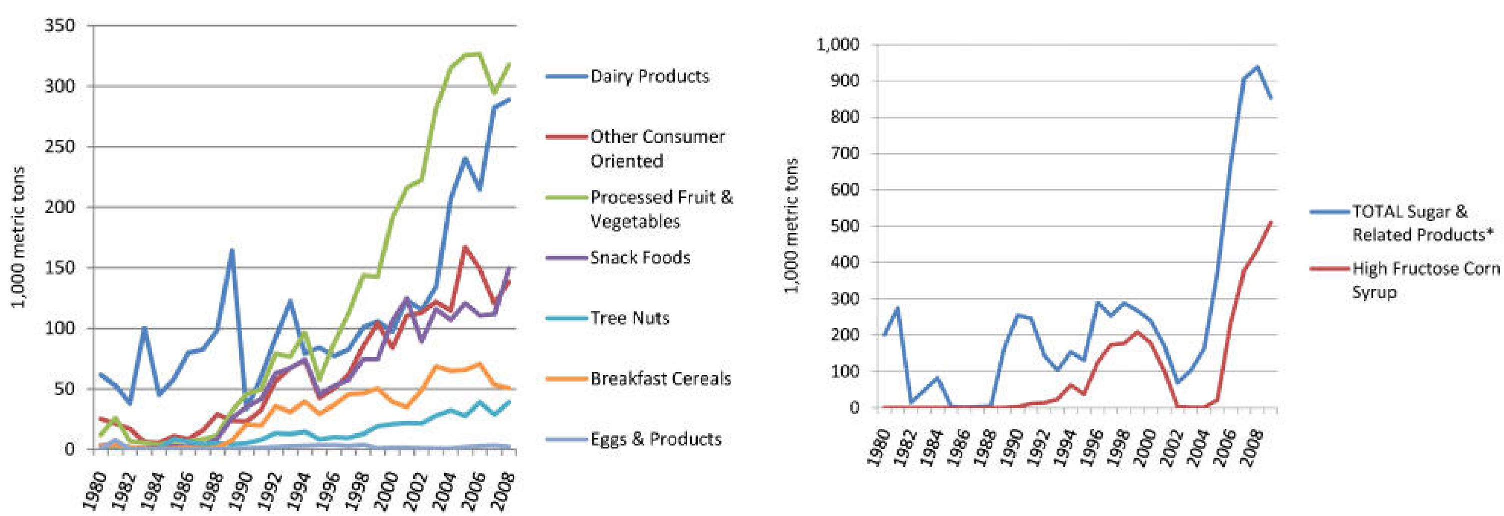
Figure 1.2: Graphical solution of the model



Note: A one-time permanent 20% decrease in price of a nutrient.

Figure 1.3: Exports to Mexico - Sugar and Related Products

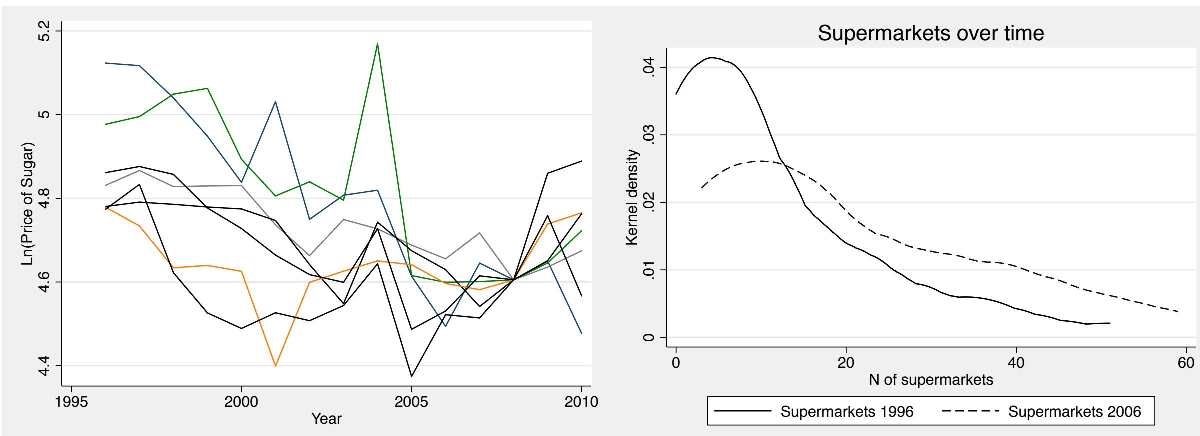
(a) (b)



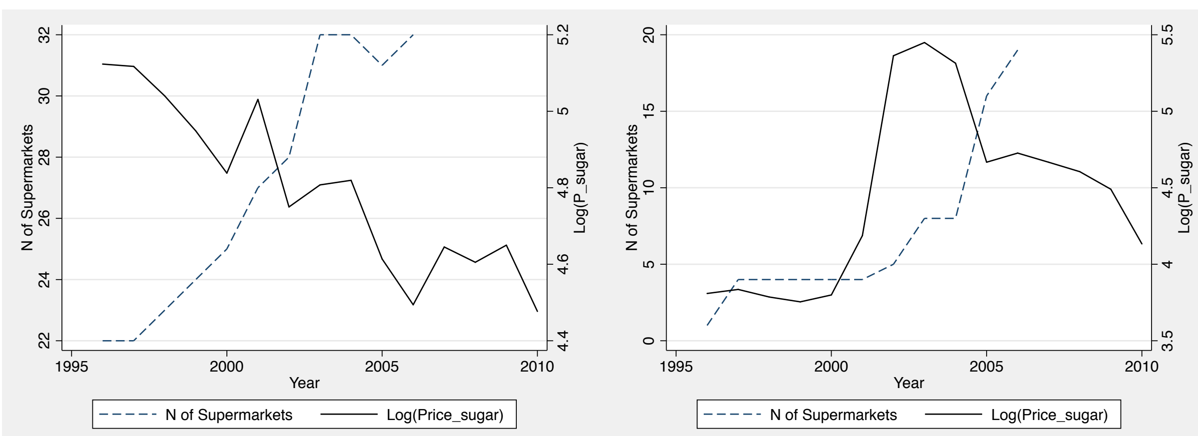
Note: Clark et al. (2012): (A) Consumer-oriented products exported to Mexico, 1980-2009. (B) Sugar and related products* exported to Mexico, 1980-2009. *Includes high fructose corn syrup and excludes honey.

Figure 1.4: Prices of Sugar-Rich Foods

(a) $Log(Price_{sugar})$ over time (b) Supermarkets per state



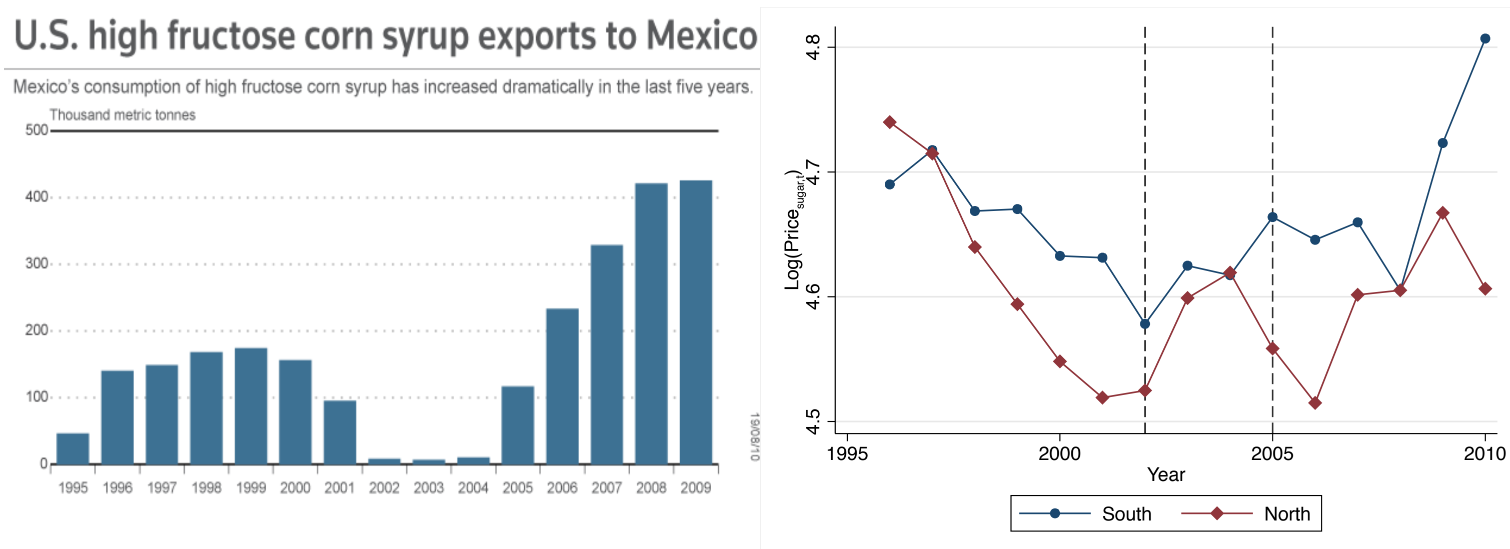
(c) Supermarkets and prices, Sinaloa (d) Supermarkets and prices, Quintana Roo



Note: Log of prices of processed foods rich in sugar, base year 2008. (A) Prices over time by states. (B) Density of supermarkets per state between 1996-2006. (C) Prices of foods rich in sugar and number of supermarkets in Sinaloa. (D) Prices of foods rich in sugar and number of supermarkets in Quintana Roo.

Figure 1.5: HFCS Tax 2002-2005

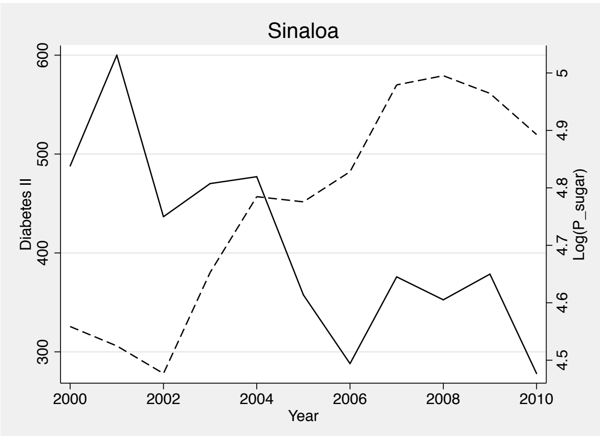
(a) (b)



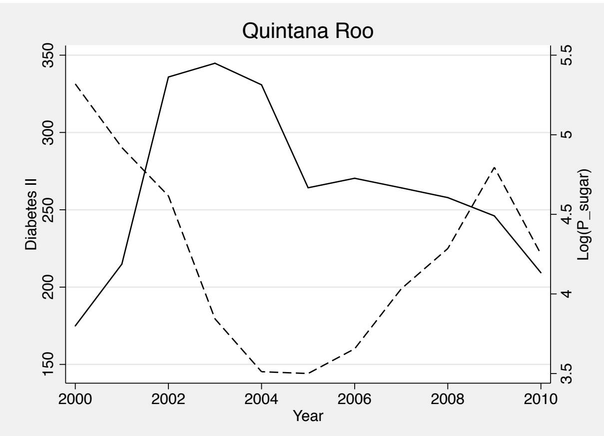
Note: Log of prices of processed foods rich in sugar. (A) Import of high fructose corn syrup from US to Mexico, Source: USDA (2010), Reuters. (B) Prices over time by Southern and Northern region. 2002-2005 HFCS 20% tax applied to products with high fructose corn syrup.

Figure 1.6: Disease Incidence and Prices of Processed Foods Rich in Sugar

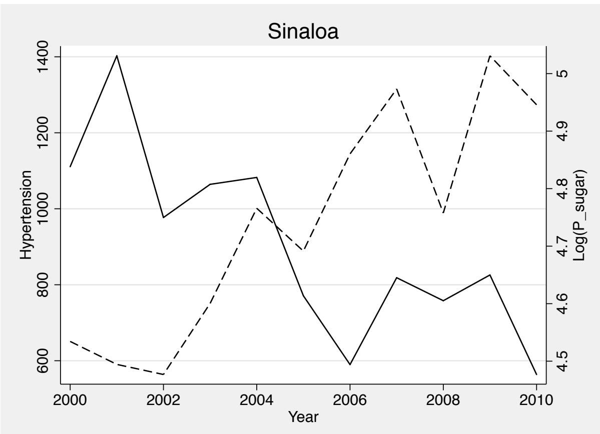
(a) Diabetes II, Sinaloa



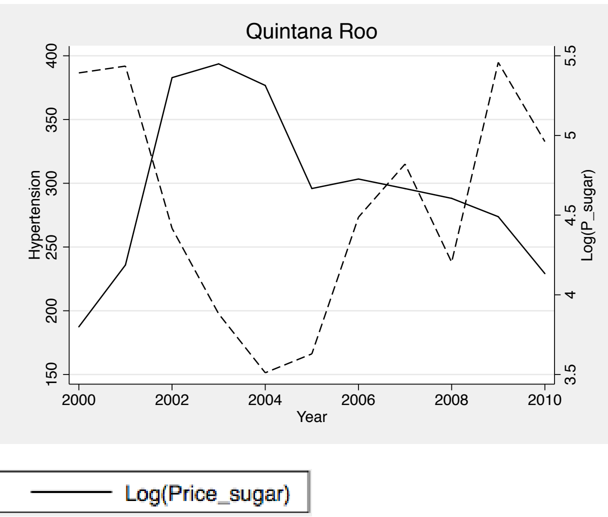
(b) Diabetes II, Quintana Roo



(c) Hypertension, Sinaloa



(d) Hypertension, Quintana Roo



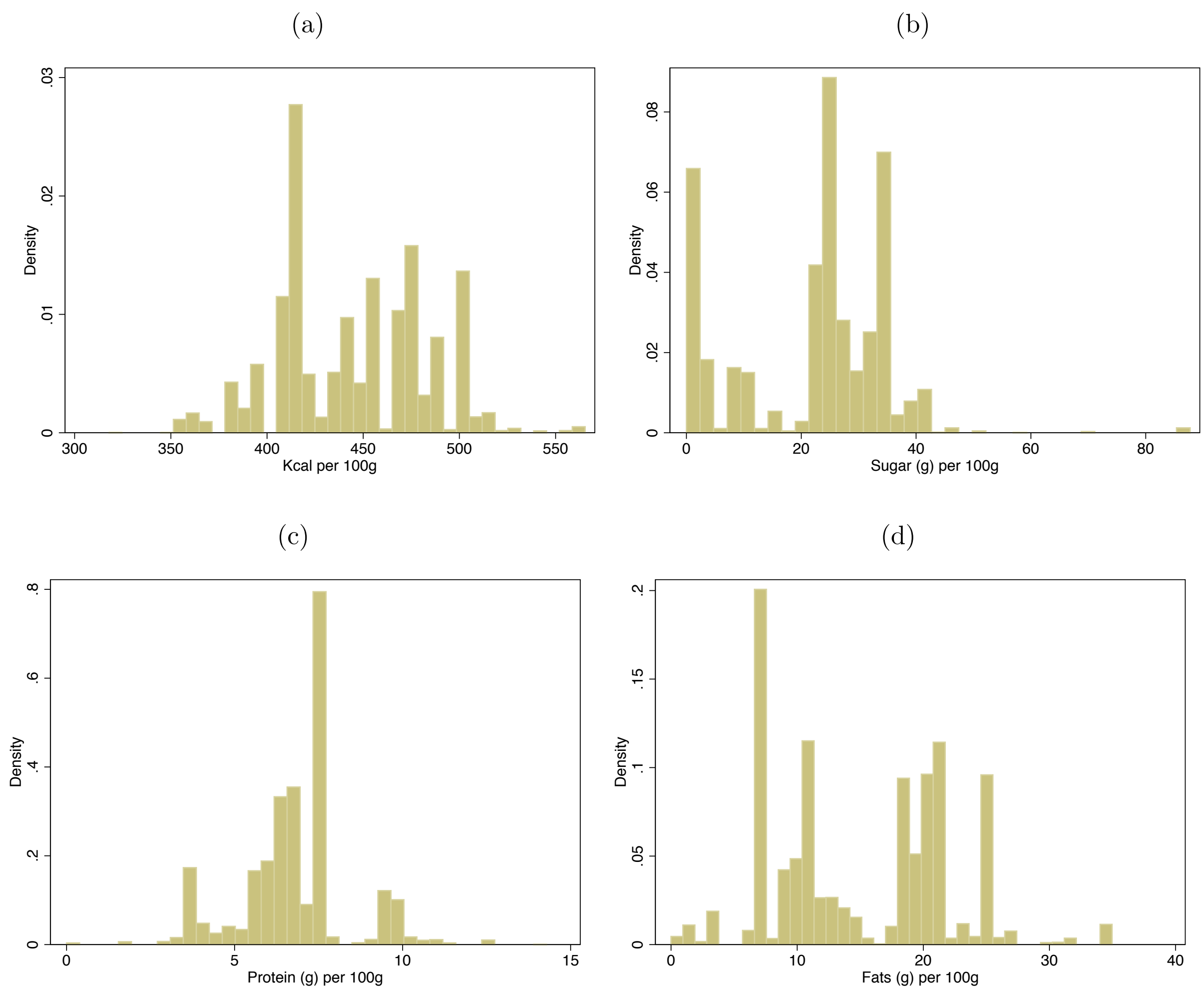
Note: Log of prices of processed foods rich in sugar and incidence per 100,000 population.

Figure 1.7: Example of price quotes in DOF

01	008001	Pasteles, pastelillos y pan dulce empaquet	74.05	1	KG	TIA ROSA, PAN DULCE, MANTECADAS, PAQ DE 105 GR
01	008002	Pasteles, pastelillos y pan dulce empaquet	111.19	1	KG	MARINELA, PASTELILLOS, CHOCO ROLES, DE PIÑA, DE 67 GR
01	008003	Pasteles, pastelillos y pan dulce empaquet	80.95	1	KG	MARINELA, PASTELILLOS, SUBMARINOS, 3 PZAS, PAQ DE 105 GR
01	008004	Pasteles, pastelillos y pan dulce empaquet	90.78	1	KG	MARINELA, PASTELILLOS, PINGÜINOS, PAQ DE 80 GR
01	008005	Pasteles, pastelillos y pan dulce empaquet	76.80	1	KG	BIMBO, PAN DULCE, MANTECADAS, PAQ DE 250 GR
01	008006	Pasteles, pastelillos y pan dulce empaquet	63.16	1	KG	BIMBO, PAN DULCE, PANQUE C/PASAS, PAQ DE 285 GR
01	008007	Pasteles, pastelillos y pan dulce empaquet	80.00	1	KG	BIMBO, PAN DULCE, MANTECADAS, PAQ DE 125 GR
01	008008	Pasteles, pastelillos y pan dulce empaquet	100.82	1	KG	SUANDY, PAN DULCE, PANQUE DE ROSCA, CAJA DE 650 GR
01	008009	Pasteles, pastelillos y pan dulce empaquet	63.16	1	KG	BIMBO, PAN DULCE, PANQUE C/PASAS, PAQ DE 285 GR
01	008010	Pasteles, pastelillos y pan dulce empaquet	92.10	1	KG	WONDER, PASTELILLOS, TUINKY, CHOCOLATE, PAQ 114 GR, C/3 PZAS
01	008011	Pasteles, pastelillos y pan dulce empaquet	95.77	1	KG	MARINELA, PASTELILLOS, CHOCO ROLES, PAQ DE 201 GR, C/6 PZAS

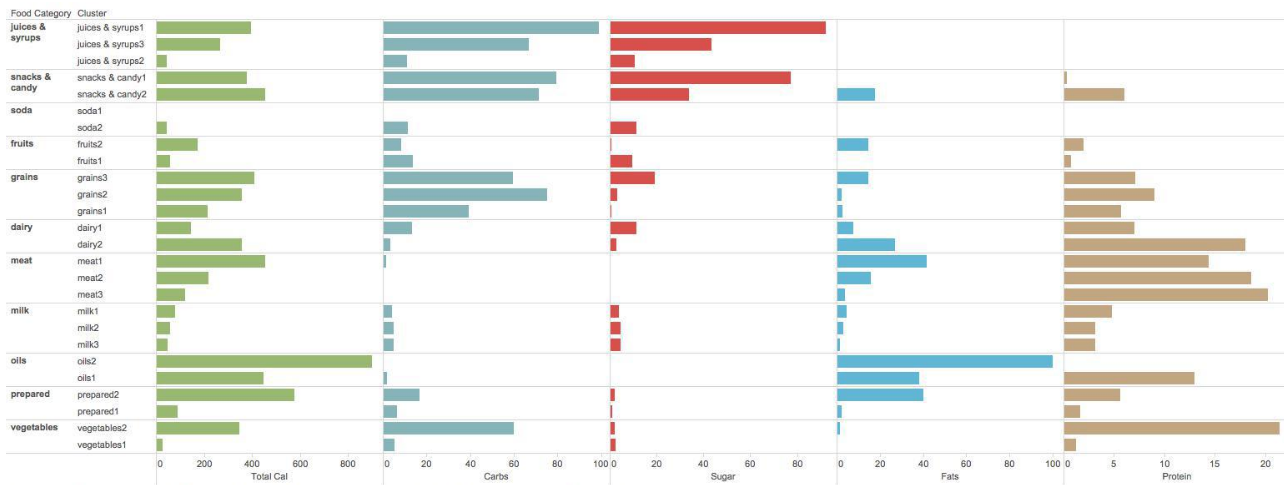
Note: http://dof.gob.mx/nota_detalle.php?codigo=5180641&fecha=04/03/2011

Figure 1.8: Variation in Nutritional Composition within Product Category – Snacks



Note: Nutritional composition for snacks, named ‘’Galletas popular” in household expenditure survey or price microdata as the lowest product category above individual products.

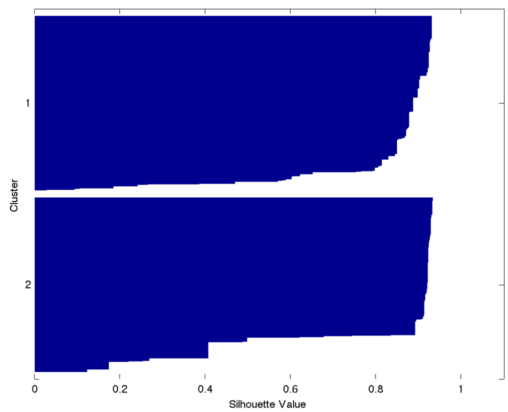
Figure 1.9: Nutritional composition by clusters



Sum of Total Cal, sum of Carbs, sum of Sugar, sum of Fats and sum of Protein for each Cluster broken down by Food Category. The view is filtered on Food Category, which excludes spices and warm beverages.

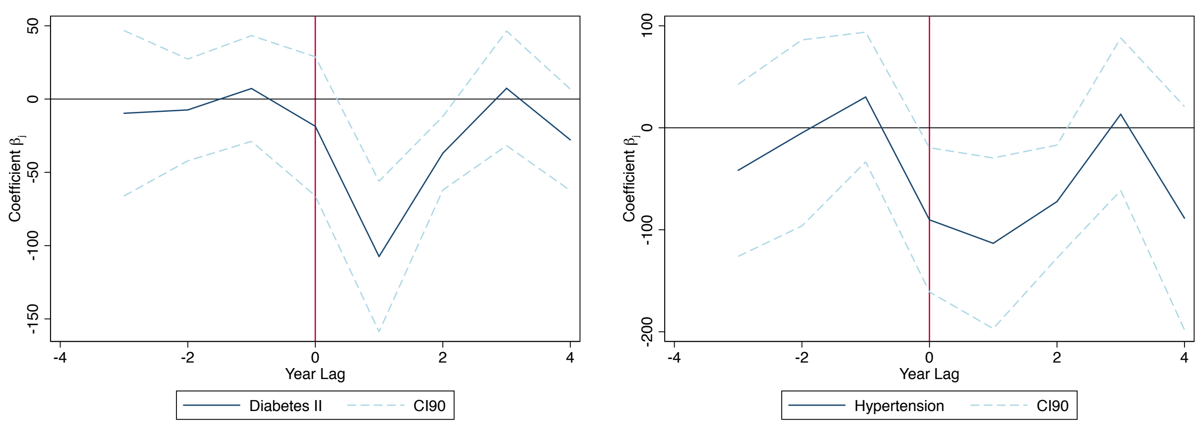
Note: Authors' calculations. Total calories, carbohydrates, protein, fats and sugar are reported in per 100 grams. Warm beverages and spices are excluded since they only consist of one cluster, represent very small share in Mexican's diet and are not relevant for my analysis.

Figure 1.10: Silhouette value for sodas



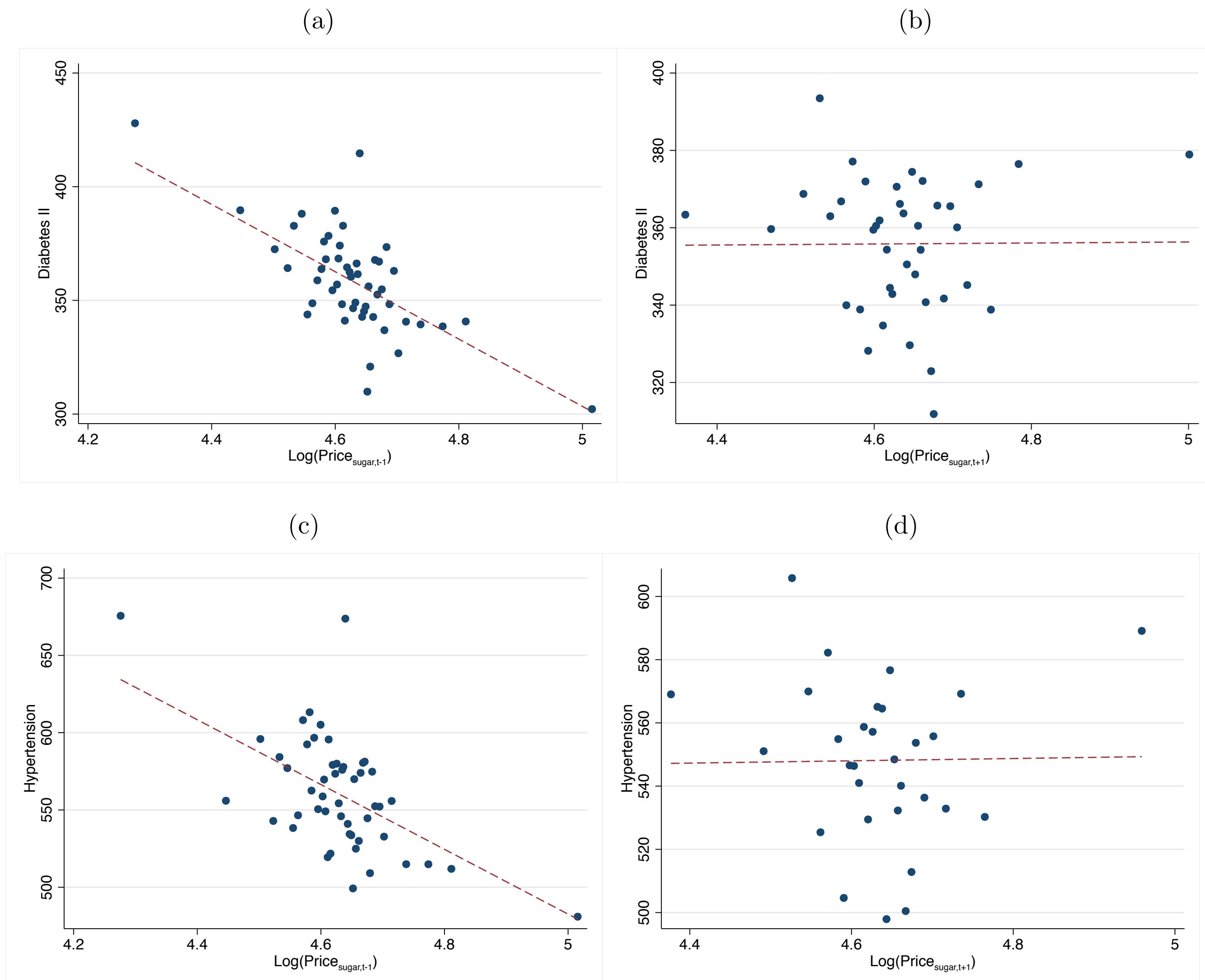
Note: Authors' calculations.

Figure 1.11: Long Run Effect and Price Leads



Note: Authors' calculations. Coefficients β_j are in absolute values, where j defines year lag.

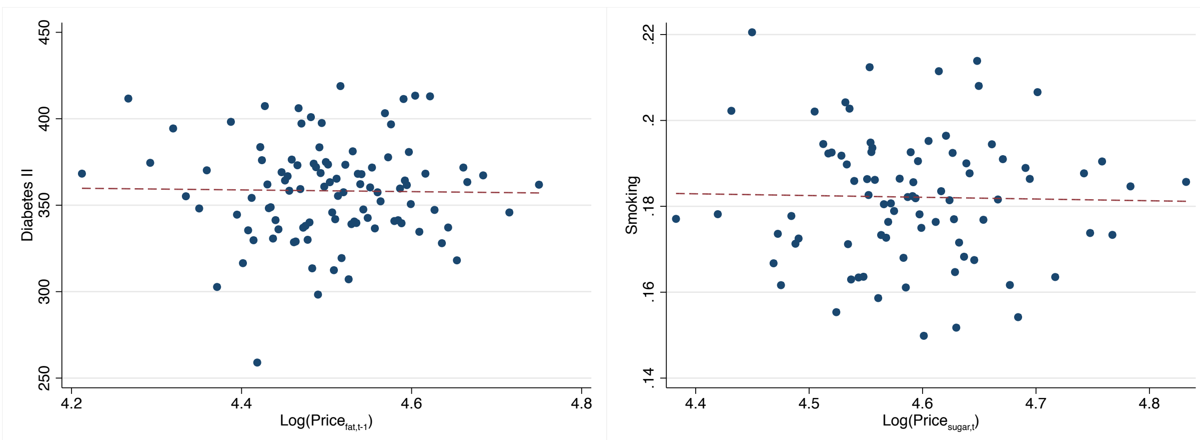
Figure 1.12: Diabetes and Hypertension - Price Leads Robustness Check



Note: (A) and (B) Diabetes II, (C) and (D) Hypertension.

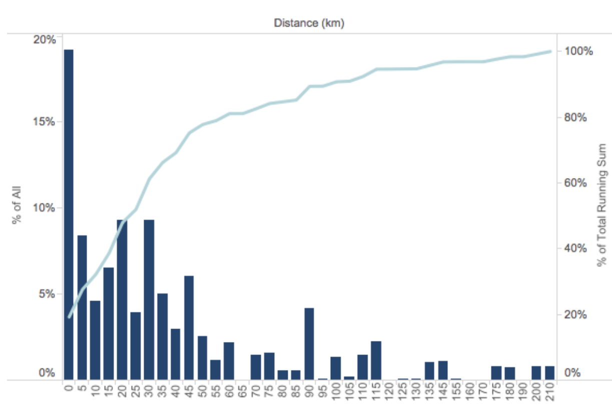
Figure 1.13: Robustness checks - Other prices, Smoking

(a) (b)



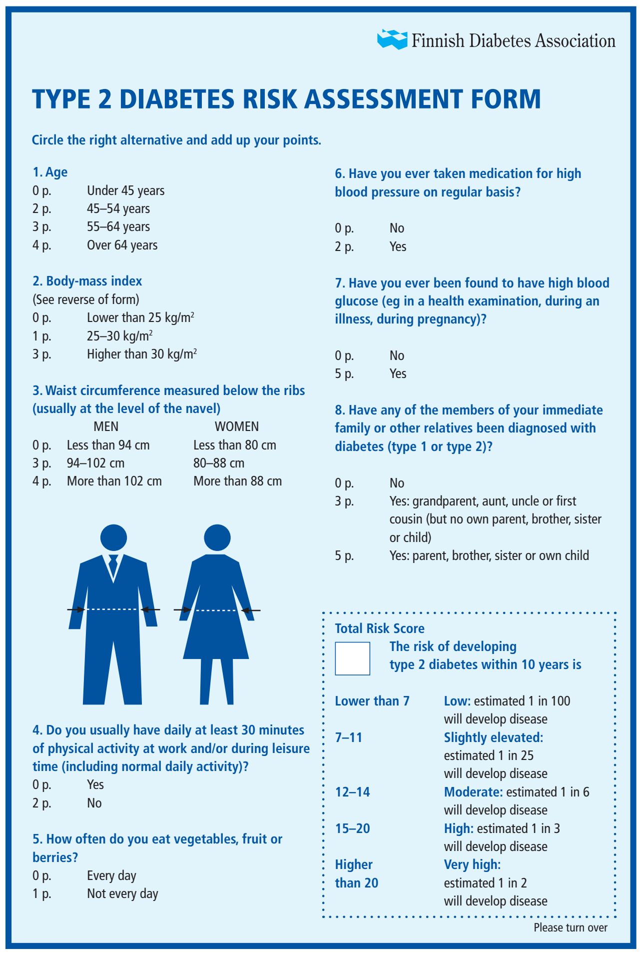
Note: (A) $Log(Price_{fat,t-1})$. (B) Smoking as dependent variable regressed on prices of foods rich in sugar.

Figure 1.14: Distance to nearest city



Note: Authors' calculations.

Figure 1.15: The Finnish Type 2 Diabetes Risk Assessment Form



Finnish Diabetes Association

TYPE 2 DIABETES RISK ASSESSMENT FORM

Circle the right alternative and add up your points.

1. Age
0 p. Under 45 years
2 p. 45–54 years
3 p. 55–64 years
4 p. Over 64 years

2. Body-mass index
(See reverse of form)
0 p. Lower than 25 kg/m^2
1 p. 25–30 kg/m^2
3 p. Higher than 30 kg/m^2

3. Waist circumference measured below the ribs (usually at the level of the navel)

	MEN	WOMEN
0 p.	Less than 94 cm	Less than 80 cm
3 p.	94–102 cm	80–88 cm
4 p.	More than 102 cm	More than 88 cm

4. Do you usually have daily at least 30 minutes of physical activity at work and/or during leisure time (including normal daily activity)?
0 p. Yes
2 p. No

5. How often do you eat vegetables, fruit or berries?
0 p. Every day
1 p. Not every day

6. Have you ever taken medication for high blood pressure on regular basis?
0 p. No
2 p. Yes

7. Have you ever been found to have high blood glucose (eg in a health examination, during an illness, during pregnancy)?
0 p. No
5 p. Yes

8. Have any of the members of your immediate family or other relatives been diagnosed with diabetes (type 1 or type 2)?
0 p. No
3 p. Yes: grandparent, aunt, uncle or first cousin (but no own parent, brother, sister or child)
5 p. Yes: parent, brother, sister or own child

Total Risk Score
The risk of developing type 2 diabetes within 10 years is

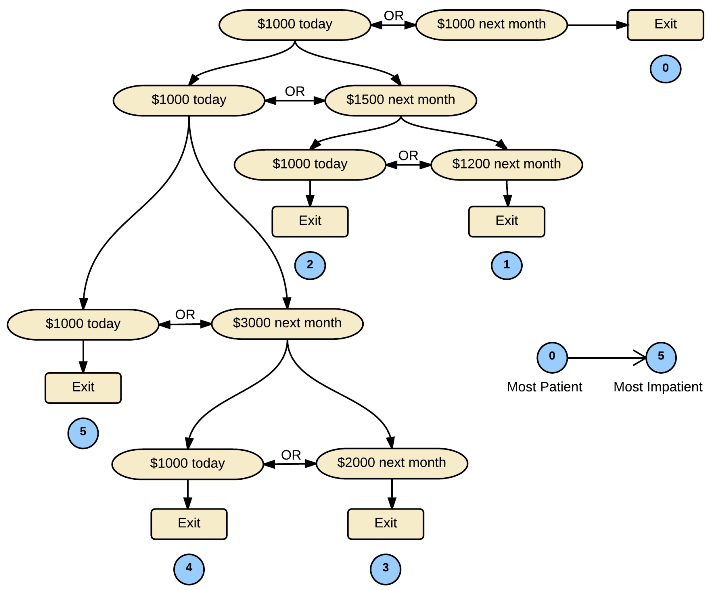
Lower than 7	**Low:** estimated 1 in 100 will develop disease
7–11	**Slightly elevated:** estimated 1 in 25 will develop disease
12–14	**Moderate:** estimated 1 in 6 will develop disease
15–20	**High:** estimated 1 in 3 will develop disease
Higher than 20	**Very high:** estimated 1 in 2 will develop disease

Please turn over

Test designed by Professor Jaakko Tuomilehto, Department of Public Health, University of Helsinki, and Jaana Lindström, MFS, National Public Health Institute.

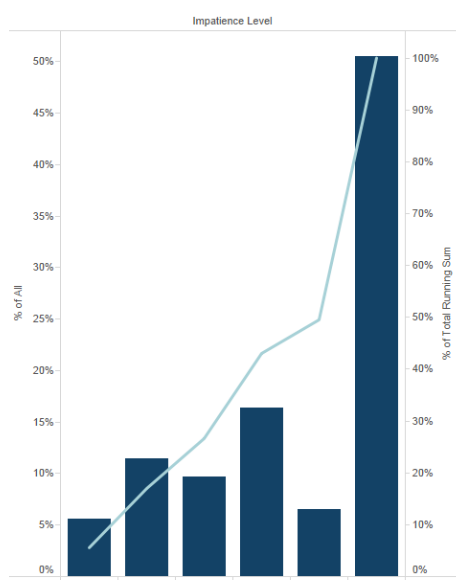
Note: Test designed by Professor Jaakko Tuomilehto, Department of Public Health, University of Helsinki, and Jaana Lindstrm, MFS, National Public Health Institute.

Figure 1.16: Flowchart illustrating elicitation of time preferences



Note: Based on MxFLS questions in Section "Preferences". Question was of the form: Imagine you have won the lottery. You can choose to get paid: \$X today or \$Y in one month. Which one do you choose?

Figure 1.17: Distribution of Individuals by Impatience



Note: Based on MxFLS questions in Section "Preferences". Question was of the form: Imagine you have won the lottery. You can choose to get paid: \$X today or \$Y in one month. Which one do you choose?

Chapter 2

Who Pays for Sins: Junk Food and Soda Tax Pass-Through in Mexico

2.1 Introduction

Over the past several decades, rates of obesity and diet-related chronic diseases in the developed and developing world has increased rapidly. The rise of this epidemic overlapped with a significant increase in the price differential between low and high-calorie dense foods. As a response, a number of academics and governments are proposing solutions, such as soda, junk food, sugar, fat tax or some combination thereof to stymie the increases in obesity levels. Hence, sin taxes are becoming increasingly popular not only due to their revenue-generating role but mainly due to their potential benefits for public health.

Even though the empirical evidence on health effects of soda consumption is well-established (Malik et al., 2006; Bray et al., 2004), evidence on the health impacts of sin taxes, be it with respect to soda, junk food or fat tax, is weak. The evidence on the impact of such taxes on obesity so far has mostly assumed a full pass-through to consumer prices, using estimated price elasticities of either soda, high-energy dense foods, or health outcomes to predict their effects. While studies using price variation show large impacts of hypothetical taxes on soda consumption and health outcomes, assuming a full pass-through (Harding and Lovenheim, 2014), literature using variation in actual taxes in contrast shows far weaker effects, both in terms of calorie consumption or health outcomes (Fletcher et al., 2010b). The latter might suggest that the actual taxes are either too low, or the tax is not fully passed through to consumer prices; depending on foods' relative elasticity of supply and demand.

The lack of implemented price-regulatory policies limits empirical assessments on the actual pass-through of sin taxes, such as soda tax, on consumer prices. The theory of pass-through suggests that the incidence of a given tax is a function of the relative demand and supply elasticities, which determines who bears the burden of the tax. While in perfectly competitive markets all of the deadweight loss falls on the consumers and the tax is fully passed forward due to a perfectly elastic supply, the prediction in markets with imperfect competition is theoretically ambiguous. For instance, in a market with monopoly or oligopoly we would expect less than a perfect pass-through, also depending on elasticity of demand. In the case of perfectly elastic demand, there is no pass-through and the tax is fully borne by the suppliers. Strategic pricing of producers or retailers in the case of oligopolistic markets with differentiated products could imply that in the case of an *ad valorem* tax, consumer prices could actually decrease due to lowered product quality, lower marginal production costs, and consequently sharpened price competition by reducing monopoly power of firms (Cremer and Thisse, 1994). On the contrary, firms enjoying market power might behave strategically in the opposite direction and end up overshifting the tax onto consumer prices instead (Fullerton and Metcalf, 2002; Lopez and Fantuzzi, 2012). Hence, how ad valorem tax passes through on consumer prices is theoretically ambiguous, and should be tested empirically.

This paper adds to the literature by providing one of the first empirical evidence on the pass-through of soda and junk food taxes. We examine the plausibility of the effectiveness of these taxes to reduce obesity levels through prices in the context of Mexico. Mexico provides a great setting for our empirical analysis. Effective January 1st, 2014, an eight percent ad valorem tax on foods with 275 or more calories per 100 grams (also called the "junk-food" tax) and a one-peso-per-liter tax on sugar sweetened drinks (diet sodas are excluded) was applied in Mexico on either domestic or imported products. The junk-food tax applies to non-basic food products, especially snacks, confectionery, chocolate and its derivatives, cakes, custards and flans, fruit jams and paste, ice creams, cereals and sweets.[1]

The analysis is made possible by rich longitudinal and nationally representative micro data on monthly retail food prices spanning between January 2011 and August 2014 with narrowly defined products' description combined with their detailed nutritional composition. This allows us to precisely identify products eligible for tax depending on their caloric content.[2] We exploit temporal variation in tax implementation and use a difference-in-difference approach, using product and month fixed effects, to test the hypothesis of no pass-through to consumer prices. First, we estimate the average pass-through on prices by various food categories. We find that average pass-through is the strongest for sodas, followed by snacks, candy and cakes. For these products, the pass through reaches about 100 percent around six months after the tax was put in place, suggesting a fairly inelastic demand. Sweets and cookies on average did not experience a full pass-through, however their prices on average increased around six percent, suggesting a more elastic demand.[3] We observe no significant change in the prices of cereal, yet do see positive increases in the point estimates four months after the tax was applied. We also observe large variation in pass-through across cities - from over- to under shifting, even documenting decreases in consumer prices. The latter is particularly obvious in the case of cereals, implying strategic price responses by retailers, which should be investigated further.

[1] Dairy products are exempt.

[2] There are more than 30,000 food price quotes available per month.

[3] Under the assumption that the supply is perfectly elastic.

In this paper, we make a number of important contributions to the literature. It is the first to provide empirical evidence on the effect of junk food and soda tax on consumer prices in the context of a middle-income country. Second, these estimates, combined with price elasticities of health outcomes, provide estimates on what we can expect from tax policies in terms of public health benefits, without relying on the assumption of a full pass-through. We estimate that an eight percent ad valorem tax on junk foods and a one-peso-per liter tax on sweetened beverages can result in between a quarter to half a percentage point decrease in type 2 diabetes and around a one and a half percentage point decrease in abdominal obesity prevalence, however, further research is needed on the heterogeneity of the pass-through. Third, by contributing to the debate on the ability of price changes to influence behavior and health (Gruber and Mullainathan, 2005; Evans and Ringel, 1999; Adda and Cornaglia, 2006; Wasserman et al., 1991), this paper complements recent evidence on proposed chronic disease management solutions, such as obesity remediation through taxes (Powell and Chaloupka, 2009; Fletcher et al., 2010b), or diabetes and obesity management by disseminating information, either through medical diagnosis (Oster, 2014), nutritional labeling (Abaluck, 2011; Bollinger et al., 2010; Downs et al., 2009) or advertising (Ippolito and Mathios, 1995). Lastly, even though this project is evaluated in the context of Mexico, this paper has implications that apply to developed countries such as the US. There, not only are lower income households at higher risk for diet-related chronic diseases, but debates on the effectiveness of soda or other sin taxes are current and still ongoing.

This paper proceeds as follows. Section 2.2 provides the theoretical background on tax incidence that motivates the empirical analysis. Section 2.3 provides the context in which the proposed research questions are answered, and presents the data of my research. Section 2.4 describes the main empirical strategy, followed by results and discussion in Section 2.5 and 2.6, respectively. I conclude in Section 2.7.

2.2 Tax Incidence

The theory of pass-through suggests that the incidence of a given tax is a function of the relative demand and supply elasticities. Depending on these elasticities, one can get a better idea on the composition of the deadweight loss incurred by the tax. That is, how much of it is coming from a change in consumer or producer surplus (see Figure 2.1, a). In perfectly competitive markets, all of the deadweight loss falls on the consumers and the tax is fully passed forward due to a perfectly elastic supply with a constant marginal cost of production (see Figure 2.1, b). Predictions in a market with imperfect competition are less straightforward, even though more likely in practice, and depend on factors such as the elasticities of supply and demand, market concentration for different goods, the free entry condition, and the type of competition assumed (Fullerton and Metcalf, 2002). For instance, we would expect less than a perfect pass-through in the beverage industry in Mexico due to its oligopolistic nature with differentiated products enjoying strong brand loyalties (Lopez and Fantuzzi, 2012; Cortes, 2009), however a lot depends on demand elasticity as well.[4]

Depending on elasticity of demand tax burden can be split differentially between producers and consumers. For instance, In the case of a perfectly elastic demand, all of tax burden falls on the producers (see Figure 2.1, c). In that case, we observe no pass-through of the tax on consumer prices, since the latter remain unchanged. The burden is fully passed backward onto suppliers (e.g. onto labor, capital, or other factors in production). In a monopoly, pass-through is greater, less than or equal to hundred percent with a demand curve having a constant elasticity of substitution, a linear or log-linear demand curve, respectively (Bulow and Pfleiderer, 1983).

When prices increase by more than the excise tax, we observe *overshifting*. While impossible in perfectly competitive markets, this can occur in imperfectly competitive ones. If faced with imperfect competition, firms might use the existence of their power and behave strategically among themselves. For instance, if firms recognize that forward shifting of the tax will decrease demand for the product, they will increase its price by more than the increase in tax rates to compensate the potential revenues lost. Overshifting will increase as product demand becomes less elastic (Fullerton and Metcalf, 2002). In an empirical exercise, Besley and Rosen (1999) investigate the impact of changes in state and local sales taxes on product prices. They observe overshifting for a number of commodities, including bread, shampoo, and soda.

The oligopoly models above, however, suffer from the restrictive assumption that goods are identical and that no distinction can be made between different brands. While this may be a reasonable assumption in some markets, such as agricultural commodity markets, in most other markets producers differentiate their products. This might create some monopoly power, and therefore, the ability to pass taxes forward depends greatly on the number of competitors in the market. In the case of oligopoly with heterogenous goods, in which firms compete over price, and product quality is endogenous, Cremer and Thisse (1994) show that a uniform *ad valorem* tax reduces the consumer price in equilibrium, partly due to a decrease in quality and hence lower marginal production costs, which in turn sharpens price competition by reducing monopoly power of firms. Hence, how *ad valorem* tax on junk food and sodas is passed through is theoretically ambiguous, and should be tested empirically.

2.3 Background and Data

In this section we first describe the background of the Mexican processed food industry, consumption, and health trends in Mexico. We then describe the introduction of sin taxes, and follow with the description of price data combined with product-specific nutritional information.

[4]In Mexico, Coca-cola is enjoying almost 50 percent in value share (Cortes, 2009).

Mexico: Consumption and health trends

Since the early 1990s, Mexico's dietary intake shifted from a traditional diet, rich in protein and fiber, to a so called "western diet", rich in fat and refined carbohydrates, especially rich in sugar. The purchase of fruits and vegetables decreased by almost 30 percent between 1988 and 1999. In particular, consumption of refined carbohydrates and soda increased by more than six and slightly less than 40 percent, respectively. Households' consumption of dairy, particularly ice cream and frozen desserts, more than trippled (Rivera et al., 2004), whereas consumption of fat increased from 23 to more than 30 percent (Clark et al., 2012). Compared to 69 liters per capita in 1991, at 172 liters per capita per year, Mexico is the largest consumer of soda today (ENSANUT, 2012).

The processed food consumption in Mexico amounted to $124.9 billion in 2012 alone. It experienced an annual growth rate of almost two percent, and is expected that between 2012 and 2020 the consumption will grow at an average annual rate of almost eight percent. Demand is growing particularly fast for sweets, snacks, ready to eat foods and especially ice cream with the fastest growth in 2013. Processed frozen foods is highly demanded particularly in the northern Mexico (Salcido, 2013).

Similarly, the processed food industry in Mexico is steadily growing as well. For instance, in 2012, processed food production exceeded a value of US$ 120,000 million, an increase of about two and a half percent compared to 2011, and present 23 percent of the manufacturing and four percent of the total gross domestic product. In terms of price sensitivity, the processed food sector in Mexico keeps prices low in order to remain competitive, especially in the product category of snacks and impulse/indulgence foods. In soft drinks market, the competitive environment in Mexico is led mainly by three companies which together hold almost 70 percent value share, with however 45 percent belonging to Coca-Cola de Mexico (Salcido, 2013).

Consumption of items rich in refined carbohydrates and/or fat are both rich in calories and are constantly linked to weight problems and diet-related chronic diseases. Indeed, parallel to nutritional transition described above, Mexico's obesity prevalence rate has become one of the highest in the world. For example, prevalence of excess weight and obesity in adults in Mexico, has gone from less than 30 to more than 70 percent between 1988 and 2012. Today, more than 75 percent are considered to be abdominally obese. Similarly, the fraction of overweight children has risen from 9 to more than 23 percent in the same period (Alberti et al., 2006). In addition to obesity, Mexico is also facing steep positive trends in prevalence and incidence of hypertension and type 2 diabetes. For instance, between 2000 and 2007 alone, the mortality rate due to type 2 diabetes increased from 77.9 to 89.2 per 100,000 people. Today, type 2 diabetes is considered the number one cause of deaths in the country, hypertension and cardiovascular diseases coming up as second (Sánchez-Castillo et al., 2005; Sánchez-Barriga, 2010). Already high growth and level of health costs due to chronic disease over the last decade, this burden is expected to increase even more in the coming years, especially due to the aging Mexican population and complications related to that (McKinsey, 2012).

Mexico: Junk Food and Soda Tax

Given the alarming health situation and its worrisome prospects, Mexican government took an active role to fight obesity and diet-related disease epidemic. Effective January 1st, 2014, an eight percent *ad valorem* tax on foods with 275 or more calories per 100 grams and a one-peso-per-liter tax on sugar-sweetened drinks was applied in Mexico on either domestic or imported products.[5] Differential tax on junk-food applies to non-basic food products, especially focusing on products within product categories such as snacks, confectionery, chocolate and chocolate derivatives, cakes, custards and flans, fruit jans and paste, peanut, hazelnut or other -nut butter, ice creams, condensed milk, and cereal based products (e.g. cookies, sweet bread, breakfast cereals). The "junk food tax" is not applicable to foods prepared or provided in restaurants or similar venues. Flavored drinks with a health record (e.g. drugs), or oral serums. Dairy products including drinkable yogurts, juices and nectars containing milk or dairy are exempt as well. In addition, some basic food products, regardless of their caloric content, are exempt as well. Those are wheat products, such as wheat flour tortillas, wheat-based pasta for soups, unsweetened bread, such as baguettes, ciabatta bread or similar including loaves of bread, wheat flour products with no added sugars, and cereal-based food products for toddlers and infants. Corn products, such as corn tortilla, tostadas, corn flour, nixtamal and corn dough, or corn cereal food product without added sugars, are excluded as well. Even though not a final product for consumption, junk food tax also applies to numerous ingredients used in bakery processing, such as dried fruits, cocoa and gelatin due to their high-caloric content.

Soda tax is applied to regular soda, sugary flavored drinks, and on every liter of product obtained through syrups, concentrates, powders, flavors, and essences with added sugar. Drinks sold in open containers using automated, mechanical or other electrical machines are included as well. For concentrated extracts tax has to be calculated based on the number of liters obtained from it in accordance with the manufacturer specifications (e.g. x grams of powder is used to make x liters of fluid). Diet sodas are excluded.

Price Data

The central dataset used for this empirical analysis is a novel dataset on monthly time series of retail food prices between January 2011 and August 2014. This retail price data consists of 30,000 food price quotes per month from a nationally representative sample of urban areas across 46 Mexican cities. Data is collected by the National Institute of Statistics and Geography (INEGI) for the purpose of computing the Mexican CPI, and is therefore representative of more than two-thirds of Mexican consumers' expenditures.

There are many reasons why this data is suitable for the purpose of my analysis. First, food prices are tracked for the same or a very similar product using a unique product identifier continously within stores over 44 months, making them comparable over time, and hence making it possible to potentially assign their temporal variation within cities to national

[5] I refer to products with more than 275 calories per 100 grams as junk-food.

changes in taxes. Second, required by Articulo 20-Bis of the Codigo Fiscal de la Federacion, INEGI publishes store price microdata together with precise item descriptions. Crucially for this project, products' price quotes are very narrowly defined. Definitions include product's name and brand, packaging type and weight, such as Kellog's Cereals, Zucaritas, box of 250 grams, sold in outlet 1100 in Mexico City. To identify products eligible for junk food and soda tax (i.e. holding more than 275 calories per 100 grams, or regular versus diet sodas), I combine retail price data consisting of barcode-equivalent food product's description with detailed nutritional information of those products, included on their nutrition label.

In particular, I obtain information on amount of energy in directly reported kilocalories (kcal), and also from reported grams of fats, protein, sodium, carbohydrates, of those grams in sugar and fiber per 100 grams, which I then convert into calories to obtain total calorie amount per 100 grams.[6] This is especially useful in the case of a soda tax, where the tax is only applied to regular (and not diet) sodas, or in the case of product categories where caloric content varies substantially across products within them. See detailed description on nutritional data collection in Gracner (2015).

Lastly, as also described in Gracner (2015), using each item's unique identifier, consisting of a product number, store, city and food category, I can assign a constant nutritional content to each product's price trajectory over time. Since INEGI reports changes in product's representation, brand, or type, I can assign an appropriate, updated nutritional composition to substitutions of existing or addition of new items. Fourth, prices of food items are mostly conveniently expressed either per 1 kg or 1 liter, which simplifies the interpretation and scaling of the nutritional composition.

Table 2.1 describes mean prices of some of the food categories which products were eligible to be taxed. Mean prices before tax are calculated over the period of August - December in 2013 and are compared to the January-August period in 2014. Simple comparison between means shows that mean prices of regular sodas, snacks, sweets or cereals increased, whereas mean prices of diet sodas, and especially mean prices of tortillas and bottled water, which were exempt from tax, remained almost the same. Figures 2.2 and 2.3 show a visible spike in their trend line for tax eligible goods. On the contrary, we observe no such jump for goods that are not eligible for tax. Figure 2.4 shows a de-trended time series, or residuals, for prices of foods, eligible for tax. We again observe a substantial jump right after the law has been passed (January 2014).

2.4 Empirical Strategy

In order to estimate the average tax pass-through on retail prices over the observed period, we exploit its temporal variation and use a difference-in-difference approach to estimate the

[6]Macronutrients are converted from grams to total calories per 100 grams by multiplying grams of carbohydrates by 4, grams of proteins by 4, and grams of fats by 9 (USDA).

following equation:

$$\log \mathrm{P}_{ism} = \alpha_m + \alpha_i + \sum_{r=1}^{n} \beta_r T_{imr} + \mathbf{z}'_{ct}\theta + \varepsilon_{ism} \tag{2.1}$$

where $logP_{ism}$ is the dependent variable, log of retail price, observed for the product i in a store s in month m, expressed in either per 1 liter or per 1 kilogram. α_m and α_i are fixed effects corresponding respectively to the months under review (January 2011 to August 2014), to product characteristics (irrespective of including more or less than 275 calories per 100 grams) and to shop characteristics, controlling for retail type, location, and competition, assuming it does not change over the study period, across n product categories r (e.g. sodas, snacks, sweets, cereal).[7] The vector $\mathbf{z}_{ct}$ controls for time variant changes at the city or state level, such as state gross domestic product, and potential local demand shifter, such as state level advertising expenditure costs and number of fast food restaurant per squared kilometer. The indicator variable T_{imr} equals 1 when product i is eligible for taxing, i.e. it contains equal to or more than 275 calories per 100 grams or belongs to a sweetened beverages product category (including more than 0 grams of sugar per serving size) and the period m is between January 1st 2011 and August 2014, yet equals 0 otherwise.

Unless stated otherwise, all parameters in this equation are estimated using fixed effects ordinary least squares. To account for correlation of the residuals ε_{ism} within cities, I report standard errors clustered by city. The key identification assumption of the equation (2.1) is that if the junk food or soda tax had not been introduced, prices of the taxed (e.g. regular soda) and non-taxed products (e.g. diet soda) would have evolved in a similar way.

We estimate two additional models. First, to allow for lags in the reaction of retailers or producers in shifting the tax onto consumer prices, we estimate the monthly effects of taxes from January to August 2014. Second, to observe geographical heterogeneity in tax pass-through of different foods, we estimate the Equation 2.1 separately for each of the 46 cities.

2.5 Results

Table 2.2 shows empirical estimates of the effect of tax law applied on log of prices of foods overall (Column 1) and by different categories (Column 2-5). Average pass-through over the period January to August 2014 seems to be the strongest for sodas, followed by snacks, candy and cakes. In particular, those two categories on average experienced a full pass-through, implying a fairly inelastic demand (assuming a perfect competition case with perfectly elastic supply). Sweets and cookies on average did not experience a full pass-through, however their prices on average increased around six percent, implying a more elastic demand.[8]

Observing changes in prices across food categories over time (see Table 2.3), however, shows that the tax is almost fully passed forward to consumers in the case of cookies and

[7]The same product is followed within the same store over time.

[8]Adding food categories in the regression one by one does not change the results.

sweets six months after the junk food tax is applied as well. Moreover, the tax seems to overshift to consumers especially in the case of snacks, candies and cakes, where a eight percent tax results in over fifteen percent increase in prices of this category within six months.[9] Similarly, prices of sodas seem to increase by around thirteen to fourteen percent by June 2014, yet decrease and remain around 11 percent higher in August 2014 relative to the pre-tax levels. These results compare to Berardi et al. (2012), who show that soda tax was fully passed-through about five months after it was applied. On average, we observe no significant change in prices of diet sodas, exempt from taxes, making us more confident in our results.

We observe no significant change in prices of cereal on average, however we do see positively increasing point estimates around four months and after since the tax was applied. Longer time series will show whether the tax is passed forward over a longer time horizon. There are several reasons why there seem to be no pass-through. On one hand, it could be that retailers adjusted their prices, but only partially to avoid potentially strong consumer reactions. On another hand, there could be large heterogeneity across retailers and their pricing responses - some transmitted the prices, perhaps even more than fully, some kept them unchanged, yet others decided to lower them instead. Recall that the latter is indeed possible especially in the case of imperfect competition and highly differentiated products, as described in Section 2.2. In fact, there is evidence on ready-to-eat cereal industry being highly-concentrated, with high price-cost margins, especially due to significant product differentiation (Nevo, 2001). In addition, Berardi et al. (2012) show that there is indeed a combination of all possible scenarios happening in the case of soda tax applied in France. For instance, they show that biggest players in the retailing trade market kept the prices low due to their large bargaining power, which might have given them better deals with suppliers.

Figure 2.5 indeed shows large variation in pass-through estimates by cities - from large negative to large positive ones. Hence, the average pass-through of the tax to prices hide some significant differences, which we plan to explore in the future; e.g. by retailer types, or by brands, and not only for cereals but across different food groups. For instance, pass-through of the tax on soda prices is positive and significant across most of the cities, yet its magnitudes vary substantially.[10] Similar holds for snacks, candies and cakes, yet results in magnitude and significance vary even more for the case of cereal, sweets and cookies.

2.6 Discussion

While there is some evidence relating changes in food prices to obesity and diet-related chronic diseases (Sturm and Datar, 2005; Datar et al., 2004; Powell et al., 2007a), such as my recent work in Gracner (2015), or recent evidence on how changes in relative food or nutrient prices alter the composition of food consumption (Dubois et al., 2013; Harding and

[9] This might suggest the presence of imperfect competition, which should be explored in the future.

[10] In some cities, the number of products is significantly smaller which might result in non-significant results.

Lovenheim, 2014), there is little rigorous evidence on health impacts of actual alternative food price policies in place (for a review, see Thow et al. (2010)). When evaluating possible price regulations, often times the assumption is made on no strategic firm behavior, hence an assumption on so called “passive pricing”, where producers and retailers do not adjust product prices in response to the tax or subsidy policy, is made. Often times evaluations also assume a full pass-through on consumers, where in fact, the policy could backfire with producers lowering product prices instead, which depends on the type of a tax (excise vs ad valorem) in particular. Taxes are unlikely to be perfectly passed through to consumers especially in the markets of fast-food, ready to eat foods, which are those we are interested in targeting. This food industry and the retail chains in it in particular are characterized by large firms with market power, and can afford markup adjustments due to consumer substitution patterns, market structure and their market power (Goldberg and Verboven, 2001; Hellerstein, 2008).

Foods that are eligible for junk food tax largely overlap with products defined in Gracner (2015). Hence, we can use the magnitude of changes in prices of junk foods, caused by the policies in place, to calculate the expected effect of junk food and soda tax on health outcomes, given the price elasticities of obesity, diabetes and hypertension incidence or prevalence in Mexico (Gracner, 2015). On average, prices of junk foods changed by roughly eight percent. This implies around one and a half percentage point decrease in the prevalence of abdominal obesity, and between a quarter and one half of a percentage point decrease in the prevalence of type 2 diabetes.

When making these calculations, we heroically assume that the consumption response is symmetric to either decrease or increase in prices. Hence, what we plan to achieve with building on current work is to first, estimate demand response across food categories to actual price changes due to the tax. In doing so, we would allow for goods to be addictive in nature. We plan to use nationally representative cross-sectionally repeated Household Expenditure Surveys (ENIGH). Product division and unit of measure in price data make it convenient to combine the store microdata with Household Expenditure Survey data (ENIGH), which will be convenient for this. Food categories in retail price data are representative of the ones in ENIGH, accounting for at least 0.02 percent of households’ expenditures, which captures well above of the 95% of Mexican households’ expenditures (Gagnon, 2009). Second, we plan to estimate the consumption responses under different tax policies. Currently, Mexico applied an ad valorem eight percent per qualified item. We plan to evaluate different policies, such as applying different tax rates per item depending on its nutritional content (e.g. sugar tax), or on its caloric value (e.g. calorie tax). Throughout our analysis, we will look for heterogeneity in responses, especially across different socioeconomic groups.

2.7 Conclusion

Despite the heated debates on price-based policies and their role in reversing the obesity and the chronic disease trends, the lack of implemented price-regulatory policies prevent for rigorous empirical evidence on how taxes on sodas and high calorie foods actually transfer to and change their prices to exist. The plentiful substitutes of foods available, elasticity of their demand, product differentiation, number of competitors and their market power all play a role in the ability to pass taxes forward to consumers.

After the implementation of junk food and soda tax in Mexico on January 1st 2014, prices of sodas and of foods rich in calories increased significantly. Using a nationally representative micro data on more than 30,000 food prices per month across Mexico, and the difference-in-differences as our preferred empirical strategy, we show that soda tax was fully shifted to prices of sodas, indicating a fairly inelastic demand, and that diet soda perhaps is not such a close (untaxed) substitute to a regular soda after all. In addition, we show an almost full pass-through for snacks, cakes and candies as well. Both, the snacks and sodas experience tax overshifting within six months after the tax was implemented. On the contrary, we observe smaller than full pass-through (yet an increasing one over time) for products, such as sweets and cookies, and no significant pass-through on prices of cereal. There are several possible reasons for undershifting. For instance, existence of untaxed substitutes and elastic demand might explain this result. Second, producers or retailers might response differently in their pricing strategy - some overshift, some do not change their prices, whereas some retailers actually decrease their prices. Especially because of the latter a careful analysis of the pass-through is necessary, to avoid the counter intuitive results, actually worsening public health.

Table 2.1: Summary statistics

Variable	Obs	Mean
Price(sodas) - before tax	2073	11.362
Price(sodas) - after tax	3312	12.971
Price(sweets) - before tax	1305	92.859
Price(sweets) - after tax	2080	103.332
Price(snacks/cakes) - before tax	905	102.093
Price(snacks/cakes) - after tax	1448	114.495
Price(cereal) - before tax	1355	68.855
Price(cereal) - after tax	2168	74
Price(diet sodas) - before tax	170	10.155
Price(diet sodas) - after tax	262	11.113
Price(tortilla) - before tax	4550	12.734
Price(tortilla) - after tax	7280	12.702
Price(water) - before tax	1820	6.381
Price(water) - after tax	2912	6.527

Notes: Mean prices before tax are calculated over months of August-December in 2013. Mean prices after tax was applied are calculated over months January-August in 2014. This explains the variation in N within product category before and after tax period.

Table 2.2: Average Pass-Through

Dep.Var: *Log*(*Price*)	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
$D_{all,m}$	0.011***				
	(0.003)				
$D_{sodas,m}$		0.114***	0.116***	0.117***	0.117***
		(0.007)	(0.007)	(0.007)	(0.007)
$D_{sweetbread,m}$			0.081***	0.082***	0.082***
			(0.005)	(0.006)	(0.006)
$D_{snaks/cakes,m}$			0.131***	0.132***	0.132***
			(0.029)	(0.029)	(0.029)
$D_{cookies,m}$				0.061***	0.061***
				(0.011)	(0.011)
$D_{cereals,m}$				0.012	0.012
				(0.016)	(0.016)
$D_{sweets,m}$				0.065***	0.065***
				(0.016)	(0.016)
$D_{dietsoda,m}$					0.016
					(0.037)
N	1302687	1302687	1302687	1302687	1302687
Adj. R^2	0.08	0.08	0.08	0.08	0.08
Y Mean					3.41
Y SD					1.06
N Clusters	46.00	46.00	46.00	46.00	46.00
Product FE	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes

$D_{i,m}$ represents an indicator whether an item is eligible for junk food (equal or more than 275 calories per 100 grams) or soda (regular, not diet sodas) tax and equals 1 after January 2014 and 0 otherwise. Diet sodas are not subject to taxation, hence hypothesis on no effect not being rejected is expected. Robust standard errors in parentheses, clustered at the city level. */**/*** denotes significant at the 10% / 5% / 1% levels.

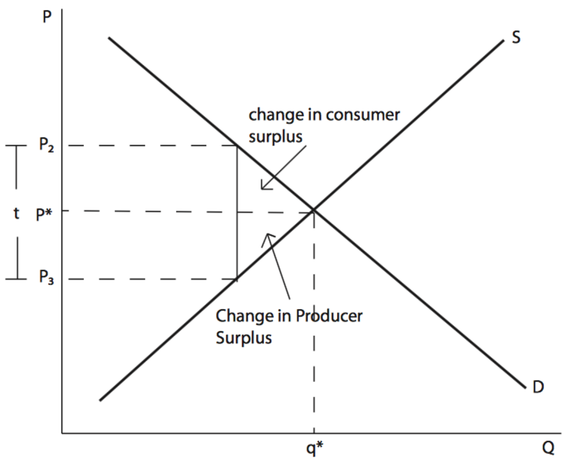
Table 2.3: Monhtly Average Pass-Through

Dep. Var: $Log(Price)$	**Sodas**	**Sweet bread**	**Snacks&Cakes**	**Cookies**	**Cereals**	**Sweets**	**Diet Sodas**
$D_{i,Jan\,2014}$	0.081***	0.037***	0.090***	0.031***	-0.010	0.043**	-0.021
	(0.007)	(0.005)	(0.027)	(0.010)	(0.013)	(0.017)	(0.037)
$D_{i,Feb\,2014}$	0.106***	0.059***	0.107***	0.046***	-0.023	0.051***	0.009
	(0.008)	(0.006)	(0.027)	(0.011)	(0.016)	(0.017)	(0.039)
$D_{i,March\,2014}$	0.109***	0.063***	0.117***	0.037***	-0.006	0.048**	0.006
	(0.008)	(0.006)	(0.030)	(0.011)	(0.016)	(0.018)	(0.038)
$D_{i,April\,2014}$	0.112***	0.074***	0.123***	0.041***	0.012	0.059***	0.005
	(0.007)	(0.006)	(0.031)	(0.013)	(0.017)	(0.016)	(0.038)
$D_{i,May\,2014}$	0.125***	0.092***	0.135***	0.065***	0.016	0.071***	0.025
	(0.008)	(0.006)	(0.031)	(0.012)	(0.018)	(0.017)	(0.037)
$D_{i,June\,2014}$	0.141***	0.106***	0.150***	0.081***	0.025	0.085***	0.044
	(0.008)	(0.006)	(0.031)	(0.013)	(0.021)	(0.016)	(0.041)
$D_{i,July\,2014}$	0.130***	0.104***	0.152***	0.087***	0.029	0.077***	0.025
	(0.008)	(0.007)	(0.031)	(0.013)	(0.022)	(0.019)	(0.041)
$D_{i,Aug\,2014}$	0.114***	0.093***	0.157***	0.074***	0.028	0.053***	0.003
	(0.009)	(0.007)	(0.031)	(0.011)	(0.020)	(0.019)	(0.040)
N	1302687	1302687	1302687	1302687	1302687	1302687	1302687
Adj. R^2	0.08	0.08	0.08	0.08	0.08	0.08	0.08
Y Mean	3.41	3.41	3.41	3.41	3.41	3.41	3.41
Y SD	1.06	1.06	1.06	1.06	1.06	1.06	1.06
N Clusters	46	46	46	46	46	46	46
Product FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

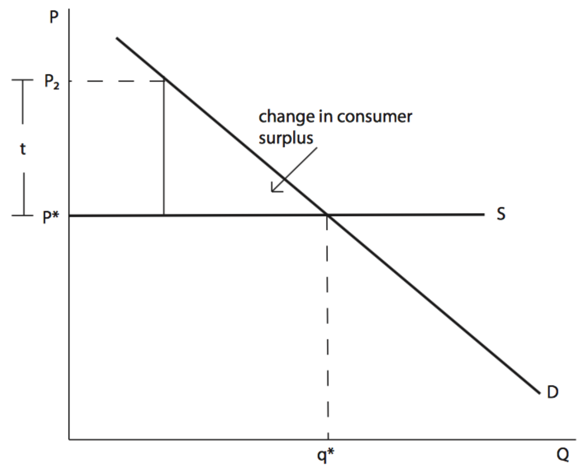
$D_{i,month\,2014}$ represents an indicator whether an item is eligible for junk food or soda tax and equals 1 after January 2014 and 0 otherwise. Robust standard errors in parentheses, clustered at the city level. */**/*** denotes significant at the 10% / 5% / 1% levels.

Figure 2.1: Tax Incidence

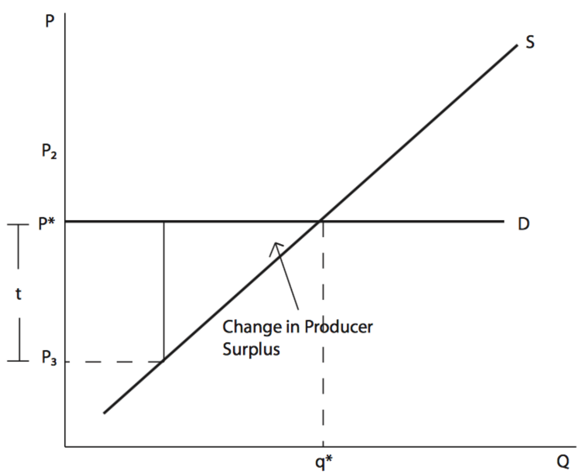
(a) In-between Case



(b) Elastic Supply



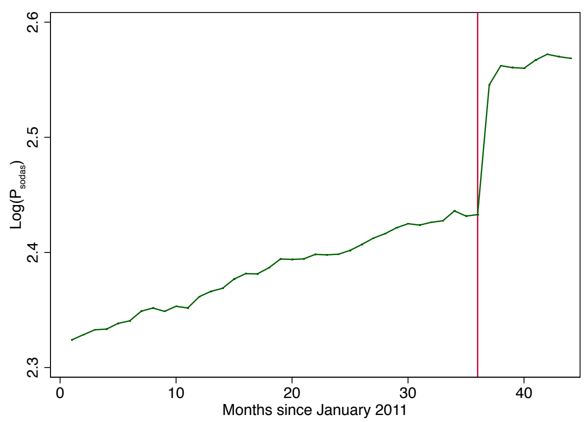
(c) Elastic Demand



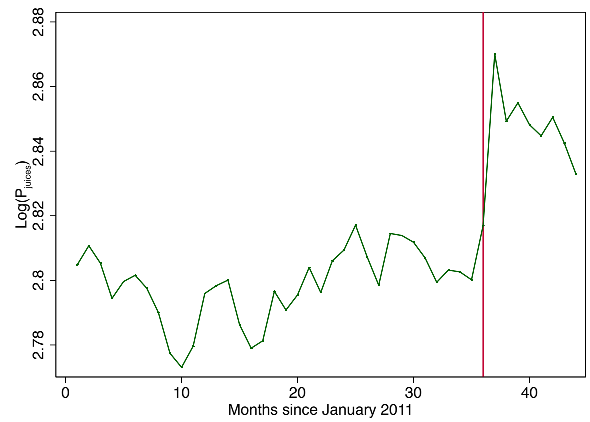
Note: Harding et al. (2010).

Figure 2.2: Taxed foods: Log(Prices) over time

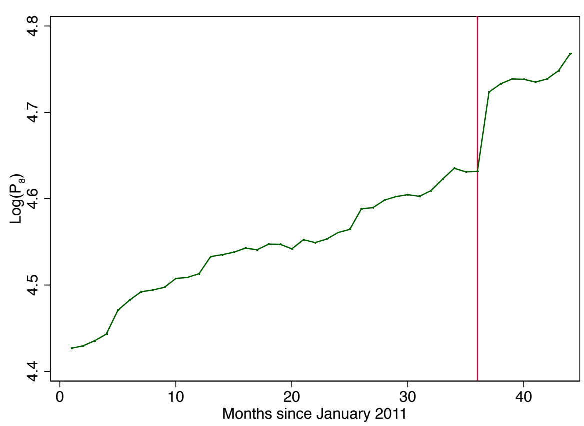
(a) Sodas



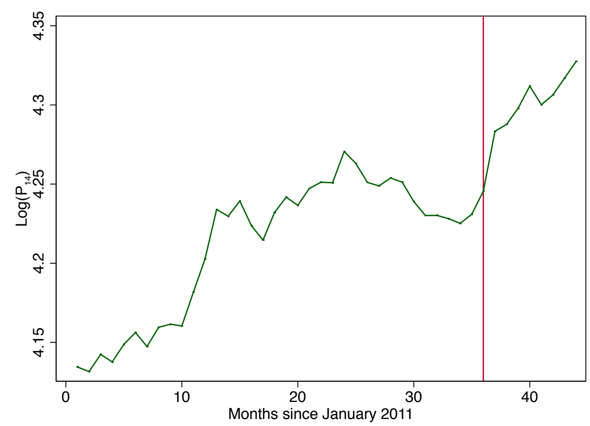
(b) Juices



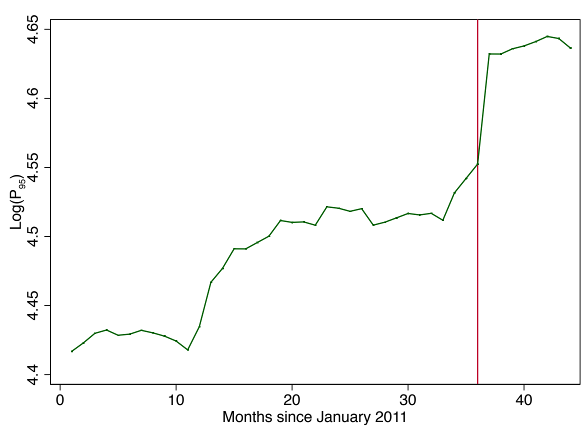
(c) Candy, small cakes, sweet bread



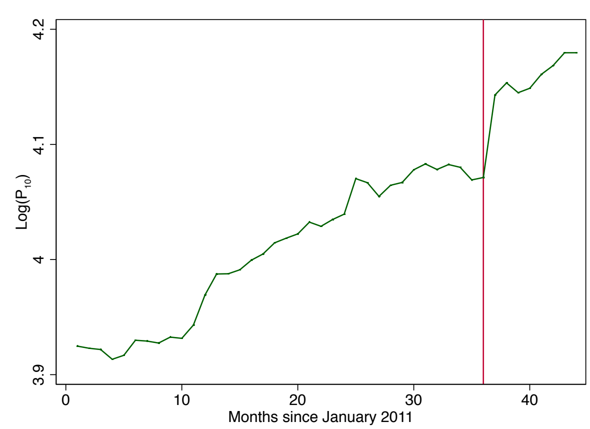
(d) Cereal



(e) Sweets



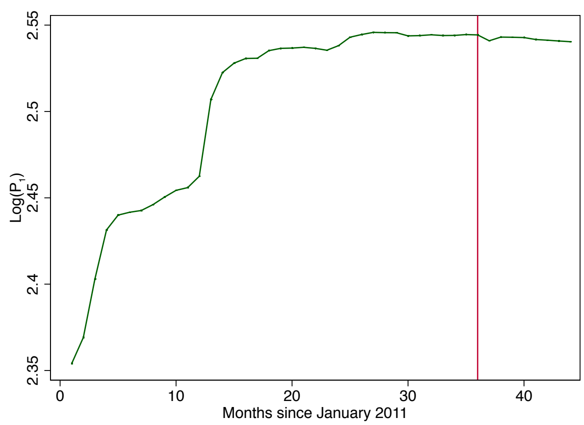
(f) Cookies



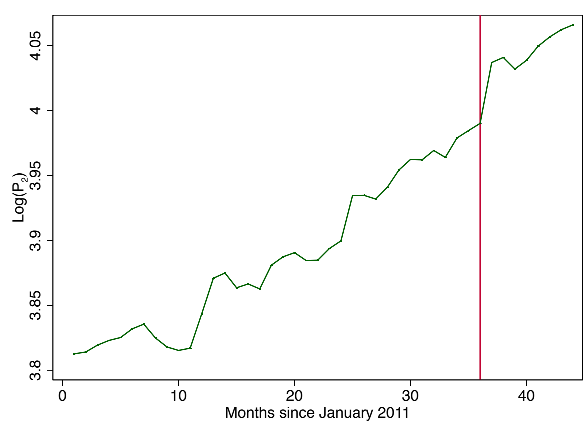
Note: Months run from January 2011 until including August 2014.

Figure 2.3: Foods Exempt from Tax: Log(Prices) over time

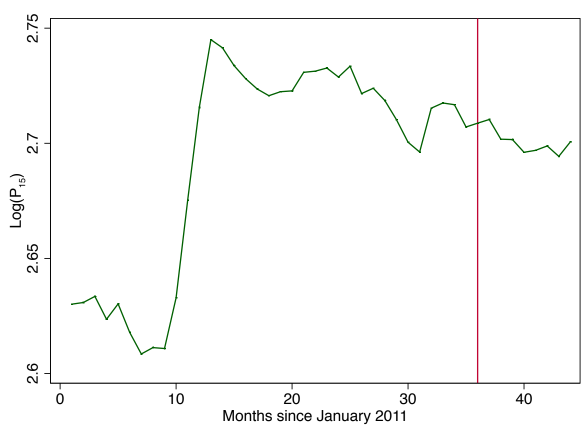
(a) Tortillas



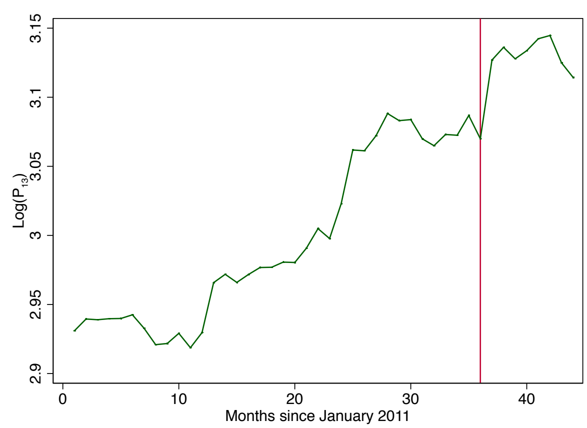
(b) Tostadas, Masa



(c) Rice



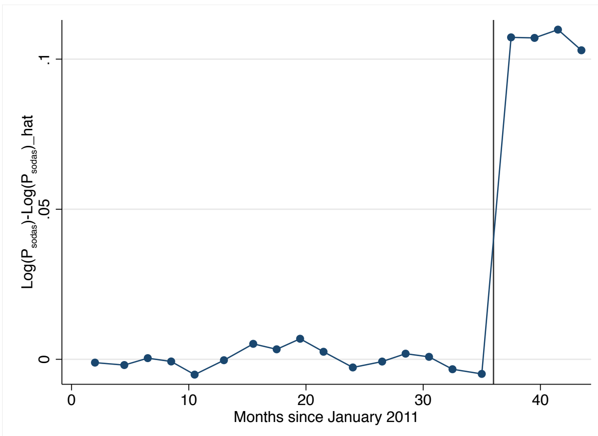
(d) Flour



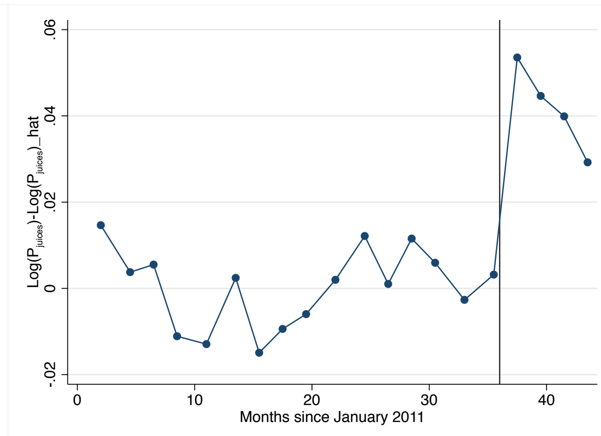
Note: Months run from January 2011 until including August 2014.

Figure 2.4: Residuals for taxed goods

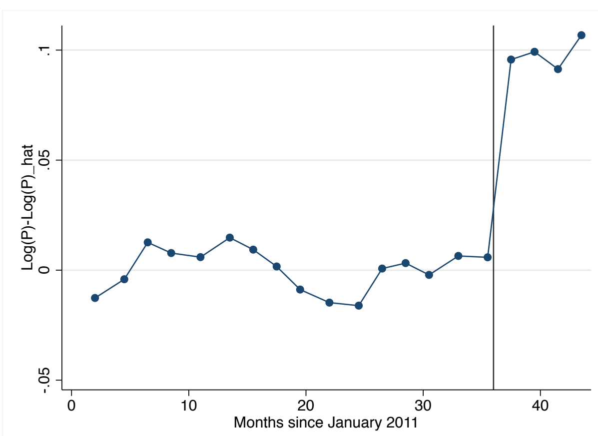
(a) Sodas



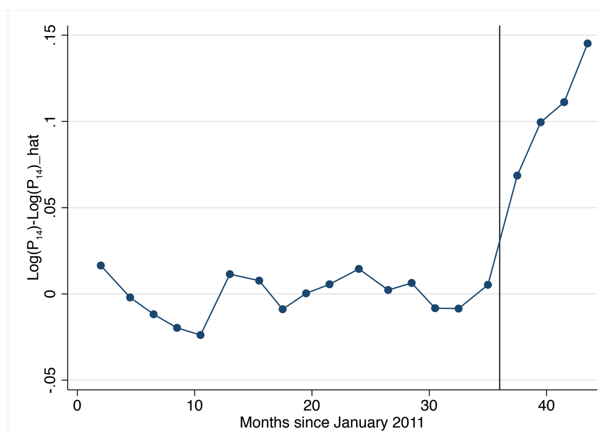
(b) Juices



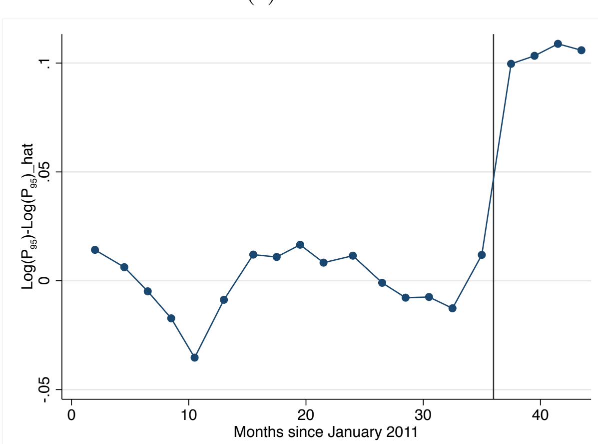
(c) Candy, cakes



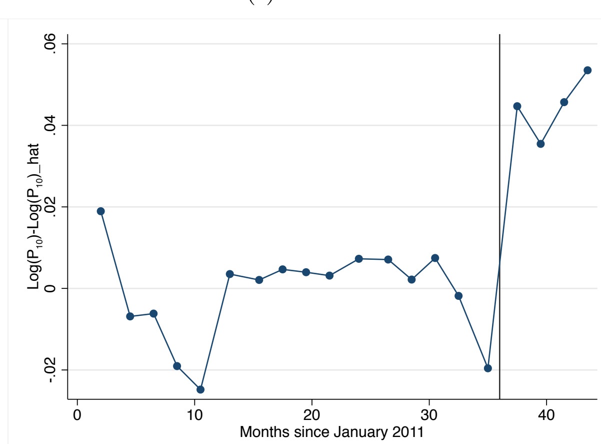
(d) Cereal



(e) Sweets



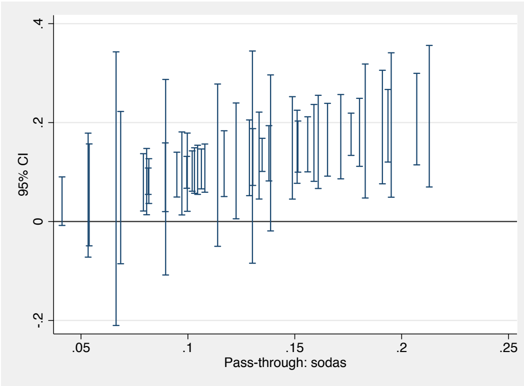
(f) Cookies



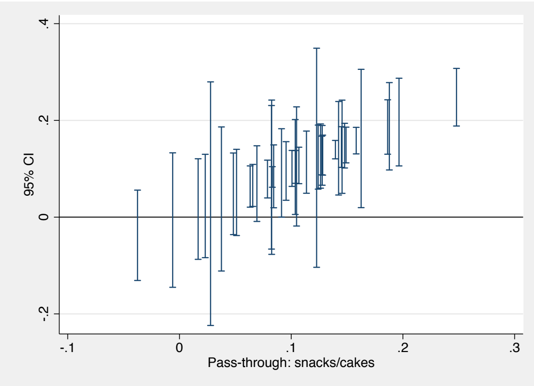
Note: Months run from January 2011 until including August 2014.

Figure 2.5: Pass-through by cities and foods

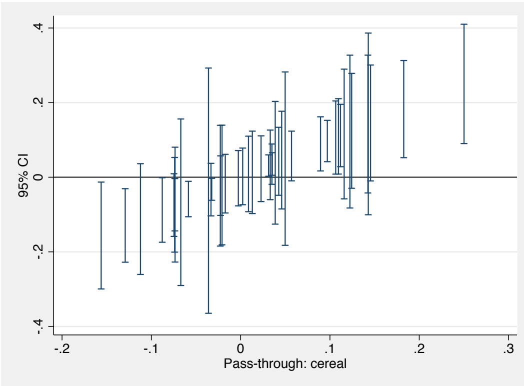
(a) Sodas



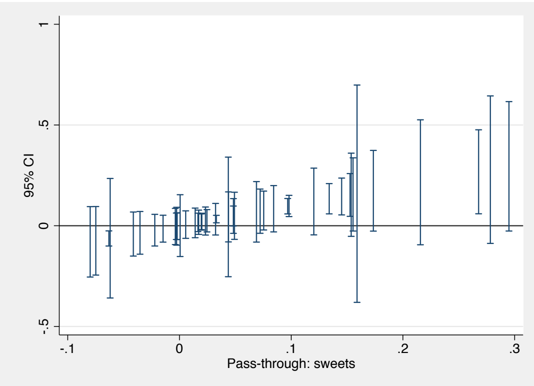
(b) Snacks, cakes



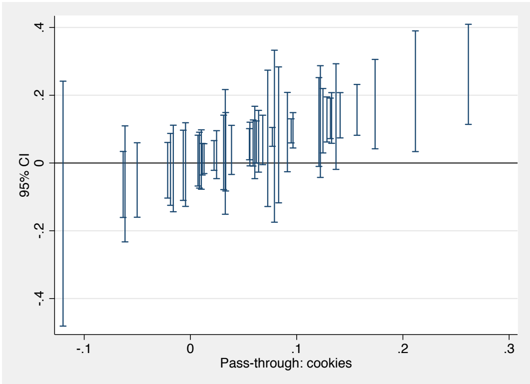
(c) Cereal



(d) Sweets



(e) Cookies



Note: Each line represents one of the 46 cities.

Chapter 3

Road Quality, Local Economic Activity and Welfare: Evidence from Indonesia's Highways

3.1 Introduction

Maintaining and improving existing road networks, a major function of government, is often justified as a public good investment in economic activity and job creation. The vast majority public expenditures on roads is actually on dedicated to maintenance and upgrading of existing roads, whereas road network expansions are a relatively rare event. The World Bank (1994) estimates that the returns to road maintenance are twice as high as those for network expansion.

Unlike most previous empirical work that focuses on gains from trade through network expansion, we posit a model based on road quality as a productive input into firm and agricultural productivity. Road quality affects the costs of inputs as farmers use road networks to take produce to markets, firms use them to acquire inputs and deliver their output, and workers use them to reach their jobs. Road quality also may affect the organization of production. Better roads allow farmers to work on plots in more distant locations and share large equipment, firms can more easily source their workers and inventory from further locations, and they can more easily combine capital and labor.

Despite its importance, there is remarkably little research on how investment in road quality affects local markets. This paper aims to fill this gap by analyzing the effects of road quality on local economic activity. Instead of focusing on gains from trade with other distant localities, we construct a model in which roads are a direct input to firm and farm production. We hypothesize that improvements to road quality increase farm and firm productivity and profits. As a result the demand for labor rises resulting in more employment at higher returns. Firms and individuals may migrate into the area to take advantage of the better roads and a stronger labor market causing the price of land to rise.

The empirical analysis is made possible by an unusually long and comprehensive road quality administrative database collected by Indonesia's road authorities. Quality is measured annually for each road segment in the country using an international road roughness measure collected by sensors in vehicles as they travel along the roads. This administrative database tracks road-segment level quality for universe of Indonesia's highways over a 20 year period. We merge these data with a nationally representative household panel database and with the annual census of manufacturing firms.

We propose a novel instrumental variable for identification that takes advantage of Indonesia's two stage budgeting procedure for roads maintenance and upgrading. While different roads are maintained by different National, Provincial and District road authorities, the budget for each authority is almost exclusively provided by the national government. The ministry of finance first allocates budgets to each authority and then the authorities decide how to allocate those funds among the road segments under their jurisdiction. Because road maintenance budgets are allocated using a formula that does not respond to short term changes in economic activity in the local economic conditions in the budget authorities' jurisdiction we can use these to construct a plausibly exogenous instrument for road quality.

We find that higher road network quality is associated with improvements in household consumption and income. More specifically, we show that one standard deviation improvement in road quality in a district results in a five percent increase in household consumption per capita, and an almost twenty percent increase in total labor earnings. In terms of mechanisms, we show that roads increase firm profitability and job creation in the manufacturing sector mostly through new firm openings. We find no effect on manufacturing wages. Although we find increased returns to agriculture, the gap in average earnings between agriculture and manufacturing employment is reduced but not eliminated. As a result, we observe an occupational shift from agriculture into manufacturing and higher profits for those who stay in agriculture. We interpret this gap in sectoral wages along with a very elastic supply of labor to manufacturing as consistent with the existence of dual labor markets in Indonesia (Lewis, 1978; Fields, 2004). Finally, we observe in-migration of households and higher land rents.

This research makes a number of important contributions to the literature. We are among the first to examine the effects of quality of existing roads rather than network expansions. Second, we focus on the effects of local roads on local markets as opposed to gains from trade due to lowering costs of transport between distant markets. Third, measuring changes in regional transport costs is difficult as there are no readily available datasets that document the evolution of the quality of transport infrastructure in developing countries.[1] In this sense, the road quality data we use represent a substantial improvement because they provide a very direct measure of transport infrastructure quality. Fourth, road improvements are typically not randomly assigned. This means that estimates of the effects of changes in transport

[1]Previous approaches, widely used in the trade literature, involve *inferring* transport costs from a gravity equation, using data on regional trade flows (e.g. Anderson and Van Wincoop, 2004) or price differences (e.g. Donaldson, 2010) which are not widely tracked and hence rarely available.

costs may be confounded by the fact that areas receiving improvements were selected by policymakers for economic growth or expected growth reasons, creating simultaneity bias. We address the simultaneity problem by making use of the new instrument described above, which has the advantage of being replicable in many countries.

This paper relates to recent contributions to the empirical literature on the evaluation of transport infrastructure in the development economics field. Qian et al (2012) find few effects of Chinese trunk roads among newly connected small counties in terms of GDP growth. Related work from Morten and Oliveira (2013) uses the construction of highways to the new capital city of Brasilia in the 1960s and documents that new roads generated growth of GDP, wages and additional migration.

Two recent studies of roads using a trade lens include Faber (2014), who presents evidence that high-capacity trunk highways linking large Chinese cities led to a reduction in industrial output from peripheral counties compared to non-connected peripheral counties. The second is Donaldson (2013) who estimates gains from trade from lower transport costs caused by the construction of India's railroads.

The urban economics field has produced substantial work using the U.S. interstate highway system expansion to analyze the effects of new highways. Duranton and Turner (2012) investigate city growth effects, Michaels (2008) analyzes skill premia changes, while Baum-Snow (2007) documents suburbanization effects.

This paper proceeds as follows: In section 3.2 we provide a theoretical framework, followed by the section 3.3, which lays out our main datasets used. Section 3.4 describes the historical background on evolution of road quality and two-step budgeting decision process in financing road improvements, which argues in favor of the instrumental variable strategy, described in sections 3.4 and 3.5. Section 3.6 presents the results and section 3.7 concludes.

3.2 Theoretical Framework

We conceptualize road quality as a village productive amenity, as in Jacoby (2000). The model we propose embeds a standard, prototypical model of the agricultural household (e.g. Singh et al., 1986; Benjamin, 1992; Bardhan and Udry, 1999) in a spatial general equilibrium model typically used in labor and urban economics (e.g. Rosen, 1979; Roback, 1982).

The economy is a collection of M discrete villages. Villages are each endowed with a single productive amenity, which we denote by A. In our application, we assume that villages with better roads have larger values of A. Without loss of generality, we can order these villages by their values of this productive amenity, as follows:

$$\underline{A} \equiv A_1 < A_2 < ... < A_M \equiv \overline{A}$$

There are two types of agents in the model: firms, which make use of land and labor in constant-returns-to-scale production, and household-workers, which decide how to allocate land and labor between farming and off-farm employment. We describe the objectives and constraints of each of these agents in the next subsections. The objective of this section is to provide a conceptual framework within which to interpret the empirical analysis.

Firms

Firms produce a composite good, X, under a constant-returns-to-scale production environment. The good is freely traded over space, and without loss of generality, we can normalize output prices to 1 (the model's numeraire). Firms use capital, labor, and land for production. For simplicity, we assume that capital is freely traded over space at rate ρ, and we also assume that each firm inelastically demands a single unit of land for production. Both rents, r, and wages, w, vary across locations and are determined in equilibrium. The firm's problem is to choose a cost-minimizing quantity of labor, L^M, to satisfy production requirements.

Crucially, we assume that there is some value of the productive amenity, $\widetilde{A}$, such that if a village has $A < \widetilde{A}$, no firm production will take place. At values of $A < \widetilde{A}$, firms will not be able to produce output with *any* amounts of capital or labor. This amenity threshold introduces non-concavity in an otherwise standard production function. We can write the firm's problem as follows:

$$\min_{L^M,K} wL^M + \rho K + r \qquad \text{subject to} \tag{3.1}$$

$$X = \begin{cases} 0 & if A < \widetilde{A} \\ G\left(L^M, K, 1;\, A\right) & if A \geq \widetilde{A} \end{cases} \tag{3.2}$$

Optimal choices of inputs lead to cost minimization. This yields a cost function, denoted by, $C(w, r; A, \rho)$, which maps wages, rents, the price of capital, and the value of the amenity into the minimized value of production costs. Because of free entry, firms in locations where $A \geq \widetilde{A}$ will produce until the point where production costs equal output prices (which are normalized to 1). This gives us the following equilibrium condition:

$$C(w, r; A, \rho) = \begin{cases} \infty & A < \widetilde{A} \\ 1 & A \geq \widetilde{A} \end{cases} \tag{3.3}$$

Note that because the price of capital is equal over space, ρ is a parameter in this expression. By applying Sheppard's Lemma to the convex portion of the cost function, we obtain the following expression for labor demand:

$$L_d^{M*} = \begin{cases} 0 & A < \widetilde{A} \\ C_w(w, r; A, \rho) & A \geq \widetilde{A} \end{cases}$$

Households

For simplicity, we assume that agricultural households are unitary, abstracting from intra-household resource allocation considerations. In each village, households inelastically demand a single unit of land, which they rent at rate r and use for farming. Households choose quantities of consumption, C, and decide how to allocate their labor endowment, E^L, between working on the farm, L^F, and working for a firm, L^M. Conditional on the choice of a location, indexed by A, the household's problem can be stated as follows:

$$\max_{C,L^M,L^F} U(C) \quad \text{subject to} \tag{3.4}$$

$$C \leq F\left(L^F, 1\,;\, A\right) - r + wL^M \tag{3.5}$$

$$E^L = L^M + L^F \tag{3.6}$$

where the utility function, $U(\cdot)$, is assumed to be continuously differentiable, with $U'(\cdot) > 0$, and $U''(\cdot) < 0$, and $F\left(L^F, 1\,;\, A\right)$ is the farm production function. We assume that $F_A > 0$, as higher A leads directly to higher farm output per unit of input, or it may lower input prices.[2] Equation (3.5) is the full income constraint, stating that the value of the household's consumption cannot exceed the value of farm profits plus off-farm labor earnings.

Equation (3.6) is the labor resource constraint, stating that the household's total labor endowment is equal to the total amounts of labor supplied to the market and to the farm. Implicitly, we are assuming that the household does not value leisure and that there is no unemployment. If individuals cannot spend their entire endowment of labor in off-farm employment, they must spend the rest of their time farming. Under this assumption changes in A have no effects on the extensive (probability of working) or intensive (hours of work) margins of labor supply.

Note that we also assume, for simplicity, that labor on the farm can only be directly supplied by the individual farmer household. This makes sense in a rural context with large families and contracting imperfections. We solve the household's problem in two cases, depending on the value of the amenity. If $A \geq \widetilde{A}$, then there is no job rationing, and the agricultural household model is separable. Following Benjamin (1992), we solve the model recursively. The household first chooses quantities of labor and land to maximize farm profits, which yields the farm's profit function:

$$\pi^*\left(w, r\,;\, A\right) = \max_{L^F} F\left(L^F, 1\,;\, A\right) - wL^F - r$$

This gives us the household's income, which we can write as:

$$M^* = \pi^*\left(w, r\,;\, A\right) + wE$$

Optimal choices of consumption lead to an indirect utility function, which maps income and prices into the maximum amount of household utility that can be attained:

$$\begin{aligned} V &= \widetilde{\psi}\left(\pi^*\left(w, r\,;\, A\right) + wE\right) \\ &\equiv \psi\left(w, r\,;\, A\right) \end{aligned} \tag{3.7}$$

[2]Note that the consumer-farmer-worker does not obtain utility for services provided by housing. Nor does the household's utility depend on A directly. The entire consumer amenity aspect of A comes through the dependence of household indirect utility on the farm production function, $F(\cdot)$.

On the other hand, if $A < \widetilde{A}$, then there are no off-farm employment opportunities for work, and $L^M = 0$. The household's supplies all of its labor to the farm, yielding the following value of income and consumption:

$$C = F\left(E^L, 1\,;\, A\right) - r$$

This implies that the farmer's indirect utility is:

$$\begin{aligned} V &= \widetilde{\psi}\left(F\left(E^L, 1\,;\, A\right) - r\right) \\ &= \psi\left(w, r\,;\, A\right) \end{aligned}$$

In locations where $A < \widetilde{A}$, at some wage levels, workers may want to supply labor to the market. This could be because at some level of farming labor, $L^{F*} \leq E^L$, the marginal product of labor is zero. However, because of the absence of non-farm employment opportunities, the household supplies its entire labor endowment to the farm. The combination of non-concavity in the manufacturing production function with respect to A and eventual zero marginal product of labor in farming creates dual labor markets (Sen, 1966; Dickens and Lang, 1985).

If migration is costless, workers will be perfectly mobile across locations, so that the following must hold:

$$\overline{V} \equiv \psi\left(w, r\,;\, A\right) \text{ for all } A \tag{3.8}$$

On the other hand, if migration is costly, there can be a wedge in utility across space.

Land Markets

Both farmers and firms inelastically demand a single unit of land, which they rent at rate r. Land is supplied on a spot market, according to marginal cost, so that:

$$r = h'\left(N_f + N\right) \tag{3.9}$$

where N_f is equal to the number of firms, N is equal to the number of agricultural households, and $h\left(\cdot\right)$ represents the total cost of clearing land for production and dividing into plots. We assume that $h\left(\cdot\right)$ is increasing and strictly convex in the total population of firms and households, $N_f + N$.

Spatial General Equilibrium

We consider spatial equilibrium conditions to hold in the long-run. Firms are created and labor reallocates, and as this occurs, wages and rents will adjust to equalize production costs and indirect utility over space. Mobility provides us with three equilibrium conditions: (3.3), (3.8), and (3.9). Using these, we can show a series of propositions from the model. We first show that because A is a productive amenity, increases in A will result in an entry of firms and an in-migration of households.

We can implicitly differentiate the equilibrium conditions, (3.3), (3.8), and (3.9), and solve for the comparative statics.

$$\begin{bmatrix} C_w & C_r & 0 \\ \psi_w & \psi_r & 0 \\ 0 & -1 & h''(\cdot) \end{bmatrix} \begin{bmatrix} \partial w / \partial A \\ \partial r / \partial A \\ \partial \left(N_f + N \right) / \partial A \end{bmatrix} = \begin{bmatrix} -C_A \\ -\psi_A \\ 0 \end{bmatrix}$$

Solving the second equation gives:

$$\frac{\partial r}{\partial A} = \frac{1}{\Delta} \left(\underset{+}{C_w} \underset{+}{\psi_A} - \underset{-}{C_A} \underset{+}{\psi_w} \right) > 0$$

where the denominator, Δ, is positive and given by:

$$\Delta = \underset{+}{\psi_w} \underset{+}{C_r} - \underset{+}{C_w} \underset{-}{\psi_r} > 0$$

So that increasing A leads to increases in land values (rents).

Solving for the first equation gives:

$$\frac{\partial w}{\partial A} = \frac{1}{\Delta} \left(\underset{-}{C_A} \underset{-}{\psi_r} - \underset{+}{C_r} \underset{+}{\psi_A} \right) \gtreqless 0$$

So that increasing A leads to ambiguous effects on wages. Note that in low amenity villages, the marginal product of agricultural labor is very low because households in these villages allocate their entire labor endowment to the farm. Road improvements in these villages can attract some labor labor away from farming without raising wages. It is straightforward to see that the model predicts an increasing number of firms and consumption when road quality increases.

3.3 Data

In this section we describe road quality data and the main measures of road roughness used for the analysis. We also describe other data sources from which we draw our main outcomes of interests at either firm, household or individual level.

Road Quality Data

One of the key problems when evaluating the impacts of changes in transport infrastructure lies in measuring its quality. There are no readily available datasets that document the evolution of the quality of transport infrastructure in developing countries.[3] We construct

[3] Previous approaches, widely used in the trade literature, involve *inferring* transport costs from a gravity equation, using data on regional trade flows (e.g. Anderson and Van Wincoop, 2004) or price differences (e.g. Donaldson, 2010) which are not widely tracked and hence rarely available. Although regarding the latter some recent papers have made use of price series from underlying CPI components. See Atkin et al 2014.

a very direct measure of road roughness, and improve the measure of road infrastructure quality used thus far by using a novel longitudinal dataset on road roughness in Indonesia.

Every year, the Department of Public Works (*Departemen Pekerjaan Umum*, or DPU) conducts annual, high resolution data to track road quality and monitor pavement deterioration.[4] These measurments were conducted by a team of surveyors, who measured the surface type and width of road segments and also collected longitudinal data for measuring road roughness. Hence, the original dataset is extremely detailed, with more than 1.2 million kilometer-post-interval-year observations, merged to the kilometer-post interval data to shapefiles of the road networks. This yields a 1990-2007 panel of road quality measures along major inter-urban roads.

The measure of road quality is based on its roughness: when faced with potholes, ragged pavement, or unpaved surfaces, drivers slow down, and this reduction in speed increases travel time and the cost of transport. Road roughness is measured by the international roughness index (IRI), a widely accepted measure of road quality in civil engineering, developed by the World Bank in the 1980s.It is defined as the ratio of vehicles accumulated suspension motion (in meters), divided by the distance traveled by the vehicle during measurement (in kilometers).[5] All else equal, when driving on gravel roads or when faced with potholes and ragged pavement, drivers prolong their travel time through decreased travel speed and consequently, road roughness increases transport costs.[6]

We merge the data of road roughness measure with a digital map of the road network, using road-link identifiers, which yields a panel of road quality measures used for the analysis. We use two measures of road quality. First, our primary measure of local road quality is a distance-weighted average of roughness for all roads in district d. Let $r = 1, ..., R(d)$ index road segments in district d, and let d_r denote the length of the road segment r. Our average roughness measure is defined as:

$$\text{Roughness}_{dt} = \frac{\sum_{r=1}^{R} d_r \text{IRI}_{rdt}}{\sum_{r=1}^{R} d_r}$$

where IRI_{rdt} denotes the road roughness of road section r in district d at time t. Second, to the extent to which road quality improves access to markets, we construct a market access measure of road quality taking into account road quality to the nearest provincial capital as well.

[4]These data are part of Indonesia's Integrated Road Management System (IRMS).

[5]Using the mapping between subjective measures of ride quality and roughness at different speeds, provided by Yu et al. (2006), one can determine the maximum speed that one can travel over a road with a given roughness level while maintaining a constant level of ride quality. Given this roughness-induced speed limit, one can calculate travel times along network arcs and to compute the time-minimizing path between different regions (Dijkstra, 1959).

[6]Although we suspect that road roughness induces greater transport costs through its effect on travel times, this could also happen through other mechanisms as well, including fuel consumption or labor costs, which both account for more than 50 percent of vehicle operating costs in Indonesia (Asia Foundation, 2008). Other significant cost factors included lubricants and tires (13%), and other maintenance costs (4%), all of which should increase as cars are driven on rougher roads.

Census of Manufacturing Firms

Our primary data source for firm-level outcomes and analysis is the manufacturing census Indonesia's Annual Census of Manufacturing Establishments (*Survei Tahunan Perusahaan Industri Pengolahan*, or SI), collected by Indonesia's central statistical agency, (*Badan Pusat Statistik* or BPS). The SI is an annual census of manufacturing plants with more than 20 employees and contains detailed information on plant's cost variables, their industry of operation, employment size and measures of value added. An Advantage of SI data is that it contains firm-level identifiers, allowing us to track changes in firm-level outcomes over time. The data contains information on the firm's starting date and its location at the district level, as well as firm-level outcomes, such as employment and wage rates, value added, output, and total factor productivity.[7]

Indonesian Family Life Survey

Our data source for the individual and household outcome level analysis, such as individual employment status (working or not) and type (formal/informal), hours worked, earnings, and household consumption per capita, is the Indonesian Family Life Survey (IFLS). The IFLS is a nationally representative longitudinal survey that was collected in 1993, 1997, 2000 and 2007. The IFLS is representative of 83 percent of Indonesia's total population and follows more than 30,000 individuals over a 14 year period. These individuals are observed in more than 300 villages (*desa*), which are located in 13 of Indonesia's 27 provinces. The IFLS is notable for its low attrition rate, as more than 87 percent of the original households are tracked through all four waves of the survey.

The IFLS contains detailed modules not only of labor market activity or household level consumption, but also data on other demographic characteristics, including age, gender, and educational attainment. Figure 3.1 shows the locations of IFLS villages used throughout our analysis.

3.4 Background

Historical Context

Indonesia's economy has been one of the fastest growing 'tiger' economies in Asia, especially between the mid 1970s and the late 1990s, enjoying annual growth rates of more than six percent and making major progress in tackling poverty, which decreased from 60 percent to 10 percent (Hill, 2000). This period of economic success overlapped with the Suharto

[7] New firms are counted when they appear in the dataset having never appeared before. Also, for the purpose of our analysis, we dropped all firms coded as state-owned enterprises (less than 3 percent of all firm-year observations). Throughout the discussion, we use plants and firms interchangeably since less than 5% of plants in the dataset are operated by multi-plant firms (Blalock and Gertler, 2008).

regime and his government's investments in major public works programs, including those that targeted road improvements.

When Suharto assumed power in 1967, rehabilitation of roads, which had been built by the Dutch colonists and left to deteriorate until his time, became a top priority of his first five-year development plans, called Repelitas (Rencana Pembangunan Lima Tahun, or *Repelita*). Due to oil price shocks and the consequent collapse of state oil revenues in the late 1970s, spending on road infrastructure stagnated until the early 1990s. Since then, however, the budget share allocated to road improvements increased substantially. Between Repelita IV (1984 - 1989) and Repelita V (1989 - 1994) [8], the total budget for road improvements almost doubled from \$2.1 to around \$3.9 billion.[9] During Repelita V, transportation constituted the single largest item of the development budget and formed almost 20 percent of total planned development expenditures and almost ten percent of GDP.[10] However, after the fall of Suharto, followed by the Asian financial crisis in 1997/1998, infrastructure investment fell to less than three percent in 2000, and increased substantially ever since (World Bank, 2012).

These investments translated into road infrastructure quality that varied both, spatially and over time. Figure 3.2 shows a significant leftward shift in the distribution of Roughness_{dt} across districts between 1990 and 2007. Similarly, Figure 3.3 documents substantial spatial variation in road improvements over time. For instance, Sulawesi's road network went from 84 percent of unpaved roads in 1990 to only 46 unpaved roads only a decade later. Between 1998s and early 2000s, Indonesia's roads started to deteriorate, which is evident in Figure 3.4, but started to pick up soon after; closely following the road expenditure pattern over time.[11]

Financing Road Improvements

In this section, we first describe and then show empirical evidence for a ***two-stage*** budgeting approach used for road improvement decisions in Indonesia (Deaton and Muellbauer, 1980). In the *first stage*, Indonesian central government distributes a share of centrally collected tax revenues required for road investments to local administrative units, such as districts or provinces, by following a national distribution formula. That is, local budgets are determined independently from recent changes in local economic conditions or from expectations thereof. In the *second stage*, agencies observe their allocated funds, and endogenously select which roads to upgrade.[12] We use this plausibly exogenous variation in budget allocation from the Center to the local authorities as instruments for local road quality to identify the parameters of interest.

---

[8]The first Repelita, Repelita I, happened between 1969-1974.

[9]These figures are expressed in 2000 U.S. dollars.

[10]These figures are taken from various planning documents describing Indonesias five year development plans .

[11]Similar trends apply to Sumatra's and Java's highway network.

[12]For more details on the institutional arrangement for the road sector in Indonesia, see Figure 3.5.

Two-Step Budgeting

Even though the provision of road infrastructure was declared a local responsibility in the early 1990s, provincial and district governments still rely heavily on transfers from the central government.[13] Own-source revenue is relatively small for local district governments in particular, accounting for less than a ten percent of their total revenue. Provinces, on the other hand, became more self-sufficient, however less than fifty percent of their revenue is own-sourced (Green, 2005; World Bank, 2012).

As briefly mentioned above, in the *first stage* of budgeting, the Center distributes nationally collected revenue to the road maintenance agencies at the provincial (33) or district (more than 400) level according to the fund allocation formula. In particular, grants are allocated according to the national formulas that utilize differently weighted objective criteria on which local governments have little control and are hence orthogonal to changes in local economic developments. Formulas are mostly designed to help equalizing the fiscal capacities of sub-national governments, their regional development, and to decrease poverty rates; all aligned with the main national priorities (Bird and Smart, 2001).

Until the late 1990s, these priorities were clearly documented in Repelitas, long-term national spatial plans. For instance, they mandated that particular regions, especially major islands and sparsely populated or poorer areas, were those that would be the first to receive infrastructure improvements. Importantly, these plans were revised very infrequently, suggesting that the road authorities did not regularly respond to changes in outcomes or economic shocks.[14]

Even though national goals were not updated regularly, distribution and amount of local budgets for roads, that is, criteria for distribution formulas, and especially their weights, were. Before 1998, the distribution criteria for provincial and district development grants consisted of some fixed, lump-sum grant and the variable part, depending on the local area and population (for provinces), or length, density, condition of the roads and the unit cost of their construction and maintenance (for districts). These allocation criteria changed over time and differ significantly in periods before and after the year 2000. Figure 3.6 shows the structure and changes of weights before the 2000s. After that, the main revenue sources for sub-national road development are broadly divided into two groups: first, earmarked financing and second, non-earmarked financing. Earmarked financing consists of DAK for roads (Daka Alokasi Khusus, or Special Allocation Grant) and a minimum of ten percent from motor-vehicle revenue. The DAK allocation has a formula component that takes into account the fiscal gap with a 10 percent matching requirement. With covering more than 85 percent of road expenditures (consisting of DAU (Dana Alokasi Umum, or General Allocation

[13]The government raises more than 90 percent of total local government revenues, dominated by the tax on natural resources, such as oil and gas, and followed by the income tax, VAT and luxury tax.

[14]We do not have access to the exact national spatial plans used, however this idea was confirmed by conversations with highway authorities at Indonesia's Department of Public Works (DPU). Also, in every national planning and budgeting document we do have access to, no information is provided at levels below the province.

Grant), revenue sharing, own revenue, and others), non-earmarked financing represents the main source of funding for sub-national road development.

DAU accounts for almost 50 percent of all sub-national revenues, and is distributed based on a national formula. It equals the sum of a portion of the sub-national budget spending on public servant salaries (also called basic allocation) and their fiscal gap. Fiscal needs are based on regional variables such as population, area, GDP per capita, and human development index.[15] Again, weights for each criteria or their implementation change every couple of years either for DAK (see Figure 3.7) or DAU (e.g. see Figure 3.8), mostly through changes in decentralization laws, in order to speed up and comply with national priorities of improving equal economic development(World Bank, 2008).[16] For more details on revenue sources over time see Figure 3.9.

The extent to which local grants and revenues are allocated for road improvements is at the discretion of provincial and district/city governments. As mentioned above, their decisions represent the *second stage* in the road improvement decision process, which is independent from the first one. One of the official criteria for the intergovernmental transfers claims that: ..."The subnational governments should have independence and flexibility in setting their priorities. They should not be constrained by the categorical structure of the programs and uncertainty associated with decision-making at the Center"... (Shah et al. (1994), p.72). This is true especially since the mid 1990s, when most of the government expenditure functions were devolved to provinces and districts, which generated an orthogonality between revenues received and locally made expenditure decisions (Green, 2005; World Bank, 2008).[17]

Instrumental Variable

In this section we use empirical evidence to confirm investment model described above by first, showing how we define local budgets, and second, demonstrating that these budgets are indeed determined exogenously to changes in local activity, yet their allocation within localities is endogenous.

We define the proxy for total budget allocated for national, provincial, and district road improvements, in the following way:

$$B_{at} = \sum_{r \in R(a)} d_r I_{rt}$$

[15]Fiscal capacity is measured by a regions own-source revenue and a fraction of total revenue-sharing.

[16]For example, until 2005, the basic allocation component of DAU consisted of a lump sum and civil service wage bill covering a portion of wage bill and used the poverty index as one of the "fiscal gap" criteria. Since 2006, however, the basic allocation component of DAU covers the full wage bill, whereas weights for the "fiscal gap" criteria, such as own revenue, shared tax revenue and natural resource revenue, changed. Also, human development index and GDP replaced the poverty index used prior to 2006.

[17]Sub-national governments are now managing over 35 percent of total public expenditures, compared to 24 percent in the mid 1990s. Figure 3.10 shows a significant and relatively constant gap between revenues and expenditures at a national and regional level over time.

where $R(a)$ denotes the set of road segments under administrative authority a (national, provincial, or district), d_r is the distance of a road segment r in kilometers, and I_{rt} is an indicator for whether or not segment r was upgraded at time t (i.e. $I_{rt} = 1$ whenever roads are improved, thus $\Delta\text{Roughness}_{rt} < 0$). In words, a proxy for a local budget for road improvements B_{at} for adminstrative authority a in year t equals total kilometers of roads upgraded in year t that are administered by authority a. In our data, there is a single national authority, 17 provinces, and 218 districts.

As described in Section 3.4, the first stage of funds distribution follows a national formula. Changes in this formula, either in the criteria or the their weights used, are, conditional on fixed effects, therefore responsible for local variation in road quality over time. If variation in our instruments is not triggered by sudden actual or expected changes in local economic conditions, then our instrument satisfies the exclusion restriction requirement. To test this hypothesis, we regress the road budget of provincial road authorities against the lagged province GDP and lagged number of firms (see Table 3.4 columns 1 and 2, respectively). Columns 3 and 4 show an analogous exercise for district road budgets against district GDP and district number of firms. In all cases we cannot reject that the coefficients on economic conditions are uncorrelated to the road budget. The national road budgets do respond to local economic conditions, hence we do not use them as an instrument for road quality. These results increase our confidence that the exclusion restriction of our proposed instrument at the province or district level is satisfied and the financing of the first stage indeed functions as described above.

Since one of the criteria for fund allocation is length, density and condition of the existing road network and local economic development (see Section 3.4), we expect local road budgets to be correlated with road quality, making them a relevant instrument for road roughness. Columns 1, 2, and 3 in Table 3.3 show that budgets for all types of roads in the province (district) are indeed strongly correlated to road quality in any given segment in the province (district). Columns 4, 5, and 6 in the same table show that budgets have a large and significant effect on road quality. The negative sign of the coefficient confirms that whenever the province or district has a more generous road investment any given road segment in the province (district) is more likely to be of better quality (less rough). These results confirm the relevance of our instrument for road quality.

3.5 Empirical Strategy

To the extent that road quality is a productive amenity, valued by farm producers and formal sector manufacturing firms, our model predicts that road improvements would encourage outcomes such as entry of firms, an increase in their output and value added, or employment. To test for the effect of road quality changes on these outcomes, we exploit a temporal variation in road roughness changes to identify the parameters of interest by estimating the following equation:

$$y_{dt} = \alpha_d + \alpha_t + \beta \log\left(\text{Roughness}\right)_{dt} + \mathbf{x}'_{dt}\theta + \varepsilon_{dt} \tag{3.10}$$

where d indexes districts and y_{dt} is the dependent variable observed for district d at time t. The variable $\log(\text{Roughness})_{dt}$ measures the log of the average roughness for all roads in a district d at a time t. The vector $\mathbf{x}_{dt}$ represents a set of time-varying controls, including non-oil district level GDP, and log of population. Panel data allow us to control for time-invariant unobservables that may be correlated with changes in road quality and outcomes of interest through district and time fixed effects, or α_d and α_t respectively. Standard errors are clustered at the district level.

Our estimates may be confounded by the fact that areas receiving improvements were selected by policymakers for economic growth or expected growth reasons, creating endogeneity bias. If road authorities routinely allocated funds for road upgrades to locations that were historically poor (or wealthy), that is, they were adopting the long-term development plans (Repelitas) (see Section 3.4), fixed effects would be sufficient to remove the targeting bias from our parameter estimates. However, our analysis shows that, even though that was the case, local road authorities did allocate funds to either faster growing areas, or areas experiencing sudden declines by responding to their local economic shocks. Hence, the worry of targeting bias remains. In addition, if substantial economic activity causes roads to deteriorate faster due to their extensive use, attenuating our results through a reverse causality bias represents an additional worry.

To address these concerns, we use a IV-GMM approach as our preferred specification, using road maintenance budgets at a district level as instruments for road quality. In Section 3.4 we confirm that this instrument is correlated to road quality changes over time but is orthogonal to local demand for roads, satisfying the relevance and exclusion restriction.[18]

3.6 Results

In this section, we first provide evidence on the relationship between changes in road roughness and household welfare. Next, we empirically evaluate the different mechanisms through which these welfare effects occur, following the discussion in Section 3.2.

Welfare Effects

In this section, we provide evidence on the effect of road roughness on welfare, measured with individual earnings and household level consumption. We estimate the following household and individual-level panel regression equation:

$$y_{it} = \alpha_i + \alpha_t + \beta \log(\text{Roughness})_{d(i)t} + \mathbf{x}'_{it}\theta + \varepsilon_{it} \tag{3.11}$$

where i indexes individuals (or households), y_{it} is the dependent variable observed for i at time t, α_i is a fixed effect for i, and α_t is a time effect. The variable $\log \text{IRI}_{d(i)t}$ measures

[18]Our first stage relationships between road roughness and budget shifters, across districts-years, households-years, and individual years, can be found in Appendix Table B1.

the log of the average roughness for all roads in i's district, denoted $d(i)$, at time t. Note that the roads are a mix of roads maintained by the national, provincial and local road authorities. The vector $\mathbf{x}_{it}$ represents a set of time-varying controls, including district-level GDP, household size, and month of survey indicators. For individual-level regressions, controls also include age and education. Unless stated otherwise, all parameters are estimated with GMM, where we instrument $\log \mathrm{IRI}_{d(i)t}$ with the budget shifting instruments described in the previous section. Whenever using GMM together with these instrument, we refer to it as IV-GMM.[19] Standard errors are clustered at the village level.

Using the log of household consumption per capita as the dependent variable, we report the results of (3.11) in Table 3.6, Panel A.[20] We find that a one percent increase in road roughness significantly decreases consumption per capita by 0.26 percent. We report the Kleibergen and Paap (2006) Wald rank F-statistic, robust to clustering and heteroskedasticity, which at the value of 35.45 rejects the null hypothesis of weak instruments.

In Panel B, we estimate the relationship between individual log total monthly earnings and road roughness. Total monthly earnings consist of wage earnings and net profits from primary and secondary occupations.[21] Controlling for hours worked allows us to interpret the estimate of β as changes in effective wages. We find that a one percent increase in road roughness significantly decreases total earnings by 0.89 percent.

In the next subsections, we examine many different mechanisms behind these effects, guided by our theoretical framework.

Inmigration of Workers and Entry of Firms

To the extent that road quality is a productive amenity, valued by farm producers and formal sector manufacturing firms, our model predicts that road improvements would encourage entry of firms and an in-migration of workers. This could result in additional employment opportunities and allow individuals to move out of agriculture into the formal manufacturing sector, enjoying relatively higher returns than before. Depending on the relative shifts in labor demand and labor supply, welfare could improve directly through increased wages as well.

To examine the hypothesis on in-migration of workers, in Table 3.7, Panel A, we report IV-GMM estimates of the following cross-sectional regression equation:

$$y_d = \alpha + \beta \Delta \log \left(\text{Roughness}\right)_d + \mathbf{x}_d' \theta + \varepsilon_d \tag{3.12}$$

[19]Our first stage relationships between road roughness and budget shifters, across districts-years, households-years, and individual years, can be found in Appendix Table B1.

[20]Household consumption per capita is defined as the sum of total monthly expenditures on food (meat, fish, dairy, spices, sugar, oil, beverages, alcohol, tobacco, snacks, eating out, staples, vegetables, dried food) and non-food items (utilities, personal goods, household goods, recreation expenses, transport expenses, lottery, clothers, furniture, medical expenses, ceremonies, taxes, other, rent or other house costs, education).

[21]Family workers are excluded from earnings regressions since their income is reported as missing or zero in most cases.

where y_d is the log of district d's share of recent migrants, calculated as the number of people from Indonesia's Household Census of 2000 in each district who reported living in another district five years prior, and $\Delta \log \text{IRI}_d$ is the change in log roughness between 1995 and 2000. Controls in this regression, denoted by $\mathbf{x}_d$, include long lags of non-oil GDP and population, both measured in 1990. In column 2, we control for province fixed effects, too.

Results suggest that a one percent increase in the road deterioriation rate significantly decreased share of migrants by 0.7 percent. For the average district, with about 6% of recent migrant population share, a one percent increase in the rate of road deterioration would result in a reduction of the migrant share to approximately 5.6%. Calculated at the average population of Indonesia's districts ($\approx$ 643,000), this represents a reduction of total recent migrants from 38,594 to 35,892, a decrease of only 2,700 individuals. This is a small effect, which is not surprising given barriers to mobility in Indonesia (Bazzi, 2013).

In Panel B, we use data from Indonesia's Annual Survey of Manufacturing Establishments (*Survei Tahunan Perusahaan Industri Pengolahan*, or SI) to test whether changes in road roughness affected the entry of manufacturing firms.

In column 1, we regress the log of the number of new firms in district d at time t on the log of road roughness in district d at time t, using district-level panel data from 1990 to 2007. We control for district and year fixed effects, controls used in (3.12). In column 2, the dependent variable is the number of manufacturing jobs, equal to the total number of workers hired by manufactures in the SI.

Our IV-GMM estimates suggest that a one percent increase in road roughness is associated with a 1.2 percent decrease in the number of new manufacturing firms, and a 1.3 percent decrease in the number of manufacturing jobs. This is equivalent to a reduction of slightly more than 1 firm, and the loss of approximately 215 jobs due to a one percent increase in road rougness. From 1990 to 2000, the average district experienced an average reduction in roughness of approximately 46%. From the Panel B estimates, this is estimated to have led to a 55.2% increase in the number of firms (42.3 more firms, on average), and a 59.8% increase in the number of workers (7,700 more workers, on average).

In summary, we find that road improvements trigger the creation of manufacturing firms and increase district level local labor demand. We also find that the worker's mobility due to better roads was much smaller than the one from firms. Next, we use data on firms directly to examine the within-firm effect of road improvements to further explore the mechanism behind the significant labor demand increase.

Firm-Level Outcomes

An advantage of using the SI data is that it contains firm-level identifiers, allowing researchers to track changes in firm-level outcomes over time. In Table 3.8, we estimate the relationship between road roughness and log firm-level employment (Panel A), log value added (Panel B), and log total factor productivity (TFP) (Panel C). Employment and value added measures are recorded directly in the SI data. We use a control-function approach to calculate TFP through productivity residuals estimation (e.g. Olley and Pakes, 1996; Levinsohn and Petrin, 2003).[22] Our regressions take the same form as (3.11), with firm-year observations ranging from 1990 to 2007.

In Column 1 of Table 3.8, we estimate regressions using only year, industry, and district fixed effects. We find that a one percent increase in road roughness translates to a 0.06 percent decrease in the number of workers, a 0.14 percent decrease in log value added, and a 0.10 percent decrease in total factor productivity. However, in Column 2, when we condition on firm fixed effects as well, the significant effects of road roughness on total workers disappear. This suggests that changes in labor demand come from the entry of new firms, rather than changes in labor demand for existing firms. This may be consistent with a perfectly competitive environment, or it may also reflect non-trivial adjustment costs; as road quality deteriorates, firms may want to be using less labor, but that would not be possible without reductions in capital, which may be costly. However, even with firm fixed effects, road deterioration reduces firms' value added and total factor productivity.

Labor Supply: Extensive and Intensive Margins

Our model makes the assumption that labor is inelastically supplied to manufacturing or agriculture. Validating this assumption in the data is important because individuals might respond to increased productivity of labor by working more hours, or entering the labor market, and we are interested in determining whether this is an avenue for the increase in household welfare. In Table 3.9, Panel A, we report results showing this is not the case by estimating our baseline specification (3.11) where the dependent variable is an indicator for whether or not individual i reports working at time t.[23] Approximately two thirds of the individual-year observations in the IFLS data reported working last week. IV-GMM estimates show that road improvements have no significant effect on probability of working at conventional significance levels.[24]

Similarly, In Panel B, we report estimates of the effect of roughness on working hours. Both hours worked and roughness are measured in logs. We again statistically insignificant effects of road roughness on hours worked.[25] Taken together, the evidence from Table 3.9 suggests that changes in road roughness do not improve total earnings or consumption through their effects on the extensive or intensive margins of labor supply.

[22]Precise details on the implementation of this approach for the SI data can be found in Poczter et al. (2014).

[23]Individuals are classified as working if in the last week they reported actually working, helping to earn income, worked at least one hour for pay, had a business and were usually working but just not last week, or worked on a family owned business (farm or nonfarm), and reported their occupation in either primary or/and secondary job. The indicator is only defined for individuals older than 15 years.

[24]The confidence interval around the point estimate reported in Panel A, Column 2 implies that if roughness increased by one standard deviation (1.887, or 0.635 log points), we could reject the hypothesis that employment would decrease by more than 3 percent, a moderate effect size.

[25]Note that this insignificance is not due to a lack of statistical power. From column 2, the confidence interval around the point estimate implies that we could reject the hypothesis that a one percent increase in road roughness causes more than a 0.15 percent reduction in hours worked.

Sector Switching

Next, we investigate the extent to which changes in road roughness result in sector switching: movement out of agriculture and services into manufacturing employment or from the informal and to the formal sector. In Table 3.10, we report estimates of the relationship between log average road roughness and measures of employment and hours worked in different sectors, using our individual-level panel regression specification (3.11). Across the columns of this table, we vary the dependent variables in the regression. In column 1, we use an indicator variable for whether the individual reported any employment in that sector. In column 2, we modify that indicator to be equal to one only if the individual reported that sector as his or her main occupation, and in column 3, we use the log of total hours worked by sector as the dependent variable. Along the rows, we vary the sector (in Panel A) or use a measure to distinguish between formal and informal employment (Panel B).[26] As before, we report GMM estimates which use the budget-share instruments, and cluster standard errors at the village level.

Despite not finding any effects on total hours worked or on the probability of being employed, we find strong evidence that changes in road roughness affect how individuals sort into sectors of employment. From the first row, we find that increases in road roughness reduce employment and hours worked in manufacturing. A one percent increase in road roughness is associated with a 0.1 percent reduction in reporting working in manufacturing and a 6.2 percent reduction in hours worked in manufacturing. The increase in manufacturing employment comes at the expense of employment in the agricultural and services sectors. The sales and services sector in this context tends to include very low-skill jobs, such as becak drivers, street vendors, and hairdressers.

To the extent that informal marginal employment is responsible for surplus labor in poor road quality villages, these results confirm the predictions of the model, particularly the dual economy interpretation. Given that we have already shown that road improvements were accompanied by an entry of manufacturing firms, this suggests that with increase in local labor demand, individuals respond by increasing their employment in manufacturing, and transitioning out of both possibly agriculture and sales and services.

In Panel B, we capture these same patterns by reporting switching behavior between the formal and informal sector. A one percent increase in road roughness is associated with a 0.08 percent increase in the probability of working in the informal sector, and a 6 percent increase in hours worked in the informal sector.

In Table 3.11, we investigate heterogeneity in the relationship between working in manufacturing and agriculture and different individual and household characteristics. Column 1 repeats the estimates reported in Table 3.10, Column 2, using all individual-year observations. In columns 2-11, we split the sample by five different characteristics: age (above and below median, columns 2 and 3), gender (columns 4 and 5), education (above and below median, columns 6 and 7), initial household per-capita consumption expenditures (above

[26] Formal and informal employment are defined in Section 3.6

and below median, column 8 and 9), and whether the household owned any farmland in the first year of the data (columns 10 and 11).

Panel A in the table provides clear evidence of heterogenous effects regarding who leaves agriculture when roads improve in the district: The effects are much stronger among males, older individuals, those less educated, the relatively poor and the landless farmers. This exercise is indicative of the characteristics of the marginal individuals are those who are most likely to be impacted by changes in road roughness.

In Panel B, we find evidence of similar elasticities across genders, age and land ownership. However, the initially poorer and less well educated are especially strongly attracted to manufacturing as a result of road quality improvements.

Total Earnings by Sector

As discussed in Section 3.2, improvements in road quality may have ambiguous effects on formal sector wages. In the short run, if firms are more mobile than workers, increases in local labor demand should offset increases in local labor supply, generating upward pressure on wages. We provide some evidence for the relative immobility of labor in Section 3.6. However, our model also suggests that dual labor market effects may suppress wages, to the extent that in areas with very poor market access, opportunities for outside employment are limited, and there is a surplus of labor in agriculture.

In Table 3.12, we provide estimates of the reduced form relationship between wages and road roughness, differentiating by sector, using a regression specification similar to (3.11). We again we use as dependent variables log total earnings by sector and control for total hours worked so that we can interpret the effects of changes in roughness as changes in wages.

From column 1, we show that when we separately estimate the relationship between road roughness and earnings by sector, we find that most of this effect comes from agricultural earnings (column 2). Although the coefficients on road roughness are negative for all other sectors, including manufacturing earnings (column 1), sales and services earnings (column 3), and other earnings (column 4), these coefficients are not significantly different from zero. However, for agriculture, the effect is large and statistically significant. A one percent increase in road roughness is associated with a 3.4 percent decrease in agricultural earnings, conditional on hours worked. As described in the model, there are many ways in which agricultural earnings and farm profits could increase because of better roads, including reductions in the costs of inputs and higher output prices. Another plausible explanation is that as surplus labor farms reduce agricultural labor because of better outside options, farm profits and total earnings per hours worked increases.

Land Prices and Land Rents

As firms and workers move into areas affected by road quality improvements, they increase the local demand for land and housing, and we would expect to see increases in the value of land in those areas. To provide evidence for this, we use household-level panel data from the

IFLS on land values and land rents and estimate hedonic rent regressions (e.g. Blomquist et al., 1988; Roback, 1982). Our regression specifications are similar to (3.11), but when using household-level data, we add a vector of housing controls, including variables for the number of rooms, indicators for access to electricity, piped water, internal toilets, and types of floor, walls, and roof material, in addition to our standard controls (district-level non-oil GDP and month or survey indicators). Standard errors are clustered at the village level.

Table 3.13 reports the results. In column 1, the dependent variable is the log of the land rent per-room. This variable is measured as the actual rent paid or, if rents were not paid, the monthly rent that the household estimates they would have been paid if they had been renting. Note that much of the data on land rents is imputed. Of the 33,049 household-years in the IFLS with rent data, 29,508 observations (89.3%) are based on imputed rents, with the remaining 3,541 observations (10.7%) representing rents actually paid. Because of this, we also control for an indicator for whether observations are based on actual rents. The coefficient is negative and statistically significant, and suggests that a one percent increase in average road roughness leads to a 0.24 percent decrease in rents.

In column 2, we use a different measure of land prices, based on self-reported land values. In the IFLS, households are asked to report the total market value for a series of different types of land they own, including land used for a farm business, land used for a non-farm business, and other owned farm or non-farm land not being used for business. This measure has limitations because it is based on self-reported data, instead of market transactions. Another issue with this measure is that because individuals were not asked about the size of their land holdings, we cannot control for total hectares owned. This leaves open the possibility that differences in total land values across households could be explained by differences in land sizes. Nevertheless, similar to the results in column 1, the coefficient on average road roughness is negative and statistically significant, suggesting that a one percent increase in average road roughness leads to a 0.66 percent decrease in land values. Taken together, both measures of land prices, despite their imperfections, provide substantial evidence for the notion that road improvements lead to increases in the value of land.

Robustness

Up to this point we have only focused on presenting our main results, in the interest of simplifying the exposition. However, in Appendix B, we subject our results to a number of robustness checks. We first present fixed-effects least squares results and find that our GMM estimates are quite similar. This is consistent with fixed effects specifications controlling for most of the endogeneity associated to road placement. We also estimate effects separately for rural areas, and although our sample sizes fall and standard errors become less precise, the point estimates are always quite similar. We also estimate effects separately for non-moving households; in our main specifications, changes in outcomes could entirely be a product of the fact that individuals were moving between locations, but we find no strong evidence of this. We also use different measures of road quality, including a roughness-induced travel time measure between village v and that village's nearest provincial capital, and measures of market access.[27] Finally, we provide GMM estimates using a Hausman IV, where we instrument district d's road quality with the average road quality of all other districts in the same province. In general, our results are very robust to these different robustness checks.

3.7 Conclusion

Even though transportation infrastructure investments typically account for a significant proportion of countries' budgets, little is known about their effects in developing countries, where spatial disparities are particularly pronounced. This paper aims to understand the role road improvement (or deterioration) can play in such countries, not only through looking at possible welfare effects, but also by investigating the different possible mechanisms through which these effects materialize. While much of the previous literature on this topic has focused on the construction of new roads, we add to the literature by evaluating the effects of substantial changes in road quality due to maintenance and upgrading of already existing roads in Indonesia.

Using a novel dataset that documents substantial variation in road quality in Indonesia, and combining this with high quality household panel data that spans years 1990 through 2007, we provide reduced form evidence that road improvements significantly increase welfare, measured either with consumption or income. Additionally, using an annual census of manufacturing firms, we show that these positive welfare effects partially materialize though increased labor market demand, generated by the entry of new firms rather than extended hiring by existing firms. However, we do not see substantial changes in the extensive or intensive margin of labor supply, but instead observe occupational shifts from agriculture into higher paying, newly available manufacturing jobs. In addition, while manufacturing wages typically don't exhibit an upward push, we do observe significant improvements in agricultural profits. This not only implies the wage gap between these two sectors is narrowed, but also confirms the predictions of our stylized model of dual labor markets. The latter shows under what conditions productive amenities, such as transport infrastructure, may translate into positive welfare effects.

The methodological contribution of this paper is in addressing the common concerns of targeting bias and reverse causality by suggesting a new instrument, replicable in many instances. We take advantage of Indonesia's institutional two-step budgeting setup for road funding, where different authorities, such as provinces or districts, are in charge of road quality and funding of different parts of the road network. This allows us to construct a time varying instrument for road quality, which equals total road funding at the provincial or district level. Thus, we identify the effects from the set of roads that get maintained when road budgets allow for it, but which get less maintenance when road budgets are tight or scaled back.

[27] For more on calculating the roughness-induced travel time measure and measures of market access, see Rothenberg (2013).

The evidence presented in this paper shows that road improvements alone can present an important stepping stone in economic development through opening up labor market opportunities and decreasing the income gap at the same time. On the flip side, deterioration of roads may have adverse affects in the opposite direction and may bring about important and unanticipated welfare effects that governments should be aware of when cutting transportation budgets.

Table 3.1: Road Quality Accumulation

Dep. Var.: Road Segment $\log \mathrm{IRI}_{\mathbf{t}}$	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
$\log \mathrm{IRI}_{t-1}$	0.831	0.852	0.857	0.616	0.698	0.708
	(0.001)***	(0.004)***	(0.006)***	(0.001)***	(0.005)***	(0.006)***
Segment Upgraded$_{t-1} = 1$	-0.484	-0.492	-0.492	-0.450	-0.449	-0.449
	(0.001)***	(0.001)***	(0.001)***	(0.001)***	(0.001)***	(0.001)***
$\log \mathrm{GDP}_{t-1}$			0.005			0.008
			(0.001)***			(0.002)***
$\log \mathrm{GDP}_{t-1} \times \mathrm{IRI}_{t-1}$		-0.002	-0.003		-0.006	-0.006
		(0.000)***	(0.000)***		(0.000)***	(0.000)***
$\log \#\mathrm{Firms}_{t-1}$			-0.007			0.010
			(0.001)***			(0.002)***
$\log \#\mathrm{Firms}_{t-1} \times \mathrm{IRI}_{t-1}$		-0.001	0.002		-0.002	-0.005
		(0.000)***	(0.001)***		(0.001)***	(0.001)***
N	972714	602872	602872	972714	602872	602872
Adjusted R^2	0.811	0.797	0.797	0.513	0.499	0.499
F Statistic	256469.40	148866.70	136271.68	31242.55	19506.45	17776.15
Road FE	No	No	No	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

GDP and # firms are measured at the district level. Robust standard errors in parentheses, clustered at the road segment level. */**/*** denotes significant at the 10% / 5% / 1% levels.

Table 3.2: Road Upgrading

	All Roads	**National Roads**		**Provincial Roads**		**District Roads**	
Dep. Var.: Segment Upgrade$_t$ (0 1)	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
$\log \text{IRI}_{t-2}$	0.017	-0.014	-0.017	-0.005	-0.009	0.023	0.022
	(0.001)***	(0.002)***	(0.002)***	(0.001)***	(0.001)***	(0.002)***	(0.003)***
Years Since Last Upgrade	-0.335	-0.436	-0.478	-0.382	-0.427	-0.258	-0.291
	(0.001)***	(0.006)***	(0.006)***	(0.004)***	(0.006)***	(0.004)***	(0.005)***
Years Since Last Upgrade^2	0.026	0.039	0.046	0.033	0.040	0.018	0.022
	(0.000)***	(0.001)***	(0.001)***	(0.001)***	(0.001)***	(0.000)***	(0.001)***
$\log \text{GDP}_{t-1}$		-0.005		-0.002		0.001	
		(0.001)***		(0.001)**		(0.001)	
$\log \#\text{Firms}_{t-1}$			-0.002		-0.003		0.001
			(0.000)***		(0.000)***		(0.001)
N	774719	237220	216508	265443	238341	27981	25165
Adjusted R^2	0.627	0.686	0.708	0.673	0.693	0.635	0.657
F Statistic	45361.69	20743.20	21132.55	28403.58	32178.52	5284.67	5374.47
Road FE	No	No	No	No	No	No	No
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Robust standard errors in parentheses, clustered at the road segment level. */**/*** denotes significant at the 10% / 5% / 1% levels.

Table 3.3: Road Quality and Budgets

Dep. Var.: Road segment log IRI(t)	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Budget (National Roads)	-0.044			-0.047		
	(0.001)***			(0.001)***		
Budget (Provincial Roads)		-0.037			-0.026	
		(0.000)***			(0.000)***	
Budget (District Roads)			-0.015			0.006
			(0.001)***			(0.001)***
$\log \mathrm{IRI}_{t-1}$				0.415	0.411	0.294
				(0.002)***	(0.002)***	(0.005)***
Years Since Last Upgrade				0.164	0.165	0.137
				(0.001)***	(0.001)***	(0.001)***
Years Since Last Upgrade^2				-0.012	-0.012	-0.009
				(0.000)***	(0.000)***	(0.000)***
N	955214	960837	306578	766335	769723	227814
Adjusted R^2	0.078	0.079	0.039	0.388	0.386	0.259
F Statistic	2520.469	2734.166	484.572	11979.469	12074.867	1777.209
Road FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

GDP and # firms at the district level. Robust standard errors in parentheses, clustered at the road segment level. */**/*** denotes significant at the 10% / 5% / 1% levels.

Table 3.4: Road Budgets and Local Economic Conditions

Dep. Var.: $\log \text{Budget}_t$	**Province Panel**				**District Panel**	
	Nat. roads **(1)**	Nat. roads **(2)**	Prov. roads **(3)**	Prov. roads **(4)**	District roads **(5)**	District roads **(6)**
$\log \text{GDP}_{t-2}$	-0.080		-0.584		-0.087	
	(0.185)		(0.506)		(0.118)	
$\log \#\text{Firms}_{t-2}$		-0.080		-0.291		-0.086
		(0.124)		(0.284)		(0.065)
N	241	256	249	265	545	568
Adjusted R^2	0.319	0.310	0.389	0.363	0.091	0.097
Road FE	No	No	No	No	No	No
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

Robust standard errors in parentheses, clustered at the road segment level. */**/*** denotes significant at the 10% / 5% / 1% levels.

Table 3.5: Summary statistics - IFLS and SI

Variable	Obs	Mean (SD)	Min	Max
Individual Level				
Age	57252	35.543 (14.31)	16	100
Female (0 1)	57218	0.51 (0.50)	0	1
Years of Schooling	57252	7.941 (3.823)	1	16
Working (0 1)	41893	0.70 (0.46)	0	1
Working in agriculture (0 1)	29342	0.38 (0.49)	0	1
Working in manufacturing (0 1)	29342	0.30 (0.46)	0	1
Hours worked last week	29221	41.75 (22.07)	0	105
Working (all but agri and manu)	29342	0.44 (0.50)	0	1
Monthly earnings	24023	141.87 (191.86)	-1.13e+04	2000.000
Household Level				
Household size	19906	4.38 (2.06)	1	20
Number of rooms	19906	5.34 (2.23)	0	30
Own the house (0 1)	19904	0.84 (0.37)	0	1
Electricity (0 1)	19906	0.85 (0.35)	0	1
Piped water (0 1)	19906	0.19 (0.39)	0	1
Dirt floor (0 1)	19906	0.18 (0.38)	0	1
Tiled floor (0 1)	19906	0.23 (0.42)	0	1
Cememt floor (0 1)	19906	0.34 (0.47)	0	1
Owns Toilet (0 1)	19906	0.60 (0.49)	0	1
Road quality data (district)				
Road Roughness	292100	4.8 (1.49)	1.69	16.54

Table 3.6: Reduced Form Effects of Road Roughness: Consumption and Total Earnings

Panel A: Dep Var: Log Consumption (Per Capita)	**(1)**
Log Avg IRI	-0.261
	(0.061)***
Adj. R^2	0.594
Regression F-Stat	191.771
Kleibergen-Paap F-Stat	35.450
DV Mean	11.046
N	22541
N Households	6928
N Desa / Kelurahan	852
Panel B: Dep Var: Log Total Earnings	**(1)**
Log Avg IRI	-0.894
	(0.328)***
Adj. R^2	0.265
Regression F-Stat	10.118
Kleibergen-Paap F-Stat	33.684
DV Mean	10.633
N	16935
N Individuals	6460
N Desa / Kelurahan	298
Household FE	Yes
Year FE	Yes
Controls	Yes

All regressions include a constant. Controls include: district GDP, month of survey indicators. Robust standard errors in parentheses, clustered at the (initial) village level. For the GMM specifications, we report the adjusted R^2 from the analagous reduced-form regression (regressing the dependent variable on the instruments). */**/*** denotes significant at the 10% / 5% / 1% levels.

Table 3.7: Road Roughness and District-Level In-Migration of Workers and Entry of Firms

Panel A: Dep Var: Log Share of Recent Migrants (2000 Census)	**(1)**	**(2)**
Log Δ IRI (2000-1995)	-1.257	-0.690
	(0.153)***	(0.225)***
N	201	201
Adjusted R^2	0.366	0.474
F Statistic	38.992	8.436
Kleibergen-Paap F-Stat	14.779	11.570
DV Mean	-3.156	-3.156
Province FE	.	Yes

Panel B: Firm Outcomes	**DV: Log # of Firms**	**DV: Log # of Workers**
	(1)	**(2)**
Log Avg IRI	-1.194	-1.294
	(0.182)***	(0.286)***
N	3121	3121
N Kabupaten / Kota	204	204
Adj. R^2	0.910	0.914
Regression F-Stat	110.553	241.978
Kleibergen-Paap F-Stat	47.239	47.239
DV Mean	2.890	6.660
Kabupaten / Kota FE	Yes	Yes
Year FE	Yes	Yes

All regressions include a constant. Panel A reports the results of cross-sectional regressions of the share of recent migrants (last five years) in 2000 on changes in road roughness. In Panel A, the unit of analysis is the kabupaten (district), and controls include logs of 1990 population and 1990 non-oil GDRP. Robust standard errors are reported in parentheses. In Panel B, we report the results of panel regressions of road roughness on the log of one plus the number of firms and workers, with data from the SI. Controls include logs of population and non-oil GDRP. Robust standard errors, clustered at the kabupaten level, are reported in parentheses. For the IV specifications, we report the adjusted R^2 from the analagous reduced-form regression (regressing the dependent variable on the instruments). */**/*** denotes significant at the 10% / 5% / 1% levels.

Table 3.8: Road Roughness and Firm-Level Outcomes

Panel A: Dep Var: Log Number of Workers	**(1)**	**(2)**
Log Avg IRI	-0.057	0.003
	(0.017)***	(0.012)
N	234272	230224
N Firms	29366	29366
Adjusted R^2	0.266	0.924
Regression F Statistic	11.563	0.069
Kleibergen-Paap F-Stat	2034.234	1821.897
Panel B: Dep Var: Log Value Added	**(1)**	**(2)**
Log Avg IRI	-0.138	-0.074
	(0.036)***	(0.031)**
N	234134	230096
N Firms	29362	29362
Adjusted R^2	0.467	0.859
Regression F Statistic	15.048	5.624
Kleibergen-Paap F-Stat	2032.975	1821.119
Panel C: Dep Var: Log TFP	**(1)**	**(2)**
Log Avg IRI	-0.099	-0.134
	(0.040)**	(0.037)***
N	126629	122859
N Firms	19073	19073
Adjusted R^2	0.480	0.725
Regression F Statistic	6.255	13.041
Kleibergen-Paap F-Stat	1190.051	1010.144
Year FE	Yes	Yes
Industry FE	Yes	Yes
Kabupaten / Kota FE	Yes	Yes
Firm FE	.	Yes

All regressions include a constant. Robust standard errors in parentheses, clustered at the firm level. For the GMM specifications, we report the adjusted R^2 from the analagous reduced-form regression (regressing the dependent variable on the instruments). */**/*** denotes significant at the 10% / 5% / 1% levels.

Table 3.9: Road Roughness and Labor Supply: Extensive and Intensive Margins

Panel A: Dep Var: Working? (0 1)	**(1)**
Log Avg IRI	-0.037
	(0.034)
Adj. R^2	0.427
Regression F-Stat	34.111
Kleibergen-Paap Wald rk F-Stat	33.594
DV Mean	0.704
N	35193
N Individuals	12459
N Desa / Kelurahan	377

Panel B: Dep Var: Log Hours Worked (Conditional on working)	**(1)**
Log Avg IRI	-0.011
	(0.060)
Adj. R^2	0.237
Regression F-Stat	3.461
Kleibergen-Paap Wald rk F-Stat	32.698
DV Mean	5.148
N	22255
N Individuals	8168
N Desa / Kelurahan	321
Individual FE	Yes
Year FE	Yes
Controls	Yes

All regressions include a constant. Controls include: district GDP, individual age, education, household size, and month of survey indicators. In Panel B, the dependent variable is only defined if the individual reported working. Robust standard errors in parentheses, clustered at the (initial) village level. For the GMM specifications, we report the adjusted R^2 from the analagous reduced-form regression (regressing the dependent variable on the instruments). */**/*** denotes significant at the 10% / 5% / 1% levels.

Table 3.10: Road Roughness and Sector Switching

Dependent Variable:	**Any Employment?**	**Primary Sector Employment?**	**Log Hours Worked**
Panel A: By Industry	**(1)**	**(2)**	**(3)**
Manufacturing	-0.114	-0.129	-6.240
	(0.036)***	(0.033)***	(1.802)***
Adj. R^2	0.462	0.493	0.395
Regression F-Stat	8.807	10.079	8.151
Kleibergen-Paap F-Stat	32.711	32.711	32.910
DV Mean	0.285	0.253	12.219
Agriculture	0.055	0.055	2.889
	(0.033)*	(0.034)	(2.494)
Adj. R^2	0.648	0.653	0.399
Regression F-Stat	6.559	4.923	2.887
Kleibergen-Paap F-Stat	32.711	32.711	32.695
DV Mean	0.419	0.367	13.013
Sales and Services	0.054	0.082	4.523
	(0.035)	(0.035)**	(1.963)**
Adj. R^2	0.498	0.533	0.443
Regression F-Stat	8.511	6.062	7.049
Kleibergen-Paap F-Stat	32.711	32.711	32.930
DV Mean	0.316	0.279	13.688
Other	-0.026	-0.015	0.026
	(0.018)	(0.011)	(0.802)
Adj. R^2	0.608	0.521	0.474
Regression F-Stat	1.903	3.954	2.160
Kleibergen-Paap F-Stat	32.711	33.594	32.771
DV Mean	0.105	0.067	3.646
Panel B: Formal vs. Informal	**(1)**	**(2)**	**(3)**
Informal	0.079	0.091	6.074
	(0.033)**	(0.032)***	(2.592)**
Adj. R^2	0.560	0.598	0.366
Regression F-Stat	11.468	6.857	7.955
Kleibergen-Paap F-Stat	32.698	32.707	33.050
DV Mean	0.636	0.576	23.298
Individual FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes

All regressions include a constant. Controls include: district GDP, age, education, household size, month of survey indicators. Dependent variables are only defined if the individual reported working. Robust standard errors in parentheses, clustered at the (initial) village level. For the GMM specifications, we report the adjusted R^2 from the analagous reduced-form regression (regressing the dependent variable on the instruments). */**/*** denotes significant at the 10% / 5% / 1% levels.

Table 3.11: Road Roughness and Sector Switching: Heterogeneity

		Age		**Gender**		**Education**		**Initial PCE**		**Own Farmland?**	
Panel A: Dep Var: Working in Agriculture (Primary) (0 1)	**All Obs (1)**	**< Med (2)**	**≥ Med (3)**	**F (4)**	**M (5)**	**< Med (6)**	**≥ Med (7)**	**< Med (8)**	**≥ Med (9)**	**No (10)**	**Yes (11)**
Log Avg IRI	0.055	0.034	0.074	0.034	0.082	0.116	-0.002	0.100	-0.008	0.118	0.079
	(0.034)	(0.082)	(0.034)**	(0.042)	(0.043)*	(0.047)**	(0.038)	(0.046)**	(0.038)	(0.036)***	(0.054)
Adj. R^2	0.664	0.579	0.678	0.708	0.640	0.646	0.627	0.650	0.630	0.633	0.622
Regression F-Stat	4.923	2.786	4.627	2.276	4.244	4.952	1.668	4.443	3.226	3.042	4.107
Kleibergen-Paap Wald rk F-Stat	32.711	19.528	33.321	31.895	31.717	26.469	30.785	27.690	29.354	26.561	26.547
DV Mean	0.367	0.314	0.382	0.328	0.392	0.491	0.212	0.494	0.226	0.197	0.529
		Age		**Gender**		**Education**		**Initial PCE**		**Own Farmland?**	
Panel B: Dep Var: Working in Manufacturing (Primary) (0 1)	**All Obs (1)**	**< Med (2)**	**≥ Med (3)**	**F (4)**	**M (5)**	**< Med (6)**	**≥ Med (7)**	**< Med (8)**	**≥ Med (9)**	**No (10)**	**Yes (11)**
Log Avg IRI	-0.129	-0.129	-0.134	-0.117	-0.138	-0.207	-0.073	-0.197	-0.071	-0.145	-0.157
	(0.033)***	(0.080)	(0.035)***	(0.044)***	(0.041)***	(0.044)***	(0.041)*	(0.046)***	(0.042)*	(0.054)***	(0.040)***
Adj. R^2	0.490	0.392	0.493	0.543	0.461	0.505	0.468	0.528	0.471	0.484	0.479
Regression F-Stat	10.079	2.862	10.382	3.439	8.778	10.285	4.021	7.270	4.635	9.754	5.366
Kleibergen-Paap Wald rk F-Stat	32.711	19.528	33.321	31.895	31.717	26.469	30.785	27.690	29.354	26.561	26.547
DV Mean	0.253	0.308	0.237	0.185	0.298	0.241	0.268	0.248	0.258	0.324	0.185
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

All regressions include a constant. Controls include: district GDP, age, education, household size, month of survey indicators. Dependent variables are only defined if the individual reported working. Robust standard errors in parentheses, clustered at the (initial) village level. For the GMM specifications, we report the adjusted R^2 from the analagous reduced-form regression (regressing the dependent variable on the instruments). */**/*** denotes significant at the 10% / 5% / 1% levels.

Table 3.12: Road Roughness and Total Earnings by Sector

	Dep Var: Earnings by Sector			
	Manu-facturing (1)	**Agri-culture (2)**	**Sales and Services (3)**	**Other (4)**
Log Avg IRI	-0.533	-3.474	-0.160	-0.132
	(0.363)	(1.224)***	(0.404)	(0.201)
Adj. R^2	0.282	0.125	0.197	0.066
Regression F-Stat	1.818	2.181	2.182	1.940
Kleibergen-Paap F-Stat	23.238	19.200	24.000	32.073
DV Mean	11.177	7.806	11.024	0.135
N	4160	5464	3692	10763
N Individuals	1722	2130	1540	4235
N Desa / Kelurahan	227	189	231	260
Individual FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes

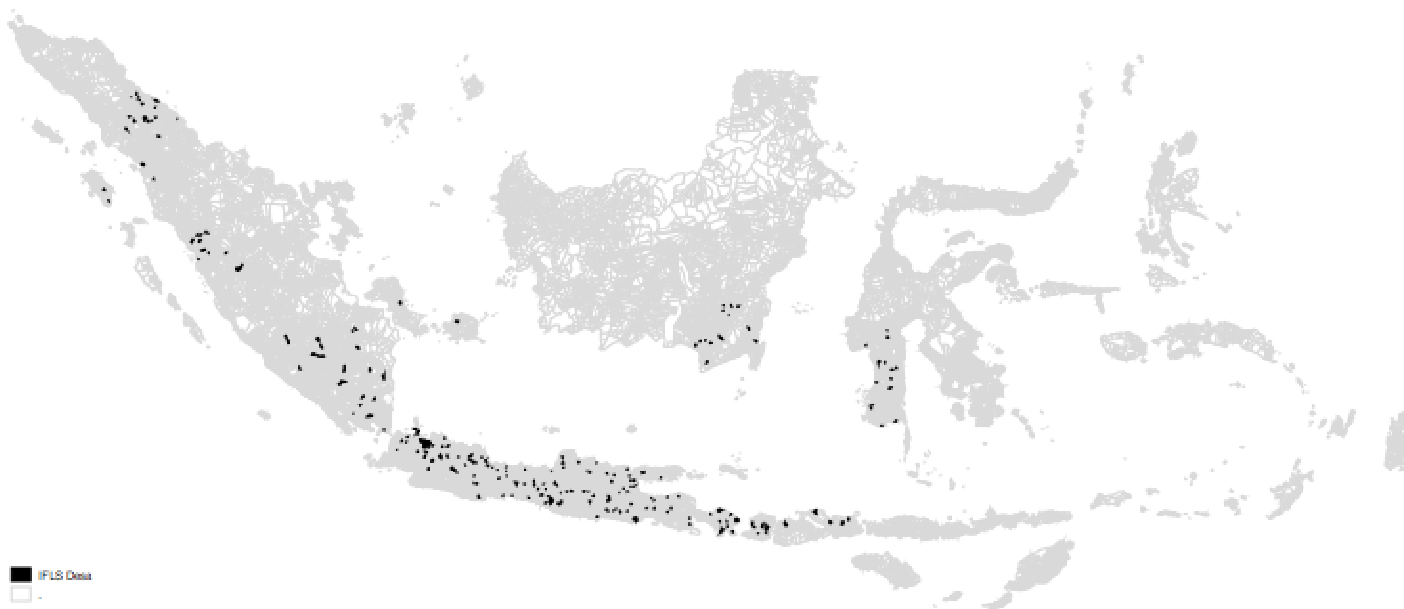
All regressions include a constant. Controls include: district GDP, age, education, household size, month of survey indicators, and hours worked. Robust standard errors in parentheses, clustered at the (initial) village level. For the GMM specifications, we report the adjusted R^2 from the analagous reduced-form regression (regressing the dependent variable on the instruments). */**/*** denotes significant at the 10% / 5% / 1% levels.

Table 3.13: Road Roughness and Land Prices

	DV: Log Rent per Room (1)	**DV: Log Land Value (2)**
Log Avg IRI	-0.244	-0.658
	(0.116)**	(0.217)***
Adj. R^2	0.360	0.488
Regression F-Stat	25.589	12.117
Kleibergen-Paap F-Stat	36.627	19.970
DV Mean	8.276	14.819
N	19115	7441
N Households	6475	2498
N Desa / Kelurahan	793	321
Household FE	Yes	Yes
Year FE	Yes	Yes
Controls	Yes	Yes

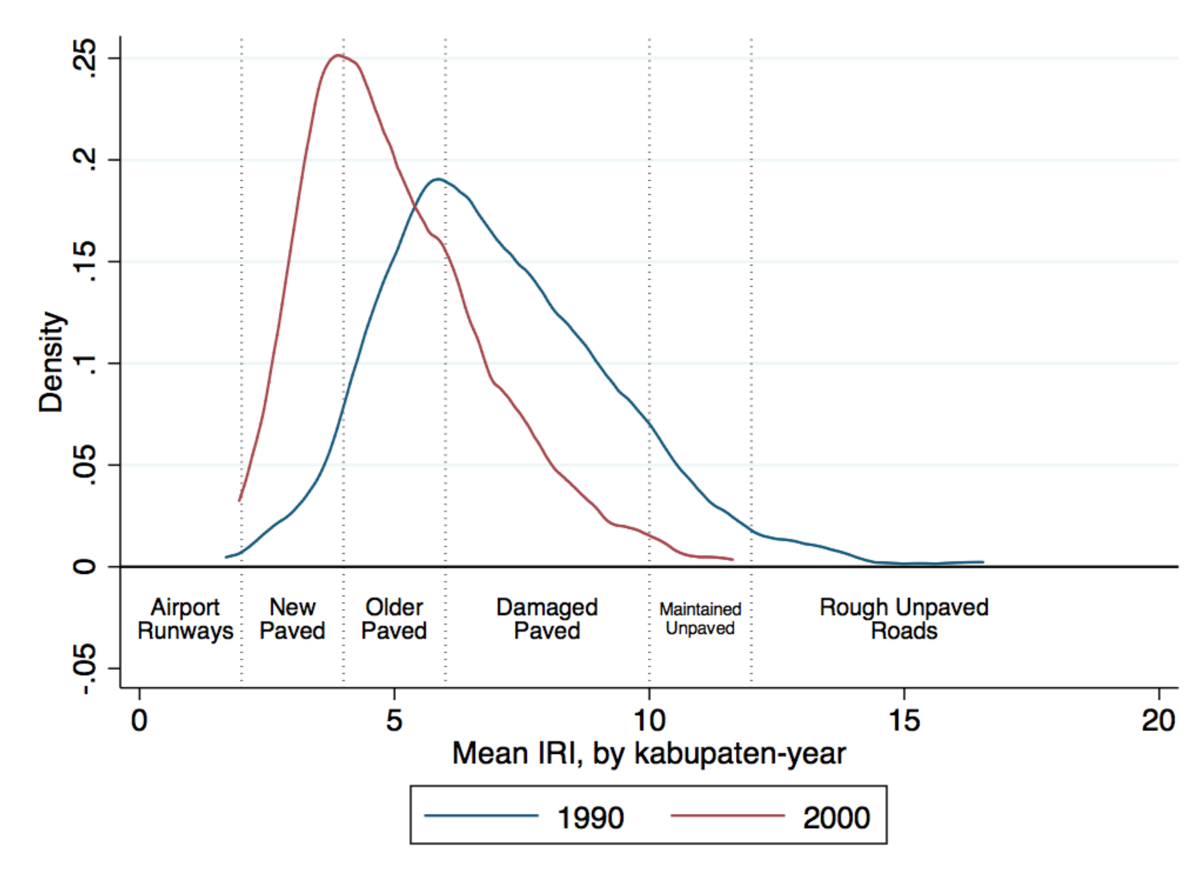
All regressions include a constant. Controls include: district GDP, month of survey indicators, number of rooms, indicators for electricity, piped water, internal toilets, types of floor (tiled, cement, dirt), mason walls, and tiled roof. Robust standard errors in parentheses, clustered at the (initial) village level. For the GMM specifications, we report the adjusted R^2 from the analogous reduced-form regression (regressing the dependent variable on the instruments). */**/*** denotes significant at the 10% / 5% / 1% levels.

Figure 3.1: IFLS Villages



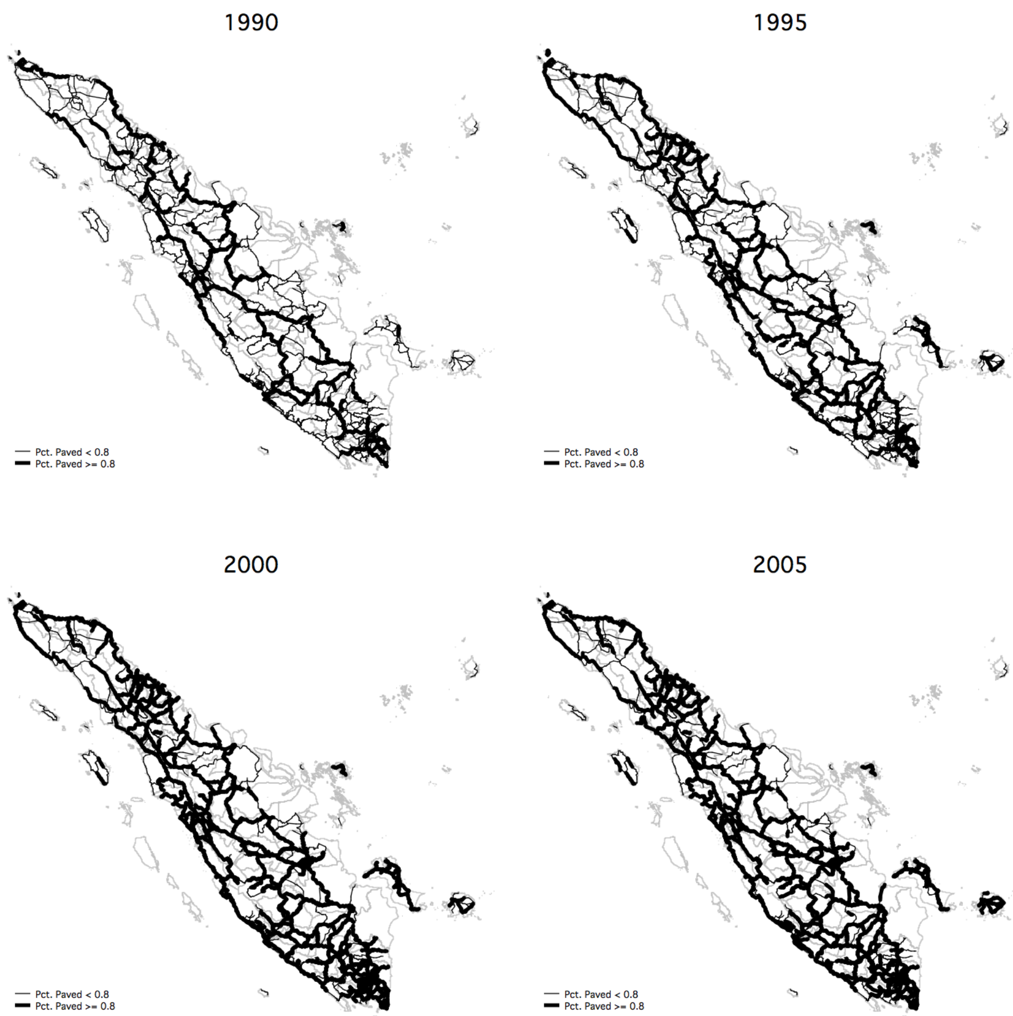
Note: Authors' calculations.

Figure 3.2: Changes in the Distribution of Road Roughness



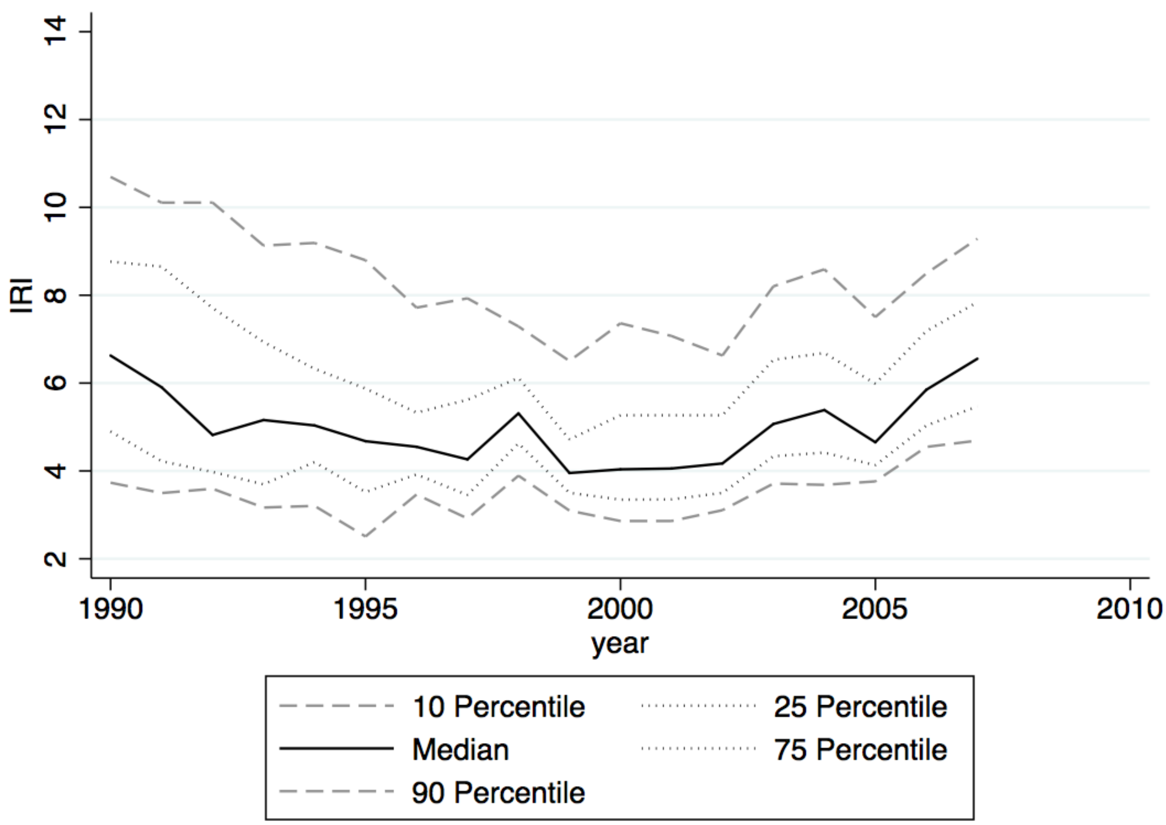
Note: Authors' calculations.

Figure 3.3: Road Roughness - Sumatra



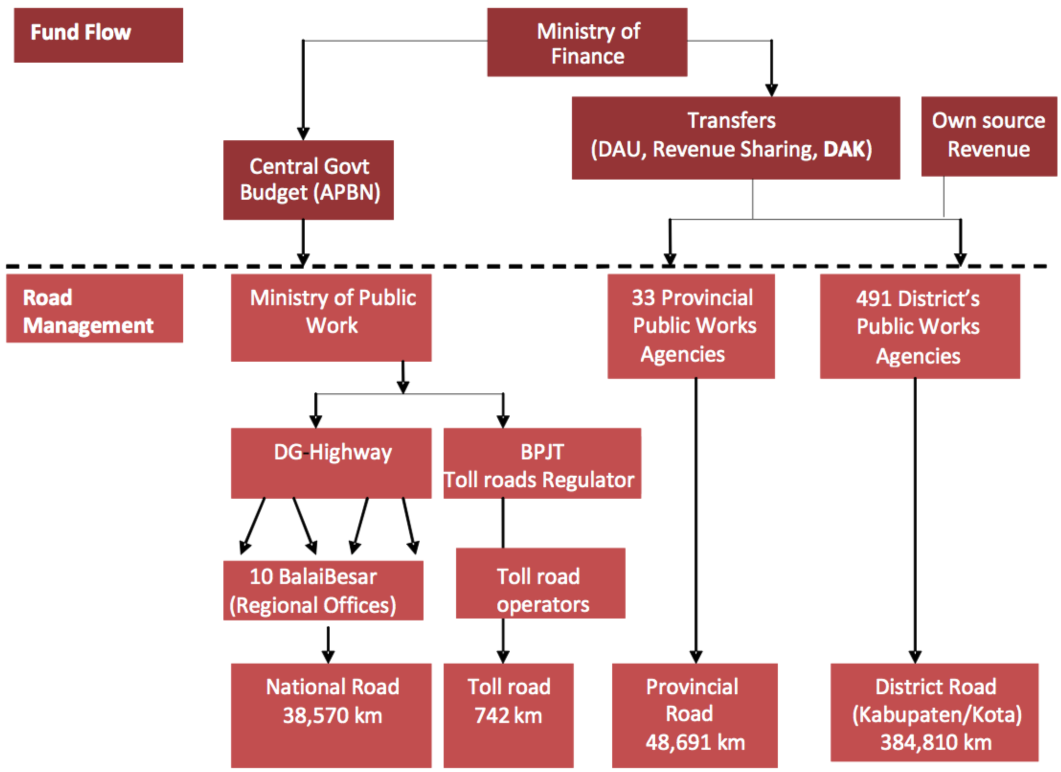
Note: Authors' calculations.

Figure 3.4: Changes in Roughness-Based Travel Time



Note: Authors' calculations.

Figure 3.5: Institutional Arrangements for the Road Sector in Indonesia



Source: (World Bank, 2012).

Figure 3.6: Allocation Criteria for District Road Improvement Grant

1989/90	1990/91	1991/92	1992/93
CRITERIA	CRITERIA	CRITERIA	CRITERIA
1. Length of road (20%) 2. Unstable & critical road (52%) 3. Area of irrigation (15%) 4. Increase of actual regional own-source receipts (PAD)(7%) 5. Actual own-source revenues compared to planned(6%) 6. Unit price correction (Ditjen Cipta Karya)	1. Length of road (15%) 2. Road condition (60%) 3. GRDP (15%) 4. Road density (10%) 5. Unit price correction	1. Length of road 2. % of good road a. Kab $<$ 28.3% good road b. 55% $>$ Kab $>$ 28.3% good road 3. Road density a. Kab $<$ 28km/1000km2 b. 100km/1000km2 $>$ Kab $>$ 28km/1000km2 4. Performance Kabupaten needs according to (a) and (b) greater than 60km, take 60km 5. Unit price correction	1. Length of road 2. % of good road a. Kab $<$ 17.68% good road b. 55% $>$ Kab $>$ 17.68% 3. Road density a. Kab $<$ 40.29km/1000km2 b. 100km/1000km2 kab $>$ 40.29km/1000km2 4. Performance Kabupaten needs according to (a) and (b) greater than 60km, take 60 km 5. Unit price correction

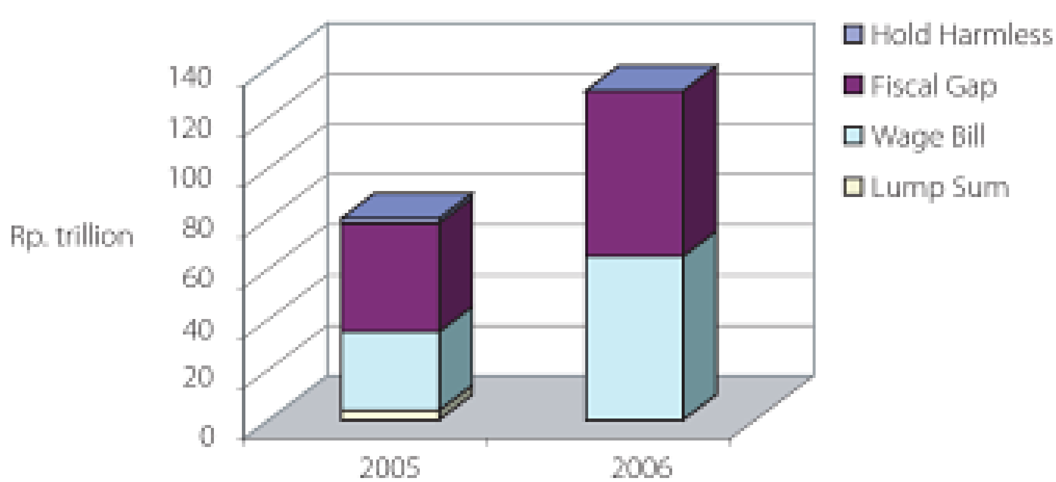
Note: (Bird and Smart, 2001).

Figure 3.7: The evolution of technical criteria in the DAK formula for roads and their respective weights

No.	Technical criteria	Description	2008	2009	2010	2011
1	Length of road	Length of road which is legally acknowledge through the decree of the head of local government	30%	25%	25%	25%
2	Road condition	Length of road with non-stable condition	30%	40%	35%	25%
3	Good road performance		20%			
4	Accessibility	Defined by the length of road divided by total area			20%	10%
5	Mobility	Length of road per 1000 population in the province/kabupaten			20%	10%
6	Ownership/concern by LG	Determined by the percentage of original APBD allocated to the road sector		20%		10%
7	Reporting	Consistency in submitting of quarterly report, physical progress, financial progress	20%	15%		20%

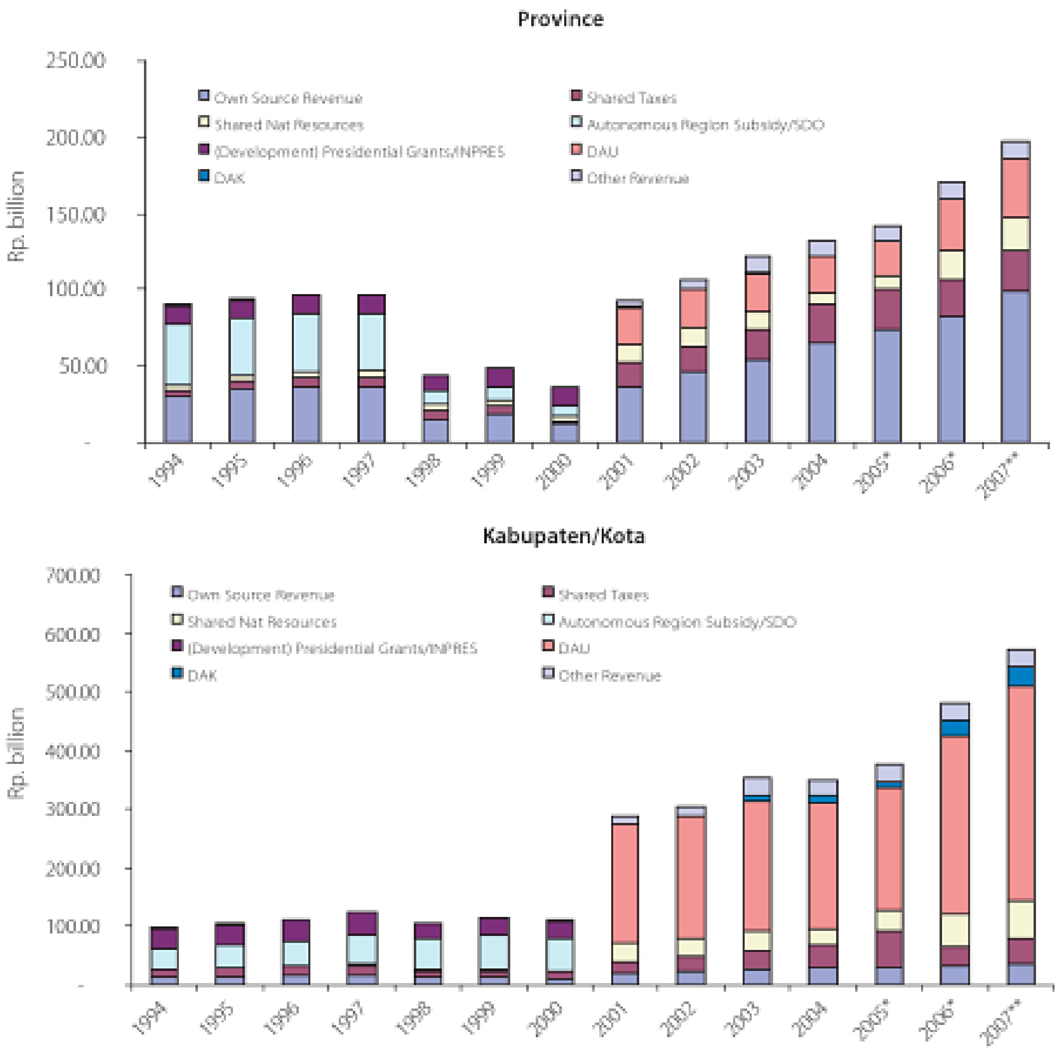
Note: (World Bank, 2012).

Figure 3.8: Changes in DAU composition over time



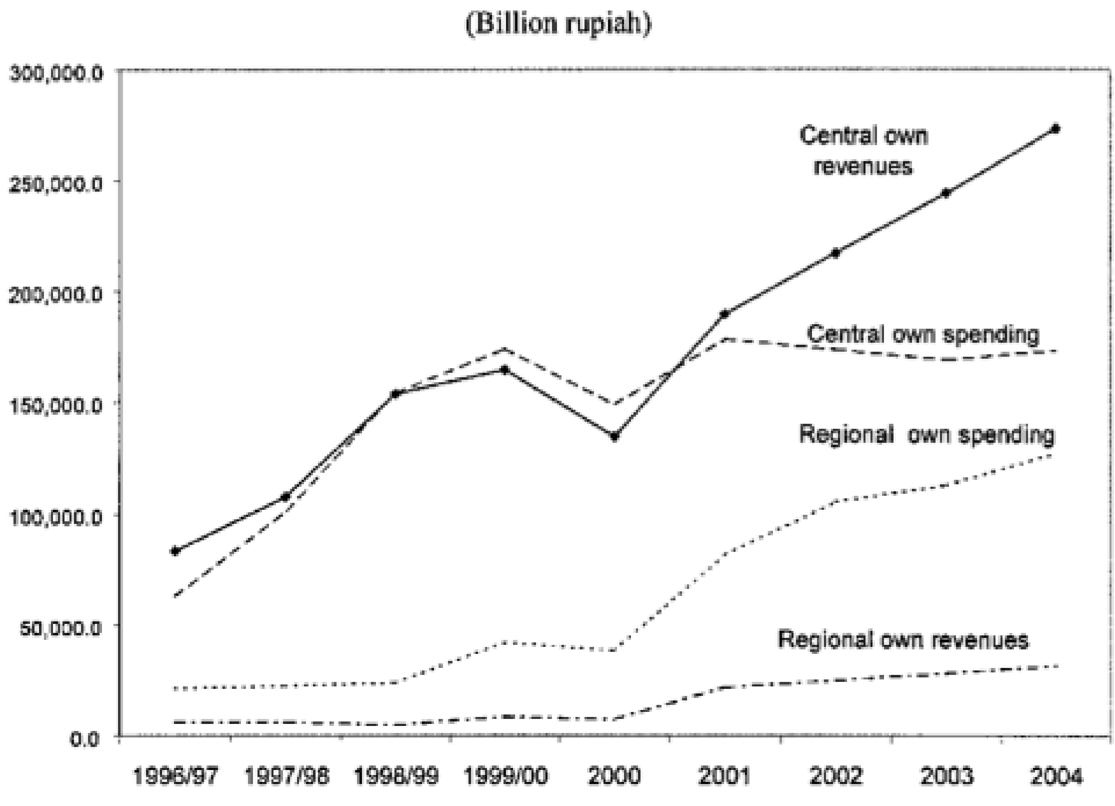
Note: World Bank staff calculation. (World Bank, 2008).

Figure 3.9: Sub-national revenue over time



Note: World Bank staff calculation. (World Bank, 2008).

Figure 3.10: Impact of Decentralization in Indonesia



Note: In billion rupiah. (Ahmad and Mansoor, 2002).

Bibliography

ABALUCK, J. (2011): "What Would We Eat if We Knew More: The Implications of a Large-Scale Change in Nutrition Labeling," *Work. Pap., Mass. Inst. Technol.*

ADDA, J. AND F. CORNAGLIA (2006): "Taxes, cigarette consumption, and smoking intensity," *The American Economic Review*, 96, 1013–1028.

AHMAD, E. AND A. M. MANSOOR (2002): *Indonesia: managing decentralization*, International Monetary Fund.

ALBERTI, K., P. ZIMMET, AND J. SHAW (2006): "Metabolic syndromea new world-wide definition. A consensus statement from the international diabetes federation," *Diabetic Medicine*, 23, 469–480.

ANDERSON, J. E. AND E. VAN WINCOOP (2004): "Trade Costs," *Journal of Economic Literature*, 42, 691–751.

ANDERSON, M. L. AND D. A. MATSA (2011): "Are restaurants really supersizing America?" *American Economic Journal: Applied Economics*, 152–188.

ASIA FOUNDATION (2008): "The Cost of Moving Goods: Road Transportation, Regulations and Charges in Indonesia," Survey report.

ATKIN, D. (2010): "Trade, tastes and nutrition in India," *Yale University Economic Growth Center Discussion Paper.*

——— (2013): "The caloric costs of culture: evidence from Indian migrants," Tech. rep., National Bureau of Economic Research.

ATKIN, D., B. FABER, AND M. GONZALEZ-NAVARRO (2014): "Retail Globalization and Household Welfare: Evidence from Mexico," *Working Paper.*

ATKINS, R. D. (2002): *Dr. Atkins' new diet revolution*, Government Institutes.

AULD, M. C. AND L. M. POWELL (2009): "Economics of food energy density and adolescent body weight," *Economica*, 76, 719–740.

Barcelo, A., C. Aedo, S. Rajpathak, and S. Robles (2003): "The cost of diabetes in Latin America and the Caribbean," *Bulletin of the World Health Organization*, 81, 19–27.

Bardhan, P. and C. Udry (1999): *Development Microeconomics*, Oxford: Oxford University Press.

Barquera, S., I. Campos-Nonato, C. Aguilar-Salinas, R. Lopez-Ridaura, A. Arredondo, and J. Rivera-Dommarco (2013): "Diabetes in Mexico: cost and management of diabetes and its complications and challenges for health policy," *Globalization and health*, 9.

Basker, E. (2007): "The causes and consequences of Wal-Mart's growth," *The Journal of Economic Perspectives*, 177–198.

Basker, E. and M. Noel (2009): "The Evolving Food Chain: Competitive Effects of Wal-Mart's Entry into the Supermarket Industry," *Journal of Economics & Management Strategy*, 18, 977–1009.

Basu, S., P. Yoffe, N. Hills, and R. H. Lustig (2013): "The relationship of sugar to population-level diabetes prevalence: an econometric analysis of repeated cross-sectional data," *PLoS One*, 8, e57873.

Baum II, C. L. and C. J. Ruhm (2009): "Age, socioeconomic status and obesity growth," *Journal of health economics*, 28, 635–648.

Bazzi, S. (2013): "Wealth Heterogeneity, Income Shocks, and International Migration: Theory and Evidence from Indonesia," Working Paper.

Benjamin, D. (1992): "Household Composition, Labor markets, and Labor demand: Testing for Separation in Agricultural Household Models," *Econometrica*, 60, 287–322.

Berardi, N., P. Sevestre, M. Tepaut, and A. Vigneron (2012): "The impact of a'soda tax'on prices: evidence from French micro data," .

Besley, T. J. and H. S. Rosen (1999): "Sales Taxes and Prices: An Empirical Analysis," *National Tax Journal*, 52, 157–178.

Beydoun, M. A., L. M. Powell, and Y. Wang (2008): "The association of fast food, fruit and vegetable prices with dietary intakes among US adults: is there modification by family income?" *Social science & medicine*, 66, 2218–2229.

Bird, R. M. and M. Smart (2001): "Intergovernmental fiscal transfers: Some lessons from international experience," in *March (revision). Paper prepared for Symposium on Intergovernmental Transfers in Asian Countries: Issues and Practices, Asian Tax and Public Policy Program, Hitotsubashi University.*

Blalock, G. and P. J. Gertler (2008): "Welfare Gains from Foreign Direct Investment Through Technology Transfer to Local Suppliers," *Journal of International Economics*, 74, 402–421.

Blomquist, G. C., M. C. Berger, and J. P. Hoehn (1988): "New Estimates of Quality of Life in Urban Areas," *American Economic Review*, 78.

Bollinger, B., P. Leslie, and A. Sorensen (2010): "Calorie posting in chain restaurants," Tech. rep., National Bureau of Economic Research.

Borghans, L. and B. H. Golsteyn (2006): "Time discounting and the body mass index: Evidence from the Netherlands," *Economics & Human Biology*, 4, 39–61.

Bray, G. A., S. J. Nielsen, and B. M. Popkin (2004): "Consumption of high-fructose corn syrup in beverages may play a role in the epidemic of obesity," *The American journal of clinical nutrition*, 79, 537–543.

Bray, G. A. and B. M. Popkin (1998): "Dietary fat intake does affect obesity!" *The American journal of clinical nutrition*, 68, 1157–1173.

Bremer, A. A., K. L. Stanhope, J. L. Graham, B. P. Cummings, W. Wang, B. R. Saville, and P. J. Havel (2011): "Fructose-Fed Rhesus Monkeys: A Nonhuman Primate Model of Insulin Resistance, Metabolic Syndrome, and Type 2 Diabetes," *Clinical and translational science*, 4, 243–252.

Bulow, J. I. and P. Pfleiderer (1983): "A note on the effect of cost changes on prices," *The Journal of Political Economy*, 182–185.

Case, A. and A. Menendez (2009): "Sex differences in obesity rates in poor countries: evidence from South Africa," *Economics & Human Biology*, 7, 271–282.

Catenacci, V. A., J. O. Hill, and H. R. Wyatt (2009): "The obesity epidemic," *Clinics in chest medicine*, 30, 415–444.

Cawley, J. and C. Meyerhoefer (2012): "The medical care costs of obesity: an instrumental variables approach," *Journal of health economics*, 31, 219–230.

Chabris, C. F., D. Laibson, C. L. Morris, J. P. Schuldt, and D. Taubinsky (2008): "Individual laboratory-measured discount rates predict field behavior," *Journal of Risk and Uncertainty*, 37, 237–269.

Chen, H.-S., T.-S. Jap, R.-L. Chen, and H.-D. Lin (2004): "A prospective study of glycemic control during holiday time in type 2 diabetic patients," *Diabetes Care*, 27, 326–330.

Chou, S.-Y., M. Grossman, and H. Saffer (2004): "An economic analysis of adult obesity: results from the Behavioral Risk Factor Surveillance System," *Journal of health economics*, 23, 565–587.

Chou, S.-Y., I. Rashad, and M. Grossman (2005): "Fast-food restaurant advertising on television and its influence on childhood obesity," Tech. rep., National Bureau of Economic Research.

Clark, S. E., C. Hawkes, S. M. Murphy, K. A. Hansen-Kuhn, and D. Wallinga (2012): "Exporting obesity: US farm and trade policy and the transformation of the Mexican consumer food environment," *International journal of occupational and environmental health*, 18, 53–64.

Cortes, P. (2009): "The Mexico Market for Soft Drinks," *USDA Foreign Agricultural Service*.

Courtemanche, C., G. Heutel, and P. McAlvanah (2014): "Impatience, incentives, and obesity," *The Economic Journal*.

Cremer, H. and J.-F. Thisse (1994): "Commodity taxation in a differentiated oligopoly," *International Economic Review*, 613–633.

Crescenzo, R., F. Bianco, P. Coppola, A. Mazzoli, M. Tussellino, R. Carotenuto, G. Liverini, and S. Iossa (2014): "Fructose supplementation worsens the deleterious effects of short-term high-fat feeding on hepatic steatosis and lipid metabolism in adult rats," *Experimental physiology*.

Currie, J., S. DellaVigna, E. Moretti, and V. Pathania (2009): "The effect of fast food restaurants on obesity and weight gain," Tech. rep., National Bureau of Economic Research.

Cutler, D., E. Glaeser, and J. Shapiro (2003): "Why have Americans become more obese?" Tech. rep., National Bureau of Economic Research.

Dasgupta, P. (1997): "Nutritional status, the capacity for work, and poverty traps," *Journal of Econometrics*, 77, 5–37.

Datar, A., R. Sturm, and J. L. Magnabosco (2004): "Childhood Overweight and Academic Performance: National Study of Kindergartners and First-Graders," *Obesity research*, 12, 58–68.

de Salud, S. (2010): "Acuerdo Nacional para la Salud Alimentaria: Estrategia contra el sobrepeso y la obesidad," .

Deaton, A. and J. Muellbauer (1980): "An almost ideal demand system," *The American economic review*, 312–326.

Dell, M. (2012): "Path dependence in development: Evidence from the Mexican Revolution," *Unpublished manuscript, Department of Economics, Harvard University.*

Dickens, W. T. and K. Lang (1985): "A Test of Dual Labor Market Theory," *American Economic Review*, 75, 792–805.

Dijkstra, E. W. (1959): "A Note on Two Problems in Connexion with Graphs," *Numerische Mathematik*, 1, 269–271.

Donaldson, D. (2010): "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure," Working Paper.

Downs, J. S., G. Loewenstein, and J. Wisdom (2009): "Strategies for promoting healthier food choices," *The American Economic Review*, 159–164.

Drewnowski, A. (2009): "Obesity, diets, and social inequalities," *Nutrition reviews*, 67, S36–S39.

Drewnowski, A. and N. Darmon (2005): "The economics of obesity: dietary energy density and energy cost," *The American Journal of Clinical Nutrition*, 82, 265S–273S.

Drewnowski, A. and S. Specter (2004): "Poverty and obesity: the role of energy density and energy costs," *The American journal of clinical nutrition*, 79, 6–16.

Dubois, P., R. Griffith, and A. Nevo (2013): "Do Prices and Attributes Explain International Differences in Food Purchases?" Tech. rep., National Bureau of Economic Research.

Evans, J., R. W. Newton, D. A. Ruta, T. M. MacDonald, and A. D. Morris (2000): "Socio-economic status, obesity and prevalence of Type 1 and Type 2 diabetes mellitus," *Diabetic Medicine*, 17, 478–480.

Evans, W. N. and J. S. Ringel (1999): "Can higher cigarette taxes improve birth outcomes?" *Journal of Public Economics*, 72, 135–154.

Ezzati, M., S. Vander Hoorn, A. Rodgers, A. D. Lopez, C. D. Mathers, and C. J. Murray (2003): "Estimates of global and regional potentil health gains from reducing muliple major risk factors," *The Lancet*, 362, 271–280.

Fernald, L. C. (2007): "Socio-economic status and body mass index in low-income Mexican adults," *Social science & medicine*, 64, 2030–2042.

Fernald, L. C., P. J. Gertler, and X. Hou (2008): "Cash component of conditional cash transfer program is associated with higher body mass index and blood pressure in adults," *The Journal of nutrition*, 138, 2250–2257.

Fields, G. S. (2004): "A guide to multisector labor market models," *Working Papers*, 86.

FINKELSTEIN, E. A., C. ZHEN, J. NONNEMAKER, AND J. E. TODD (2010): "Impact of targeted beverage taxes on higher-and lower-income households," *Archives of Internal Medicine*, 170, 2028–2034.

FLETCHER, J. M., D. FRISVOLD, AND N. TEFFT (2010a): "Can soft drink taxes reduce population weight?" *Contemporary Economic Policy*, 28, 23–35.

FLETCHER, J. M., D. E. FRISVOLD, AND N. TEFFT (2010b): "The effects of soft drink taxes on child and adolescent consumption and weight outcomes," *Journal of Public Economics*, 94, 967–974.

FUCHS, V. R. (1982): "Time preference and health: an exploratory study," *National Bureau of Economic Research Cambridge, Mass., USA*.

FULLERTON, D. AND G. E. METCALF (2002): "Tax incidence," *Handbook of public economics*, 4, 1787–1872.

GAGNON, E. (2009): "Price setting during low and high inflation: Evidence from Mexico," *The Quarterly Journal of Economics*, 124, 1221–1263.

GOLDBERG, P. K. AND F. VERBOVEN (2001): "The evolution of price dispersion in the European car market," *The Review of Economic Studies*, 68, 811–848.

GREEN, K. (2005): "Decentralization and good governance: The case of Indonesia," .

GRIFFITH, R. AND M. O'CONNELL (2009): "The Use of Scanner Data for Research into Nutrition*," *Fiscal Studies*, 30, 339–365.

GROSSMAN, M. (1972): "On the concept of health capital and the demand for health," *The Journal of Political Economy*, 223–255.

GROSSMAN, M., E. TEKIN, AND R. WADA (2014): "Food prices and body fatness among youths," *Economics & Human Biology*, 12, 4–19.

GRUBER, J. AND M. FRAKES (2006): "Does falling smoking lead to rising obesity?" *Journal of health economics*, 25, 183–197.

GRUBER, J. AND B. KOSZEGI (2000): "Is addiction" rational"? Theory and evidence," Tech. rep., National bureau of economic research.

GRUBER, J. H. AND S. MULLAINATHAN (2005): "Do cigarette taxes make smokers happier," *Advances in Economic Analysis & Policy*, 5.

HARDING, M., E. LEIBTAG, AND M. LOVENHEIM (2010): "The heterogeneous geographic and socioeconomic incidence of cigarette and beer taxes: evidence from Nielsen Homescan data," .

Harding, M. and M. Lovenheim (2014): "The Effect of Prices on Nutrition: Comparing the Impact of Product-and Nutrient-Specific Taxes," Tech. rep., National Bureau of Economic Research.

Hawkes, C. (2006): "Uneven dietary development: linking the policies and processes of globalization with the nutrition transition, obesity and diet-related chronic diseases," *Globalization and health*, 2, 4.

Hellerstein, R. (2008): "Who bears the cost of a change in the exchange rate? Pass-through accounting for the case of beer," *Journal of international economics*, 76, 14–32.

Hill, H. (2000): *The Indonesian Economy*, Cambridge: Cambridge University Press.

Ippolito, P. M. and A. D. Mathios (1995): "Information and advertising: The case of fat consumption in the United States," *The American Economic Review*, 91–95.

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

Johnson, R. J., M. S. Segal, Y. Sautin, T. Nakagawa, D. I. Feig, D.-H. Kang, M. S. Gersch, S. Benner, and L. G. Sánchez-Lozada (2007): "Potential role of sugar (fructose) in the epidemic of hypertension, obesity and the metabolic syndrome, diabetes, kidney disease, and cardiovascular disease," *The American journal of clinical nutrition*, 86, 899–906.

Kim, D. and I. Kawachi (2006): "Food taxation and pricing strategies to thin out the obesity epidemic," *American journal of preventive medicine*, 30, 430–437.

Kleibergen, F. and R. Paap (2006): "Generalized reduced rank tests using the singular value decomposition," *Journal of Econometrics*, 133, 97–126.

Knaul, F. M., E. González-Pier, O. Gómez-Dantés, D. García-Junco, H. Arreola-Ornelas, M. Barraza-Lloréns, R. Sandoval, F. Caballero, M. Hernández-Avila, M. Juan, et al. (2012): "The quest for universal health coverage: achieving social protection for all in Mexico," *The Lancet*, 380, 1259–1279.

Lakdawalla, D. and T. Philipson (2002): "The growth of obesity and technological change: a theoretical and empirical examination," Tech. rep., National Bureau of Economic Research.

Levinsohn, J. and A. Petrin (2003): "Estimating Production Functions Using Inputs to Control for Unobservables," *Review of Economic Studies*, 70, 317–342.

Lewis, W. A. (1978): *The evolution of the international economic order*, vol. 151, Princeton University Press Princeton.

Lloyd-Sherlock, P., J. Beard, N. Minicuci, S. Ebrahim, and S. Chatterji (2014): "Hypertension among older adults in low-and middle-income countries: prevalence, awareness and control," *International journal of epidemiology*, 43, 116–128.

Lopez, R. A. and K. L. Fantuzzi (2012): "Demand for carbonated soft drinks: implications for obesity policy," *Applied Economics*, 44, 2859–2865.

Lustig, R. H. (2013): *Fat Chance: The bitter truth about sugar*, Fourth Estate.

Maersk, M., A. Belza, H. Stødkilde-Jørgensen, S. Ringgaard, E. Chabanova, H. Thomsen, S. B. Pedersen, A. Astrup, and B. Richelsen (2012): "Sucrose-sweetened beverages increase fat storage in the liver, muscle, and visceral fat depot: a 6-mo randomized intervention study," *The American journal of clinical nutrition*, 95, 283–289.

Malik, V. S., M. B. Schulze, and F. B. Hu (2006): "Intake of sugar-sweetened beverages and weight gain: a systematic review," *The American journal of clinical nutrition*, 84, 274–288.

Miljkovic, D. and W. Nganje (2008): "Regional obesity determinants in the United States: A model of myopic addictive behavior in food consumption," *Agricultural Economics*, 38, 375–384.

Mischel, W., Y. Shoda, and M. I. Rodriguez (1989): "Delay of gratification in children," *Science*, 244, 933–938.

Monteiro, C. A., W. L. Conde, and B. M. Popkin (2007): "Income-specific trends in obesity in Brazil: 1975-2003," *American Journal of Public Health*, 97, 1808.

Monteiro, C. A., E. C. Moura, W. L. Conde, and B. M. Popkin (2004): "Socioeconomic status and obesity in adult populations of developing countries: a review," *Bulletin of the World Health Organization*, 82, 940–946.

Nevo, A. (2001): "Measuring market power in the ready-to-eat cereal industry," *Econometrica*, 69, 307–342.

Nicita, A. (2009): "The price effect of tariff liberalization: Measuring the impact on household welfare," *Journal of Development Economics*, 89, 19–27.

Olley, G. S. and A. Pakes (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64, 1263–1297.

Oster, E. (2014): "Diabetes and Diet: Behavioral Response and the Value of Health," .

Pitt, M. M. and M. R. Rosenzweig (1984): "Agricultural prices, food consumption and the health and productivity of farmers," *Economic Development Center*.

Poczter, S., P. Gertler, and A. D. Rothenberg (2014): "Financial Crisis and Productivity Evolution: Evidence from Indonesia," *The World Economy*, 37, 705–731.

Popkin, B. M. (1994): "The nutrition transition in low-income countries: an emerging crisis," *Nutrition reviews*, 52, 285–298.

Popkin, B. M., L. S. Adair, and S. W. Ng (2012): "Global nutrition transition and the pandemic of obesity in developing countries," *Nutrition reviews*, 70, 3–21.

Powell, L. M., M. C. Auld, F. J. Chaloupka, P. M. OMalley, and L. D. Johnston (2007a): "Access to fast food and food prices: relationship with fruit and vegetable consumption and overweight among adolescents," *Advances in health economics and health services research*, 17, 23–48.

——— (2007b): "Access to fast food and food prices: relationship with fruit and vegetable consumption and overweight among adolescents," *Advances in health economics and health services research*, 17, 23–48.

Powell, L. M. and F. J. Chaloupka (2009): "Food prices and obesity: evidence and policy implications for taxes and subsidies," *Milbank Quarterly*, 87, 229–257.

Reaven, G. M. (1991): "Relationship between insulin resistance and hypertension," *Diabetes care*, 14, 33–38.

Rivera, J. A., S. Barquera, T. González-Cossío, G. Olaiz, and J. Sepulveda (2004): "Nutrition transition in Mexico and in other Latin American countries," *Nutrition reviews*, 62, S149–S157.

Roback, J. (1982): "Wages, Rents, and the Quality of Life," *Journal of Political Economy*, 90, 1257–1278.

Rosen, S. (1979): "Wage-Based Indexes of Urban Quality of Life," in *Current Issues in Urban Economics*, ed. by P. Mieszkowski and M. Straszheim, Johns Hopkins University Press.

Rothenberg, A. D. (2013): "Transport Infrastructure and Firm Location Choice in Equilibrium: Evidence from Indonesia's Highways," Working Paper.

Saffer, H. and F. Chaloupka (2000): "The effect of tobacco advertising bans on tobacco consumption," *Journal of health economics*, 19, 1117–1137.

Salcido, V. (2013): "Mexico's Food Processing and Ingredients Report," *USDA Foreign Agricultural Service*.

Samuelson, P. A. (1937): "A note on measurement of utility," *The Review of Economic Studies*, 4, 155–161.

Sánchez-Barriga, J. J. (2010): "Mortality trends from diabetes mellitus in the seven socioeconomic regions of Mexico, 2000-2007," *Revista Panamericana de Salud Pública*, 28, 368–375.

Sánchez-Castillo, C. P., O. Velásquez-Monroy, A. Lara-Esqueda, A. Berber, J. Sepulveda, R. Tapia-Conyer, and W. P. T James (2005): "Diabetes and hypertension increases in a society with abdominal obesity: results of the Mexican National Health Survey 2000," *Public health nutrition*, 8, 53–60.

Sen, A. K. (1966): "Peasants and Dualism With and Without Surplus Labor," *Journal of Political Economy*, 74, 425–450.

Shah, A., Z. Qureshi, A. Bagchi, B. Binder, and H.-f. Zou (1994): "Intergovernmental fiscal relations in Indonesia: issues and reform options," Tech. rep., China Economics and Management Academy, Central University of Finance and Economics.

Singh, I., L. Squire, and J. Strauss (1986): *Agricultural Household Models: Extensions, Applications, and Policy*, Baltimore: Johns Hopkins University Press.

Smith, P. K., B. Bogin, and D. Bishai (2005): "Are time preference and body mass index associated?: Evidence from the National Longitudinal Survey of Youth," *Economics & Human Biology*, 3, 259–270.

Stanhope, K. L., A. A. Bremer, V. Medici, K. Nakajima, Y. Ito, T. Nakano, G. Chen, T. H. Fong, V. Lee, R. I. Menorca, et al. (2011): "Consumption of fructose and high fructose corn syrup increase postprandial triglycerides, LDL-cholesterol, and apolipoprotein-B in young men and women," *The Journal of Clinical Endocrinology & Metabolism*, 96, E1596–E1605.

Strauss, J. and D. Thomas (1998): "Health, nutrition, and economic development," *Journal of economic literature*, 766–817.

Sturm, R. and A. Datar (2005): "Body mass index in elementary school children, metropolitan area food prices and food outlet density," *Public health*, 119, 1059–1068.

Tapia-Conyer, R., P. Kuri-Morales, L. González-Urbán, and E. Sarti (2001): "Evaluation and Reform of Mexican National Epidemiological Surveillance System," *American journal of public health*, 91, 1758–1760.

Taubes, G. (2007): *Good calories, bad calories*, Random House Digital, Inc.

Teff, K. L., J. Grudziak, R. R. Townsend, T. N. Dunn, R. W. Grant, S. H. Adams, N. L. Keim, B. P. Cummings, K. L. Stanhope, and P. J. Havel (2009): "Endocrine and metabolic effects of consuming fructose-and glucose-sweetened beverages with meals in obese men and women: influence of insulin resistance on plasma triglyceride responses," *The Journal of Clinical Endocrinology & Metabolism*, 94, 1562–1569.

Thomas, D. and E. Frankenberg (2000): "The measurement and interpretation of health in social surveys," Tech. rep., DTIC Document.

Thomas, D. and J. Strauss (1992): "Prices, infrastructure, household characteristics and child height," *Journal of Development Economics*, 39, 301–331.

Thow, A. M., S. Jan, S. Leeder, and B. Swinburn (2010): "The effect of fiscal policy on diet, obesity and chronic disease: a systematic review," *Bulletin of the World Health Organization*, 88, 609–614.

Tobenna, D. A. and I. Rahkovsky (2014): "The Association Between Food Prices and the Blood Glucose Level of US Adults With Type 2 Diabetes," *American Journal of Public Health*, 104, 678–685.

Varela, J. M. (2013): "The Costs of Growth: Estimating Entry Costs During Rollouts," *Working Paper*.

Wardle, J., J. Waller, and M. J. Jarvis (2002): "Sex differences in the association of socioeconomic status with obesity," *American journal of public health*, 92, 1299–1304.

Wasserman, J., W. G. Manning, J. P. Newhouse, and J. D. Winkler (1991): "The effects of excise taxes and regulations on cigarette smoking," *Journal of health economics*, 10, 43–64.

Whiting, D. R., L. Guariguata, C. Weil, and J. Shaw (2011): "IDF diabetes atlas: global estimates of the prevalence of diabetes for 2011 and 2030," *Diabetes research and clinical practice*, 94, 311–321.

WHO (2002): "The World health report: 2002: Reducing the risks, promoting healthy life," *Geneva: World Health Organization*.

World Bank (2008): *Spending for Development: Making the Most of Indonesia's New Opportunities*, World Bank.

——— (2012): *Investing in Indonesia's roads : improving efficiency and closing the financing gap*.

Yu, J., E. Chou, and J. Yau (2006): "Development of Speed-Related Ride Quality Thresholds using International Roughness Index," *Transportation Research Record*, 1974, 47–53.

Appendix A: Chapter 1

Table A1: Soda Consumption

Dep.Var: Log(Soda Consumption pc)	**(1)**	**(2)**	**(3)**	**(4)**
$Log(P_{sodas})_t$	-0.403	-0.391	-0.385	-0.390
	(0.137)***	(0.142)***	(0.143)***	(0.101)***
N	16135	16135	16135	14259
Adj. R^2	0.27	0.27	0.29	0.05
F-Stat	10.91	3.75	7.39	15.04
Y Mean			6.63	6.63
Y SD			26.05	26.05
N Clusters	239	239	239	164
Household FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Same Community	No	No	No	Yes
HH Controls	.	.	Yes	Yes
Other food prices control	.	Yes	Yes	Yes

Dependent variable is log of liters of monthly soda consumption per capita per household. Prices are proxied with median community unit values (expenditure/quantity), to avoid outliers and quality issue related to unit values. Controls include household average years of schooling, household size, number of rooms, log of annual household's income, ownership indicator variables (motor vehicle, bike, house, tap in the dwelling, telephone, electricity), state level log(GDP), state drought index, and number of fast food services per km^2 and their real advertising expenditures per capita at municipality level. Robust standard errors in parentheses, clustered at the community level. */**/*** denotes significant at the 10% / 5% / 1% levels.

Table A2: Diabetes II and Hypertension - Risk for disease (long)

DV	**Diabetes II**		**Hypertension**	
	(1)	**(2)**	**(3)**	**(4)**
$Log(P\_sugar)\_t$	0.028 (0.039)	-0.003 (0.052)	0.041 (0.057)	0.059 (0.069)
Med Risk Diab x $Log(P\_sugar)\_t$	-0.033 (0.065)	-0.013 (0.061)		
High Risk Diab x $Log(P\_sugar)\_t$	-0.194 (0.084)**	-0.210 (0.107)*		
$Log(P\_fats)\_t$		0.080 (0.049)		-0.002 (0.067)
Med Risk Diab x $Log(P\_fats)\_t$		-0.122 (0.040)***		
High Risk Diab x$Log(P_fats)_t$		-0.012 (0.096)		
$Log(P\_fiber)\_t$		0.001 (0.027)		-0.044 (0.050)
Med Risk Diab x $Log(P\_fiber)\_t$		-0.029 (0.036)		
risk2_all=2x$Log(P_fiber)_t$		-0.065 (0.056)		
$Log(P\_protein)\_t$		-0.040 (0.057)		0.059 (0.092)
Med Risk Diab x $Log(P\_protein)\_t$		0.016 (0.049)		
High Risk Diab x $Log(P\_protein)\_t$		-0.036 (0.118)		
High Risk Hyper x $Log(P\_sugar)\_t$			-0.209 (0.089)**	-0.153 (0.137)
High Risk Hyper x $Log(P\_fats)\_t$				0.011 (0.106)
High Risk Hyper x $Log(P\_fiber)\_t$				0.075 (0.109)
High Risk Hyper x $Log(P\_protein)\_t$				0.095 (0.183)
N	10524	10524	10727	10727
Adj. R^2	0.06	0.06	0.03	0.03
Y Mean		0.12		0.20
Y SD		0.32		0.40
N Clusters		29		29

Individual and year fixed effects, controls 1 and controls 2 are included in all regressions. Robust standard errors in parentheses, clustered at the city level. */**/*** denotes significant at the 10% / 5% / 1% levels.

Table A3: Long Run Effect by Initial Risk for Disease

Panel A				
DV: Diabetes II	**High Risk**		**Low Risk**	
	(1)	**(2)**	**(3)**	**(4)**
$Log(P_{sugar})_t$	-0.188		0.043	
	(0.088)**		(0.043)	
$Log(P_{sugar})_{t-3}$		-0.006		-0.060
		(0.078)		(0.029)**
N	2273	1474	4308	3451
Adj. R^2	0.06	0.08	0.05	0.06
F-Stat	28.55	53.28	245.07	40.28
Y Mean	0.28	0.29	0.04	0.05
Y SD	0.45	0.45	0.20	0.23
N Clusters	29	29	29	29

Panel B				
DV: Hypertension	**High Risk**		**Low Risk**	
	(1)	**(2)**	**(3)**	**(4)**
$Log(P_{sugar})_t$	-0.102		0.008	
	(0.084)		(0.060)	
$Log(P_{sugar})_{t-3}$		0.092		-0.145
		(0.059)		(0.054)**
N	3511	2295	7167	5392
Adj. R^2	0.03	0.05	0.04	0.06
F-Stat	54.76	19.49	38.44	127.29
Y Mean	0.25	0.26	0.17	0.16
Y SD	0.43	0.44	0.38	0.37
N Clusters	29	29	29	29
Individual FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Controls 1	Yes	Yes	Yes	Yes
Controls 2	Yes	Yes	Yes	Yes
Surveys	2002-2009	2005-2009	2002-2009	2005-2009

People who scored at least 10 or below 6 points on the Finnish Diabetes Risk Score, are considered as high and low risk group for type 2 diabetes, respectively. (See Figure 1.15). Parental history is adjusted so that 3 points are given in the case of more than one household member having diabetes. The question on your daily vegetable consumption is excluded, and question 5 only related to high glucose level during pregnancy. People at high risk for hypertension are those scoring at least 4 points, each risk factor counting 1 point (initially obese, abdominally obese, smoking and not exercising, experiencing sleeping problems and stress, being diabetic, prehypertensive and older than 45 years old). Results are not particularly sensitive on cut-offs or group of risk factors included. Robust standard errors in parentheses, clustered at the city level. */**/*** denotes significant at the 10% / 5% / 1% levels.

Table A4: Impatience proxied with spending/saving decisions

Panel A				
DV: Diabetes II	**All**		**High Risk**	
	(1)	**(2)**	**(3)**	**(4)**
$Log(P_{sugar})_t$	-0.025	-0.024	-0.053	0.019
	(0.024)	(0.032)	(0.072)	(0.087)
$Log(P_{sugar})_t$ x Spent90+	-0.079		-0.168	
	(0.050)		(0.072)**	
$Log(P_{sugar})_t$ x Spent50+		-0.046		-0.210
		(0.044)		(0.099)**
N	10026	10026	2254	2254
Adj. R^2	0.05	0.05	0.06	0.06
Y Mean	0.12		0.27	
Y SD	0.32		0.45	
N Clusters	29	29	29	29

Panel B				
DV: Hypertension	**All**		**High Risk**	
	(1)	**(2)**	**(3)**	**(4)**
$Log(P_{sugar})_t$	0.014	0.054	-0.016	0.015
	(0.052)	(0.073)	(0.096)	(0.134)
$Log(P_{sugar})_t$ x Spent90+	-0.025		-0.123	
	(0.077)		(0.135)	
$Log(P_{sugar})_t$ x Spent50+		-0.080		-0.123
		(0.083)		(0.144)
N	10247	10247	2039	2039
Adj. R^2	0.04	0.04	0.05	0.05
Y Mean	0.20		0.25	
Y SD	0.40		0.43	
N Clusters	29	29	29	29
Individual FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Controls 1	Yes	Yes	Yes	Yes
Controls 2	Yes	Yes	Yes	Yes

Indicator variable Spend90+ and Spend50+ equals 1 if one would spend more than 90 or 50 percent of $1000, in case offered, instead of saving either all or part of it. Indicator variable is constant over time and is assigned to individuals based on this question in 2005 survey. Controls 1 and Controls 2 are the same as in previous tables. Robust standard errors in parentheses, clustered at the city level. */**/*** denotes significant at the 10% / 5% / 1% levels.

Table A5: Impatience proxied with spending/saving decisions

Abdominal Obesity - Panel C				
DV:	**Log(waistline)**		**Abd. Obese**	
	(1)	**(2)**	**(3)**	**(4)**
$Log(P_{sugar})_t$	-0.039	-0.045	-0.168	-0.162
	(0.021)*	(0.024)*	(0.042)***	(0.051)***
$Log(P_{sugar})_t$ x Spent90+	0.010		0.084	
	(0.018)		(0.082)	
$Log(P_{sugar})_t$ x Spent50+		0.014		0.036
		(0.016)		(0.064)
N	18055	18055	18055	18055
Adj. R^2	0.32	0.32	0.13	0.13
Y Mean	93.58		0.78	
Y SD	12.57		0.42	
N Clusters	31	31	31	31
Individual FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Controls 1	Yes	Yes	Yes	Yes
Controls 2	Yes	Yes	Yes	Yes

Indicator variable Spend90+ and Spend50+ equals 1 if one would spend more than 90 or 50 percent of $1000, in case offered, instead of saving either all or part of it. Indicator variable is constant over time and is assigned to individuals based on this question in 2005 survey. Controls 1 and Controls 2 are the same as in previous tables. Robust standard errors in parentheses, clustered at the city level. */**/*** denotes significant at the 10% / 5% / 1% levels.

Table A6: Other Interactions

DV	**Log(waist)**	**Abd.Obese**	**Diab II**	**Hyper**
	(1)	**(2)**	**(3)**	**(4)**
$Log(P_{sugar})_t$	0.030	-0.247	0.392	-0.288
	(0.107)	(0.426)	(0.514)	0.544)
$Log(P_{sugar})_t$ *x Impatience*(5)	-0.04	-0.105	-0.184	-0.175
	(0.0163)**	(0.062)*	(0.076)*	(0.074)*
$Log(Income)$	0.010	-0.074	0.162	-0.114
	(0.467)	0.181)	(0.219)	(0.215)
$Log(P_{sugar})_t$ *x* $Log(Income)$	-0.002	0.017	-0.035	0.029
	(0.101)	(0.039)	(0.476)	(0.047)
$Log(P_{sugar})_t$ x Sex(1=M)	-0.065	-0.043	0.047	0.032
	(0.017)**	(0.049)	(0.050)	(0.084)
N	16797	16797	1982	3051
Adj. R^2	0.296	0.13	0.07	0.06
N Clusters	32	32	29	29

Individual, year FE, and controls are included. Interactions of food prices with work status, education, or risk aversion do not change the results and are not significant at conventional levels. Robust standard errors in parentheses, clustered at the city level. */**/*** denotes significant at the 10% / 5% / 1% levels.

Table A7: Diabetes II

Dep. Var: Diabetes II	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
$Log(P_{sugar})_t$	-100.980	-14.990	-12.381	-22.945	-23.625
	(0.004)	(0.499)	(0.570)	(0.258)	(0.244)
$Log(P_{sugar})_{t-1}$		-82.856	-83.909	-83.986	-82.625
		(0.004)***	(0.004)***	(0.004)***	(0.004)***
$Log(P_{sugar})_{t-2}$		-33.886	-35.177	-30.925	-31.447
		(0.024)**	(0.019)**	(0.026)**	(0.030)**
$Log(P_{sugar})_{t-3}$		-15.886	-15.345	-27.117	-27.822
		(0.319)	(0.368)	(0.127)	(0.097)*
$Log(P_{sugar})_{t-4}$		-13.735	-15.102	-2.134	-3.510
		(0.574)	(0.548)	(0.938)	(0.901)
$Log(P_{fats})_{t-1}$					-6.587
					(0.889)
$Log(P_{protein})_{t-1}$					-4.722
					(0.917)
$Log(P_{fiber})_{t-1}$					-10.476
					(0.764)
N	480	383	383	351	351
Adj. R^2	0.72	0.77	0.77	0.78	0.77
Macro Controls	.	.	Yes	Yes	Yes
FastFood&Adv	.	.	.	Yes	Yes
N Clusters	32	32	32	32	32

State and year FE included. Wild bootstrapped standard errors (1000 repetitions), p-values reported in parentheses. */**/*** denotes significant at the 10% / 5% / 1% levels. Diabetes II incidence rate mean is 358.

Table A8: Hypertension

Dep. Var: Hypertension	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
$Log(P_{sugar})_t$	-183.838	-82.147	-75.344	-87.511	-89.801
	(0.004)***	(0.219)	(0.228)	(0.066)*	(0.0664)*
$Log(P_{sugar})_{t-1}$		-72.738	-85.281	-73.398	-69.101
		(0.009)***	(0.009)***	(0.054)*	(0.091)*
$Log(P_{sugar})_{t-2}$		-47.263	-34.432	-49.660	-50.413
		(0.037)**	(0.272)	(0.109)	(0.111)
$Log(P_{sugar})_{t-3}$		-67.712	-74.155	-87.832	-89.356
		(0.105)	(0.079)*	(0.041)**	(0.043)**
$Log(P_{sugar})_{t-4}$		-51.613	-51.844	-22.775	-26.903
		(0.301)	(0.284)	(0.635)	(0.599)
$Log(P_{fats})_{t-1}$					-11.813
					(0.885)
$Log(P_{protein})_{t-1}$					-18.374
					(0.631)
$Log(P_{fiber})_{t-1}$					-27.738
					(0.542)
N	480	383	383	351	351
Adj. R^2	0.79	0.85	0.85	0.85	0.85
Macro Controls	.	.	Yes	Yes	Yes
FastFood&Adv	.	.	.	Yes	Yes
N Clusters	32	32	32	32	32

State and year FE included. Wild bootstrapped standard errors (1000 repetitions), p-values reported in parentheses. */**/*** denotes significant at the 10% / 5% / 1% levels. Diabetes II incidence rate mean is 548.55.

Proofs:

Proposition 1. *Increase/decrease in price p_t improves/deteriorates health if $p_t < \lambda$. The effect is increasing in λ.*

Proof. Proved in text. ☐

Proposition 2. *Individual's health is increasing in one's discount factor, that is, those more impatient have lower health H_t compared to the more patient ones: $\frac{\partial H_t}{\partial \delta} > 0$. Health response to change in p_t is decreasing in δ, hence $\frac{\partial H}{\partial p \partial \delta} < 0$.*

Proof. $\frac{\partial H_t}{\partial \delta} > 0$ proved in text. ☐

Proof.

$$\frac{\partial H_{t+1}}{\partial p_t \partial \delta} = -((\lambda - p_t)\frac{\partial^2 n_t^*}{\partial p_t \partial \delta} - \frac{\partial n_t^*}{\partial \delta}) < 0$$

- $\frac{\partial}{\partial \delta}\left(\frac{\partial F(n_t,p_t)}{\partial n}\Big|_{n=n^*}\right) = 0$
- $\frac{\partial n_t^*}{\partial \delta} = -\left(\frac{\partial F}{\partial n_t}(n_t,p_t)\Big|_{n=n^*}\right)^{-1}\frac{\partial D}{\partial \delta}(-\lambda + p_t) < 0$
- $\frac{\partial}{\partial \delta}\left(\frac{\partial F(n_t,p_t)}{\partial p_t}\Big|_{n=n^*}\right) = \underbrace{\frac{\partial D}{\partial \delta}}_{>0} + \underbrace{\left(\frac{\partial A}{\partial n_t^*}\right)}_{<0} \cdot \underbrace{\left(\frac{\partial n_t^*}{\partial \delta}\right)}_{<0} + \underbrace{\left(\frac{\partial A}{\partial p_t}\right) \cdot \left(\frac{\partial p_t}{\partial \delta}\right) + \left(\frac{\partial A}{\partial w_t}\right) \cdot \left(\frac{\partial w_t}{\partial \delta}\right)}_{=0} > 0$
- $\frac{\partial^2 n^*}{\partial p \partial \delta}$

$$\frac{\partial^2 n_t^*}{\partial p_t \partial \delta} = \underbrace{-\left(\frac{\partial F}{\partial n_t}(n_t,p_t)\Big|_{n=n^*}\right)^{-2}}_{<0} \cdot$$

$$\left(\underbrace{\left(\frac{\partial F(n_t,p_t)}{\partial n}\Big|_{n=n^*}\right)}_{<0} \cdot \underbrace{\frac{\partial}{\partial \delta}\left(\frac{\partial F(n_t,p_t)}{\partial p_t}\Big|_{n=n^*}\right)}_{>0} - \underbrace{\left(\frac{\partial F(n_t,p_t)}{\partial p_t}\Big|_{n=n^*}\right) \cdot 0}_{=0}\right) > 0$$

☐

Proposition 3. *Increase/decrease in price p_t improves/deteriorates health H_{t+1} more for those less healthy, that is, those with lower H_t: $\frac{d}{dH_t}\frac{dH_{t+1}}{dp_t} < 0$.*

Proof.

$$\begin{aligned}
\frac{\partial}{\partial H_t}\frac{\partial H_{t+1}}{\partial p_t} &= \frac{\partial}{\partial \delta}\frac{\partial H_{t+1}}{\partial p_t}\cdot\frac{\partial \delta}{\partial H_t} = \frac{\frac{\partial^2 H_{t+1}}{\partial\delta\partial p_t}}{\frac{\partial H_t}{\partial \delta}} \\
&= \frac{-\left((\lambda - p_t)\frac{\partial n_t^*}{\partial p_t \partial \delta} - \frac{\partial n^*}{\partial \delta}\right)}{\sum_{s=0}^{t-1}(1-d)^{t-s}\underbrace{\left(-\left((\lambda - p_s)\dfrac{\partial n_s^*}{\partial \delta} - n_s^*\right)\right)}_{\frac{\partial H_s}{\partial \delta}>0}} < 0
\end{aligned}$$

$\square$

Appendix B: Chapter 3

Table B1: Budget IV First Stage

	Kabupatens (1)	**Households (2)**	**Individuals (3)**
Log Budget Proxy (Nat'l Roads)	0.122 (0.010)***	0.123 (0.024)***	0.169 (0.012)***
Log Budget Proxy (Prov Roads)		0.046 (0.025)*	
Log Budget Proxy (Kabu Roads)	-0.023 (0.004)***		
Log Budget Proxy (Nat'l Roads, t-1)	0.091 (0.012)***	-0.036 (0.014)***	-0.054 (0.015)***
Log Budget Proxy (Prov Roads, t-1)	-0.034 (0.005)***	-0.093 (0.015)***	-0.064 (0.012)***
Log Budget Proxy (Kabu Roads, t-1)		-0.052 (0.014)***	-0.013 (0.006)**
Log Budget Proxy (Nat'l Roads, t-2)	0.099 (0.014)***		
Log Budget Proxy (Prov Roads, t-2)	-0.034 (0.006)***	-0.078 (0.018)***	-0.065 (0.020)***
Log Budget Proxy (Kabu Roads, t-2)		-0.045 (0.013)***	-0.036 (0.013)***
Log B-Proxy (Nat'l Roads)^2	-0.014 (0.001)***	-0.012 (0.003)***	-0.018 (0.002)***
Log B-Proxy (Prov Roads)^2	-0.005 (0.000)***	-0.008 (0.003)**	-0.002 (0.001)*
Log B-Proxy (Kabu Roads)^2			-0.004 (0.002)**
Log B-Proxy (Nat'l Roads)^2 (t-1)	-0.010 (0.001)***	0.005 (0.002)**	0.007 (0.002)***
Log B-Proxy (Prov Roads)^2 (t-1)		0.013 (0.003)***	0.009 (0.002)***
Log B-Proxy (Kabu Roads)^2 (t-1)	-0.002 (0.001)**	0.011 (0.004)**	
Log B-Proxy (Nat'l Roads)^2 (t-2)	-0.011 (0.001)***	-0.004 (0.001)***	-0.003 (0.001)**
Log B-Proxy (Prov Roads)^2 (t-2)		0.012 (0.003)***	0.009 (0.003)***
Log B-Proxy (Kabu Roads)^2 (t-2)	-0.004 (0.001)***	0.015 (0.004)***	0.011 (0.004)***
N	3121	22617	66254
Adj. R^2	0.413	0.672	0.644
Regression F-Stat	77.708	149.049	97.222
DV Mean	1.000	1.831	1.828
Kabupaten-Years	Yes	.	.
Household-Years	.	Yes	.
Individual-Years	.	.	Yes
Year FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes

All regressions include a constant. Kabupaten-year controls include logs of population and non-oil GDRP. Household controls include: district GDP and month of survey indicators. Individual controls include: district GDP, age, education, household size, month of survey indicators. Robust standard errors in parentheses, clustered at the kabupaten level in column 1, and the village level in columns 2 and 3. */**/*** denotes significant at the 10% / 5% / 1% levels.

Table B2: Variable Names for Robustness Tables

Variable	**Description**	**Original Table**
Y_1	Working? (0 1)	Table 3.9
Y_2	Log Hours Worked (0 1)	Table 3.9
Y_3	Log Total Earnings	Table 3.12
Y_4	Log Manufacturing Earnings	Table 3.12
Y_5	Log Agricultural Earnings	Table 3.12
Y_6	Log Sales and Services Earnings	Table 3.12
Y_7	Log Other Sector Earnings	Table 3.12
Y_8	Working in Agriculture (0 1)	Table 3.10
Y_9	Working in Agriculture, Primary (0 1)	Table 3.10
Y_{10}	Log Hours Worked, Agriculture	Table 3.10
Y_{11}	Working in Manufacturing (0 1)	Table 3.10
Y_{12}	Working in Manufacturing, Primary (0 1)	Table 3.10
Y_{13}	Log Hours Worked, Manufacturing	Table 3.10
Y_{14}	Working in Sales and Services (0 1)	Table 3.10
Y_{15}	Working in Sales and Services, Primary (0 1)	Table 3.10
Y_{16}	Log Hours Worked, Sales and Services	Table 3.10
Y_{17}	Working in Other (0 1)	Table 3.10
Y_{18}	Working in Other, Primary (0 1)	Table 3.10
Y_{19}	Log Hours Worked, Other	Table 3.10

This table contains the names and descriptions of variables, for reference to individual and household-level robustness tables, Appendix Tables B3.

Table B3: Robustness: Individual-Level Results (Part 1)

Panel A: Preferred Spec. (GMM)	Y_1 **(1)**	Y_2 **(2)**	Y_3 **(3)**	Y_4 **(4)**	Y_5 **(5)**	Y_6 **(6)**	Y_7 **(7)**
Log Avg IRI (Budget IV)	-0.037	-0.011	-0.894	0.058	-1.303	-3.474	0.080
	(0.034)	(0.060)	(0.328)***	(0.105)	(0.664)**	(1.224)***	(0.215)
Panel B: OLS Results	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
Log Avg IRI	0.016	0.010	-0.465	0.030	-1.149	-1.284	0.072
	(0.020)	(0.035)	(0.190)**	(0.059)	(0.389)***	(0.770)*	(0.132)
Log Avg IRI, Non-Movers	0.034	-0.011	-0.382	0.012	-1.128	-1.401	0.141
	(0.022)	(0.038)	(0.215)*	(0.064)	(0.431)***	(0.839)*	(0.161)
Log Avg IRI, Rural Areas	0.081	-0.018	-0.563	0.032	-0.709	-0.638	0.351
	(0.027)***	(0.052)	(0.354)	(0.113)	(0.601)	(0.851)	(0.268)
Log Travel Time to Prov Capital	0.016	0.027	-0.147	-0.021	-0.580	-1.674	-0.025
	(0.009)*	(0.031)	(0.114)	(0.032)	(0.268)**	(0.533)***	(0.064)
Log Local Market Access (GDRP weights)	0.017	0.028	-0.181	-0.015	-0.665	-1.699	-0.030
	(0.010)	(0.032)	(0.127)	(0.035)	(0.294)**	(0.570)***	(0.070)
Log Local Market Access (Pop weights)	0.018	0.027	-0.179	-0.013	-0.650	-1.642	-0.035
	(0.010)*	(0.032)	(0.127)	(0.034)	(0.290)**	(0.567)***	(0.069)
Panel C: GMM Results	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
Log Avg IRI, Non-Movers	-0.014	-0.013	-0.805	-0.002	-1.224	-3.782	0.051
	(0.036)	(0.069)	(0.371)**	(0.124)	(0.748)	(1.367)***	(0.238)
Log Avg IRI, Rural Areas	0.027	-0.015	-0.819	-0.100	-0.607	-1.867	-0.135
	(0.037)	(0.083)	(0.482)*	(0.152)	(0.880)	(1.283)	(0.318)
Log Avg IRI, Hausman IV	-0.022	-0.018	-0.842	-0.101	-1.807	-2.599	-0.276
	(0.031)	(0.058)	(0.306)***	(0.102)	(0.644)***	(1.237)**	(0.220)
Log Travel Time to Prov Capital, Budget IV	0.014	-0.052	-0.267	-0.080	-0.571	-2.184	-0.052
	(0.019)	(0.039)	(0.268)	(0.077)	(0.566)	(1.243)*	(0.141)
Log Local Market Access (GDRP weights), Budget IV	-0.004	-0.053	-0.312	-0.087	-0.595	-1.488	-0.075
	(0.025)	(0.047)	(0.304)	(0.083)	(0.624)	(1.293)	(0.163)
Log Local Market Access (Pop weights), Budget IV	-0.003	-0.052	-0.291	-0.084	-0.567	-1.226	-0.082
	(0.024)	(0.046)	(0.297)	(0.081)	(0.612)	(1.266)	(0.163)

Refer to Appendix Table B2 for variable names and definitions. All regressions include a constant, individual and year fixed effects. Individual controls include: district GDP, age, education, household size, month of survey indicators. Robust standard errors in parentheses, clustered at the village level. */**/*** denotes significant at the 10% / 5% / 1% levels.

Table B4: Robustness: Individual-Level Results (Part 2)

Panel A: Preferred Spec. (GMM)	Y_8 **(1)**	Y_9 **(2)**	Y_{10} **(3)**	Y_{11} **(4)**	Y_{12} **(5)**	Y_{13} **(6)**
Log Avg IRI (Budget IV)	0.055 (0.033)*	0.055 (0.034)	2.889 (2.494)	-0.114 (0.036)***	-0.129 (0.033)***	-6.240 (1.802)***
Panel B: OLS Results	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Log Avg IRI	0.062 (0.018)***	0.055 (0.018)***	1.458 (1.349)	-0.085 (0.022)***	-0.089 (0.020)***	-4.559 (1.130)***
Log Avg IRI, Non-Movers	0.066 (0.019)***	0.055 (0.020)***	1.230 (1.450)	-0.087 (0.025)***	-0.100 (0.023)***	-5.001 (1.231)***
Log Avg IRI, Rural Areas	0.069 (0.029)**	0.071 (0.030)**	1.020 (2.273)	-0.089 (0.032)***	-0.106 (0.028)***	-4.781 (1.469)***
Log Travel Time to Prov Capital	0.026 (0.010)**	0.021 (0.011)*	0.840 (0.603)	-0.050 (0.012)***	-0.055 (0.012)***	-1.477 (0.569)***
Log Local Market Access (GDRP weights)	0.029 (0.011)**	0.024 (0.012)**	1.161 (0.719)	-0.048 (0.013)***	-0.055 (0.012)***	-1.699 (0.613)***
Log Local Market Access (Pop weights)	0.028 (0.011)**	0.022 (0.011)*	1.123 (0.713)	-0.046 (0.013)***	-0.053 (0.012)***	-1.610 (0.602)***
Panel C: GMM Results	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Log Avg IRI, Non-Movers	0.056 (0.036)	0.046 (0.037)	1.498 (2.611)	-0.121 (0.038)***	-0.147 (0.036)***	-6.515 (2.010)***
Log Avg IRI, Rural Areas	0.046 (0.045)	0.040 (0.045)	1.926 (3.601)	-0.120 (0.042)***	-0.128 (0.035)***	-7.331 (2.090)***
Log Avg IRI, Hausman IV	0.086 (0.032)***	0.090 (0.033)***	3.639 (2.066)*	-0.166 (0.037)***	-0.162 (0.033)***	-9.519 (1.840)***
Log Travel Time to Prov Capital, Budget IV	0.017 (0.020)	0.010 (0.018)	-1.502 (1.198)	-0.099 (0.029)***	-0.125 (0.031)***	-5.232 (1.632)***
Log Local Market Access (GDRP weights), Budget IV	0.028 (0.027)	0.006 (0.021)	-1.863 (1.500)	-0.119 (0.035)***	-0.144 (0.034)***	-5.891 (1.859)***
Log Local Market Access (Pop weights), Budget IV	0.028 (0.027)	0.005 (0.021)	-1.831 (1.468)	-0.117 (0.034)***	-0.141 (0.034)***	-5.755 (1.826)***

Refer to Appendix Table B2 for variable names and definitions. All regressions include a constant, individual and year fixed effects. Individual controls include: district GDP, age, education, household size, month of survey indicators. Robust standard errors in parentheses, clustered at the village level. */**/*** denotes significant at the 10% / 5% / 1% levels.

Table B5: Robustness: Individual-Level Results (Part 3)

Panel A: Preferred Spec. (GMM)	Y_{14} **(1)**	Y_{15} **(2)**	Y_{16} **(3)**	Y_{17} **(4)**	Y_{18} **(5)**	Y_{19} **(6)**
Log Avg IRI (Budget IV)	0.054	0.082	4.523	-0.026	-0.015	0.026
	(0.035)	(0.035)**	(1.963)**	(0.018)	(0.011)	(0.802)
Panel B: OLS Results	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Log Avg IRI	0.046	0.053	3.064	-0.020	-0.014	-0.304
	(0.020)**	(0.020)***	(1.319)**	(0.012)*	(0.009)	(0.628)
Log Avg IRI, Non-Movers	0.054	0.059	2.928	-0.018	-0.007	-0.320
	(0.021)**	(0.020)***	(1.334)**	(0.013)	(0.010)	(0.651)
Log Avg IRI, Rural Areas	0.033	0.048	0.738	-0.023	-0.005	0.381
	(0.025)	(0.027)*	(1.381)	(0.013)*	(0.008)	(0.556)
Log Travel Time to Prov Capital	0.029	0.029	1.492	0.005	0.003	0.024
	(0.020)	(0.015)*	(1.351)	(0.014)	(0.009)	(0.469)
Log Local Market Access (GDRP weights)	0.030	0.030	1.702	-0.000	0.001	-0.170
	(0.020)	(0.015)**	(1.339)	(0.015)	(0.010)	(0.530)
Log Local Market Access (Pop weights)	0.030	0.030	1.709	-0.001	0.001	-0.216
	(0.020)	(0.015)**	(1.347)	(0.015)	(0.010)	(0.538)
Panel C: GMM Results	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Log Avg IRI, Non-Movers	0.062	0.104	5.937	-0.032	-0.011	-0.038
	(0.037)*	(0.036)***	(1.919)***	(0.018)*	(0.013)	(0.872)
Log Avg IRI, Rural Areas	0.079	0.093	3.143	-0.025	-0.006	-0.779
	(0.041)*	(0.038)**	(2.061)	(0.016)	(0.011)	(0.550)
Log Avg IRI, Hausman IV	0.091	0.096	5.513	-0.021	-0.011	-0.325
	(0.034)***	(0.032)***	(2.037)***	(0.020)	(0.014)	(0.948)
Log Travel Time to Prov Capital, Budget IV	0.100	0.111	7.353	-0.014	-0.005	-0.465
	(0.033)***	(0.032)***	(1.839)***	(0.014)	(0.009)	(0.513)
Log Local Market Access (GDRP weights), Budget IV	0.099	0.127	7.742	-0.016	-0.004	-0.533
	(0.037)***	(0.036)***	(2.081)***	(0.015)	(0.010)	(0.580)
Log Local Market Access (Pop weights), Budget IV	0.098	0.125	7.515	-0.015	-0.004	-0.547
	(0.037)***	(0.036)***	(2.062)***	(0.015)	(0.010)	(0.573)

Refer to Appendix Table B2 for variable names and definitions. All regressions include a constant, individual and year fixed effects. Individual controls include: district GDP, age, education, household size, month of survey indicators. Robust standard errors in parentheses, clustered at the village level. */**/*** denotes significant at the 10% / 5% / 1% levels.

Table B6: Road Roughness and Probability of Working: Heterogeneity

		Age		**Gender**		**Education**		**Initial PCE**		**Own Farmland?**	
	All Obs	**< Med**	**≥ Med**	**F**	**M**	**< Med**	**≥ Med**	**< Med**	**≥ Med**	**No**	**Yes**
Dep Var: Working? (0 1)	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**	**(11)**
Log Avg IRI	-0.037	-0.123	0.002	-0.020	-0.066	-0.011	-0.095	-0.013	-0.071	-0.062	-0.023
	(0.034)	(0.070)*	(0.033)	(0.048)	(0.033)**	(0.044)	(0.040)**	(0.042)	(0.044)	(0.041)	(0.045)
Adj. R^2	0.454	0.350	0.459	0.374	0.448	0.443	0.462	0.445	0.474	0.471	0.436
Regression F-Stat	34.111	59.174	9.186	23.800	24.119	8.171	42.791	18.410	16.287	21.484	17.711
Kleibergen-Paap F-Stat	33.594	26.782	34.343	33.836	32.212	26.660	34.965	29.379	29.570	27.925	27.021
DV Mean	0.679	0.514	0.748	0.528	0.837	0.714	0.641	0.696	0.662	0.644	0.717
N	35193	7920	27273	18249	16932	19483	15710	17521	16106	17080	16910
N Individuals	12459	3383	9076	6354	6099	6547	5912	6129	5840	6208	5876
N Desa / Kelurahan	377	301	320	328	297	305	311	276	324	316	278
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

All regressions include a constant. Controls include: district GDP, age, education, household size, month of survey indicators. Dependent variables are only defined if the individual reported working. Robust standard errors in parentheses, clustered at the (initial) village level. For the GMM specifications, we report the adjusted R^2 from the analagous reduced-form regression (regressing the dependent variable on the instruments). */**/*** denotes significant at the 10% / 5% / 1% levels.

Table B7: Road Roughness and Hours Worked: Heterogeneity

		Age		**Gender**		**Education**		**Initial PCE**		**Own Farmland?**	
	All Obs	**< Med**	**≥ Med**	**F**	**M**	**< Med**	**≥ Med**	**< Med**	**≥ Med**	**No**	**Yes**
Dep Var: Log Hours Worked	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**	**(11)**
Log Avg IRI	-0.011	-0.219	0.043	0.089	-0.086	0.018	-0.034	-0.036	0.042	0.012	0.048
	(0.060)	(0.147)	(0.062)	(0.100)	(0.066)	(0.087)	(0.070)	(0.083)	(0.073)	(0.065)	(0.083)
Adj. R^2	0.329	0.282	0.312	0.318	0.276	0.316	0.352	0.319	0.345	0.323	0.342
Regression F-Stat	3.461	5.349	2.778	1.579	4.117	1.314	3.512	1.737	2.113	1.843	3.102
Kleibergen-Paap F-Stat	32.698	19.528	33.307	31.877	31.713	26.456	30.790	27.691	29.295	26.550	26.547
DV Mean	5.119	5.046	5.140	4.965	5.220	5.117	5.120	5.104	5.135	5.170	5.070
N	22255	3050	19205	8321	13930	13114	9141	11547	9765	9987	11573
N Individuals	8168	1374	6794	3122	5044	4662	3506	4181	3654	3770	4152
N Desa / Kelurahan	321	250	292	262	291	271	275	245	285	272	259
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

All regressions include a constant. Controls include: district GDP, age, education, household size, month of survey indicators. Dependent variables are only defined if the individual reported working. Robust standard errors in parentheses, clustered at the (initial) village level. For the GMM specifications, we report the adjusted R^2 from the analagous reduced-form regression (regressing the dependent variable on the instruments). */**/*** denotes significant at the 10% / 5% / 1% levels.

Table B8: Road Roughness and Working in Sales and Services (Primary): Heterogeneity

		Age		**Gender**		**Education**		**Initial PCE**		**Own Farmland?**	
DV: Working in Sales and Services (Primary) (0 1)	**All Obs (1)**	**$<$ Med (2)**	**$\geq$ Med (3)**	**F (4)**	**M (5)**	**$<$ Med (6)**	**$\geq$ Med (7)**	**$<$ Med (8)**	**$\geq$ Med (9)**	**No (10)**	**Yes (11)**
Log Avg IRI	0.082	0.193	0.084	0.148	0.061	0.095	0.144	0.077	0.114	0.027	0.102
	(0.035)**	(0.087)**	(0.036)**	(0.053)***	(0.041)	(0.044)**	(0.051)***	(0.044)*	(0.054)**	(0.056)	(0.044)**
Adj. R^2	0.534	0.418	0.547	0.618	0.432	0.574	0.479	0.541	0.518	0.519	0.522
Regression F-Stat	6.062	1.353	5.356	4.751	3.993	2.495	8.776	3.612	4.178	7.006	1.773
Kleibergen-Paap Wald rk F-Stat	32.711	19.528	33.321	31.895	31.717	26.469	30.785	27.690	29.354	26.561	26.547
DV Mean	0.279	0.296	0.274	0.377	0.213	0.248	0.316	0.213	0.352	0.360	0.201
N	22259	3050	19209	8323	13932	13117	9142	11548	9768	9991	11573
N Individuals	8169	1374	6795	3123	5044	4663	3506	4181	3655	3771	4152
N Desa / Kelurahan	321	250	292	262	291	271	275	245	285	272	259
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

All regressions include a constant. Controls include: district GDP, age, education, household size, month of survey indicators. Dependent variables are only defined if the individual reported working. Robust standard errors in parentheses, clustered at the (initial) village level. For the GMM specifications, we report the adjusted R^2 from the analagous reduced-form regression (regressing the dependent variable on the instruments). */**/*** denotes significant at the 10% / 5% / 1% levels.

Table B9: Road Roughness and Working in Other Sectors (Primary): Heterogeneity

Dep Var: Working in Other Sectors	**All Obs (1)**	**Age** $<$ **Med (2)**	**Age** $\geq$ **Med (3)**	**Gender F (4)**	**Gender M (5)**	**Education** $<$ **Med (6)**	**Education** $\geq$ **Med (7)**	**Initial PCE** $<$ **Med (8)**	**Initial PCE** $\geq$ **Med (9)**	**Own Farmland? No (10)**	**Own Farmland? Yes (11)**
Log Avg IRI	-0.015	-0.001	-0.015	-0.015	-0.024	-0.002	-0.036	-0.009	-0.022	-0.009	-0.020
	(0.011)	(0.032)	(0.015)	(0.013)	(0.020)	(0.008)	(0.023)	(0.014)	(0.021)	(0.017)	(0.015)
Adj. R^2	0.515	0.175	0.590	0.616	0.428	0.169	0.491	0.433	0.515	0.498	0.545
Regression F-Stat	3.954	6.657	3.132	5.038	3.301	1.370	5.334	1.910	4.190	2.748	3.222
Kleibergen-Paap Wald rk F-Stat	33.594	26.782	34.343	33.836	32.212	26.660	34.965	29.379	29.570	27.925	27.021
DV Mean	0.067	0.041	0.078	0.057	0.078	0.011	0.129	0.029	0.107	0.075	0.059
N	35193	7920	27273	18249	16932	19483	15710	17521	16106	17080	16910
N Individuals	12459	3383	9076	6354	6099	6547	5912	6129	5840	6208	5876
N Desa / Kelurahan	377	301	320	328	297	305	311	276	324	316	278
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

All regressions include a constant. Controls include: district GDP, age, education, household size, month of survey indicators. Dependent variables are only defined if the individual reported working. Robust standard errors in parentheses, clustered at the (initial) village level. For the GMM specifications, we report the adjusted R^2 from the analagous reduced-form regression (regressing the dependent variable on the instruments). */**/*** denotes significant at the 10% / 5% / 1% levels.

Table B10: Road Roughness and Working in Informal Sector (Primary): Heterogeneity

		Age		**Gender**		**Education**		**Initial PCE**		**Own Farmland?**	
DV: Working in Informal Sector	**All Obs (1)**	**$<$ Med (2)**	**$\geq$ Med (3)**	**F (4)**	**M (5)**	**$<$ Med (6)**	**$\geq$ Med (7)**	**$<$ Med (8)**	**$\geq$ Med (9)**	**No (10)**	**Yes (11)**
Log Avg IRI	0.091	-0.114	0.109	0.078	0.097	0.139	0.040	0.119	0.056	0.043	0.101
	(0.032)***	(0.107)	(0.033)***	(0.041)*	(0.046)**	(0.043)***	(0.043)	(0.044)***	(0.042)	(0.045)	(0.040)**
Adj. R^2	0.601	0.439	0.615	0.689	0.548	0.542	0.623	0.571	0.641	0.570	0.588
Regression F-Stat	6.857	2.051	8.569	2.929	6.864	4.411	5.546	4.140	4.109	5.310	5.360
Kleibergen-Paap Wald rk F-Stat	32.707	19.474	33.295	31.991	31.671	26.425	30.925	27.860	29.087	26.569	26.578
DV Mean	0.576	0.455	0.611	0.643	0.531	0.669	0.460	0.621	0.525	0.442	0.703
N	22147	3033	19114	8300	13843	13042	9105	11487	9724	9943	11523
N Individuals	8133	1366	6767	3114	5017	4640	3493	4160	3643	3755	4138
N Desa / Kelurahan	320	249	292	262	290	270	275	244	285	272	258
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

All regressions include a constant. Controls include: district GDP, age, education, household size, month of survey indicators. Dependent variables are only defined if the individual reported working. Robust standard errors in parentheses, clustered at the (initial) village level. For the GMM specifications, we report the adjusted R^2 from the analagous reduced-form regression (regressing the dependent variable on the instruments). */**/*** denotes significant at the 10% / 5% / 1% levels.