Essays in Empirical Microeconomics

by

Raymond P. Guiteras

B.A. Economics and History
Amherst College, 1998

Submitted to the Department of Economics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2008

Signature of Author...
Department of Economics
August 1, 2008

Certified by... ....
Michael Greenstone
3M Professor of Environmental Economics
Thesis Supervisor

Certified by............ ....
Esther Duflo
Abdul Latif Jameel Professor of Poverty Alleviation and Development Economics
Thesis Supervisor

Accepted by................. ....
Peter Temin
Elisha Gray II Professor of Economics
Chairman, Departmental Committee on Graduate Studies

Essays in Empirical Microeconomics

by

Raymond P. Guiteras

Submitted to the Department of Economics on August 1, 2008,

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Economics

ABSTRACT

This thesis consists of three essays addressing open empirical questions in applied microeconomics.

Chapter 1 attempts to quantify the impact of climate change on Indian agriculture. I use historical data on past yearly weather fluctuations and crop yields to measure the effect of these weather fluctuations on output, then use climate change prediction models to derive projections of the impact of future climate change on future productivity. I find that even moderate climate change could be seriously detrimental to productivity, with a consensus prediction for warming over the period 2010-2039 reducing productivity 4.5 to 9 percent.

Chapter 2 provides a new tool for analysis of distributional, or quantile, effects in regression discontinuity (RD) models. RD has become increasingly popular over the last decade as a method of obtaining quasiexperimental estimates of mean treatment effects. This paper extends the methodology to the measurement of quantile treatment effects. I provide simulation evidence on the effectiveness of the estimator and an empirical application to returns to compulsory schooling in the United Kingdom.

Chapter 3, written jointly with Esther Duflo and Michael Greenstone, examines the impact of a water and sanitation intervention in Orissa, India, on health outcomes, in particular the monthly incidence of severe cases of diarrhea and malaria. The design of the intervention, in particular the fact that the water system is activated suddenly, unpredictably and simultaneously for all households in a given village, allow us to overcome several empirical challenges that have impeded credible estimation in the past. We find large effects: the arrival of services appears to reduce severe cases of diarrhea by as much as forty percent, with similar effects on severe cases of malaria. Furthermore, these effects appear to be persistent, as they continue to be apparent in the data after three and even five years.

Thesis Supervisor: Michael Greenstone
Title: 3M Professor of Environmental Economics

Thesis Supervisor: Esther Duflo
Title: Abdul Latif Jameel Professor of Poverty Alleviation and Development Economics

ACKNOWLEDGEMENTS

I am deeply indebted to my thesis supervisors, Michael Greenstone and Esther Duflo, for their advice and guidance throughout graduate school and especially during the writing of this dissertation. Their dedication to outstanding scholarship has been and continues to be inspirational.

Many other members of the MIT faculty have made me a better economist. I would like to recognize in particular Victor Chernozhukov for helping me develop as a student, teaching assistant and researcher and Nancy Rose for her invaluable contributions during the job search process.

My fellow MIT graduate students have been an integral part of my learning experience. There are too many to name, but I am especially grateful to Suman Basu, Jim Berry, Neil Bhutta, Tal Gross, Rick Hornbeck, Jeanne Lafortune, Konrad Menzel, Trang Nguyen, Joe Shapiro, Chris Smith, Jose Tessada, Patrick Warren, Tom Wilkening and Maisy Wong for their friendship and intellectual gifts. I would like to recognize in particular Jessica Cohen and Greg Fischer for helping me to defy the laws of anatomy by simultaneously keeping my chin up and my nose to the grindstone.

I had the special privilege of teaching a wonderful group of graduate students in both econometrics and development economics. In particular, I would like to thank Leopoldo Fergusson, Cynthia Kinnan, Monica Martinez-Bravo, Matt Notowidigdo, Sahar Parsa, Pablo Querubin and Jeremy Shapiro for helping me to become a better teacher and scholar. I am eager to see where your talents will take you and am confident of your future success; I hope it will not be unseemly to take great pride in thinking that I have made a small contribution.

I gratefully acknowledge financial support from the George and Obie Schultz Fund and the National Science Foundation's Graduate Research Fellowship.

Finally, I am thankful beyond words to my family for their love and support throughout these five years. PDGTQM.

PAPER-SPECIFIC ACKNOWLEDGEMENTS

Chapter 1: I thank C. Adam Schlosser of MIT's Center for Energy and Environmental Policy Research for guidance with the NCC data set, Kavi Kumar of the Madras School of Economics for guidance with the World Bank India Agriculture and Climate Data Set and Daniel Sheehan of MIT's Rotch Library for help with GIS programming. Nivedhitha Subramanian and Henry Swift provided excellent research assistance. I thank Daron Acemoglu, Miriam Bruhn, Jessica Cohen, Greg Fischer, Rick Hornbeck, Jeanne LaFortune, Ben Olken, José Tessada, Robert Townsend, Maisy Wong and participants in the MIT Applied Microeconomics Summer Lunch and the Harvard Environmental Economics Lunch for their comments.

Chapter 2: I thank Josh Angrist, David Autor, Jessica Cohen, Victor Chernozukhov, Greg Fischer, Brigham Fransden, Christian Hansen, Amanda Kowalski, Konrad Menzel, Whitney Newey, Joe Shapiro and participants in the MIT Econometrics Lunch and the MIT Summer Applied Microeconomics Lunch for useful comments and suggestions. I especially thank Phil Oreopolous for many helpful suggestions for acquiring, understanding and using the UK GHHS data and for sharing his Stata code.

Chapter 3: We thank Gram Vikas for granting us access to their data and for many discussions on the program's details. In particular, Chitraleka Choudhary, Joe Madiath and Sojan Thomas were extraordinarily generous with their time. Silvia Robles, Shobhini Mukerji, Stefanie Stantcheva, Niveditha Subramanian and, especially, Joe Shapiro provided outstanding research assistance.

Chapter 1:

The Impact of Climate Change on Indian Agriculture

1 Introduction

As the scientific consensus grows that significant climate change, in particular increased temperatures and precipitation, is very likely to occur over the 21st century (Christensen and Hewitson, 2007), economic research has attempted to quantify the possible impacts of climate change on society. Since climate is a direct input into the agricultural production process, the agricultural sector has been a natural focus for research. The focus of most previous empirical studies has been on the US, but vulnerability to climate change may be greater in the developing world, where agriculture typically plays a larger economic role. Credible estimates of the impact of climate change on developing countries, then, are valuable in understanding the distributional effects of climate change as well as the potential benefits of policies to reduce its magnitude or promote adaptation. This paper provides evidence on the impact of climate change on agriculture in India, where poverty and agriculture are both salient. I find that climate change is likely to reduce agricultural yields significantly, and that this damage could be severe unless adaptation to higher temperatures is rapid and complete.

Most previous studies of the economic effects of climate change have followed one of two methodologies, commonly known as the *production function approach* and the *Ricardian approach*. The production function approach (also known as *crop modeling*) is based on controlled agricultural experiments, where specific crops are exposed to varying climates in laboratory-type settings such as greenhouses, and yields are then compared across climates. This approach has the advantage of careful control and randomized application of environmental conditions. However, these laboratory-style outcomes may not reflect the adaptive behavior of optimizing farmers. Some adaptation is modeled, but how well this will correspond to actual farmer behavior is unclear. If farmers' actual practices are more adaptive, the production function approach is likely to produce estimates with a negative bias. On the other hand, if the presumed adaptation overlooks constraints on farmers' adaptations or does not take adjustment costs into account, these estimates could be overoptimistic.

The Ricardian approach, pioneered by Mendelsohn et al. (1994), attempts to allow for the full range of compensatory or mitigating behaviors by performing cross-sectional regressions of land prices on county-level climate variables, plus other controls. If markets are functioning well, land prices will reflect the expected present discounted value of profits from all, fully adapted uses of land, so, in principle, this approach can account for both the direct impact of climate on specific crops as well farmers' adjustment of production techniques, substitutions of different crops and even exit from agriculture. However, the success of the Ricardian approach depends on being able to account fully for all factors correlated with climate and influencing agricultural productivity. Omitted variables, such as unobservable farmer or soil quality, could lead to bias of unknown sign and magnitude. The possibility of omitted variables bias and the inconsistent results obtained from Ricardian studies of climate change in the US have lead to a search for new estimation strategies.

More recently, economists studying the US have turned to a panel data approach, using presumably random year-to-year fluctuations in realized weather across US counties to estimate the effect of weather on agricultural output and profits (Deschênes and Greenstone, 2007; Schlenker and Roberts, 2006). This fixed-effects approach has the advantage of controlling for time-invariant district-level unobservables such as farmer quality or unobservable aspects of soil quality. Furthermore, unlike the production function approach, the use of data on actual field outcomes, rather than outcomes in a laboratory environment, means that estimates from panel data will reflect intra-year adjustments by farmers, such as changes in inputs or cultivation techniques. However, by measuring effects of annual fluctuations, the panel data approach does not reflect the possibility of longer-term adaptations, such as crop switching or exit from farming.

Agriculture typically plays a larger role in developing economies than in the developed world. For example, agriculture in India makes up roughly 20% of GDP and provides nearly

52% of employment (as compared to 1% of GDP and 2% of employment for the US), with the majority of agricultural workers drawn from poorer segments of the population (FAO, 2006). Furthermore, it is reasonable to expect that farmers in developing countries may be less able to adapt to climate change due to credit constraints or less access to adaptation technology. However, the majority of the economics literature on the impact of climate change has focused on developed countries, in particular the US, presumably for reasons of data availability. Most research in developing countries has followed the production function approach, finding alarmingly large possible impacts (Cruz et al., 2007). A true Ricardian study would be difficult to carry out in a developing country context, because land markets are less likely to be well-functioning and data on land prices are not generally available. Instead, a *semi-Ricardian* approach has used data on average profits instead – the idea is that the land price, if it were available, would just be the present discounted value of profits. The major developing country semi-Ricardian studies, of India and Brazil, found significant negative effects, with a moderate long-run climate change scenario (an increase of $2.0°C$ in mean temperature and seven percent increase in precipitation by the end of the 21st century) leading to losses on the order of 10% of agricultural profits (Sanghi et al., 1998b, 1997).

This paper applies the panel data approach to agriculture in India, using a panel of over 200 districts covering 1960-1999.[1] The basic estimation strategy, following Deschênes and Greenstone (2007), is to regress yearly district-level agricultural outcomes (in this case, yields) on yearly climate measures (temperature and precipitation) and district fixed effects. The resulting weather parameter estimates, then, are identified from district-specific deviations in yearly weather from the district mean climate. Since year-to-year fluctuations in the weather are essentially random and therefore independent of other, unobserved determinants of agricultural outcomes, these panel estimates should be free of the omitted variables prob-

[1] Auffhammer et al. (2006) also employ the panel data methodology to study Indian agriculture, rice in particular. Their study uses state-level data on rice output and examines the impact of climate as well as atmospheric brown clouds, the byproduct of emissions of black carbon and other aerosols. They find a negative impact of increased temperature, as does this paper.

lems associated with the hedonic approach. The use of district-level data is important to obtain adequate within-year climate variation, thereby distinguishing climate impacts from other national-level yearly shocks. I also include smooth regional time trends so that the effect of a slowly warming climate over the second half of the twentieth century is not confounded with improvements in agricultural productivity over the same period. The predicted mean impact of climate change is then calculated as a linear combination of the estimated weather parameters and the predicted changes in climate.

The paper finds significant negative impacts, with medium-term (2010-2039) climate change predicted to reduce yields by 4.5 to nine percent, depending on the magnitude and distribution of warming. Long-run climate change (2070-2099) is even more detrimental, with predicted yields falling by 25 percent or more. Because these large changes in long-run temperatures will develop over many decades, farmers will have time to adapt their practices to the new climate, likely lessening the negative impact. However, estimates from this panel data approach may be more relevant for the medium-run scenario, since, as the paper's theoretical section argues, developing country farmers face significant barriers to adaptation, which may prevent rapid and complete adaptation.

This negative impact of climate change on agriculture is likely to have a serious impact on poverty: recent estimates from across developing countries suggest that one percentage point of agricultural GDP growth increases the consumption of the three poorest deciles by four to six percentage points (Ligon and Sadoulet, 2007). The implication is that climate change could significantly slow the pace of poverty reduction in India.

2 Theoretical Framework

Because this paper will attempt to estimate the impacts of climate change based on the effects of annual fluctuations in the weather, it is worthwhile to consider the relationship between the two.

2.1 Short-run weather fluctuations versus long-term climate change

Consider the following simple model of farmer output. A representative farmer's production function is $f(T, L, K)$, where T represents temperature, L represents an input that can be varied in the short run, which we shall call labor for concreteness, and K represents an input that can only be varied in the long run, which we call capital. Labor and capital should not be thought of literally, nor are the distinctions between inputs that are flexible in the short and long run so sharp in reality. The point is that some inputs, such as fertilizer application or labor effort are relatively easy to adjust, while other inputs, such as crop choice or irrigation infrastructure, may be more difficult to adjust or may be effectively fixed at the start of the growing season. The farmer, taking price and temperature as given, solves the following program:

$$\max \{p \cdot f(T, L, K) - wL - rK\}$$

where for simplicity we assume linear input costs. For a given temperature T, with all inputs fully flexible, the farmer will choose profit-maximizing $L(T)$ and $K(T)$ and obtain a maximized profit of $\pi(T, L(T), K(T))$. Now consider a small change in temperature to $T' > T$. First, consider the case where the farmer is not allowed to make any changes, i.e. L and K are held fixed at $L(T)$ and $K(T)$, respectively. In this case, the farmer obtains profit $\pi(T', L(T), K(T))$. To the extent that the production function approach discussed in the introduction understates or ignores the possibility of adaptation, that approach estimates the effect of climate change on profits as $\widehat{\Delta\pi}_{PF} = \pi(T', L(T), K(T)) - \pi(T, L(T), K(T))$.

Next, consider the case where the farmer can carry out short-run adjustments, which in this model we capture as reoptimizing L, but is constrained from long-run adjustments of K. In this case, the farmer obtains $\pi(T', L(T'), K(T))$. The panel data approach followed in this paper, where farmers are free to make all intra-season adjustments but not longer-run adjustments, estimates the effect of climate change as $\widehat{\Delta\pi}_{FE} = \pi\left(T', L\left(T'\right), K\left(T\right)\right) - \pi\left(T, L\left(T\right), K\left(T\right)\right)$.

Finally, consider the case where the farmer is allowed to reoptimize all factors. In this case, the farmer obtains $\pi\left(T', L\left(T'\right), K\left(T'\right)\right)$ and the true effect of climate change is $\Delta\pi = \pi\left(T', L\left(T'\right), K\left(T'\right)\right) - \pi\left(T, L\left(T\right), K\left(T\right)\right)$. Since greater choice can only help the farmer, we have

$$\Delta\pi \geq \widehat{\Delta\pi}_{FE} \geq \widehat{\Delta\pi}_{PF}$$

This framework is illustrated in Figure 1. The first point to note is that the panel data approach should better approximate the true effect of climate change than a production function approach that does not allow for adaptation. The second point is that, for small changes in climate, the panel data approach may provide a reasonable approximation to the true effect of climate change. However, for large changes in climate, the panel data approach will overstate the costs of climate change relative to the true long-run cost, when farmers have re-optimized.

Furthermore, the panel data approach may also provide a reasonable approximation if farmers are unable to reoptimize along some margins or do so only slowly. If long-term reoptimization is slow or incomplete, it is plausible that the panel data approach will provide a good estimate of the costs incurred over the medium run, while not all adjustments have been carried out. There are several reasons to expect that agricultural practice may adapt slowly to climate change. First, the signal of a changing mean climate will be difficult to extract from the year-to-year weather record. The IPCC calculates that a discernible signal of a warmer mean climate for the South Asian growing season will take 10-15 years to emerge from the annual noise (Christensen and Hewitson, 2007). This is for South Asia as a whole, greater noise in particular locations will slow the signal's emergence further. If farmers' practices are based on mean climates, then this difficulty in discerning climate change could lead to farming practices significantly out of phase with the true optimum.

Second, many of the investments associated with long-term reoptimization – new irrigation, new crop varieties, or migration – involve both fixed costs and irreversibilities, both of which can delay investment in the presence of uncertainty (Bertola and Caballero, 1994; Dixit and Pindyck, 1994).

Additionally, it is reasonable to expect developing country agriculture to face even greater difficulties adjusting. Incomplete capital markets, poor transmission of information, and low levels of human capital are all pervasive and likely to slow adaptation. Topalova (2004) provides evidence that factors, especially labor, are relatively immobile in India. Furthermore, slow adaptation of profitable agricultural practices is a long-standing puzzle in the economics literature (Foster and Rosenzweig, 1995; Duflo et al., 2005).

2.2 Caveats

Several important caveats may limit the applicability of the above model. First, data on annual agricultural profits are not available.[2] This paper will use data on annual yields (output per hectare) as a proxy instead and explore the impact on those inputs for which annual data are available. This may overstate the impact on welfare if farmers reduce their use of inputs in response to a negative weather shock. The empirical analysis explores the impacts on those inputs for which yearly, district-level data are available. Second, it is not possible for a panel study to assess the impact of weather on output through its effects on stock inputs. For example, if climate change hurts agriculture by depleting aquifers but one year's drought does not appreciably deplete an aquifer, the panel data approach will not capture this effect. Finally, the panel approach cannot assess the impact of variables that vary only slowly over time. For example, it is believed that the same increased levels of carbon dioxide (CO_2) that are causing global warming may be beneficial to agriculture, since carbon

[2] Sanghi et al. (1998b) use average profits over a 20-year period. Their imputed labor inputs are based on agricultural labor quantities measured by decadal censuses, with linear interpolations for non-census years. This is appropriate for their purpose, which is to asses the relationship between average climate and average profits, but not appropriate for this paper, where the emphasis is on annual fluctuations.

dioxide is important to plant development.[3] Since the level of CO_2 changes only slowly, it is not possible to separate its effect from that of, for example, smooth technological progress over time. However, since CO_2 levels are roughly constant across space, the Ricardian approach is not able to capture this effect either.

3 Data Sources and Summary Statistics

The analysis is performed on a detailed 40-year panel of agricultural outcomes and weather realizations covering over 200 districts. Although Indian districts are generally somewhat larger than US counties, the district is the finest administrative unit for which reliable data are available. This section describes the data and provides some summary statistics.

3.1 Agricultural outcomes

Detailed district-level data from the Indian Ministry of Agriculture and other official sources on yearly agricultural production, output prices and acreage planted and cultivated for 271 districts over the period 1956-1986 have been collected into the "India Agriculture and Climate Data Set" by a World Bank research group, allowing computation of yield (revenues per acre) and total output (Sanghi et al., 1998a). This dataset covers the major agricultural states with the exceptions of Kerala and Assam. Also absent, but less important agriculturally, are the minor states and Union Territories in northeastern India, and the northern states of Himachal Pradesh and Jammu-Kashmir. These 271 districts are shown in Figure 2.A. The production, acreage and price data for major crops were extended through 1999 by Duflo and Pande (2007), allowing computation of yields (output per acre) for these major crops.[4] 218 districts have data for all years 1960-1999; these are the districts that will

[3] Recent research in the crop modelling school has cast doubt on the magnitude of beneficial effects from CO_2 fertilization (Long et al., 2006).

[4] The six major crops are rice, wheat, jowar (sorghum), bajra (millet), maize and sugar. These comprise roughly 75% of total revenues.

be included in the regressions. These districts are mapped in Figure 2.B. The bulk of the districts lost are in the East, in particular Bihar and West Bengal.

Because markets are not well-integrated, local climate shocks could affect local prices. These price effects make estimating effects on revenue undesirable. While the price response to a negative climate shock will reduce the impact on farmers, calculating the effect of climate on revenues will ignore the effect on consumer surplus. In this context, the impact on *yields* better approximates the overall welfare effects, as pointed out by Cline (1992). To avoid these potential pitfalls from endogenous prices, then, I hold prices fixed at their 1960-1965 averages.

The World Bank dataset also includes input measures, such as tractors, plough animals and labor inputs, as well as prices for these inputs. However, many of these inputs, in particular the number of agricultural workers, are only measured at each 10-year census, with annual measures estimated by linear interpolation. This precludes construction of annual profits data, a theoretically preferable measure. This paper will use data on fertilizer inputs, the agricultural wage and the extent of double-cropping, each of which is measured annually at the district level, to estimate the extent of within-year adaptation to negative climate shocks.

3.2 Weather data

Recent research in economics and agricultural science has pointed to the importance of daily fluctuations in temperature for plant growth (Schlenker and Roberts, 2006). Commonly available data, such as mean monthly temperature, will mask these daily fluctuations, so it is important to obtain daily temperature records. Recent economics research in the US has used daily records from weather stations to construct daily temperature histories for US counties. However, the publicly available daily temperature data for India are both sparse and erratic. The main clearinghouse for daily data, the Global Summary of the Day (GSotD, compiled by the US National Climatic Data Center on behalf of the World Meteorological Organization) has at most 90 weather stations reporting on any one day and contains major gaps in the record – for example, there are no records at all from 1963–1972. Furthermore, these individual stations' reports come in only erratically – applying a reasonable sample selection rule such as using stations that report at least 360 days out of the year or 120 days out of the 122 day growing season would yield a database with close to zero observations.

To circumvent this problem, I use data from a gridded daily dataset that use non-public data and sophisticated climate models to construct daily temperature and precipitation records for $1° \times 1°$ grid points (excluding ocean sites). This data set, called NCC (NCEP/NCAR Corrected by CRU), is a product of the Climactic Research Unit, the National Center for Environmental Prediction / National Center for Atmospheric Research and the Laboratoire de Météorologie Dynamique, CNRS. NCC is a global dataset from which Indian and nearby gridpoints were extracted, providing a continuous record of daily weather data for the period 1950-2000 (Ngo-Duc et al., 2005). To create district-level weather records from the grid, I use a weighted average of grid points within 100 KM of the district's geographic center.[5] The weights are the inverse square root of the distance from the district center.

I employ two methods to convert these daily records to yearly weather metrics for analysis. The first, *degree-days*, reflects the importance of cumulative heat over the growing season, but may fail to capture important nonlinear effects. The second, less parametric approach, counts the number of growing-season days in each one-degree C temperature bin. This approach is more flexible, but imposes a perhaps-unrealistic additive separability assumption. However, the results are similar between the two approaches. Details of the methods follow.

[5] Alternative radii did not appreciably affect the district-level records.

3.2.1 Temperature: Degree-days

Agricultural experiments suggest that most major crop plants cannot absorb heat below a temperature threshold of $8°C$, then heat absorption increases roughly linearly up to a threshold of $32°C$, and then plants cannot absorb additional heat above this threshold. I follow the standard practice in agronomics, then, by converting daily mean temperatures to *degree-days* by the formula

$$D(T) = \begin{cases} 0 & \text{if } T \leq 8°C \\ T-8 & \text{if } 8°C < T \leq 32°C \\ 24 & T \geq 32°C \end{cases}$$

Degree-days are then summed over the summer growing season, which for India is defined as the months of June through September, following Kumar et al. (2004). Fixing the growing season avoids endogeneity problems with farmers' planting and harvesting decisions. It should be noted that the degree-day thresholds were developed in the context of US agriculture. Crops cultivated in a warmer climate may have different thresholds, in particular a higher upper threshold. For comparability with other research, I use the standard $8°C$ and $32°C$ thresholds in the empirical results that follow, but the results are not sensitive to the use of alternative upper thresholds $(33°C, 34°C)$. I also allow for the possibility that heat in excess of a threshold may be damaging by including a separate category of *harmful degree-days.* Each day with mean temperatures above $34°C$ is assigned difference between that day's mean temperature and $34°C$; these harmful degree-days are then summed over the growing season. Again, the results are not sensitive to alternate thresholds ($33°C$, $35°C$).

3.2.2 Temperature: One-degree bins

Schlenker and Roberts (2006) emphasize the importance of using daily records in the context of nonlinear temperature effects. Consider the following simple example: imagine that increased temperature is initially beneficial for plants, but then drastically damaging above $30°C$. Consider two pairs of days, the first pair with temperatures of $(30°C, 30°C)$ and the second pair with temperatures of $(29°C, 31°C)$. Although both pairs have the same *mean* temperature, their contributions to growth will be very different, with the second much less beneficial. To capture such potential nonlinearities, I employ a nonparametric approach, counting the total number of growing season days in each one-degree C interval and including these totals as separate regressors. That is, for each grid point g, I construct

$$T_{c,g,y} = \{\#\text{ of growing season days with mean temperature in the interval } ((c-1)\,°C, c°C]\}$$

for year y and for each of $c = 1, \ldots, 50$. To obtain district-level measures from these measures at each grid point, I again take the weighted average of the number of days in that bin for each grid point within 100KM of the district center.

It is important to emphasize that the district-level bins are constructed by averaging over grid point temperature bins rather than constructing bins of district center temperatures. Again, this is necessary to account for potential nonlinear temperature effects. To understand the reasoning, consider the following simplified example of a district center equidistant between two grid points. Suppose these are the only two grid points within 100 KM of the district center. As above, imagine that increased temperature is initially beneficial for plants, but then drastically damaging above $30°C$. Now suppose that one of the two grid points has a mean temperature of $29°C$ every day while the other grid point has a mean temperature of $31°C$. The mean temperature calculated at the district center will be $30°C$ each day, but the bin-by-bin experience of the district as a whole would be better captured by assigning half a day to each of the bins corresponding to $29°C$ and $31°C$. This methodology does lead to districts having fractional number of days in bins, but the total over all bins still sums to 122, the number of days in the growing season, for each district.

The mean number of days in each bin across all districts is plotted in Figure 3. Because of the scarcity of observations above $38°C$ and below $22°C$, each of these will be collected into single bins. The tradeoff here is between precision of estimation (aided by grouping these observations) and estimation of nonlinearities at extreme temperatures.

3.2.3 Precipitation

Precipitation data are summed by month to form total monthly precipitation for each month of the growing season, during which the vast majority of annual precipitation occurs. Including separate monthly measures, rather than merely summing over the growing season, allows the timing of precipitation, in particular the arrival of the annual monsoon, to affect output. To test robustness, I also run regressions with total growing season precipitation.

3.3 Climate change predictions

I compute estimated impacts for three climate change scenarios. First, I examine the short-term (2010-2039) South Asia scenario of the Intergovernmental Panel on Climate Change's latest climate model (Cruz et al., 2007), which is an increase of $0.5°C$ in mean temperature and four percent precipitation for the growing season months of June–September. This scenario corresponds to the "business-as-usual" or highest emissions trajectory, denoted A1F1 in the IPCC literature. However, because most of the short-run component of climate change is believed to be "locked-in", i.e. already determined by past emissions, these short run projections are not very sensitive to the emissions trajectory. For example, the short-term South Asia scenario associated with the lowest future emissions trajectory, denoted B1 in the IPCC literature, differs by less than $0.05°C$ for the growing season months. The impact of this scenario on the distribution of growing season temperatures is plotted in Figure 4.

The IPCC does not report higher moments of predictions for this consensus scenario. However, considering just a mean shift in temperatures would overlook the potentially important effects of the distribution of temperatures, in particular nonlinearities at temperature extremes. Furthermore, this consensus scenario is given as a uniform change across all regions, whereas it is likely that climate change will develop differently across different regions of India. To assess the effects of changing distributions of temperatures and to account for regional differences, I use daily predictions from the Hadley Climate Model 3 (HadCM3) data produced by the British Atmospheric Data Centre for the A1F1 business-as-usual scenario. These predictions are given for points on a 2.5° latitude by 3.5° longitude grid. I calculate the average number of days in each one-degree interval, by region, for the years 1990-1999, 2010-2039 and 2070-2099. The changes in the distribution of temperatures are then applied to the district-level temperature distributions derived from the historical NCC data to obtain district-level changes in temperature distributions.[6] These changes are plotted in Figures 5 and 6. The contrast with the mean-shift scenario of the IPCC is apparent in the greater relative mass in the right tails. Significantly, the *increase* in the mean number of growing-season days with temperatures above $38°C$ is greater than the mean *number* of such days in the historical data: while the average district experienced just 0.4 such days observed per year in the historical data, the mean number of days is expected to increase by nearly 2 for the period 2010-2039 and nearly 10 for the period 2070-2099. Because the effect of these extreme temperatures is only imprecisely estimated, this will add uncertainty to the estimated impacts.

3.4 Summary statistics

Summary statistics of the key variables of interest are presented in Table 1.A. This table presents the sample used in the analysis, covering 1960-1999 and including only the 218 districts with full records of output and yields. Noteworthy points in this table include the

[6]The Hadley data for 1990-1999 display both a higher mean and variance than the NCC data for the same period. Since the estimation of temperature effects is performed with the NCC data, calculating projected impacts using temperature changes based on the Hadley data would not be properly scaled. In the projections, I rescale the level of the Hadley data so that the 1990s means by region match the 1990s NCC data. I also rescale the spread so that the root mean squared errors around each gridpoint's monthly mean match for the 1990s.

high productivity, irrigation and use of high-yield varieties (HYV) of the Northern states (Haryana, Punjab and Uttar Pradesh). Significant poverty reduction, defined relative to state- and sector-specific thresholds for minimum adequate calorie consumption, is also visible, although poverty remains high, especially in the Eastern states. Panel I of Table 1.B compares the 218 districts analyzed in this paper to the sample of 271 districts for the period 1966–1986 studied in Sanghi et al. (1998b) (referred to as SMD98 hereafter). The two samples are very similar. Panel II of Table 1.B looks at the 218 districts over time. Noteworthy trends include the increase in agricultural productivity revealed by increasing yields, the large increase in irrigation an high-yield varieties, and warming (observed in mean temperatures and degree-days).

3.5 Residual variation

Because this paper uses district fixed-effects to strip out time-invariant unobservables that could be confounded with mean climate, it is important to consider how much variation in climate will be left over after these fixed effects and other controls have been removed. This section assess the extent of this residual variation.

3.5.1 Mean temperatures, degree-days and precipitation

Table 2.A reports the results of an exercise designed to assess the extent of residual variation in mean temperatures, degree-days and precipitation. I regress each weather measure on various levels of fixed effects – none, district, district and year, district and region*year, district and state*year. The residual from this regression is a measure of remaining variation. For example, the residual from the regression with no fixed effects is simply the deviation of that district*year observation from the grand mean of the sample, the residual from the regression with district fixed effects is the deviation of that district*year observation from the district mean, etc. I then count how many observations have residuals of absolute value

greater than certain cutoffs – for mean growing season temperature, for example, steps of $0.5°C$ up to $2.5°C$. Ideally, there should be a substantial number of observations with deviations greater than the predicted change in climate. If this is the case, then the effect of weather variation of similar magnitude to the predicted climate change would be identified from the data, rather than from functional form extrapolations.

Unfortunately, the fixed effects do wipe out a great deal of variation. Consider the sixth row of Panel 2, which examines the results for district and year fixed effects for the sample that will be the focus of the regression analysis: the 218 districts with output data for all years 1960-1999. Here, we see that just 15 district*year observations differ from the predicted value – which, in this case, is the district mean plus the deviation of the national mean for that year from the national mean for the sample period – by more than 120 degree-days (which corresponds roughly to a $1.0°C$ mean temperature increase), while no observations differ from the predicted value by more than 180 degree-days.

These findings are less than ideal, since they mean that only a few observations are available to identify even small weather fluctuations. To recapture some of this variation, I retreat from year fixed effects and add smooth time trends (linear, quadratic, cubic) to district fixed effects. This way, I remove possible confounding from correlated trends in temperature and technological progress. If yearly weather fluctuations are indeed random, then in expectation they will be uncorrelated with other economic shocks and therefore the consistency of the estimates will not be affected. Looking at the fifth row of Panel 2, we see that we now have 161 observations differing from the predicted value by more than 120 degree-days. Although this is an improvement relative to the year fixed-effects, there is still not an overwhelming amount of variation: we still have no observations differing from the predicted value by more than 240 degree-days. However, not much variation is lost relative to the district fixed-effects alone (the first row of each panel). In Appendix Table 2, I experiment with alternative upper bounds for the degree-day measure, but this does not revive much variation.

These results should lead to caution in interpreting predicted impacts for large changes in climate, since these will depend on functional form assumptions. However, in the case of precipitation, there is no lack of underlying variation, as is made clear by Panel 3. Estimates of precipitation effects will be well-identified from the data.

3.5.2 Temperature bins

To assess the extent of the residual variation within temperature bins, I calculate the sum of the absolute value of the residuals from a regression of the value of the bin variable on different levels of fixed effects. That is, for each bin $c = < 20, 21, \ldots, 40, > 40$, I estimate

$$T_{c,d,t} = \sum_f FE_f + \varepsilon_{c,d,t}$$

where $\{FE_f\}$ is some set of fixed effects (e.g. none, district, district and year, district and region*year, state*year) and calculate the average value of the absolute residuals,

$$\overline{AVR_c} = \frac{1}{D \times T} \sum_{d,t} |\hat{\varepsilon}_{c,d,t}|$$

I also perform similar calculations for regression models incorporating smooth functions of time rather than year fixed effects, e.g.

$$T_{c,d,t} = \sum_d FE_d + \gamma_1 Y + \gamma_2 Y^2 + \gamma_3 Y^3 + \varepsilon_{c,d,t}$$

The results of these calculations of mean sums of absolute residuals are reported in Table 2.B. Each entry represents the mean across districts and years, so the mean times the number of district-by-year observations (here $218 \times 40 = 8270$) yields the number of observations available to identify the effect of that interval. For example, looking at the fifth row, corresponding to the regression model with district fixed effects and a cubic time trend, there are roughly $0.05 \times 8720 \approx 435$ observations available to identify the extremal bin collecting all days with mean temperatures above $40°C$. Because of the scarcity of observations above $38°C$ and below $22°C$, each of these will be collected into single bins. The tradeoff here is between precision of estimation (aided by grouping these observations) and estimation of nonlinearities at extreme temperatures. The results for the specification of district fixed effects and a cubic year trend are plotted in Figure 7.

4 Econometric Strategy

4.1 Semi-Ricardian method

This section describes the econometric framework used in the *semi-Ricardian approach* of Sanghi et al. (1998b) in order to make clear the difference between that approach and the panel approach considered here. The cross-sectional model is

$$\bar{y}_d = \mathbf{X}_d'\boldsymbol{\beta} + \sum \theta_i f_i \left(\bar{W}_{id}\right) + \varepsilon_d \tag{1}$$

where $\bar{y}_d$ is the mean agricultural outcome of interest for district d, $\mathbf{X}_d$ is a vector of observable district characteristics (such as urbanization, soil quality, etc.), $\bar{W}_{id}$ is a climate variable of interest (temperature, precipitation) and ε_d is the error term. In SDM98, the climate variables are monthly mean temperature and precipitation for the months of January, April, July and October, as well as their squares and within-month interactions. As noted above, SDM98 diverge from the traditional Ricardian or hedonic approach by using an average of profits, output and other flow variables rather than land values in a year, for reasons of data availability. The regressions are weighted by area in cropland in each district, with the motivation being that estimates of output from larger districts will be measured more precisely.

For the coefficients of interest θ_i to be estimated consistently, it is necessary that

$$E\left[f_i\left(\bar{W}_{id}\right)\varepsilon_d \,|\, \mathbf{X}_d\right] = 0$$

for all i. Intuitively, climate must be uncorrelated with unobserved determinants of agricultural productivity, after controlling for observed determinants of agricultural productivity. Note that this requires that the true influence of the $\mathbf{X}_d$ is linear (or, alternatively, that the true, nonlinear relationship has been correctly specified). SMD98 include measures of soil quality, population density and other plausible determinants of agricultural productivity. However, the possibility remains that unobserved determinants of output, such as unobserved soil quality, farmer ability, or even government institutions are correlated with the error term ε_d, which would bias the estimated coefficients $\hat{\theta}_i$ and therefore the imputed impact of climate change.

In the language of the model in Section 2.1, the semi-Ricardian method estimates the impact of a shift in climate from T to T' by comparing observed $\pi\left(T', L\left(T'\right), K\left(T'\right)\right)$ with observed $\pi\left(T, L\left(T\right), K\left(T\right)\right)$, with the observations taking place in two different districts. However, there may be other unobserved components of the profits function, so in truth the semi-Ricardian method would be comparing $\pi\left(T', L\left(T'\right), K\left(T'\right), \tilde{\varepsilon}\right)$ with $\pi\left(T, L\left(T\right), K\left(T\right), \varepsilon\right)$, while the true long-run impact of climate change for the district currently at climate T would be $\pi\left(T', L\left(T'\right), K\left(T'\right), \varepsilon\right) - \pi\left(T, L\left(T\right), K\left(T\right), \varepsilon\right)$.

4.2 Panel approach

This paper follows the panel data approach, estimating

$$y_{dt} = \alpha_d + g_r\left(t\right) + \mathbf{X}_{dt}'\boldsymbol{\beta} + \sum \tilde{\theta}_i f_i\left(W_{idt}\right) + \varepsilon_{dt} \tag{2}$$

There are a number of important differences between equation (2) and equation (1). First, note that the dependent variable, y_{dt}, is a yearly measure rather than an average. In the models estimated below, this is annual yields (output per hectare). Second, the regressors of interest are (functions of) yearly realized weather W_{idt}, rather than climate averages. Third, as discussed in the theory section, the coefficients on short run fluctuations need not be the same as those on long run shifts, i.e. $\tilde{\theta}_i \neq \theta_i$. Finally, the district fixed effects α_d will absorb any district-specific time-invariant determinants of y_{dt}.

The consistency of fixed-effects estimates of $\tilde{\theta}_i$ rests on the following assumption:

$$E\left[f_i\left(W_{idt}\right)\varepsilon_{dt} \,|\, \mathbf{X}_{dt}, \alpha_d, g_r\left(t\right)\right] = 0$$

Intuitively, $\tilde{\theta}_i$ is identified from district-specific deviations in weather about the district averages after controlling for a smooth time trend. This variation is presumed to be orthogonal to unobserved determinants of agricultural outcomes, so it provides a potential solution to the omitted variables bias problems that impede estimation of equation (1).

Because outcomes are likely autocorrelated between years for a given district, I perform feasible generalized least squares (FGLS) estimation of the fixed-effects model, estimating the autocorrelation structure of the dependent variable using the bias-corrected method of Hansen (2007). Examining the residuals from the fixed effects regression reveals that an AR(2) process best fits the data. However, as Hansen (2007) emphasizes, conventional estimation of the parameters of the autocorrelation model are biased in a fixed effects framework, so I compute these parameters using Hansen's bias-corrected method.

While an AR(2) process describes the observed data best, it is unlikely that the true underlying error-generating process is literally AR(2). Therefore, I construct cluster-robust standard errors for the FGLS estimates in the spirit of the Huber-White heteroscedasticity-robust variance-covariance matrix. That is, rather than computing the standard error as $\hat{\sigma}^2\left(\widetilde{X}'\hat{\Omega}^{-1}\widetilde{X}\right)^{-1}$ (where $\widetilde{X}$ denotes the regressors with fixed effects removed), which would be appropriate if the data truly were governed by an AR(2) process, I compute $\left(\widetilde{X}'\hat{\Omega}^{-1}\widetilde{X}\right)^{-1}\hat{W}\left(\widetilde{X}'\hat{\Omega}^{-1}\widetilde{X}\right)^{-1}$, where $\hat{W}$ is the robust sum of squared residuals matrix.[7] This procedure combines the best of both worlds – the FGLS procedure is more efficient than fixed effects estimation alone, because the AR(2) process does approximate the true autocorrelation structure, but the robust standard errors are conservative (Wooldridge, 2003; Hansen, 2007).

5 Results

5.1 Regression results

5.1.1 Modelling temperatures with degree-days

The first set of regressions models temperature using growing season degree-days (and its square) and harmful growing season degree-days. As noted above, this reflects the agronomic emphasis on cumulative heat over the growing season, but may be overly restrictive in its functional form. In particular, I estimate

$$\begin{aligned} y_{dt} &= \alpha_d + g_r(t) + \tilde{\theta}_{DD} GSDD_{dt} + \tilde{\theta}_{DD^2} GSDD_{dt}^2 + \tilde{\theta}_{HDD} HDD_{dt} \\ &\quad + \tilde{\theta}_P P_{dt} + \tilde{\theta}_{P^2} P_{dt}^2 \\ &\quad + \tilde{\theta}_{I_{DD,P}} \left(GSDD_{dt} \cdot P_{dt}\right) + \tilde{\theta}_{I_{HDD,P}} \left(HDD_{dt} \cdot P_{dt}\right) + \varepsilon_{dt} \end{aligned} \tag{3}$$

where the dependent variable is the major crop yield (output per hectare in 2005 US dollars). I include district fixed effects and region-specific cubic time trends. The time trends allow productivity to improve as the climate slowly warms over the latter half of the 20th century, avoiding confounding of temperature warming with technological progress. Ta-

[7]Precisely, $\hat{W} = \sum_{j=1}^{N} \hat{u}_j' \hat{u}_j$, where $\hat{u}_j = \sum_{t=1}^{T} \hat{e}_{jt} \tilde{x}_{jt}^*$. $\hat{e}_{jt}$ is the residual from the FE regression and and $\tilde{x}_{jt}^*$ is the j, t row of the matrix $\hat{\Omega}^{-1/2}\tilde{X}$.

ble 3.A, column (1) reports the results of this FGLS regression. The results are overall as expected: yields are increasing in the linear temperature and precipitation terms, but decreasing in the squares. Harmful degree-days are indeed very harmful, although when considering the magnitude of the coefficient it should be kept in mind that an increase of 100 harmful degree-days would be quite a radical increase in temperatures. Interestingly, the interaction term between degree-days and precipitation is negative, which runs counter to the received agronomic wisdom that extra moisture helps shield plants from extra heat. However, precipitation does appear to shield plants from extreme heat, at least if we take seriously the positive (but insignificant) point estimate of positive interaction between harmful degree-days and precipitation.

In column (2) of Table 3.A, I report results for a regression that includes monthly precipitation (and squares), in an attempt to capture the importance of the timing of precipitation. The estimates on precipitation are generally sensible, with yields increasing in linear precipitation terms and decreasing in their squares. The exception is August precipitation, where the signs are reversed, presumably reflecting the negative impact of a late-arriving monsoon.

5.1.2 Temperatures in nonparametric bins

Because of concerns that the degree-day specification may be overly restrictive, I also estimate models where the regressors are the number of days in each of 20 temperature bins. In particular, I estimate

$$y_{dt} = \alpha_d + g(t) + \sum_c \tilde{\theta}_{T_c} T_{c,dt} + \tilde{\theta}_P P_{dt} + \tilde{\theta}_{P^2} P_{dt}^2 + \varepsilon_{dt} \tag{4}$$

where the temperature bins are $c =< 22, 21, \ldots, 38, > 38$. The $(29, 30]$ bin is omitted as the reference category, so each coefficient $\tilde{\theta}_{T_c}$ represents the impact of an additional day in the bin $(c-1, c]$ compared to a day in the $(29, 30]$ bin. The main functional form restriction this framework imposes on the temperature effects is that the effect is constant within each bin. This seems a reasonable approximation for the interior bins but of course cannot be true for the extremal bins (< 22 and > 38). As above, I run models with aggregate growing season precipitation (and its squares) and with monthly precipitation (and squares). Again, I include district fixed effects and regional cubic time trends.

The coefficient estimates are given in Table 3.B, but are perhaps more easily assessed in graphical form, presented in Figure 8. The clear pattern is that cooler temperatures increase agricultural productivity and warmer temperatures are harmful, relative to the $(29, 30]$ bin. For example, imagine that climate change shifts one day from the P29 bin (temperatures in $(28, 29]$ C) to the P31 bin. Since the estimated coefficients on these bins are 0.013 and -0.082, respectively, the total estimated impact of such a shift would be -0.095. Notably, the effects of the highest temperature bins, while negative, are imprecisely estimated.

This additional flexibility does come with a cost: as is readily visible in equation (4), there is a strong assumption of additive separability. That is, I am implicitly assuming that the marginal effect of, for example, a day in the $(34, 35]$ bin is the same in a relatively warm year as in a relatively cool year. This is unlikely to be true. However, as will be apparent in the predicted impacts, the results from this nonparametric approach are reasonably close to those obtained from the degree-day approach, which does take into account cumulative heat over the growing season.

5.2 Predicted Impact of Climate Change

To incorporate the estimated coefficients into a climate change prediction, I calculate the discrete difference in predicted yields at the projected temperature and precipitation scenario from the predicted yield at the historical mean. That is, in the case of the nonparametric bins, I calculate

$$
\begin{aligned}
\widehat{\Delta y} &= \widehat{y_1} - \widehat{y_0} \\
&= \sum_c \left\{ \hat{\theta}_{T_c} \left(\overline{T_{c,1}} - \overline{T_{c,0}} \right) \right\} \\
&\quad + \sum_m \left\{ \hat{\theta}_{P_m} \left(\overline{P_{m,1}} - \overline{P_{m,0}} \right) + \hat{\theta}_{P_m^2} \left(\overline{P_{m,1}^2} - \overline{P_{m,0}^2} \right) \right\}
\end{aligned}
$$

where $\hat{\theta}_{T_c}$ is the estimated coefficient on temperature bin c, $\hat{\theta}_{P_m}$ on precipitation in month m, and $\hat{\theta}_{P_m}$ on squared month-m precipitation. $\overline{T_{c,1}}$ represents the mean number of days in bin c in the projected climate, $\overline{T_{c,0}}$ the mean in the historical climate, and similarly for the precipitation variables. The calculation is similar for the degree-days approach.

Table 4.A presents results for the degree-days approach, using both total precipitation and monthly precipitation. The underlying regression coefficients are taken from the corresponding columns in Table 3.A. In each case, impacts are estimated for each of three climate change scenarios: the IPCC 2010-2039 consensus A1F1 (business-as-usual) scenario ($+0.5°C$ uniform temperature increase, +4% precipitation increase), the Hadley 2010-2039 A1F1 temperature predictions with +4% precipitation, and the Hadley 2070-2099 A1F1 temperature predictions with +10% precipitation. The aggregate impact is negative for all three scenarios, with mildly positive precipitation effects outweighed by negative temperature effects. Even the moderate IPCC 2010-2099 scenario reduces yields by roughly 4.5%. The Hadley scenarios are even more detrimental. In the medium run (2010-2039), yields are predicted to fall by approximately nine percent, while the long run effect is over 40% of yields. However, this latter estimate is in the absence of long-run adaptation, and therefore likely represents an upper bound on damages.

Table 4.B presents results for the temperature bins approach. The national results, reported in column (1) are broadly similar to those from the degree-days approach: the mild (IPCC) medium-run scenario reduces yields by roughly 4.5 percent, and damages increase for

the Hadley scenarios, emphasizing the importance of the shift into the highest temperature bins. Notably, the long-run Hadley scenario is not nearly as damaging as in the degree-days model, although yields are still predicted to fall by 25 percent. The difference can be attributed to the degree-days model's reliance on functional form: as temperatures increase, the negative coefficient on the quadratic degree-day term pushes yields far down.

To explore potentially heterogeneous impacts, columns (2)–(5) reports results from separate estimates by region. Effects are negative across all regions with the exception of the East, which is very imprecisely estimated due to the small sample size. The estimated long-run effect for the Northwest region is perhaps implausibly large (over 60% of yields), although this region also has a small sample size and this estimate is not very precise.

Table 4.C explores the possibility of heterogenous impacts over time. I split the sample into 1960-1979 and 1980-1999 and run the temperature bins regressions separately. The coefficients on the temperature bins are plotted in Figures 9.A and 9.B. Inspection of these figures reveals the same pattern of beneficial lower temperatures and harmful higher temperatures. overall decline in temperature. As above, these coefficients (along with the coefficients on monthly precipitation) are combined with climate projections to obtain predicted impacts, reported in Table 4.C. For comparison, column (1) re-reports the results for the full sample, 1960-1999. Interestingly, the later period (reported in column (3)) shows greater sensitivity to climate than the earlier period (reported in column (2)), both absolutely and as a percentage of average yields. One possible explanation for this increased vulnerability is the higher prevalence of high-yield varieties (HYVs) in the latter period, as HYVs are believed to provide greater yields on average but are more sensitive to climate fluctuations. Another potential explanation is that temperatures in the second half of the period were somewhat higher. The important message is that technological progress need not reduce climate vulnerability.

5.3 Evidence on Adaptation

Three margins for adaptation can be explored with the available data. First, the application of fertilizer can be adjusted in the face of a harmful weather shock. The fertilizers reported in the data are nitrogen, phosphorus, and potassium, which are aggregated at mean 1960-1965 prices. Column (1) of Table 5 shows the estimated impact of a $1°C$ mean temperature increase on the quantities of fertilizers used per hectare. Fertilizer use falls by roughly 4.5 percent. This suggests that the true welfare impact of a climate shock may be slightly overstated by the effect on yields, since farmers can reduce their input use.

Second, farmers could respond to a harmful shock by planting in the second, winter season. However, column (2) shows that the extent of double-cropping[8] is not significantly affected by a one-degree temperature shock. This margin for adaptation may be limited, at least in the short run.

Finally, although yearly data on labor inputs are not available, yearly wages are available at the district level. If we assume that the temperature affects the agricultural labor market mainly through the channel of labor demand rather than labor supply, the behavior of wages in response to temperature shocks can be informative about the response of labor demand to temperature. Column (3) shows that the wage falls by nearly two percent in response to a one-degree temperature increase.[9] In a full-employment context, we could interpret this as reducing the welfare effect of a climate shock, since farmers use fewer scarce resources, much as with fertilizer. However, given chronic unemployment in India, it is likely that these resources are left unemployed, so the welfare effect is ambiguous. The results of Topalova (2005), demonstrating the slow response of factor quantities in India to shocks, suggest that this effect could be persistent.

[8] As measured by the ratio of gross cropped area to net cropped area.

[9] This is consistent with the effects for rainfall found by Jayachandran (2006).

6 Conclusion

This paper employs a panel data methodology to show that the impact of climate change on Indian agriculture is likely to be negative over the short- to medium-term. The medium-term (2010-2039) impact on yields is estimated to be negative 4.5 to nine percent. Since agriculture makes up roughly 20 percent of India's GDP, this implies a cost of climate change of 1 to 1.8 percent of GDP per year over the medium run. Furthermore, agricultural productivity is particularly important for the well-being of the poor. A back-of-the envelope calculation using the estimate of Ligon and Sadoulet (2007) that each percentage point of agricultural GDP growth increases consumption of the lowest three deciles by four to six percent would imply that climate change could depress consumption among India's poor by at least 18 percent. In the absence of rapid and full adaptation, the consequences of long-run climate change could be even more severe, up to 25 percent of crop yields. The results of this paper pose two important questions for future research. First, what are the factors explaining the difference between these negative consequences for a developing country and the mildly positive results for the U.S. found by Deschênes and Greenstone (2007)? Second, and crucial for the welfare of Indian agriculture, how quickly will developing country farmers be able to adjust their farming practices to adapt to the changing climate and what policies or technologies will enable rapid adaptation?

References

Maximilian Auffhammer, V. Ramanathan, and Jeffrey R. Vincent. Intergrated model shows that atmospheric brown clouds and greenhouse gases have reduced rice harvests in India. *Proceedings of the National Academy of Sciences*, 103(52):19668–19672, December 2006.

Guiseppe Bertola and Ricardo J. Caballero. Irreversibility and aggregate investment. *The Review of Economic Studies*, 61(2):223–246, April 1994. ISSN 0034-6527. URL `http://links.jstor.org/sici?sici=0034-6527%28199404%2961%3A2%3C223%3AIAAI%3E2.0.CO%3B2-W`.

Jens Hesselbjerg Christensen and Bruce Hewitson. Regional climate projections. In *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, chapter 11, pages 847–940. Cambridge University Press, 2007.

William R. Cline. *The Economics of Global Warming.* Peterson Institute for International Economics, 1992.

R.V. Cruz, H. Harasawa, M. Lal, S. Wu, Y. Anokhin, B. Punsalmaa, Y. Honda, M. Jafari, C. Li, and N. Huu Ninh. Asia. In M.L. Parry, O.F. Canziani, J.P. Palutikof, P.J. van der Linden, and C.E. Hanson, editors, *Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovenmental Panel on Climate Change.*, chapter 10, pages 470–506. Cambridge University Press, Cambridge, UK, 2007.

Olivier Deschênes and Michael Greenstone. The economic impacts of climate change: Evidence from agricultural profits and random fluctuations in the weather. *American Economic Review*, 97(1):354–385, March 2007.

Avinash Dixit and Robert S. Pindyck. *Investment Under Uncertainty.* Princeton University Press, 1994.

Esther Duflo and Rohini Pande. Dams. *Quarterly Journal of Economics*, pages 601–646, May 2007.

Esther Duflo, Michael Kremer, and Jonathan Robinson. Understanding fertilzer adoption: Evidence from field experiments. MIT mimeo, 2005.

FAO. *FAO Statistical Yearbook 2005–2006.* Food and Agricultural Organization, 2006. URL `http://www.fao.org/statistics/yearbook/`.

Andrew D. Foster and Mark R. Rosenzweig. Learning by doing and learning from others: Human capital and technical change in agriculture. *Journal of Political Economy*, 103 (6):1176–1209, December 1995. URL `http://links.jstor.org/sici?sici=0022-3808%28199512%29103%3A6%3C1176%3ALBDALF%3E2.0.CO%3B2-5`.

Christian Hansen. Generalized least squares inference in panel and multilevel models with serial correlation and fixed effects. *Journal of Econometrics*, 140:670–694, 2007.

Seema Jayachandran. Selling labor low: Wage responses to productivity shocks in developing countries. *Journal of Political Economy*, 114(3):538–575, 2006.

K. Krishna Kumar, K. Rupa Kumar, R.G. Ashrit, N.R. Deshpande, and J.W. Hansen. Climate impacts on Indian agriculture. *International Journal of Climatology*, 24:1375–1393, 2004. doi: 10.1002/joc.1081.

Ethan Ligon and Elizabeth Sadoulet. Estimating the effects of aggregate agricultural growth in the distribution of expenditures. Background paper for the World Development Report 2008, 2007.

Stephen P. Long, Elizabeth A. Ainsworth, Andrew D. B. Leakey, Josef Nosberger, and Donald R. Ort. Food for thought: Lower-than-expected crop yield stimulation with rising CO2 concentrations. *Science*, 312:1918–1921, June 2006. doi: 10.1126/science.1114722.

Robert Mendelsohn, William D. Nordhaus, and Daigee Shaw. The impact of global warming on agriculture: A Ricardian analysis. *American Economic Review*, 84(4):753–771, 1994.

T. Ngo-Duc, J. Polcher, and K. Laval. A 53-year forcing data set for land surface models. *Journal of Geophysical Research*, 110(D06116):1–16, March 2005. doi: 10.1029/2004JD005434.

Apruva Sanghi, D. Alves, R. Evenson, and R. Mendelsohn. Global warming impacts on Brazilian agriculture: Estimates of the Ricardian model. *Economia Aplicada*, 1997.

Apurva Sanghi, K.S. Kavi Kumar, and James W. McKinsey Jr. India agriculture and climate dataset. Technical report, World Bank, 1998a. Compiled and used in the study "Measuring the Impact of Climate Change on Indian Agriculture," World Bank Technical Paper No. 402. Funded jointly by the World Bank Research Budget and by the Electric Power Research Institute in Palo Alto, California. Available online at the UCLA / BREAD website.

Apurva Sanghi, Robert Mendelsohn, and Ariel Dinar. The climate sensitivity of Indian agriculture. In *Measuring the Impact of Climate Change on Indian Agriculture*, chapter 4, pages 69–139. World Bank, 1998b.

Wolfram Schlenker and Michael J. Roberts. Estimating the impact of climate change on crop yields: The importance of non-linear temperature effects. URL `http://www.columbia.edu/~ws2162/SchlenkerRoberts.pdf`. Working paper, September 2006.

Petia Topalova. Factor immobility and regional effects of trade liberalization: Evidence from India. MIT mimeo, 2004.

Petia Topalova. Trade liberalization, poverty and inequality: evidence from Indian districts. URL http://www.nber.org/papers/w11614. NBER Working Paper #11614, Sep 2005.

Jeffrey Wooldridge. Cluster-sample methods in applied econometrics. *The American Economic Review*, 93(2):133–188, 2003.

Table 1.A: Descriptive Statistics

Variable	Units	All	North	NorthWest	East	South
Output	2005 USD (000)	4,564.2	7,580.2	2,289.0	4,098.6	3,682.8
		(4,813.8)	(6,860.9)	(2,567.6)	(2,396.3)	(3,092.3)
Yield (output per hectare)	2005 USD	15.2	19.6	10.4	10.9	14.8
		(11.0)	(11.5)	(9.0)	(4.8)	(10.9)
Mean temperature (growing season)	Deg. C	28.3	29.6	29.1	27.8	27.3
		(2.3)	(2.3)	(2.0)	(1.4)	(2.0)
Degree-days (growing season)	Deg. C	2,438.4	2,577.2	2,526.2	2,403.5	2,336.2
		(253.6)	(255.6)	(220.7)	(166.3)	(220.8)
Precipitation (growing season)	mm	760.8	791.2	682.8	996.1	746.0
		(309.1)	(279.3)	(273.7)	(197.7)	(330.9)
Share of cropland irrigated		0.28	0.48	0.21	0.21	0.20
		(0.23)	(0.22)	(0.12)	(0.16)	(0.19)
Share of cropland HYV		0.23	0.31	0.16	0.16	0.21
		(0.19)	(0.22)	(0.13)	(0.12)	(0.19)
Share below poverty line (1973)		0.45	0.36	0.42	0.70	0.49
		(0.14)	(0.11)	(0.16)	(0.14)	(0.10)
Share below poverty line (1999)		0.23	0.15	0.17	0.42	0.27
		(0.12)	(0.10)	(0.08)	(0.15)	(0.10)
Number of districts		218	61	36	11	110
Number of observations		8,720	2,440	1,440	440	4,400

Notes: Regression sample: 1960-1999, 218 districts with output data for all years 1960-1999. Regions defined as: North (Haryana, Punjab and Uttar Pradesh); Northwest (Gujarat, Rajasthan); East (Bihar, Orissa, West Bengal); South (Andhra Pradesh, Karnakata, Madhya Pradesh, Maharastra, Tamil Nadu). Standard deviations in parentheses.

Table 1.B: Supplemental Descriptive Statistics

Panel I: Comparison of Regression Sample with World Bank Sample

Variable	Units	Regression Sample, 1960-1999	World Bank Sample, 1966-1986	Regression Sample, 1966-1986
Output	2005 USD (000)	4,564.2	4,296.8	4,245.7
		(4,813.8)	(3,806.1)	(3,912.0)
Yield (output per hectare)	2005 USD	15.2	13.3	13.7
		(11.0)	(8.3)	(8.6)
Mean temperature (growing season)	Deg. C	28.3	28.0	28.0
		(2.3)	(2.6)	(2.6)
Degree-days (growing season)	Deg. C	2,438.4	2,408.2	2,404.9
		(253.6)	(294.3)	(292.3)
Precipitation (growing season)	mm	760.8	801.9	754.2
		(309.1)	(351.8)	(315.8)
Share of cropland irrigated		0.28	0.26	0.28
		(0.23)	(0.23)	(0.23)
Share of cropland HYV		0.23	0.22	0.22
		(0.19)	(0.19)	(0.19)
Share below poverty line (1973)		0.45	0.47	0.45
		(0.14)	(0.15)	(0.14)
Share below poverty line (1999)		0.23		
		(0.12)		
Number of districts		218	271	218
Number of observations		8,720	5,670	4578

Panel II: Regression Sample By Decade

Variable	Units	1960-1969	1970-1979	1980-1989	1990-1999
Output	2005 USD (000)	2,777.3	3,956.2	5,365.1	6,158.1
		(2,110.4)	(3,346.6)	(5,025.6)	(6,713.1)
Yield (output per hectare)	2005 USD	10.2	12.9	17.3	20.4
		(5.8)	(7.6)	(10.6)	(14.8)
Mean temperature (growing season)	Deg. C	27.6	28.3	28.6	28.6
		(3.3)	(1.8)	(1.8)	(1.8)
Degree-days (growing season)	Deg. C	2,356.4	2,443.5	2,476.5	2,477.3
		(368.8)	(196.0)	(192.6)	(190.0)
Precipitation (growing season)	mm	734.2	788.6	766.5	754.1
		(327.2)	(302.6)	(308.9)	(294.6)
Share of cropland irrigated		0.22	0.27	0.32	
		(0.19)	(0.21)	(0.25)	
Share of cropland HYV		0.05	0.20	0.35	
		(0.07)	(0.16)	(0.19)	
Share below poverty line			0.45	0.39	0.27
			(0.14)	(0.17)	(0.15)
Number of districts		218	218	218	218
Number of observations		2,180	2,180	2,180	2,180

Notes: Regression sample includes all 218 districts with output data for all years 1960-1999. Word Bank sample includes all 271 districts of the Sanghi, Mendelsohn and Dinar (1998) World Bank study. Standard deviations in parentheses.

Table 2.A: Residual Variation in District Weather Variables

Panel 1: Growing Season Mean Temperatures (C)

Mean: 28.5; N:8720		District*year observations differing from predicted value by more than									
		0.5 deg C		1.0 deg C		1.5 deg C		2.0 deg C		2.5 deg C	
Regressors	RMSE	Number	Share	Number	Share	Number	Share	Number	Share	Number	Share
Constant only	1.81	6853	0.786	5200	0.596	3619	0.415	2214	0.254	1230	0.141
District FEs	0.50	2624	0.301	358	0.041	69	0.008	6	0.001	0	0.000
District FEs, Linear Year	0.49	2551	0.293	330	0.038	46	0.005	4	0.000	0	0.000
District FEs, Quadratic Year	0.49	2557	0.293	303	0.035	53	0.006	5	0.001	0	0.000
District FEs, Cubic Year	0.49	2531	0.290	298	0.034	44	0.005	4	0.000	0	0.000
District and Year FEs	0.33	1011	0.116	32	0.004	0	0.000	0	0.000	0	0.000
District and Year*Region FEs	0.26	517	0.059	9	0.001	0	0.000	0	0.000	0	0.000
District and Year*State FEs	0.20	174	0.020	3	0.000	0	0.000	0	0.000	0	0.000

Panel 2: Growing Season Degree-Days (C)

Mean: 2,464.6; N:8720		District*year observations differing from predicted value by more than									
		60 deg-days (C)		120 deg-days (C)		180 deg-days (C)		240 deg-days (C)		300 deg-days (C)	
Regressors	RMSE	Number	Share	Number	Share	Number	Share	Number	Share	Number	Share
Constant only	194.95	6752	0.774	4914	0.564	3034	0.348	1441	0.165	867	0.099
District FEs	50.47	1926	0.221	189	0.022	25	0.003	0	0.000	0	0.000
District FEs, Linear Year	49.32	1783	0.204	168	0.019	20	0.002	0	0.000	0	0.000
District FEs, Quadratic Year	48.92	1735	0.199	168	0.019	22	0.003	0	0.000	0	0.000
District FEs, Cubic Year	48.89	1751	0.201	161	0.018	21	0.002	0	0.000	0	0.000
District and Year FEs	33.89	617	0.071	15	0.002	0	0.000	0	0.000	0	0.000
District and Year*Region FEs	27.27	323	0.037	3	0.000	0	0.000	0	0.000	0	0.000
District and Year*State FEs	20.49	85	0.010	2	0.000	0	0.000	0	0.000	0	0.000

Panel 3: Growing Season Precipitation (mm)

Mean: 775.0; N:8720		District*year observations differing from predicted value by more than									
		2 percent		4 percent		6 percent		8 percent		10 percent	
Regressors	RMSE	Number	Share	Number	Share	Number	Share	Number	Share	Number	Share
Constant only	302.47	8403	0.964	8058	0.924	7721	0.885	7382	0.847	7028	0.806
District FEs	183.38	8093	0.928	7439	0.853	6775	0.777	6220	0.713	5638	0.647
District FEs, Linear Year	182.15	8091	0.928	7473	0.857	6850	0.786	6219	0.713	5631	0.646
District FEs, Quadratic Year	182.15	8096	0.928	7462	0.856	6853	0.786	6222	0.714	5635	0.646
District FEs, Cubic Year	182.14	8091	0.928	7464	0.856	6851	0.786	6225	0.714	5637	0.646
District and Year FEs	149.57	8003	0.918	7228	0.829	6517	0.747	5802	0.665	5125	0.588
District and Year*Region FEs	133.05	7785	0.893	6912	0.793	6009	0.689	5229	0.600	4460	0.511
District and Year*State FEs	105.00	7382	0.847	6094	0.699	4952	0.568	3970	0.455	3232	0.371

Notes: Table counts residuals from regressions of district*year observations on regressors listed in row headings. Cell entries are number of residuals of absolute value greater than or equal to the cutoffs given in the column headings.Years: 1960-1999; Sample: 218 districts with output data for all years.

Table 2.B: Residual Variation in District Temperature Bins

Regressor(s)	<20	21	22	23	24	25	26	27	28	29	30
Constant	0.30	0.06	0.13	0.85	2.89	6.08	12.16	18.77	19.47	18.13	13.68
District FEs	0.04	0.03	0.09	0.42	0.92	2.24	4.08	4.85	5.65	4.92	4.06
District FEs, Linear Year	0.06	0.04	0.11	0.50	1.05	2.32	4.15	4.85	5.64	4.89	4.06
District FEs, Quadratic Year	0.06	0.04	0.11	0.52	1.05	2.32	4.15	4.84	5.63	4.90	4.05
District FEs, Cubic Year	0.06	0.04	0.11	0.52	1.05	2.31	4.16	4.84	5.64	4.90	4.05
District and Year FEs	0.07	0.04	0.14	0.61	1.12	2.43	4.19	4.68	5.48	4.79	4.01
District and Region*Year FEs	0.07	0.04	0.15	0.62	1.09	2.21	3.72	4.41	4.89	4.24	3.28
District and State*Year FEs	0.07	0.05	0.13	0.47	0.89	2.04	3.26	4.07	4.41	3.89	2.93

Regressor(s)	31	32	33	34	35	36	37	38	39	40	>40
Constant	7.48	5.20	4.33	3.91	3.35	2.51	1.54	0.76	0.30	0.09	0.03
District FEs	2.70	2.02	1.77	1.59	1.40	1.22	1.00	0.68	0.36	0.14	0.04
District FEs, Linear Year	2.70	2.02	1.77	1.59	1.40	1.22	1.01	0.69	0.36	0.15	0.05
District FEs, Quadratic Year	2.69	2.02	1.77	1.60	1.40	1.22	1.01	0.69	0.36	0.15	0.05
District FEs, Cubic Year	2.69	2.02	1.77	1.60	1.41	1.23	1.02	0.71	0.37	0.15	0.05
District and Year FEs	2.56	1.93	1.67	1.51	1.34	1.17	0.98	0.70	0.39	0.17	0.06
District and Region*Year FEs	2.22	1.75	1.54	1.41	1.24	1.08	0.90	0.64	0.37	0.16	0.06
District and State*Year FEs	1.99	1.54	1.35	1.25	1.09	0.93	0.74	0.53	0.32	0.14	0.05

	Alternative Extremal Bins			
Regressor(s)	<21	<22	>38	>39
Constant	0.36	0.49	0.42	0.12
District FEs	0.05	0.11	0.51	0.18
District FEs, Linear Year	0.05	0.12	0.52	0.19
District FEs, Quadratic Year	0.06	0.13	0.52	0.19
District FEs, Cubic Year	0.06	0.13	0.53	0.19
District and Year FEs	0.07	0.17	0.57	0.21
District and Region*Year FEs	0.07	0.18	0.53	0.20
District and State*Year FEs	0.08	0.17	0.45	0.18

Notes: This table assesses the extent of residual variation available after removing district fixed effects and other controls. For each bin, the number of days in that bin is regressed on the controls given in the row heading. The absolute value of the residual is then averaged over all district*year observations. The result can be interpreted as the mean number of days per district*year available to identify the effect of that bin. Years: 1960-1999; Sample: 218 districts with output and yield data for all years 1960-1999 (8720 total year*district observations)

Table 3.A: FGLS Estimates of Weather Variables' Effects on Major Crop Yields

	(1)	(2)
Growing Season Degree-days (100, C)	5.418	3.536
	(2.325)	(2.310)
GSDD Squared	-0.125	-0.094
	(0.047)	(0.048)
Harmful GSDD (100, C) with threshold 34	-3.508	-2.687
	(1.540)	(0.706)
Total Growing Season Precipitation (100 mm)	1.620	
	(0.471)	
TotalGrowSeasonPrecip Squared	-0.021	
	(0.007)	
GrowSeasonDegreeDays*TotalGrowSeasonPrecip	-0.048	
	(0.017)	
HarmfulGSDD34*TotalGrowSeasonPrecip	0.068	
	(0.173)	
June Precipitation (100 mm)		0.520
		(0.216)
June Precipitation Squared		-0.034
		(0.049)
July Precipitation (100 mm)		0.272
		(0.206)
July Precipitation Squared		-0.061
		(0.027)
August Precipitation (100 mm)		-0.450
		(0.207)
August Precipitation Squared		0.050
		(0.030)
September Precipitation (100 mm)		0.533
		(0.234)
September Precipitation Squared		-0.010
		(0.051)
N	8,720	8,720

Notes: Dependent variable: major crop yields (2005 USD / HA). Regressions include district fixed effects and region*year cubic time trends (coefficients not reported). Years: 1960-1999. Sample: 218 districts with output data for all years. FGLS estimator uses bias-corrected AR(2) parameter estimates; standard errors are calculated from the robust variance-covariance matrix.

Table 3.B: FGLS Estimates of Effect of Days in One-Degree (C) Temperature Bins on Major Crop Yields

	Total GS Precipitation	Monthly GS Precipitation
	(1)	(2)
Days in <=22 bin	0.192	0.126
	(0.120)	(0.121)
Days in P23 bin	0.081	0.094
	(0.061)	(0.060)
Days in P24 bin	0.032	0.026
	(0.037)	(0.037)
Days in P25 bin	0.040	0.038
	(0.019)	(0.018)
Days in P26 bin	0.037	0.028
	(0.014)	(0.014)
Days in P27 bin	0.031	0.025
	(0.015)	(0.014)
Days in P28 bin	0.006	-0.001
	(0.015)	(0.015)
Days in P29 bin	0.017	0.013
	(0.017)	(0.017)
Days in P30 bin (omitted category)	-	-
	-	-
Days in P31 bin	-0.073	-0.082
	(0.041)	(0.040)
Days in P32 bin	-0.001	-0.013
	(0.041)	(0.043)
Days in P33 bin	0.012	0.004
	(0.047)	(0.047)
Days in P34 bin	-0.091	-0.082
	(0.037)	(0.038)
Days in P35 bin	0.040	0.044
	(0.041)	(0.042)
Days in P36 bin	-0.114	-0.123
	(0.045)	(0.046)
Days in P37 bin	-0.021	-0.014
	(0.059)	(0.060)
Days in P38 bin	-0.225	-0.209
	(0.132)	(0.133)
Days in >38 bin	-0.092	-0.092
	(0.076)	(0.076)
Total Growing Season Precipitation (100 mm)	0.387	
	(0.144)	
Growing Season Precipitation Squared	-0.016	
	(0.007)	
June Precipitation (100mm)		0.602
		(0.253)
June Precipitation Squared		-0.050
		(0.055)
July Precipitation (100mm)		0.361
		(0.205)
July Precipitation Squared		-0.075
		(0.028)
August Precipitation (100mm)		-0.441
		(0.217)
August Precipitation Squared		0.050
		(0.032)
September Precipitation (100mm)		0.614
		(0.235)
September Precipitation Squared		-0.030
		(0.051)
N	8720	8720

Notes: Dependent variable: major crop yields (2005 USD / HA). Robust FGLS standard errors in parentheses. Each bin is identified as its upper limit (e.g. P35 includes temperatures in (34,35] C). FGLS regressions include district fixed effects and a cubic time trend (coefficients not reported). Years: 1960-1999. Sample: 218 districts with output data for all years.

Table 4.A: Projected Impact of Climate Change on Major Crop Yields from Aggregate Weather Regressions

	Total GS Precipitation (1)	Monthly GS Precipitation (2)
Mean of Dependent Variable	15.215	15.215
Number of Observations	8,720	8,720
Panel A: IPCC Medium-Run (2010-2039) S. Asia Scenario (Uniform +0.5 deg C, +4% precipitation)		
Temperature Effect	-0.566	-0.691
	(0.100)	(0.073)
Precipitation Effect	0.373	0.026
	(0.122)	(0.011)
Interaction Effect	-0.520	
	(0.180)	
Total Effect	-0.712	-0.665
	(0.076)	(0.079)
Panel B: Hadley A1F1 Medium-Run (2010-2039) Scenario		
Temperature Effect	-1.358	-1.414
	(0.315)	(0.165)
Precipitation Effect	0.373	0.026
	(0.122)	(0.011)
Interaction Effect	-0.514	
	(0.308)	
Total Effect	-1.498	-1.387
	(0.168)	(0.171)
Panel C: Hadley A1F1 Long-Run (2070-2099) Scenario		
Temperature Effect	-6.755	-6.506
	(1.608)	(0.847)
Precipitation Effect	0.926	0.064
	(0.303)	(0.026)
Interaction Effect	-1.329	
	(1.453)	
Total Effect	-7.158	-6.441
	(0.870)	(0.860)

Notes: Projections are calculated as the discrete difference in yields (output per hectare) at the projected climate versus the historical climate. Coefficients are obtained from bias-corrected FGLS regressions of yields on growing season weather variables, regional cubic time trends and district fixed effects, weighted by area cropped. Weather variables in column (1) are growing-season degree-days, its square, harmful growing season degree days, total growing season precipitation, its square and the interaction of precipitation with growing-season degree-days and harmful degree-days. Column (2) substitutes monthly precipitation (and squares) for aggregate precipitation, and drops the interactions. Sample: 218 districts with output data for all years 1960-1999.

Table 4.B: Projected Impact of Climate Change on Major Crop Yields from Bins Regressions

	National	North	Northwest	East	South
	(1)	(2)	(3)	(4)	(5)
Mean of Dependent Variable	15.215	19.555	10.436	10.858	14.809
Number of Observations	8720	2440	1440	440	4400
Panel A: IPCC Medium-Run (2010-2039) S. Asia Scenario (Uniform +0.5 deg C, +4% precipitation)					
Temperature Effect	-0.727	-1.467	-0.888	-0.256	-0.452
	(0.065)	(0.201)	(0.169)	(0.207)	(0.080)
Precipitation Effect	0.029	0.059	-0.015	-0.070	0.037
	(0.011)	(0.020)	(0.026)	(0.066)	(0.012)
Total Effect	-0.699	-1.408	-0.903	-0.326	-0.415
	(0.072)	(0.208)	(0.193)	(0.249)	(0.087)
Panel B: Hadley A1F1 Medium-Run (2010-2039) Scenario					
Temperature Effect	-1.225	-1.479	-0.871	0.255	-1.676
	(0.298)	(1.389)	(0.399)	(0.330)	(0.447)
Precipitation Effect	0.029	0.059	-0.015	-0.070	0.037
	(0.011)	(0.020)	(0.026)	(0.066)	(0.012)
Total Effect	-1.196	-1.420	-0.886	0.185	-1.639
	(0.301)	(1.395)	(0.417)	(0.348)	(0.451)
Panel C: Hadley A1F1 Long-Run (2070-2099) Scenario					
Temperature Effect	-3.983	-4.514	-6.624	2.276	-2.615
	(0.919)	(4.643)	(1.667)	(2.926)	(0.986)
Precipitation Effect	0.070	0.129	-0.046	-0.160	0.093
	(0.026)	(0.049)	(0.065)	(0.175)	(0.028)
Total Effect	-3.913	-4.385	-6.670	2.116	-2.521
	(0.924)	(4.659)	(1.693)	(3.075)	(0.999)

Notes: Projections are calculated as the discrete difference in yields (output per hectare) at the projected climate versus the historical climate. Coefficients are obtained from bias-corrected FGLS regressions of yields on growing season days in one-degree (C) temperature bins, monthly precipitation (and squares), regional cubic time trends and district fixed effects, weighted by area cropped. Sample: 218 districts with output data for all years 1960-1999.

Table 4.C: Comparing Climate Resiliency Over Time: 1960-1979 vs. 1980-1999

	1960-1999	1960-1979	1980-1999
	(1)	(2)	(3)
Mean of Dependent Variable	15.215	11.567	18.863
Number of Observations	8720	4360	4360
Panel A: IPCC Medium-Run (2010-2039) S. Asia Scenario (Uniform +0.5 deg C, +4% precipitation)			
Temperature Effect	-0.727	-0.242	-1.189
	(0.065)	(0.042)	(0.127)
Precipitation Effect	0.029	0.051	-0.006
	(0.011)	(0.006)	(0.019)
Total Effect	-0.699	-0.191	-1.195
	(0.072)	(0.046)	(0.137)
Panel B: Hadley A1F1 Medium-Run (2010-2039) Scenario			
Temperature Effect	-1.225	-0.299	-2.034
	(0.298)	(0.139)	(0.583)
Precipitation Effect	0.029	0.051	-0.006
	(0.011)	(0.006)	(0.019)
Total Effect	-1.196	-0.247	-2.040
	(0.301)	(0.144)	(0.584)
Panel C: Hadley A1F1 Long-Run (2070-2099) Scenario			
Temperature Effect	-3.983	-0.496	-7.007
	(0.919)	(0.457)	(1.694)
Precipitation Effect	0.070	0.123	-0.020
	(0.026)	(0.015)	(0.046)
Total Effect	-3.913	-0.372	-7.026
	(0.924)	(0.466)	(1.695)

Notes: Projections are calculated as the discrete difference in yields (output per hectare) at the projected climate versus the historical climate. Coefficients are obtained from bias-corrected FGLS regressions of yields on growing season days in one-degree (C) temperature bins, monthly precipitation (and squares), regional cubic time trends and district fixed effects, weighted by area cropped. Sample: 218 districts with output data for all years 1960-1999.

Table 5: Projected Impact of Climate Change on Major Crop Yields

	Impact of Uniform One-Degree (C) Temperature Increase On:		
	Fertilizer Use	Double-Cropping	Agricultural Wage
	(1)	(2)	(3)
Effect	-5.556	-0.002	-0.118
	(0.828)	(0.001)	(0.008)
Mean of Dependent Variable	125.3	1.2	7.0
N	7588	7570	7588

Notes: Projections are calculated as the discrete difference in yields (output per hectare) at the projected climate versus the historical climate. Coefficients are obtained from bias-corrected FGLS regressions of yields on growing season days in one-degree (C) temperature bins, monthly precipitation (and squares), regional cubic time trends and district fixed effects. Sample: all 271 districts, 1960-1987.

Figure 1: Impact of Climate Change With Various Degrees of Adaptation

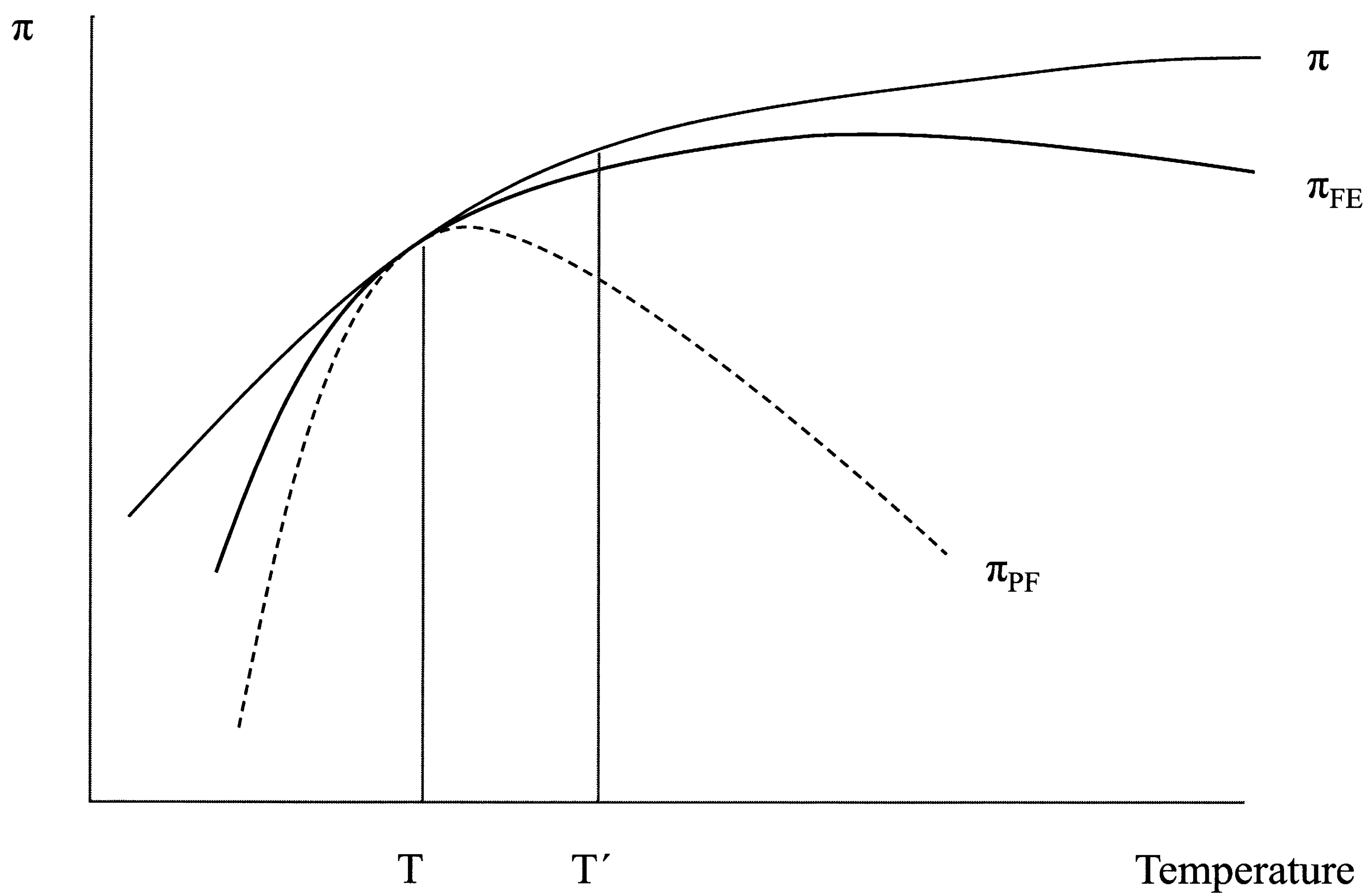
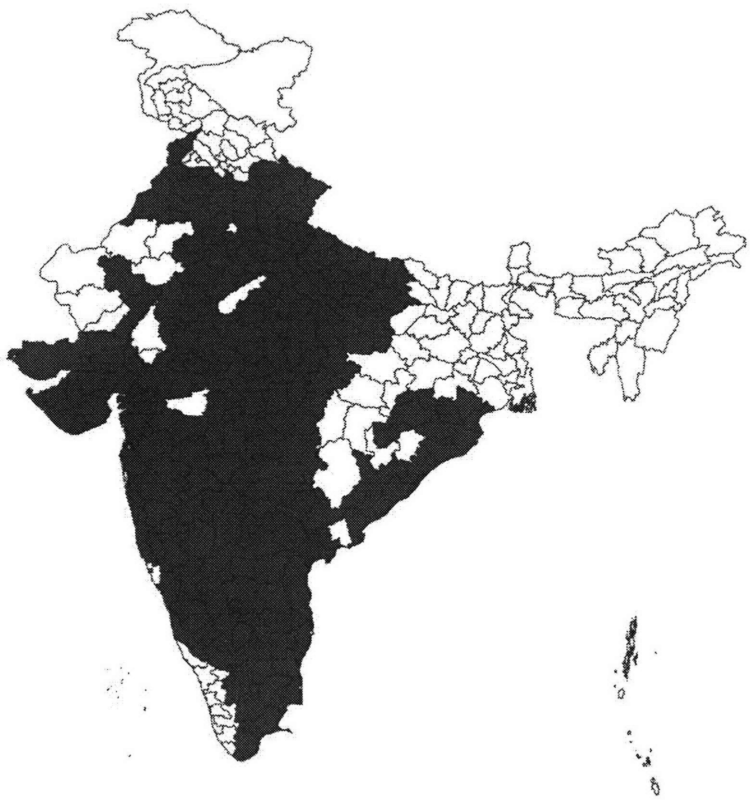


Figure 2.A: Districts Included in SMD98 Study

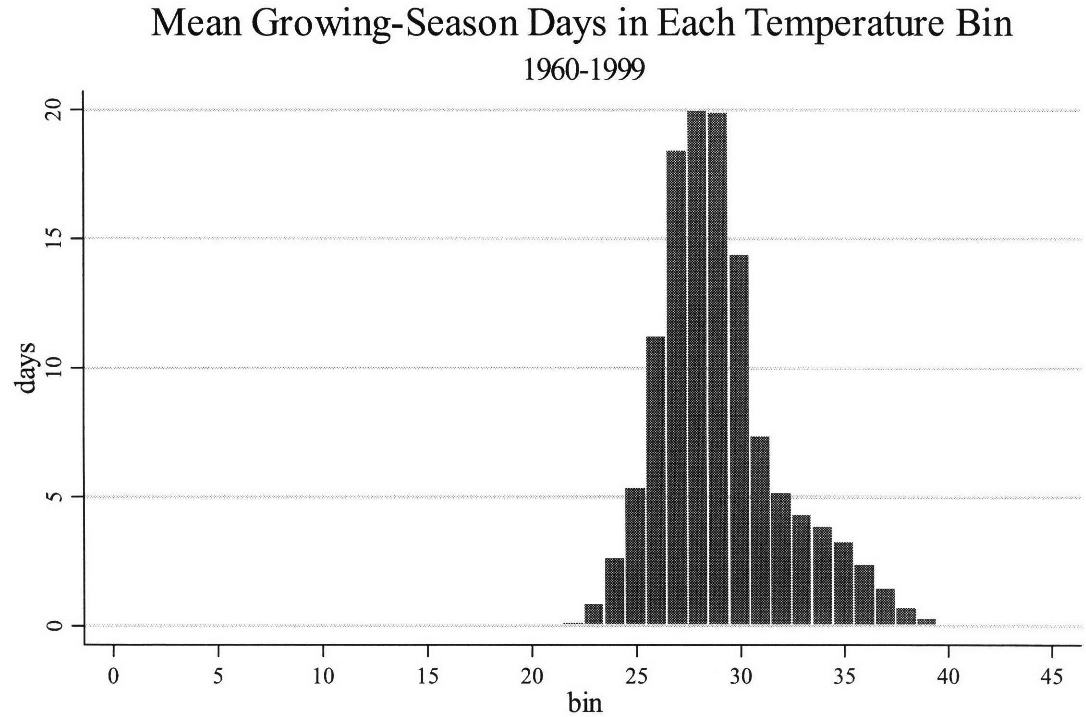
Notes: This map shows the 271 districts included in the Sanghi, Mendelsohn and Dinar 1998 study. The states included are: Haryana, Punjab and Uttar Pradesh (North); Gujarat, Rajasthan (Northwest); Bihar, Orissa, West Bengal (East); Andhra Pradesh, Karnakata, Madhya Pradesh, Maharastra, Tamil Nadu (South). The major agricultural state excluded is Kerala.

Figure 2.B: Districts Included in Regressions



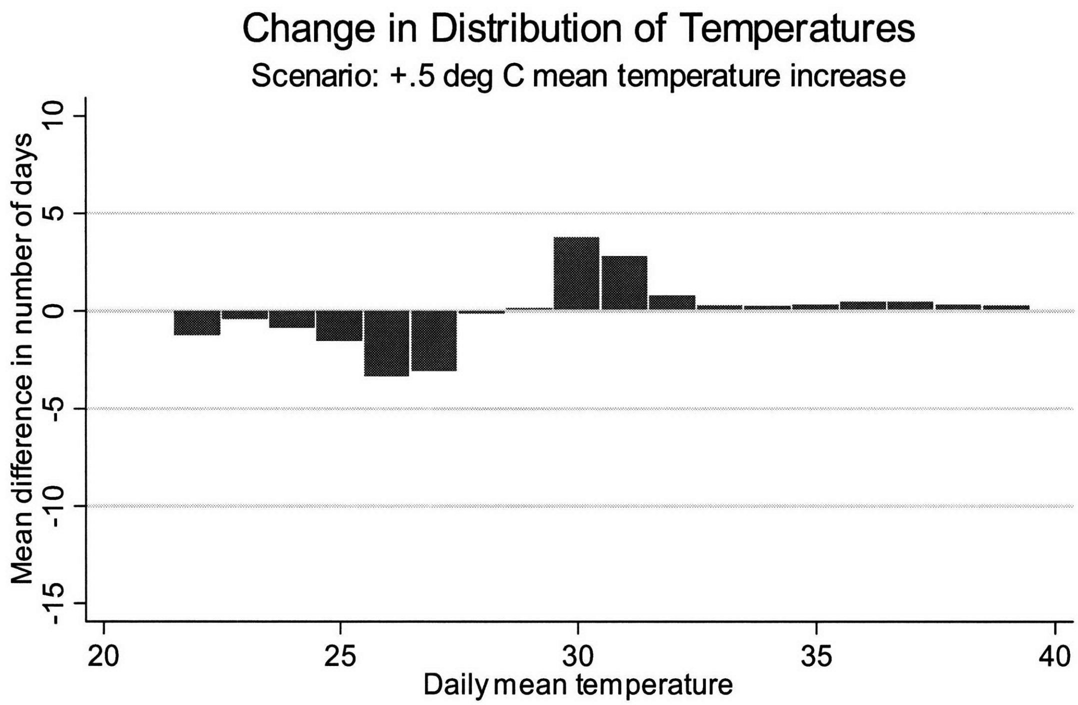
Notes: This map shows the 218 districts with output data for all years 1960-1999. The bulk of the lost districts (relative to the SMD98 dataset) are from the East, especially Bihar and West Bengal.

Figure 3:



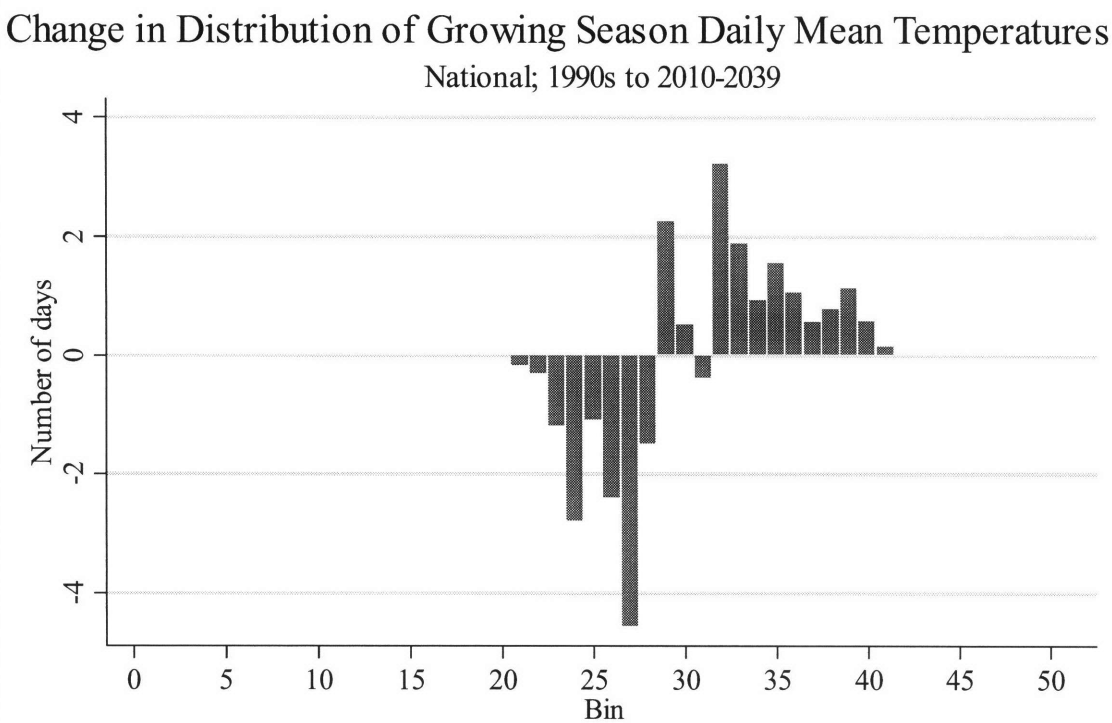
Notes: Each one-degree C bin is indicated by the upper limit, e.g. 20 indicates a day with mean temperature in (19,20]. Growing season is June-September (122 days).

Figure 4:



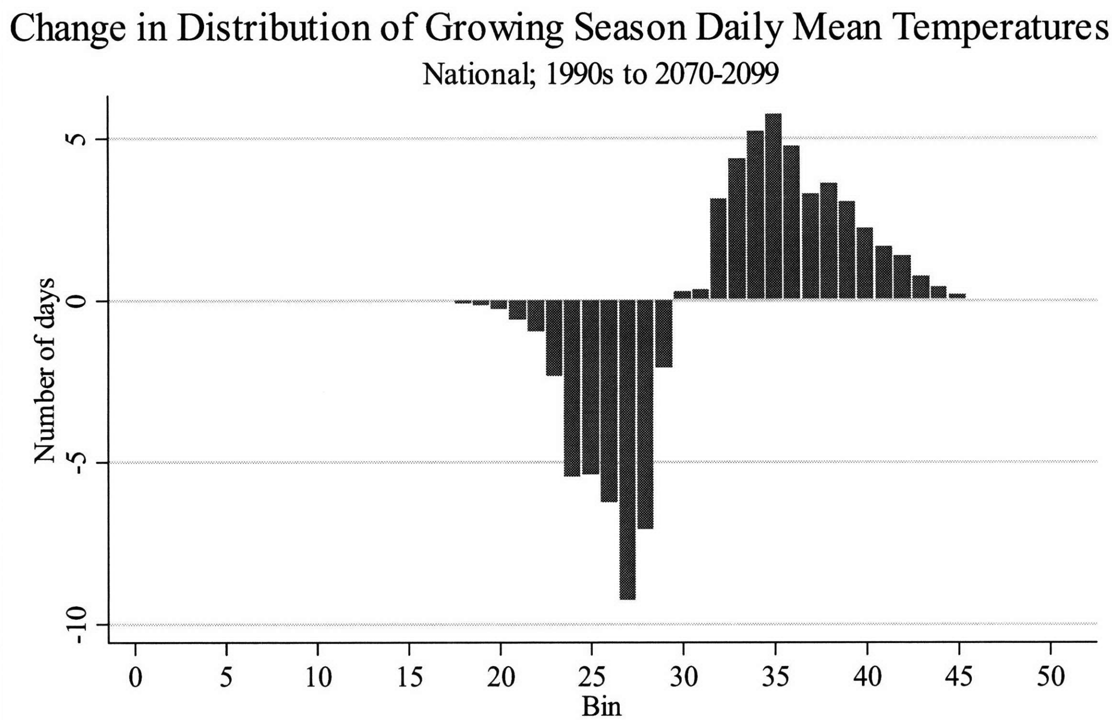
Notes: Each one-degree C bin is indicated by the upper limit, e.g. 30 corresponds to mean tempera
All temperatures below 22C and above 38C are grouped. Growing season is June-September (122
Years: 1960-1999; Sample: districts with output data for all years.

Figure 5:



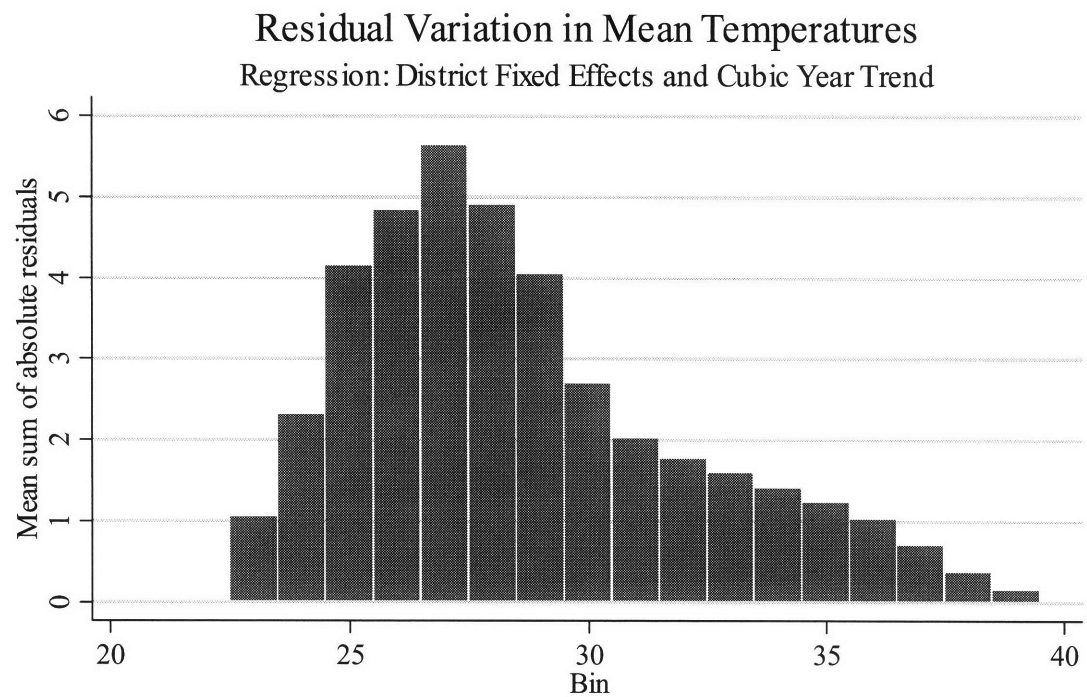
Notes: Each one-degree C bin is indicated by its upper limit, e.g. 30 corresponds to mean temperature (29,30].
Source: Hadley A1F1 model, rescaled. Growing season is June-September (120 days).

Figure 6:



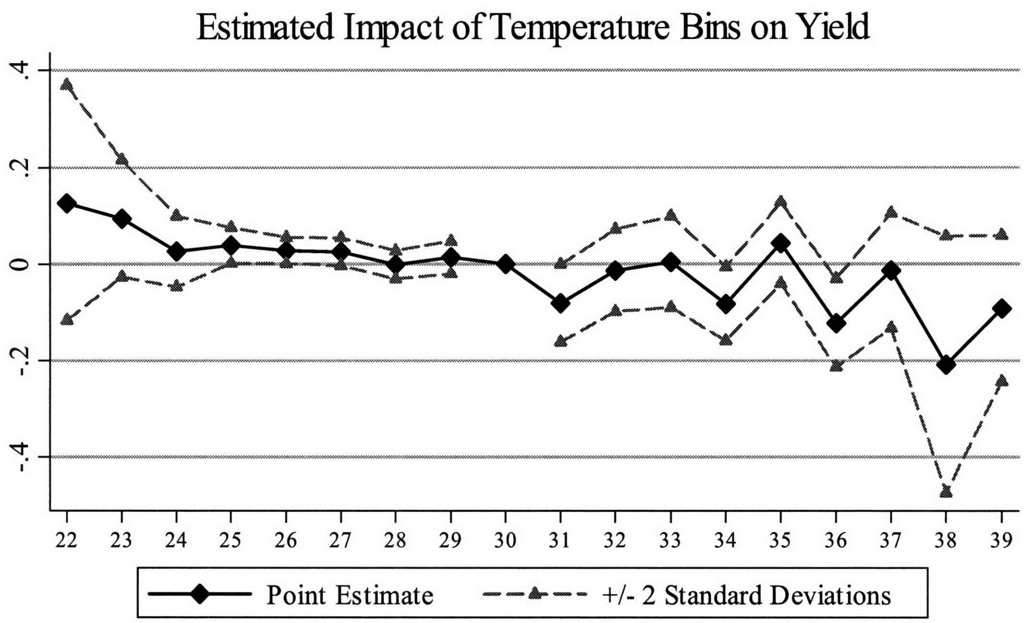
Notes: Each one-degree C bin is indicated by its upper limit, e.g. 30 corresponds to mean temperature (29,30].
Source: Hadley A1F1 model, rescaled. Growing season is June-September (120 days).

Figure 7:



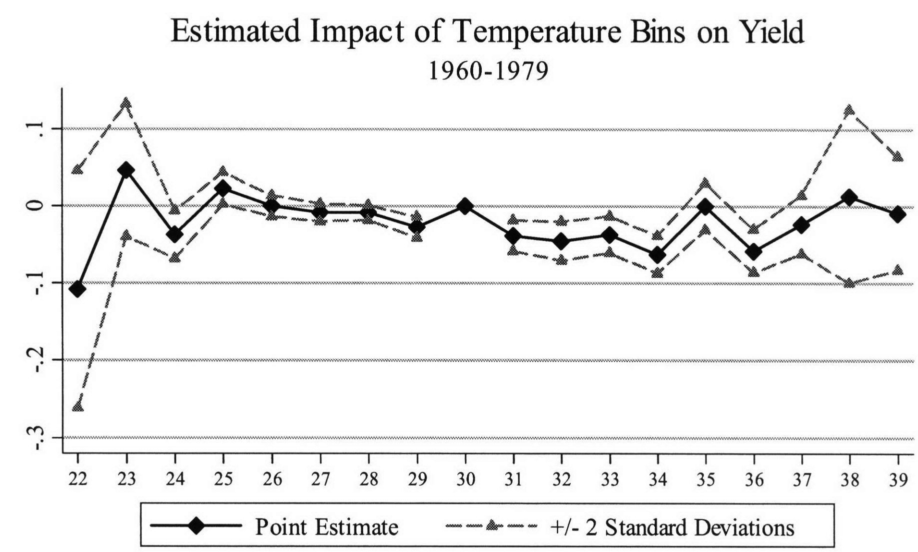
Notes: Each one-degree C bin is indicated by the upper limit, e.g. 30 corresponds to mean temperature (29,30]. All temperatures below 22C and above 38C are grouped. Growing season is June-September (122 days). Years: 1960-1999; Sample: districts with output data for all years.

Figure 8:



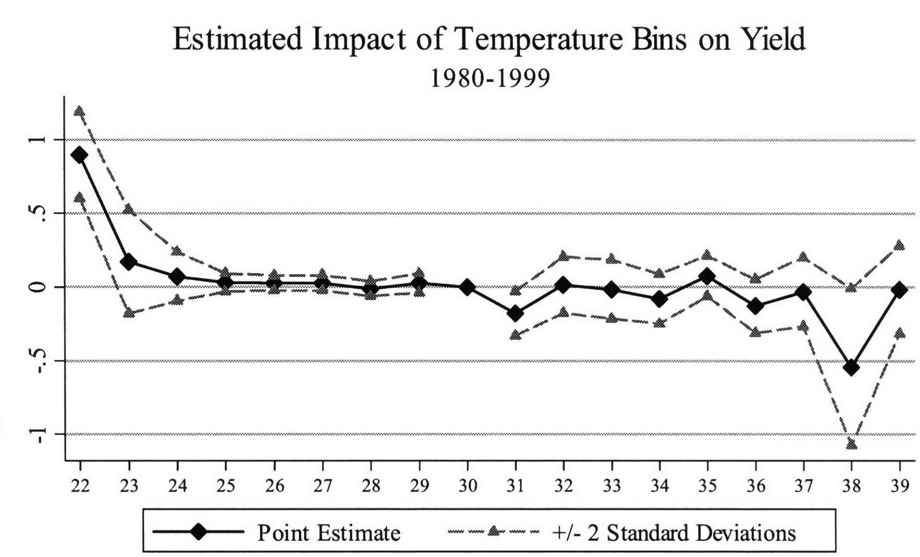
Notes: Plots coefficients from FGLS regression of yield on number of growing season days in each bin, plus district fixed effects, monthly precipitation (and squares) and regional cubic time trends. Each 1-degree C bin is indicated by upper limit, e.g. 24 indicates a day in with mean temperatures in (23,24] C. The (29,30] bin is the omitted category, so estimates are relative to the effect of a day in (29,30]. Temperatures below 22 C and above 38C are grouped. Sample: 218 districts with output data for all years 1960-1999.

Figure 9.A:



Notes: Plots coefficients from FGLS regression of yield on number of growing season days in each bin, plus district fixed effects, monthly precipitation (and squares) and regional cubic time trends. Each 1-degree C bin is indicated by upper limit, e.g. 24 indicates a day in with mean temperatures in (23,24] C. The (29,30] bin is the omitted category, so estimates are relative to the effect of a day in (29,30]. Temperatures below 22 C and above 38C are grouped. Sample: 218 districts with output data for all years 1960-1999.

Figure 9.B:



Notes: Plots coefficients from FGLS regression of yield on number of growing season days in each bin, plus district fixed effects, monthly precipitation (and squares) and regional cubic time trends. Each 1-degree C bin is indicated by upper limit, e.g. 24 indicates a day in with mean temperatures in (23,24] C. The (29,30] bin is the omitted category, so estimates are relative to the effect of a day in (29,30]. Temperatures below 22 C and above 38C are grouped. Sample: 218 districts with output data for all years 1960-1999.

Appendix Table 2.A: Residual Variation in Growing Season Degree-Days with Alternative Upper Bounds

Panel 1: Degree-Day Upper Bound of 33 C

		District*year observations differing from predicted value by more than									
Mean: 2,479.1; N:8720		60 deg-days (C)		120 deg-days (C)		180 deg-days (C)		240 deg-days (C)		300 deg-days (C)	
Regressors	RMSE	Number	Share	Number	Share	Number	Share	Number	Share	Number	Share
Constant only	204.38	6826	0.783	5069	0.581	3304	0.379	1730	0.198	994	0.114
District FEs	53.72	2168	0.249	235	0.027	40	0.005	2	0.000	0	0.000
District FEs, Linear Year	52.58	2058	0.236	206	0.024	33	0.004	0	0.000	0	0.000
District FEs, Quadratic Year	52.15	2031	0.233	212	0.024	36	0.004	1	0.000	0	0.000
District FEs, Cubic Year	52.10	2025	0.232	208	0.024	33	0.004	0	0.000	0	0.000
District and Year FEs	35.52	753	0.086	18	0.002	0	0.000	0	0.000	0	0.000
District and Year*Region FEs	28.51	374	0.043	3	0.000	0	0.000	0	0.000	0	0.000
District and Year*State FEs	21.29	104	0.012	2	0.000	0	0.000	0	0.000	0	0.000

Panel 2: Degree-Day Upper Bound of 34 C

		District*year observations differing from predicted value by more than									
Mean: 2,489.3; N:8720		60 deg-days (C)		120 deg-days (C)		180 deg-days (C)		240 deg-days (C)		300 deg-days (C)	
Regressors	RMSE	Number	Share	Number	Share	Number	Share	Number	Share	Number	Share
Constant only	211.53	6855	0.786	5146	0.590	3489	0.400	1984	0.228	1093	0.125
District FEs	56.43	2379	0.273	295	0.034	47	0.005	3	0.000	0	0.000
District FEs, Linear Year	55.28	2287	0.262	247	0.028	41	0.005	2	0.000	0	0.000
District FEs, Quadratic Year	54.83	2271	0.260	245	0.028	44	0.005	3	0.000	0	0.000
District FEs, Cubic Year	54.76	2264	0.260	244	0.028	42	0.005	2	0.000	0	0.000
District and Year FEs	37.00	853	0.098	22	0.003	0	0.000	0	0.000	0	0.000
District and Year*Region FEs	29.62	418	0.048	5	0.001	0	0.000	0	0.000	0	0.000
District and Year*State FEs	22.03	130	0.015	2	0.000	0	0.000	0	0.000	0	0.000

Panel 3: Degree-Day Upper Bound of 35 C

		District*year observations differing from predicted value by more than									
Mean: 2,495.9; N:8720		60 deg-days (C)		120 deg-days (C)		180 deg-days (C)		240 deg-days (C)		300 deg-days (C)	
Regressors	RMSE	Number	Share	Number	Share	Number	Share	Number	Share	Number	Share
Constant only	216.36	6876	0.789	5193	0.596	3602	0.413	2144	0.246	1165	0.134
District FEs	58.51	2517	0.289	325	0.037	55	0.006	5	0.001	0	0.000
District FEs, Linear Year	57.33	2419	0.277	282	0.032	48	0.006	3	0.000	0	0.000
District FEs, Quadratic Year	56.88	2424	0.278	278	0.032	52	0.006	5	0.001	0	0.000
District FEs, Cubic Year	56.78	2427	0.278	270	0.031	47	0.005	3	0.000	0	0.000
District and Year FEs	38.24	948	0.109	30	0.003	1	0.000	0	0.000	0	0.000
District and Year*Region FEs	30.55	477	0.055	8	0.001	0	0.000	0	0.000	0	0.000
District and Year*State FEs	22.71	143	0.016	3	0.000	0	0.000	0	0.000	0	0.000

Panel 4: Degree-Day Upper Bound of 36 C

		District*year observations differing from predicted value by more than									
Mean: 2,499.5; N:8720		60 deg-days (C)		120 deg-days (C)		180 deg-days (C)		240 deg-days (C)		300 deg-days (C)	
Regressors	RMSE	Number	Share	Number	Share	Number	Share	Number	Share	Number	Share
Constant only	219.12	6881	0.789	5224	0.599	3660	0.420	2246	0.258	1233	0.141
District FEs	59.91	2601	0.298	356	0.041	64	0.007	7	0.001	0	0.000
District FEs, Linear Year	58.72	2521	0.289	318	0.036	50	0.006	6	0.001	0	0.000
District FEs, Quadratic Year	58.24	2517	0.289	300	0.034	53	0.006	5	0.001	0	0.000
District FEs, Cubic Year	58.13	2519	0.289	300	0.034	48	0.006	5	0.001	0	0.000
District and Year FEs	39.13	1011	0.116	32	0.004	1	0.000	0	0.000	0	0.000
District and Year*Region FEs	31.23	519	0.060	9	0.001	0	0.000	0	0.000	0	0.000
District and Year*State FEs	23.24	170	0.019	3	0.000	0	0.000	0	0.000	0	0.000

Notes: Table counts residuals from regressions of district*year observations on regressors listed in row headings. Cell entries are number of residuals of absolute value greater than or equal to the cutoffs given in the column headings.Years: 1960-1999; Sample: 218 districts with output data for all years.

Chapter 2:

Estimating Quantile Treatment Effects in a Regression-Discontinuity Design

1 Introduction

In the last 10 years, regression discontinuity (RD) has become increasingly popular in economics as a quasi-experimental estimator of treatment effects. Interesting recent applications include van der Klaauw (2002) on the effect of financial aid offers on college enrollment, Black (1999) on valuation of school quality, Angrist and Lavy (1999) on the effect of class size on student achievement, Black et al. (2003) on the impact of job training on earnings and the duration of unemployment spells, Jacob and Lefgren (2004) on the educational impact of summer school and grade retention and Oreopoulos (2006) on the effect of compulsory education on earnings. Each of these papers exploits an abrupt change in the probability of receiving some treatment to estimate the mean causal effect of that treatment on some outcome for those who, on the basis of some covariate, are close to this jump in probability. The pioneering theoretical work of Hahn et al. (2001) has shown that this approach can produce unbiased estimates of mean treatment effects, and comparisons of RD estimates with the "gold standard" of randomized trials have shown that the RD method is often successful at approximating the experimental measurements (Black et al., 2007; Cook and Wong, 2007).

However, for many of the applications mentioned above, the mean treatment effect is not the only object of interest. For outcomes such as student achievement or earnings, we might be interested in the distribution of effects. In the case of summer school, for example, Jacob and Lefgren show that remedial education increases the mean of test scores, but the distribution of the effect is also important. For example, it could be that weaker students benefit most from the additional schoolwork while stronger students are negatively impacted by peer effects. On the other hand, summer school could benefit the relatively strong students and be less useful for the weaker students. In fact, Jacob and Lefgren write that "*there also exist methods (such as quantile regression analysis) for examining the distribution of outcomes in treated and untreated states for individuals with similar observable characteristics. Unfortunately, it is not straightforward to apply these methods in the context* *of our regression discontinuity analysis.*"

The goal of this paper is to provide a straightforward method to estimate quantile treatment effects in a regression discontinuity (RD) model. The link between RD and instrumental variables (IV) developed by Angrist and Lavy (1999) and Hahn et al. (2001) suggests using quantile IV to estimate treatment effects in a quantile RD model. I use the inverse quantile regression (IQR) method of Chernozhukov and Hansen (2006) to provide a simple, implementable estimator for quantile regression discontinuity (QRD). I provide simulation results on the effectiveness of the estimator using a simulated example based on Jacob and Lefgren (2004) and an empirical example applying the estimator to the increase in the minimum drop-out age studied by Oreopoulos (2006). I also revisit the results of Oreopoulos (2006) and show that, due to a data-processing error, his estimates are incorrect.[1] Not only is the RD estimate of the effect on mean earnings insignificant, QRD estimates suggest no detectable effect on other points of the earnings distribution.

This paper proceeds as follows. Section 2 provides a brief review of mean regression discontinuity and establishes notation. Section 3 shows that quantile RD can be incorporated into a quantile IV framework and discuss estimation procedures. Section 4 provides the simulation example and Section 5 provides the empirical example. Section 6 concludes and discusses avenues for further work.

2 Review of Regression Discontinuity Model

This treatment closely follows Imbens and Lemieux (2007). For simplicity, I provide a basic setup with a binary treatment and no covariates. Consider an outcome variable Y such as educational attainment or earnings. Each unit i in the population has potential outcomes $\{Y_i(0), Y_i(1)\}$ depending on whether or not she received the treatment. As usual, we do not

[1] This error was discovered independently by myself, Joe Shapiro of the MIT Department of Economics and Paul Devereux and Robert Hart of University College Dublin and University of Stirling, respectively.

observe potential outcomes but the realized outcome, which we write as

$$Y_i = T_i Y_i(1) + (1 - T_i) Y_i(0)$$

where T_i is an indicator of whether i received the treatment:

$$T_i = 1\{i \text{ received treatment}\}$$

We are interested in the average treatment effect, $\beta_1 = E[Y_i(1) - Y_i(0)]$, where this averaging may take place over certain subpopulations. However, because treatment may be chosen endogenously, simple comparisons of means for the treated and untreated units will yield biased estimates of β_1.

Now, suppose that the probability of receiving the treatment changes discontinuously depending on whether the value of a covariate Z, which we call the forcing variable or running variable, exceeds some exogenously determined threshold, which we will label c:

$$\lim_{z \downarrow c} P(T_i = 1 \mid Z_i = z) \neq \lim_{z \uparrow c} P(T_i = 1 \mid Z_i = z)$$

It is important to note that unit i's treatment status is not completely determined by whether the value of unit i's forcing variable exceeds the threshold. This distinguishes this model, commonly referred to as a "fuzzy regression discontinuity" from a "sharp regression discontinuity," in which treatment depends deterministically on whether the value of unit i's forcing variable exceeds the threshold. Without loss of generality, let the probability of treatment increase when $Z > c$. For example, the probability of receiving an offer of financial aid from a university may increase if an applicant's SAT score exceeds 1200.

There are two key assumptions for identification in the RD model. The first is that the conditional regression functions for the outcome variable are continuous in the forcing variable, i.e.

$$E\left[Y\left(1\right)\mid Z=z\right] \text{ and } E\left[Y\left(0\right)\mid Z=z\right] \text{ are continuous in } z \tag{1}$$

That is, outcomes are allowed to depend on the value of the forcing variable, but this dependence must be continuous, in contrast with the discontinuous dependence of the probability of treatment on the forcing variable. This assumption could be violated if there is some other treatment that also depends on the position of the forcing variable relative to the threshold. For example, eligibility for Medicare leads to a discontinuous jump in health care coverage at age 65, but there may be other government services for which eligibility or access change discretely at age 65 (Card et al., 2004).

The second key assumption is that treatment status does not depend on potential outcomes within some neighborhood of the cutoff point. This is often referred to as "ignorability of treatment" and is written formally as

$$c-h<Z_i<c+h \Rightarrow Y_i\left(0\right), Y_i\left(1\right) \perp T_i \mid Z_i \tag{2}$$

where h is a positive number defining the neighborhood of the cutoff point. This assumption is most plausible if one believes that Z is somewhat random, for example a student's performance on an exam, or if the subjects do not know the cutoff point c. It could be violated if participants are able to manipulate their Z to sort onto the more desirable side of the cutoff.

When these two assumptions and a monotonicity assumption hold, Hahn et al. (2001) show that the local average treatment effect (LATE), i.e. the average effect on those whose treatment status was affected by their position relative to the cutoff, is identified. The simplest estimator for the LATE is the familiar Wald estimator using just observations in the h-neighborhood of the cutoff, i.e.

$$\hat{\beta}_1 = \frac{\hat{E}\left[Y \mid c-h \leq Z < c\right] - \hat{E}\left[Y \mid c \leq Z \leq c+h\right]}{\hat{E}\left[T \mid c-h \leq Z < c\right] - \hat{E}\left[T \mid c \leq Z \leq c+h\right]} \tag{3}$$

However, this assumes that the forcing variable has no effect on the outcome within the h-neighborhood of the cutoff, which may be too strong.

To relax this assumption, Imbens and Lemieux (2007) recommend estimation by local linear regression. For exposition, it is useful at this point to switch to a regression framework rather than the potential outcomes framework used above. Write the structural model as a simple linear model

$$Y_i = \beta_0 + \beta_1 T_i + \beta_{2,l} 1\left\{Z_i < c\right\}\left(Z_i - c\right) + \beta_{2,r} 1\left\{Z_i > c\right\}\left(Z_i - c\right) + \varepsilon_{iy} \tag{4}$$

which allows the forcing variable to affect Y and for the slope of this relationship to vary on either side of the cutoff. We obtain the LATE estimate by estimating four separate pieces with local linear regression. For the outcome variable y, estimate the reduced form as

$$\begin{aligned}
\left(\hat{\beta}_{1,y,l}, \hat{\beta}_{2,y,l}\right) &= \arg\min \sum_{\{i:c-h \leq Z_i < c\}} \left(Y_i - \beta_{1,y,l} - \beta_{2,y,l}\left(Z_i - c\right)\right)^2 \\
\left(\hat{\beta}_{1,y,r}, \hat{\beta}_{2,y,r}\right) &= \arg\min \sum_{\{i:c < Z_i \leq c\}} \left(Y_i - \beta_{1,y,r} - \beta_{2,y,r}\left(Z_i - c\right)\right)^2 \\
\hat{\beta}_{1,y} &= \hat{\beta}_{1,y,r} - \hat{\beta}_{1,y,l}
\end{aligned}$$

and for the treatment variable, estimate the first stage as

$$\begin{aligned}
\left(\hat{\beta}_{1,t,l}, \hat{\beta}_{2,t,l}\right) &= \arg\min \sum_{\{i:c-h \leq Z_i < c\}} \left(Y_i - \beta_{1,t,l} - \beta_{2,t,l}\left(Z_i - c\right)\right)^2 \\
\left(\hat{\beta}_{1,t,r}, \hat{\beta}_{2,t,r}\right) &= \arg\min \sum_{\{i:c < Z_i \leq c\}} \left(Y_i - \beta_{1,t,l} - \beta_{2,t,r}\left(Z_i - c\right)\right)^2 \\
\hat{\beta}_{1,t} &= \hat{\beta}_{1,t,r} - \hat{\beta}_{1,t,l}
\end{aligned}$$

Then the LATE estimate is calculated as the ratio of the reduced form to the first stage:

$$\hat{\beta}_T = \frac{\hat{\beta}_{1,y}}{\hat{\beta}_{1,t}} = \frac{\hat{\beta}_{1,y,r} - \hat{\beta}_{1,y,l}}{\hat{\beta}_{1,t,r} - \hat{\beta}_{1,t,l}} \tag{5}$$

Imbens and Lemieux (2007) provide details on the choice of bandwidth h and weighting of observations within the window.

There have been many extensions of this basic RD model, but two will be particularly important for the exposition of quantile RD here. First, Hahn et al. (2001) show that the Wald-type estimator in (5) is equivalent to an instrumental variables estk atlr. In this approach, the structural model (4) is estimated by two-stage least squares, using $1\{Z_i > c\}$ as an instrument for the endogenous T_i. The idea is that, once we have controlled for the (piecewise) linear effect of Z on the outcome variable, the side of the cutoff on which the unit happens to fall can be excluded from the structural equation. In other words, having controlled for the influence of Z in a smooth way, whether or not Z exceeds the threshold c affects the outcome *only* through its effect on the probability of treatment. Formally, and suppressing the piecewise linearity of the influence of Z in (4) for simplicity of notation, we have

$$E\left[1\{Z_i > c\}\, \varepsilon_{iy} = 0 \mid p_1(Z_i)\right]$$

where we use the notation $p_n(Z)$ to denote an n^{th}-order polynomial in Z.

Second, when the forcing variable is discrete, as in Card et al. (2004) and Oreopoulos (2006), it is not possible to restrict the data to arbitrarily small windows around the cutoff point. Instead, standard practice is to control for a flexible polynomial in the forcing variable and use data that are not so near the cutoff point. Lee and Card (2008) point out that conventional or even heteroscedasticity-robust standard errors will overstate the precision of the estimated coefficient, because they will not account for the error in estimating the polynomial function. Instead, Lee and Card (2008) recommend clustering at the discrete values of the forcing variable to obtain consistent standard errors.

3 Extending RD to Quantile Analysis

3.1 Quantile RD model

In this subsection, I show that an RD estimation strategy can be incorporated into the quantile IV framework of Chernozhukov and Hansen (2005). In the next subsection, I then discuss how this framework can be used to compute quantile RD estimates. Again, for simplicity I provide a basic setup with a binary treatment variable and no covariates. The most basic object of interest is the quantile treatment response (QTR) function $q(t,z,\tau)$, which is defined by the relationship

$$q(t,z,\tau) : P(Y \leq q(t,z,\tau) \mid T=t, Z=z) = \tau \tag{6}$$

Note that, as in the structural regression equation (4) above, the forcing variable Z enters the QTR function directly. We assume strict monotonicity, i.e. $q(t,z,\tau)$ is strictly increasing in τ. This rules out discrete outcome variables. It is convenient to rewrite the QTR function as

$$Y_t = q(t,z,U_t) \;\; ; \;\; U_t \sim U(0,1)$$

where U_t is a rank variable that describes the relative position of individuals with the same observables (here, $T=t$ and $Z=z$). U_t can be thought of as unmeasured ability. It is important to note that U_t in general depends on the treatment status, as its t-subscript indicates. That is, if we consider all units with the same value of $Z=z$, the median individual when all these units are exposed to the treatment, i.e. the individual with $U_{i,1}=0.5$, need not be the median individual when the treatment is withheld from these units, in which case $U_{i,0} \neq 0.5$. Restricting this evolution of ranks across treatment states will be key in identification below.

Treatment status is determined by the selection equation

$$T \equiv \delta\left(1\left\{Z>c\right\},Z,V\right) \tag{7}$$

where δ is an unknown function and V is an unobserved disturbance term. As above, Z is the forcing variable and the presence of $1\left\{Z>c\right\}$ allows for a discontinuous jump in treatment probability at a cutoff threshold c.

We wish to obtain estimates of the quantile treatment effect (QTE), which for a binary treatment variable is

$$q\left(1,z,\tau\right)-q\left(0,z,\tau\right) \tag{8}$$

If there is correlation between U and V, the disturbance terms in the structural and selection equations, then ordinary quantile regression will not provide consistent estimates of the QTE because of selection bias.

As in the mean RD model, it is the presence of the discontinuity in the probability of treatment that provides an excludable instrument and allows us to identify the QTE. We make three key assumptions:

A1. CONTINUITY: $q\left(1,z,\tau\right)$ and $q\left(0,z,\tau\right)$ are continuous in z for all τ.

A2. INDEPENDENCE: Conditional on $Z=z\in\left(c-h,c+h\right)$, $\{U_t\}$ are independent of $1\left\{Z>c\right\}$.

A3. RANK SIMILARITY: Conditional on (a) $Z=z\in(c-h,c+h)$, (b) $1\left\{Z>c\right\}$ and (c) V, the random variables $\{U_t\}=\{U_1,U_0\}$ have the same distribution.

The first assumption is exactly parallel to the continuity assumption in the mean RD model. The second is analogous to the exclusion restriction in the mean RD model. It says that, within the window around the cutoff, whether or not an individual happened to fall above or below the cutoff is not related to her unobserved ability.

The third assumption, which does not have a counterpart in the mean IV framework, says that the ranks of individuals with the same observables Z does not change *systematically* between the untreated and treated states. As a simple example, imagine two students, A and B, with the same end-of-year test score z. Student A has more ability than student B, but happened to have a bad day on the end-of-year test. Assumption A3 says that if we expect student A to outperform student B on an end-of-summer test if neither attends summer school, we would also expect student A to outperform student B if both attend summer school. It is important to note that this can still be satisfied if student A benefits more from summer school than student B, if the two benefit the same amount, or if student B benefits more. The main restriction imposed by Assumption A3 is that the weaker student B cannot benefit by so much more than student A that, even though we expect student A to outperform student B if both of them do not attend summer school, we expect student B to outperform student A if they both attend summer school.

More generally, there should be no systematic "crossings" in the distribution when comparing treated and untreated states, although the distribution may expand or contract (indeed, could expand in some regions and contract in others). This need not always hold: one might imagine a summer school so specifically targeted at weaker students that it damaged stronger students enough that the quantiles crossed.

With these assumptions, the conditions for Theorem 1 of Chernozhukov and Hansen (2005) are satisfied, yielding the following testable restriction: for all $\tau \in (0,1)$,

$$P\left(Y \leq q\left(T, Z, \tau\right) \mid Z, 1\left\{Z > c\right\}\right) = \tau \tag{9}$$

almost surely. This conditional moment restriction will form the basis of estimation. It is important to note that we are conditioning not just on the forcing variable Z but also on the excluded indicator $1\left\{Z > c\right\}$. The parallel in a non-RD quantile IV would be to condition on covariates X and an exogenous instrument W. This "conditioning on instruments" approach is only valid because of the restrictions imposed by our assumptions, especially the assumption of rank similarity.

3.2 Estimation

In this subsection, I discuss estimation of quantile treatment effects in the quantile RD model described above. For exposition, I simplify the structural quantile response function to a piecewise linear model, as in the mean RD case:

$$q(t,z,\tau) = \beta_0(\tau) + \beta_1(\tau) T_i + \beta_{2,l}(\tau) 1\{Z_i < c\}(Z_i - c) + \beta_{2,r}(\tau) 1\{Z_i > c\}(Z_i - c) \quad (10)$$

The coefficients are now written as functions of the quantile τ to remind us that these parameters can differ across quantiles. For example, the effect of summer school for the student at the first quartile of the distribution (conditional on $Z = z$) may be greater than the effect for a student at the third quartile, in which case $\beta_1(0.25) > \beta_1(0.75)$.

This notation can help in understanding the rank similarity condition A3 above: although $\beta_1(\cdot)$ can be decreasing as a function of τ, it cannot be so sharply decreasing that lower-ranking units overtake higher-ranking units.[2]

A tempting, intuitive estimator is the quantile equivalent of the RD estimator given by (5) above, i.e. use ordinary quantile regression to obtain estimates of each of the four pieces for a given quantile of interest and then calculate the ratio of the differences as above. This will not work for quantile RD, for two reasons. First, with a binary treatment, quantile estimation of the first stage is not really sensible, since all quantiles will be either zero or

[2]The formal condition is: for all z, τ, τ', if $q(0,z,\tau) > q(0,z,\tau')$ then $q(1,z,\tau) > q(1,z,\tau')$. To make this concrete, first simplify the structural equation to $q(t,z,\tau) = \beta_0(\tau) + \beta_1(\tau) T_i + \beta_2(\tau) Z_i$ for ease of exposition. The formal condition is now: if $\beta_0(\tau) + \beta_2(\tau) Z_i > \beta_0(\tau') + \beta_2(\tau') Z_i$, then $\beta_0(\tau) + \beta_1(\tau) + \beta_2(\tau) Z_i > \beta_0(\tau') + \beta_1(\tau') + \beta_2(\tau') Z_i$, or, rearranging, $\beta_1(\tau) - \beta_1(\tau') > (\beta_0(\tau') + \beta_2(\tau') Z_i) - (\beta_0(\tau) + \beta_2(\tau) Z_i)$. When $\beta_1(\tau) > \beta_1(\tau')$, this is automatically satisfied, because $q(0,z,\tau) > q(0,z,\tau')$ implies that the right-hand side of the previous inequality is negative. When $\beta_1(\tau) < \beta_1(\tau')$, it is convenient to rewrite the condition as $\beta_1(\tau') - \beta_1(\tau) > (\beta_0(\tau) + \beta_2(\tau) Z_i) - (\beta_0(\tau') + \beta_2(\tau') Z_i)$, which is to say that the treatment effect for the τ'-quantile is not so much greater than the effect for the τ-quantile that the τ-quantile's initial advantage, given by the right-hand side of this inequality, is overcome.

one. Second, and more fundamentally, the derivation of (5) depends heavily on the linearity of the expectations operator. In particular, (5) requires that

$$E\left[\beta_{1,y,r}-\beta_{1,y,l}\right]=E\left[\beta_{1,y,r}\right]-E\left[\beta_{1,y,l}\right]$$

for the reduced form in the numerator and similarly for the first stage in the denominator. However, the parallel is not true in the case of quantiles: it is not true that the quantile of the difference is the difference of the quantiles.

To proceed with estimation, we must instead follow the Hahn et al. (2001) IV approach to RD, but here, use quantile IV. In the previous subsection, I showed that, under the appropriate conditions, the quantile RD model can be expressed in terms of the conditional moment restriction given by (9). The analogy principle suggests that we could obtain parameter estimates by minimizing the empirical analog of the moment condition, which for our simple linear case is

$$\frac{1}{N^*}\sum_{\{i\,:\,c-h\leq Z_i\leq c+h\}} 1\left\{Y_i-\beta_1(\tau)T_i-\gamma(\tau)1\{Z_i>c\}-X_i'\theta(\tau)\right\}=\tau$$

where $N^*=\{\#i\,:\,c-h\leq Z_i\leq c+h\}$ and, for simplicity of notation, the exogenous variables are written as $X_i=\left(1\{Z_i<c\}(Z_i-c),1\{Z_i>c\}(Z_i-c),1\right)'$ with parameters $\theta(\tau)=\left(\beta_{2,l}(\tau),\beta_{2,r}(\tau),\beta_0(\tau)\right)'$ and $\gamma(\tau)$ is the coefficient on the instrument. Chernozhukov and Hansen (2006) show that this is equivalent to minimizing the criterion function

$$\hat{Q}(\tau,\theta,\beta_1,\gamma)=\frac{1}{N^*}\sum_{\{i\,:\,c-h\leq Z_i\leq c+h\}}\rho_\tau\left(Y_i-\beta_1(\tau)T_i-\gamma(\tau)1\{Z_i>c\}-X_i'\theta(\tau)\right)\tag{11}$$

where $\rho_\tau(\cdot)$ is the standard τ-quantile "check function":

$$\rho_\tau(s)=(\tau-1\{s<0\})s$$

Note that "conditioning on instruments" is readily apparent here, since the instrument $1\{Z_i > c\}$ appears directly in the objective function.

The coefficient of interest, $\beta_1(\tau)$, can be estimated using the "inverse quantile regression" method of Chernozhukov and Hansen (2006): $\hat{\beta}_1(\tau)$ is chosen as the value of β_1 that minimizes γ, the coefficient on the excluded instrument $1\{Z_i > c\}$, in an ordinary quantile regression of Y on T, $1\{Z_i > c\}$ and X. The intuition for this procedure is that, since the instrument can be excluded from the structural equation, its coefficient in a regression of the outcome on the treatment should be zero once we have conditioned on the exogenous variables. In practice, the algorithm is as follows: for a given quantile τ, (1) fix a value of β_1, which we will call β_1^j; (2) perform an ordinary quantile regression of $Y - \beta_1^j T$ on $1\{Z > c\}$ and X, obtaining $\tilde{\gamma}(\beta_1^j)$ and $\tilde{\theta}(\beta_1^j)$ and; (3) calculate a quadratic form in $\tilde{\gamma}(\beta_1^j)$; (4) repeat over a suitable grid of values for β_1, i.e. $\beta_1^1, \ldots, \beta_1^j, \ldots, \beta_1^J$; (5) let $\hat{\beta}_1(\tau)$ be the value of β_1^j that minimizes the quadratic form in $\tilde{\gamma}(\beta_1^j)$.

Three approaches to conducting inference are possible. First, confidence intervals can be constructed via an "inverse Wald statistic" procedure: if the middle matrix in the quadratic form in $\tilde{\gamma}(\beta_1^j)$ is consistent for the asymptotic variance of $\tilde{\gamma}(\beta_1^j)$, then this quadratic form is a proper Wald statistic taking a $\chi^2_{(1)}$ distribution under the null hypothesis that $\gamma_0 = 0$. The confidence interval for $\hat{\beta}_1(\tau)$ consists of those values of are those values of β_1^j such that the Wald statistic for $\tilde{\gamma}(\beta_1^j)$ does not reject $\gamma_0 = 0$ at the confidence level chosen. Formally,

$$\begin{aligned} CI &= \left\{\beta_1^j : W\left(\tilde{\gamma}\left(\beta_1^j\right)\right) < C_{\chi^2_{(1)}}\right\} \\ W\left(\tilde{\gamma}\left(\beta_1^j\right)\right) &= \tilde{\gamma}\left(\beta_1^j\right)' \left(\hat{V}\left[\tilde{\gamma}\left(\beta_1^j\right)\right]\right)^{-1} \tilde{\gamma}\left(\beta_1^j\right)' \end{aligned}$$

where $C_{\chi^2_{(1)}}$ is the appropriate critical value for the $\chi^2_{(1)}$ distribution.[3] This has theoretical

[3]The Wald statistic need not be convex in β_1^j and in some cases it may happen that the set CI will incorporate nonadjacent values of β_1^j. In these cases, we set the left-hand endpoint of the confidence interval to be the *minimum* β_1^j such that the Wald statistic does not reject the null and the right-hand endpoint to be the *maximum* such β_1^j.

advantage of being robust to weak instruments and partial identification (Chernozhukov and Hansen, 2008) and the practical advantage that the inputs are calculated during the estimation process. Second, analytical estimated variance-covariance matrices can be calculated following the formulae given in Section 4.3 of Chernozhukov and Hansen (2008). These have the advantage of being computable for multiple quantiles simultaneously, which allows testing of hypotheses across quantiles (e.g. uniform effects, increasing effects, etc.). However, while the computed standard errors will be heteroscedasticity-robust, it is difficult to adapt them to other complications such as clustering. Finally, standard errors can be computed by bootstrapping. This can be easily adapted to accommodate clustering by sampling at the cluster level, as recommended by Horowitz (2001). However, this approach may be computationally intensive, since $B \times J$ quantile regressions will be required to perform B bootstrap repetitions over a grid of J points just for a single quantile.

4 Simulated Example

This section presents a series of simulation results based on the Jacob and Lefgren (2004) study of the effects of summer school on educational achievement. Because students select into summer school, a simple regression of a measure of educational achievement on summer school and other observables likely would not provide an unbiased estimate of the effects of summer school. However, Jacob and Lefgren take advantage of a policy change in the Chicago Public Schools to implement a regression discontinuity design. In 1995, the Chicago Public Schools (CPS) instituted a strict requirement for promotion from third, sixth and eighth grades: all students had to score above grade-level thresholds on standardized math and reading tests. Those who did not meet the threshold were required to attend six weeks of summer school and re-take the test at the end of the summer to be promoted to the next grade. Jacob and Lefgren use this sharp discontinuity in the probability of treatment to identify the causal effects of summer school and grade retention. They find that summer school increased academic achievement in reading and math for both third and sixth graders, with the effect persisting for two years. The effects of grade retention are less clear: third graders do not suffer and may benefit in the short run; sixth graders' math scores are unaffected but their reading scores fall. The simulations in this section will show how this mean RD study could be extended to examine quantile treatment effects. I first provide a simple example of QRD estimation of a single quantile, the median, to introduce the simulation procedure and describe the implementation of QRD in some detail. I then show results of simulations over all nine deciles for a constant treatment effect, an increasing treatment effect and a decreasing treatment effect.

The simulation is performed as follows. Each of 10,000 units i is randomly assigned a test score $Z_i \sim U(0,5)$ and a pair of disturbance terms

$$\begin{pmatrix} \varepsilon_i \\ \eta_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right)$$

where ε_i is the disturbance to the structural (outcome) equation and η_i is the disturbance to the selection equation. Unit i's quantile index is then $\tau_i = \Phi(\varepsilon_i)$, where $\Phi(\cdot)$ is the standard normal CDF.[4] Given these random disturbances, unit i's treatment status is given by the sign of the latent variable

$$T_i^* = \pi_0 + \pi_1 Z_i + \pi_2 1\{Z_i > c\} + \eta_i$$

i.e. $T_i = 1\{T_i^* > 0\}$. The cutoff point c is chosen to be 2 so that roughly 40% of the units will be below the cutoff and the parameters π_0, π_1 and π_2 are chosen so that the assignment probabilities roughly match those in the Jacob and Lefgren data: the probability of treatment is about 0.9 just above the cutoff and roughly 0.2 just above the cutoff, with the

[4] Because Z_i is generated independently of the disturbance terms, we do not have to calculate the quantile conditional on Z_i.

probabilities going towards zero (one) as the test score decreases (increases).[5] The outcome variable is then determined by the structural equation

$$Y_i = \beta_1(\tau_i) T_i + \beta_2(\tau_i) Z_i + \varepsilon_i$$

where the functions $\beta_1(\tau_i)$ and $\beta_2(\tau_i)$ will depend on the particular simulation. In this simplest expository case, we are interested only in the median and set $\beta_1(0.5) = 0.5$ and $\beta_2(0.5) = 1$. In the context of the schooling example, this means that test scores next year increase one-for-one (in terms of grade level) with this year's test score, and summer school improves test scores by half a grade level. Note that for simplicity the slope to the right and to the left of the cutoff are now constrained to be equal, i.e. $\beta_{2,l}(0.5) = \beta_{2,r}(0.5) = \beta_2(0.5) = 1$.

The parameter ρ is particularly important, because it induces correlation between the selection and outcome equations and thus leads to bias. In this simulation, I set $\rho = -0.3$, which corresponds to negative selection: for example, students whose parents anticipate a worse outcome next year are more likely to encourage their students to attend summer school. In this case, simple OLS or quantile regression would underestimate the effect of summer school.

Figures (1) and (2) plot the treatment probability and mean of the outcome variable, respectively, against the running variable, from a randomly chosen repetition of the simulation. As one would hope, discontinuities in both the first stage and reduced form are plainly visible.

For each repetition of the simulation, I run OLS, RD / two-stage least squares, ordinary quantile regression and QRD/IVQR. Each of these is calculated over the full 10,000 observations and within a window of ± 0.25 units around the cutoff of 2. Again, this window corresponds roughly to that employed by Jacob and Lefgren and yields about 1,000 observations per repetition. The mechanics of QRD estimation are as follows. I use a grid with

[5]In the simulation, I set $\pi_0 = 1.4$, $\pi_1 = -0.2$ and $\pi_2 = -1.4$.

a width of 4 grade-level equivalents, with each grid is centered on the mean RD estimate for that sample. There are 401 grid points in total, separated evenly by 0.01 grade-level equivalents. I chose this spacing because it is an order of magnitude below what we might consider the minimum economically significant increment of 0.1 grade-level equivalents but also leads to a computationally feasible number of grid points. I follow the algorithm given in Section 3.2 above: for each grid point β_1^j, I perform an ordinary median regression of $Y_i - \beta_1^j T_i$ on $1\{Z_i > c\}$, Z_i, and a constant. I then compute a Wald statistic in $\tilde{\gamma}\left(\beta_1^j\right)$, set $\hat{\beta}_1$ equal to the minimizing β_1^j, and compute a confidence interval containing all β_1^j such that the Wald statistic in $\tilde{\gamma}\left(\beta_1^j\right)$ is less than the 90% critical value. In Figure (3), I plot these $W\left(\tilde{\gamma}\left(\beta_1^j\right)\right)$ over the set of grid points for an arbitrary repetition of the Monte Carlo exercise. This figure should make clear the idea of inverse quantile regression estimation: the point estimate, indicated by the solid vertical line, corresponds to the minimum value of the objective function; the confidence interval, indicated by the dashed vertical lines, contains all values of β_1 for which the objective function falls below the critical value of the $\chi^2_{(1)}$ distribution, which is indicated by the solid horizontal line.

The results of this Monte Carlo exercise are summarized in Table 1. First, both OLS and ordinary quantile regression exhibit large downward bias, as we would expect given negative selection. Mean RD / 2SLS and QRD / IVQR both perform well. When all the observations are used, the confidence intervals for mean RD are on average a bit narrower and the RMSE smaller than for quantile RD, but this is not surprising since 2SLS is in fact the efficient estimator in this case. Still, QRD is certainly competitive, as Figure (3), which plots kernel densities of the two estimators, indicates. QRD also displays good size properties, rejecting the truth 10.8% of the time when the intended size is 10%. When the data are restricted to the ± 0.25 window around the cutoff, the results are broadly similar but the gap between the performance of RD and QRD widens.

I perform three variants of this Monte Carlo simulation looking at estimates across deciles:

first, for a constant treatment effect across deciles; then for a treatment effect that is increasing in τ; and finally for a treatment effect that is decreasing in τ. In the first, the treatment effect is 0.5 for all quantiles, i.e. $\beta_1(\tau) = \beta_1 = 0.5$. All other parameters are the same as in the previous simulation. Table 2 and Figures (5) and (6) summarize the results of this exercise. The QRD estimator performs well on average for all quantiles, but its performance is best at central quantiles. For high and low quantiles, 0.1 and 0.9 in particular, the distribution of the estimator tends to widens noticeably, as indicated by the widening 25/75 bars and 5/95 whiskers in Figures (5) and (6). In the second variant of the Monte Carlo exercise, the treatment effect is increasing linearly in the quantile index, i.e.

$$\beta_1(\tau) = \tau$$

with all other parameters the same. Table 3 and Figures (7) and (8) summarize the results of this exercise, which are broadly similar to the results of the previous one: the estimator performs well on average for all quantiles, but its distribution becomes increasingly dispersed at high and low deciles. Finally, I run a Monte Carlo simulation with a decreasing treatment effect:

$$\beta_1(\tau) = 0.75 - 0.5\tau$$

Table 4 and Figures (9) and (10) summarize the results, which largely mimic the results of the previous two.

5 Empirical Application

5.1 Review of Oreopoulos (2006) results

The returns to education is one of the most heavily researched topics in economics. Because an individual's level of schooling is likely correlated with unobserved characteristics that also affect earnings, much recent research has focused on finding valid instruments for schooling in hopes of obtaining consistent IV estimates of the returns. One strand of this literature has used variation induced by compulsory schooling laws as an instrument for schooling (Angrist and Krueger, 1991; Acemoglu and Angrist, 2000; Oreopoulos, 2003). However, many of these instruments affect less than ten percent of the population. Since IV estimates yield estimates reflecting the causal effect only for the subpopulation affected by the instrument, it is not clear how applicable these estimates are to the larger population.

In a recent paper, Oreopoulos (2006) uses the increase in the minimum dropout age in Great Britain from 14 to 15 in 1947, which, because of historically high dropout rates, impacted nearly half the population.[6] Because the effect of the legal change was so broad, it is likely that estimates from this quasi-experiment will be more generally applicable. This change took place in a single year and affected all cohorts born during or after 1933 (who turned 14 in 1947 or after) and none born before, making a regression discontinuity design applicable. Since the forcing variable, year of birth, is discrete, it is necessary to control for a polynomial in year of birth and cluster standard errors by year of birth.

Oreopoulos uses data from the 1983-1998[7] UK General Household Surveys (GHHS) to estimate the following set of equations: a first stage equation

$$E_{i,t} = \delta_1 1\left\{YOB_i > 1933\right\} + \delta_2' p_4\left(YOB_i\right) + \delta_3' p_n\left(AGE_{it}\right) + \varepsilon_{i,t,2} \tag{12}$$

where $E_{i,t}$ is the age left full-time schooling for individual i surveyed in year t (the data are a repeated cross-section, not a panel), YOB_i is the year of birth for individual i, and $AGE_{i,t}$ is individual i's age when surveyed in year t. Oreopoulos includes a quartic in year of birth and varies the controls for age (none, a quartic and a full set of fixed effects); a reduced form

[6] Oreopoulos also uses a similar change for Northern Ireland in 1957. My discussion focuses on the results for Great Britain, because the Northern Ireland income data are reported only by category, which I cannot use in my quantile RD estimates.

[7] No survey was conducted in 1997.

equation

$$\log y_{it} = \pi_1 1\left\{YOB_i > 1933\right\} + \pi_2' p_4\left(YOB_i\right) + \pi_3' p_n\left(AGE_{it}\right) + \varepsilon_{it,3} \tag{13}$$

where y_{it} is real weekly earnings for individual i in year t and all other variables are as in the first-stage equation (12); and a structural equation

$$\log y_{it} = \beta_1 E_{i,t} + \beta_2' p_4\left(YOB_i\right) + \beta_3' p_n\left(AGE_{it}\right) + \varepsilon_{i,t,1} \tag{14}$$

which is estimated by OLS (for comparison) and two-stage least squares, with $1\left\{YOB_i > 1933\right\}$ instrumenting for $E_{i,t}$ In estimating equation (14), Oreopoulos collapses his data to year-of-birth by age cell means, runs the regressions on the cell means, weighting by cell size, and clusters the standard errors by year of birth. The point estimates computed in this way will be numerically identical to the point estimates obtained by the Lee and Card (2008) procedure of running the regressions on the micro data and clustering standard errors by year of birth. However, collapsing to year-of-birth by age cell means will yield even more conservative standard errors, since variation at the age level is lost.

Although Oreopoulos finds positive and significant effects of education on earnings, these estimates are incorrect, likely due to an error either in downloading the raw individual-level data from GHHS or in converting these micro-data to cell means. The exact cause of the error cannot be determined, since Oreopoulos's raw data are no longer available. My estimates of the first stage and reduced form are given in Table 5.[8] I do find a positive and highly significant first-stage relationship, although my point estimates are roughly 15 percentage points lower than the estimates in his Table 1. However, my estimates of the reduced-form equation (13) yield much smaller point estimates and slightly larger standard errors in two of three specifications. Some of these differences may be due to differences in the

[8]To maximize comparablility with Oreopoulos's results, I also collapse the data to year-of-birth by age cell means and run the regressions on the cells, weighting by cell size and clustering on year of birth.

raw data used: I discovered that a coding error at the UK Data Archive, which houses and disseminates the GHHS data, had caused all observations of earnings from the 1983 GHHS to be coded as missing. By correcting this coding error, I was able to restore the 1983 data to the sample. However, this turns out not to make much of a difference for the first stage or reduced form, as is apparent in Table 7, where I restrict the sample to observations from the 1984-1996 GHHS, making my sample as comparable to Oreopoulos's as possible. The first stage estimates are virtually unaffected, while the reduced form estimates do increase but not to the level of statistical significance and still well below Oreopoulos's estimates.

Table 6 reports the OLS and RD estimates of equation (14). Interestingly, my OLS estimates are substantially higher than Oreopoulos's, given in his Table 2. As expected from the first stage and reduced form results, my RD estimates are substantially smaller – roughly 8.5 to 10 log points as compared to his point estimates of 14.5 to 15 log points. Furthermore, my estimated standard errors have doubled, eliminating any statistical significance. When I restrict to the 1984-1998 sample, my reduced form point estimates increase, approaching and even exceeding Oreopoulos's in one specification. However, the standard errors remain high, so no results are statistically significant. These results are largely unchanged when I restrict the sample to men only and when I reduce the order of the year-of-birth polynomial to a quadratic.[9]

5.2 Applying Quantile RD

Although the preceding section shows that no statistically significant effect can be detected on the mean, we could still see effects at particular quantiles, for example if the benefits of education tended to be high for those lower in the earnings distribution (conditional on their

[9]I took the latter step because the condition index of the $\left[1\{YOB > 1933\} \vdots p_4(YOB)\right]$ data matrix is over 1,000, indicating severe multicollinearity and likely instability of estimates. The condition index of $\left[1\{YOB > 1933\} \vdots p_2(YOB)\right]$ was just 42, which is not too far in excess of the rule of thumb that a condition index should not exceed 30 (Belsley, 1991).

covariates, in this case year of birth and age at time of survey). To explore this possibility, I apply the quantile RD estimator developed in Section 2 to the raw GHHS data. The details of the estimation procedure are as follows: for each quantile, I use a grid centered at 0.15 and extending 0.5 units in either direction. Since the units are log points, this should encompass any realistic impact on wages. The grid has 201 points, spaced 0.005 units apart. I use only a second-degree polynomial in year of birth, since the multicollinearity of higher-order terms with the indicator function $1\{YOB > 1933\}$ causes even more computational difficulty in quantile regression than it does in least squares. Similarly, I do not include age fixed-effects in these models for computational reasons: in quantile regression, fixed effects cannot be simply "partialled out" as in least squares, but must be computed directly, increasing the computational burden tremendously. Since the age fixed effects made little difference in the mean RD estimates, it seems likely that the costs of including them in the QRD estimates exceed any benefits.

No strong results emerge from the quantile analysis, which is summarized in Table 9. The causal effect of education on earnings is not estimated to be significant at any decile. One interesting pattern is that the analytical standard errors dramatically overstate the precision of the estimates when compared to the robust (i.e. dual) confidence interval. This points to the need to check that the analytical standard errors are similar to the dual confidence intervals before using the analytical variance-covariance matrix to test hypotheses across quantiles. Another interesting phenomenon is that the dual confidence interval of the lowest decile ($\tau = 0.1$) encompasses the right endpoint of the grid and that of the highest decile ($\tau = 0.9$) encompasses the left endpoint of the grid. This points to only partial identification of the treatment effects at these quantiles, which is somewhat surprising given the strength of the first stage.

6 Conclusion

This paper has shown that the quantile IV framework of Chernozhukov and Hansen (2005) can be adapted to a regression discontinuity design, allowing estimation of quantile treatment effects. Simulations show the estimator to be reasonably well-behaved, but an unexpectedly weak reduced-form relationship in the empirical application limited the ability of the estimator to provide insight.

This project could be usefully extended in two directions. The first is to apply the estimator in an RD setup where the forcing variable is continuous, rather than discrete as in this example. Jacob and Lefgren (2004) is an natural potential application, because its design is relatively transparent and quantile treatment effects are certainly interesting in the context of educational achievement. The second direction is to assess the validity of the method by comparing QRD estimates to quantile estimates obtained from a randomized trial. Here, the job training study Black et al. (2003) could be interesting, since Black et al. (2007) have shown that mean RD estimates match experimental estimates from the same sample reasonably well.

References

Daron Acemoglu and Joshua Angrist. How large are human-capital externalities? evidence from compulsory schooling laws. *NBER Macroeconomics Annual*, 15:9–59, 2000. ISSN 08893365. URL `http://www.jstor.org/stable/3585383`.

Joshua D. Angrist and Alan B. Krueger. Does compulsory school attendance affect schooling and earnings? *Quarterly Journal of Economics*, 106(4):979–1014, November 1991. URL `http://www.jstor.org/stable/2937954`.

Joshua D. Angrist and Victor Lavy. Using Maimonides' rule to estimate the effect of class size on scholastic achievement. *Quarterly Journal of Economics*, 114(2):533–575, May 1999. URL `http://links.jstor.org/sici?sici=0033-5533%28199905%29114%3A2%3C533%3AUMRTET%3E2.0.CO%3B2-3`.

David A. Belsley. *Conditioning diagnostics, collinearity and weak data in regression.* John Wiley and Sons, 1991.

Daniel Black, Jeffrey Smith, Mark Berger, and Brett Noel. Is the threat of reemployment services more effective than the services themselves? Evidence from random assignment in the UI system. *American Economic Review*, 93(4):1313–1327, 2003.

Daniel Black, Jose Galdo, and Jeffrey Smith. Evaluating the bias of the regression discontinuity design using experimental data. URL `http://socserv.mcmaster.ca/galdojo/docs/rddpaper.pdf`. Working paper, January 2007.

Sandra Black. Do better schools matter? Parental valuation of elementary education. *Quarterly Journal of Economics*, 114:577–599, 1999.

David Card, Carlos Dobkin, and Nicole Maestas. The impact of nearly universal insurance coverage on health care utilization and health: Evidence from Medicare. URL `http://www.nber.org/papers/w10365`. NBER Working Paper 10365, March 2004.

Victor Chernozhukov and Christian Hansen. An IV model of quantile treatment effects. *Econometrica*, 73(1):245–261, January 2005. URL `http://dx.doi.org/10.1111/j.1468-0262.2005.00570.x`.

Victor Chernozhukov and Christian Hansen. Instrumental quantile regression inference for structural and treatment effect models. *Journal of Econometrics*, 132(2):491–525, June 2006. URL `http://dx.doi.org/10.1016/j.jeconom.2005.02.009`.

Victor Chernozhukov and Christian Hansen. Instrumental variable quantile regression: A robust inference approach. *Journal of Econometrics*, 142:379–398, 2008.

Thomas Cook and Vivian Wong. Empirical tests of the validity of the regression discontinuity design. URL `http://www.northwestern.edu/ipr/publications/papers/2007/wp0702.pdf`. Northwestern University Institute for Policy Research Working Paper, 2007.

Jinyong Hahn, Petra Todd, and Wilbert Van der Klaauw. Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69(1):201–209, January 2001.

Joel L. Horowitz. The bootstrap. In J.J. Heckman and E. Leamer, editors, *Handbook of Econometrics*, volume V, chapter 52, pages 3159–3228. Elsevier, 2001. URL `http://dx.doi.org/10.1016/S1573-4412(01)05005-X`.

Guido W. Imbens and Thomas Lemieux. Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, 2007. URL `http://dx.doi.org/10.1016/j.jeconom.2007.05.001`.

Brian A. Jacob and Lars Lefgren. Remedial education and student achievement: A regression-discontinuity analysis. *Review of Economics and Statistics*, 86(1):226–244, January 2004. URL `http://dx.doi.org/10.1162/003465304323023778`.

David Lee and David Card. Regression discontinuity inference with specification error. *Journal of Econometrics*, 142(3):655–674, February 2008. URL http://dx.doi.org/10.1016/j.jeconom.2007.05.003.

Phillip Oreopoulos. Estimating average and local average treatment effects of education when compulsory schooling laws really matter. *American Economic Review*, 96(1):152–175, March 2006.

Phillip Oreopoulos. Do dropouts drop out too soon? International evidence from changes in school-leaving laws. URL http://www.nber.org/papers/W10155. NBER Working Paper, 2003.

Wilbert van der Klaauw. Estimating the effect of financial aid offers on college enrollment: a regression-discontinuity approach. *International Economic Review*, 43:1249–1287, 2002.

Figure 1:

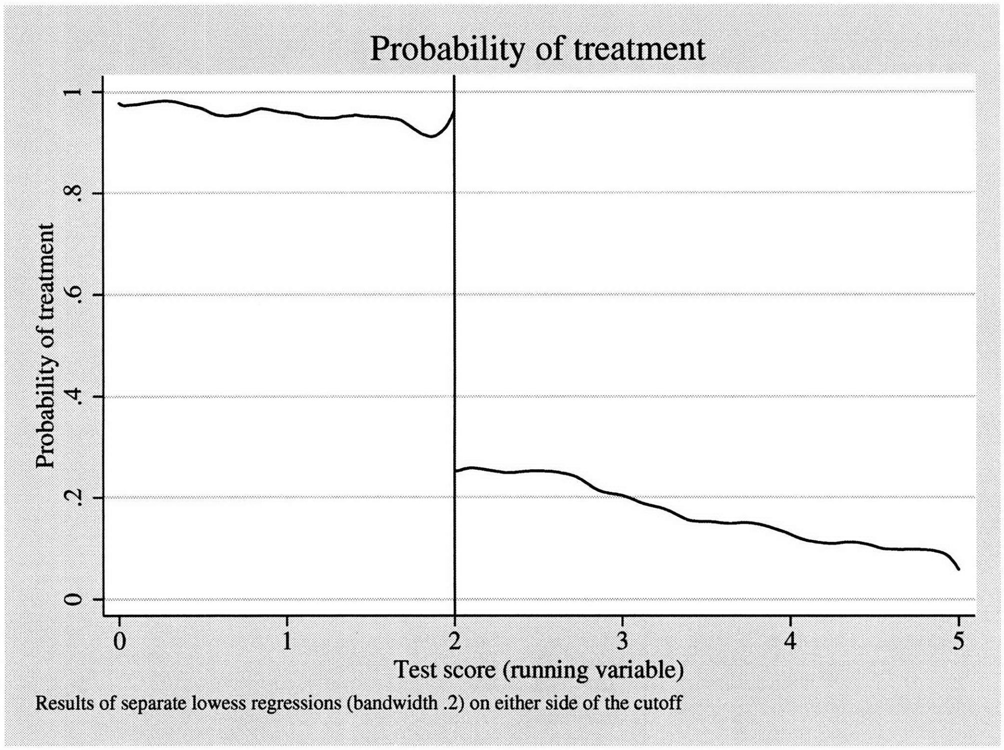


Figure 2:

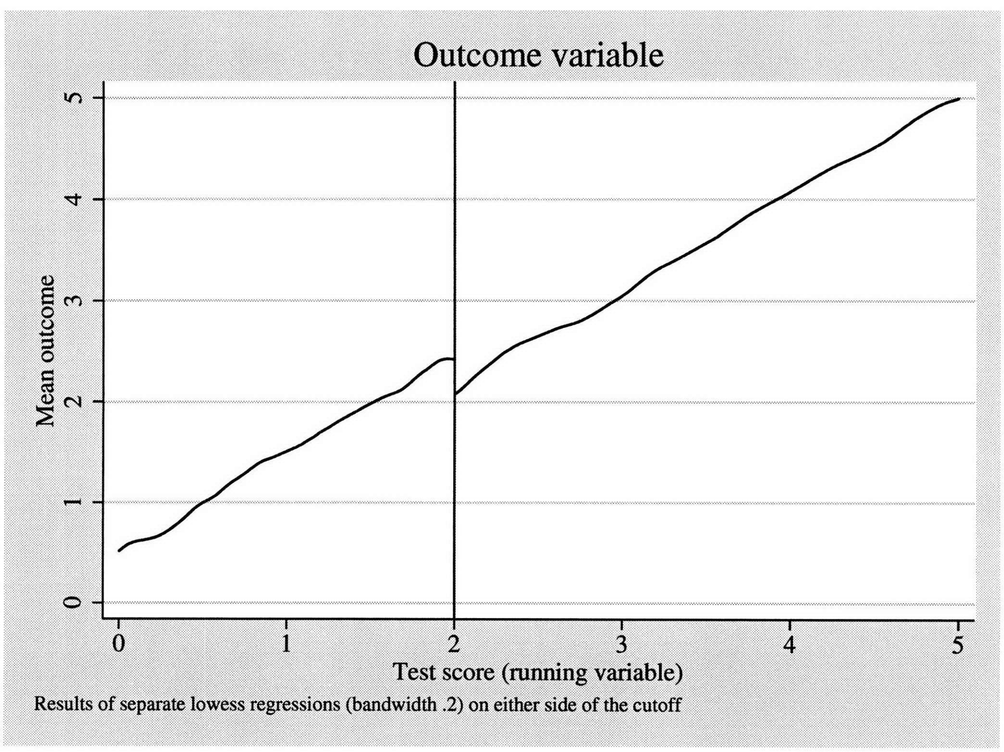
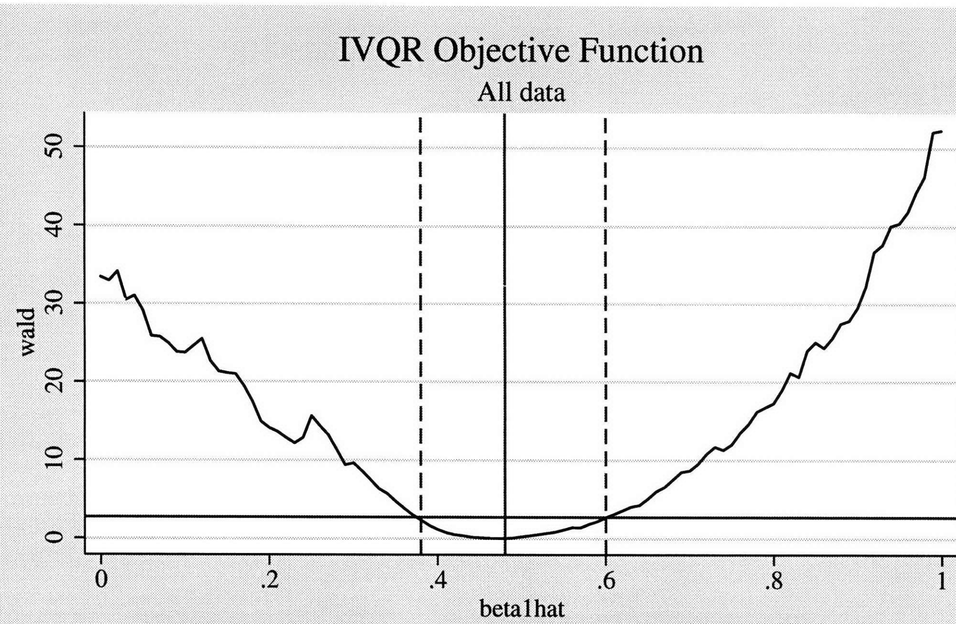
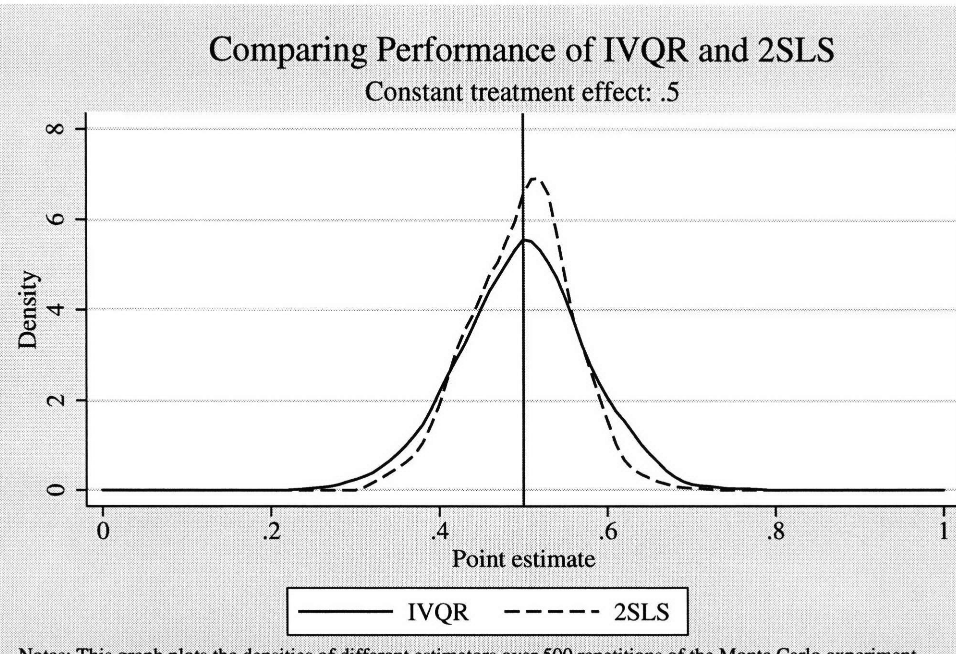


Figure 3:



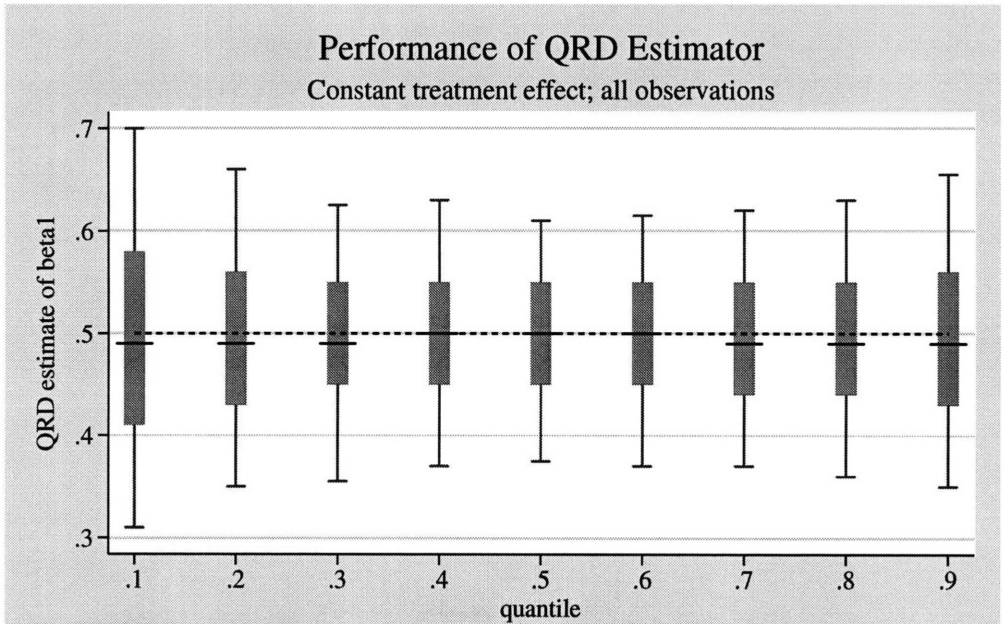
Notes: The solid vertical line marks the point estimate, .48; the dashed vertical lines show the 90 percent confidence interval, (.38,.6); the horizontal line shows the critical value for the chi–squared statistic, 2.706. The objective function, critical value line and the endpoint lines do not intersect exactly because of the discrete nature of the estimator.

Figure 4:



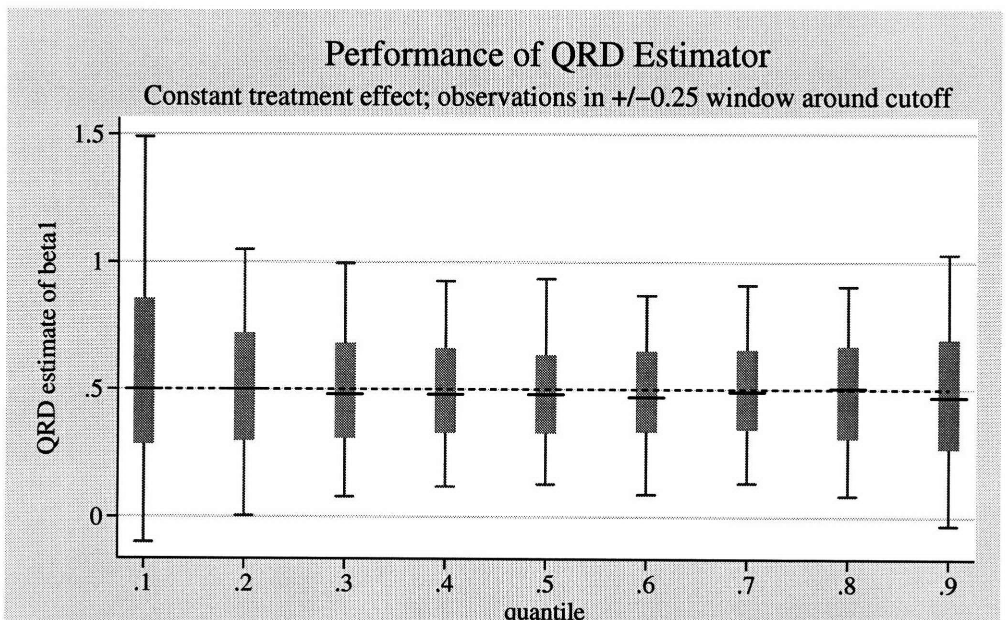
Notes: This graph plots the densities of different estimators over 500 repetitions of the Monte Carlo experiment. The sample size for each repetition is 10000. The vertical line marks the true value of the treatment effect, .5. The densities are kernel estimates.

Figure 5:



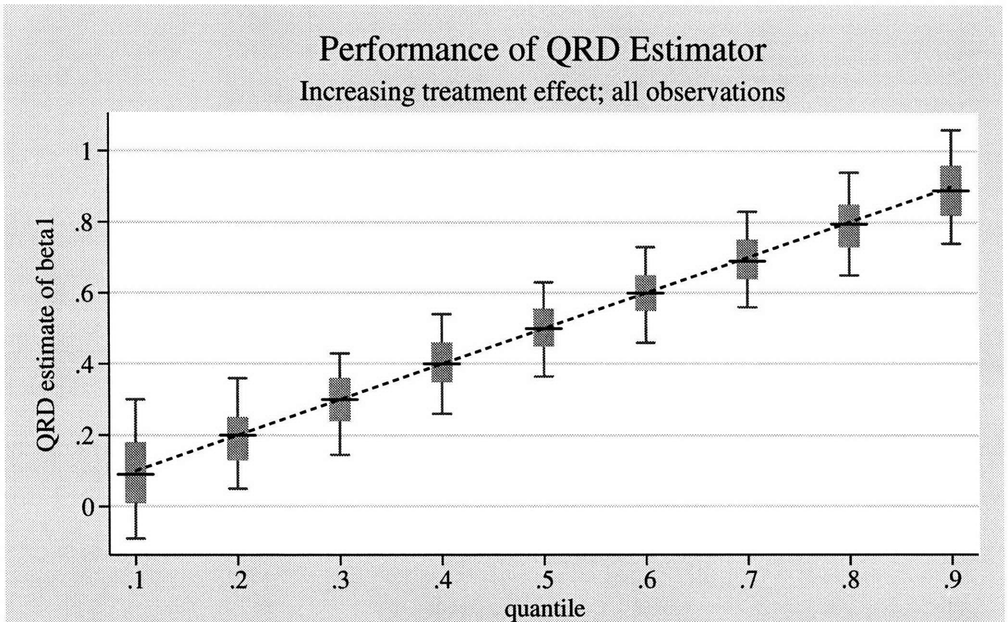
Notes: Box–and–whiskers plot of the distribution of the QRD estimate of beta1, by quantile, over 500 repetitions of the Monte Carlo experiment. The median IVQR estimate is indicated by the horizonal lines; the boxes indicate the 25th and 75th percentiles; the whiskers indicate the 5th and 95th percentiles. The true value of beta1 is indicated by the dashed line.

Figure 6:



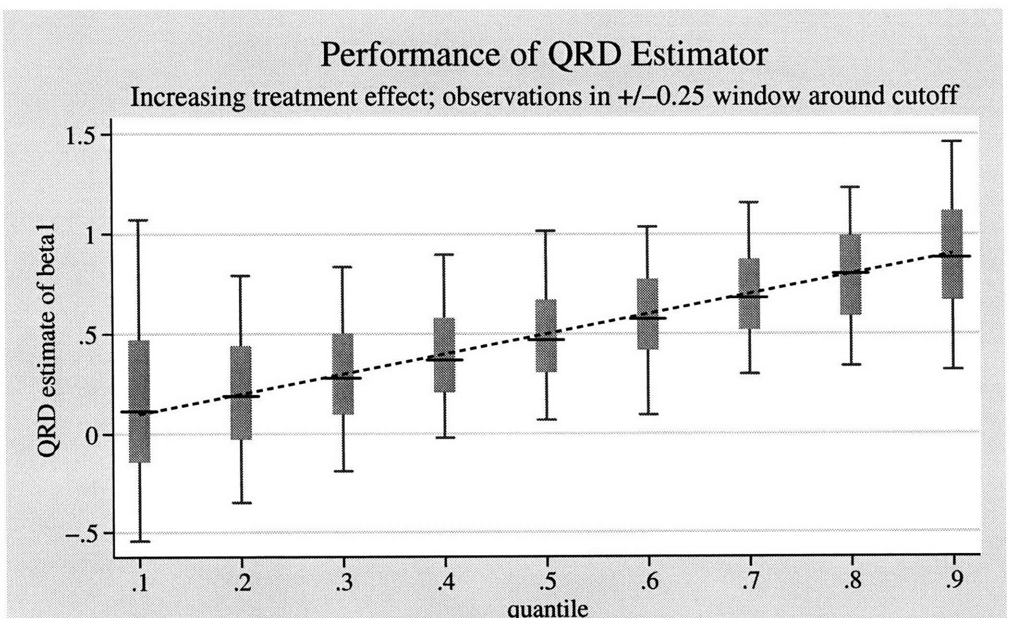
Notes: Box–and–whiskers plot of the distribution of the QRD estimate of beta1, by quantile, over 500 repetitions of the Monte Carlo experiment. The median IVQR estimate is indicated by the horizonal lines; the boxes indicate the 25th and 75th percentiles; the whiskers indicate the 5th and 95th percentiles. The true value of beta1 is indicated by the dashed line.

Figure 7:



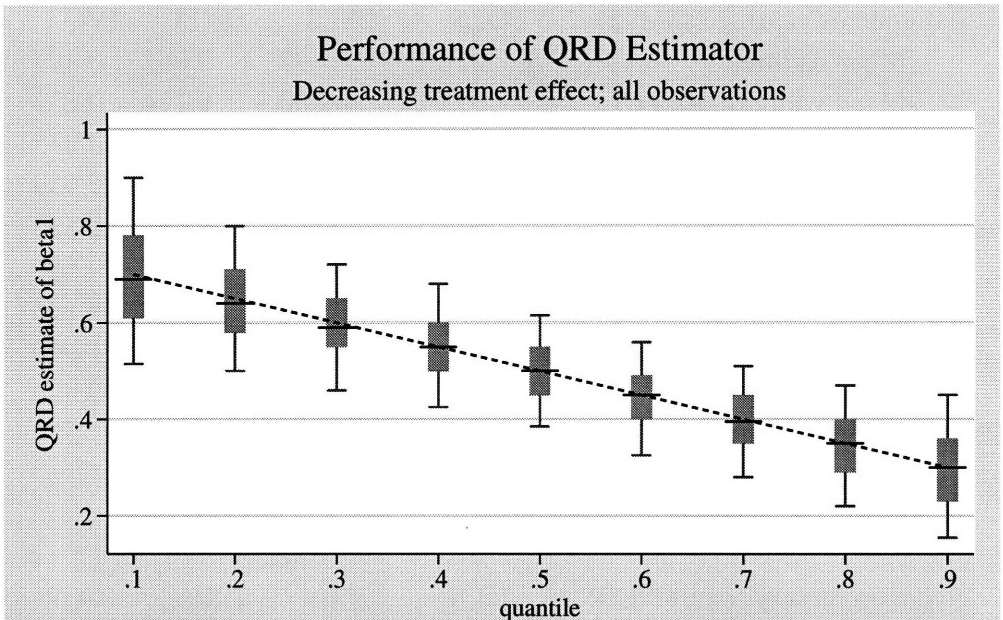
Notes: Box–and–whiskers plot of the distribution of the QRD estimate of beta1, by quantile, over 500 repetitions of the Monte Carlo experiment. The median IVQR estimate is indicated by the horizonal lines; the boxes indicate the 25th and 75th percentiles; the whiskers indicate the 5th and 95th percentiles. The true value of beta1 is indicated by the dashed line.

Figure 8:



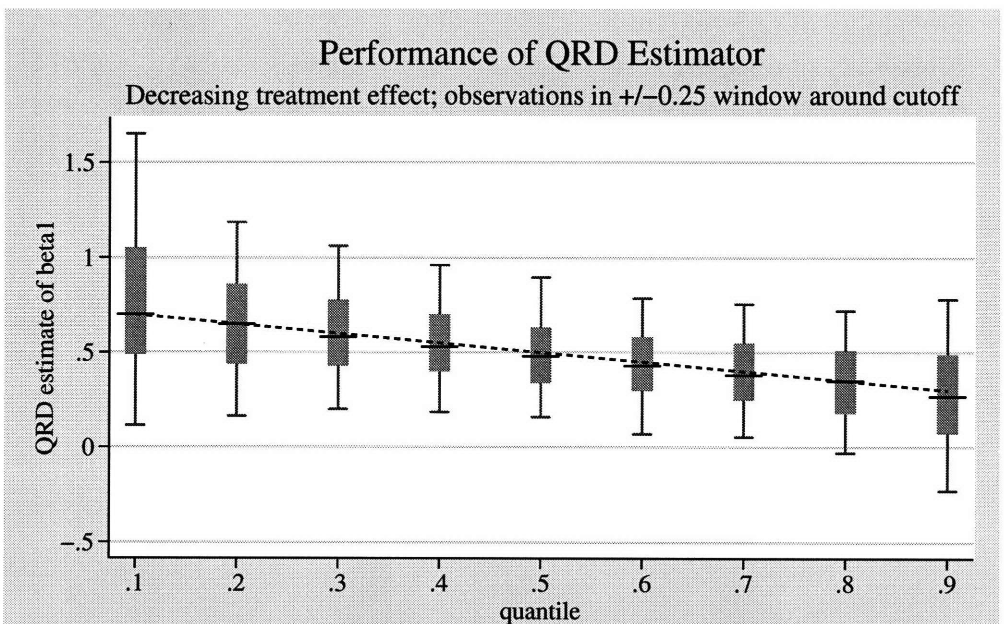
Notes: Box–and–whiskers plot of the distribution of the QRD estimate of beta1, by quantile, over 500 repetitions of the Monte Carlo experiment. The median IVQR estimate is indicated by the horizonal lines; the boxes indicate the 25th and 75th percentiles; the whiskers indicate the 5th and 95th percentiles. The true value of beta1 is indicated by the dashed line.

Figure 9:



Notes: Box–and–whiskers plot of the distribution of the QRD estimate of beta1, by quantile, over 500 repetitions of the Monte Carlo experiment. The median IVQR estimate is indicated by the horizonal lines; the boxes indicate the 25th and 75th percentiles; the whiskers indicate the 5th and 95th percentiles. The true value of beta1 is indicated by the dashed line.

Figure 10:



Notes: Box–and–whiskers plot of the distribution of the QRD estimate of beta1, by quantile, over 500 repetitions of the Monte Carlo experiment. The median IVQR estimate is indicated by the horizonal lines; the boxes indicate the 25th and 75th percentiles; the whiskers indicate the 5th and 95th percentiles. The true value of beta1 is indicated by the dashed line.

Table 1: Monte Carlo Results -- Median Only

A: Estimation using all observations

Estimate of beta1 (truth=0.5)	Estimator			
	OLS	2SLS	QREG	IVQR
Mean estimate	0.061	0.497	0.062	0.499
Mean width of 90% CI	0.091	0.200	0.114	0.234
RMSE	0.440	0.060	0.440	0.075
25th percentile	0.043	0.457	0.040	0.450
Median	0.060	0.503	0.060	0.500
75th percentile	0.081	0.535	0.087	0.550
Probability of rejecting truth	1.000	0.092	1.000	0.108
Probability of rejecting zero		1.000		1.000
Observations	10,000	10,000	10,000	10,000
Monte Carlo repetitions	500	500	500	500

B: Estimation using observations in +/- 0.25 window around cutoff

Estimate of beta1 (truth=0.5)	Estimator			
	OLS	2SLS	QREG	IVQR
Mean estimate	0.049	0.496	0.043	0.503
Mean width of 90% CI	0.258	0.662	0.329	0.815
RMSE	0.457	0.196	0.467	0.243
25th percentile	-0.006	0.357	-0.020	0.350
Median	0.052	0.501	0.042	0.500
75th percentile	0.101	0.623	0.104	0.655
Probability of rejecting truth	1.000	0.096	1.000	0.096
Probability of rejecting zero		0.802		0.614
Mean number of observations	998	998	998	998
Monte Carlo repetitions	500	500	500	500

Table 2: Comparison of IVQR Estimator Across Quantiles
Constant treatment effect: true beta1 = 0.5 for all quantiles

A. IVQR estimation using all observations

	Quantile								
Estimate of beta1	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Mean estimate	0.500	0.495	0.496	0.500	0.501	0.496	0.495	0.493	0.494
Mean width of 90% CI	0.460	0.311	0.270	0.247	0.232	0.230	0.233	0.248	0.295
RMSE	0.122	0.092	0.081	0.078	0.072	0.073	0.074	0.081	0.095
25th percentile	0.410	0.430	0.450	0.450	0.450	0.450	0.440	0.440	0.430
Median	0.490	0.490	0.490	0.500	0.500	0.500	0.490	0.490	0.490
75th percentile	0.580	0.560	0.550	0.550	0.550	0.550	0.550	0.550	0.560
Probability of rejecting truth	0.092	0.082	0.080	0.124	0.094	0.114	0.110	0.128	0.100
Probability of rejecting zero	0.996	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Observations	10,000	10,000	10,000	10,000	10,000	10,000	10,000	10,000	10,000
Monte Carlo repetitions	500	500	500	500	500	500	500	500	500

B: IVQR estimation using observations in +/- 0.25 window around cutoff

	Quantile								
Estimate of beta1	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Mean estimate	0.590	0.517	0.501	0.497	0.493	0.485	0.503	0.497	0.488
Mean width of 90% CI	2.016	1.187	0.960	0.867	0.816	0.806	0.829	0.910	1.125
RMSE	0.521	0.339	0.272	0.247	0.239	0.234	0.232	0.253	0.318
25th percentile	0.285	0.300	0.310	0.330	0.330	0.335	0.345	0.310	0.270
Median	0.500	0.500	0.480	0.480	0.480	0.470	0.490	0.505	0.470
75th percentile	0.855	0.720	0.680	0.660	0.635	0.650	0.655	0.670	0.695
Probability of rejecting truth	0.090	0.088	0.072	0.076	0.104	0.096	0.086	0.074	0.084
Probability of rejecting zero	0.362	0.500	0.564	0.592	0.598	0.610	0.618	0.532	0.420
Mean number of observations	1,001	1,001	1,001	1,001	1,001	1,001	1,001	1,001	1,001
Monte Carlo repetitions	500	500	500	500	500	500	500	500	500

Table 3: Comparison of IVQR Estimator Across Quantiles
Increasing treatment effect: beta1(τ) = τ

A. IVQR estimation using all observations

	Quantile								
Estimate of beta1	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Mean estimate	0.101	0.195	0.297	0.401	0.501	0.596	0.695	0.792	0.893
Mean width of 90% CI	0.491	0.322	0.284	0.264	0.253	0.251	0.254	0.269	0.312
RMSE	0.126	0.096	0.086	0.085	0.078	0.080	0.080	0.087	0.099
25th percentile	0.010	0.130	0.240	0.350	0.450	0.550	0.640	0.730	0.820
Median	0.090	0.200	0.300	0.400	0.500	0.600	0.690	0.795	0.890
75th percentile	0.180	0.250	0.360	0.460	0.555	0.650	0.750	0.850	0.960
True value	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Probability of rejecting truth	0.094	0.088	0.098	0.120	0.106	0.118	0.114	0.126	0.094
Probability of rejecting zero	0.202	0.632	0.948	1.000	1.000	1.000	1.000	1.000	1.000
Observations	10,000	10,000	10,000	10,000	10,000	10,000	10,000	10,000	10,000
Monte Carlo repetitions	500	500	500	500	500	500	500	500	500

B: IVQR estimation using observations in +/- 0.25 window around cutoff

	Quantile								
Estimate of beta1	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Mean estimate	0.195	0.210	0.299	0.397	0.493	0.585	0.702	0.796	0.887
Mean width of 90% CI	2.401	1.280	1.060	0.972	0.932	0.921	0.943	1.011	1.216
RMSE	0.549	0.356	0.303	0.283	0.274	0.270	0.265	0.286	0.348
25th percentile	-0.140	-0.025	0.100	0.210	0.310	0.420	0.520	0.590	0.670
Median	0.115	0.190	0.280	0.370	0.470	0.575	0.680	0.800	0.880
75th percentile	0.470	0.440	0.500	0.580	0.670	0.770	0.870	0.990	1.115
True value	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Probability of rejecting truth	0.104	0.094	0.076	0.082	0.104	0.098	0.094	0.068	0.094
Probability of rejecting zero	0.108	0.158	0.244	0.360	0.504	0.672	0.760	0.816	0.766
Mean number of observations	1,001	1,001	1,001	1,001	1,001	1,001	1,001	1,001	1,001
Monte Carlo repetitions	500	500	500	500	500	500	500	500	500

Table 4: Comparison of IVQR Estimator Across Quantiles
Decreasing treatment effect: beta1(0) = 0.75, linearly declining to beta1(1) = 0.25

A. IVQR estimation using all observations

	Quantile								
Estimate of beta1	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Mean estimate	0.701	0.645	0.597	0.550	0.500	0.446	0.396	0.344	0.294
Mean width of 90% CI	0.523	0.303	0.260	0.236	0.218	0.216	0.220	0.236	0.286
RMSE	0.121	0.090	0.078	0.074	0.068	0.070	0.071	0.077	0.092
25th percentile	0.610	0.580	0.550	0.500	0.450	0.400	0.350	0.290	0.230
Median	0.690	0.640	0.590	0.550	0.500	0.450	0.395	0.350	0.300
75th percentile	0.780	0.710	0.650	0.600	0.550	0.490	0.450	0.400	0.360
True value	0.7	0.65	0.6	0.55	0.5	0.45	0.4	0.35	0.3
Probability of rejecting truth	0.088	0.084	0.086	0.122	0.102	0.116	0.118	0.138	0.100
Probability of rejecting zero	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.960
Observations	10,000	10,000	10,000	10,000	10,000	10,000	10,000	10,000	10,000
Monte Carlo repetitions	500	500	500	500	500	500	500	500	500

B: IVQR estimation using observations in +/- 0.25 window around cutoff

	Quantile								
Estimate of beta1	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Mean estimate	0.789	0.662	0.601	0.547	0.491	0.436	0.398	0.345	0.285
Mean width of 90% CI	2.128	1.164	0.897	0.803	0.749	0.738	0.763	0.855	1.089
RMSE	0.510	0.321	0.254	0.229	0.219	0.216	0.216	0.236	0.310
25th percentile	0.490	0.440	0.430	0.400	0.340	0.300	0.250	0.180	0.075
Median	0.700	0.650	0.580	0.530	0.480	0.430	0.380	0.350	0.270
75th percentile	1.050	0.860	0.775	0.700	0.630	0.580	0.550	0.510	0.490
True value	0.7	0.65	0.6	0.55	0.5	0.45	0.4	0.35	0.3
Probability of rejecting truth	0.096	0.086	0.078	0.086	0.108	0.094	0.080	0.070	0.090
Probability of rejecting zero	0.626	0.758	0.764	0.724	0.688	0.616	0.516	0.352	0.220
Mean number of observations	1,001	1,001	1,001	1,001	1,001	1,001	1,001	1,001	1,001
Monte Carlo repetitions	500	500	500	500	500	500	500	500	500

Table 5 -- Estimated Effect of Minimum School-Leaving Age on Age Finished Full-Time Education and Log Earnings

Great Britain, ages 32-64, born 1921-1951, General Household Surveys 1983-1998

Sample population	First stage -- dependent variable: age finished full-time school			Reduced form -- dependent variable: log annual earnings			Initial sample size
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Great Britain	0.298 ***	0.289 ***	0.296 ***	0.025	0.026	0.030	66115
	[0.067]	[0.068]	[0.080]	[0.031]	[0.030]	[0.033]	
Number of cells	422	422	422	422	422	422	
Birth cohort polynomial controls	Quartic	Quartic	Quartic	Quartic	Quartic	Quartic	
Age polynomial controls	No	Quartic	No	No	Quartic	No	
Age fixed effects	No	No	Yes	No	No	Yes	

Notes: Sample: 32- to 65-year olds from the 1983 through 1998 General Household Surveys, aged 14 between 1935 and 1965. Data are aggregated into cell means and regressions are weighted by cell size. Robust standard errors clustered by birth cohort are in brackets.

Table 6 -- OLS and RD Returns to (Compulsory) Schooling, Estimates for Log Annual Earnings

Great Britain, ages 32-64, born 1921-1951, General Household Surveys 1983-1998

Sample population	Returns to schooling: OLS			Returns to compulsory schooling: RD			Initial sample size
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Great Britain	0.177 ***	0.175 ***	0.178 ***	0.086	0.096	0.099	66115
	[0.014]	[0.012]	[0.012]	[0.113]	[0.116]	[0.116]	
Number of cells	422	422	422	422	422	422	
Birth cohort polynomial controls	Quartic	Quartic	Quartic	Quartic	Quartic	Quartic	
Age polynomial controls	No	Quartic	No	No	Quartic	No	
Age fixed effects	No	No	Yes	No	No	Yes	

Notes: Sample: 32- to 65-year olds from the 1983 through 1998 General Household Surveys, aged 14 between 1935 and 1965. Data are aggregated into cell means and regressions are weighted by cell size. Robust standard errors clustered by birth cohort are in brackets.

Table 7 -- Estimated Effect of Minimum School-Leaving Age on Age Finished Full-Time Education and Log Earnings

Great Britain, ages 32-64, born 1921-1951, General Household Surveys 1984-1998

Sample population	First stage -- dependent variable: age finished full-time school			Reduced form -- dependent variable: log annual earnings			Initial sample size
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Great Britain	0.305 ***	0.293 ***	0.293 ***	0.041	0.041	0.039	60666
	[0.057]	[0.060]	[0.077]	[0.033]	[0.031]	[0.037]	
Number of cells	391	391	391	391	391	391	
Birth cohort polynomial controls	Quartic	Quartic	Quartic	Quartic	Quartic	Quartic	
Age polynomial controls	No	Quartic	No	No	Quartic	No	
Age fixed effects	No	No	Yes	No	No	Yes	

Notes: Sample: 32- to 65-year olds from the 1984 through 1996 General Household Surveys, aged 14 between 1935 and 1965. Data are aggregated into cell means and regressions are weighted by cell size. Robust standard errors clustered by birth cohort are in brackets.

Table 8 -- OLS and RD Returns to (Compulsory) Schooling, Estimates for Log Annual Earnings

Great Britain, ages 32-64, born 1921-1951, General Household Surveys 1984-1998

Sample population	Returns to schooling: OLS			Returns to compulsory schooling: RD			Initial sample size
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Great Britain	0.170 ***	0.170 ***	0.173 ***	0.125	0.130	0.167	60666
	[0.013]	[0.012]	[0.012]	[0.108]	[0.103]	[0.132]	
Number of cells	391	391	391	391	391	391	
Birth cohort polynomial controls	Quartic	Quartic	Quartic	Quartic	Quartic	Quartic	
Age polynomial controls	No	Quartic	No	No	Quartic	No	
Age fixed effects	No	No	Yes	No	No	Yes	

Notes: Sample: 32- to 65-year olds from the 1984 through 1996 General Household Surveys, aged 14 between 1935 and 1965. Data are aggregated into cell means and regressions are weighted by cell size. Robust standard errors clustered by birth cohort are in brackets.

Table 9 -- QRD Estimates of Effect of (Compulsory) Education on Earnings Quantiles

Great Britain, ages 32-64, born 1921-1951, General Household Surveys 1983-1998

Model A -- Exogenous Regressors: Year of Birth Quadratic

	Quantile								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Point Estimate	0.040	0.130	-0.020	-0.025	0.000	0.020	-0.015	-0.315	-0.045
Dual CI Left Endpoint	-0.100	-0.140	-0.110	-0.070	-0.065	-0.065	-0.135	-0.350	-0.350
Dual CI Right Endpoint	0.650	0.650	0.120	0.215	0.050	0.065	0.070	0.040	0.080
Analytical Standard Error	0.031	0.070	0.023	0.031	0.023	0.020	0.021	0.044	0.166

Model B -- Exogenous Regressors: Year of Birth Quadratic, Age Quartic

	Quantile								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Point Estimate	-0.015	0.125	0.055	0.010	0.020	0.030	0.010	0.005	-0.185
Dual CI Left Endpoint	-0.090	-0.050	-0.090	-0.055	-0.040	-0.010	-0.090	-0.350	-0.350
Dual CI Right Endpoint	0.650	0.640	0.175	0.215	0.070	0.080	0.105	0.060	0.085
Analytical Standard Error	0.032	0.044	0.047	0.032	0.018	0.014	0.019	0.011	0.020

Notes: Sample: 32- to 65-year olds from the 1983-1996 and 1998 General Household Surveys, aged 14 between 1935 and 1965. Results are from a QRD/QIV regression of log annual earnings on age leaving full-time school, controlling for the exogenous regressors noted. An indicator for year of birth after 1931 instruments for age leaving school. QRD/QIV implemented using a grid of 201 points at 0.005 log point intervals, ranging from -0.35 to 0.65.

Chapter 3:

The Short- and Medium-Run Benefits of Sanitation

Co-authored with Esther Duflo and Michael Greenstone, both of the MIT Department of Economics

Introduction

Lack of access to clean water and sanitation are two of the most significant threats to health and welfare in the developing world. Approximately 17 percent of the world's population lack access to improved water sources and 42 percent lack access to sanitary facilities. These privations have serious consequences: diarrheal diseases kill roughly 1.8 million people each year, and diarrhea causes 17% of all deaths of children under five years in developing counties (WHO 2005). Poor water quality also contributes to other major diseases, including malaria, schisosomiasis, trachoma and worms (WHO 2004). Indeed, the Millenium Development Goals call for a fifty percent reduction in the share of households worldwide without access to safe drinking water (UN 2007).

While a strong consensus exists regarding the severity of the problem, there is little robust evidence on the effectiveness of potential solutions relevant to the rural poor. Piped water and sanitation infrastructure have proved effective when the correct institutional incentives are present (Cutler and Miller 2005; Watson 2006; Galiani, Gertler and Schagrodsky 2005). However, the large-scale infrastructure investments required to provide in-home solutions are commonly believed to be too expensive to construct or too difficult to maintain in a rural context, where homes are dispersed and institutional support necessary for ongoing maintenance may be weak (Clasen and Haller 2008). As a result, most studies have focused on "community-level infrastructure," solutions such as communal well improvement that are short of piped water. The results have not been promising: in fact, a recent literature review concludes that "there is little evidence that providing community-level rural water infrastructure substantially reduces diarrheal disease or that this infrastructure can be effectively maintained. Investments in communal water infrastructure short of piped water may serve other needs and may reduce diarrhea in particular circumstances, but the case for prioritizing communal infrastructure provision needs to be made rather than assumed" (Zwane and Kremer 2008).[1]

---

[1] One promising avenue for further research is to investigate the effectiveness of *household* technologies, such as in-home chlorination (Ashraf, Berry and Shapiro 2007; Kremer, Miguel, Null and Zwane 2008), or improving

This paper presents the first medium-run evaluation of piped water and sanitation infrastructure in a rural context. We examine a series of village-level interventions conducted by an Indian NGO, Gram Vikas (GV), in Orissa, one of the poorest states in India. The program, called the Rural Health and Environment Program (RHEP), consists of construction of a village water tank, household toilet and bathing facilities with a tap for drinking water and piping to connect the household facilities to the central water tank.

Several unusual features of the program make it compelling to analyze. First, the program requires all households in affected villages to participate, so the estimated impacts capture all within-village externalities. Second, service is initiated suddenly, unpredictably and at different times for different villages, reducing concerns of confounding with other trends. Finally, and perhaps most importantly, data on all outcomes are available beginning several years before the sanitation project arrives and data collection continues for five years or more after service begins, which makes it possible to assess medium-run impacts.

We find that RHEP led to substantial reductions in water-related disease. Episodes of severe diarrhea declined by roughly one-third and serious cases of malaria declined by 20-40%. These results are evident in the very first month that the sanitation system became operational. Further, they are lasting as they are still evident five years later. These benefits were achieved at an annual cost per household of roughly US$40 per year. While these results are notable, the program was implemented in a selected sample of villages that may have stronger institutional capacity than the typical village.

The paper proceeds as follows. Section I describes the Rural Health and Environment Program and how the features of the program noted above may allow for a credible impact evaluation. Section II details the data sources and provides some summary statistics. Section III describes the econometric methods and reports the empirical findings. Section IV discusses the results, provides a back-of-the-envelope cost assessment and concludes.

---

individual hygienic practice, such as hand-washing (Luby et. al. 2006).

I. The Rural Health and Environment Program

A. Program History, Goals and Components[2]

The Rural Health and Environment Program is an initiative of Gram Vikas, an NGO located in Orissa, India. Orissa is perhaps the poorest state in India: over 46% of its population has Below Poverty Line status, compared to the national average of 27.5%.[3] GV works with rural populations, many of them members of Scheduled Castes and Scheduled Tribes[4], among which poverty rates are even higher: in every RHEP village, at least 70% of residents are below the poverty line. Having worked with many of these villages in the late 1980s and early 1990s in the development of biogas plants, GV determined that waterborne diseases were major sources of mortality and morbidity, and that the establishment of protected drinking water and improvement of sanitary conditions had the potential to improve villagers' well-being.

As a response, GV developed the intervention package now known as RHEP. The basic components of the program are a water tank storing water pumped from a safe source, piping from the tank to each home in the village, and toilet and bathing rooms for each home. The bathing house is perhaps unusual in a water intervention. Because of the high prevalence of gynecological and skin diseases, GV, attributing these diseases to the need to bathe in the polluted waters of local ponds, deemed it important for women to have a private space and clean water for bathing.

The program was piloted in 5 villages (340 households) in 1992. It was then expanded in four phases: adding 40 villages (3,000 households) from 1995-1998, 27 villages (2,000 households) from 1999-2001, 38 villages (3,000 households) from 2001-2003 and 160 villages (8,000 households) from

---

[2] Information in this section comes from Gram Vikas documents (Gram Vikas 2001, Gram Vikas 2004 and Gram Vikas 2005) and interviews of Gram Vikas management and personnel by the authors.

[3] Below Poverty Line (BPL) is an official status that confers access to certain grants and subsidies. The precise definition varies from location to location, but can be generally understood as lacking the income necessary for subsistence. For rural Orissa in 2004-2005, the official poverty line was 326 Rs. / month, or roughly US$0.24 per day at market exchange rates or US$0.72 per day at purchasing power parity exchange rates. Source: "Poverty Estimates for 2004-5," Planning Commission, Government of India, March 2007. http://www.planningcommission.gov.in/news/prmar07.pdf. Accessed: 8 July, 2008.

[4] Scheduled Castes, or *dalits*, are castes that have faced discrimination and receive government affirmative action; Scheduled Tribes, or *adivasis*, are indigenous groups and also receive affirmative action. Both are official designations.

2003-2006.

Because the implementation timeline is important for our identification strategy, we provide some detail on the typical sequence of events for implementing RHEP in a village. First, GV extension workers identify a village well-suited to the program. Desirable characteristics are a strong sense of community unity and good village leadership. This information may come from GV contacts with local elected leaders, state government officials or pre-existing relationships with villages established through other programs such as biogas. Sometimes representatives of villages neighboring existing RHEP villages will actively seek out Gram Vikas.

Next, GV representatives meet informally with village leadership. If there is interest in starting a program, a series of village meetings will be held to build participation and obtain 100% consensus for participation in the program. This process of consensus-building usually takes 3-6 months, but can take up to 18 months. GV insists on 100% participation because of the externalities inherent in sanitation: if even a few villagers do not participate, they can transmit disease to their neighbors, reducing the benefits of the program.

Once the village has reached 100% consensus, GV and the village enter a formal agreement. At this time, data collection begins. The villagers begin construction of the toilet and bathing facilities, following GV designs and guidance but using their own labor and locally acquired materials. Simultaneously, the village must raise a "corpus fund" with average contributions of $20 per family.[5] Wealthier households are expected to subsidize poorer households, although the exact subsidy program is left up to village agreement. This fund is invested, with the interest set aside to be used to extend service to any households that may be formed in the village after the initial construction and activation is complete. Once all households in the village complete the walls of the toilet and bathing house (i.e. the external structure is complete except for the roof), GV releases a cash subsidy of $50-60 per household, which is used to provide inputs the villagers cannot provide themselves (e.g. ceramics, piping materials,

[5] All monetary figures are given in 2005 U.S. dollars, converted from Indian rupees at the 1997-2006 average market exchange rate of 45Rs./USD.

etc.). The total cost of each structure is roughly $140, including the value of the villagers' unskilled labor (priced at a wage rate of one dollar per day) and the materials gathered (rough stone, etc.).

Construction of the water supply usually begins when GV releases the subsidy and runs in parallel to the completion of the toilet and bathing houses. The water supply consists of a safe water source, a central water tank for the village, and piping from the tank to individual households. Since the groundwater in Orissa is generally of good quality, the water source is usually a deep tubewell, although in some hilly areas higher-elevation natural springs can feed the tank via gravity, making a motorized pump unnecessary. The water tank (usually 80,000 liter capacity) is piped to each house, with each house receiving three water taps: one each for the toilet and bathing rooms, and one inside the home. As with the sanitary houses, villagers supply unskilled labor and locally available materials while Gram Vikas provides skilled labor and specialized materials. Construction of the water supply requires 3-4 months of work, but usually takes roughly one calendar year to complete because work is not continuous – villagers supply their own labor outside of planting and harvest seasons. The cost of the water tank is about $175 per household, again including the value of villagers' labor and materials they collect.

Although some families inevitably complete their sanitary houses sooner than others, GV as a rule does not activate the water supply until *every* households have completed its sanitary facility. GV believes that households are primarily interested in obtaining running water for their homes, so turning on the water to the home before the sanitary facility is complete could lead to villagers not completing the project. Furthermore, GV believes that social pressure or cross-subsidization can help push lagging households towards completion, and these forces would be muted if leading households received water service. On the other hand, this delay does come with a cost, since some households with completed sanitary houses are unable to connect to the water tank until the other households in the village complete their structures.

Finally, villagers are responsible for operational costs, such as electricity to power the water pump, maintenance and repairs, usually amounting to about $10 per household per year. GV provides some training for villagers to provide maintenance and advises village committees on governance, but does not directly intervene after the project is complete, other than to collect data.

B. Challenges of Evaluation and Research Design

An evaluation of RHEP offers several advantages over previous efforts to measure the causal effect of water and sanitation programs. Accurate measurement of the impact of water and sanitation programs has proven difficult for several reasons, in particular the absence of a valid counterfactual for outcomes in the absence of the sanitation program, the problem of measurement in the presence of externalities, small sample sizes and short followup periods.[6] The design of RHEP and GV's data-collection efforts allow us to address each of these concerns, as we discuss in turn.

First and perhaps foremost, valid counterfactuals are not easy to obtain. Comparisons of communities with and without a package of water and sanitation services are likely to be biased, because communities that receive water and sanitation programs frequently differ significantly from those that do not, both along observable and unobservable dimensions. Comparisons of outcomes before and after an intervention, while controlling for time-invariant community unobservables, may confound program effects with the evolution of other determinants of health. However, in the case of RHEP, the package of water and sanitation services arrived suddenly, in principle on a single day for all households. Since it is unlikely that other determinants of health outcomes would change simultaneously and just as suddenly, an event study in this context has greater validity.[7] In addition, because the time at which the water will be turned on is not known in advance, we are less concerned that anticipatory adjustment of compensating behaviors (such as in-migration of sicklier relatives) will mask the program's effect.

Second, studies that compare households with and without sanitation services within a given community will overlook potential external benefits to non-recipient households, which may be less

---

[6] Our discussion is brief. For a more complete survey of the literature and associated issues, see Zwane and Kremer (2007).

[7] Gram Vikas works with RHEP villages on other projects, such as schoolbuilding, skills training, aquaculture and sustainable forestry. While these projects are generally less directly related to the health outcomes we study, there could be indirect effects, for example if increased income leads to the ability to purchase better health care services. However, because the timing of these programs is not coordinated with switching on the water and because these programs are more gradual, we are less concerned that we are confounding the effect of RHEP with the effects of these programs. Furthermore, as detailed in the econometric methods section, we control for a linear trend in health outcomes prior to the arrival of the water, which will pick up impact of these other programs to the extent that they arrive gradually and have roughly linear effects on outcomes.

likely to contract communicable disease.[8] Because RHEP is implemented and our data are recorded at the village level, we can capture the total effect of the intervention within the village, i.e. direct benefits as well as within-village externalities. Because villages are usually distant from each other, it is reasonable to assume that most of the external benefits to improved sanitation will be within the village, i.e. between neighbors rather than from residents of one village to another. This is especially likely in the Orissan context because children, who are potential disease vectors, usually attend school within their own village.[9] However, it is possible that there are some external benefits we are not capturing, for example via rivers or interactions on market days.[10]

Third, many studies of sanitation programs suffer from small sample sizes and short spans of data. Small samples make precise measurement difficult, while short follow-up periods make it impossible to assess anything other than the program's short-run impact. We have usable data from roughly 100 RHEP villages (see the following section for a detailed discussion of which RHEP villages are included in our sample), which improves the precision of our estimates relative to studies based on just a handful of sites. Additionally, because data collection begins when GV and the village enter the formal agreement – usually at least a year before the water is turned on and continues for three years or more after the water arrives, we can assess the medium-run impact of the program rather than being limited just to a short-run evaluation.

II. Data Sources and Summary Statistics

A. Data Sources[11]

The data for our study come from Gram Vikas's internal "Monthly Progress Reports" (MPRs),

---

[8] In fact, simply comparing the health status of the two households would doubly undercount the benefits. Rather than *adding* the improvement of the neighbor to the program's measured benefits, the neighbor's improvement would be *subtracted* from the program's tally, since we are comparing the outcomes of the two households with each other.

[9] See Miguel and Kremer (2004) for a study of how to measure cross-village externalities when children from different villages attend the same school. Their study is of Busia in western Kenya.

[10] While we would undercount these cross-village benefits, we would only doubly undercount (as in the discussion in footnote 8) if the benefits redounded to someone from another RHEP village included in the study. This is somewhat unlikely, since our study is of roughly 100 villages out of literally thousands in the state.

[11] As in section I.A, information in this section comes from GV documents (Gram Vikas 2001; Gram Vikas 2004; Gram Vikas 2005) and interviews of GV management and personnel by the authors.

which we scanned and had double-entered by a data-entry firm. These MPRs were compiled on a monthly basis by GV personnel, either during monthly (or more frequent) visits or by GV staff residing on-site. The MPRs contain detailed information on the status of RHEP (e.g. water tank construction, number of households with and without competed toilet and bathing houses) and incidence of disease, in particular water- and sanitation-related diseases such as diarrhea. Two aspects of this data-collection process are particularly useful for our study. First, as mentioned above, data collection typically begins a year or more before the water is activated and continues for at least three years thereafter. Second, the purpose of these records is to provide information on the project's progress and conditions in the village. Because the records are collected for internal purposes and not with an external evaluation in mind, we are confident that they do not exaggerate the benefits of the program.

While the monthly data collection effort covers many aspects of village life, we focus our exposition on the key data for our analysis: the stage of completion of RHEP and the occurrence of water-related diseases. When a formal RHEP agreement is reached with a village, GV conducts a census of all households and assigns an employee to supervise the project in that village. This village supervisor visits the village frequently: usually once per week, although he or she may even take up part- or full-time residence in the village. The supervisor goes house-to-house and records each household's progress towards completing its individual sanitary house. The supervisor also records progress in construction of the village water tank and piping. In our analysis, we code water improvement as beginning in the first month when the water tank is recorded as operational.[12]

Data on disease prevalence are also recorded monthly by the GV village supervisor, who draws on a variety of sources. The primary source is the records of the village health clinic, but the supervisor supplements this with discussions with the village health committee, village leadership and individual visits with villagers themselves. Our analysis focuses on diarrhea and malaria, since these are the most common water- and sanitation-related diseases for which data are recorded.[13] It is important to note that

[12] For robustness, we also perform the analysis using alternative indicators for water improvement, such as water tank completion and sanitary facility completion. The results are not substantially affected. See the results section.
[13] Other diseases monitored include fever, typhoid, jaundice, cold and cough, night blindness, scabies, TB and leprosy. We do not include these in the main analysis because they are either too scarce for any reasonably-sized

disease episodes are only recorded if a resident is "checked" or "treated". "Checked" means that the villager went for a consultation with a health worker or a doctor. "Treated" means that the villager received treatment for the condition, again either from a health worker or from a doctor. This excludes cases for which a villager does not seek a medical consultation, or cases when a villager seeks treatment via home remedies or local healers. As a result, our measures likely understate the true prevalence of these ailments, since time and expense mean that villagers are likely to seek treatment only in the most serious cases.[14] We focus on the "treated" measure in our analysis but the results using the "checked" variable are very similar, which is not surprising since the two measures have at least a 90% correlation with each other in our data (see Appendix Table 1).

After our data-entry firm converted the paper records to electronic format, we performed some cleaning operations to check for outliers, reconcile inconsistencies, etc. These operations are detailed in the Data Appendix.

B. Sample Selection and Summary Statistics

The primary criterion for inclusion in our sample is that we can identify the month in which the program "switches on," i.e. the water supply is activated. We will use τ to indicate the month relative to the month in which the water switches on, so $\tau = -1$ is one month before the water turns on, $\tau = 0$ is the month in which the water turns on, $\tau = 1$ is one month after. We have 97 such villages, with 5,999 village-by-month observations total. Some early RHEP villages are excluded because recordkeeping did not begin until 1997. For many of these villages, we know only that the water was running at the beginning of 1997, but we do not know when the water supply started nor do we know any pre-intervention outcomes.

Figure 1 provides information on the distribution of these observations. The upper left-hand graph

---

treatment effect to be detected or are only weakly or indirectly related to water and sanitation conditions. Similarly, data are recorded on mortality, child mortality, pregnancies carried to term and other outcomes of interest, but given the rarity of these events, we do not have statistical power to detect any reasonably-sized effect.

[14] In principle, medical treatment is available at Primary Health Centres (PHCs), which are typically no farther than five to 10 kilometers from a village. However, medical personnel and doctors are frequently absent from these PHCs, so villagers usually travel to the nearest town to seek treatment. This may involve a journey of up to 70 kilometers each way, as well as a minimum expenditure of 200 Rs. for treatment. For documentation of these and other difficulties in obtaining health care in rural India (albeit in Rajasthan, not Orissa), see Banerjee, Deaton and Duflo (2004).

shows the number of observations month-by-month. Our sample grows over time, as more villages enter RHEP and data collection continues for earlier participants. The lower left-hand graph also plots the number of village-by-month observations for this sample, but this time arranged relative to the month in which the water began to flow, i.e. in terms of τ. By definition, the graph has its mode at $\tau = 0$, since a report in month $\tau = 0$ is necessary for the village to be included in the analysis. The graph shows that it is not uncommon to have data as much as a year before and four years after the water arrives. Finally, the upper right-hand graph shows the distribution of months in which water supplies were activated for villages in our sample. Importantly, there is no visible pattern of seasonality which might bias our results. Nevertheless, we will include month-by-year fixed effects in our analysis.

Table 1 provides descriptive statistics for this sample and two sub-samples: columns (1)-(3) correspond to Sample A, the "all observations with known τ" sample described above; columns (4)-(6) correspond to Sample B, in which we use only observations from 2 years prior to 5 years after the activation of the water supply (i.e. $\tau \in [-24,59]$) and only use observations from villages from which we have at least six observations before and six observations after the activation of the water supply; columns (7)-(9) correspond to Sample C, in which we further restrict the sample to observations within one year of activation of the water supply (i.e. $\tau \in [-12,11]$), with the same villages as in columns (4)-(6).

In sample A, the average village contains roughly 85 households and 477 residents, of which 178 are children. Reported school attendance is low, averaging less than 7 days per month, which is not surprising considering high rates of teacher absenteeism (Chaudhury et. al. 2006; Duflo, Hanna and Ryan 2008). The prevalence of the water-related diseases of interest are perhaps lower than expected, likely because only the most severe cases are reported. For example, the typical village-by-month observation has less than one case of diarrhea checked or treated. This is surprising considering the overall poor state of health in the villages: for example, 33 percent of children are borderline malnourished and an additional 12 percent are severely malnourished. Samples B and C are broadly similar to Sample A, although the number of villages and observations necessarily fall.

III. Econometric Methods and Empirical Results

A. Econometric Methods: Our estimates of program impacts follow two main strategies: an event study analysis in which we estimate month-by-month effects; and a panel approach in which we summarize the overall impact of the program in the difference between mean outcomes before and after the intervention. The event study analysis involves estimation of regression equations similar to

$$y_{vt} = \sum_{s=-12}^{+11} \alpha_s \tau_{vs} + \gamma V_v + \delta T_t + \varepsilon_{vt} \quad (1)$$

where y_{vt} denotes the outcome of interest, e.g. monthly cases of severe diarrhea, for village v in month t, τ_{vs} is an indicator for the s^{th} month after the water improvement begins (starting at zero, i.e. s=0 is the month in which the water turns on) in village v, V_v is a fixed effect for village v, T_t is a fixed effect for month t and ε_{vt} is the disturbance term for village v in month t. We omit the $\tau_{v,-1}$ indicator from the regression, so our estimates of the coefficients α_s are interpreted as the mean of the outcome variable relative to the month before the water turns on.

A few aspects of equation (1) merit discussion. First, the presence of the village fixed effect V_v allows us to control nonparametrically for any mean differences across villages. Second, having multiple observations in each month (from different villages) allows us to include year-by-month fixed effects T_t to control nonparametrically for aggregate monthly shocks across villages in our sample, for example from a particularly heavy monsoon in a given month. Third, and most important, the combination of (a) our knowledge of the sudden change in the availability of sanitation in month $\tau_{v,0}$ and (b) the quick biological response of our main outcome variables of interest, diarrhea and malaria, to a change in the environment together give us a sharp prediction of a near-immediate change in the prevalence of these diseases, which would be apparent in our results by negative estimates of α_1, α_2, etc. Since we have relatively few observations for each τ, as the lower left-hand graph in Figure (1) indicates, and we have not imposed any restrictions at all on the month-by-month effects, we would not expect these month-by-month estimates to be very precise. However, this regression will be informative about the integrity of the identification strategy, since it may provide information on pre-program trends that might threaten our experimental validity, and also will allow us to examine visually whether program impacts appear to persist.

Our second approach is to estimate regressions such as

$$y_{vt} = \alpha POST_{vt} + \gamma V_v + \delta T_t + \varepsilon_{vt} \tag{2}$$

where $POST_{vt}$ indicates that the water supply has turned on in village v in month t, and all other variables are as in equation (1). This will collapse the month-by-month (relative to program initiation) estimates from equation (1) into a single, succinct measure of the program's impact. For example, in Sample C, $\widehat{\alpha}$ represents our estimate of the program's impact in the first year, relative to the mean in the year leading up to the activation of services.

We also estimate an important variant of the equation (2) in which we add a linear trend in τ_{vt}, as in

$$y_{vt} = \alpha POST_{vt} + \beta \tau_{vt} + \gamma V_v + \delta T_t + \varepsilon_{vt} \tag{3}$$

where τ_{vt} denotes how many months have passed since the water tank turned on village v. This allows us to control for any linear trend in outcomes around the arrival of water services. For example, imagine that the consensus-building necessary to launch RHEP improves local institutions, which in turn improves health incomes (e.g. villagers trust each other more and therefore insure each other better, leading to smoother consumption and therefore less malnutrition). Simply comparing outcomes before and after the water turns on would then incorporate these health improvements, which should not be attributed to RHEP *per se*. By adding a linear trend in τ, we can control for these possible confounding effects, providing they are approximately linear in the neighborhood of the intervention. Other programs or behaviors that tend to "switch on" abruptly and simultaneously with RHEP would not be adequately controlled for by this approach, but we have no reason to believe that this has taken place.

In order to formalize our analysis of the persistence of program effects, we also estimate models separating the program effect in various sub-periods. For example, to see whether program effects persist after 3 years, we estimate the following variant of equation (2):

$$y_{vt} = \alpha_0 1\{0 \le \tau_{vt} < 36\} + \alpha_1 1\{\tau_{vt} \ge 36\} + \gamma V_v + \delta T_t + \varepsilon_{vt} \tag{4}$$

where α_0 represents the impact of the program in the first three years and α_1 the impact thereafter.

Several additional details relevant to the precision of the resulting estimates are worth noting. First by using village-level aggregates as our unit of observation, we do not commit the error of treating each household in the village as an independent observation (Kloek 1981; Moulton 1990). Second, because village outcomes are likely correlated over time, we compute standard errors clustered at the village level, making our estimated standard errors robust to arbitrary autocorrelation patterns (Bertrand, Duflo and Mullainathan 2004). Third, we weight each village-by-month observation by the inverse of the number of observations of that village, as in probability-weighted sampling. We do this because an observation from an infrequently-sampled village contains more information relative to an observation from a frequently-sampled village and this is likely to improve the precision of our estimates.

B. Empirical Results

Our empirical analysis will focus on treated cases of diarrhea and malaria. We focus on these two variables for three reasons. First and foremost, the health literature gives us strong *ex-ante* reason to believe they are linked to water and sanitary conditions (WHO 2004). Second, the response of diarrhea and malaria to improvements in water quality is likely to be rapid, unlike other plausibly liked health outcomes such as malnutrition. This is important because our identification strategy is cleanest near the month in which the water switches on. Third, these conditions are relatively common in our data, at least relative to other important outcomes deaths and infant deaths. This gives us sufficient statistical power to detect program effects of a plausible magnitude – detecting effects on very rare events would require a much larger sample than we have.

We begin with discussion of the event-study methodology described by equation (1) above. We run regressions using cases of diarrhea treated and cases of malaria treated as the independent variables and over Samples A, B and C as defined in section I.B.[15] The estimation results are tabulated in Table 2, but are perhaps more clearly communicated in graphical form, provided in Figure 2, which graphs the point estimates and 95% confidence intervals for $\beta_{-12},\ldots,\beta_{+11}$ for Sample C.[16] Although the individual estimates are not too precise, we see a clear downward shift after the water turns on. Reassuringly for the validity of our approach, we do not see evidence of a downward trend prior to the start of improved water,

[15] As discussed in Section II.A, we report only results for the "cases treated" measures here. However, we run all regressions using the "cases checked" measures as well, and the results are very similar. For brevity, we do not report these results here, but all results are available in the Online Supplement, which can be downloaded at http://web.mit.edu/guiteras/Public/RHEP-supplement.pdf.

[16] $\beta_{-1} = 0$ by definition, since $\tau_{v,-1}$ is omitted.

although again these estimates are not very precise. Results for Samples A and B are tabulated in Table 2 and are broadly similar.

Figure 3 plots the results of a similar exercise, this time including indicators for each of the 24 months prior to and each of the 60 months following the beginning of water improvements, i.e. results from estimating

$$y_{vt} = \sum_{s=-24}^{+59} \beta_s \tau_{vs} + \gamma V_v + \delta T_t + \varepsilon_{vt} \tag{5}$$

The results are similar to those in Figure 2, in that we do not see any clear pre-intervention trend of improvement (although there is some evidence of relatively poor outcomes more than a year before the water quality improves) and a discernable, if not statistically significant month-by-month, improvement after the water turns on. Furthermore, this improvement appears to persist with similar magnitude, although the estimates become ever noisier as time goes on. These regression results suggest that RHEP did have a discernable impact on health outcomes, a finding which we explore further in the remainder of this section.

To obtain a more precise estimate of overall program impact, we turn to the pre- versus post-intervention measures obtained from estimation of equations (2) and (3). Again, we run regressions using treated cases of diarrhea and malaria as the independent variables and over Samples A, B and C as defined in section I.B. Table 3 reports the results. The upper panel, labeled Model 1, reports estimates of α from equation (2) above. With respect to diarrhea, we find statistically significant and practically meaningful reductions across all three specifications, with cases reduced from 0.356 (from a pre-intervention mean of 1.11) to 0.59 (from a base of 1.18) per month depending on the sample and definition of the dependent variable (cases checked versus cases treated). The evidence that RHEP reduced cases of malaria is also strong, with all estimates negative, significant and large relative to the pre-intervention mean.

The lower panel, labeled Model 2, reports estimates of α and β from equation (3) above, where β represents the coefficient on a linear trend in τ_{vt}. The results are quite robust to the addition of this linear trend, as can be seen from the similarity of the point estimates and standard errors of the "level change" variable ($POST_{vt}$ in equation (3) above) to our previous estimates of the "mean shift" variable in model 1 variable ($POST_{vt}$ in equation (2)). Furthermore, the estimates of the "linear trend" variable (τ_{vt} in equation (3)) are small in all 6 models and statistically significant in only 2 of 6 models, suggesting that there is not much of a trend around the arrival of RHEP. This finding was foreshadowed by the event-study figures.

We conduct several robustness checks to confirm that our results are not simply the product of a certain specification or choice of variable definition. Tables 4 and 5 report results of our robustness checks for the primary diseases of interest, diarrhea and malaria. Table 4 examines the results as we alter our specification of equation (2). The original estimates of equation (2) are reproduced on the top row for comparison with the estimates in the robustness checks. We alter the specification in the following ways: different weighting schemes (unweighted, weight by mean village population); include additional fixed effects (district-by-year); different samples (include villages with unknown τ, include only villages with 3 years of post-intervention data or more, include only villages with less than 3 years of post-intervention data, include only "early" program villages, include only "late" program villages); different methods of data-cleaning (replace missing outcomes with zeros, replace zero outcomes with missing, perform no cleaning on the data). While there are too many specifications to discuss individually, the clear pattern is that results of statistically and practically significant health improvements persist. This is visible in Figure 4, which plots the 95% confidence interval from each specification using Sample B: the centers of all confidence intervals are less than zero and most of the intervals are entirely contained in the negative domain.

Table 5 contains the results of another series of robustness checks, in which we compare results when we use different definitions of our key independent variable, the month in which the water supply is activated. We use three different definitions: the month in which the water tank turns on (as in our main estimates reported in Table 3); the month in which the tank is completed; the month is which 90% of the latrines are complete. As before, we use Samples A, B and C as defined in Section I.B. In addition, we cut the sample in the two new ways. Sample I considers all villages for which τ is defined according to the measure given in the row heading, with the restriction that no other indicator of the program beginning come before this indicator. That is, in the row labeled "tank complete," a village is included if and only if (a) the month in which the tank is completed is known (and therefore τ is defined) and (b) no other indicator of water improvement come before the month in which the tank is complete. Sample II adds the additional restriction that this measure be within 3 months of the water tank completion date recorded in the ex-post inventory provided to us by Gram Vikas in 2007. As in Table 4, the results are reasonably robust. Again, this is perhaps best communicated in graphical form, as in Figure 5, which plots the 95% confidence interval from each specification using Sample B. There is slightly more variation here than in Figure 4, since a few confidence intervals are centered above zero, but the overall pattern again is that, in most cases, most or all of the confidence interval is below zero, indicating a program benefit.

Table 6 examines the estimated impact of RHEP using different outcome variables. First, we use more detailed diarrhea measures that are only recorded from 1997-1999. In particular, the Monthly Progress Reports for these years report cases of diarrhea among children (ages 0-18), number of days children suffer from diarrhea and the number of workdays lost by adults to diarrhea. The evidence for a policy benefit to children is weaker: only one estimate of six is statistically significant and two point estimates have a positive sign. This is surprising, since the health literature emphasizes the benefits of sanitation for children. This failure to find the expected result could be due to the limited sample size or the rarity of the outcome in our data (villages report an average of 0.04-0.05 child cases of diarrhea per month). We also look at other recorded health outcomes, such as malnutrition and fever, as well as total population and number of children (which may proxy for mortality or in-migration toward environmental amenities) and school attendance. The results are not particularly strong for any of these outcomes, although the point estimates are generally indicative of improvement. It is perhaps not surprising that these results are not as strong, since these are outcomes that may be only weakly linked to water quality (e.g. fever), take a relatively long time to respond to water quality improvements (e.g. malnutrition) or are constrained by other institutional factors (e.g. school attendance may be limited by teacher absence). This may be a result of the limited sample size.

In Table 7, we explore the question of whether village characteristics are correlated with improvements within our sample. First, we calculate the predicted improvement in outcomes village-by-village by interacting the $POST_{vt}$ dummy with indicator variables for each village (limiting the sample to villages with at least six observations before and six observations after the water improvement), and including these regressors in our basic estimation equation (5). That is, we estimate

$$y_{vt} = \sum_{v}\left\{\alpha_v\left(POST_t \times V_v\right)\right\}_t + \sum_{v}\gamma_v V_v + \sum_{t}\delta_t T_t + \alpha POST + \varepsilon_{vt} \tag{6}$$

from which we $\hat{\alpha}_v$, the estimated impact of the program in village v. Ninety-five percent confidence intervals for these estimates are plotted in increasing order in Figure 6.

Most villages are estimated to have benefitted from the program, although improvement is not universal. We then regress these village-level improvements on village characteristics, including the share of the village in the scheduled tribe or scheduled caste category, the pre-intervention prevalence of the condition and the calendar month in which the program started in that village. No strong results emerge, except for that the program appears to have been more successful at reducing malaria in villages where pre-program malaria prevalence was high.

Finally, in Table 8 we report formal tests of the persistence of the program effect with regressions such as that in equation (4) above. Panel (A) presents regressions in which we estimate the effect separately for each of the first five years after the program is initiated, i.e.

$$y_{vt} = \sum_{j=1}^{5}\left(\alpha_j \times 1\left\{\tau_{vt} \in Y_j\right\}\right) + \left(\alpha_6 \times 1\left\{\tau_{vt} \geq 36\right\}\right) + \gamma V_v + \delta T_t + \varepsilon_{vt} \tag{7}$$

where Y_j indicates that month t is in the j^{th} year after the intervention for village v. That is, $\hat{\alpha}_1$ is the estimated impact in the first year after service begins (months 0-11), $\hat{\alpha}_2$ is the estimated impact in the first year after service begins (month 12-23), and so on. The results for diarrhea are quite striking: the effect persists in each of the subsequent five years and even afterwards, although the effect more than five years out is no longer statistically significant. The impact on malaria, though, does seem to fade over time. In Panel (B), we split the post-intervention period into just two sub-periods, the 36 months immediately after the intervention and all months thereafter (this is exactly equation (4) above). Again, the effects appear to be persistent, even increasing (i.e. a larger negative point estimate) in the case of malaria. Formally, we find no evidence of decreasing effects, based on the F-statistics and p-values for tests of the equality of the estimated $\widehat{\alpha}_1$ and $\widehat{\alpha}_2$ in panel (B).

IV. Caveats, Cost Assessment and Conclusion

IV.A: Caveats

There are four main caveats to keep in mind when considering these results. First, as mentioned in sections II.A and II.B, we have reason to believe that our data under-report the true prevalence of the diseases in question, since it is likely that only the most severe cases are reported. Therefore, we do not know how the program affects the *overall* burden of disease in the program villages. It could be that the program's effects are concentrated largely on the most severe cases, so our results capture the majority of the program's benefits. Alternatively, the program affects both severe cases (those recorded in our data) and less-severe cases equally, in which case we could use our estimates to predict that the overall burden of these diseases is reduced by the same proportion, i.e. 33-50 percent. Even more optimistically, it could be that the program is *least* effective at reducing the number of severe cases, in which case even predicting an equivalent reduction in less-severe cases would *understate* the benefits of the program. At present, we do not have reason to favor any one of these three scenarios.

Second, our estimates are from a program that is provided to all households in a given village. As we have discussed above, we view this as a methodological advantage, since we capture all intra-village externalities. However, our results may be less informative about the impacts of a program that does not reach all residents of target localities.

Third, the program we evaluate has a number of components. In particular, RHEP involves provision of a communal water tank, piped water to the home, and household toilets and bathing facilities. Our evaluation is of this package of services and therefore does not provide information on how effective individual components might be.

Fourth, while we have argued that our event study methodology in likely to produce unbiased estimates of the program's effects in these villages, it is important to note that this is a selected sample of villages. GV partners with villages it believes capable of building the institutional capacity to complete the project and to sustain the facilities after construction is finished. Furthermore, these villages are sufficiently internally cooperative that they are able to produce 100% participation, a strictly necessary condition for GV's involvement. These are important and perhaps unusual characteristics of a village, and it should be kept in mind that a program technically similar to RHEP but targeted at villages lacking these characteristics may be less successful.

IV.B: Costs

In order to consider how best to direct scarce development resources, it is important to estimate the cost of the program. This necessarily involves some guesswork and imputation, in particular in considering the value of villagers' labor and the basic materials villagers provide, but should yield a ballpark estimate that, at a minimum, could be used to compare the cost of a program like RHEP with other potential interventions. Table 9 presents a rough breakdown of costs per household in a "typical" village of 50 households among construction of toilets and bath houses, construction of water supply and project development costs (primarily GV staff time spent in organizing and capacity-building) and between villagers' contributions and inputs purchased with external funds.[17] The value of villagers' labor is assessed at US$1 per day, the local agricultural day labor wage. This may be an overstatement since most work takes place outside the planting and harvest seasons, i.e. in times in which the opportunity cost of villagers' time is perhaps lower. The approximate construction cost per household is $440 with annual maintenance costs of roughly $11. If we assume an interest rate of 10% per year and given the intended useful 20-year life of the tank and sanitary facilities, we can approximate the per-year cost at $58 per household per year. For comparison, a yearly supply of dilute sodium hypochlorite disinfectant used to treat water in many developing countries costs roughly $7 per year, and has been shown to reduce diarrhea in young Kenyan children by up to 40% (Kremer et. al. 2008). Of course, a full cost-benefit assessment of RHEP would require a more precise monetization of benefits, which we cannot provide, for

[17] For more details on inputs and costs, see Gram Vikas 2001.

several reasons: first, our measures are of just a subset of the health conditions likely impacted by RHEP; second, there is no agreed-upon, stable mapping from health improvements to monetary values; third, we have no data on other benefits of the program, such as water collection time saved or the amenity value of private toilet and bathing facilities.

IV.C: Conclusion

Our study makes two key contributions, one methodological and one substantive. We show that it is possible to obtain scientific evidence on the impact of a water and sanitation program even in the absence of a randomized trial. To accomplish this, we combine detailed understanding of the implementation of the program with ex-ante understanding of epidemiological mechanisms to create a quasi-experimental research design.

On substantive grounds, we show that it is indeed possible for rural piped water to reduce the prevalence of water-borne disease. This is perhaps contrary to the beliefs of the community of clean water researchers (Clasen and Haller 2008). This suggests that, in addition to continuing research on the benefits of household and point-of-use interventions, there is a place for continued research into the provision of rural piped water infrastructure.

References

Nava Ashraf, James Berry, and Jesse M. Shapiro. Can higher prices stimulate product use? Evidence from a field experiment in Zambia. Working Paper 13247, National Bureau of Economic Research, July 2007. URL http://www.nber.org/papers/w13247.

Abhijit Banerjee, Angus Deaton, and Esther Duflo. Wealth, health, and health services in rural Rajasthan. *American Economic Review Papers and Proceedings*, 94(2):326–330, May 2004. URL http://www.ingentaconnect.com/content/aea/aer/2004/00000094/00000002/art00059.

Marianne Bertrand, Esther Duflo, and Sendhil Mullainathan. How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics*, 119(1):249–275, 2004. doi: 10.1162/003355304772839588. URL http://www.mitpressjournals.org/doi/abs/10.1162/003355304772839588.

Nazmul Chaudhury, Jeffrey Hammer, Michael Kremer, Karthik Muralidharan, and F. Halsey Rogers. Missing in action: Teacher and health worker absence in developing countries. *Journal of Economic Perspectives*, 20(1):91–116, 2006. doi: 10.1257/089533006776526058.

Thomas F. Clasen and Laurence Haller. Water quality interventions to prevent diarrhoea: Cost and cost-effectiveness. Technical report, World Health Organization, February 2008.

David Cutler and Grant Miller. Water, water, everywhere: Municipal finance and water supply in American cities. Working Paper 11096, National Bureau of Economic Research, January 2005. URL http://www.nber.org/papers/w11096.

Esther Duflo, Rema Hanna, and Stephen Ryan. Monitoring works: Getting teachers to come to school. URL http://www.cepr.org/pubs/dps/DP6682.asp. CEPR Discussion Paper 6682, January 2008.

Sebastian Galiani, Paul Gertler, and Ernesto Schargrodsky. Water for life: the impact of the privatization of water services on child mortality. *Journal of Political Economy*, 113:83–120, 2005.

Annual Report 2004–2005. Gram Vikas, 2005. URL http://www.gramvikas.org/PDF/Annual_Reports/Annual%20Report%202004-2005.pdf.

Rural Health and Environment Programme. Gram Vikas, May 2001. URL http://www.gramvikas.org/PDF/unpublished/Rural_Health_and_Environment_Programme.pdf. Technical Report.

Rural Health and Environment Programme in Orissa. Gram Vikas, 2004. URL http://www.gramvikas.org/PDF/published/Rural_Health_and_Environment_Programme.pdf. Technical Report.

T. Kloek. OLS estimation in a model where a microvariable is explained by aggregates and contemporaneous disturbances are equicorrelated. *Econometrica*, 49(1):205–207, January 1981. ISSN 00129682. URL http://www.jstor.org/stable/1911134.

Michael Kremer, Edward Miguel, Clair Null, and Alix Peterson Zwane. Trickle down: Diffusion of chlorine for drinking water treatment in Kenya. University of California Berkeley Mimeo, February 2008.

Stephen P. Luby, Mubina Agboatwalla, John Painter, Arshad Altaf, Ward L. Billhimer, and Robert M. Hoekstra. Effect of intensive handwashing promotion on childhood diarrhea in high-risk communities in Pakistan. *Journal of the American Medical Association*, 291:2547–2554, 2004.

Edward Miguel and Michael Kremer. Worms: Identifying impacts on education and health in the presence of treatment externalities. *Econometrica*, 72(1):159–217, January 2004. ISSN 00129682. URL `http://www.jstor.org/stable/3598853`.

Brent R. Moulton. An illustration of a pitfall in estimating the effects of aggregate variables on micro units. *The Review of Economics and Statistics*, 72(2):334–338, May 1990. ISSN 00346535. URL `http://www.jstor.org/stable/2109724`.

Millenium Development Goals Report. United Nations, 2007. URL `http://www.un.org/millenniumgoals/docs/UNSD_MDG_Report_2007e.pdf`.

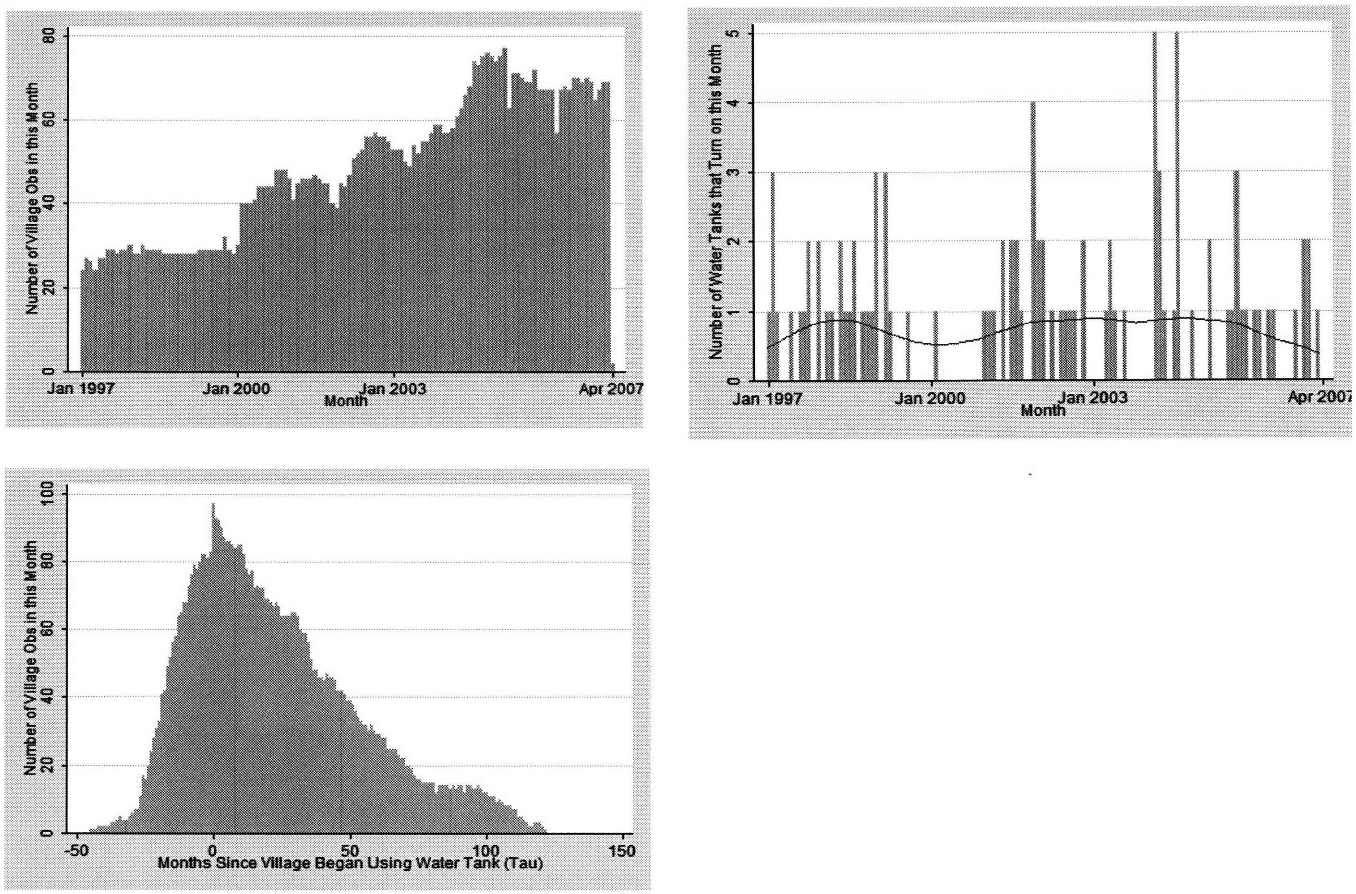
Tara Watson. Public health investments and the infant mortality gap: Evidence from federal sanitation interventions on U.S. Indian reservations. *Journal of Public Economics*, 90(8-9):1537–1560, September 2006. URL `http://dx.doi.org/10.1016/j.jpubeco.2005.10.002`.

Water, sanitation and hygiene links to health. World Health Organization, November 2004. URL `http://www.who.int/water_sanitation_health/publications/facts2004/en/index.html`.

World Health Report. World Health Organization, 2005. URL `http://www.who.int/entity/whr/2005/whr2005_en.pdf`.

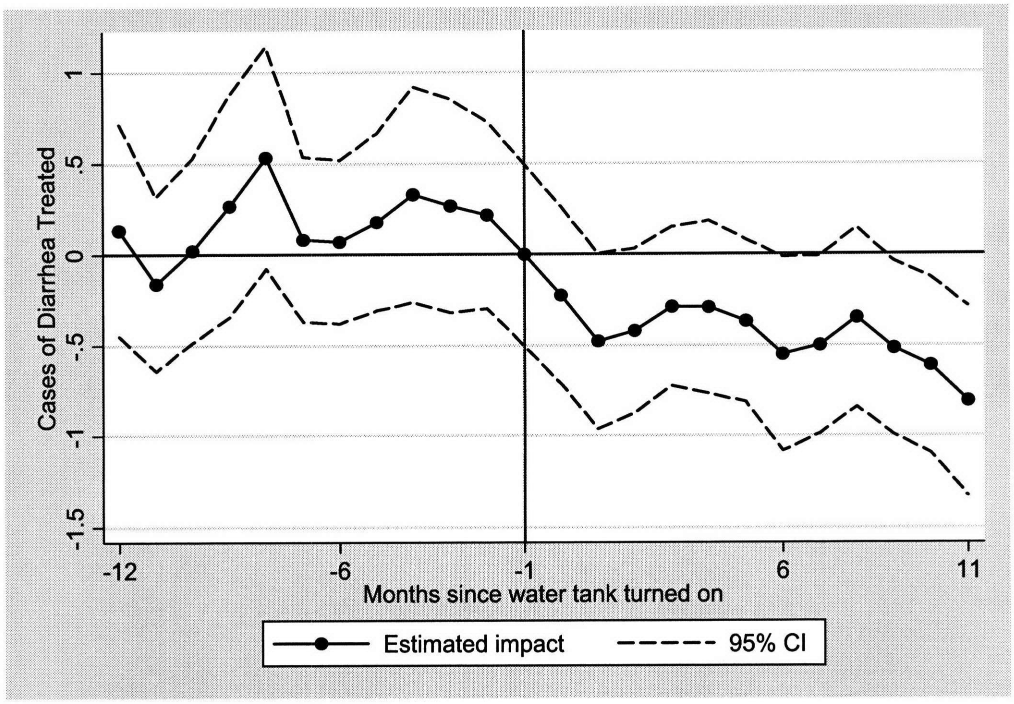
Alix Peterson Zwane and Michael Kremer. What works in fighting diarrheal diseases in developing countries? A critical review. *World Bank Research Observer*, 22(1):1–24, 2007. doi: 10.1093/wbro/lkm002.

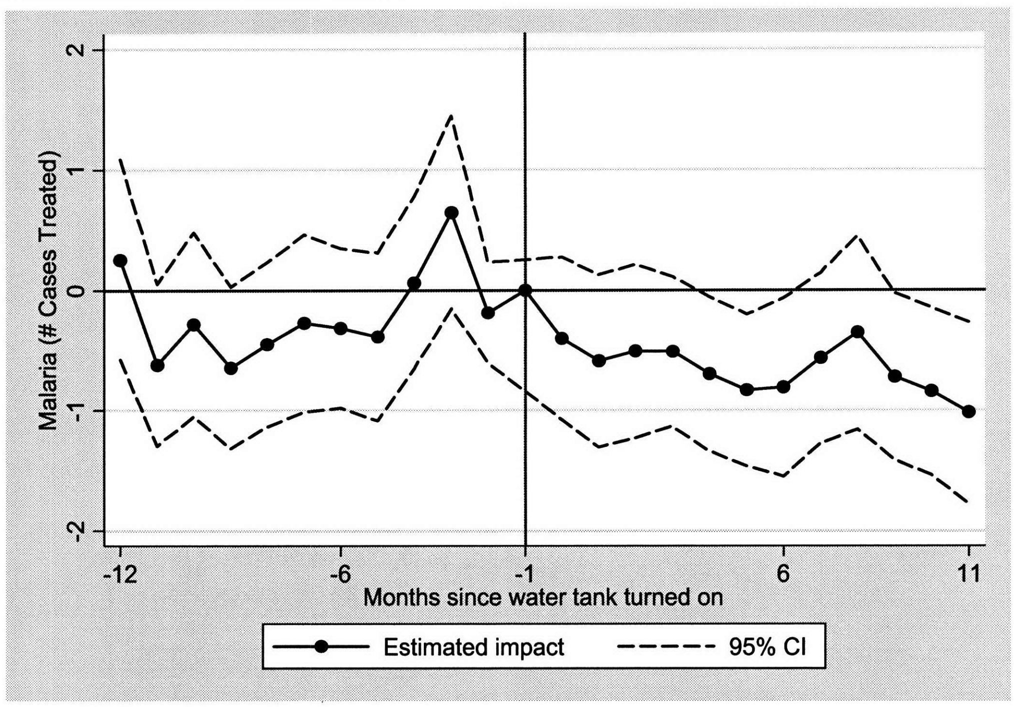
Figure 1. The Timing of Observations



Notes: all graphs show all observations with known τ. The upper right-hand graph overlays a kernel density plot which uses an Epinechnikov kernel with a half-width of 10 months.

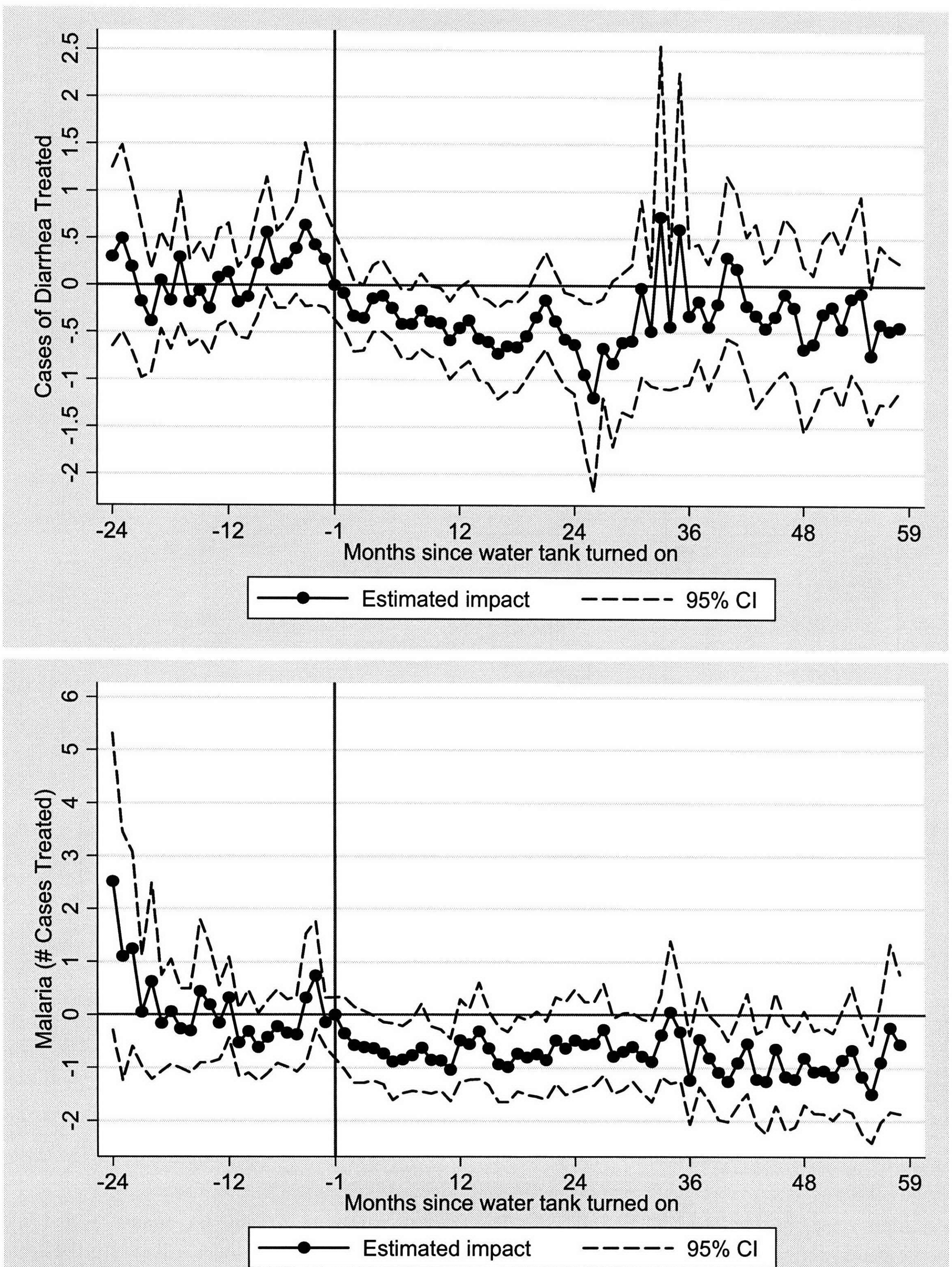
Figure 2: Event Analysis Within Twelve Months of Beginning Water Improvement





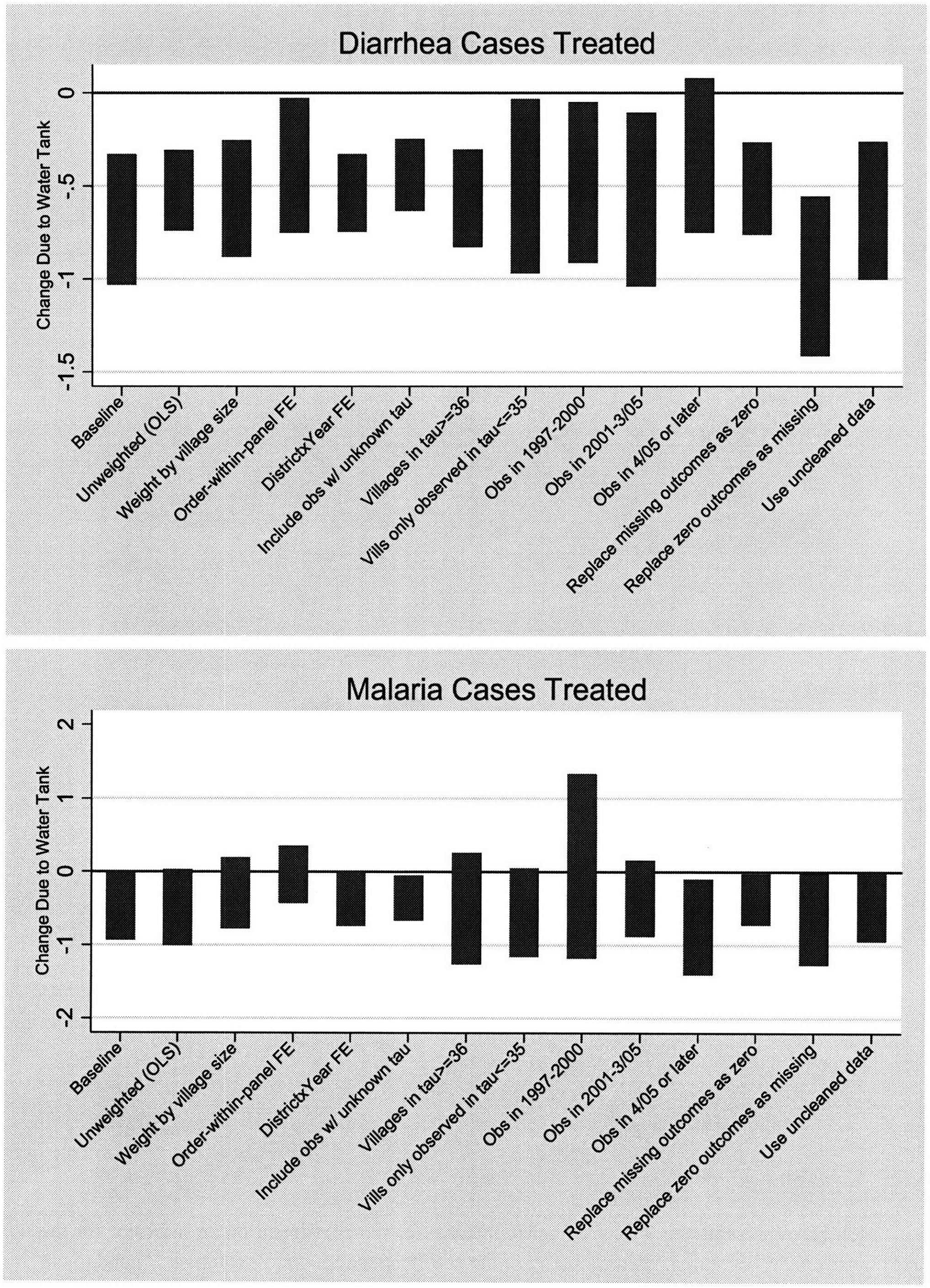
Notes: these figures use observations with τ in [-12,11]. The graphs correspond to Table 6 columns (3) and (6). The points are parameter estimates for each value of τ in a GLS regression which includes village fixed effects and year fixed effects, with weights equal to the inverse of the number of months in which a village has non-missing outcomes.

Figure 3: Event Analysis for up to Five Years Beyond the Beginning of Water Improvement



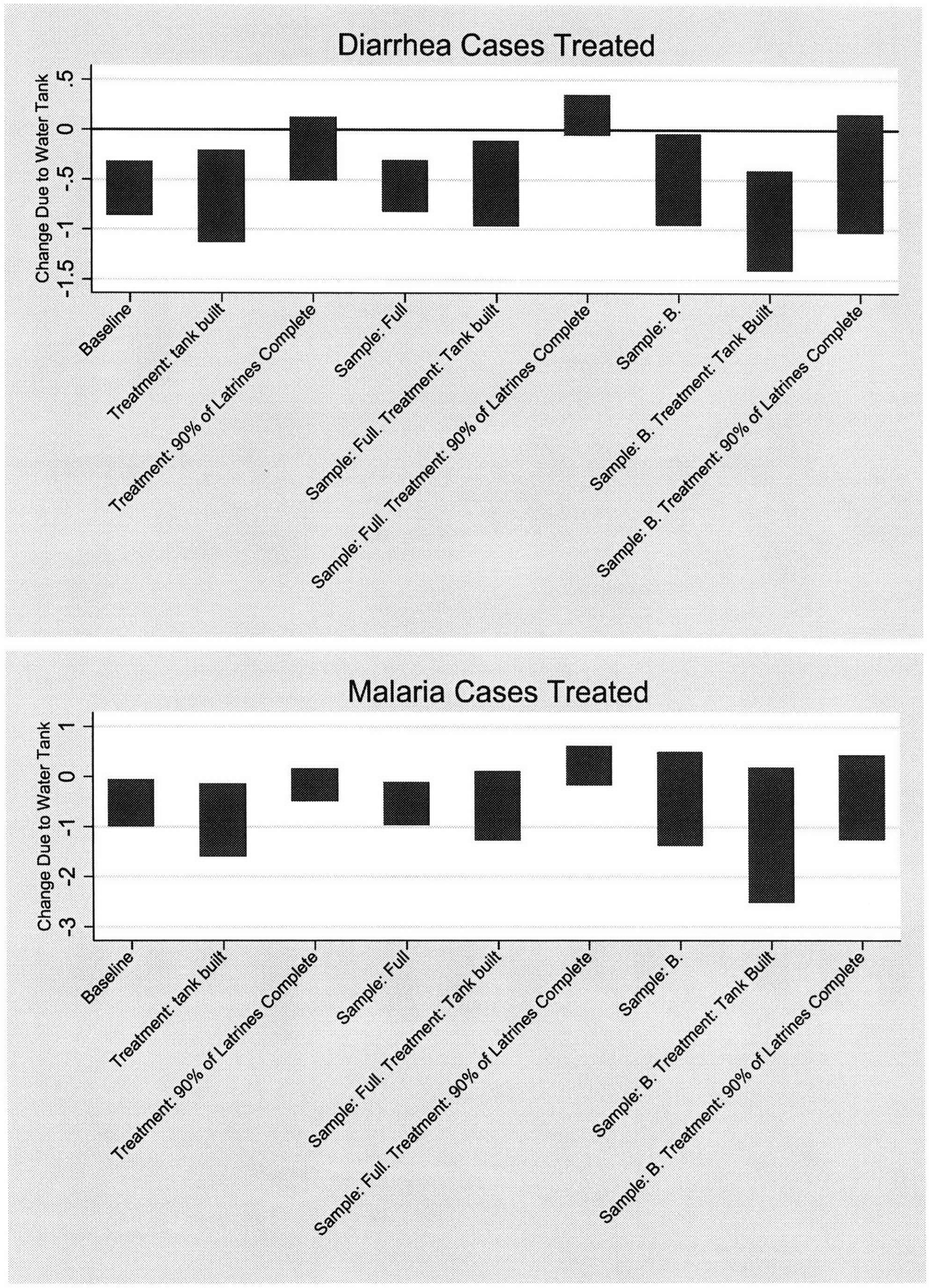
Notes: these figures use observations with τ in [-24,59]. The graphs correspond to Table 6 columns (2) and (5). The points are parameter estimates for each value of τ in a GLS regression which includes village fixed effects and year fixed effects, with weights equal to the inverse of the number of months in which a village has non-missing outcomes.

Figure 4: Distribution of Estimates of Program Impacts, by Estimation Method



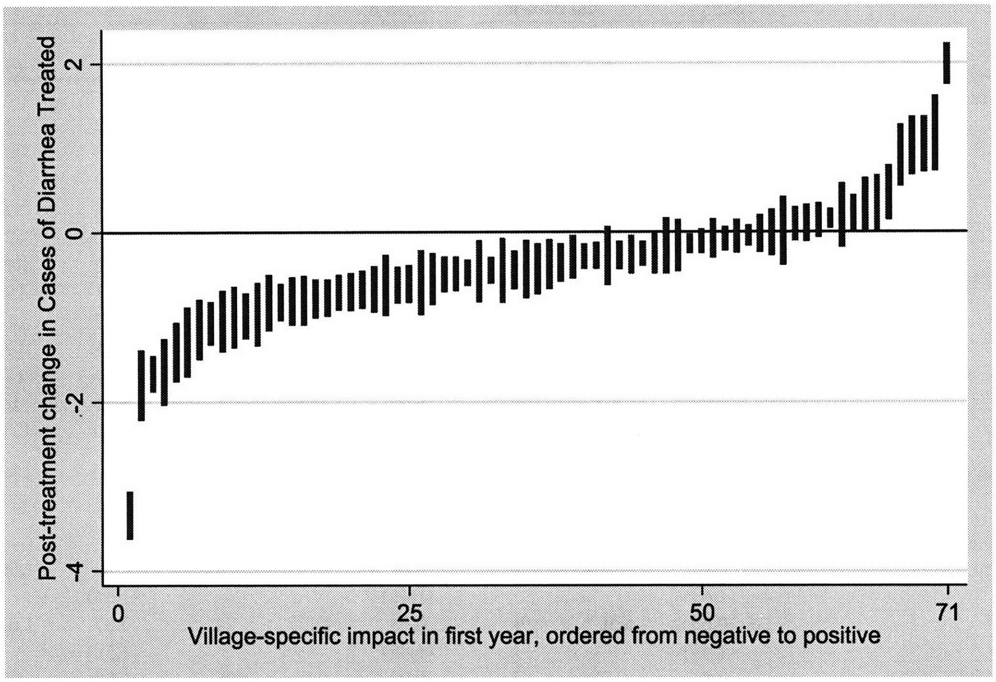
Notes: each bar represents the 95% confidence interval for the coefficient on an indicator for the water improvement beginning shown in columns (2) and (5) of Table 3, corresponding to Sample A (observations with τ in [-24,59]). In the baseline case, the coefficient is obtained from a GLS regression which includes year and time fixed effects and which has weights equal to the inverse of the number of observations. Each regression uses a different methodology. Additional information in the notes accompanying Table 3.

Figure 5: Distribution of Estimated Program Impact, by Definition of Program Start Month

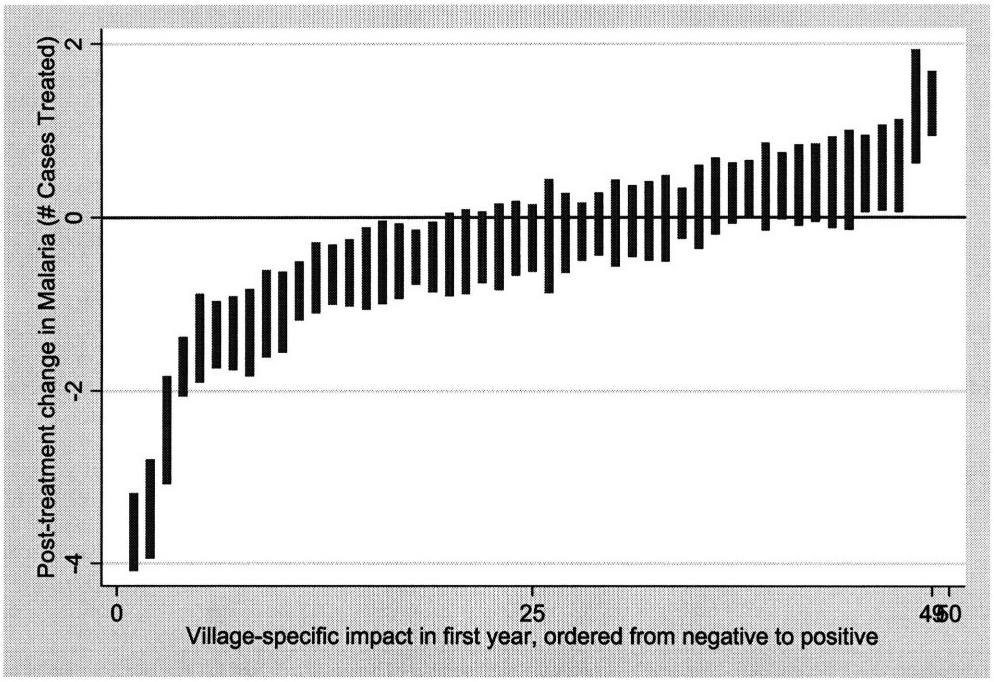


Notes: each bar represents the 95% confidence interval for the coefficient on an indicator for the water improvement beginning, corresponding to the coefficients reported in columns (2) and (5) of Table 4. The coefficient is obtained from a GLS regression which includes year and time fixed effects and which has weights equal to the inverse of the number of observations. Each regression uses a different definition of when the water improvement began. The baseline defines the treatment variable according to when the tank began operation. The baseline regression uses observations with τ in [-24,59] from the sample of villages which have at least 6 non-missing observations before and 6 non-missing observations after water improvement begins. The baseline sample also uses the restriction that when the monthly data report other measures of when water treatment begins, those measures must be no later than the measure used in this estimate. Additional samples and restrictions explained in the notes accompanying Table 4.

Figure 6: Distribution of Health Impacts by Village



Share of village-specific coefficients that are negative: 79%; share neg. and stat. signif.: 65%



Share of village-specific coefficients that are negative: 67%; share neg. and stat. signif.: 39%

Notes: Values below the x-axis represent cases of diarrhea or malaria prevented by the tank. Each vertical bar represents the 95% confidence interval for the coefficient on the interaction of a village fixed effect with a τ in [0,11] indicator in a GLS regression with weights equal to the inverse of the number of village-specific observations on the outcome of interest. The village-specific treatment effects are estimated for all villages which have at least six $\tau<0$ observations and six $\tau\geq 0$ observations. Regressions also include village fixed effects, year fixed effects, and one indicator for $\tau\geq 12$. Huber-White covariance matrix clustered within villages.

Table 1. Description of the RHEP Village Data

Sample	Sample A: Obs with known τ			Sample B: τ ∈[-24,59]			Sample C: τ ∈[-12,11]		
	Mean	N	Villages	Mean	N	Villages	Mean	N	Villages
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Share of observations in each year:									
Year 1997	0.05	329	30	0.06	249	23	0.07	125	16
Year 1998	0.06	342	30	0.06	273	23	0.14	242	23
Year 1999	0.06	347	33	0.07	286	25	0.09	157	18
Year 2000	0.09	511	50	0.09	413	41	0.04	76	16
Year 2001	0.09	532	50	0.10	431	40	0.10	173	16
Year 2002	0.10	628	59	0.12	533	51	0.12	212	24
Year 2003	0.11	649	62	0.11	490	50	0.11	194	24
Year 2004	0.13	803	78	0.13	547	51	0.12	213	26
Year 2005	0.14	845	78	0.13	556	50	0.11	193	26
Year 2006	0.13	806	72	0.11	482	47	0.07	115	14
Year 2007	0.03	207	69	0.02	105	36	0.01	10	4
Total	1.00	5999	97	1.00	4365	75	1.00	1710	75
Months since began using water tank (τ)	23.22	5999	97	13.92	4365	75	-0.23	1710	75
	(31.73)			(20.88)			(6.81)		
Monthly heath and education outcomes:									
Diarrhea cases checked	0.63	4,694	96	0.67	3,511	74	0.85	1,505	73
	(1.43)			(1.45)			(1.54)		
Diarrhea cases treated	0.63	5,507	96	0.68	4,077	74	0.87	1,628	73
	(1.43)			(1.45)			(1.59)		
Days with diarrhea, children 0-18	0.81	929	35	0.90	740	25	1.00	473	24
	(1.43)			(1.52)			(1.55)		
Cases of diarrhea, children 0-18	0.05	609	35	f	468	25	0.04	283	24
	(0.26)			(0.22)			(0.21)		
Workdays lost to diarrhea	3.16	1,018	35	3.55	808	25	3.87	524	24
	(6.80)			(7.24)			(7.05)		
Malaria cases checked	0.88	3,797	94	0.93	2,792	73	0.86	1,042	52
	(2.06)			(2.05)			(1.73)		
Malaria cases treated	0.95	4,635	95	0.99	3,360	75	0.93	1,145	56
	(2.09)			(2.12)			(1.95)		
Fever cases checked	2.33	3,692	94	2.43	2,732	75	2.36	1,013	56
	(3.61)			(3.80)			(3.46)		
Fever cases treated	2.66	4,634	95	2.66	3,359	75	2.56	1,145	56
	(3.66)			(3.80)			(3.68)		
Borderline or severely malnourished[1]	0.46	1,120	70	0.47	917	56	0.44	441	44
	(0.20)			(0.20)			(0.17)		
Severely malnourished[1]	0.12	1,187	71	0.12	971	56	0.11	457	44
	(0.14)			(0.15)			(0.10)		
Households	85.19	3,830	93	87.68	2,815	73	91.29	1,037	52
	(54.71)			(53.59)			(53.10)		
People aged over 18	298.64	3,837	93	304.36	2,863	75	311.35	1,033	56
	(199.92)			(195.31)			(186.17)		
Children	178.49	3,114	89	180.01	2,400	74	179.96	889	52
	(115.58)			(110.33)			(105.17)		
Days of school attended	6.57	1,776	83	6.85	1,358	69	6.56	519	48
	(3.83)			(4.01)			(4.11)		

Notes: Each observation represents one village-month-year, and N denotes the total number of observations used to calculate a particular mean. Standard deviations for non-binary variables appear in parentheses. Sample A (columns (1), (2), and (3)) includes all villages where the the data report the month when the water tank turned on (i.e. τ is known) and where this date is no earlier than other measures of when the policy began (90% of latrines built, or water tank built) from the monthly data. Sample B (columns (4), (5), and (6)) further restrict the sample to observations with τ ∈[-24,59] in villages which have at least 6 observations before and 6 observations after the water turns on. Sample C (columns (7), (8), and (9)) further restrict the sample to observations with τ ∈[-12,11].

[1] Malnutrition data are the share of tested children that are malnourished. All tested children are aged 0-5 years. Borderline malnutrition denotes grade I or II malnutrition by WHO standard, while malnourished indicates grade III or IV malntutrition by WHO standard.

Table 2. Health Outcomes Relative to Month when Water Improvement Begins

Outcome	Diarrhea cases treated			Malaria cases treated		
Sample	All with known τ	τ in [-24,59]	τ in [-12,11]	All with known τ	τ in [-24,59]	τ in [-12,11]
Months since water tank operational (τ)	(1)	(2)	(3)	(4)	(5)	(6)
τ=-12	0.454	0.322	0.134	0.238	0.308	0.253
	(0.230)*	(0.186)*	(0.297)	(0.330)	(0.364)	(0.425)
τ=-11	-0.01	0.003	-0.162	-0.565	-0.54	-0.623
	(0.140)	(0.135)	(0.244)	(0.182)***	(0.195)***	(0.345)*
τ=-10	-0.017	0.057	0.024	-0.267	-0.32	-0.284
	(0.171)	(0.182)	(0.258)	(0.312)	(0.339)	(0.392)
τ=-9	0.24	0.423	0.267	-0.618	-0.615	-0.644
	(0.203)	(0.230)*	(0.310)	(0.213)***	(0.248)**	(0.344)*
τ=-8	0.319	0.759	0.533	-0.519	-0.393	-0.452
	(0.302)	(0.275)***	(0.310)*	(0.232)**	(0.217)*	(0.349)
τ=-7	0.112	0.373	0.084	-0.04	-0.168	-0.275
	(0.238)	(0.213)*	(0.230)	(0.311)	(0.274)	(0.378)
τ=-6	0.5	0.429	0.069	-0.291	-0.268	-0.316
	(0.224)**	(0.203)**	(0.229)	(0.258)	(0.269)	(0.339)
τ=-5	0.59	0.606	0.178	-0.19	-0.301	-0.388
	(0.254)**	(0.253)**	(0.248)	(0.264)	(0.266)	(0.357)
τ=-4	0.405	0.861	0.329	0.043	0.395	0.064
	(0.354)	(0.407)**	(0.302)	(0.507)	(0.649)	(0.365)
τ=-3	0.538	0.659	0.267	0.542	0.858	0.646
	(0.243)**	(0.300)**	(0.299)	(0.395)	(0.479)*	-(0.409)
τ=-2	0.386	0.517	0.216	-0.203	-0.012	-0.186
	(0.219)*	(0.237)**	-(0.261)	(0.253)	(0.262)	-(0.215)
τ=-1	-	-	-	-	-	-
τ=0	-0.054	0.167	-0.224	-0.456	-0.187	-0.401
	(0.150)	(0.163)	(0.248)	(0.284)	(0.281)	(0.346)
τ=1	-0.252	-0.068	-0.478	-0.712	-0.386	-0.589
	(0.150)*	(0.171)	(0.247)*	(0.282)**	(0.263)	(0.366)
τ=2	-0.267	-0.09	-0.419	-0.791	-0.411	-0.507
	(0.126)**	(0.156)	(0.231)*	(0.281)***	(0.269)	(0.369)
τ=3	-0.087	0.112	-0.288	-0.857	-0.426	-0.511
	(0.146)	(0.148)	(0.223)	(0.263)***	(0.193)**	-(0.318)
τ=4	-0.102	0.148	-0.291	-0.772	-0.503	-0.7
	(0.150)	(0.187)	(0.242)	(0.258)***	(0.224)**	(0.327)**
τ=5	-0.142	0.024	-0.366	-0.692	-0.628	-0.833
	(0.106)	(0.124)	(0.228)	(0.206)***	(0.221)***	(0.322)**
τ=6	-0.233	-0.125	-0.548	-0.774	-0.581	-0.808
	(0.109)**	(0.110)	(0.272)**	(0.202)***	(0.198)***	(0.378)**
τ=7	-0.265	-0.115	-0.498	-0.727	-0.49	-0.565
	(0.125)**	(0.127)	(0.249)**	(0.243)***	(0.238)**	(0.362)
τ=8	-0.082	0.025	-0.347	-0.192	-0.323	-0.355
	(0.096)	(0.129)	(0.251)	(0.280)	(0.305)	(0.409)
τ=9	-0.102	-0.064	-0.514	0.404	-0.518	-0.721
	(0.122)	(0.129)	(0.244)**	(0.632)	(0.182)***	(0.354)**
τ=10	-0.074	-0.074	-0.609	-0.4	-0.509	-0.844
	(0.135)	(0.153)	(0.247)**	(0.215)*	(0.192)***	(0.355)**
τ=11	-0.08	-0.248	-0.807	-0.143	-0.66	-1.022
	(0.150)	(0.104)**	(0.266)***	(0.481)	(0.160)***	(0.385)**
Observations	5507	4077	1628	4635	3360	1145
Villages	96	74	73	95	75	56
R-squared	0.35	0.38	0.46	0.28	0.29	0.45
Village FE	YES	YES	YES	YES	YES	YES
Month x year FE	YES	YES	YES	YES	YES	YES

Notes: Each observation is one village-month-year. Table presents GLS regressions with weight of each observation in village i equal to 1/(total number of observations of village i). Includes only villages where the date when water tank turned on is known. Huber-White standard errors in parentheses are clustered within villages. τ=-1 is omitted as the reference category. All regressions include a constant.

Table 3. The Effect of Water Improvement on Health

Outcome	Diarrhea Cases Treated			Malaria Cases Treated		
Sample	A: All with known τ	B: τ in [-24,59]	C: τ in [-12,11]	A: All with known τ	B: τ in [-24,59]	C: τ in [-12,11]
	(1)	(2)	(3)	(4)	(5)	(6)
Model 1:						
Mean Shift	-0.59	-0.587	-0.453	-0.473	-0.528	-0.479
	(0.125)***	(0.139)***	(0.178)**	(0.191)**	(0.241)**	(0.280)*
Observations	5507	4077	1628	4635	3360	1145
Villages	96	74	73	95	75	56
R-squared	0.35	0.37	0.45	0.27	0.29	0.43
Mean outcome τ<0	1.08	1.09	1.18	1.33	1.28	1.15
Model 2:						
Level Change	-0.59	-0.587	-0.453	-0.473	-0.528	-0.479
	(0.125)***	(0.139)***	(0.178)**	(0.191)**	(0.241)**	(0.280)*
Linear Trend	0.007	-0.001	-0.014	-0.011	-0.008	-0.014
	(0.002)***	-0.004	-0.012	(0.003)***	-0.005	-0.018
Observations	5507	4077	1628	4635	3360	1145
Villages	96	74	73	95	75	56
R-squared	0.35	0.37	0.45	0.27	0.29	0.43
Mean outcome τ<0	1.08	1.09	1.18	1.33	1.28	1.15
Village FE	YES	YES	YES	YES	YES	YES
Month x yr FE	YES	YES	YES	YES	YES	YES

Note: Each observation is one village-month-year. Table presents GLS regressions with weight of each observation in village i equal to 1/(total number of observations of village i). Asterisks denote statistical significance at 90% (*), 95% (**), and 99% (***) confidence. Huber-White standard errors clustered within villages. All regressions include a constant.

Table 4. The Effect of Water Improvement on Diarrhea: Robustness to Different Specifications

Outcome	Diarrhea Cases Treated			Malaria Cases Treated		
Sample	A: All with known τ	B: τ in [-24,59]	C: τ in [-12,11]	A: All with known τ	B: τ in [-24,59]	C: τ in [-12,11]
	(1)	(2)	(3)	(4)	(5)	(6)
Original estimates	-0.59	-0.587	-0.453	-0.473	-0.528	-0.479
	(0.125)***	(0.139)***	(0.178)**	(0.191)**	(0.241)**	(0.280)*
Observations	5507	4077	1628	4635	3360	1145
Villages	96	74	73	95	75	56
	-0.757	-0.807	-0.286	-0.707	-0.646	-0.406
Include controls for # kids, #adults	(0.213)***	(0.240)***	(0.296)	(0.304)**	(0.335)*	(0.338)
Observations	2923	2301	853	3021	2369	879
Villages	85	71	50	88	74	52
Change Weighting:						
Unweighted (OLS)	-0.539	-0.524	-0.426	-0.453	-0.489	-0.479
	(0.107)***	(0.110)***	(0.178)**	(0.214)**	(0.265)*	(0.291)
Observations	5507	4077	1628	4635	3360	1145
Villages	96	74	73	95	75	56
Weight by village size[1]	-0.589	-0.567	-0.458	-0.244	-0.288	-0.174
	(0.139)***	(0.160)***	(0.185)**	(0.225)	(0.246)	(0.342)
Observations	5434	4041	1604	4620	3357	1142
Villages	92	73	72	93	74	55
Use Additional Fixed Effects:						
Include order-within-panel FE[2]	-0.423	-0.39	-0.397	-0.254	-0.04	-0.451
	(0.155)***	(0.184)**	(0.182)**	(0.231)	(0.200)	(0.270)
Observations	5507	4077	1628	4635	3360	1145
Villages	96	74	73	95	75	56
Include district x year FE	-0.616	-0.537	-0.441	-0.413	-0.363	-0.493
	(0.100)***	(0.107)***	(0.179)**	(0.174)**	(0.190)*	(0.277)*
Observations	5507	4077	1628	4635	3360	1145
Villages	96	74	73	95	75	56
Modify the Sample Used:						
Include obs with unknown τ[3]	-0.515	-0.44	-0.232	-0.489	-0.355	-0.386
	(0.115)***	(0.098)***	(0.091)**	(0.207)**	(0.156)**	(0.137)***
Observations	11230	11230	11230	10156	10156	10156
Villages	342	342	342	340	340	340
Villages observed through τ=36 or later	-0.609	-0.566	-0.386	-0.522	-0.499	-0.421
	(0.142)***	(0.133)***	(0.213)*	(0.333)	(0.386)	(0.401)
Observations	4334	3138	1105	3527	2465	647
Villages	58	49	49	58	49	32
Villages only observed through τ=35 or earlier	-0.444	-0.502	-0.428	-0.335	-0.554	-0.47
	(0.283)	(0.238)**	(0.179)**	(0.260)	(0.310)*	(0.368)
Observations	1173	939	523	1108	895	498
Villages	38	25	24	37	26	24
Obs dates: 1997-2000	-0.509	-0.482	-0.706	-14.801	0.075	-0.617
	(0.223)**	(0.220)**	(0.287)**	(0.112)***	(0.642)	(1.160)
Observations	1519	1211	599	507	409	72
Villages	52	41	35	50	41	16
Obs dates: 2001-3/05	-0.541	-0.573	-0.264	-0.474	-0.362	-0.503
	(0.209)**	(0.237)**	(0.270)	(0.226)**	(0.265)	(0.330)
Observations	2736	2085	825	2838	2151	855
Villages	90	70	46	93	73	48
Obs dates: 4/05 or later	-0.355	-0.337	-0.421	-0.383	-0.752	-0.685
	(0.175)**	(0.211)	(0.251)	(0.256)	(0.333)**	(0.607)
Observations	1252	781	204	1290	800	218
Villages	73	47	17	75	48	18

Table 4 (Continued)

Outcome	Diarrhea Cases Treated			Malaria Cases Treated		
Sample	A: All with known τ	B: τ in [-24,59]	C: τ in [-12,11]	A: All with known τ	B: τ in [-24,59]	C: τ in [-12,11]
	(1)	(2)	(3)	(4)	(5)	(6)
Modify Data Cleaning:						
Replace missing outcomes as zero	-0.48	-0.514	-0.416	-0.354	-0.373	-0.311
	(0.109)***	(0.127)***	(0.171)**	(0.144)**	(0.177)**	(0.180)*
Observations	5999	4365	1710	5999	4365	1710
Villages	97	75	75	97	75	75
Replace zero outcomes as missing	-0.975	-0.984	-0.974	-0.505	-0.64	-0.512
	(0.189)***	(0.218)***	(0.355)***	(0.289)*	(0.318)**	(0.597)
Observations	1492	1192	575	1609	1223	404
Villages	90	72	67	92	73	50
Use uncleaned data[4]	-0.677	-0.632	-0.466	-0.504	-0.471	-0.303
	(0.182)***	(0.188)***	(0.260)*	(0.206)**	(0.242)*	(0.208)
Observations	4826	3606	1361	4822	3602	1357
Villages	71	60	60	71	60	60
Village FE	YES	YES	YES	YES	YES	YES
Month x year FE	YES	YES	YES	YES	YES	YES

Note: Each observation is one village-month-year. GLS regression with weight of each observation in village i equal to 1/(total number of observations of village i). Table entries present coefficients on Water Tank is in Use. Huber-White standard errors in parentheses are clustered within villages. All regressions include a constant.

[1] Village size measured by mean number of households, where mean is taken across all months in which a village appears.

[2] An observation's order within a panel is the date on which the observation is recorded minus the first date in which the

[3] These regressions use all observations in the data and the table presents the coefficient on a variable which equals one for the

[4] The version of the data set used in these regressions have cleaning of village name and time codes and conversion of character strings to numeric values but no cleaning of numeric values for any variables.

Table 5: The Effect of Water Improvement on Diarrhea and Malaria, According to Different Definitions of When Water Improvement Began

	Outcome	Diarrhea Cases Treated			Malaria Cases Checked		
	Sample	A: All with known τ	B: τ in [-24,59]	C: τ in [-12,11]	A: All with known τ	B: τ in [-24,59]	C: τ in [-12,11]
Sample	Treatment begins when	(1)	(2)	(3)	(4)	(5)	(6)
Original	Tank turns on	-0.59	-0.587	-0.453	-0.372	-0.348	-0.236
		(0.125)***	(0.139)***	(0.178)**	(0.194)*	(0.233)	(0.233)
	Observations	5507	4077	1628	3797	2792	1042
	Villages	96	74	73	94	73	52
Original	Tank built	-0.574	-0.668	-0.192	-0.703	-0.79	-0.243
		(0.196)***	(0.237)***	(0.291)	(0.344)**	(0.386)**	(0.352)
	Observations	5507	3029	1091	3797	2068	642
	Villages	96	51	50	94	50	33
Original	90% of latrines	-0.2	-0.192	0.179	-0.505	-0.141	0.247
	complete	(0.137)	(0.163)	(0.322)	(0.212)**	(0.162)	(0.283)
	Observations	5507	2747	1132	3797	2158	883
	Villages	96	53	53	94	53	46
I	Tank turns on	-0.548	-0.562	-0.353	-0.363	-0.409	-0.171
		(0.122)***	(0.132)***	(0.154)**	(0.180)**	(0.198)**	(0.219)
	Observations	11230	4924	1955	7683	3440	1294
	Villages	342	90	89	297	89	66
I	Tank built	-0.483	-0.536	-0.323	-0.646	-0.712	-0.997
		(0.174)***	(0.217)**	(0.237)	(0.296)**	(0.396)*	(0.645)
	Observations	11230	4415	1639	7683	3150	1077
	Villages	342	77	76	297	75	55
I	90% of latrines	0.092	0.15	0.268	0.197	0.108	0.476
	complete	(0.094)	(0.104)	(0.169)	(0.283)	(0.179)	(0.217)**
	Observations	11230	5562	2628	7683	4337	2137
	Villages	342	126	126	297	126	116
II	Tank turns on	-0.684	-0.497	-0.168	-0.606	-0.485	-0.489
		(0.257)**	(0.233)**	(0.408)	(0.470)	(0.416)	(0.332)
	Observations	1625	1108	403	1217	839	308
	Villages	30	19	18	29	19	14
II	Tank built	-0.764	-0.91	-0.068	-0.997	-1.027	-0.353
		(0.226)***	(0.256)***	(0.385)	(0.755)	(0.746)	(0.443)
	Observations	1625	934	329	1217	780	286
	Villages	30	16	15	29	16	14
II	90% of latrines	-0.549	-0.438	-0.091	-0.903	-0.432	0.503
	complete	(0.255)**	(0.303)	(0.367)	(0.622)	(0.541)	(0.934)
	Observations	1625	717	267	1217	641	242
	Villages	30	12	12	29	12	12
Village FE		YES	YES	YES	YES	YES	YES
Month x year FE		YES	YES	YES	YES	YES	YES

Note: Each observation is one village-month-year. GLS regression with weight of each observation in village i equal to 1/(total number of observations of village i). Huber-White standard errors in parentheses are clustered within villages. All regressions include a constant.

Original Sample: as in Tables 2 and 3

Sample I: in villages where the monthly data report other measures of when water treatment begins, those measures must be no later than the measure used in this estimate

Sample II: as in sample I with the additional restriction that the report of when water treatment begins which is used in this estimate must be within 3 months of when Gram Vikas' inventory indicates that the water tank is complete.

Table 6. The Effect of Water Improvement on Other Outcomes

Sample Outcome	A: All with known τ (1)	B: τ in [-24,59] (2)	C: τ in [-12,11] (3)
Specific Diarrhea Measures Available in 1997-1999 Only:			
Days with diarrhea, kids 0-18	-0.367	-0.28	-0.466
	(0.174)**	(0.166)	(0.298)
Observations	929	740	473
Villages	35	25	24
Mean outcome τ<0	1.19	1.23	1.29
Cases of diarrhea, kids 0-18	-0.008	0.003	0.086
	(0.039)	(0.028)	(0.049)*
Observations	609	468	283
Villages	35	25	24
Mean outcome τ<0	0.09	0.07	0.05
Workdays lost to diarrhea	-1.586	-1.382	-2.661
	(0.806)*	(1.050)	(1.109)**
Observations	1018	808	524
Villages	35	25	24
Mean outcome τ<0	4.96	5.06	5.19
Other Health Outcomes:			
Borderline malnourished[1]	-0.059	-0.072	-0.042
	(0.037)	(0.042)*	(0.033)
Observations	1120	917	441
Villages	70	56	44
Mean outcome τ<0	0.46	0.46	0.45
Severely malnourished[1]	-0.015	-0.007	-0.01
	(0.017)	(0.016)	(0.021)
Observations	1187	971	457
Villages	71	56	44
Mean outcome τ<0	0.13	0.13	0.12
Fever: cases treated	-0.399	-0.622	-0.336
	(0.305)	(0.351)*	(0.419)
Observations	4634	3359	1145
Villages	95	75	56
Mean outcome τ<0	3.16	2.89	2.75
Population aged over 18	3.351	1.783	3.439
	(2.546)	(2.592)	(3.606)
Observations	3837	2863	1033
Villages	93	75	56
Mean outcome τ<0	337.03	325.46	323.07
Number of children in village	1.068	-0.282	0.286
	(2.291)	(1.910)	(0.531)
Observations	3114	2400	889
Villages	89	74	52
Mean outcome τ<0	194.55	184.94	183.28

Table 6 (continued)

Outcome	Sample A: All with known τ (1)	B: τ in [-24,59] (2)	C: τ in [-12,11] (3)
Education Outcomes:			
School: total days attended per child	0.374	0.416	1.026
	(0.547)	(0.543)	(0.341)***
Observations	1776	1358	519
Villages	83	69	48
Mean outcome τ<0	6.49	6.61	6.53
Village FE	YES	YES	YES
Month x year FE	YES	YES	YES

Note: Each observation is one village-month-year. GLS regression with weight of each observation in village i equal to 1/(total number of observations of village i). "Borderline malnourished" includes also children who are severely malnourished. Huber-White standard errors in parentheses are clustered within villages. All regressions include a constant.
1 Malnutrition is share of tested children, all of whom are aged 0-5 years. Borderline malnutrition denotes grade I or II malnutrition by WHO standard, while malnourished indicates grade III or IV maIntutrition by WHO standard.

Table 7: Correlates of Village-Specific Effects of Water Improvement on Health

	Diarrhea Cases Treated	Malaria Cases Treated
	(1)	(2)
Scheduled tribe or scheduled	-0.009	0.489
caste (share of households)	(0.260)	(0.521)
Number of households	0.011	-0.004
	-(0.007)	(0.008)
Number of children	-0.006	0.004
	(0.003)	(0.004)
Pre-treatment outcome	0.057	-0.257
	-(0.153)	(0.099)**
First month in panel[1]	0.006	0.011
	-(0.008)	(0.012)
# pre-treatment observations	0.027	-0.015
	(0.014)*	-(0.025)
# post-treatment observations	0.009	0.017
	(0.009)	(0.014)
Villages	63	48
R-squared	0.1	0.26

The dependent variable the estimated coefficient on the interaction of a village fixed effect with a $\tau \in [0,11]$ indicator (as plotted in Figure 6) in a GLS regression with weights equal to the inverse of the number of village-specific observations on the outcome of interest. Regressions also include village fixed effects, time fixed effects, and one indicator for $\tau \geq 12$. ST/SC denotes the mean share of village households that are scheduled tribe or scheduled caste. Population denotes the mean number of housheolds in the village. Children denotes the mean number of children in the village. Pre-treatment Diarrhea denotes the mean number of cases of diarrhea in the village before water improvement began. All means are taken across all non-missing observations within the village. All regressions include a constant.

[1] First month in panel: March 1997=1, April 1997=2, …, April 2007=126

Table 8: Effect of Program on Health Outcomes, by Post-Intervention Period

Outcome	Diarrhea Cases Treated		Malaria Cases Treated	
Sample	All with known τ	τ in [-24,59]	All with known τ	τ in [-24,59]
	(1)	(2)	(3)	(4)
Panel A: By 12-Month Period				
τ in [0,11]	-0.568	-0.743	-0.435	-0.433
	(0.161)***	(0.203)***	(0.178)**	(0.258)*
τ in [12,23]	-0.704	-1.154	-0.117	-0.149
	(0.235)***	(0.345)***	(0.28)	(0.42)
τ in [24,35]	-0.77	-1.353	0.239	0.215
	(0.292)***	(0.461)***	(0.38)	(0.59)
τ in [36,47]	-0.605	-1.321	0.083	-0.026
	(0.373)	(0.584)**	(0.46)	(0.74)
τ in [48,59]	-0.79	-1.77	0.308	0.272
	(0.450)*	(0.714)**	(0.56)	(0.85)
τ ≥ 60	-0.759		0.539	
	(0.59)		(0.69)	
Observations	5408	4077	4572	3360
Villages	95	74	94	75
R-squared	0.35	0.38	0.27	0.29
Mean outcome τ<0	1.08	1.09	1.33	1.28
F test: coeffs are equal	2.3	3.71	1.12	1.48
p-value: coeffs are equal	0.05	0.01	0.36	0.22
Panel B: Before 36 Months and After 36 Months				
τ in [0,35]	-0.552	-0.539	-0.492	-0.68
	(0.135)***	(0.148)***	(0.189)**	(0.232)***
τ ≥ 36	-0.423	-0.406	-0.584	-1.033
	(0.212)**	(0.231)*	(0.309)*	(0.359)***
Observations	5408	4077	4572	3360
Villages	95	74	94	75
R-squared	0.35	0.37	0.27	0.29
Mean outcome τ<0	1.08	1.09	1.33	1.28
F test: coeffs are equal	1.04	1.00	0.21	2.08
p-value: coeffs are equal	0.31	0.32	0.65	0.15
Village FE	YES	YES	YES	YES
Month x yr FE	YES	YES	YES	YES

Note: Each observation is one village-month-year. Table presents GLS regressions with weight of each observation in village i equal to 1/(total number of observations of village i). Asterisks denote statistical significance at 90% (*), 95% (**), and 99% (***) confidence. Huber-White standard errors clustered within villages. All regressions include a constant.

Table 9: RHEP Costs per Household (US$)

	Household	External	Total
Construction costs:			
Toilet and bath house	66.7	77.8	144.4
Water supply	72.2	105.6	177.8
Project development		116.7	116.7
Total	138.9	300.0	438.9
Maintenance (annual)			11.1

Notes: Information from Gram Vikas (2001) and Gram Vikas (2004), as well as authors' interviews with Gram Vikas staff. Costs calculated for a "typical" village of 50 households. Figures in Indian rupees converted to US dollars at 45 INR/USD. Value of village labor assumed to be $1/day, roughly the wage for agricultural day labor. "Project development" costs are primarily wages paid to Gram Vikas staff during startup and data collection.

DATA APPENDIX

This appendix describes preparation of the data for analysis. The original data were collected in three separate formats – one format for 1997-1999, one for 2000-2001, and another for 2002-2005 – where each format presented a slightly different set of variables. The original handwritten survey data were scanned and transmitted to a data firm in India, which entered them into electronic records using double-entry for accuracy. The original records called for English entries, but data were sometimes recorded in Oriya (the predominant language in Orissa). Our data entry company assigned Oriya-literate employees to our project. Gram Vikas also converted paper records to electronic format starting in 2002. Comparison of the 2002-2005 data entered by GV to that entered by the data firm indicates that the data firm's version had fewer errors, so for 2002-2005 we mainly use the Indian firm's version of the data, though we fill in some gaps (presumably from paper records that were lost) with the GV-entered data. Because we only scanned the paper records through 2005, all 2006 and 2007 data come from GV's data-entry. Although all forms reported deaths, the total number of reported deaths was extremely small and hence not used in analysis.

The most relevant differences between the survey forms are diarrhea variables included in 1997-1999 and excluded from other forms, and other variables which appear in all forms except those used in 1997-99. The 1997-1999 form does not report on any illnesses other than diarrhea, but it does report the number of cases of diarrhea for children, the number of days that children suffer diarrhea, and the number of working days lost to diarrhea. Other forms do not report these more detailed diarrhea measures. Also, the 1997-1999 data do not report school attendance, the number of households, number of individuals, or the number of children.

Basic data cleaning steps included the following:

- Correction of village names and code numbers
- Correction of incorrectly entered dates
- Conversion of numbers entered as strings.
- While most villages have continuous panels – every consecutive month is observed – some observations were distant temporal outliers and appeared a year or more after all other observations from the same village. Those outliers generally have values differing substantially from the other values in the village, so we exclude them from the analysis.

By comparing against images of the original handwritten monthly survey documents we conclude that most of these anomalies came from the original paper data forms recorded by the supervisor, rather than from data-entry errors.

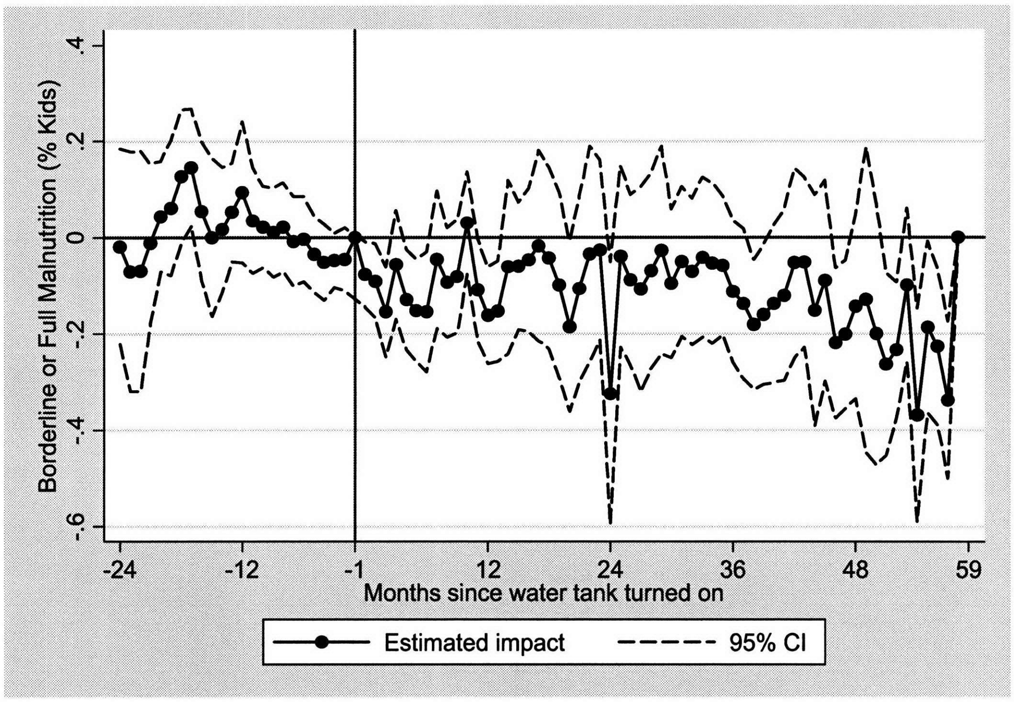
The data offer several differing reports of when the month in which the water improvement began, and the paper's analysis focuses on the report which appears to be most reliable, though some specifications use different definitions of when water improvement began. The different reports are the following:

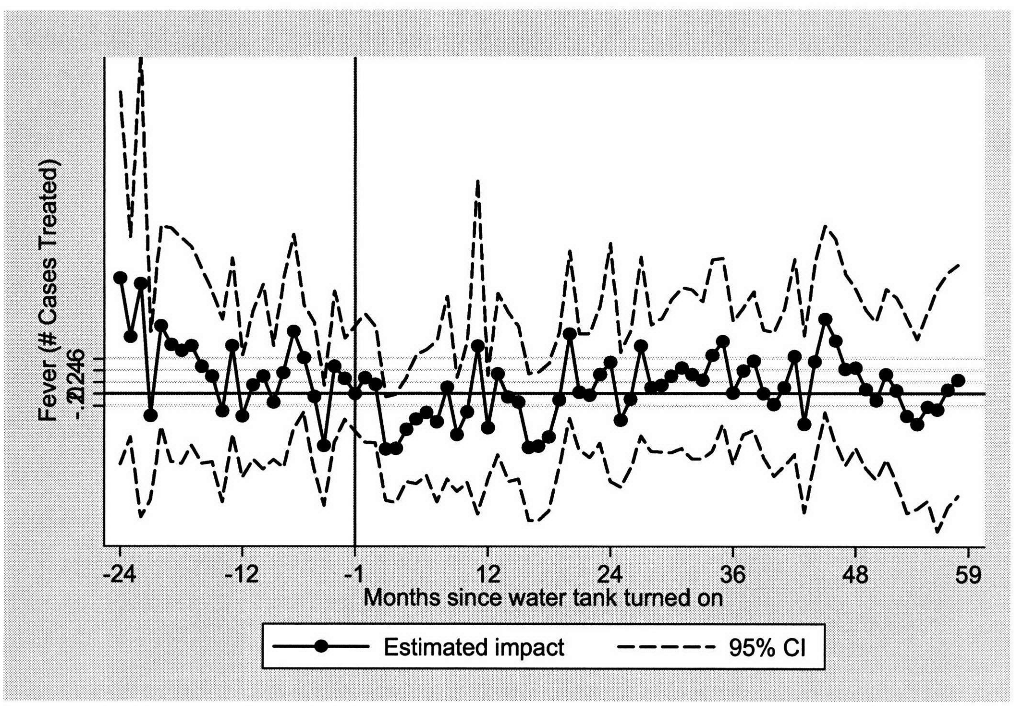
- Each month, the surveyor indicated whether the village water tank had begun operation
- Each month, the surveyor indicated whether construction of the village water tank had completed
- Each month, the surveyor recorded the number of latrines the village had built.
- In 2007, Gram Vikas compiled a list reporting, for each village that had a water tank built on 2005 or before, the month when the village's water tank finished construction.

The first seems most reliable because it is operation rather than mere construction of the water tank which should affect welfare measures. Also, monthly reports by a surveyor visiting the village are likely to more accurately reflect the state of the village water tank than a retrospective summary written ten years later by GV will. While the first two reports have fairly high correlation, the correlation with these dates and the ex-post GV dates is low. Some villages have reports that the water tank began operation before water tank construction was complete or before latrine construction is complete. Since this should in principle be impossible, we exclude these villages from analysis.

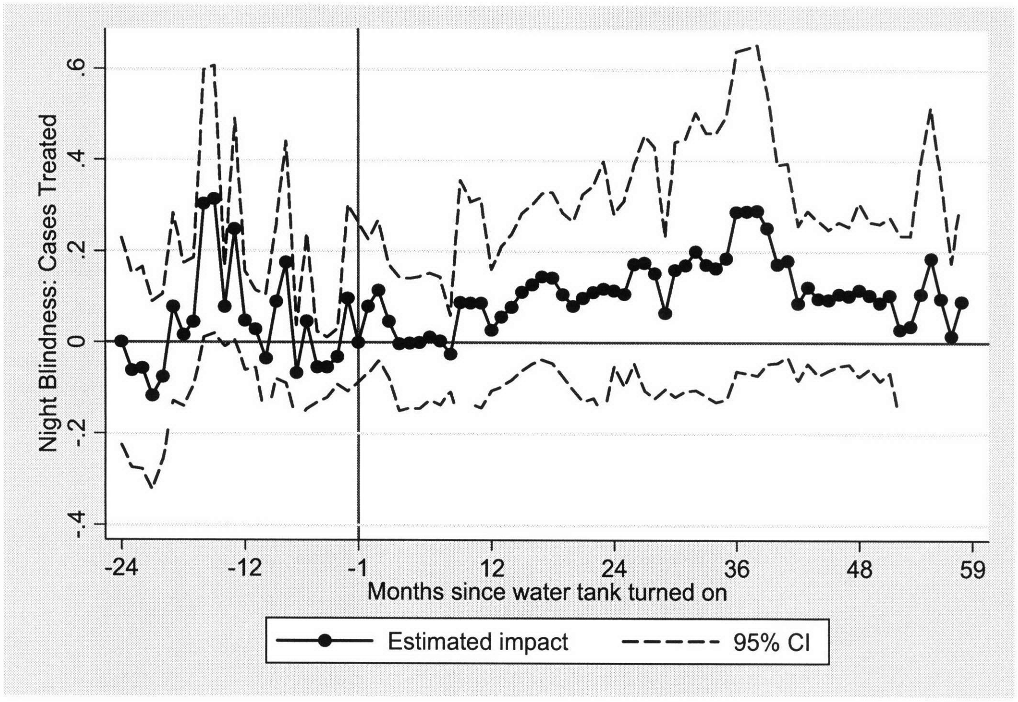
The main estimates also reflect cleaning of the underlying data for all variables. Much of the cleaning uses redundant information in the village-level data. For example, population levels in a village are smooth, so comparing adjacent time periods clarifies when outlier population values are typos. Specifically, the following general rules for cleaning the data represent some of the more important:

- For all variables, extreme outlier values which can be identified as obvious typos are corrected, while outliers which cannot be assigned a certain correct value are coded to be missing.
- Water tank variables (tank construction, usage, latrine construction) are defined to be nondecreasing, so the first village report of when a tank is being used implies that the tank is in use for all future months. This rule is the predominant reason why the cleaned data have slightly more usable observations than the uncleaned data have.
- Some of the variables contain three separate reports for men, women, and the entire population. When the values for men and women do not sum to the total village value, we identify the error by comparing against subsequent or preceding months for the same village. Similarly, the total number of malnutrition measurements taken should equal the sum of normal, borderline, and malnourished children. Also the number of children tested for malnutrition should not exceed the number of children in the village that month.
- In some years the data report the number of illnesses detected, and the number of cases of a particular illness which are detected should not exceed the number checked or treated.
- Although some villages have both a village school and a non-formal education program, the school attendance data only refer to the village school. School attendance must roughly correspond to the number of days in the month when the village school was open.
- Population values (households, adults, children) which have the value zero are recoded to missing. The total village population must roughly equal the number of children plus the number of adults.


.4
.2
0
-.2
-.4
-.6
Borderline or Full Malnutrition (% Kids)
-24
-12
-1
12
24
36
48
59
Months since water tank turned on
Estimated impact
95% CI

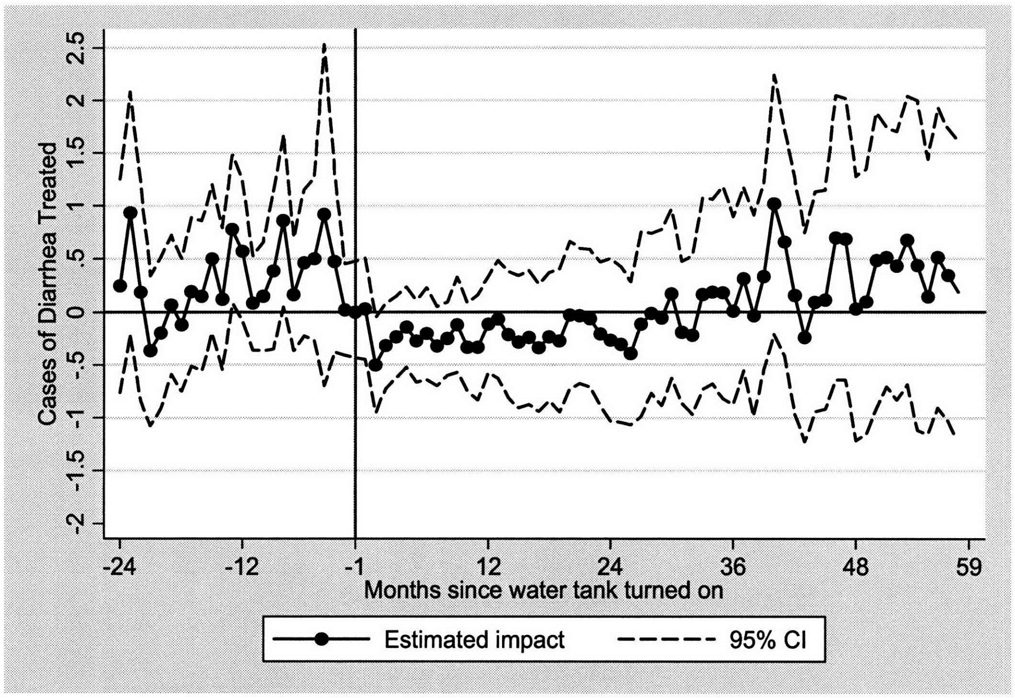

Fever (# Cases Treated)
-24
-12
-1
12
24
36
48
59
Months since water tank turned on
Estimated impact
95% CI

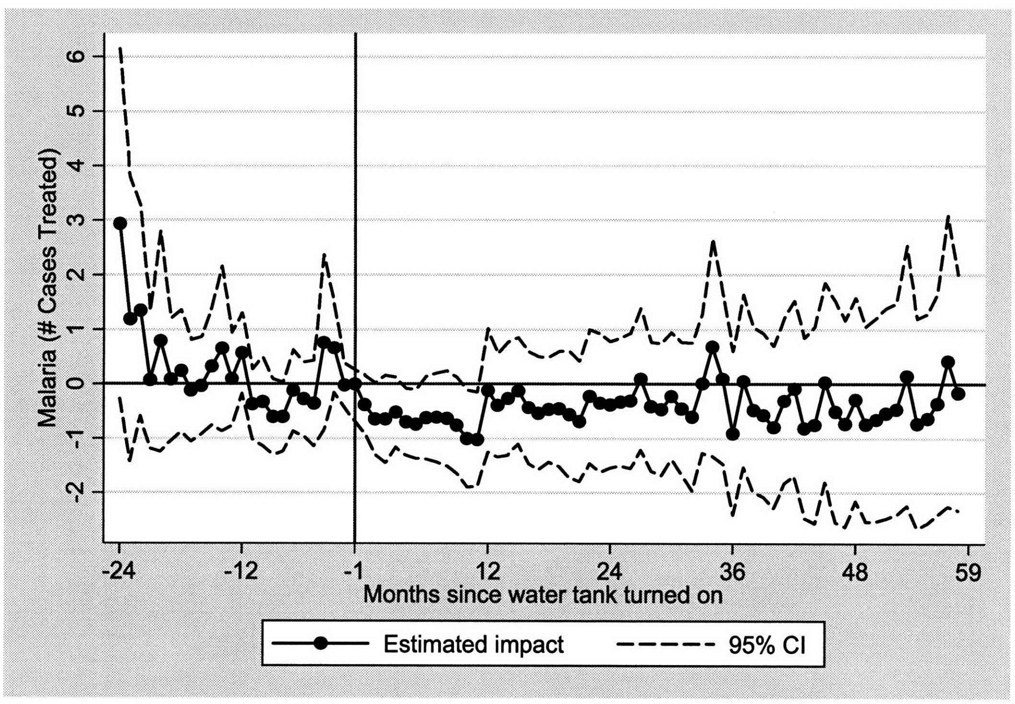
Appendix Figure 1 (Continued)



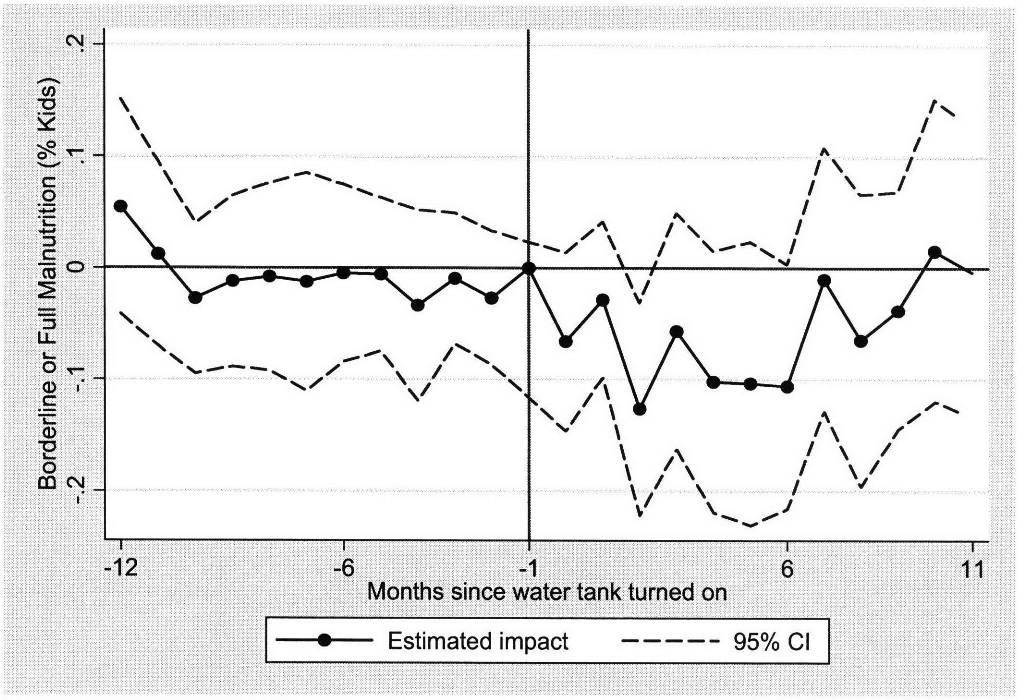
Note: these figures use observations with τ in [-24,59]. The points are parameter estimates for each value of τ in a GLS regression which includes village fixed effects and year fixed effects, with weights equal to the inverse of the number of months in which a village has non-missing outcomes. Huber-White covariance matrix allows within-village clustering.

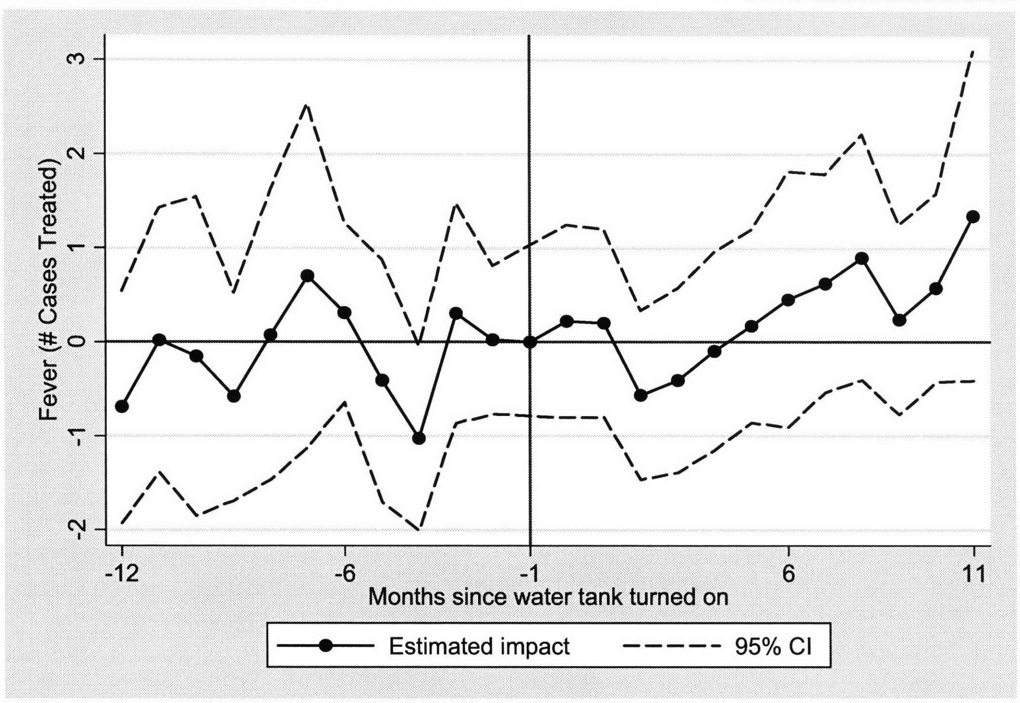
Appendix Figure 1b: Impact While Controlling for Number of Children and Adults.



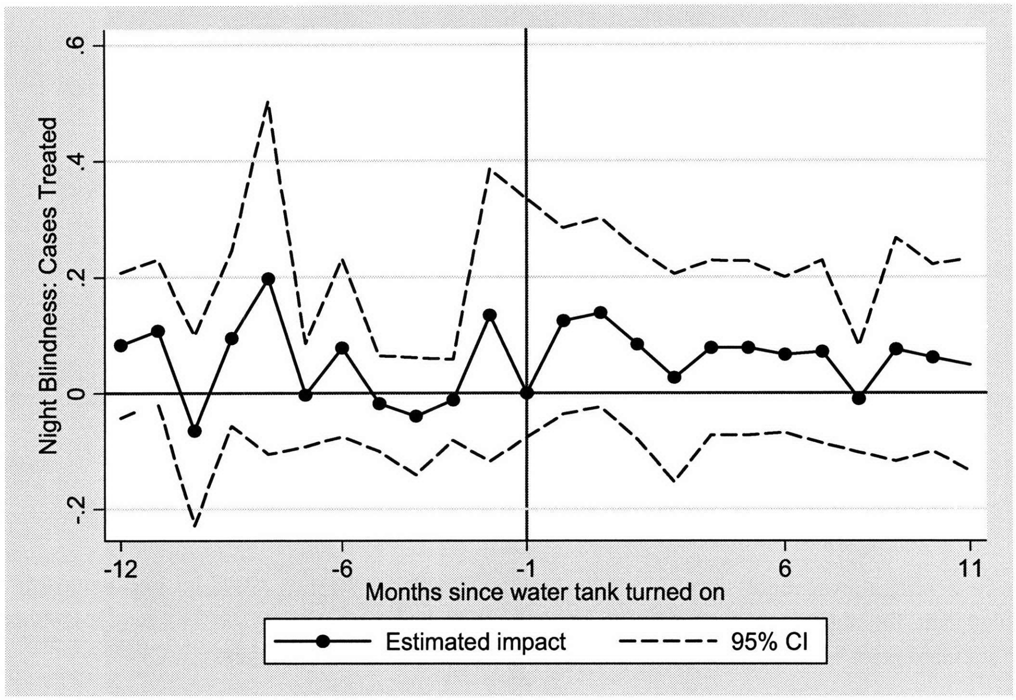


Appendix Figure 2: -12/11 month graphs for other outcomes



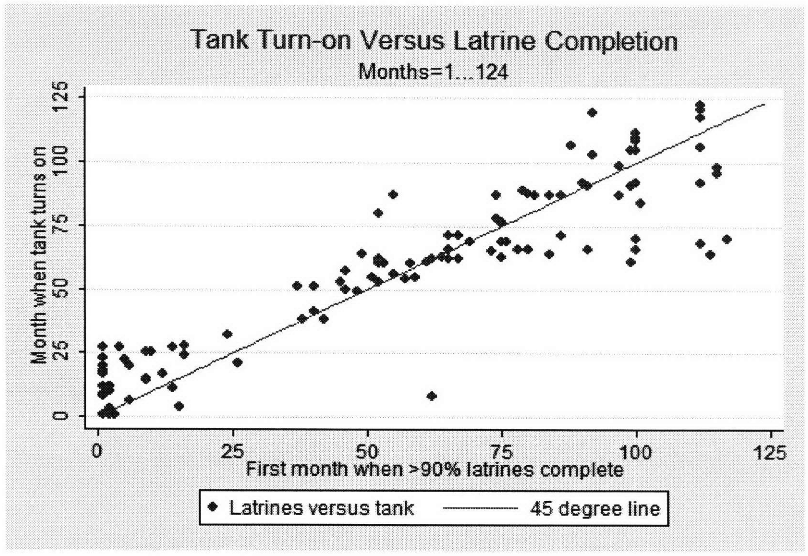


Appendix Figure 2 (Continued)

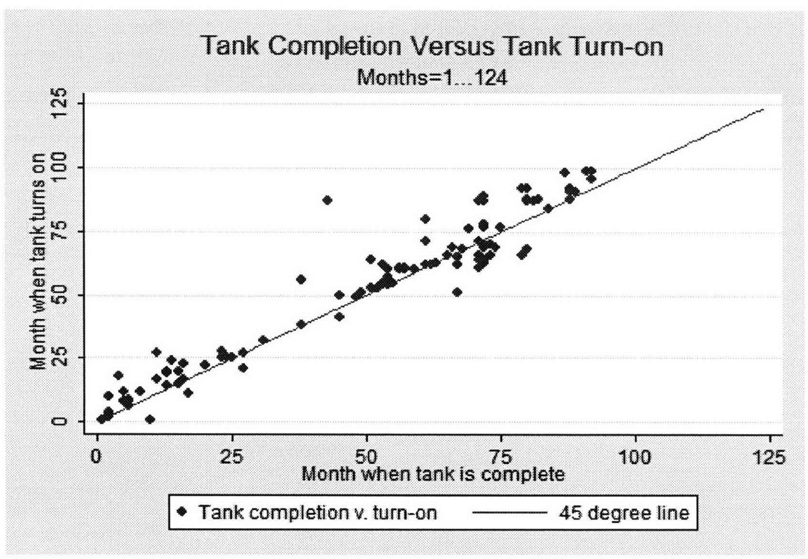


Notes: these figures use observations with τ in [-12,11]. The points are parameter estimates for each value of τ in a GLS regression which includes village fixed effects and year fixed effects, with weights equal to the inverse of the number of months in which a village has non-missing outcomes. Huber-White covariance matrix allows within-village clustering.

Appendix Figure 3: Comparison of Measures of Program Initiation



Notes: in a village-level regression of "Month tank turns on" on "Month >90% latrines complete" and a constant, the regression coefficient on "Month >90% latrines complete" is 0.82 with a Huber-White standard error of 0.035. The regression uses observations on 120 villages.



Notes: in a village-level regression of "Month tank turns on" on "Month tank complete" and a constant, the regression coefficient on "Month tank complete" is 0.98 with a Huber-White standard error of 0.021. The regression uses observations on 109 villages. Beginning after March 2005 (month number 99), the data no longer report "tank completion" separately from "tank turn-on."

Appendix Table 1: Pairwise Correlation Coefficients Between Health Outcome Variables	Diarrhea cases checked	Diarrhea cases treated	Malaria cases checked	Malaria cases treated	Fever cases checked	Fever cases treated	Borderline or severely malnourished	Severely malnourished
Panel A. All Observations								
Diarrhea cases checked	1.00							
Diarrhea cases treated	0.96	1.00						
Malaria cases checked	0.13	0.14	1.00					
Malaria cases treated	0.11	0.16	0.99	1.00				
Fever cases checked	0.22	0.25	0.28	0.26	1.00			
Fever cases treated	0.18	0.27	0.27	0.27	0.92	1.00		
Borderline or severely malnourished	-0.07	-0.07	0.05	0.00	0.08	0.06	1.00	
Severely malnourished	0.09	0.05	0.04	0.00	-0.04	-0.05	0.38	1.00
Panel B. τ in [-24,59]								
Diarrhea cases checked	1.00							
Diarrhea cases treated	0.97	1.00						
Malaria cases checked	0.13	0.13	1.00					
Malaria cases treated	0.11	0.17	0.99	1.00				
Fever cases checked	0.18	0.19	0.28	0.27	1.00			
Fever cases treated	0.15	0.25	0.28	0.27	0.91	1.00		
Borderline or severely malnourished	-0.09	-0.09	0.00	-0.04	0.04	0.03	1.00	
Severely malnourished	0.08	0.04	0.01	-0.02	-0.06	-0.06	0.41	1.00
Panel C. τ in [-12,11]								
Diarrhea cases checked	1.00							
Diarrhea cases treated	0.98	1.00						
Malaria cases checked	0.08	0.08	1.00					
Malaria cases treated	0.10	0.26	0.97	1.00				
Fever cases checked	0.18	0.17	0.14	0.11	1.00			
Fever cases treated	0.12	0.15	0.09	0.12	0.94	1.00		
Borderline or severely malnourished	-0.06	-0.06	-0.01	-0.11	0.03	0.05	1.00	
Severely malnourished	0.30	0.22	0.21	0.07	0.18	0.15	0.07	1.00