Essays in Development and Labor Economics

by

Claire Elise Duquennois

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Agricultural and Resource Economics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Jeremy Magruder, Chair
Professor Elisabeth Sadoulet
Assistant Professor Supreet Kaur

Spring 2020

Essays in Development and Labor Economics

Abstract

Essays in Development and Labor Economics

by

Claire Elise Duquennois

Doctor of Philosophy in Agricultural and Resource Economics

University of California, Berkeley

Professor Jeremy Magruder, Chair

This manuscript presents three essays on development and labor economics. Each chapter examines a factor that affects the educational or labor market outcomes of individuals in low income environments. All three papers take advantage of detailed secondary data sources ranging from detailed household surveys, to data from a computerized homework platform. This data is then used to conduct careful examinations of the educational outcomes and labor market choices made by individuals in low-income environments. In Chapter 1, I show that students of lower socio-economic status (SES) under perform on mathematic exams that feature a higher share of monetary themed questions due to an attention capture effect of poverty. In Chapter 2, we conduct a detailed characterization of the role that the seasonality of labor demand in agriculture plays in the persistence of rural poverty in Malawi. Finally in Chapter 3, I evaluate whether nominal wage rigidity created smaller distortionairy effects on rural labor markets in irrigated areas of India that experience less production volatility.

In my first manuscript, I examine how students with lower SES indicators perform worse when randomly given an exam or assignment that features a larger share of mathematics questions in which money is salient. I examine this effect and find a similar pattern in three different data sets: a homework platform in the United States, a national assessment exam in Mexico and an international cross-country standardized assessment examination. I find that this pattern begins as early as in the fourth grade, is largest for the most disadvantaged and is responsive to income shocks. For students with SES indicators below the national median, a 10 percentage point increase in the share of monetary themed questions depresses exam performance by 0.026 to 0.038 standard deviations. The magnitude of the effect represents about 6% of the overall performance gap for below median SES students. Evidence from a homework platform shows that acquiring a mathematical skill takes differentially more time and effort for low SES students when it is practiced using monetary prompts. Using question-level data, I confirm the role of financial salience by comparing performance on monetary and highly similar non-monetary questions. Furthermore, by leveraging the randomized ordering

of questions, I identify an attention capture effect on directly subsequent questions, providing evidence that the attention capture effects of poverty affect policy relevant outcomes outside of experimental settings as performance on examinations is a significant determinant of educational and economic opportunities.

In my second manuscript, using data from Malawi, we explore how the seasonality of agriculture contributes to the persistence of rural poverty in Sub-Saharan Africa. We show that labor calendars for rural households offer similar employment opportunities as for urban households in terms of time worked at peak planting time, but much lower opportunities throughout the rest of the year. Given the high level of urban unemployment in Malawi, an urban-based structural transformation does not offer a clear alternative to rural poverty through rural-urban migration. By contrast, we show that elements of both an agricultural and a rural transformation can help fill-in and smooth-out labor calendars, providing a pathway to rural poverty reduction. Importantly, in this paper, we develop a methodology to map labor time use calendars across an agricultural season that can be applied broadly to agricultural household surveys in other contexts.

Finally, in the third manuscript, I investigate whether downward nominal wage rigidity is more binding and has larger distortionary effects in Indian districts that face greater production volatility. I use the instrument developed for dam construction in Duflo and Pande (2007), and apply the identification strategy used in Kaur (2019) to estimate whether dam presence influences the magnitude of the distortions created by nominal wage rigidities. Despite suggestive evidence, results are inconclusive. Results are likely confounded by nonlinearity in the effect of dams, the effects of lagged rainfall on contemporaneous production as well as substitution towards water intensive farming strategies in irrigated areas that exposes producers to significant production drops in severe drought years.

To John Haugarth

you keep me happy, healthy, silly, and well fed
happy 10 years
<3

Contents

List of Figures

List of Tables

Acknowledgments

Coauthors and data: Chapter 2 of this dissertation is coauthored with Alain de Janvry and Elisabeth Sadoulet. I would like to to thank the ASSISTments team and in particular Korinn Ostrow for her help with the ASSISTments data and Andrew Dustan for his assistance with the COMIPEMS data. All errors are my own.

Personal thanks: I am deeply grateful for the advice and support I received from Jeremy Magruder, Elisabeth Sadoulet, Supreet Kaur, Alain de Janvry, and Hani Mansour. I am deeply grateful for all the guidance, encouragement and time you shared with me over the past years. Each in your own way, you have shown me how to be the kind of economist I want to become. Thank you.

I would also like to thank my fellow PhD students, particularly the 2015 ARE cohort, for being by my side through thick and thin. I thank you all for so much great advice and all the good times we got to share.

Finally, Mom, Dad, Annie and John, you are always there when I need you, reminding me that I have the patience and perseverance to do this.

Chapter 1

Fictional Money, Real Costs: Impacts of Financial Salience on Disadvantaged Students

Performance on examinations is a significant determinant of educational and economic opportunities. Using three data sets, I find evidence that students with lower socio-economic status (SES) indicators perform worse when randomly given an exam or assignment that features a larger share of mathematics questions in which money is salient. This pattern begins as early as in the fourth grade, is largest for the most disadvantaged and is responsive to income shocks. For students with SES indicators below the national median, a 10 percentage point increase in the share of monetary themed questions depresses exam performance by 0.026 to 0.038 standard deviations. The magnitude of the effect represents about 6% of the overall performance gap for below median SES students. Evidence from a homework platform shows that acquiring a mathematical skill takes differentially more time and effort for low SES students when it is practiced using monetary prompts. Using question-level data, I confirm the role of financial salience by comparing performance on monetary and highly similar non-monetary questions. Furthermore, by leveraging the randomized ordering of questions, I identify an attention capture effect on directly subsequent questions, providing evidence that the attention capture effects of poverty affect policy relevant outcomes outside of experimental settings.

1.1 Introduction

Performance on examinations matters. Test results are commonly used for assessment of students and schools, as an allocation criterion or admission requirement, and for licensing and certification. Student performance on examinations can thus have significant economic implications and determine future educational and economic opportunities.

Examinations may be an efficient mechanism to benchmark and rank a population based on a specific set of skills. The notion that they are fair, however, has increasingly been questioned. A significant concern is that performance differences reflect inequities in the testing process itself, rather than differences in underlying skills. Students of the same ability, but from different backgrounds, are known to respond differently to questions,[1] though there is limited understanding as to why (Freedle (2010), Editors (2010)). This paper explores one possible reason.

I investigate whether differential performance may be generated by the frequent use of monetary themed questions on mathematics examinations.[2] Open any first grade mathematics workbook and you will undoubtedly see simple algebra problems centered around the buying and selling of various items. These types of monetized scenarios are frequently used in early mathematical education the world over and are commonly featured on tests as well. I exploit the natural variation in the financial salience of mathematics exams that is generated from monetary questions. I begin by documenting that disadvantaged students differentially underperform on mathematics exams and assignments when they feature a larger share of monetary themed questions. I observe this result in three different datasets spanning three different contexts: a homework platform in the US, an international cross-country standardized exam and a national educational assessment exam in Mexico. Using data from the two examinations, I find that a 10 percentage point increase in the financial salience of the exam depresses the performance of students with socio-economic status (SES) indicators below the national median[3] by 0.026 and 0.038 standard deviations depending on the context. This is a non-negligible effect representing about 6% of the overall performance gap for below median SES students. This effect manifests as early as in the fourth grade, is largest for the most disadvantaged and is responsive to income shocks caused by rainfall. Furthermore, evidence from the homework platform shows that acquiring a mathematical skill requires differentially more time and effort for disadvantaged students when it is practiced using monetary themes.

There are a number of reasons why we might expect lower income students to perform differentially on these types of exercises. The literature on poverty and cognition has proposed that, for low income individuals, attention can become focused on scarcity and lead to stress and inattention, particularly when choices about money and finances are being considered. I investigate an attention capture mechanism that draws on recent experimental findings from this literature. I identify the effect of attention capture by matching monetary themed questions to similar non-monetary themed questions, and by exploiting the randomized ordering

[1] Differential performance by different ethnic and socio-economic groups has been documented on the SAT for instance. A proposed alternative scoring mechanism could shrink the performance differential between white and African-American test takers by a third (Freedle (2003), Santelices and Wilson (2010)).

[2] I define monetary themed questions as questions that involve topics such as buying, selling, making payments, saving and spending money or calculations using currency. Examples of monetary themed questions for the three datasets are presented in figure 1.6 and figures A.2 and A.3 in the appendix.

[3] The datasets I use feature different SES indicators: parental education levels, a school marginalization index and the share of students in a school receiving free or reduced price lunch.

of questions on the homework platform. This analysis of the itemized question level response data shows evidence of an attention capture effect. Comparison of student responses on monetary themed questions to highly similar non-monetary themed questions provides evidence that disadvantaged students underperform on questions that feature a monetary theme. Furthermore, by leveraging the randomized ordering of questions in the homework data, I observe a pattern of underperformance on questions that are placed subsequent to a monetary themed question. This pattern is consistent with an attention capture effect on subsequent questions for the low SES students and manifests in the other itemized exam data as well. A relationship between poverty and cognition has been observed in experimental settings using psychological tests. These findings show policy relevant impacts on student performance using real homework and examination scores.

The proposed mechanism draws heavily from recent ideas in the psychology of poverty literature regarding the relationship between cognitive functioning and poverty. This literature has suggested that poverty captures attention, generates intrusive and distracting thoughts that reduce an individual's cognitive resources (Mani et al. (2013), Shah, Mullainathan, and Shafir (2012), Shah, Mullainathan, and Shafir (2018), Tomm and Zhao (2016)). Though hard to differentiate, several mechanisms have been investigated. The limited cognition mechanism posits that economic decisions are more difficult for the poor as they face more difficult trade-offs which deplete their cognitive resources, leaving them with less cognitive control. This mechanism has been tested in a number of lab and field experiments (Mani et al. (2013), Shah, Mullainathan, and Shafir (2012), Spears (2011), Kaur et al. (2019)). The limited attention mechanism differs from the limited cognition mechanism in that it does not require a cognitively taxing economic decision. Rather, it simply suggests that, under conditions of poverty, attention becomes focused on scarcity, leading to stress and inattention to other issues. There have been a number of works evaluating the relationship between poverty and stress. Haushofer and Fehr (2014) provide an extensive review of this literature, concluding that the majority of findings support a causal link. The impacts on cognition, however are not as well established, with some contradictory results (Mani et al. (2013), Carvalho, Meier, and Wang (2016), Kaur et al. (2019)). A particular challenge to identifying this mechanism is the difficulty in using actual income variation, as it correlates with changes in nutrition which are known to generate cognitive effects (particularly for children) even in the short run (Anderson, Gallagher, and Ritchie (2018), Gassman-Pines and Bellows (2018)). The mechanism that I propose, while drawing heavily on the limited attention mechanism, adds the caveat that something must capture attention to activate temporary inattention and errors. Even if the effects are temporary, the fact that this distraction occurs precisely when a low income individual is required to make potentially cognitively demanding decisions about financial resources makes any such effect important to understand for scholars who study decision making in the context of poverty.

This mechanism is reminiscent of the stereotype threat effect first posited by Steele and Aronson (1995), who suggested that an individual's performance on an examination is sensitive to priming about a stereotype of their group. This hypothesis has generated a significant amount of research, primarily in lab and field-lab settings (Spencer, Logel, and

Davies (2016), Fryer Jr, Levitt, and List (2008)). Empirical challenges and research preferences within disciplines, has limited field research on stereotype threats[4] and on poverty's effects on cognitive functioning. By utilizing real examination data, I address this gap in the literature and alleviate concerns of experimenter demand effects. I also remove concerns about sensitivity to specifically designed wording of priming statements that may not be reflective of typical examination conditions. By using secondary sources for my examination and homework data, I am able to estimate the effects of the tested mechanism under normal exam and homework conditions and show that the experimental results on the cognitive effects of poverty have external validity beyond the experimental setting. Though the effects on attention may be temporary, the impacts are economically meaningful because exam scores are frequently used to determine important economic opportunities such as eligibility for further education, placement in schools or access to scholarships. Furthermore, the homework effects I find imply impacts on the entire learning process. In this regard, I am addressing a gap in the cognitive functioning literature by investigating real costs and showing that effects that have thus far been measured using psychological tests also impact exam scores, a policy relevant metric.

The rest of the paper is organized as follows: Section 1.2 describes the three primary datasets used in my analysis. Section 1.3 presents estimation methods and results on aggregate exam and assignment performance, using exam level variation in financial salience. Section 1.4 investigates potential mechanisms using itemized question level data, providing evidence of an attention capture effect. Section 1.5 discusses implications for high stakes examinations and simulates the effects on exam performance and high school placement using data from a high school entrance exam in Mexico City. Finally, section 1.6 concludes.

1.2 Data

This paper uses data from three different sources to provide evidence that the share of monetary questions featured on an exam or assignment differentially affects the performance of low socio-economic status (SES) students. In addition to confirming the replicability of this result across a variety of contexts, each of these three datasets has distinct attributes allowing for a more thorough understanding of the mechanisms behind the general result. The ASSISTments homework platform in the US allows me to show effects on learning and effort and to exploit the randomized ordering of questions to identify attention capture effects. The cross-country Trends in International Mathematics and Science Study (TIMSS) exam provides examination setting evidence and shows that these results on exam performance and attention capture are widely generalizable. The Mexican Evaluación Nacional de Logros Académicos en Centros Escolares (ENLACE) exam provides evidence from a more tradi-

[4] A few researchers have experimented with placement of demographic questions around actual AP exams (Stricker and Ward (2004), Danaher and Crandall (2008)), while Wei (2012) exploits natural variation in pretest background questions to detect a stereotype reactance effect in the NAEP math test.

tional examination setting and allows me to exploit the panel nature of the data to show that the effects on examinations respond to income shocks. Table 1.1 summarizes the key attributes of each of these datasets.

Homework Platform Micro Data: ASSISTments

ASSISTments is a free online homework platform in the US operated by the Worcester Polytechnic Institute's Computer Science Department. Teachers create accounts on ASSISTments and then use the platform to assign homework to their students. Teachers can generate their own problems sets or use existing material. The most widely used format on ASSISTments is called 'skill builders.' Skill builders consist of a large pool of questions meant to practice a specific skill. When assigned a skill builder, students are expected to respond to questions until they answer three in a row correctly. Several hints are attached to each question; the students can consult the hints and can make several attempts at answering each question. Importantly, there is no set order to the questions a student will face, as questions are randomly drawn without replacement from the question pool of the assigned skill builder.

Though ASSISTments is not a widely used homework aid, it is partially funded by the NSF as a research platform and assignment data is available for research purposes. Student level user data includes the sequence of questions a student faced, the amount of time spent on each question, the number of attempts made,[5] the number of hints requested, and whether they completed the skill builder by answering three questions in a row correctly. ASSISTments data does not include any socio-economic indicators, though ASSISTments made an exception and agreed to match the schools in their user pool to National Center for Educational Statistics (NCES) data in order to supply me with de-identified school level SES indicators. These include school enrollment and a count of students enrolled in free and reduced price lunch programs.

I use the data for the 13 skill builders that feature both monetary and non-monetary themed practice questions. The main sample consists of 23,208 different student assignments covering thirteen different skill builders, featuring 1690 questions, of which 519 are coded as monetary themed.[6] Figure 1.1 shows that there is significant variation in the proportion of monetary questions featured on student assignments and that there is also substantial variation in the share of schoolmates receiving free or reduced price lunch, the two key sources of variation I exploit in this dataset.

[5]To focus on students who are actually engaged in completing the assignment, time spent on a question is coded as NA if the student spends more than 8.8 minutes (the 90th percentile) or less than 5 seconds on a question. Outlier attempt counts beyond 8 attempts (the 90th percentile) are also coded as NA.

[6]Cleaning primarily involved limiting the sample to student assignments for which the SES indicator is observed and the monetary indicator is defined for all questions. Furthermore, though rarely exercised, teachers have the option of fixing the ordering of questions. For each assignment, I test that monetary questions are not correlated with a particular sequential positioning and drop any assignments where this correlation is significant at the 10% level. I also drop any skill builder that features multiple part questions.

The ASSISTments data has the distinct advantage of random question ordering. This is key to identification of attention capture effects on subsequent questions, because this alleviates the concern that systematic placement of questions may be impacting estimates. Furthermore, the ASSISTments data provides insight into the learning process which can shed some light on the performance gaps that become evident in the examination data.

Cross-Country Exam Micro Data: TIMSS

The Trends in International Mathematics and Science Study (TIMSS) is an international standardized test in math and science administered by the International Association for the Evaluation of Educational Achievement (IEA) to a random sample of 4th and 8th graders in participating countries.[7] [8] These examinations have been taking place every 4 years since 1995. The TIMSS tests are one of the main sources reported by the World Bank for international learning outcomes data. Sampling follows a stratified two-stage cluster sample design. First, a probability weighted stratified random sample of schools is selected and then a random sample of classes is selected from within each school. This procedure generally results in the selection of approximately 150 schools and 4000 students per country.

The advantage of the TIMSS data is that it features question level responses so that I observe student answers to each question on their exam. In addition to student responses, student, teacher and school surveys are also administered. Importantly, since 2011, most countries also administered a parental questionnaire for the 4th grade exam which reports basic occupational and educational categories of the parents.[9] For the 4th grade exams, 53 countries participated in 2015 and 60 in 2011, though parental questionnaires were only administered in 50 and 37 respectively. Most of these countries are middle to high income (see figure 1.4).[10] For this dataset I opt to use the highest reported parental education

[7] TIMSS exam design, sampling and implementation is executed in coordination with participating countries via country representatives and national statistical organizations. For instance, the NCES, part of the U.S. Department of Education, is responsible for the collaboration and implementation of TIMSS in the US. In collaboration with TIMSS sampling experts, participating countries define their national target population, apply the TIMSS requirements to construct the country's sampling frame, and select a nationally representative sample of schools and students (see LaRoche, Joncas, and Foy (2016)).

[8] For brevity I use the term country, though the IEA also works with regional authorities that wish to benchmark their performance.

[9] The 8th grade TIMSS data also provides these indicators but they are elicited from the students. I opt to use the 4th grade data out of concern that 8th grade student misreporting could correlate with exam performance.

[10] Countries participating in 2015 4th grade exams that administered parental questionnaires include: Abu Dhabi, Australia, Bahrain, Flemish Belgium, Buenos Aires, Bulgaria, Canada, Chile, Chinese Taipei, Croatia, Cyprus, Czech Republic, Denmark, Dubai, Finland, France, Georgia, Germany, Hong-Kong, Hungary, Indonesia, Ireland, Iran, Italy, Japan, Kazakhstan, Republic of Korea, Kuwait, Lithuania, Morocco, Northern Ireland, Netherlands, Norway, New Zealand, Oman, Ontario, Poland, Portugal, Qatar, Quebec, Russian Federation, Saudi Arabia, Singapore, Serbia, Slovak Republic, Slovenia, Spain, Sweden, Turkey, United Arab Emirates.

Countries participating in 2011 4th grade exams that administered parental questionnaires include: Abu category as my primary SES indicator.[11] Furthermore, since SES may be associated with different education levels in different national contexts, I also generate an indicator variable for whether the highest parental education reported for a student falls below the national median as observed in the TIMSS data.[12] 469,849 students have taken the examination over the two rounds of 2011 and 2015. My main sample will consist of the 379,468 students for whom parental education is available. There is selection into the main sample due to non-random parental non-response. Nevertheless, because treatment is random within the sample, estimates are internally valid.[13]

Each year, a student taking a TIMSS exam is assigned one of 14 possible booklets. Each booklet consists of three components: a mathematics and a sciences section, followed by the student survey, all of which are separated by short breaks. For my estimations I will focus exclusively on the mathematics section of the exams. For clarity throughout the remainder of this paper, I use the term 'prompt' to refer to a unique query, while 'question' will refer to a prompt in a specific booklet and year. Each mathematics section consists of two blocks of prompts that permutate throughout the 14 booklets so that each block of prompts is featured in two different booklets. Prompt order within a prompt block does not vary. Among other goals, the TIMSS exams are designed to measure time trends in learning outcomes; therefore, eight blocks of mathematics questions get re-administered between 2011 and 2015. Thus, a unique prompt is either featured in two questions if in a non-readministered block or four questions if in a readministered block. Figure 1.2 illustrates the structure of the 14 TIMSS booklets that could be handed to a student in a given year. For example, a student handed booklet 1 would first complete their math section, which would consist of prompt blocks M01 and M02, and then move on to their science section after a short break. A student handed booklet 2 would complete their science section first and after a break complete their math section consisting of prompt blocks M02 and M03.

I do not observe the exact text of most of the prompts.[14] Information is available on each prompt, including some prompt characteristics such as the answer type (completed response

Dhabi, Australia, Austria, Azerbaijan, Botswana, Chinese Taipei, Croatia, Czech Republic, Dubai, Finland, Germany, Georgia, Honduras, Hong-Kong, Hungary, Ireland, Iran, Italy, Lithuania, Morocco, Malta, Northern Ireland, Norway, Oman, Poland, Portugal, Qatar, Romania, Russian Federation, Saudi Arabia, Singapore, Slovak Republic, Slovenia, Spain, Sweden, Quebec, United Arab Emirates.

[11]Occupational categories are more difficult to compare and interpret given the cross-country nature of this data. Nonetheless, results using highest family occupational category are broadly similar and reported in table A.2 of the appendix.

[12]Because the national median in Honduras is for parents to have primary or no education, I set this indicator to one for Honduran students who are at the national median in order to have a comparison group.

[13] My estimation exploits the fact that the random assignment of booklets to students is orthogonal to parental non-response, as demonstrated by the first column in table A.1. Columns two through four in table A.1 show that students whose parents do not complete the parental questionnaire perform worse than their peers, even when controlling for classroom fixed effects, and these students also do worse on their exam if it features a higher share of monetary questions.

[14]TIMSS readministers prompts across examination waves and thus does not releases the full set of prompts that were used.

or multiple choice),[15] topic area and cognitive domain and a brief thematic descriptor. I flag as monetary any question whose prompt or prompt descriptor contains terms such as 'money', 'buy', 'sell', 'cost', 'pay' or 'zeds' (the fictional currency used for this international exam). Furthermore, I also flag the four directly subsequent non-monetary questions as 'post' questions to track persistence of effects.

Pooling the 2015 and 2011 4th grade data gives me 28 different exam booklets. On average students face 25.32 different math questions, making each question worth approximately 4% of the math exam score, which I calculate as simply a student's mean performance on all of the mathematics question in their booklet.[16] Out of 708 questions, 44 are flagged as monetary questions and feature 14 unique prompts. Figure 1.2 shows that there is variation in the proportion of monetary questions featured in the booklets as well as variation in the reported parental education categories, the two key sources of variation I exploit in this dataset.

National School Panel: ENLACE

The Mexican Evaluación Nacional de Logros Académicos en Centros Escolares (ENLACE) exams were administered throughout the country each June from 2006 to 2013. While ENLACE started out as a low stakes test, ENLACE results were broadly diffused, becoming one of the main metrics for school performance and eventually being linked to teacher salary bonuses (Vivanco (2013), Hoyos (2014)). ENLACE was eventually discontinued because the growing performance incentives, combined with lack of implementation oversight, led to concerns about cheating.

School level subject results for all tested grades in all schools in Mexico are publicly available. The data also includes the school's marginalization index (1 to 5)[17] as defined by Mexico's National Population Council.[18] ENLACE examination booklets are also publicly available. Within each booklet, I tally the total number of mathematics prompts and the number featuring a monetary theme. Figure 1.3 show that there is variation in the proportion of monetary questions featured on exams within each grade. Figure 1.3 also shows

[15] TIMSS exams feature both multiple choice and completed response questions. Most of the questions only allow for a single correct answer, but occasionally multiple answers are considered correct and some questions allow for partially correct answers. For simplicity I do not count partially correct answers as correct.

[16] TIMSS exams are designed to measure the distribution of proficiency in a population rather than accurately measure the proficiency of a single individual, thus the exam mean differs from the official TIMSS achievement measure, which is generated using a complex parameterized imputation procedure.

[17] Although the marginalization index does not change over time for most schools, there is some year on year variation. I opt to treat this index as time invariant, calculating the average for each school and rounding to the closest index category.

[18] Mexico's National Population Council (CONAPO) calculates marginalization indices using a principal components method based on percentage indicators of social exclusion collected in the census. Indicators include illiteracy, incomplete primary education, lack of running water, sewage systems, and electricity, dirt floors, household overcrowding, geographic isolation, and low incomes in employment. Further details are available at http://www.conapo.gob.mx.

the variation in the marginalization indicator across schools, the other source of variation necessary for my estimation.

I use a panel of school performance for 135,307 different schools between the years 2009 and 2013.[19] While the ENLACE data has the disadvantage of not being at the question level, nor even at the individual level, the panel structure presents certain advantages. In particular, by incorporating additional data, it allows some insight into how estimated effects respond to income shocks.

Rainfall has been shown to generate income shocks in the Mexican context (Munshi (2003)). I obtain the coordinates for Mexican municipalities and match these to rainfall data.[20] I use data from the Tropical Rainfall Measuring Mission (TRMM) [21] to calculate a drought indicator for each examination year. The drought indicator is set to one if the cumulative rainfall in the previous agricultural season (July-February) falls in the lowest decile of a locality's rainfall realizations between 1998 and 2018.

1.3 Impacts of Financial Salience on Aggregate Performance

I begin by presenting student level estimations that look at how the variation in the proportion of monetary questions featured on an exam or assignment impacts effort and performance. I find that more financially salient exams differentially depress the exam scores of lower SES students and that this effect is responsive to income shocks. I also identify that lower SES students have to exert differentially greater learning effort when faced with more financially salient homework assignments.

Impacts in Examination Settings: TIMSS

For the estimation using the TIMSS data, I exploit the random assignment of test booklets to students and the variation in the number of monetary questions between booklets. Columns 1 through 4 of table 1.3 provide a randomization check, confirming that within a year there

[19] In many schools, examinations were administered in several sessions throughout the day. Performance data is reported for each session. I construct a single school level subject result for each grade by calculating a weighted average of the performance in the different sessions using the number of tested students as weights. Though some data is available for the earlier years, the number of examined students is not included in the 2006 and 2007 data. Furthermore, the data in 2008 does not disaggregate performance by subject. Analysis is thus focused on the years 2009-2013. Finally, in 2011 two different test booklets were used for the 3rd and 4th grades in certain regions. As the data does not indicate which booklet was used, these observations are also dropped from the final dataset.

[20] Municipality coordinates are available from the Instituto Nacional de Estadística y Geografía (INEGI).

[21] Specifically, I use the TRMM Multi-Satellite Precipitation Analysis (TMPA) Rainfall Estimate Product 3B43 Version 7, which merges satellite and gauge data to generate a monthly estimate on a 0.25° by 0.25° spatial resolution.

is no correlation between a student's SES categories and the share of monetary questions in the booklet they receive, overall and within a classroom.

The effect on a low SES student of receiving a financially salient booklet is estimated as follows,

$$E_{ib} = \Theta_1 + \Theta_2 LowP_i + \Theta_3 LowP_i * PM_b + \kappa_b + c_i + \epsilon_{bi}, \tag{1.1}$$

$$E_{ib} = \theta_1 + \sum_{p=2}^{5} \theta_{2p} P_i + \sum_{p=2}^{5} \theta_{3p} P_i * PM_b + \kappa_b + c_i + \epsilon_{bi}. \tag{1.2}$$

I regress standardized exam scores (E_{ib})[22] on the SES indicator and the interaction between the SES indicator and the proportion of monetary themed questions (PM_b) featured in the randomly assigned booklet (b). For the TIMSS estimations, I use an indicator for whether the reported parental education category is lower than the national median $(LowP_i)$ as observed in the TIMSS data in equation 1.1 or parental education category dummies(P_i) as specified in equation 1.2. I also include booklet fixed effects (κ_b) and country or class fixed effects (c_i) as controls.

Results are reported in table 1.3. Estimates in columns 5 and 7 imply that a 10 percentage point increase in the share of monetary questions featured on an exam differentially depresses the performance of students whose parental education falls below the national median, by 0.026 standard deviations. Note that on the TIMSS exams the proportion of monetary questions featured in a booklet ranges from 0 to 0.217. Columns 6 and 8 show that this effect is negatively related to parental education, with the largest effect for the most disadvantaged students. Columns 7 and 8 include classroom fixed effects. The addition of classroom fixed effects does not significantly change the magnitudes of the θ_3 coefficients of interest, although the overall variation in performance due to parental education levels as estimated by θ_2 is significantly smaller within a classroom than within a country. This is likely due to selection across schools.

The magnitude of the effect of monetary questions is not small. It is informative to compare this effect to the general performance gap between these students as measured by the θ_2 coefficients. The 0.026 standard deviation decrease resulting from a 10 percentage point increase in the proportion of monetary questions is equivalent to about 6% of the within country performance differential between students whose parental education is at or above the national median and those below. This increases to about 10% when considering the within classroom performance differential.

Impacts in Examination Settings: ENLACE

I apply a similar estimation approach to the ENLACE exam data,

[22] I use crude exam scores calculated as the mean performance on the questions in the question level data.

$$E_{sgy} = \Theta_1 + \Theta_2 LowZ_s * PM_{gy} + \kappa_{gy} + \tau_{sy} + \rho_{gs} + \epsilon_{sgy}, \tag{1.3}$$

$$E_{sgy} = \theta_1 + \sum_{z=2}^{5} \theta_{2z} Z_s * PM_{gy} + \kappa_{gy} + \tau_{sy} + \rho_{gs} + \epsilon_{sgy}. \tag{1.4}$$

I regress the standardized school average[23] for each grade and year (E_{gys}) on SES indicators interacted with the proportion of questions on that grade's exam that featured a monetary theme that year (PM_{gy}). Here, SES indicators include an indicator for whether a school's marginalization index falls below the national median ($LowZ_s$) in equation 1.3 or the school marginalization index dummies (Z_s) as specified in equation 1.4. I include a grade by year fixed effect (κ_{gy}) to control for overall difficulty of each particular exam booklet, school by year fixed effects (τ_{sy}) to control for local shocks that might affect overall performance in a school, and grade by school fixed effects (ρ_{gs}) to control for time invariant performance of a grade in a school.

Results are reported in table 1.4 and are qualitatively consistent with the results using the TIMSS data. Students in disadvantaged schools see their mathematics exam scores further depressed when more monetary questions are featured on the exam. As illustrated in figure 1.3, the percentage of monetary questions featured on an exam can vary by up to 18 percentage points within a grade level. These estimates suggest that a 10 percentage point increase in the share of monetary themed questions differentially reduces performance in below median schools by 0.038 standard deviations and up to 0.126 standard deviations in very disadvantaged schools. The overall performance gap between above and below median schools is 0.265 standard deviations. Thus the effect of a 10 percentage point increase in monetary salience represent about 14% of the overall performance gap.

Impacts in Examination Settings: Response to Income Shocks

In addition to confirming the TIMSS results, the ENLACE data has the advantage of being a school level panel, allowing me to observe how the estimated effects respond to fluctuations in income. In the context of Mexico, drought conditions have been shown to generate economically significant income variation (Munshi (2003)). To consider whether annual income variation impacts this effect, I add the relevant interaction terms with the drought indicator (D_{sy}),

$$E_{sgy} = \delta_1 + \delta_2 LowZ_s * PM_{gy} + \delta_3 LowZ_s * PM_{gy} * D_{sy} + \delta_4 PM_{gy} * D_{sy} + \kappa_{gy} + \tau_{sy} + \rho_{gs} + \epsilon_{sgy}. \tag{1.5}$$

Results are reported in table 1.5. Student performance fluctuations in response to income shocks are consistent with the hypothesis that income scarcity amplifies the negative effect of monetary questions on exam performance. δ_2, reported in the first row, shows that students in below median schools perform worse on exams that feature a higher percent of monetary

[23] The standardization of the school averages is weighted by the number of students who took the exam.

questions. δ_3, reported in the second row, shows that this negative effect on exam scores is amplified, such that it more than doubles in magnitude in below median schools during drought years. Thus, in a drought year, facing an exam with 10 percentage points more monetary questions differentially depresses the performance of students in below median schools by 0.077 standard deviations. As one might expect, δ_4 is small and insignificant, as droughts do not affect performance on monetary questions in above median schools.

Impacts on Learning: ASSISTments

Analysis of the TIMSS and ENLACE data presents evidence that monetary questions differentially depress the exam performance of lower SES students. These impacts are of concern because examination performance often determines educational and economic opportunities. Yet, while the estimated effects are non-negligible, the underlying performance gap that exists between high and low SES students is substantially larger.

In this section, I present evidence that monetary questions may also contribute to this underlying performance gap through their impact on learning. Open any elementary school level math textbook and you will invariably find monetary themed examples being used to teach mathematical concepts. Thus, the same mechanism that depresses exam performance may also affects learning and skill acquisition. Using the user data from the ASSISTments homework platform, I find that lower SES students must exert differentially more effort to complete an assignment when their assignment features a greater proportion of monetary themed questions.

When assigned a skill builder by their teachers, students must log in and answer randomly selected questions from the skill builder's question pool until they answer three correctly in a row, at which point the system registers that they have mastered the assignment. For each student, I calculate the proportion of monetary questions they faced on their assignment as well as the mean number of attempts and hints they requested per question and the total time spent on the assignment.

In the following estimations, I use the assignment (a) data for all students (i) with complete question level data to estimate the effect of the proportion of monetary questions on several different dependent variables (Y_{ia}),

$$Y_{ia} = \alpha_1 + \alpha_2 PFR_s * PM_{ia} + \alpha_3 PM_{ia} + c_{ac} + \epsilon_{ia}. \tag{1.6}$$

I am interested in α_2, the interaction between the proportion of students in the school receiving free or reduced price lunch (PFR_s) and the proportion of questions the student faced that features a monetary theme (PM_{ia}), controlling for the proportion of monetary themed questions, as these are on average easier, and an assignment by class fixed effect (c_{ac}), which captures the general performance of students in that class on the assignment.

I first run a conditional logit to estimate the likelihood of mastering the assignment. I also estimate the equation above as a linear probability model. Next, I restrict the data to students who master the assignment and are actively engaged throughout the assignment [24] and run the same estimation on other dependent variables that capture learning effort such as the total time spent on the assignment, the number of questions the student answered, the mean number of hints they requested per question and the mean number of attempts they made on each question.

Results are presented in table 1.6. The estimates of α_2 using the conditional logit and the linear probability model reported in columns 1 and 2 are not statistically significant. Though the coefficients move in the hypothesized direction, I cannot reject that the proportion of monetary questions a student faces on their assignment has the same effect on the likelihood of mastering the assignment for students in advantaged and disadvantaged schools. However, the remaining columns in table 1.6 do present evidence that in order to achieve mastery on the assignment, students in disadvantaged schools have to differentially exert more learning effort when faced with an assignment featuring a higher share of monetary questions.

The coefficients in the first and second row suggest that any learning benefit monetary questions have for students in more advantaged schools is smaller, and even a disadvantage, for students in lower income schools where more students receive free and reduced price lunches. The mean value of the percent of students in the school receiving free or reduced price lunch is 0.252 while the minimum value is 0.012 (close to $PFR_s = 0$) and the maximum is 0.946 (close to $PFR_s = 1$). When $PFR_s = 0$, a 10 percentage point increase in the proportion of monetary questions a student faces decreases the time spent on mastering the assignment by 7.05 seconds (-0.020 sd). For the mean school where 25% of students receive free or reduced price lunch, assignment time is decreased by 1.8 seconds. However, in a school where all students receive free or reduced price lunch, much like the most disadvantaged school in the data, $PFR_s = 1$ and assignment time would increase by 13.75 seconds (+0.04 sd) over a mean value of 330 seconds. A similar pattern holds for the other measures of learning effort. A 10 percentage point increase in the share of monetary questions faced by a student reduces the number of questions needed to complete the assignment by 0.063 (-0.023 sd) questions if $PFR_s = 0$, 0.032 at the mean and increases the number of questions to mastery by 0.063 (+0.023 sd) if $PFR_s = 1$. Similarly, the mean number of attempts made on a question is reduced by 0.01 (-0.02 sd) when $PFR_s = 0$, 0.005 at the mean and increases by 0.012 (+0.024 sd) when $PFR_s = 1$. Finally, a 10 percentage point increase in the share of monetary questions faced by a student reduces the mean number of hints requested by 0.016 (-0.021 sd) if $PFR_s = 0$, 0.006 at the mean and increases the mean number of requested hints by 0.024 (+0.0323 sd) if $PFR_s = 1$. Thus, when comparing two classmates completing the same assignment, the student who randomly faces a larger share of monetary questions does not need to exert as much effort to complete the assignment in the wealthier schools. By contrast, in the more disadvantaged schools, the advantage presented by a larger share of monetary questions is reduced and even a net disadvantage.

These estimates are consistent with the findings using the TIMSS and ENLACE data.

[24] I define active engagement as students whose time spent on each assigned problem falls between 5 seconds and 8.8 minutes (the 90th percentile).

The results on time suggest that monetary questions would impact lower income students' performance when placed under a time constraint, as is common in examination settings. These effects on learning would impact much of the learning process in early mathematics education. This suggests that, while these monetary themed questions may present certain pedagogical advantages, these advantages are not evenly distributed and are even a disadvantage for the most vulnerable students, creating an unrelenting drag on their learning process. Over the course of an education, the few seconds of extra effort on each 6 minute assignment would certainly add up into a non-negligible effort cost that would contribute to explaining the underlying performance gap in educational outcomes between low and high SES students.

1.4 Identifying Attention Capture

While the evidence presented so far does suggest a differential impact of monetary questions on exam performance and learning effort, it does not clearly identify the mechanisms or show evidence of an attention capture effect. Monetary questions may differ from other mathematics questions. They may be used to test different skills in which low SES students face a disadvantage. Or, the effects may be entirely driven by the fact that monetary questions are more difficult for low SES students as they may have fewer opportunities to engage in monetary transactions. In the following sections, I exploit the itemized question level responses of the ASSISTments and TIMSS datasets to present evidence of an attention capture mechanism. I begin with an analysis of the itemized ASSISTments data where I exploit the precise thematic content of skill builders and the large number of questions to conduct a matching exercise to investigate the possibility that effects are driven solely by monetary questions being used to test different skills. Next, I exploit the random ordering of questions to clearly identify attention capture effects by looking at the lagged effects of monetary questions on subsequent questions. Finally, I show evidence that these effects are generalizable, as similar results are observable in the itemized TIMSS data.

Controlling for Question Characteristics: Matching

Figure 1.5 shows how question level performance and effort metrics vary by the percent of students on free and reduced price lunch in a school based on whether a question is monetary themed or not. These plots suggest that at all levels of free and reduced price lunch shares, the monetary questions in the ASSISTments skill builders are easier for students. However as the estimates in section 1.3 indicate, this advantage is smaller for students in disadvantaged schools. In schools where few students receive free or reduced price lunch, students are more likely to answer monetary questions correctly and request fewer hints, make fewer attempts, and spend less time on these questions. For students in schools where most students receive free or reduced price lunches, the advantages presented by monetary questions are much smaller if not nonexistent.

Figure 1.5 is constructed using all of the questions in the ASSISTments data. An important concern may be that monetary questions are used to test a very different set of mathematical skills in which low SES students are disadvantaged. For instance, these questions may be more likely to test numerical operations rather than geometric reasoning. While this is undoubtedly the case in most settings, including in the ENLACE and TIMSS exams, it is worth noting that the ASSISTments skill builders are very narrow in thematic content, as teacher use them to practice very specific mathematical skills such as 'Writing a Linear Equation from a Situation', 'Finding the Whole from the Percent and Part in a Word Problem' or 'Percent Increases and Decreases'. Nevertheless, one may still be concerned that monetary questions require a different skill set. These questions may involve more reading than, for instance, algebraic formula problems. To address this concern, within each skill builder I match monetary themed questions to almost identical non-monetary questions. Questions are matched if they are formulated similarly and involve the application of the same mathematical process. Figure 1.6 shows two examples of matched monetary and non-monetary questions.

Figure 1.7 plots the performance metrics by the share of students receiving free or reduced price lunch and monetary theme for the matched sub-sample. Note that restricting the data to matched questions significantly reduces sample size from 133,997 to 33,295 question observations. Nonetheless, figure 1.7 shows that performance on these matched questions is very similar for students in the most advantaged schools. Monetary and non-monetary themed questions are about equally likely to be answered correctly and require about the same number of hints and attempts, though the monetary questions do appear to take a little longer. For students in the most disadvantaged schools, the differences are much more substantial. They are much less likely to answer the monetary questions correctly and require more hints and attempts, and differentially more time.

To more formally estimate the difference between matched monetary and non-monetary questions, I estimate the following,

$$Y_{iq} = \gamma_1 + \gamma_2 PFR_s * M_q + \gamma_3 M_q + ms_{qi} + \epsilon_{iq}. \tag{1.7}$$

I regress question level performance metrics (Y_{iq}) on the interaction between the proportion of students in the school receiving free or reduced price lunch (PFR_s) and an indicator for monetary themed questions (M_q) controlling for the monetary indicator. I include a school by matched question group fixed effect ms_{qi}. This fixed effect is important as it restricts my variation so that I am comparing student performance within a school on questions that are nearly identical except for their thematic content. Question level performance metrics include whether the student answered the questions correctly,[25] how many hints were requested, how many attempts were made, and the time spent on the question.

Results are reported in table 1.7 and reflect the pattern observed in figure 1.7. Although the results are somewhat under-powered in this small sample, estimates of the differential, γ_2, are significant at the 5% level for hint requests and at the 10% level for completion time.

[25] ASSISTments does allow for partial credit; however, most of the data is either a 0 or 1.

The coefficients on answering correctly and attempts are consistent with underperformance and increased effort on monetary questions, although they are not statistically significant at conventional levels. Estimates for γ_3 are insignificant except for completion time. Thus, when a school has a low free and reduced price lunch rate, students within the same school perform similarly on monetary questions (though they require 11.42 seconds more time) as compared to their performance on almost identical non-monetary themed questions. However in more disadvantaged schools, compared to their performance on almost identical non-monetary themed questions, students within the same school request more hints and require even more time. In a school where all students received free or reduced lunch, students experience an additional disadvantage as compared to students in wealthier schools, and request an additional 0.518 (+0.24 sd)hints, and spend an additional 23.16 seconds (0.26 sd) (in addition to the additional 11.42 seconds experienced in all schools) on a monetary question as compared to their schoolmates answering a matched non-monetary question. This is evidence that underperformance on monetary questions by low SES students cannot be fully explained by the possibility that monetary questions require a different set of mathematical or question answering skills beyond those implied by their topical content.

Evidence of Attention Capture

The evidence presented in table 1.7 shows that monetary themed questions present a greater challenge to students in disadvantaged schools. This evidence, however is insufficient to clearly identify attention capture. Students in lower income households may not have as many opportunities to apply mathematical skills to monetized situations. For instance, they may be less likely to receive an allowance with which they can make purchases or they may be less likely to be put in charge of making small purchases in a shop or market where they must collect change. This explanation could lead to underperformance on monetary questions and generate the pattern of results in table 1.7 and in the aggregate effects estimated in section 1.3. Attention capture cannot be disentangled from this possible explanation by looking only at performance on the monetized questions.

If non-negligible, the attention capture effect can be identified by looking at performance on subsequent questions. These questions are not monetary themed but are potentially affected by the attention capture effect generated by the preceding monetary question. The randomized ordering of questions in the ASISSTments data can be exploited to identify whether there is such a lagged performance effect on subsequent questions for low SES students, as random ordering alleviates any concerns that question placement might be based on question unobservables.

I leverage this randomized ordering to identify the attention capture effect by comparing the performance of students in the same school on a question when it is placed subsequent to a monetary question versus when it is placed after a matched non-monetary question. To avoid having to consider the effects of repeated exposures and selection as students complete their assignments, I limit my sample to questions that are positioned between the first and second matched question a student encounters and no more than 4 questions after the first matched question.[26] I estimate the following,

$$Y_{iq} = \beta_1 + \beta_2 PFR_s * Post_{iq} + \beta_3 Post_{iq} + m_{q_{pre}} + \nu_{qs} + \epsilon_{iq}. \tag{1.8}$$

I regress question level performance metrics (Y_{iq}) on the interaction between the proportion of students in the school receiving free or reduced price lunch (PFR_s) and an indicator for being placed subsequent to a monetary themed question ($Post_{iq}$). I include a fixed effect for the leading matched group of questions ($m_{q_{pre}}$) and a question by school fixed effect (ν_{qs}). These fixed effects allow me to compare the performance of students in the same school on the same question when it is placed after a monetary themed question or a very similar non-monetary question.

Results are reported in table 1.8 and are consistent with an attention capture effect. Compared to their peers answering the same question, students in a school where everyone receives free or reduced price lunch request 0.389 (+0.15 sd) more hints and spend an additional 27 seconds (+0.28 sd) on the question if it follows a monetary themed question rather than a similar non-monetary themed question. The β_2 coefficients on answering correctly and attempts are also consistent with an attention capture effect, though not statistically significant. This stands in sharp contrast to the effect of monetary questions on students in more advantaged schools as estimated with the β_3 coefficients. These students experience reduced effort and better performance on questions subsequent to monetary themed questions suggesting that these questions are particularly effective learning tools in the more advantaged schools.

Attention Capture versus Cognitive Fatigue

Because of the randomized question order in the ASSISTments data, the lagged effect of a monetary question on subsequent questions must be due to their positioning relative to a monetary question. There is a possible alternative mechanism to the attention capture explanation. If low SES students find monetary questions differentially difficult, this might affect their performance on subsequent questions if they are differentially fatigued when they face them. This explanation could have the same implications for exam and assignment performance as estimated above but the underlying explanatory mechanism would be subtly different.

To distinguish attention capture from fatigue effects, I create a measure of lagged difficulty that is adjusted for different SES groups. I divide the students by quartile based on the share of students in their school receiving free or reduced price lunch. I then calculate the mean time spent by students in each quartile on the preceding matched questions. This measure

[26] More formally, let S_{qi} be the position of question q in student i's sequence of questions. S_{M1i} and S_{M2i} are the positions of the first and second matched questions faced by student i. I subset the data to observations where $S_{M1i} < S_{qi} < S_{M2i}$ and $S_{qi} \leq (S_{M1i} + 4)$

of differential difficulty of the preceding question is added as a control to the estimation strategy used in equation 1.8.

Table 1.9 presents the results of this estimation. The positive and significant coefficient in the third row suggests that students do spend a little bit more time on questions that follow questions that were differentially more difficult for them. This effect is small, however, and controlling for it does not does not meaningfully alter the coefficients or significance levels of the β_2 coefficients of interest, supporting the hypothesis that an attention capture mechanism is driving these results.

Attention Capture versus Stalled Learning

If low SES students struggle with monetary themed questions, they may not benefit as much from practicing using these questions, compared to similar non-monetary questions. Thus low SES students answering questions that follow a monetary themed question have not received as much effective practice as their peers who answered a similar non-monetary question. This could plausibly generate results similar to those in table 1.8, where the depressed performance of low SES students after monetary questions is simply due to having had one less practice question than their schoolmates.

To consider this possibility, I add a sequence control to the estimation strategy used in equation 1.8. Results are presented in table 1.10. The coefficients in the third row show that, as a student proceeds through a homework assignment, the likelihood that they answer a question correctly increases with each problem and the effort they must expend on each problem is reduced, in that they require fewer hints, make fewer attempts and spend less time on each subsequent question. If low SES students receive no learning gains from the monetary questions, the sequence position of the subsequent questions is effectively reduced by one. This could result in β_2 coefficients in the first row that offset the sequence coefficients in the third row.

Inspection of the relative magnitudes of the estimates in the first row as compared to the estimates in the third row shows that the depressed performance of low SES students is substantially larger than the effect of having answered one less previous practice question. Thus, underperformance following monetary themed questions truly is underperformance and not simply lack of improvement.

This evidence supports the attention capture hypothesis rather than a stalled learning effect. This is further supported by the evidence presented in the next section. Indeed, the TIMSS examinations are not designed as learning tools and question topics and themes can drastically change from one question to the next. In this setting, we would not expect improvement from learning as students proceed through the examination, ruling out the stalled learning explanation for the evidence we will consider next.

Confirming Attention Capture in Cross-Country Examination Data

The randomized ordering used on ASSISTments skill builders had the distinct advantage of allowing the estimation of the effect on subsequent questions without the concern that placement after a monetary question may correlate with some question unobservable that differentially depresses low SES performance. Nevertheless, ASSISTments is a small platform used primarily in the United States. In the following sections, I examine the TIMSS exam data to consider whether this effect is generalizable to an examination setting and a cross-county dataset. I find results consistent with an attention capture effect in the TIMSS data as well.

Suggestive Evidence Using Question Aggregates

Before imposing structure on my estimation methods, a simple approach suggests that there is indeed a performance gap on monetary and subsequent questions in the TIMSS data as well. A simple regression on aggregate TIMSS question statistics suggests a pattern of underperformance on monetary and subsequent questions for lower SES students. For different groups of students, I estimate the following,

$$\bar{C}_{q,p<nm} = \Phi_1 + \Phi_2 M_q + \Phi_3 Post_q + \Phi_4 \bar{C}_{q,p>=nm} + \epsilon_{q,p<nm}, \tag{1.9}$$

$$\bar{C}_{q,p} = \phi_{1p} + \phi_{2p} M_q + \phi_{3p} Post_q + \phi_{4p} \bar{C}_{q,uni} + \epsilon_{q,p}. \tag{1.10}$$

For students whose parental education falls below the national median, I regress the mean performance on each question, as measured by correct answers, ($\bar{C}_{q,p<nm}$) on the monetary indicator (M_q), an indicator for non-monetary questions placed within four questions after a monetary question ($Post_q$), and the mean performance of students with parental education above the national median ($\bar{C}_{q,p>=nm}$) to control for question difficulty. I repeat the same procedure for each of the parental education categories, p, other than university graduates, using the mean performance of students with university educated parents ($\bar{C}_{q,uni}$) to control for question difficulty.

Results are reported in table 1.11. Φ_2 estimates are negative and the magnitude of the penalty increases for lower parental education. Estimates for Φ_3 follow the same pattern, consistent with the hypothesized attention capture effect.

Confirming Attention Capture in Cross-Country Exams

While the above results are indeed suggestive, I can exploit the itemized TIMSS student level micro data to compare an individual student's performance on monetary and subsequent questions to their performance on other mathematics questions to see whether a pattern consistent with an attention capture effect is also present in the TIMSS data.

To do so, I estimate the following,

$$C_{iq} = \Lambda_1 + \Lambda_2 LowP_i * M_q + \Lambda_3 LowP_i * Post_q + \mu_q + \eta_i + \epsilon_{qi} \quad (1.11)$$

$$C_{iq} = \lambda_1 + \sum_{p=2}^{5} \lambda_{2p} P_i * M_q + \sum_{p=2}^{5} \lambda_{3p} P_i * Post_q + \mu_q + \eta_i + \epsilon_{qi}. \quad (1.12)$$

I regress an indicator for a correct response (C_{qi}) on the interaction between an SES indicator (having a parental education level below the national median ($LowP_i$) or parental education category dummies(P_i)) and the monetary indicator (M_q) as well as the post-monetary indicator ($Post_q$) for the four questions directly subsequent a monetary question. All specifications use student (η_i) and question (μ_q) fixed effects to control for student and question unobservables. Note that because the sequence of TIMSS questions is fixed within a booklet, question fixed effects directly capture the effect of placement within a booklet.

Results are reported in table 1.12, columns 1 and 2. Students with above median educated parents (column 1) or with university educated parents (column 2) are the omitted categories. Both sets of Λ_2 and Λ_3 coefficients are of interest. As on the ASSISTments platform, lower SES students' performance is differentially depressed on monetary questions and the questions that follow them. Furthermore, the sets of λ_2 and λ_3 coefficients are negative and inversely related to parental education, indicating a larger effect for the most disadvantaged students.

Though the results in columns 1 and 2 are consistent with attention capture, it is important to clarify that, because question ordering on the TIMSS exam is not randomized across students, it is not possible to cleanly identify the attention capture effect. Indeed, many of the concerns that were mentioned in the previous sections also apply to the TIMSS data. Monetary questions may be used to test mathematical topics that are differentially difficult for low SES students. Disadvantaged students may have fewer opportunities to apply their math skills to monetized situations. Importantly, unlike in the ASSISTments data, because question order is not random, we should further be concerned that subsequent questions may also systematically cover different mathematical skills or topics in a way that is unobservable. Though we cannot use the TIMSS data as conclusively as the ASSISTments data to identify attention capture effects – because of the non-random ordering of questions – there are a number of approaches that can be applied to improve upon the estimation strategy outlined in equations 1.11 and 1.12 and to appease some of the concerns above.

I begin by comparing question observables by question type. Because many TIMSS questions are not released, I do not observe all of the actual questions. TIMSS does, however, disseminate some information about each prompt on the exam, including a topical descriptor,[27] the question type of the prompt, the topical area the prompt is designed to test and the cognitive domain exercised by the prompt. Comparisons of observable prompt characteristics across the question indicators of interest are illustrated in figure 1.8. Unsurprisingly, it is clear that monetary questions differ in topical content, in that they are never used to test

[27]Each prompt is labeled with a name that broadly describes the topic and theme of the prompt. Examples include 'Total number of people on a ship', 'Multiply 23 and 19', 'Cost of ice cream', or 'Stickers bought by Mr. Brown'.

geometry topics, which represent almost half of the other questions. The four questions that follow monetary questions, however, do not share the unifying monetary theme and have a distribution of question characteristics that is broadly similar to the other questions in the TIMSS booklets.

I use these question observables to augment the controls used in equation 1.11. I construct additional fixed effects designed to capture differential performance due to placement,[28] difficulty,[29] question type and topic.[30] Columns 3 and 4 of table 1.12 give estimates of equations 1.11 and 1.12 using these additional fixed effects. Adding the additional controls does not change the qualitative features of the estimates. There is a small impact on the magnitude of the estimated effect, which actually becomes more pronounced.

One may be concerned that the effect stems from differences in teaching patterns between more advantaged and disadvantaged classrooms. In columns 5 and 6, I add additional fixed effects controlling for classroom performance on the questions of interest. These additional controls do not change the qualitative features of the estimates, though the magnitude of the estimated effect is slightly reduced. Thus the individual SES indicator of parental education still has explanatory power for performance on these questions, even controlling for class performance.

My preferred specifications are those in columns 3 and 4. Aggregated as in table 1.3, these estimates suggest that each additional monetary question featured on an 25 question exam would depress the score of a student whose parents have an education level below the national median by 0.193 percentage points or about 0.008 standard deviations[31] compared to students whose parents have an education level above the national median. Since each question represents about 4% of the mathematics section, this is equivalent to a 0.02 standard deviation decrease for a 10 percentage point increase in the share of monetary questions, which is consistent with the estimates in column 5 of table 1.3. As these exams can feature up to five monetary questions in a single booklet, this could amount to an exam score impact of between 0.04 and 0.05 standard deviations for students given a monetary intensive booklet who have below median parental education, and up to a 0.063 standard deviation decrease if their parents have a primary education or less.

I explore different combinations of the above fixed effects in table A.3 of the appendix.

[28] I include parental education by sequence fixed effects, where sequence is a constructed categorical variable indicating whether a question is featured in the first five questions of the exam, second five and so forth.

[29] I include parental education by difficulty fixed effects, where difficulty is a constructed categorical variable that uses the mean performance on a question by students with university educated parents to categorize questions into 20 difficulty bins.

[30] Parental education by country by topic and parental education by country by question type address the possibility that certain education systems may differentially prepare students in different mathematical topics or use different testing methods.

[31] Since each question is worth approximately 4% of the exam score, using the estimates from column 5, I calculate the direct effect as $-1.206 * 0.04$ with an additional effect on four subsequent questions of $-0.906 * 0.04 * 4$ for a total of 0.193 percentage points or about 0.008 standard deviations as the standard deviation of the exam scores is 23.56.

Though the magnitude of the estimated effects is somewhat sensitive to the choice of fixed effects, the effect on both monetary and subsequent questions remains negative and statistically significant in all estimations. Table A.4 in the appendix investigates whether unanswered questions, which are coded as incorrect in the estimations above, could be driving the effects. Interestingly, low SES students seem to be slightly less likely to leave a monetary question, or subsequent questions, unanswered.

Monetary Questions as Events

It is possible to think of students proceeding though exams and encountering 'events' in the form of monetary themed questions and to graphically visualize these effects. As students in these different datasets often encounter multiple monetary questions, setting up the estimation as an event study is not entirely straightforward. By limiting the TIMSS data to the subset of booklets that only feature one monetary question or two consecutive monetary questions, so as to have clearly defined pre and post periods, an event study approach is possible. This subset of the data covers 19 of the 24 booklets that include monetary questions. Figure 1.9 plots the differential performance on questions based on question placement relative to the monetary question in the booklet.[32] The figure clearly illustrates the sharp drop in performance on monetary question for students whose parental education is below the national median, and the continued effect on subsequent questions as well.

Attention Capture versus Cognitive Fatigue

The estimates in table 1.12 and figure 1.9 are very much consistent with the attention capture hypothesis. Nevertheless, there remains the possibility that the estimates on subsequent questions reflect the effect of cognitive fatigue as discussed in the previous section. For each question, I calculate the share of students from each SES category that answered the question correctly and use this as an indicator for how difficult a question is for a student from a particular SES category. I generate four lags of this indicator to control for differential difficulty of the four questions leading up to a question. Results are reported in table 1.13. Controlling for the differential difficulty of leading questions in columns 3 and 4 seems to slightly reduce the magnitude of the estimates on subsequent questions by a small amount, though they remain negative and statistically significant, suggesting that this explanation cannot explain the entirety of the effect.

[32] Figure 1.9 plots estimates for coefficients π_{2t} from the following estimating equation:

$$C_{iq} = \pi_1 + \sum_{t=-6,t\neq-1}^{10} \pi_{2t}(T_q = t) * LowP_i + \mu_q + \eta_i + ... + \epsilon_{iq} \tag{1.13}$$

Where T_q is a question's position relative to the monetary question. Additional fixed effects include the same fixed effects used in my preferred specification in column 3 of table 1.12: Below median by difficulty, below median by sequence, below median by country by question type and below median by country by question topic. Standard errors are clustered at the student level. Questions more than 5 questions prior to or 9 questions subsequent to the monetary event are binned together and coded as -6 and 10.

As an alternative method to estimate whether the effect on subsequent questions is due to cognitive fatigue from the differential difficulty of preceding questions, I generate 1000 placebo estimates from the data. Instead of flagging the true monetary questions, I flag a random set of questions as monetary and the 4 questions following this random set as post questions.[33] I then estimate my preferred specification of equation 1.11 as estimated in column 3 of table 1.12.

The resulting pairs of $\hat{\Lambda}_3^{placebo}$ and $\hat{\Lambda}_2^{placebo}$ coefficients are plotted in figure 1.10. The scatter plot suggests that it is highly unlikely that the two coefficients would both be jointly negative and of such a large magnitude by random chance, confirming the results above. In addition to verifying the above results, looking at the correlation between the coefficient pairs can also help decompose the role of cognitive fatigue due to the differential difficulty of preceding questions in explaining the effect on subsequent questions. Suppose the differential difficulty of preceding questions generates differential cognitive fatigue and thus differential performance on subsequent questions. Under these conditions, if the randomly selected placebo monetary questions happen to be differentially difficult for the low SES students, then we would expect them to perform differentially worse on subsequent questions and vice versa. Thus we would expect the correlation between $\hat{\Lambda}_3^{placebo}$ and $\hat{\Lambda}_2^{placebo}$ to be positive. To investigate this, I estimate the following regression.

$$\hat{\Lambda}_{3p}^{placebo} = \psi_1 + \psi_2 \hat{\Lambda}_{2p}^{placebo} + \epsilon_p \tag{1.14}$$

Results are reported in table 1.14 and plotted in figure 1.10. ψ_2 is indeed positive and statistically significant, suggesting that cognitive fatigue due to the differential difficulty of preceding questions does explain part of the magnitude of the estimated effect on subsequent questions. Nonetheless, as visible in figure 1.10, the predicted value of the coefficient on subsequent questions using the estimated placebos ($\hat{\hat{\Lambda}}_3^{placebo}$) is significantly smaller in magnitude than the estimate using the actual monetary questions, $\hat{\Lambda}_3$. I can reject that cognitive fatigue due to the differential difficulty of preceding questions explains the entirety of the effect on subsequent questions, supporting the attention capture hypothesis. When decomposed, I estimate that cognitive fatigue due to the differential difficulty of the previous questions explains approximately 29% of the estimated effect on subsequent questions.[34] I interpret the remainder as evidence of attention capture.

1.5 Implications for High Stakes Exams

Performance differences on high stakes entrance exams can significantly affect access to secondary and higher education and thus to economic opportunities. If exam design on high

[33]Some question blocks are repeated across the two years. To ensure that the distribution is representative of the actual distribution of monetary questions, I make sure to randomly select 6 questions from the non-repeated blocks and 8 questions from the repeated blocks.

[34]Estimates in table 1.14 imply that $\mathbb{E}(\hat{\Lambda}_3^{placebo} | \hat{\Lambda}_2^{placebo} = -1.207) = -0.254$ or 29% of $\hat{\Lambda}_3 = -0.891$.

stakes examinations puts vulnerable students at a disadvantage, these tests could aggravate socio-economic disparities in access to education. As monetary questions are regularly featured on high stakes exams, the effects identified in this paper have the potential to significantly impact the educational opportunities of low SES students.

I cannot identify effects using a high stakes exam for two reasons. First, most high stakes entrance exams use only one examination booklet per examination wave, making it difficult to control for contemporaneous shocks that might differentially affect different socio-economic groups. Secondly, administered booklets and itemized question data are not generally publicly available. Nonetheless, using available information about high stakes tests, I am able to project my estimates onto high stakes exam scores and simulate the potential impact on access to further education.

The scholastic assessment test (SAT) is an important component of student applications to universities in the United States. A survey of current official practice exams suggests that monetary questions are regularly featured on the exam and can account for up to 20% of the questions on the quantitative section of the SAT.[35] Using the estimates in table 1.3 and official assessment statistics,[36] a 20 percentage point reduction in the share of monetary questions on the SAT could improve expected performance by students with below median parental education levels by 6 points (0.052 standard deviations). This represents about 7.2% of the quantitative section's performance gap between these groups.

In the US, SAT scores are generally only one of many components in a complex admission process. It is thus difficult to anticipate exactly how a change in score would affect access to higher education beyond the prediction that it would make access more equitable. Globally, though, there are many high stakes exams where scores are the sole determinant of eligibility for further education. In the following sections, I use my estimates from the TIMSS and ENLACE exams to generate counterfactual exam scores on Mexico City's high school entrance exam. I then perfectly replicate the placement algorithm used to allocate students to high schools across Mexico City. Finally, I use the counterfactual scores to simulate how the change in exam scores would affect student allocation.

Simulating Effects on Mexico City's High School Entrance Exam

In response to an inefficient high school enrollment process, a consortium of public schools in Mexico City known as the Comisión Metropolitana de Instituciones Públicas de Educación Media Superior (COMIPEMS) adopted a competitive centralized admissions process. All

[35]Ten official practice tests for the SAT were accessed on the college board website in September 2019. Monetary questions on these practice tests ranged from 8.6% to 20% of the questions on the quantitative portion of the exam with a median of 13.8%.

[36]The SAT's 2018 Annual Report shows a standard deviation of 114 points on the quantitative section. This report also shows performance by parental education categories that can be used to determine that the median level of parental education of test takers was a bachelor's degree. The mean quantitative score for students with below median levels of parental education was 495, while the median score for those with parental education at or above the median level was 578.

ninth graders wishing to attend one of these schools submit a ranked list of up to 20 high school programs and subsequently take a comprehensive standardized exam. After exams have been scored, students are ranked and assigned to schools according to a serial dictatorship mechanism (see Abdulkadiroglu and Sonmez (2003)).

The COMIPEMS exams consists of 128 multiple choice questions covering multiple subjects including mathematics, Spanish, history and the natural sciences and is administered to about 250,000 students each year. Though I do not observe the exam booklets, practice COMIPEMS mathematics questions do feature monetary themed questions. I use data from the 2004 and 2005 COMIPEMS entrance exam in which I observe student rankings of preferred high schools and performance on the COMIPEMS exam in the different subjects. I also observe parental education levels and current junior high school, which I match to the school marginalization levels reported in the ENLACE data.

I use two approaches to consider how a 10 percentage point decrease in the share of monetary questions would change the scores of students on the mathematics portion of the COMIPEMS test. The first uses the estimates derived from the ENLACE data in column 4 of table 1.4. These estimates suggest that a 10 percentage point decrease in the share of monetary questions should increase the mean score of students in very disadvantaged schools by 0.126 standard deviations. The math portion on the COMIPEMS had a standard deviation of 5.12 points in 2004 and 5.26 points in 2005. Because scores on the COMIPEMS use round numbers only, I use these values to generate a random binomial and add 1 point to the math COMIPEMS score of randomly selected students in very disadvantaged schools such that their aggregate performance on the mathematics section is improved in a manner consistent with the ENLACE estimate. I repeat the same procedure for students at each marginalization level using the relevant estimates.[37] Having simulated a new counterfactual mathematics score, I calculate their new counterfactual COMIPEMS score which I then use to generate a new counterfactual ranking of the students. The second approach is similar but uses parental education and employs the estimates from column 6 of table 1.3. The difference between actual and simulated counterfactual math scores and the mean difference between the actual and counterfactual rank for students in each SES category using these two different SES indicators is reported in table 1.15.

The effect of a 10 percentage point decrease in the share of monetary questions improves the performance of disadvantaged students on the mathematics portion of the exam. The gap in mean performance between the highest and lowest SES group is reduced by 3 to 20% on the mathematics section, depending on the SES indicator used.[38] Because the mathematics section covers only 20% of the exam, the effect on aggregate exam scores

[37]Exam scores are not adjusted for students with a missing SES indicator.

[38]Mexico City is relatively wealthy compared to much of Mexico and the vast majority of test takers are attending junior high schools that are considered very advantaged on the national scale.

is proportionally smaller.[39] Once ranked using the counterfactual scores,[40] the ranking of students in disadvantaged groups improves, at a cost to those in the more advantaged groups.

Mexico City's High School Placement Algorithm

Students in Mexico City are assigned to high schools according to a serial dictatorship mechanism based on their ranked exam performance and the list of preferred schools each student submits prior to taking the exam.

High schools first set the maximum number of students they will accept.[41] Students who fail to score above 30 or who fail to complete middle school are disqualified from attending high school. A computer program then proceeds through the ranked list of students, starting with the highest scoring student, and allocates each student to their top-ranked school with open seats remaining. If no seats remain at any of the schools listed by the student, the student is unassigned. After the first assignment process is complete, these students undergo a secondary selection process over several days that allocates unassigned students to the remaining open slots (Dustan, De Janvry, and Sadoulet (2017)).

Following these rules, I replicate the placement algorithm used by the COMIPEMS's centralized admission system. Because I observe the school to which each student was actually assigned, I can verify that high school placement in Mexico City actually follows the rules described above.[42] My replication of the placement algorithm perfectly replicates actual student assignment when I use the students' true exam scores. I can thus accurately simulate student placements using my replicated algorithm and the counterfactual ranking based on the counterfactual scores in table 1.15.

Simulated High School Placements in Mexico City

Since so few schools are considered disadvantaged in Mexico City, I focus here on the results using parental education categories. Tables 1.16 and 1.17 present estimated impacts for the

[39] I only simulate the effect of monetary questions on the mathematics portion of the exam as this paper has focused on mathematics, and all of the estimates are derived using mathematics questions. I elected to focus on mathematics questions because monetary questions are a common feature in mathematics instruction and the structure of many mathematics exams and assignments (multiple short, distinct questions) helps with identification. Nevertheless, though not identified in this paper, it is possible these effects may apply to other subjects.

[40] Ranking among students with identical exam scores is generated randomly.

[41] Many students receive the exact same COMIPEMS score. In the actual assignment process, once a school's available slots are filled, the school must elect to admit all or none of the students who receive the marginal score and would otherwise be assigned to that school based on the student's stated school preferences. Since I do not observe this rounding process, I cannot replicate it in the simulation. For competitive schools, where the lowest exam score of an admitted student was above 31, I use the number of students who were admitted into the school in each year as the maximum number admissible. I do not constrain the number of admissions for non-competitive schools.

[42] In addition to the matching conditions above, UNAM and IPN affiliated schools have an additional minimum GPA requirement.

two years of data. Equivalent results using school marginalization levels are presented in tables A.5 and A.6 of the appendix. Of primary interest is the change in the number of examined students who are ineligible to be assigned to a high school because they fail to meet the 31 point cutoff. For these students, the counterfactual of being eligible to go to high school, as opposed to being ineligible, has the potential to significantly alter the course of their lives. As illustrated in table 1.16, the simulation suggests that reducing the share of monetary themed questions on the exam by 10 percentage points would reduce the number of ineligible students by 2.2% for households with below median parental education levels.[43] The most impacted group comprises students whose parents are primary educated; they experience a 2.9% reduction in ineligible students. Overall, an additional 128 students pass this cutoff using the counterfactual exam scores. Having additional low SES students pass this threshold cutoff score is the most straightforward effect of the policy simulation and likely the most meaningful effect in terms of improving educational opportunities for low SES students.

In addition to the impact on high school eligibility, the counterfactual scores also change which students get assigned to a high school that they requested.[44] Because more students are able to meet the 31 point cutoff, the total number of assigned and unassigned students both increase, though there is significant heterogeneity across parental education groups. Unlike the requirement of meeting the 31 point cutoff, assignment to high demand high schools is a zero sum game where improved performance by lower SES students results in some displacement of higher SES students. Overall, students from higher SES groups are more likely to remain unassigned and less likely to receive an assignment in the simulation. By contrast, in the more disadvantaged groups, students are more likely to get assigned and less likely to remain unassigned.

Finally, the counterfactual scores also change whether students get to attend a more highly preferred school as summarized in table 1.17. Overall, students who received an assignment using both the real and counterfactual scores, on average get assigned to slightly less preferred schools. This is not surprising, as more students are passing the 31 point threshold, generating more competition and some displacement. Here again, heterogeneity is important. Students with highly educated parents experience the bulk of this negative effect. For the children of university educated parents, since their exam scores are unaffected, it is rare that they get assigned to a more preferred school in the simulated data while 111 are displaced into less preferred schools and 30 become unassigned. Conversely, students with the least educated parents are on net more likely to receive a preferred school assignment: 308 receive a preferred assignment while 180 receive a less preferred assignment. Furthermore, in this group 143 of previously unassigned students receive an assignment, while only 48 become unassigned. Note that I do not use any outside metric of school quality and rely solely on the preference ranking elicited from the students. This listing of school preferences

[43] The median parental education level in this data is Upper Secondary.

[44] Recall that not receiving an assignment means that the student did not score sufficiently high to be placed in any of the schools they listed on their application. In this event, students go through another secondary selection process that allocates unassigned students to the remaining open slots.

may be endogenous to student expectations about their performance, and these expectations would incorporate expectations regarding monetary questions. The effect of this endogeneity is not reflected in the simulation.

Overall, as the simulation shrinks the test score differential between higher and lower SES students, the allocation of educational opportunities becomes, predictably, more equitable. Note that the effects in this context are relatively small as I only generate counterfactual scores on the mathematics portion of the exam, which accounts for only 20% of a student's score. In contexts where mathematics is weighted more heavily, impacts could be substantially larger.

1.6 Discussion and Conclusion

Every year, millions of people around the world take examinations that have the potential to significantly impact their future economic outcomes. Performance on an exam may determine whether they receive a degree or get licensed, which school they can attend and even whether they are eligible to continue their schooling. Societies rely on examinations because they are a relatively efficient way of assessing and ranking a population by ability. The legitimacy of this approach, however, relies heavily on the perception of examinations as fair and objective, and a belief that the skills tested are good proxies for the skills assessors are actually interested in.

In this paper, I show that lower SES students perform differentially worse on mathematics exams that feature higher shares of monetary themed questions. This performance differential increases with socio-economic disadvantage and responds to negative income shocks. Furthermore, the evidence suggests that a similar pattern holds for homework assignments. Lower SES students must exert differentially greater effort and spend differentially more time completing homework assignments when they feature a larger share of monetary questions. As monetary questions are a common tool in the instruction of mathematics, this potentially affects much of the learning process in mathematics education. Investigation of question level response data shows evidence of depressed performance on monetary questions, even when compared to questions that are virtually identical. Furthermore, performance is depressed on subsequent questions as well, indicating an attention capture effect as posited in the psychology of poverty literature.

Should monetary themed questions be used in the teaching of mathematical concepts? It depends on what the ultimate teaching goals are. It is clear that lower SES students face a disadvantage when confronted with these topics; however, being able to apply mathematical concepts to monetary transactions is an important, even critical, skill. To the extent that equipping students with critical life skills is an important goal of early education, then one might argue that lower SES students may benefit from more practice using monetized examples to help overcome this disadvantage.

Should monetary themed questions be featured on mathematics exams? It depends on what the examination is supposed to be assessing. If assessing the ability to engage in monetary transactions is a primary goal of the examination, then it would be appropriate.[45] Most high stakes academic mathematics exams are designed to evaluate student preparation for more advanced mathematics studies. To the extent that more advanced mathematics studies do not necessarily center around monetary themes, opting for questions featuring non-monetary content would likely improve examination equity.[46]

Beyond the implications for educational testing, I present non-experimental evidence of attention capture due to poverty, and show that it affects a policy relevant outcome. This evidence that lower SES students underperform and make errors when distracted by a monetary theme has implications beyond the educational setting. Despite being temporary, this effect would impact financial choices made under conditions of scarcity, as it would mechanically be activated each time a disadvantaged individual must make a financial decision. These findings support the recommendations made by Mani et al. (2013) that policy makers be cautious of imposing cognitive taxes on the poor, with the additional caveat that this is particularly relevant for financially salient bureaucratic processes.

Policymakers may not be able to prevent this attention capture effect from creating a cognitive cost and inducing errors. However, minimizing the potential to make errors and the possible consequences of these errors is a conceivable avenue for policy intervention. Further research identifying cognitively demanding decisions and processes in which such errors are being committed is warranted. Similarly, educators cannot fully insulate low SES students from the disadvantage generated by the use of monetary examples without depriving them of an important life skill. Given this, it would be valuable to better understand how these effects might be shaping educational choices, aspirations and outcomes. Furthermore, adjusting assessment goals and strategies, by avoiding these monetary topics on high stakes exams where financial literacy is not explicitly being assessed, is a feasible and relatively simple policy. This could prevent these effects from limiting the long run educational opportunities of disadvantaged students.

[45] In its statement on testing fairness, the Educational Testing Service in the US frequently discusses the idea of 'construct-irrelevant variance', differences between test takers' scores that are caused by factors other than differences in the knowledge, skills, abilities, or traits the test is intended to measure (ETS (2014)). Thus, if a question or test is intentionally designed to test applications to monetary themes, differential performance would not be considered unfair.

[46] In fact, figure 1.3 shows that monetary themed questions become less common in higher grade examinations, likely because the subject becomes more abstract and conceptual.

Tables

Table 1.1: Dataset Features

	ASSISTments	TIMSS	ENLACE
Setting	Homework	Exam	Exam
Variation in Financial Salience	Yes	Yes	Yes
SES Indicator	School	Student	School
Panel			Yes
Itemized Question Data	Yes	Yes	
Question Matching	Yes		
Randomized Question Ordering	Yes		

Table 1.2: TIMSS Booklet Structure

	Part 1		**Part 2**	
	First Block	Second Block	First Block	Second Block
Booklet 1	M01	M02	S01	S02
Booklet 2	S02	S03	M02	M03
Booklet 3	M03	M04	S03	S04
Booklet 4	S04	S05	M04	M05
Booklet 5	M05	M06	S05	S06
Booklet 6	S06	S07	M06	M07
Booklet 7	M07	M08	S07	S08
Booklet 8	S08	S09	M08	M09
Booklet 9	M09	M10	S09	S10
Booklet 10	S10	S11	M10	M11
Booklet 11	M11	M12	S11	S12
Booklet 12	S12	S13	M12	M13
Booklet 13	M13	M14	S13	S14
Booklet 14	S14	S01	M14	M01

Note: A student handed booklet one would complete their math section first in part 1 and after a short break their science section in part 2. The math component of their exam would consist of prompt blocks M01 and M02. In contrast, a student handed booklet two would complete their science section first in part 1 followed by their math section in part 2. Their math section would consist of prompt blocks M02 and M03. Thus about half of the math prompts are identical between booklets 1 and 2.

Table 1.3: Financial Salience and Aggregate Performance in TIMSS

	Proportion Mon Q. in Booklet				Standardized Score			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Below Nat. Median	-0.000141		-0.000234		-0.438***		-0.268***	
	(0.000198)		(0.000240)		(0.00407)		(0.00390)	
Post Secondary		-0.000266		-0.000270		-0.315***		-0.209***
		(0.000262)		(0.000293)		(0.00523)		(0.00489)
Upper Secondary		-0.000315		-0.000367		-0.493***		-0.334***
		(0.000243)		(0.000291)		(0.00493)		(0.00471)
Lower Secondary		0.000185		0.0000258		-0.712***		-0.488***
		(0.000349)		(0.000421)		(0.00715)		(0.00686)
Primary or None		0.000178		0.0000918		-0.779***		-0.523***
		(0.000380)		(0.000504)		(0.00802)		(0.00784)
Below Nat. Median x Prop Mon Q.					-0.260***		-0.262***	
					(0.0471)		(0.0427)	
Post Sec x Prop Mon Q.						-0.0812		-0.0515
						(0.0609)		(0.0558)
Upper Sec x Prop Mon Q.						-0.0865		-0.0844
						(0.0569)		(0.0519)
Lower Sec x Prop Mon Q.						-0.160*		-0.147*
						(0.0825)		(0.0755)
Prim/No x Prop Mon Q.						-0.219**		-0.242***
						(0.0895)		(0.0817)
Constant	0.0617***	0.0618***	0.0618***	0.0618***	0.160***	0.314***	0.100***	0.213***
	(0.000117)	(0.000153)	(0.000128)	(0.000178)	(0.00165)	(0.00215)	(0.00153)	(0.00211)
FE: Year	Yes	Yes	Yes	Yes	.	.	.	.
FE: Booklet x Year	No	No	No	No	Yes	Yes	Yes	Yes
FE: Country	.	.	.	.	Yes	Yes	.	.
FE: Class	No	No	Yes	Yes	No	No	Yes	Yes
N	379468	379468	379160	379160	379468	379468	379160	379160

Note: Standard errors in parentheses. *$p < 0.10$,** $p < 0.05$, *** $p < 0.01$. Observations are at the student by examination level with a student level SES indicator: parental education. Omitted categories are students with parental education at or above the national median for columns 1, 3, 5 and 7 and university educated parents for columns 2, 4, 6 and 8. The proportion of monetary questions in a booklet is a value from 0 to 1.

Table 1.4: Financial Salience and Aggregate Performance in ENLACE

	Standardized Score			
	(1)	(2)	(3)	(4)
Below Median x Prop Mon Q.	-0.241***		-0.376***	
	(0.0333)		(0.0464)	
Advantaged x Prop Mon Q.		-0.160***		-0.0903**
		(0.0312)		(0.0408)
Middle x Prop Mon Q.		-0.296***		-0.220***
		(0.0448)		(0.0580)
Disadvantaged x Prop Mon Q.		-0.217***		-0.277***
		(0.0367)		(0.0507)
Very Disadvantaged x Prop Mon Q.		-0.837***		-1.256***
		(0.0823)		(0.118)
FE: Grade x Year	Yes	Yes	Yes	Yes
FE: Year x School	Yes	Yes	Yes	Yes
FE: School x Grade	No	No	Yes	Yes
N	1912259	1912259	1870964	1870964

Note: Standard errors in parentheses are clustered at the school level. Observations are weighted by the number of tested students. $^{*}p < 0.10$,** $p < 0.05$, *** $p < 0.01$. Observations are at the school by grade by year level with a school level SES indicator: the school's marginalization index. Omitted categories are schools at or above the median marginalization level for columns 1 and 3 and very advantaged schools for columns 2 and 4. The proportion of monetary questions in a booklet is a value from 0 to 1.

Table 1.5: Financial Salience and Rainfall Effects in ENLACE

	Standardized Score	
	(1)	(2)
Below Median x Prop Mon Q.	-0.376***	-0.315***
	(0.0601)	(0.0626)
Below Median x Prop Mon Q. x Drought		-0.453**
		(0.181)
Prop Mon Q. x Drought		0.0881
		(0.0824)
FE: Grade x Year	Yes	Yes
FE: Year x School	Yes	Yes
FE: School x Grade	Yes	Yes
N	1870964	1870964

Note: Standard errors in parentheses are clustered at the municipality level. Observations are weighted by the number of tested students. $^{*}p < 0.10$,** $p < 0.05$, *** $p < 0.01$. Observations are at the school by grade by year level with a school level SES indicator: the school's marginalization index. Omitted categories are schools at or above the median marginalization level. Drought is an indicator variable set to 1 if rainfall during the prior agricultural season (Jul- Feb) falls in the lowest decile of a locality's rainfall realizations between 1998-2018. The proportion of monetary questions in a booklet is a value from 0 to 1.

Table 1.6: Financial Salience and Aggregate Performance in ASSISTments

	Mastery		Effort Conditional on Mastered			
	Cond. Logit	Linear	Questions	Time (Sec)	Mean Hints	Mean Attempts
Prop. Free/Red. Lunch x Prop. Mon Q.	-0.607	-0.0318	1.256**	207.9***	0.396**	0.216**
	(0.466)	(0.0683)	(0.543)	(74.21)	(0.166)	(0.107)
Prop. Mon Q.	0.707***	0.0744***	-0.627***	-70.45***	-0.161***	-0.0905***
	(0.163)	(0.0196)	(0.156)	(20.72)	(0.0564)	(0.0254)
FE: Classroom x Assignment	Yes	Yes	Yes	Yes	Yes	Yes
Dependent Mean	0.757	0.787	4.669	329.5	0.386	2.339
Dependent SD	0.429	0.410	2.735	348.0	0.749	0.489
N	19043	22962	12125	12125	12125	10894

Note: Standard errors in parentheses clustered at the school level. $*p < 0.10$,** $p < 0.05$, *** $p < 0.01$. Observations are at the student by assignment level with a school level SES indicator: the share of students receiving free or reduced lunch. Mastery is a dummy variable set to 1 if a student has mastered the assignment by sequentially answering three questions in a row correctly. Questions is the number of questions a student answered on their assignment. Time is the amount of time in seconds spent on the assignment. Mean hints is the mean number of hints requested by the student on questions in the assignment and mean attempts is the mean number of attempts made by the student on questions in the assignment. Students who exceed more than 8 attempts on any single question are not included in the attempts estimation. Effort measures are estimated on the sub-sample of student assignments where the student masters the assignment and is actively engaged throughout the assignment (defined as the time spent on any question is such that 5 sec $<$ Time $<$ 8.8 min).

Table 1.7: Matched Monetary Questions in ASSISTments

	Correct	Hints	Attempts	Time (Sec)
	(1)	(2)	(3)	(4)
Prop. Free/Red. Lunch x Mon Q.	-0.0609	0.518**	0.153	23.16*
	(0.0471)	(0.239)	(0.162)	(12.18)
Monetary Question	0.0272	-0.0690	-0.0890	11.42**
	(0.0191)	(0.0983)	(0.0628)	(5.071)
FE: Matched Group by School	Yes	Yes	Yes	Yes
Dependent Mean	0.736	0.986	2.695	90.58
Dependent SD	0.428	2.121	1.434	88.78
N	29277	29277	28207	29277

Note: Standard errors in parentheses clustered at the school level. $^*p < 0.10$,** $p < 0.05$, *** $p < 0.01$. Observations are at the student by question level with a school level SES indicator: the share of students receiving free or reduced lunch. Observations are limited to monetary questions and questions that have been matched to a monetary question. Inactive observations are dropped (5 sec < Time < 8.8 min). Monetary question is a dummy variable set to 1 if the question features a monetary theme. The omitted category is non-monetary questions.

Table 1.8: Questions After Matched Monetary Questions in ASSISTments

	Correct	Hints	Attempts	Time (Sec)
	(1)	(2)	(3)	(4)
Prop. Free/Red. Lunch x 4 Post Matched Mon Q.	-0.106	0.791**	0.402	35.14**
	(0.0753)	(0.361)	(0.291)	(15.22)
4 Post Matched Mon Q.	0.0695***	-0.402***	-0.188	-8.223**
	(0.0255)	(0.101)	(0.114)	(3.685)
FE: Leading Matched Q. Type	Yes	Yes	Yes	Yes
FE: Question x School	Yes	Yes	Yes	Yes
Dependent Mean	0.651	1.433	2.891	101.1
Dependent SD	0.466	2.530	1.600	96.40
N	5409	5409	4963	5409

Note: Standard errors in parentheses clustered at the school level. $^{*}p < 0.10$,** $p < 0.05$, *** $p < 0.01$. Observations are at the student by question level with a school level SES indicator: the share of students receiving free or reduced lunch. Observations are limited to questions positioned between the first and second matched question a student faces and no more than 4 questions following the first matched question. Inactive observations are dropped (5 sec < Time < 8.8 min). 4 Post matched monetary question is a dummy variable set to 1 if a question follows a monetary themed question. The omitted category are questions that follow matched non-monetary themed questions.

Table 1.9: Questions After Matched Monetary Questions in ASSISTments with Controls for Preceding Differential Difficulty

	Correct	Hints	Attempts	Time (Sec)
	(1)	(2)	(3)	(4)
Prop. Free/Red. Lunch x 4 Post Matched Mon Q.	-0.110	0.817**	0.413	33.11**
	(0.0735)	(0.351)	(0.284)	(16.02)
4 Post Matched Mon Q.	0.0632**	-0.361***	-0.173	-11.34***
	(0.0269)	(0.0938)	(0.123)	(4.244)
Quartile Mean Time on Leading Matched Q.	0.000225	-0.00144	-0.000549	0.111***
	(0.000204)	(0.00165)	(0.000944)	(0.0396)
FE: Leading Matched Q. Group	Yes	Yes	Yes	Yes
FE: Question x School	Yes	Yes	Yes	Yes
Dependent Mean	0.651	1.433	2.891	101.1
Dependent SD	0.466	2.530	1.600	96.40
N	5409	5409	4963	5409

Note: Standard errors in parentheses clustered at the school level. $^{*}p < 0.10$,** $p < 0.05$, *** $p < 0.01$. Observations are at the student by question level with a school level SES indicator: the share of students receiving free or reduced lunch. Observations are limited to questions positioned between the first and second matched question a student faces and no more than 4 questions following the first matched question. Inactive observations are dropped (5 sec $<$ Time $<$ 8.8 min). 4 Post matched monetary question is a dummy variable set to 1 if a question follows a monetary themed question. The omitted category are questions that follow matched non-monetary themed questions.

Table 1.10: Questions After Matched Monetary Questions in ASSISTments with Sequence Controls

	Correct	Hints	Attempts	Time (Sec)
	(1)	(2)	(3)	(4)
Prop. Free/Red. Lunch x 4 Post Matched Mon Q.	-0.107	0.792**	0.406	35.27**
	(0.0753)	(0.359)	(0.290)	(16.73)
4 Post Matched Mon Q.	0.0692***	-0.401***	-0.187	-8.092**
	(0.0251)	(0.101)	(0.113)	(3.713)
Sequence Positon	0.0152**	-0.0471*	-0.0427**	-5.432***
	(0.00589)	(0.0242)	(0.0192)	(0.625)
FE: Leading Matched Q. Type	Yes	Yes	Yes	Yes
FE: Question x School	Yes	Yes	Yes	Yes
Dependent Mean	0.651	1.433	2.891	101.1
Dependent SD	0.466	2.530	1.600	96.40
N	5409	5409	4963	5409

Note: Standard errors in parentheses clustered at the school level. $^{*}p < 0.10, ^{**}\, p < 0.05, ^{***}\, p < 0.01$. Observations are at the student by question level with a school level SES indicator: the share of students receiving free or reduced lunch. Observations are limited to questions positioned between the first and second matched question a student faces and no more than 4 questions following the first matched question. Inactive observations are dropped (5 sec < Time < 8.8 min). 4 Post matched monetary question is a dummy variable set to 1 if a question follows a monetary themed question. The omitted category are questions that follow matched non-monetary themed questions.

Table 1.11: Regressions on Question Means in TIMSS

	Mean Performance (Correct=100) by Students with Parental Education Level:				
	Below Nat. Median	Primary/No	Lower Sec	Upper Sec	Post Sec
Monetary Question	-1.529*	-3.042*	-2.265*	-1.057	-0.511
	(0.908)	(1.749)	(1.293)	(0.717)	(0.429)
4 Post Question	-1.001*	-1.594	-1.045	-0.844	-0.425
	(0.576)	(1.215)	(0.903)	(0.666)	(0.376)
Q. Mean for Par. Edu. Above Nat. Median	0.939***				
	(0.0118)				
Question Mean for Univ. Parental Edu.		0.784***	0.942***	1.026***	1.036***
		(0.0227)	(0.0166)	(0.0112)	(0.00684)
N	706	706	706	706	706

Note: Standard errors in parentheses clustered at the prompt level. *$p < 0.10$,** $p < 0.05$, *** $p < 0.01$. Observations are at the question level with a student level SES indicator: parental education. Monetary Question is dummy variable set to 1 if a question is monetary themed. 4 Post is a dummy variable set to 1 if a question is non-monetary themed and positioned within 4 questions following a monetary question. Question means by parental education level are the mean of correct responses for each group where a correct response equals 100 and an incorrect response equals 0.

Table 1.12: Monetary and Subsequent Questions in TIMSS

	Question Answered Correctly (=100)					
	(1)	(2)	(3)	(4)	(5)	(6)
Below Nat. Median x Mon Q.	-0.885***		-1.207***		-0.753***	
	(0.123)		(0.131)		(0.151)	
Post Sec. x Mon Q.		-0.800***		-0.0641		-0.459**
		(0.165)		(0.177)		(0.188)
Upper Sec. x Mon Q.		-1.351***		-0.502***		-0.338*
		(0.152)		(0.163)		(0.185)
Lower Sec. x Mon Q.		-1.948***		-1.682***		-0.919***
		(0.213)		(0.229)		(0.262)
Primairy/No x Mon Q.		-1.786***		-2.548***		-1.514***
		(0.227)		(0.241)		(0.306)
Below Nat. Median x 4 Post	-0.680***		-0.891***		-0.395***	
	(0.0880)		(0.0969)		(0.109)	
Post Sec. x 4 Post		-0.614***		-0.397***		-0.482***
		(0.119)		(0.129)		(0.136)
Upper Sec. x 4 Post		-1.012***		-0.717***		-0.677***
		(0.109)		(0.120)		(0.134)
Lower Sec. x 4 Post		-0.926***		-0.998***		-0.653***
		(0.153)		(0.170)		(0.191)
Primairy/No x 4 Post		-0.925***		-1.213***		-0.534**
		(0.161)		(0.180)		(0.222)
FE: Student	Yes	Yes	Yes	Yes	Yes	Yes
FE: Question	Yes	Yes	Yes	Yes	Yes	Yes
FE: Below Med. x Diff.	No	.	Yes	.	Yes	.
FE: Below Med. x Seq.	No	.	Yes	.	Yes	.
FE: Below Med. x QType x Country	No	.	Yes	.	Yes	.
FE: Below Med. x QTopic x Country	No	.	Yes	.	Yes	.
FE: Par. Edu. x Diff.	.	No	.	Yes	.	Yes
FE: Par. Edu. x Seq.	.	No	.	Yes	.	Yes
FE: Par. Edu. x QType x Country	.	No	.	Yes	.	Yes
FE: Par. Edu. x QTopic x Country	.	No	.	Yes	.	Yes
FE: Class x Mon Q.	No	No	No	No	Yes	Yes
FE: Class x 4 Post	No	No	No	No	Yes	Yes
Dependent Variable Mean	49.56	49.56	49.56	49.56	49.56	49.56
Dependent Variable SD	23.56	23.56	23.56	23.56	23.56	23.56
N	9564201	9564201	9564201	9564201	9563918	9563918

Note: Standard errors in parentheses clustered at the student level. $^*p < 0.10$,** $p < 0.05$, *** $p < 0.01$. Observations are at the question by student level with a student level SES indicator: parental education. When a question is answered correctly the indicator is set to 100, 0 otherwise. Omitted categories are students with parental education at or above the national median for columns 1, 3 and 5 and university educated parents for columns 2, 4 and 6. Difficulty is a 20 bin binned indicator based on the performance on a question by students with university educated parents. Sequence is a 5 bin binned indicator based on the the position of a question within the exam booklet. Question type indicates whether a question is multiple choice or completed response. Question topic indicates categorized questions based on the topics listed in panel b of figure 1.8.

Table 1.13: Monetary and Subsequent Questions in TIMSS with Controls for Preceding Differential Difficulty

	Question Answered Correctly (=100)							
	(1)		(2)		(3)		(4)	
Below Nat. Median x Mon Q.	-1.108***	(0.151)			-1.202***	(0.153)		
Post Sec. x Mon Q.			0.190	(0.202)			0.199	(0.202)
Upper Sec. x Mon Q.			-0.450**	(0.187)			-0.461**	(0.187)
Lower Sec. x Mon Q.			-1.599***	(0.265)			-1.640***	(0.265)
Primairy/No x Mon Q.			-2.249***	(0.280)			-2.352***	(0.282)
Below Nat. Median x 4 Post	-1.005***	(0.102)			-0.878***	(0.104)		
Post Sec. x 4 Post			-0.393***	(0.136)			-0.370***	(0.136)
Upper Sec. x 4 Post			-0.793***	(0.126)			-0.754***	(0.127)
Lower Sec. x 4 Post			-1.002***	(0.179)			-0.904***	(0.180)
Primairy/No x 4 Post			-1.306***	(0.191)			-1.156***	(0.192)
Below Med. Performance on q-1					0.0612***	(0.0102)		
Below Med. Performance on q-2					0.0312***	(0.0107)		
Below Med. Performance on q-3					-0.0130	(0.00993)		
Below Med. Performance on q-4					-0.0498***	(0.0103)		
Par. Edu. Group Performance on q-1							0.0327***	(0.00644)
Par. Edu. Group Performance on q-2							-0.0160**	(0.00660)
Par. Edu. Group Performance on q-3							0.0359***	(0.00638)
Par. Edu. Group Performance on q-4							-0.0576***	(0.00662)
FE: Student	Yes		Yes		Yes		Yes	
FE: Question	Yes		Yes		Yes		Yes	
FE: Below Med. x Diff.	Yes		.		Yes		.	
FE: Below Med. x Seq.	Yes		.		Yes		.	
FE: Below Med. x QType x Country	Yes		.		Yes		.	
FE: Below Med. x QTopic x Country	Yes		.		Yes		.	
FE: Par. Edu. x Diff.	.		Yes		.		Yes	
FE: Par. Edu. x Seq.	.		Yes		.		Yes	
FE: Par. Edu. x QType x Country	.		Yes		.		Yes	
FE: Par. Edu. x QTopic x Country	.		Yes		.		Yes	
Exam Mean	49.56		49.56		49.56		49.56	
Exam SD	23.56		23.56		23.56		23.56	
N	8046329		8046329		8046329		8046329	

Note: Standard errors in parentheses clustered at the student level. $^{*}p < 0.10$,** $p < 0.05$, *** $p < 0.01$. Observations are at the question by student level with a student level SES indicator: parental education. When a question is answered correctly the indicator is set to 100, 0 otherwise. Omitted categories are students with parental education at or above the national median for columns 1, 3 and 5 and university educated parents for columns 2, 4 and 6. Difficulty is a 20 bin binned indicator based on the performance on a question by students with university educated parents. Sequence is a 5 bin binned indicator based on the the position of a question within the exam booklet. Question type indicates whether a question is multiple choice or completed response. Question topic indicates categorized questions based on the topics listed in panel b of figure 1.8. Sample mechanically does not include the first for questions on an exam for which the differential difficulty controls are undefined.

Table 1.15: COMIPEMS Simulation Summary Statistics

SES Indicator	Group	Observations	Math Mean Actual	Math Mean Simulated	Mean Change in Rank
School Indicator	Missing	16,437	13.10	13.10	−39
	Very Advantaged	391,249	14.15	14.15	−38
	Advantaged	90,112	13.56	13.61	144
	Middle	3,445	13.51	13.63	424
	Disdvantaged	1,666	12.61	12.75	537
	Very Disadvantaged	34	10.76	11.44	2,291
Parental Education	Missing	81,164	13.20	13.20	−197
	University	62,592	16.19	16.19	−167
	Upper Secondary	130,618	14.73	14.78	−13
	Lower Secondary	137,216	13.51	13.58	88
	Primary or Less	91,353	12.89	12.99	175

Table 1.14: Regressions on Placebo Coefficients

	Placebo Post Estimates ($\hat{\Lambda}_3^{placebo}$)
Placebo Mon. Estimates ($\hat{\Lambda}_2^{placebo}$)	0.142***
	(0.0202)
N	1000

Note: *$p < 0.10$,** $p < 0.05$, *** $p < 0.01$. $\hat{\Lambda}_2^{placebo}$ are the estimates for Λ_2 from equation 1.11 when randomly selected questions are flagged as placebo monetary questions. $\hat{\Lambda}_3^{placebo}$ are the estimates for Λ_3 on the corresponding placebo subsequent questions.

Table 1.16: COMIPEMS Simulation using Parental Education

		Ineligible		Eligible			
		Score under 31 points		Not Assigned		Assigned	
Total	Actual	8,373	1.67%	84,513	16.80%	410,057	81.53%
	Simulated	8,245	1.64%	84,584	16.82%	410,114	81.54%
Missing	Actual	1,756	2.16%	14,878	18.33%	64,530	79.51%
	Simulated	1,756	2.16%	14,937	18.40%	64,471	79.43%
University	Actual	349	0.56%	10,565	16.88%	51,678	82.56%
	Simulated	349	0.56%	10,592	16.92%	51,651	82.52%
Upper Secondary	Actual	1,334	1.02%	22,590	17.30%	106,694	81.68%
	Simulated	1,315	1.01%	22,594	17.30%	106,709	81.70%
Lower Secondary	Actual	2,637	1.92%	22,717	16.56%	111,862	81.52%
	Simulated	2,594	1.89%	22,727	16.56%	111,895	81.55%
Primary or Less	Actual	2,297	2.51%	13,763	15.07%	75,293	82.42%
	Simulated	2,231	2.44%	13,734	15.03%	75,388	82.52%

Table 1.17: COMIPEMS Simulation using Parental Education: Movement Detail

	Remain Ineligible or Unassigned	Become Assigned	More Prefered Assignment	Unchanged Assignment	Less Prefered Assignment	Become Unassigned	Change in Mean Preference Rank*
Total	92,510	376	987	407,732	1,019	319	-.00028
	18.39%	0.08%	0.20%	81.07%	0.20%	0.06%	
Missing	16,625	9	15	64,282	165	68	-.00453
	20.48%	0.01%	0.02%	79.20%	0.20%	0.08%	
University	10,911	3	3	51,534	111	30	-.00445
	17.43%	0.00%	0.00%	82.33%	0.18%	0.05%	
Upper Secondary	23,834	90	242	106,102	275	75	-.00082
	18.25%	0.07%	0.19%	81.23%	0.21%	0.06%	
Lower Secondary	25,223	131	419	111,057	288	98	.00234
	18.38%	0.10%	0.31%	80.94%	0.21%	0.07%	
Primary or Less	15,917	143	308	74,757	180	48	.00308
	17.42%	0.16%	0.34%	81.83%	0.20%	0.05%	

Note: For students who are assigned in both the actual and simulated data.

Figures

Figure 1.1: ASSISTments Variation

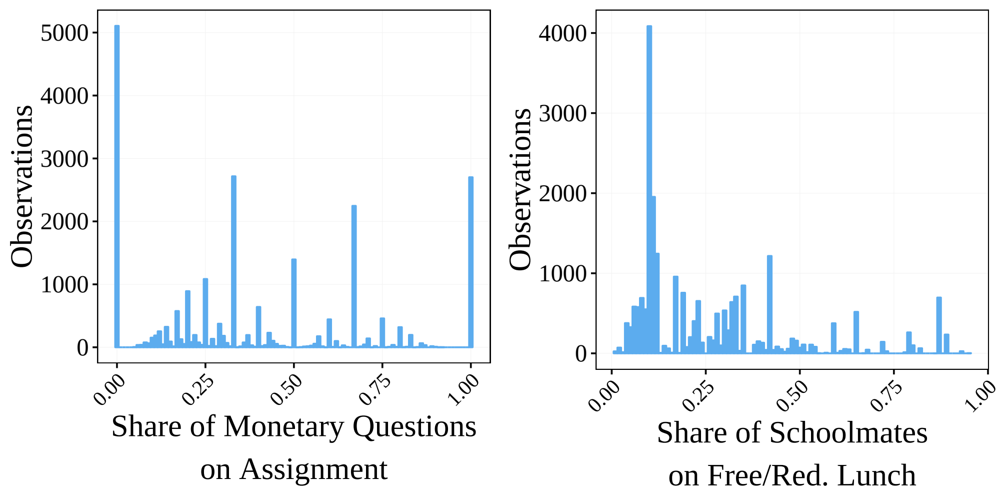


Figure 1.2: TIMSS Variation

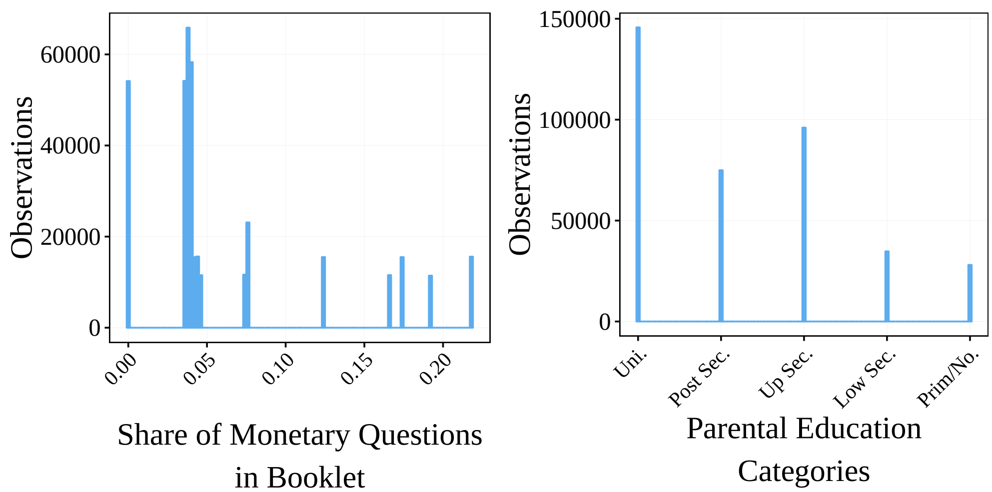


Figure 1.3: ENLACE Variation

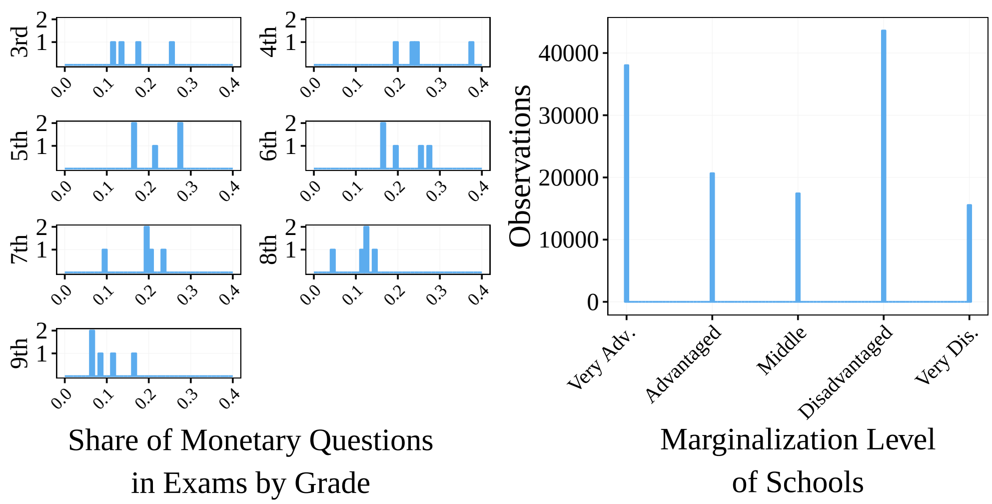
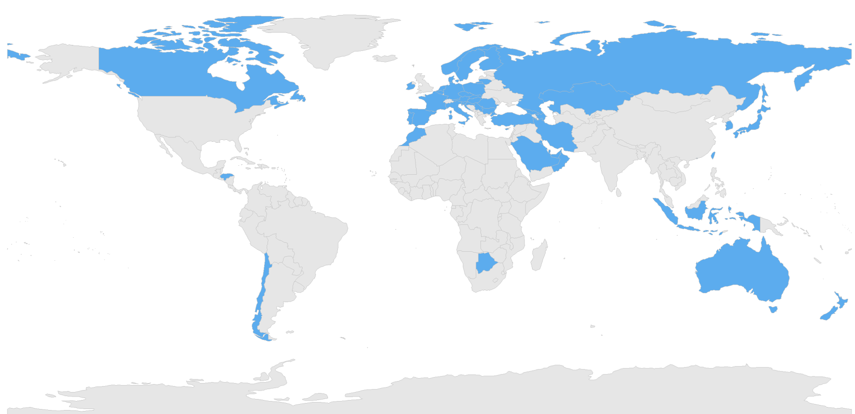


Figure 1.4: Countries participating in 4th grade TIMSS



Note: Mapped countries only show countries participating in the 4th grade TIMSS in 2011 and 2015 in which parental questionnaires were administerd.

Figure 1.5: ASSISTments Question Statistics by Question Type

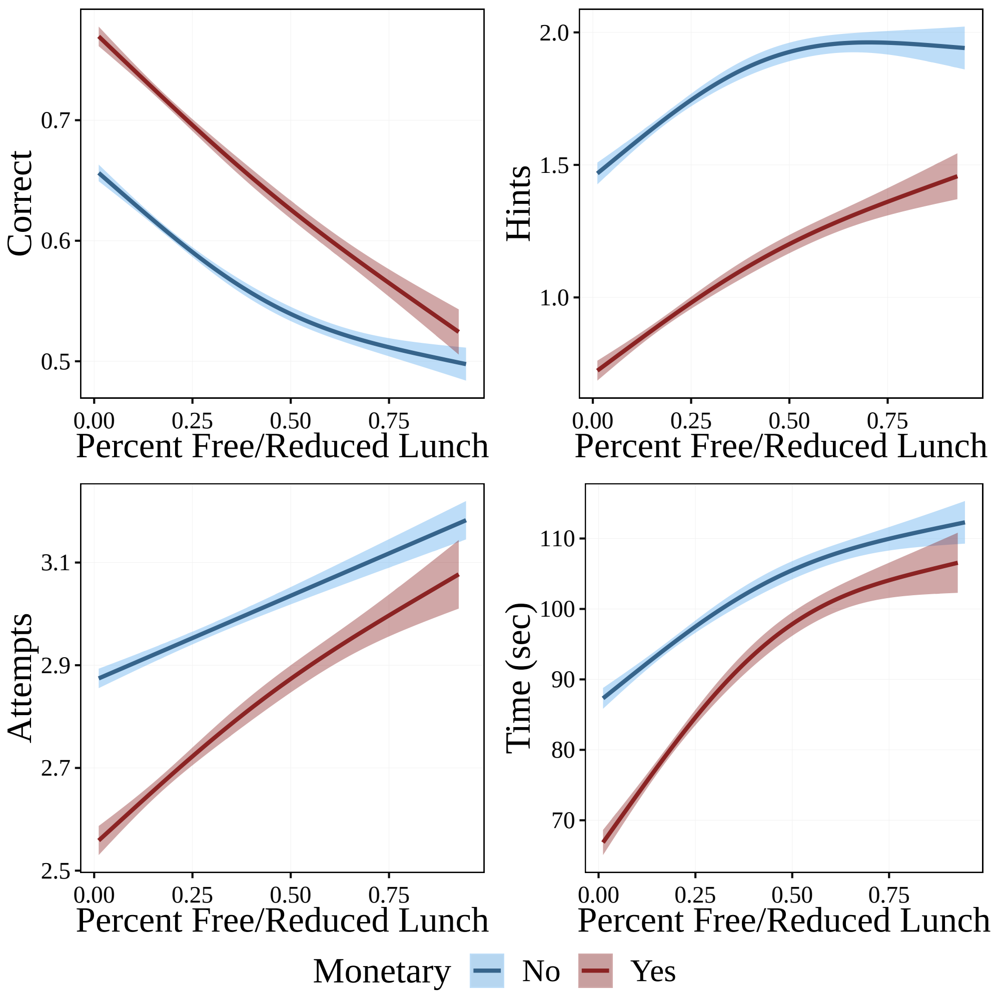


Figure 1.6: Examples of Matched ASSISTments Questions

Kate went shopping with $72 in her pocket, but she didn't want to spend it all. She decided to spend 25% of her money at most, and save the rest for later. How much was Kate willing to spend?

David has 840 cookies. He decides to give 96% of them to a friend as a birthday present. How many cookies does David give away?

A charity is performing a fund raising campaign, below are the amounts of money raised each week:

$683, $1357, $352, $1946, $301, $1577

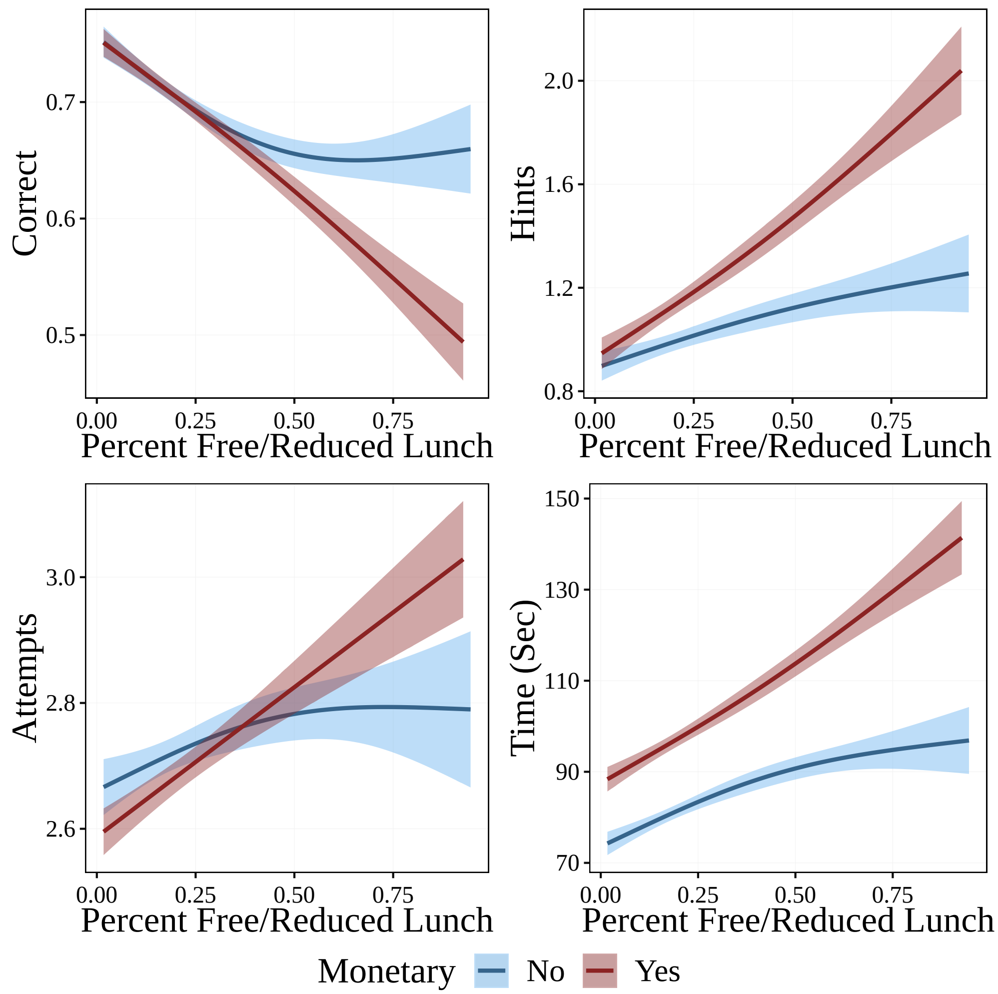
Calculate the **mean** dollar amount of money raised per week (round to nearest dollar).

Below are the number of spam emails filtered each week over the past few weeks by a school email system:

1073, 538, 964, 514, 273, 340

What is the **mean** number of spam emails filtered per week?

Figure 1.7: ASSISTments Matched Question Statistics by Question Type



Note: Data is limited to questions that are matched to a monetary themed question.

Figure 1.8: Question Characteristics by Category

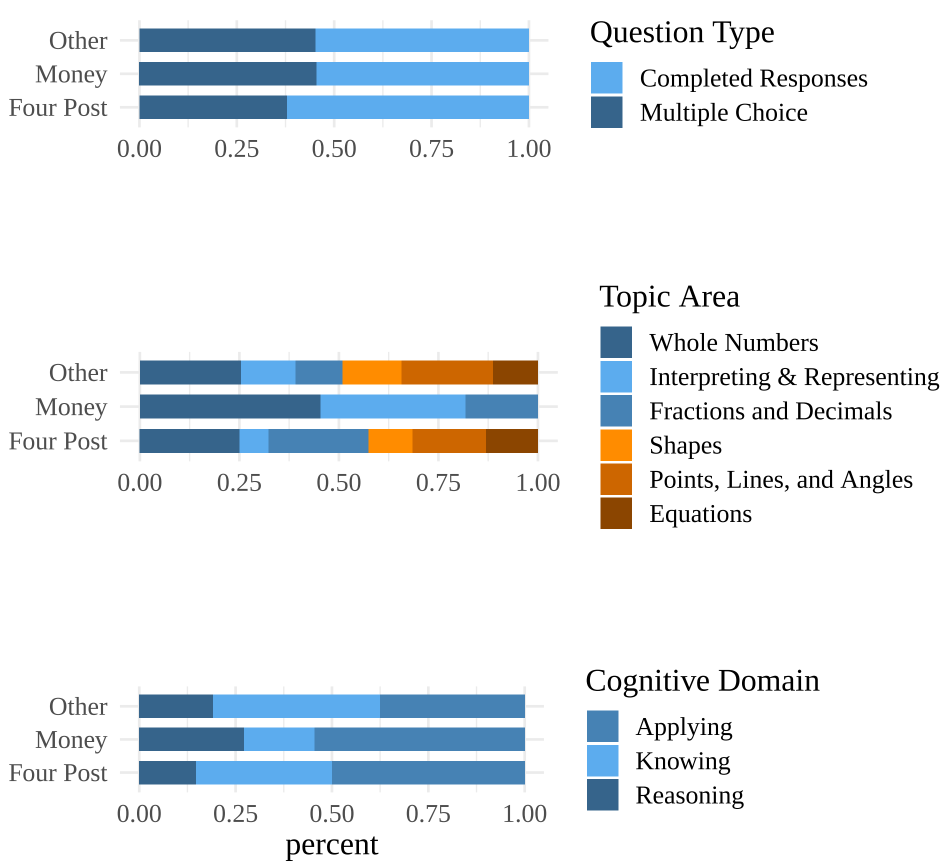
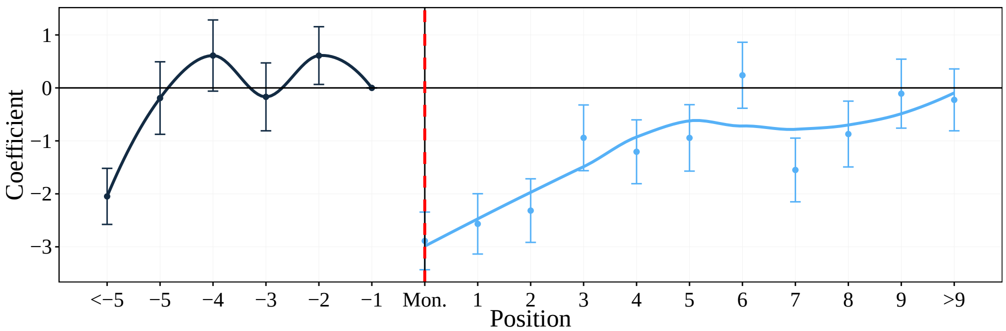
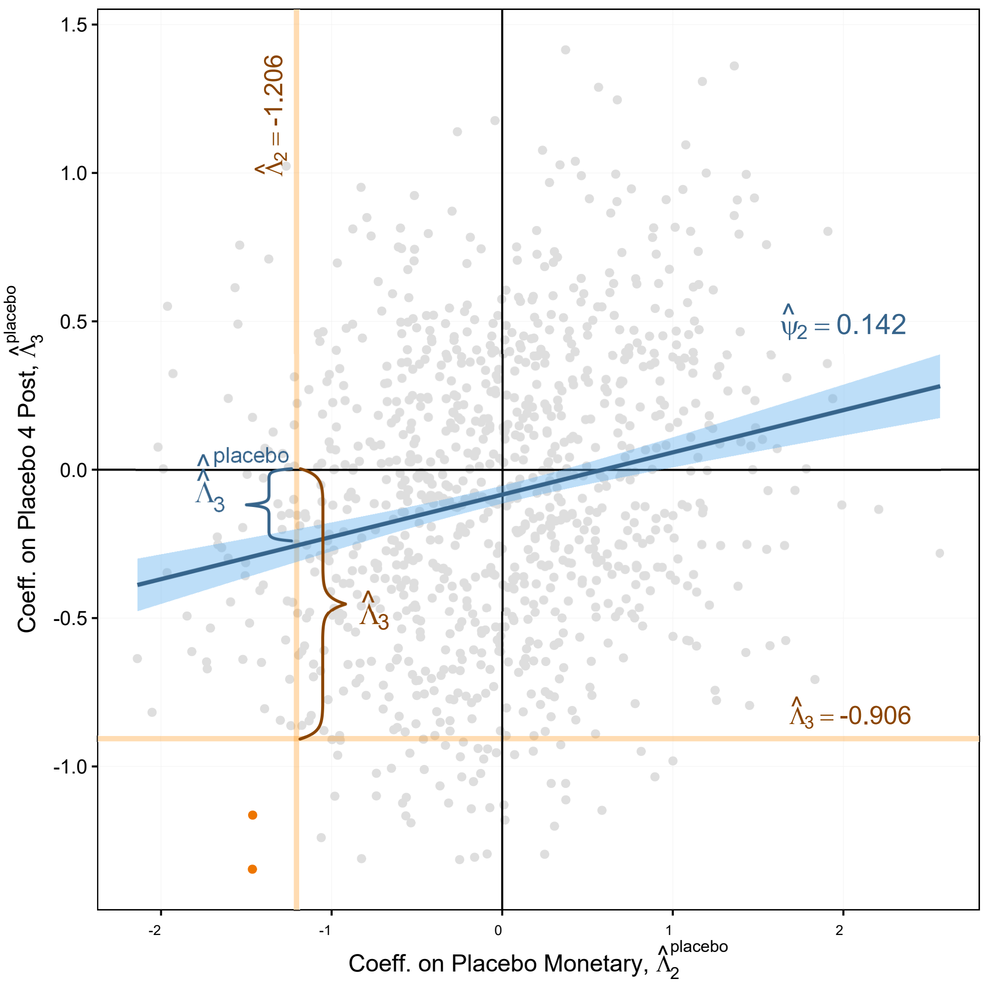


Figure 1.9: Differential Performance by Position Relative to Monetary Event for Below National Median Students in TIMSS



Note: Data is limited to booklets that feature a single monetary question or booklets that feature only two sequential monetary questions. Estimating equation includes Student, Question, Below Med. x Diff., Below Med. x Seq., Below Med. x QType x Country, Below Med. x QTopic x Country fixed effects.

Figure 1.10: Estimates from 1000 Placebo Estimations



Chapter 2

Labor Calendars and Rural Poverty: A case study for Malawi

The persistence of rural poverty in Sub-Saharan Africa is a major challenge to meet the Sustainable Development Goal on poverty eradication. Using data for Malawi, we investigate the contribution of seasonality to this phenomenon by showing that labor calendars for rural households offer similar employment opportunities as for urban households in terms of time worked at peak planting time, but much lower opportunities throughout the rest of the year. Due to a high level of urban unemployment, an urban-based structural transformation is not the current solution to rural poverty. By contrast, we show that elements of both an agricultural and a rural transformation can help fill-in and smooth-out labor calendars, providing a pathway to rural poverty reduction.

2.1 Introduction

Rural poverty is the most prevalent form of extreme poverty at a world scale, and it is increasingly concentrated in Sub-Saharan Africa and among households dependent on agriculture. In 2013, Sub-Saharan Africa accounted for 51% of the total number of poor at a PPP \$1.90/day poverty line, up from 18% in 1993.[1] In 2013, 82% of the Sub-Saharan Africa poor lived in rural areas, and 75% of Sub-Saharan Africa rural households income was obtained in agriculture (Beegle, Christiaensen, et al. (2016)). With eradication of extreme poverty the number one Sustainable Development Goal, there is much interest in understanding why poverty in Sub-Saharan Africa is associated with agriculture and rural areas and in exploring what can be done with agriculture and the rural economy in reducing poverty.

Labor is the main asset of the poor in generating income. The conversion of the poor's labor endowments into income depends on both the productivity of labor when people work and on the amount of time they work. Previous research has focused on differences in la-

[1] Using data from PovcalNet http://iresearch.worldbank.org/PovcalNet.

bor productivity on an annual basis between the agricultural and non-agricultural sectors (McMillan and Rodrik (2011) and Gollin, Lagakos, and Waugh (2013)). When labor productivity has been measured on a per hour worked basis, the measurement has been done contrasting agriculture and non-agriculture by McCullough (2017). Poverty is however not defined by sector, but determined by the portfolio of activities that households develop in the rural and urban sectors. In this paper, we make two advances relative to this literature. First, we focus on rural vs. urban households, and second we assess productivity both in terms of returns to labor when people work and in terms of time worked. Our attention is consequently drawn to the role of labor calendars in understanding poverty based on the observation that there is a strong correlation between poverty, as measured by per capita consumption, and the use of household labor across months of the year. While the time worked each year relative to capacity depends on overall under-employment levels in both rural and urban areas, the seasonality of rural labor calendars further reduces labor hours in rural areas, greatly aggravating household under-employment for the vast majority of months. Recognizing seasonal under-employment as a major contributor to low consumption levels opens up new pathways for rural development focused on transforming and diversifying agricultural and rural economies.

There has been a long standing controversy as to whether rural poverty will be reduced through the structural transformation of the economy–shifting labor out of agriculture into the urban economy–or through agricultural and rural transformations–increasing labor productivity and intensifying labor use in agriculture and in the rural non-farm economy where the rural poor reside. The debate started with Lewis (1954) who argued that labor incomes will only rise with resorption of surplus labor in agriculture as employment is created through capital accumulation in the urban industrial sector. Gollin, Lagakos, and Waugh (2013) showed that annual value added per worker is much higher in the non-agricultural sector than in agriculture, with a productivity gap typically in excess of four in developing countries and reaching six in the African countries. Based on these results, labor is considered to be greatly misallocated in most developing countries, with too much labor in agriculture and a large agricultural productivity deficit. This led many authors to argue that structural transformation would be an important source of aggregate output growth and an instrument for poverty reduction. Collier (2008), Collier and Dercon (2014), and Dercon and Gollin (2014) have endorsed this agro-pessimist viewpoint, whereby agriculture offers limited promise as a source of employment growth and poverty reduction compared to urban-industrial growth.

Using the LSMS-ISA data for four African economies, McCullough (2017) makes the important distinction between annual per capita labor productivity (used by Gollin, Lagakos, and Waugh (2013)) and labor productivity per hour worked in comparing agriculture and non-agriculture. For Malawi, she finds that labor productivity per person per year is 4.8 times higher in non-agriculture than in agriculture, confirming the Gollin, Lagakos, and Waugh (2013) observation.[2] By contrast, labor productivity per hour worked is only 40% higher in

[2]This is computed as the total revenue generated in each sector–net returns to farming and livestock

non-agriculture than in agriculture. The discrepancy between the two ratios is due to the high seasonality of agricultural labor calendars, with a few months of high employment at peak time (planting in this case) and a large number of months with high hidden unemployment. On an annual basis, the ratio of the two productivity measures says that the number of hours worked in a sector divided by the number of persons that declare their main activity to be in that sector is 3.3 times higher in non-agriculture than in agriculture. This suggests that lack of opportunities for year-round labor use in agriculture is a more important contributor to differentially higher poverty than lower labor productivity when working.

The policy implication of this observation is stark. As opposed to the ineluctable need for a structural transformation to secure growth and eradicate extreme poverty, other transformations become a possibility. One is an agricultural transformation following which farming systems become more diversified, correspondingly diversifying sources of income in agriculture and smoothing-out labor calendars throughout the year. The other is a rural transformation following which a rural non-farm economy largely linked to agriculture emerges locally, allowing rural households to diversify their sources of income outside of agriculture and further smooth-out labor calendars. This approach focusing on agriculture and rural areas has been recently advocated by IFAD (2016) and by Beegle and Christiaensen (2019). McMillan, Rodrik, and Verduzco-Gallo (2014) observed that structural change in countries like Malawi has been growth reducing as it shifted labor from low productivity agriculture to even lower productivity urban informality. As a consequence, focusing on productivity growth in agriculture and rural areas where the poor live is an appealing option for as long as extensive urban unemployment prevails.

A key question for implementation of an agricultural and rural transformation is to determine whether there is full employment at peak time in rural areas and how much under-employment there is in other months. Full employment at peak time would put emphasis on the importance of labor productivity gains in agriculture (such as through mechanization) to increase value added in the peak labor-constrained tasks (principally planting). Large under-employment in other periods of the year stresses the need to pursue agricultural and rural transformations to smooth labor calendars in agriculture and the rural non-farm economy.

In this article, we use the 2004, 2010, and 2016 LSMS data for Malawi, with most of the analysis done with the 2010 data which are of higher quality for our purpose. Data were collected monthly over a 13-month period, allowing the measure of seasonality in labor use. We analyze the use of rural and urban labor, estimating unemployment in high and low seasons of labor demand and throughout the year. We find that unemployment affects both urban and rural areas, throughout the year. Unemployment in Malawi is thus an overall problem, limiting the potential gains for rural poverty reduction coming from seasonal or permanent rural-urban migration as advocated for example by Lagakos, Mobarak, and Waugh (2018) for Bangladesh. In analyzing the case of Malawi, Evidence Action (2014) thus concluded that, "there are insufficient potential migration destinations to absorb excess

or non-farm enterprises and wage income by sector of employment– divided by the number of household members sorted by their primary sector of occupation labor from rural areas."

We decompose total unemployment in rural areas between what we call peak unemployment (the high-season unemployment level extended throughout the year) and seasonal unemployment (the additional unemployment in other months of the year). We find that seasonal unemployment amounts to 2/3 of total rural unemployment and peak unemployment to 1/3. We then explore in detail elements in both the agricultural and rural transformations that could help increase high-season employment and smooth out seasonal employment.

We find that that there is no silver bullet to smooth out rural labor calendars and that a broad array of instruments need to be mobilized to have impact, as each of them only makes a (small) contribution to addressing the seasonal under-employment problem. For the agricultural transformation, raising livestock and, in a limited way, dry-season planting permitted by irrigation and crop diversification reduce the variability of hours worked across months. Growing tobacco may smooth out labor demand in the growing and harvest seasons but its high planting season labor demand corresponds to that of the main staples. Rural transformation includes labor market participation and engagement in a non-farm enterprise. Both of these activities add labor use throughout the year, with labor market participation more effective as a counter-cyclical activity.

Our results are in sharp contrast with the conclusion drawn by Wodon and Beegle (2006) who analyzed the same 2004 LSMS data for Malawi. Like us, they find an important seasonality in labor use and substantial under-employment during most of the year in rural areas. But contrary to us, they find labor shortages in some months of the cropping season that, they conclude, limits households' ability to fully use their productive endowments such as land. Part of the difference in overall employment is due to changing conditions over time, with a large decline in farm size, as we will see below. But there is also a methodological difference with our analysis, as they include in total time worked not only productive activities in agriculture (on-farm self-employment, labor exchange, and wage labor) and the rural non-farm economy (off-farm self-employment and wage labor), but also domestic chores and the production of z-goods such as fetching water and firewood collection. These activities roughly add 23 hours to women's weeks and 4 hours to men's in both rural and urban contexts, with almost no variation across months of the year. We opted for a narrower definition of total work that solely includes income generating activities (productive activities in agriculture and the rural non-farm economy), more in line with the focus of our article on the poverty consequences of under-employment. This choice does not negate the long hours that households have to spend on these other activities, with their potential gender imbalance, nor the time and cost that workers may have to spend getting to their employment. In that sense, our measure of under-employment is strictly a measure of lack of opportunities for income generating activities, not of leisure.

The outline of the article is as follows. The first two sections present the data and the context of rural poverty in Malawi. The following three sections represent the core of the paper, where we construct and contrast labor calendars for rural and urban households, measure the share of unemployment that is due to seasonality, and verify that the differential welfare between urban and rural household is driven by underemployment rather than productivity differences. In the last two sections we explore elements of the agricultural transformation that can help smooth out labor calendars and the following section does the same for the rural transformation. The final section concludes and draws policy implications.

2.2 Data

To investigate labor market seasonality, we utilize principally data from Malawi's Third Integrated Household Survey (IHS3) collected in 2010-11. This is a living standards measurement survey (LSMS) covering a cross section of more than twelve thousands households. The IHS3 uses a stratified two-stage sample design, first sampling enumeration areas (EA) in the 2008 Population and Housing Census stratified by rural/urban location and then sampling households from a list that was constructed for each sampled EA. A minimum of 24 EAs were sampled in each district. For practical reasons, a multiple of 12 EAs were sampled in each stratum in order to distribute the sample evenly across the 12 months. The IHS3 is a very comprehensive household survey designed to monitor conditions in Malawian households.

The rural labor supply can be observed using the time use questions featured in the employment module of the household questionnaire. The questions ask each member of the household above the age of five to report the number of hours spent in the past seven days on several different activities which we group into four categories: agriculture (agricultural activities including livestock and fishing), business (running a household business and helping in a household business), casual labor, and regular wage-paying labor.[3] In this article, weekly work hours will be analyzed at both the household and individual levels. Household labor hours per week aggregates the hours reported by all members of the household over the age of five thereby capturing the labor of household members that may not be the primary bread winners. Our main household sample consists of 12,266 households of which 10,037 are rural and 2,229 are urban. Analysis of individual labor hours per week only includes individuals of working-age (15 to 65 years old) who report that they are not attending school, which we will refer to as 'individuals' or 'adults' without further reference to these selection criteria. Our adult sample consists of 23,324 individuals in 11,492 households as 774 households have no working-age adults. Of theses adults, 18,699 are rural and 4,625 are urban. Since interviews were spread throughout the year, we can observe the seasonality of activities and establish labor calendars for the whole population or subgroups of the population, at both an individual and a household level.

[3]The survey questions distinguish between "casual, part-time or ganyu labor", and "for a wage, salary, commission, or any payment in kind, excluding ganyu". It is this second category that we name 'regular wage-paying labor' or 'wage labor' as 93% of the respondents declare working at least 35 hours last week, while the majority of those under casual labor worked less than 15 hours. The survey also asks about unpaid apprenticeships but we drop this category as very few respondents engage in it. The time use survey also asks respondents how much time was spent yesterday on collecting firewood and water which we omit from our analysis.

These two levels of analysis correspond to different ways of looking at labor use. Employment is normally measured at the individual level, leading to clear measures of unemployment (no hours worked) and underemployment (comparing hours worked to a norm of full employment). However, the averaging of these hours worked does not necessarily measure the aggregate availability of work in any particular area in the given month. This is because working age members of a household may temporarily leave or come back in response to availability of income generating opportunities where the household resides. At the extreme, a fully unemployed person migrating during the low season would raise the average per individual employment while obviously it does not increase the household's in situ employment. It is much less likely that any household would entirely leave an area for seasonal migration. Additionally, fluctuations in labor demand may induce the young and elderly to provide supplemental labor in times of need which would not be reflected in the individual analysis. For these reasons the average number of hours worked by households in a week of a given month should give a better measure of aggregate work availability during that month, where the household resides.

We also use the other surveys in this series, the second and fourth Integrated Household Survey collected using the same methods in 2004 and 2016, respectively. However, we rely primarily on the 2010 results as the 2010 survey features both a large number of EAs and the most even spread of the timing of EA interviews across calendar months. We use the data from the 2004 and 2016 waves to observe aggregate trends over these 12 years in some household characteristics, and as robustness checks for the results established with the 2010 survey.

2.3 Rural poverty in Malawi

Malawi, with a population estimated at 18 million in 2016, is one of the least developed countries in the world ranking 170 out of 188 countries on the UNDP's Human Development Index.[4] Though Malawians have experienced significant improvements in life expectancy and education since 1990, estimated GNI per capita has not grown proportionally during this time period, contributing to the reproduction of monetary poverty.[4] While 71% of the population lived below the international absolute poverty line of US$1.90 PPP per day in 2010, this percentage was still equal to 70% in 2016.[5]

Representing about 30% of the country's GDP, agriculture is central to livelihoods.[5] 92% of rural households and 38% of urban households surveyed report farming at least one plot of land. In all three of Malawi's regions–North, Central, and South–the agricultural sector is characterized by smallholder farms primarily cultivating maize on rainfed plots during the rainy season, the main agricultural cycle, that runs from October to June. Irrigation is rare leaving crops vulnerable to floods and droughts and limiting farming in the dry season (Chafuwa (2017)). During the rainy season, 99% of plots in our sample are rainfed. Only

[4]Human Development Report 2016, *UNDP*, www.hdr.undp.org, accessed 5th Feb. 2018.

[5]*World Bank Country Data: Malawi,* www.worldbank.org/en/country/malawi, accessed 5th Feb. 2018.

10% of households report planting during the dry season that runs from June to October and those that do so rely primarily on bucket irrigation.

Farms are small, with a mean holding of 2.38 acres though it is slightly higher in the central region where it reaches 3.47 acres. Maize and intercropped maize account for the majority of farmed acreage, accounting for 72% of the area cultivated by the mean household.[6] Tobacco is an important cash crop, particularly in the central region, accounting for 51% of national export revenues in 2010.[7]

63% of farming households in our sample report relying solely on household labor. 27% make use of hired labor and 14% of labor that was "free of charge, as exchange laborers, or to assist for nothing in return," with 4% using both. Off farm employment opportunities are limited mostly to small scale entrepreneurship and casual day labor (referred to as "ganyu" labor).

Regular wage-paying jobs are scarce, even in the cities, which experience high levels of unemployment which we will characterize in the next section. A feasibility analysis by Evidence Action in 2014 for a migration subsidy intervention interviewed 81 respondents who reported very low success rates at finding urban jobs leading the report to conclude that "there are insufficient potential migration destinations to absorb excess labor from rural areas" (Evidence Action (2014)). Overall, unemployment in both rural and urban areas is a serious issue in Malawi.

Continued demographic pressure on the land and lack of urban employment opportunities has resulted in a dramatic decline in farm size and in time worked by households across surveys. Farm size declined from 2.29 acres per household engaged in agriculture in 2004 to 1.38 in 2016. Total household labor hours declined from 59.2 per week in 2004, to 41 in 2010, and 31.7 in 2016, while the number of adults in the household declined from 2.0 in 2004 to 1.8 in 2016.[8] This means that land per adult decreased by 18% from 1.13 to 0.93 acres. Malawi thus epitomizes countries stuck in a Malthusian trap.

2.4 Comparing rural and urban labor calendars

In this section we build and compare the labor calendar of rural and urban households. Overall results are reported in Figure 2.1 and Table 2.1.

Figure 2.1 reports the estimated average hours work per week by households throughout the year from the estimation of:

[6] Table B.1 in the appendix gives the average acreage planted per household by crop or intercropped combination for surveyed households for the country and each of the three regions. Categories were defined by first grouping varietals of the same crop (i.e., hybrid maize, local maize, etc.) and then looking for common crop combinations as multiple crops are commonly grown on the same plot.

[7] The Atlas of Economic Complexity, http://atlas.cid.harvard.edu, accessed 5th Sep. 2018.

[8] Table B.3 in the appendix shows the evolution of farm size over time. Because GPS measures are not available for 2004, to make the comparison over time we use self reported areas in all three years. Furthermore, because there are far more outliers in the self reported area (mainly due to what are likely miscoding of the unit of measurement (m2 vs acres), we winsorized the area at 5 pct.

$$L_h = \beta_1 Mar10_h + \beta_2 Apr10_h + ... + \beta_{12} Feb11_h + \beta_{13} Mar11_h + \epsilon_h, \tag{2.1}$$

where L_h is total hours spent engaged in labor activities by household h during the reference week, calculated as the sum of hours spent on all four productive activities (agriculture, business, casual and wage labor), summed over all household members, and the regressors are dummy variables set to one if the reference week for the time use questionnaire of household h falls in that month. Estimated parameters $\widehat{\beta_m}$ with 95% confidence intervals are reported for rural households and urban households separately. We observe that urban households have a relatively stable employment level through the year, of 50 to 60 hours per week. In contrast household employment in rural areas shows a clear seasonal pattern.

Table 2.1 reports several summary statistics from these calendars. In column 1 of panel a, the total annual hours worked is calculated using the estimates from equation 2.1, which are multiplied by the number of weeks in the month, and then summed across months,[9] or

$$\text{Estimated Annual Household Total} = \widehat{LL} = \sum_{m=1}^{12} \widehat{\beta_m} * \# \text{ weeks in } m. \tag{2.2}$$

Observing the marked seasonal pattern of rural employment in figure 2.1, we define the high season as the months of December and January, during which planting takes place, and the low season as the months of July and August, where labor use is at its low point. Weekly hours in the high and low seasons are calculated by taking the mean of the corresponding $\widehat{\beta_m}$ coefficients from equation 2.1. The reported standard deviation is the standard deviation of the 13 $\widehat{\beta_m}$, and the coefficient of variation the ratio of this standard deviation to the mean value of the estimated coefficients, multiplied by 100. Panel b of Table 2.1 reports similar statistics for the binary variable of whether the household provides any labor hours, which we refer to household labor engagement. These statistics exhibit some striking patterns that we now analyze.

There is significantly more variability in rural than in urban labor calendars.

Notable in these urban-rural contrasts in labor calendars is that high season activity offers similar work opportunities for rural and urban households, both in terms of hours worked (both 57-58 hours as observable in column 2) and the percent of active households (column 2 shows a statistically insignificant difference in household labor engagement, with urban areas higher by only 4 percentage points).[10] There is however a large significant discrepancy in the rest of labor calendar months, with labor per week for rural households 57% of that for urban households and engagement 10 percentage points lower among rural households in the low season as noted in column 3. This higher variability of rural calendars is captured

[9]Since the survey lasted 13 months, we have two observations for the month of March, in 2010 and 2011. Figures report them separately, but for all calculations that refer to one year, we use the mean of the two observations for March in our calculations.

[10]Households are considered active if they report spending any time in labor activities. Figure B.1 of the appendix displays the percent of active households by month of interview for rural and urban areas.

by comparing the coefficients of variation of work over the different months of the year. The coefficients of variation in hours worked is 136% higher for rural compared to urban households, as noted in the third row of column 5.

We can decompose the difference in the coefficient of variation between rural and urban households into the difference in mean values and the difference in standard deviations as follows:

$$\frac{\Delta CV}{CV} \approx \frac{\Delta St.Dev.}{St.Dev.} - \frac{\Delta Mean}{Mean} \tag{2.3}$$

In this case, rural households have both a higher standard deviation in work across months of the year (70 and 50% for hours and participation, respectively) and, for hours work, a lower mean value (by 28%). Both of these contribute to the very large difference in variability of labor calendars.[11]

Figure 2.2 disaggregates the labor hours reported in figure 2.1 by activity. It shows that agriculture is by far the largest and the most cyclical source of work for rural households, and that employment in the other activities–household business, casual labor, and wage labor–is relatively stable throughout the year. Importantly, they are not countercyclical to agriculture. Their contributions to overall smoothing of the labor calendar (reduction of the coefficient of variation of labor across months) is thus by adding labor opportunities in less seasonal activities throughout the year rather than by complementing work in agriculture when the latter is low.

There is significant underemployment in rural areas even in the high season.

Looking now at effective unemployment, we turn to individual level observations. There is a dramatic seasonal contrast in the distribution of hours worked by rural adults in all activities. Close to 50% of surveyed rural adults report working no or a very low number of hours in the low seasons. [12] To obtain labor calendars at the individual level we estimate an equation similar to equation 2.1 at the individual level, and report corresponding statistics in table 2.1. We observe that unemployment rises from 7% in high season to 36% in low season and the average number of hours worked per week falls by half from 24.6 to 12.4. However, even in high season, underemployment prevails. With 25 hours per week, underemployment is 38% for a benchmark full-employment of 40 hours per week.

There is also significant unemployment in urban areas.

[11] We verify the results in Table 2.1 obtained with the 2010 data for household labor supplied and for individuals participation in tables B.4 and B.5 using the 2004 and 2016 LSMS-ISA data. We see that results are broadly consistent to those of 2010. Rural household labor calendars for hours worked have a CV which is larger than their urban counterparts. The same applies to individual labor engagement, with exception of the 2004 result. Pooled data across the three surveys show a CV which is almost three times higher for rural household hours worked and double for individual labor participation compared to their urban counterparts.

[12] Figure B.2 in the appendix shows histograms of hours worked in the past week for rural adults in all activities. Panel a shows the distribution for the high season while panel b shows the distribution for the low season.

Referring to table 2.1 panel b, we see significant unemployment in urban areas too. The mean individual unemployment rate is 36%, and it remains high throughout the year. Hence although urban adults work more hours on average than rural adults, this high urban unemployment rate limits the opportunity for rural workers to use seasonal or permanent migration to fill in their unused time given the challenges to finding productive employment in the urban economy. Labor displacement to the urban sector is not accompanied by productive employment, but by accumulation of labor in urban slums and no effect on growth. This phenomenon was observed in the 2008 World Development Report WorldDev2008 for many Sub-Saharan Africa countries where a decline in the share of the labor force employed in agriculture is not accompanied by a corresponding increase in GDP per capita. Malawi was one of them.

Low employment is associated with dependence on agriculture

In this section we try to get some insights on who is most affected by high unemployment, especially in rural areas. Figure 2.3 compares the employment structure across the four major categories of activities for rural individuals based on working hours reported when interviewed in the low season (July and August). Note that 34% of individuals report no work at all and are not included in this table. We see that individuals severely underemployed in the low season are less likely to be working in occupations else than agriculture. Hence, despite working very few hours in agriculture, they depend on agriculture for 68% of their work time compared to 38% for those working over 30 hours. Work in household non-agricultural businesses and in casual labor gains some importance as we move from households that work less than 10 hours to those working more than 30 hours. The main activity that makes a difference for those working full time is engagement in the wage labor market. As a group, these individuals work on average 18 hours in agriculture, 8 in their businesses, 9 in casual labor, and 14 in wage labor. It becomes apparent that while low employment may be a problem throughout Malawi's economy, it is particularly pronounced for rural households that are dependent on agriculture as their primary occupation, highlighting the importance of not only an agricultural transformation but the need for opportunities in the rural non-farm economy that would emerge in a rural transformation.

Urban-rural labor time equilibrates in the high season.

Against this backdrop of large unemployment in both rural and urban areas, one should notice that in the high season, hours worked in the two sectors are not very different (table 2.1a). Households work 56.9 hours per week in the rural areas and 58.2 in the urban areas. Individuals work 24.6 hours per week in rural areas and 28.1 in urban areas. Yet, participation rates for individuals show a striking contrast, with 93% of the population employed in rural areas, indicating extensive work sharing, while all work available in urban areas is shared among 67% of the population, leaving 33% unemployed.

The role of unemployment in understanding differences in welfare between rural and urban households.

We decompose the differences in labor productivity between rural and urban areas. We consider how much of the rural-urban income gap is due to differences in hourly productivity and how much is due to underemployment in rural areas.

The IHS3 survey administers a consumption module to each household. The first row of table 2.2 compares urban and rural household consumption levels, C_h, at the mean and median levels, showing a low rural/urban ratio of 0.42 for means and 0.54 for medians. Similarly to McCullough (2017)'s adjustments for sectoral productivity, we proceed to adjust the mean household consumption by our estimate of households total labor hours worked, $\hat{L}_{yh}$, as calculated in equation 2.2. For this we calculate $\bar{C_h}/\hat{L}_{yh}$, at the mean and median, for rural and urban households. Results are reported in row 2 of table 2.2. Since rural households work on average 72% of the hours worked by urban households, annually, calculating consumption on a per hour worked basis leads the the rural/urban ratio raises sharply to 0.58 for means and 0.75 for medians.

One issue with this adjustment by average hours worked at the household level is that it assumes the same annual employment level for all households.[13] If employment and consumption levels are correlated (which we expect, larger households having higher employment and consumption), this would not be correct. An alternative is then to compare household consumption per working age individuals, by calculating the mean and median of $C_i = C_h/I_h$ for urban and rural areas, where I_h is the number of working age individuals in household h. These results are report this in row 3 of table 2.2. As above, we further adjust this value by the annual hours worked per individual, in rural vs. urban areas, by calculating $\bar{C_i}/\hat{L}_i$ in row 4.

These per adult calculations have a potential opposite bias if adults in a household share work opportunities, and the employment rate of adults decreases with the number of adults in the household. Results reported in table 2.2 show our results to be very robust to the method used. Because the number of adults per household is lower in rural than in urban areas, the rural/urban ratio in consumption per adult is a bit higher in per adult terms, but the main adjustment comes from measuring it on a per-hour basis. The rural/urban ratio of consumption per hour worked is 0.66 for the mean and 0.81 for the median.

This result is similar to McCullough (2017) comparing the sectoral productivity contrast between agriculture and non agriculture. It stresses the fact that urban-rural consumption gaps come not so much from a differential return per hour worked than from a differential in number of hours worked, much to the advantage of the urban population.

[13] Recall that we only observe each household labor use for one week in a very seasonal calendar. Hence we cannot infer its own annual labor use, and need to resort to an average over the population or a segment of the population.

2.5 Decomposing rural unemployment between peak and seasonal deficits

In the previous section, we observed substantial unemployment in rural areas throughout the year, characterized by an important seasonal pattern. In this section, we propose an approach to measure the share of unemployment faced by rural households that results from seasonality. Any measure of unemployment is based on a definition of full employment. We thus start with a definition of full employment appropriate to this context, and proceed to decompose annual unemployment into what we call peak unemployment and seasonal unemployment.

Malawi distinguishes itself as having a large deficit in employment opportunities. If we define full employment as 48 weeks per year (to allow for unexpected shocks such as illness and political disruptions) and 40 hours per week (to allow time for household maintenance and reproduction), annual hours reported in table 2.1, panel a, show urban individuals to be at 67.1% of the 1920-hours work potential and rural individuals at 47.3%. Looking at the high season, urban workers work 28.1 hours per week and rural workers 24.6. Urban workers are thus still only at 70.2% of a 40 hour week, and rural workers at 61.5%. Hence, a deficit in work opportunities applies to both urban and rural workers, and exists throughout the year. It is this large and pervasive urban work deficit that limits the possibility of using rural-urban migration as a major instrument for poverty reduction Evidence Action (2014). Solving the deficit in work opportunities, basically through labor-intensive aggregate economic growth, remains the key issue for large scale poverty reduction in Malawi.

Given this important deficit, what is the importance of seasonality in rural households labor calendars in their opportunities to work? Since full employment as defined above is completely out of reach, we propose to consider the current high-season urban workload as the benchmark employment for rural adults throughout the year. Using the numbers reported in column 2 of table 2.1, we see that the high-season urban workload is 28.05 hours a week per adult, which amounts to a benchmark of 1459 annual hours for the year. We then define the peak deficit as the annualized difference between the high season work load in rural areas and this potential maximum. Since the high-season rural work load is 24.61 hours a week, the peak deficit accounts for 3.44 hours a week, for an annual deficit of 179 hours. In other words, this is the under-employment level that would prevail in rural areas assuming that high-season employment was constant throughout the year. Seasonal under-employment is then defined as the difference between the observed labor hours in the year and this annualized high-season level. We estimate 909 annual labor hours per rural adult, as noted in column 1 of table 2.1. When compared to our benchmark of 1459 annual hours, this gives us a deficit of 550 hours. Since the peak deficit accounts for 179 annual hours, we attribute the remaining 371 hours to the seasonal deficit. The seasonal deficit is then 67% of the total deficit. Beyond addressing the high-season deficit for urban and rural workers, the seasonality of rural labor calendars is indeed a big issue. Finding ways of smoothing rural labor calendars through agricultural and rural transformations is thus a key policy problem in addressing rural poverty. This is what we explore in the following section.

2.6 Elements of an agricultural transformation that can help smooth labor calendars

In this section, we explore the timing of agricultural labor requirements to better characterize the reason behind the seasonality in labor demand that rural household face. We begin by looking at the timing of the labor requirements associated with the main crops grown in Malawi. We then consider the timing of labor in other agricultural activities and activities associated with the rural non-farm economy to consider how rural households may smooth their labor throughout the year by engaging in counter cyclical activities.

Agricultural labor calendars

In order to better understand the extreme seasonality of labor demand in rural Malawi, and to validate our results using the time use survey, we use information in the agriculture questionnaire of the LSMS to construct an estimate of labor demand by crop per acre for each day of the agricultural season. We construct labor demand calendars for the most common types of crops and intercropping combinations reported in the 2009/2010 rainy season.[14]

Constructing these crop level labor demand calendars is not trivial as it effectively entails calculating the household labor used each week on each plot in the dataset so that we can then generate a representative calendar for each crop. While non-trivial, we find this exercise both informative and methodologically interesting. Informative because this allows us to observe how crops agronomy contributes to the seasonality of labor demand. Methodologically interesting because this approach could easily be applied to other contexts and datasets that include agricultural modules similar to the one found in the LSMS. Indeed, unlike our results using the time use modules, the approach that follows does not require that the survey be conducted continuously across the calendar year as it relies on retrospective data commonly found in agricultural modules.

Estimates of the mean weekly labor demand per acre of a crop are generated by constructing plot level labor demand calendars for each plot farmed. These household plot calendars are constructed using two key pieces of information reported in the agricultural questionnaire for the plot. The timing of planting and harvest activities as well as the amount of household labor that was applied to the plot.

Respondents are asked about the timing of planting and harvesting. Using this information, for each plot j we estimate the duration in weeks D_j^p, beginning date p_j^b, and end

[14]Maize and intercropped maize is the main crop grown in Malawi followed by tobacco and groundnuts. Table B.1 of the appendix gives the average acreage planted per household for the main crops.

date p_j^e of planting activities on the plot [15] as well as the duration in weeks D_j^h, beginning date h_j^b, and end date h_j^e of harvest activities[16] and define the period between these as the growing season such that $p_j^e = g_j^b$ and $g_j^e = h_j^b$ with a duration in weeks of D_j^g.[17]

For each of these three activities (planting, growing and harvesting), respondents are also asked about household labor,[18] reporting the number of weeks, the days per week and the hours per day each household member was engaged on the plot. We can thus calculate L_j^a, the total amount of household labor hours applied by n household members to the plot j for activity a, adjusted for plot size, as

$$L_j^a = \frac{\sum_{i=1}^{n} weeks_{ij}^a * days/week_{ij}^a * hours/day_{ij}^a}{Acres_j}. \tag{2.4}$$

Plot level, acreage adjusted weekly labor hour demand for each of the three activities, l_j^a, is then estimated as

$$l_j^a = \frac{L_j^a}{D_j^a}. \tag{2.5}$$

For each plot we can then assign l_j^a, to the each day of the calendar year in which the household is engaged in activity a. This defines, ℓ_{dj}, the acreage adjusted weekly labor hour

[15]The start and end dates of a household's planting activities are determined using two elements reported in the LSMS survey. First, the survey asks respondents the month in which they planted the seed on the plot. Second, the survey asks each household member the number of weeks they were engaged in planting activities on the the plot. We select the maximum number of weeks reported by any of the n household members and set this as the duration of the household's engagement in planting on plot j, $D_j^p = \max_{i \in n}(weeks_{ij}^p)$. We randomly select a day in the month in which seeds were reported to be planted and set this as the midpoint of planting activities. We use this date and the duration of planting activities, D_j^p, to calculate the beginning date p_j^b and end date p_j^e of planting.

[16]The start and end dates of a household's harvest activities are determined using two elements reported in the LSMS survey. First, the survey asks respondents the month in which they started harvesting the plot. Second, the survey asks each household member the number of weeks they were engaged in harvesting activities. We select the maximum number of weeks reported by any of the n household members and set this as the duration of the household's engagement in harvesting on plot j, $D_j^h = \max_{i \in n}(weeks_{ij}^h)$. We randomly select a day in the month in which the harvest started and set this as h_j^b and then use the duration of harvest activities, D_j^h, to calculate h_j^e.

[17]The timing of growing season activities is not specified in the survey. We opt to define the duration of growing season activities on plot j, D_j^g, as the number of weeks between the end of planting, p_j^e, and the beginning of harvest activities, h_j^b, though the number of weeks people actually report working in growing season activities during that period suggest that these hours are often lumped together over a few weeks rather than spread evenly across the growing months.

[18]In order to build a representative calendar of labor demand by crop we use the 69.4% of plots that rely solely on household labor. We exclude households that engage in hiring and exchanging labor as non-household labor is not disaggregated by task and is measured in days rather than hours, making comparisons to household labor difficult. We verified that while these households typically farm fewer acres, their crop composition is broadly comparable to that of households hiring and exchanging labor. Estimates of the timing of farming activities and the labor hours required for each task and crop using this subset consisting of 10,253 plots farmed by 6,260 households should thus be generalizable to the full sample.

demanded for the week of day d on plot j, such that

$$\ell_{dj} = \begin{cases} 0 & \text{if } d \leq p_j^b \\ l_j^p & \text{if } p_j^b \leq d < p_j^e \\ l_j^g & \text{if } p_j^e \leq d < h_j^b \\ l_j^h & \text{if } h_j^b \leq d < h_j^e \\ 0 & \text{if } h_j^e < d. \end{cases} \tag{2.6}$$

We then calculate the average number of hours $\bar{\ell}_d$ for each day of the agricultural season to generate a representative calendar for a one acre plot of that crop. Estimated labor calendars are plotted in figures 2.4a and 2.4b for the most common crop and intercropped combinations.[19]

We see that the November-December planting period is the peak of labor demand. Maize and intercropped maize accounts for over 70% of the acreage of the typical household farm,[20] thus the timing of maize planting and harvest as illustrated in figure 2.4a governs the fluctuation in the labor demand calendar of the typical households. The other commonly grown crops, tobacco and groundnuts, also compete for labor hours during the same high demand planting season. Labor demand at harvest time is much lower and exhibits more dispersion between different crops. Peak harvesting for maize happens in April. Plots that are intercropped with pigeon-peas continue to require labor inputs until the late pigeon-peas harvest in July and August. As seen in figure 2.4b groundnut harvesting is more labor intensive than the maize harvest but still does not require a substantial labor input as compared to planting activities.[21] The timing of the groundnut harvest is also more spread out running from April to June. The only crop that has a very different pattern in the timing of labor demand as compared to the maize staple is tobacco. Tobacco leaves start to get harvested quite early in the agricultural season and continues until the end of March, right before the maize harvest begins. The tobacco harvest is highly labor intensive, including of child labor (Xia and Deininger (2019)), requiring 2.5 times more labor hours than harvesting maize. [22] Finally, while the tobacco harvest is counter cyclical to the maize harvest, the peak labor

[19]Generating the plot level labor calendar for intercropped plots is more complicated. We limit our calculation of daily labor calendars to plots with no more than four intercropped crops. The questionnaires elicit timing questions for each crop on the plot, however labor applied to the plot is not differentiated by crop. We opt to divide the reported planting and harvesting labor hours equally across crops such that $L_{jc}^h = \frac{L_j^h}{C}$ and $L_{jc}^p = \frac{L_j^p}{C}$ where C is the total number of crops planted on a plot. Furthermore, we also divide the number of weeks households report being engaged in planting and harvesting activities by the number of crops. We then use the crop specific timing question responses to calculate the beginning p_{jc}^b and end p_{jc}^e, of planting activities for each crop, as well as the beginning h_{jc}^b, and end h_{jc}^e, of harvest activities of each crop using the same approach as above. The growing period captures any remaining undefined days between the earliest planting and last harvesting day.

[20]See appendix B.1.

[21]See appendix table B.2 for total labor demand estimates for each activity by crop.

[22]See appendix table B.2.

demand for tobacco is also its planting season which coincides with the planting of other crops.

For each household h, [23] we can then re-weight the plot level labor calendar ℓ_{dj} by the acres of plot j and sum across the household's J plots to generate $\mathcal{L}_{dh}$, the weekly agricultural labor hours demanded for household h in the week of day d. Thus for each day, we calculate

$$\mathcal{L}_{d=x,h} = \sum_{j=1}^{J} \ell_{d=x,j} * Acres_j. \tag{2.7}$$

From these daily household labor calendars, we then calculate the average number of hours across households, $\bar{\mathcal{L}}_d$, for each day of the agricultural season to generate a representative calendar for household agricultural labor demand, plotted in figure 2.5.

The agricultural labor demand calendar generated with this procedure covers the 2009/2010 agricultural season (rather than the 2010/2011 survey season) and relies on retrospective recalls of significant agricultural dates and labor requirements. Nonetheless, this calendar is consistent with the labor hours in agriculture reported in figure 2.2a.[24] Figure 2.5 shows a sharply concentrated labor calendar, particularly at planting time. These concentrated labor demands in agriculture are at the origin of the high seasonality in rural households' labor calendars. Else than planting (and to a lesser extent harvesting), labor demand per household in agriculture is minimal given the small size of the average family holding.

Specific contributors to labor smoothing

We saw in figure 2.2a that agricultural activities have a very strong seasonal pattern of labor use, largely responsible for the seasonality in rural labor calendars. In this section, we look into more specific activities or characteristics of agricultural production that could contribute to smoothing the agricultural labor calendars. In order to do this, we contrast the time use survey labor supply calendars of rural households that do or do not participate in these activities. Note that undertaking an activity may or may not generate higher employment depending on whether it fully substitutes or not to the other household activities, which we can check by comparing total annual hours worked. In terms of its contribution to smoothing the labor calendar, best would be that the activity be counter-cyclical to the other activities in which households are engaged, as it will then generate a decline in the standard deviation (SD) of labor use across months. Nonetheless, even if it is not counter-cyclical, an activity that generates a constant amount of labor through the year will induce no change in SD but

[23] We select only households that do not hire or exchange labor on any plots leaving 8,543 plots farmed by 5,094 households. We do this to avoid concerns about substitution of household and outside labor between plots.

[24] Differences between these two graphs could be due to differences between years and recall errors. In addition, the phrasing of recall questions about labor hours induce respondents to report in a lumpy way which creates some arbitrariness in the way we define the length and intensity of work when there are different members of the household working different lengths of time. Finally, figure 2.2a also include hours spent on other activities not associated with specific crops (eg livestock).

a decline in the coefficient of variation (CV) of the labor calendar, as illustrated by equation (2.3).

Table 2.3 reports total hours worked, high and low season work, SD and CV of hours worked across months of the year for households that do or do not participate in these activities. Because we are looking at potential smoothing of the agricultural work calendar, the sample used in this table consists of the 9,389 rural households (93.5% of all rural households) that are directly engaged in agriculture by cultivating a plot of land and/or owning livestock. We use this grid of indicators to assess in this section the contributions of livestock, tobacco, crop diversity, farm area, irrigation, and use of non-family labor to smoothing the agricultural labor calendar.

Livestock. About 56% of rural households engaged in agriculture own livestock. Of the households that own livestock, the mean is of 10.7 heads, of which 62% are poultry, 24% are sheep or goats, 7% pigs, and 3% cattle. Figure 2.6a shows working hours for households that own livestock compared to those that do not. The figures show households that raise livestock have higher household work hours throughout the year, with no seasonal effect, except possibly during the harvesting period when livestock has to be herded away from crops. This is reflected in a 33% increase in total hours worked with almost no difference in the SD (table 2.3). By adding to work opportunities, livestock reduces the CV of the agricultural labor calendar by 23-24% for both households and individuals.

Tobacco. Most of the tobacco in Malawi is cultivated by smallholder farmers (Lea and Hanmer (2009)). As observed by Orr (2000) and by Xia and Deininger (2019), tobacco is highly labor intensive, especially at harvest time. Comparing hours worked in households that grow tobacco compared to those that do not shows that tobacco adds a significant 33% to household labor. Because the labor intensive planting season coincides with that of other crops, tobacco provides limited smoothing opportunities. Nonetheless, as visible in figures 2.4b and 2.6b, the labor intensive harvest season of tobacco does create an increase in labor requirements during the early period of the growing season prior to the harvest of other crops. The net of these two effects results in an increase of the SD, and the CV of agricultural labor calendars is 2% higher for tobacco growing households than for the other households.

Crop diversity. A similar analysis applies to crop diversification. Here we compare households with three or more crops to households planting only one crop. In general one expects crop diversity to smooth the agricultural calendar. Yet here, as with the case of tobacco, the seasonal patterns of rain implies that planting of all crops happen at the same time, and hence multiple crops provide substantially more work but no relief from seasonality of demand for labor.

Farm area. Comparing reported hours for rural households in the top 25% of farmed area compared to the bottom 25% shows that land area is a major determinant of household time worked. By increasing labor a bit more in the low season than in the high season, larger farms have an 11% lower CV of labor calendar than smaller farms.

Irrigation and dry season cultivation. We compare household labor hours in households that report planting a plot in the previous year's dry season. This is generally done with bucket irrigation. What is interesting is that households that irrigate have higher labor demand not only in the dry period, but also during the wet season, suggesting that it is associated with intensification of land use. Irrigation decreases the CV of agricultural labor calendars by 7%.

Use of hired labor. The last two comparisons look at the use of non-family labor in periods of high labor demand. Only 25% of the households ever hire labor. Among those that do hire labor, they hire on average 16 days of labor per year, although the distribution has a long tail with 1% of the households hiring more than 60 days. These numbers are small relative to annual work, although they are certainly critical at particular times of the year. We see very little difference in family labor between households that hire and those that do not hire labor. The interpretation is that households can easily hire labor when their demand is higher than what they would like to supply, so that households maintain their own labor supply in either case. There could have been some difference by selection, as households that do not hire labor include households that are always in surplus of labor. This is likely very marginal as we see that total hours worked is also very similar across these two groups.

Use of exchange labor. The contrast between the roles of exchange labor and hired labor is interesting. Labor exchange is a within season arrangement between households. Typically, instead of having a short very intense few days of work on your own field, you get neighbors to come and help you and then go on to help them. This helps spread each household's work over a longer period of time if there is some heterogeneity in the exact timing of the operation, or if the operation is for technical reason difficult to spread over more days. The CV of monthly hours worked in agriculture is 34% lower for household that use labor exchange and this is all due to spreading labor rather than adding any labor.

In conclusion, agricultural activities on the farm have little countercyclical patterns of labor use with the main crops that could contribute to smooth the labor calendars. Only households raising livestock and to a lesser extent having irrigation that allows intensification of agriculture or more crop diversification have a lower variability in hours worked across months, and this is mostly due to increased labor use throughout the year. In contrast, using labor exchange seems to allow smoothing labor calendars, without any change in aggregate annual labor.

Elements of rural transformation that can help smooth labor calendars

While the agricultural transformation may affect labor calendars through agricultural activities, the rural transformation seeks to affect labor calendars through decisions beyond agriculture such as engagement in non-farm activities. We look into the effect of seasonal participation to labor market activities by household members and the role of household enterprises. Results are summarized in table 2.4.

Agricultural Labor Markets. Participation in the labor market is associated with a large increase of annual hours worked by 37%. It decreases a bit the SD of monthly hours worked by adding a few more hours in the low season than in the high season, but the very large 33% decline in the CV is largely due to the increased overall level of employment. Ricker-Gilbert (2014) shows that fertilizer subsidies, as extensively used in Malawi, can increase labor absorption in the home plot, demand for hired labor, and create a small spillover benefit on all farm workers through higher agricultural wage rates.

Household Enterprises. Figure 2.6c compares reported hours worked by rural households that run a household enterprise to those that do not. Most of the households that run an enterprise are engaged in retail or trade selling consumer products or services. With the exception of some basket weaving, brick making, mat weaving, and tailors there is very little manufacturing of non-perishable goods. Household enterprises increase work hours throughout the year (by an average 36%) with no evidence of counter-cyclical smoothing, to the contrary (the SD is higher by 22%). Work in household enterprises reduces the CV of labor calendars by 11%.

In conclusion, participation to the labor market and having a non-farm enterprise are both associated with a large increase in total employment, and through this with a decrease in the seasonality of work. Participation in the labor market is also associated with some counter-cyclical opportunities that allow a large decrease in the overall seasonality of the labor calendar, which the non-farm enterprises do not provide.

2.7 Conclusion

Structural transformation has been advocated as an engine of growth and poverty reduction for the agriculture-based economies, which include most of the Sub-Saharan Africa countries. In that perspective, land and labor productivity growth in agriculture enables the transfer of labor out of rural areas at no opportunity cost on the price of food. Released labor can then be employed at a higher level of productivity in the urban industrial and services economy. As a consequence, the shares of agriculture in employment and GDP decline while the engine of aggregate growth and poverty reduction is found in capital accumulation and employment creation in the urban economy. The analysis of rural household data permitted by some LSMS surveys shows that this approach to growth and poverty reduction is less evident in countries like Malawi where there is a large deficit of urban employment. Labor transfers from the rural sector are less likely to stimulate GDP growth than to displace poverty to the urban environment. As a consequence, we have focused on growth and poverty reduction in the rural areas themselves through agricultural and rural transformations. Key in using these transformations for rural poverty reduction is to reduce seasonality in labor calendars. We have seen that, taking the urban high season employment rate as the maximum workload that could be attained by rural households under current circumstances, the seasonal work deficit explains 2/3 of the total work deficit for rural households. Smoothing rural labor calendars can be achieved in the agricultural transformation through a variety of instruments including livestock, crop diversity, irrigation, and use of non-family labor, especially exchange labor. Smoothing of labor calendars through the rural transformation includes labor market participation and rural non-farm enterprise development. We have shown that there is no single magic bullet among these various instruments to smooth out labor calendars, requiring instead a comprehensive agenda focusing on all available instruments. Activities that contribute to labor smoothing are however not countercyclical to the labor demands of staple crops agriculture. They instead add to labor opportunities throughout the year. As a consequence, family members are likely to each specialize in one or several of these new activities, rather than engaging in seasonal job switching. In any case, our main result is that the increasingly prevalent agro-pessimism needs revisiting and that, for agriculture-based countries like Malawi, facilitating the engagement of rural households in agricultural and rural transformations seems to be the most effective policy option for growth and poverty reduction.

Tables

Table 2.1: Rural-Urban Contrasts in Labor Calendars: Labor Supply and Engagement

Panel 1a: Labor supplied (*hours worked*)

	Contrast	Total annual hrs	High weekly hrs	Low weekly hrs	Standard deviation	Coeff. of variation (%)
Rural vs. urban, household	Rural	2,065	56.93	29.23	9.58	24.26
	Urban	2,863	58.21	51.38	5.62	10.26
	Rural/urban	0.72***	0.98	0.57***	1.70	2.36
Rural vs. urban, individual	Rural	909	24.61	12.39	4.14	23.85
	Urban	1,288	28.05	23.26	2.63	10.67
	Rural/urban	0.71***	0.88**	0.53***	1.57	2.24

Panel 1b: Labor engagement (*indicator set to 1 if any labor hours are reported*)

	Contrast	Mean % active	High % active	Low % active	Standard deviation	Coeff. of variation (%)
Rural vs. urban, household	Rural	0.88	0.97	0.78	0.06	7.31
	Urban	0.91	0.93	0.87	0.04	3.88
	Rural/urban	0.97***	1.04	0.90***	1.50	1.88
Rural vs. urban, individual	Rural	0.79	0.93	0.64	0.10	13.08
	Urban	0.64	0.67	0.61	0.05	8.58
	Rural/urban	1.23***	1.39***	1.05	2	1.52

Note: Sample consists of 23324 working age individuals who are not in school and 12266 households of which 10037 are rural. Tests for statistical significance of the ratio between the comparison groups being different from 1 are reported for columns 1-3 with * $p<0.1$, ** $p<0.05$ and ***$p<0.01$. Mean percent active is the mean value over the year of the percentage of households that report positive working hours in any given month. High season is December and January, low season is July and August.

Table 2.2: Consumption per Hour Worked

Household consumption units		Rural	Urban	Rural/urban
Per household	Mean	197,000	468,000	0.42
	Median	152,000	284,000	0.54
Per household hour worked	Mean	95	163	0.58
	Median	74	99	0.75
Per individual	Mean	109,000	233,000	0.47
	Median	86,000	151,000	0.57
Per individual hour worked	Mean	120	181	0.66
	Median	95	117	0.81

Note: The adjustment for hours worked is done by dividing consumption by the estimated annual hours worked for the relevant group reported in Table 2.1.

Table 2.3: Agricultural Labor Supply by Participation in Activities of Agricultural Transformation

	Contrast	Obs	Total annual hrs	High weekly hrs	Low weekly hrs	Standard deviation	Coeff. of variation (%)
Livestock	Livestock	5,275	1,667	50.37	20.43	10.40	32.62
	No livestock	4,114	1,252	41.87	13.93	10.15	42.36
	Liv/NoLiv		1.33***	1.20***	1.47***	1.02	0.77
Livestock (individual)	Livestock	10,552	641	19.08	7.46	3.96	32.33
	No livestock	7,158	550	18.16	5.75	4.47	42.46
	Liv/NoLiv		1.17***	1.05	1.30***	0.89	0.76
Tobacco	Tobacco	1,255	1,870	54.32	20.25	13.01	36.32
	No tobacco	8,134	1,404	44.43	17.19	9.59	35.72
	Tob/NoTob		1.33***	1.22**	1.18	1.36	1.02
Crop diversity	More diverse	1,920	1,899	61.43	22.92	14.31	39.41
	Less diverse	2,510	1,133	36.62	9.84	8.86	40.87
	More/Less		1.68***	1.68***	2.33***	1.62	0.96
Large farm	Highest quartile	2,343	1,926	58.33	20.96	13.13	35.61
	Lowest quartile	2,379	1,001	32.76	10.70	7.68	40.12
	Highest/Lowest		1.92***	1.78***	1.96***	1.71	0.89
Dry season planting	Planting	1,287	1,903	57.27	23.36	12.72	34.95
	No planting	8,102	1,408	45.15	16.38	10.10	37.49
	Plant/NoPlant		1.35***	1.27***	1.43**	1.26	0.93
Uses hired labor	Hires	2,309	1,493	45.07	20.31	9.76	34.16
	No hiring	7,080	1,460	46.38	16.17	10.41	37.28
	Hires/NoHires		1.02	0.97	1.26**	0.94	0.92
Uses exchanged labor	Exchanges	1,242	1,460	37.34	17.43	6.98	24.97
	No exchange	8,147	1,464	46.82	17.55	10.59	37.81
	Exch/NoExch		1	0.80***	0.99	0.66	0.66

Note: Sample consists of rural households that report cultivating a plot or owning livestock. Household crops are considered more diversified if they report planting three or more crops and less if they report planting a single crop. Tests for statistical significance of the ratio between the comparison groups being different from 1 are reported for columns 1-3 with * $p<0.1$, ** $p<0.05$ and *** $p<0.01$. High season is December and January, low season is July and August.

Table 2.4: Rural Household Labor Supply by Participation in Activities of Rural Transformation

	Contrast	Obs	Total annual hrs	High weekly hrs	Low weekly hrs	Standard deviation	Coeff. of variation (%)
Work as paid labor	Paid work	6,077	2,323	61.13	35.49	9.40	21.17
	No paid work	3,960	1,698	50.70	21.15	10.21	31.45
	Paid/NoPaid		1.37***	1.21***	1.68***	0.92	0.67
Non-farm enterprise	Enterprise	1,755	2,659	70.96	40.17	11.46	22.53
	No enterprise	8,282	1,948	54.34	27.13	9.43	25.31
	Ent/NoEnt		1.36***	1.31***	1.48***	1.22	0.89

Note: Sample consists of all rural households. Household are categorized as working as paid labor if any houshold member reports working for a wage, salary or in casual labor in the past 12 months. Tests for statistical significance of the ratio between the comparison groups being different from 1 are reported for columns 1-3 with * $p<0.1$, ** $p<0.05$ and ***$p<0.01$. High season is December and January, low season is July and August.

Figures

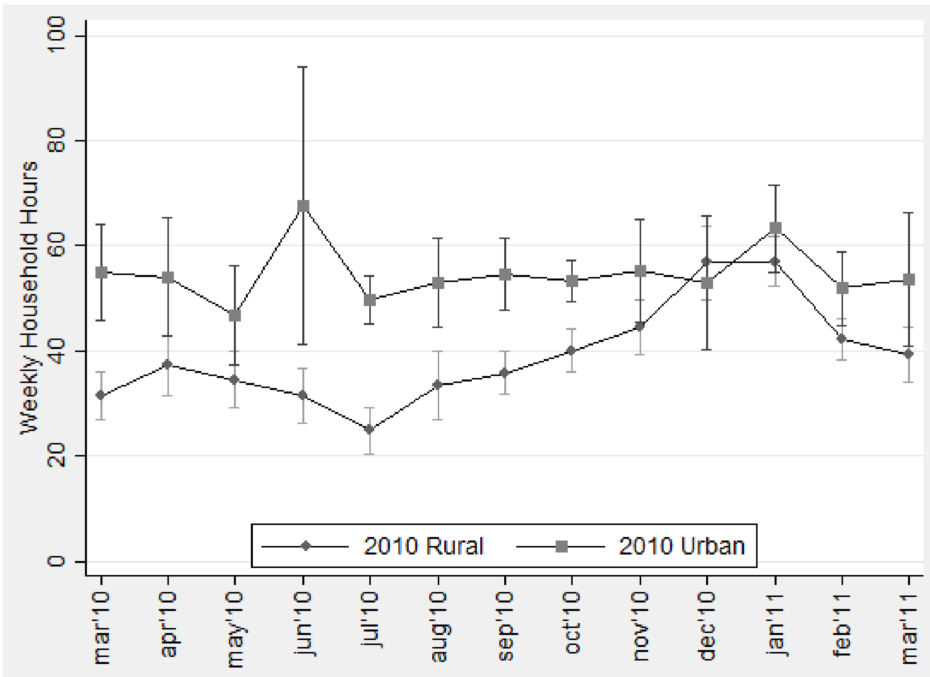
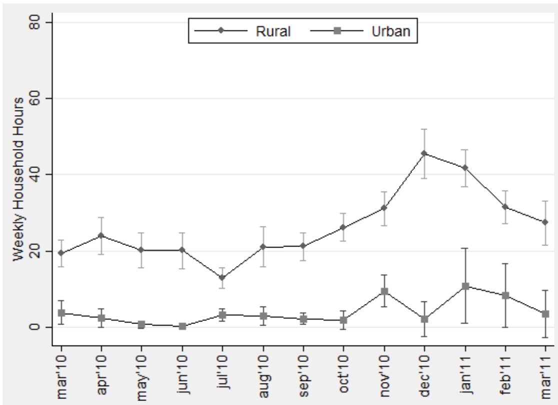
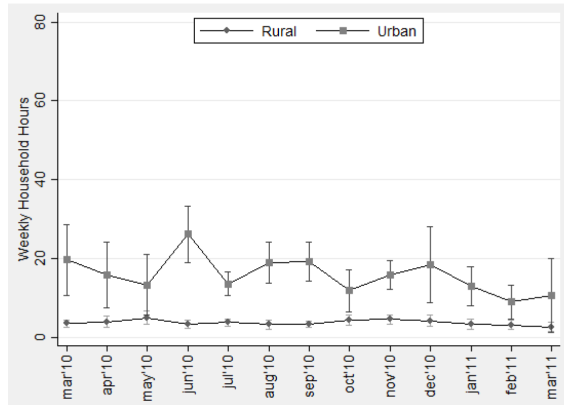


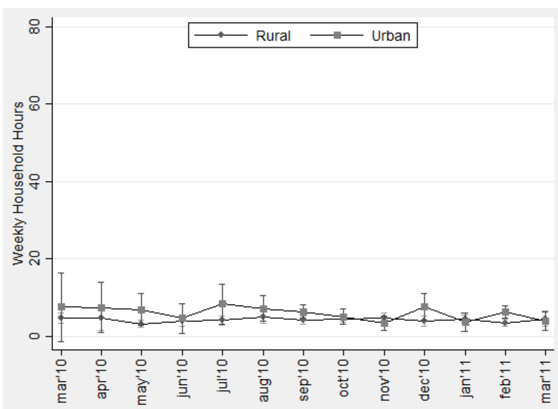
Figure 2.1: Total household labor hours worked last week



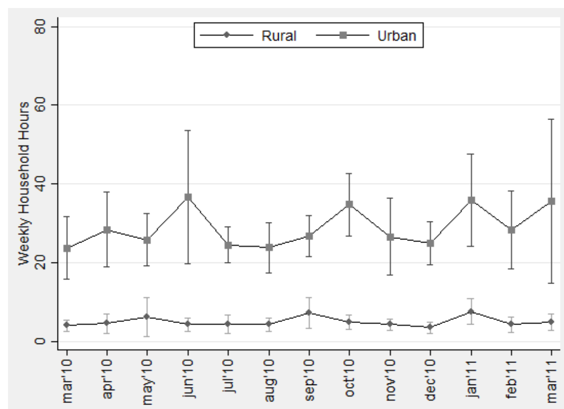
(a) Hours supplied to agriculture



(b) Hours supplied to household businesses



(c) Hours supplied to casual labor



(d) Hours supplied to wage labor

Figure 2.2: Household labor supplied last week by activity

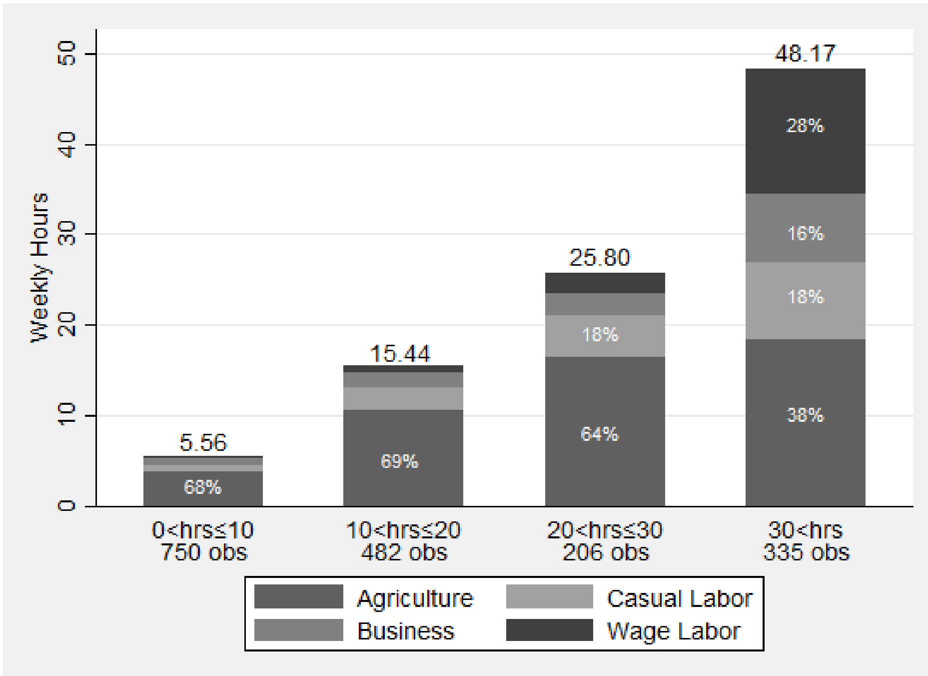
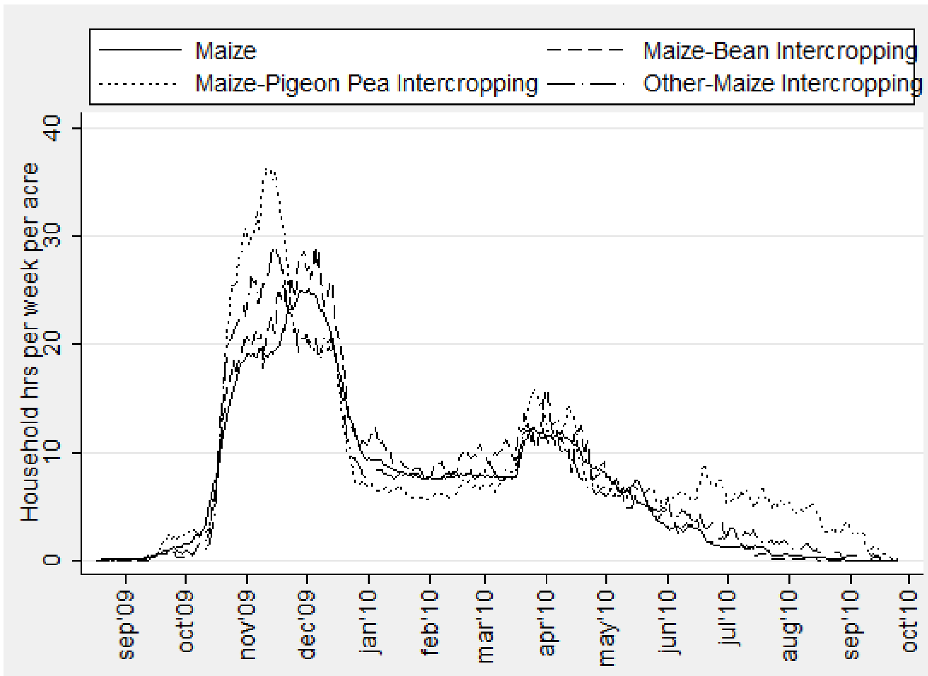
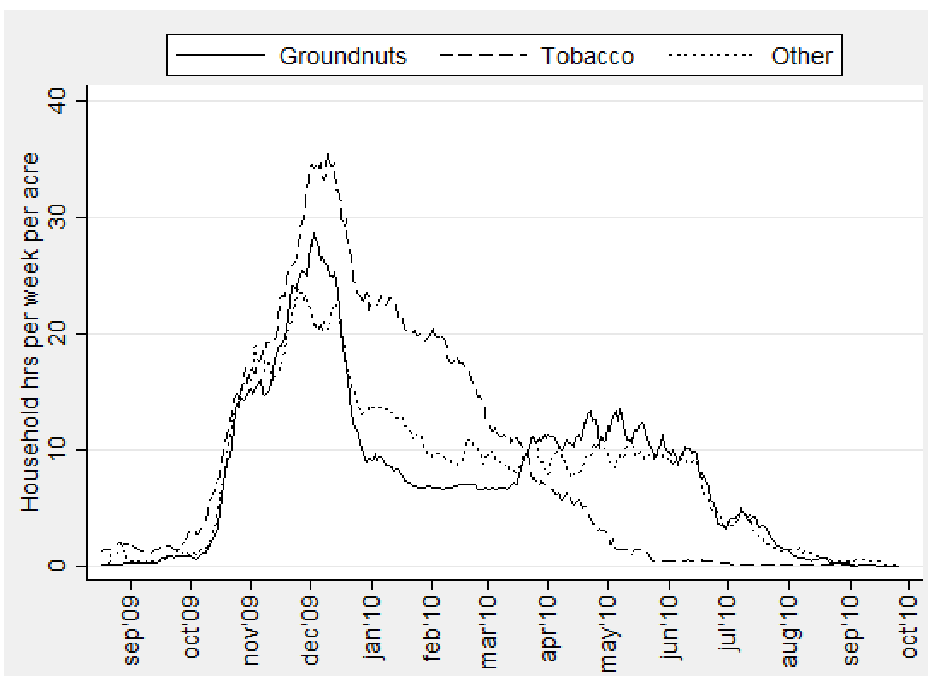


Figure 2.3: Allocation of time across activities in rural areas during the low season

Note: Sample consists of 2703 rural individuals interviewed in July and August. 930 individuals (34 % of the sample) who report working no hours are not included in the table.



(a) Maize and intercropped maize



(b) Non-maize

Figure 2.4: Estimated labor demand per week for an acre of the crop

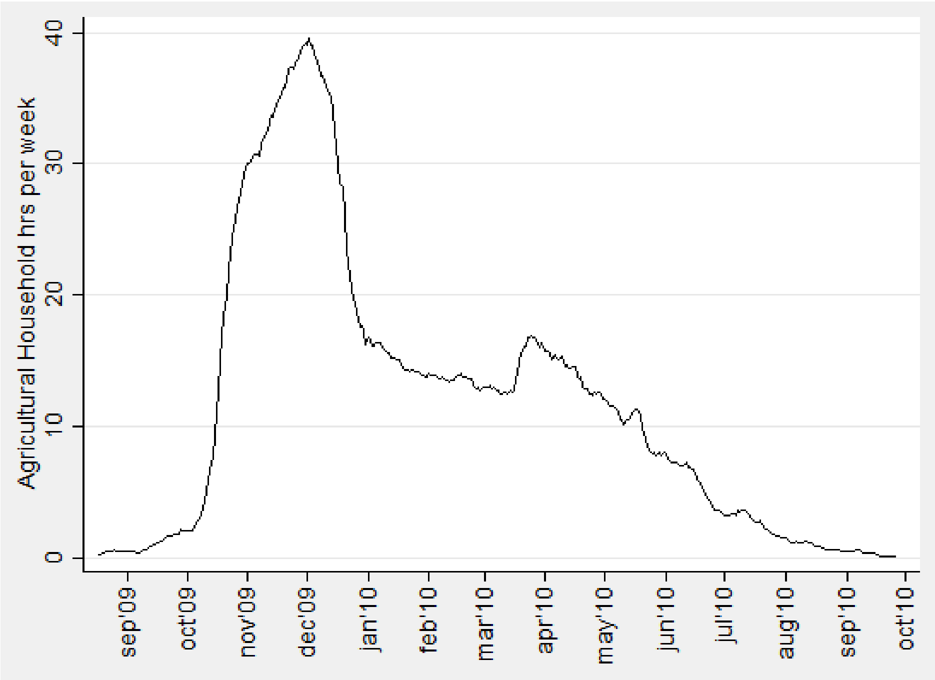
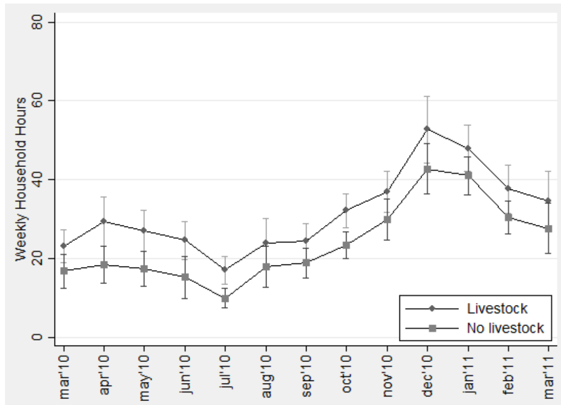
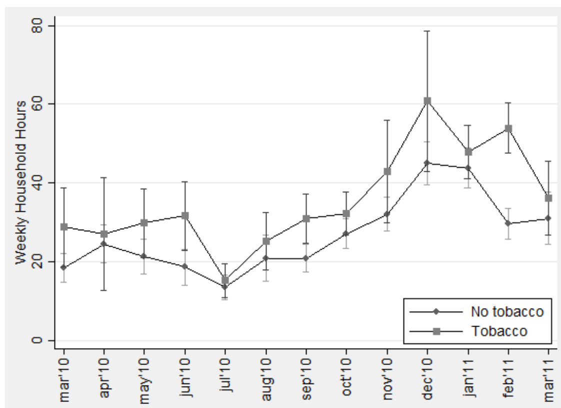


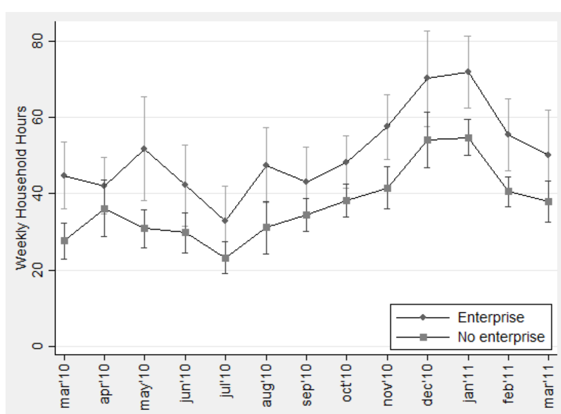
Figure 2.5: Estimated household agricultural labor demand per week for farming households using the retrospective agricultural questionnaire



(a) Household hours in agriculture by ownership of livestock



(b) Household hours in agriculture by tobacco cropping



(c) Total household hours by presence of household enterprise

Figure 2.6: Labor supply by household activites

Chapter 3

Production Volatility and Nominal Wage Rigidities: Dams and Village Labor Markets

This paper investigates whether downward wage rigidity is more binding and has larger distortionary effects in Indian districts that face greater production volatility. Using the instrument developed for dam construction in Duflo and Pande (2007), I apply the identification strategy used in Kaur (2019) to estimate whether dam presence influences the magnitude of the distortions created by wage rigidities. Despite suggestive evidence, results are inconclusive. Results are likely confounded by non-linearity in the effect of dams, the effects of lagged rainfall on contemporaneous production, substitution towards water intensive farming strategies in irrigated areas that exposes producers to significant production drops in severe drought years, as well as high inflation in the sample years.

3.1 Introduction

Understanding the impact of risk is central to the study of economic development. As in all economies, the economic impacts of risk in the developing world are altered by behavioral responses. These behavioral responses can in turn induce economic distortions. It is thus plausible that behavioral responses to unexpected changes would have more profound consequences in economies where individuals are more exposed to risk. While this may be evident, studying the interaction between risk and behavior is challenging, particularly if the focus is on a behavioral response that influences prices rather than an individual decision. While researchers may have methods to induce changes to individual risk exposure and thus influence individual response to risk, exogenous manipulation of the risk faced by an entire economy is much more difficult. Because of this it is challenging to identify how risk may be interacting with nominal wage rigidities to influence an aggregate variable such as wages.

In this paper I analyze the impact of risk on nominal wage responses to exogenous rainfall shocks by employing an instrumental variable strategy to identify risk reduction from irrigation dams. I use the instrument for irrigation dam construction developed in Duflo and Pande (2007) that uses exogenous geographic and topographic attributes of districts in India to predict the cost and availability of irrigation. Since irrigation should change the magnitudes of production deviations, the distortions created by nominal wage rigidities identified in Kaur (2019) would vary in severity. Working with these tools I investigate whether nominal wage rigidity has a different effect in districts containing dams, districts downstream of dams and districts that are not affected by dams due to changed variability in agricultural productivity. Building on the results of Kaur (2019) and Duflo and Pande (2007), I examine how rainfall shocks affect production in different areas and incorporate this into an analysis of the asymmetric response of wages to productivity shocks.

The paper is structured as follows: Section 3.2 presents background on the findings of Duflo and Pande (2007) as well as those of Kaur (2019) and then outlines the hypothesis under investigation. Section 3.3 presents the data. Section 3.4 outlines my empirical strategy. Section 3.5 presents results starting with suggestive evidence followed by the estimation of the main specification. Section 3.6 then presents an investigation of confounding factors starting with a discussion of the changing marginal returns to dams, and followed by an exploration of the lagged effects of rainfall on productivity. Section 3.7 concludes. Readers unfamiliar with Duflo and Pande (2007) and Kaur (2019) are advised to consult the appendices. Appendix C.1 provides a more in depth discussion of the instrument for dams. Appendix C.2 replicates relevant results from Duflo and Pande (2007). Appendix C.3 replicates relevant results from Kaur (2019) and Appendix C.4 presents a discussion of inflation's effects on my estimates.

3.2 Background and Hypothesis

In her job market paper on downward wage rigidity, Kaur shows that rainfall shocks induce productivity shocks which in turn lead to wage adjustments. She then shows that these wage adjustments are asymmetric: wages adjust upward following positive shocks but do not symmetrically adjust downwards following negative productivity shocks. Furthermore, she finds evidence of wage ratcheting as wages do not downward adjust in the year that follows a positive shock. Thus wage increases induced by positive rainfall shocks are persistent. The failure to downward adjust leads to above equilibrium real wages and generates a 9% reduction in employment levels. Production variability induced by rainfall shocks underlies and triggers these observed effects of nominal wage rigidity. Thus, technologies that change the variability of agricultural production should interact with this mechanism, leading to different impacts on wages and unemployment.

Dams would be one such technology. Duflo and Pande show that Indian districts downstream of a dam have higher agricultural production and that production there is less sensitive to rain shocks. Conversely, districts containing the dam and thus part of the dam's catchment area experience smaller non-significant increases in production and increased sensitivity to rain shocks. Given these difference in production variability, we would expect that the effects of wage rigidity would differ in dam affected areas compared to areas unaffected by dams. Figure 3.1 illustrates this hypothesis. Let the distribution illustrated in panel b represent the percent change in nominal wages from one year to the next in a district that is not affected by dams. In an economic environment with positive real economic growth and inflation, the mean of this distribution will be positive and centered on the long run nominal wage growth rate. Year to year production volatility, induced in this case by rainfall shocks, implies that in some years, nominal wages should be lower than those in the previous years. These observations fall in the shaded area of the distribution. In the presence of nominal wage rigidities as modeled by Kaur, employers in a competitive labor market may opt not to cut nominal wages in these years to avoid employee effort reductions. This behavior would induce bunching at 0% (illustrated as the black bar).

Under different production volatility regimes, the magnitude of this effect should vary. Less production volatility, such as what is observed in districts downstream of dams, should tighten this distribution as illustrated in panel a. The number of observations affected by downward wage rigidity should be smaller. The converse should hold for dam containing districts that experience higher production volatility as illustrated in panel c.

Given this hypothesis, I would expect districts downstream from dams to experience smaller wage increases (decreases) following positive (negative) rainfall shocks and the effects of lagged positive rainfall shocks on wages to be smaller. The converse would hold for dam containing districts. Finally, as in Kaur (2019), these effects should be mitigated by higher inflation which is explored in appendix C.4.

This mechanism potentially adds another dimension to the redistributive effects of dams discussed by Duflo and Pande. Indeed, Kaur identified significant employment costs induced by these distortions. In the same way that in developed countries, workers in durable goods industries, a sector particularly sensitive to business cycles, are more vulnerable to recession induced unemployment, the employment costs identified in Kaur (2019) may be particularly severe in districts highly exposed to rainfall induced production volatility. Establishing whether nominal wage rigidity distortions differ based on dam presence is the first step in estimating the distribution of these significant employment costs. It is also worth noting that while dams are being used to identify this mechanism, this mechanism may hold for any technology that changes aggregate production volatility and thus findings are not necessarily unique to this particular context.

3.3 Data

Analysis of this question is done using the datasets that were constructed for Duflo and Pande (2007) and the World Bank Agriculture and Climate Dataset data used in Kaur (2019). Though the World Bank Agriculture and Climate Dataset is available for 1956-1987, the data on dam construction from 1956-1970 is used to generate the instrument for dams. Thus my analysis will be limited to data from 1971-1987. Detailed descriptions of the primary sources of data that were used in the construction of these datasets are below.

Districts: In her study of the effect of rain shocks on wages, Kaur focuses on districts in the World Bank Agriculture and Climate Dataset where over 0.5% of crop area is planted with rice. This removes 31 of the 271 World Bank districts, principally in the states of Gujarat and Rajasthan from her analysis. It is worth noting too that the World Bank dataset does not cover the states of Kerala, Assam, Himachal Pradesh, Jammu and Kashmir, Manipur, Meghalaya and Tripura. Districts covered by the world bank data cover over 85% of India. Most of the omitted states, with the exception of Kerala and Assam, are some of the least important from the agricultural perspective (Sanghi, Kumar, and McKinsey Jr (1998)). Analysis is conducted using 1961 district boundaries as in the World Bank Dataset.

Geography: Geographic data likely to influence dam construction was assembled by Duflo and Pande (2007) and is available on the Harvard Dataverse. This data was originally sourced from GIS files processed by CIESIN, Earth Institute Columbia University. Data includes measures of river gradient and district topography. These same variables are also identified for upstream and downstream districts as well as other neighboring districts.

Dams: Data on dam construction was assembled by Duflo and Pande (2007). This data was originally sourced from the World Registry of Large Dams maintained by the International Commission on Large Dams. Dams qualify as large if they are 15 m or more in height or, if between 5 and 15 m have a reservoir capacity greater than 3 million cubic meters. The assembled data covers yearly dam construction between 1956 and 1999 with an additional year of data for 2004.

Agriculture: Agricultural data is from the World Bank Agriculture and Climate Dataset (Sanghi, Kumar, and McKinsey Jr (1998)). It provides agricultural data for 271 districts for agricultural years 1956-1987 using 1961 district boundaries. Variables include output, input, yield, area, prices and other indicators for 20 crops. This data also includes annual data on district level nominal wages for agricultural laborers.

Prices: Kaur (2019) includes several district level measures of inflation in her dataset. These are built using the state level price index from Ozler, Datt, and Ravallion (1996). District level inflation indicators are constructed to capture the national level of inflation without being affected by local shocks. This is calculated by taking the average level of inflation over all states excluding the state in which a district is located in.

Rainfall: The rainfall data used in Kaur (2019) comes from the University of Delaware Air Temperature and Precipitation dataset (version 4.01). Data was constructed by Kenji Matsuura and Cort J. Willmott using a spatial interpolation algorithm of neighboring weather station data. Monthly precipitation data is available for 1901-2014 on a 0.5 by 0.5 degree coordinate grid. Measures of rainfall focus on rainfall in the first month when the monsoon typically hits a district as the rainfall realization and timing in this month are important determinants of agricultural outcomes. Shocks are defined as rainfall realizations that fall in the highest and lowest quantiles of rainfall realizations for that district.

3.4 Empirical Strategy

To identify the differential impacts of production shocks, I instrument for irrigation dams following Duflo and Pande. I estimate

$$D_{dst} = \alpha_1 + \sum_{k=2}^{4} \alpha_{2k}(RGr_{kd} * \overline{D_{st}}) + \alpha_3(M_d * \overline{D_{st}}) + \sum_{k=2}^{4} \alpha_{4k}(RGr_{kd} * T_t) + \nu_d + \mu_{st} + \omega_{dst}, \tag{3.1}$$

where D_{dst} is the number of dams in district d, in state s at time t and RGr_{kd} is the fraction of district d's river gradient falling in category k(four gradient categories are used). These are interacted with $\overline{D_{st}}$, predicted dam incidence in the state and M_d is a vector of geographic controls including district elevation, gradients, river length and area. RGr_{kd} is also interacted with year dummies T_t to account for national time varying characteristics. Finally district and state by year fixed effects are included. The same strategy is estimated on upstream dams D^U_{dst}. The parameters estimated above are then used to generate a set of predicted values $\widehat{D_{dst}}$ and $\widehat{D^U_{dst}}$. Appendix C.1 presents a more complete description and explanation of the construction of this instrument and estimates generated by equation 1. Appendix C.2 replicates relevant results from Duflo and Pande (2007), verifying the impact of dams on the variability of agricultural production in response to rainfall.

To identify nominal wage rigidity, I follow the identification approach used in Kaur (2019) which evaluates wage responses to lagged rainfall shocks. Kaur defines a positive (negative) rainshock as rainfall falling within the upper (lower) 20% of a district's rain realizations at the start of the monsoon season. Over a two year period there are 9 possible sequences of rain shocks: (0,0), (-,0), (+,0), (0,-), (-,-), (+,-), (0,+), (-,+), (+,+). Under assumptions of downward nominal wage rigidity, several of theses realizations yield similar predicted effects on current nominal wages and are thus grouped together in 5 shock groups (S_d): Pos_{dt}, which will equal 1 in districts experiencing contemporaneous positive shocks (i.e realizations (+,+); (0,+); (-,+)), $NonPos_{d,t-1} * Neg_{dt}$ whcih will equal 1 in districts experiencing (0,-) and (-,-) shock sequences; $Pos_{d,t-1} * Neg_{dt}$ for districts with (+,-) shock sequences and $Pos_{d,t-1} * Zero_{dt}$ when a district faces (+,0), and finally the omitted category is districts facing the shock sequences (0,0) and (-,0). lnw_{dt}, the log of nominal wages in district d in year t is then regressed on these groups of rainfall realizations with controls for district (ν_d) and year (τ_t) fixed effects and $\sum_{k=2}^{K} \phi_k Pos_{d,t-k}$ which controls for positive shocks 2 and 3 years ago,

$$lnw_{dt} = \beta_0 + \beta_1 Pos_{dt} + \beta_2 NonPos_{d,t-1} Neg_{dt} + \beta_3 Pos_{d,t-1} Neg_{dt} + \beta_4 Pos_{d,t-1} Zero_{dt} + \sum_{k=2}^{K} \phi_k Pos_{d,t-k} + \nu_d + \tau_t + \epsilon_{dt}. \tag{3.2}$$

Kaur predicts that the coefficient on positive shocks will be positive, but that with sufficiently severe downward rigidity, the coefficients on contemporaneous negative shocks will be zero and the coefficients on lagged positive shocks will be positive. Appendix C.3 replicates Kaur's estimation for relevant samples. Appendix C.4 replicates and discusses additional results from Kaur (2019) that are important to understanding the mitigating effects that inflation has on my 1971-1987 sample.

In order to estimate how the effects of wage rigidity might be dampened or amplified by dams, I combine the estimation technique of Duflo and Pande with Kaur's, adding interactions with instrumented upstream dams and within district dams to evaluate their effects. I regress the log of nominal wages, lnw_{dt}, on the 5 rainfall shock groups, S_{dt}, $\widehat{D_{ist}}$ and $\widehat{D^U_{ist}}$, the instruments for dams (D_{dt}) and upstream dams (D^U_{dt}), the interactions between the rainfall groups and the dam instruments, Z_{ist} and Z^U_{ist} which are the right hand side variables in equation 3.1 (interactions $RGr_{ki} * \overline{D_{st}}$ excepted), controls for positive shocks 2 and 3 years ago, as well as district and state by year fixed effects. Thus I estimate

$$\begin{aligned} Lnw_{dt} = \delta_0 &+ \sum_{k=2}^{5} \delta_{1k} S_{kdt} + \delta_3 D_{dt} + \delta_4 D^U_{dt} \\ &+ \sum_{k=2}^{5} \delta_{4k} D_{dt} * S_{kdt} + \sum_{k=2}^{5} \delta_{5k} D^U_{dt} * S_{kdt} \\ &+ \delta_6 Z_{dt} + \delta_7 Z^U_{dt} + \sum_{k=2}^{K} \delta_8 k Pos_{d,t-k} + \nu_d + \mu_{st} + \epsilon_{dt} \end{aligned} \tag{3.3}$$

If risk exposure affects nominal wage distortions, the interaction term for in district dams should have the same sign as the corresponding rain shock (thus amplifying the effects on nominal wage) while the interaction term for upstream dams should have the opposite sign (mitigating the distortions). Indeed, in districts that are irrigated thanks to an upstream dam, we would expect contemporaneous rainfall shocks to have a reduced effect on production, and thus wages. Similarly, because of this, the pattern of wages being ratcheted up after a positive shock would also be mitigated.

3.5 Results

Dams and the Distribution of Wage Changes

Given the potential difficulties in identifying evidence of how dams might interact with nominal wage rigidity, it is informative to consider some suggestive evidence. I follow the methodology employed by Sarsons (2015) and construct indicator variables $\widehat{nodams}$, $\widehat{damsin}$ and $\widehat{damsup}$ based on the predicted number of dams in a district and it's upstream districts. The $\widehat{nodams}$ indicator is set to 1 if a district is predicted to have fewer than 10 dams within and upstream of the district. The $\widehat{damsin}$ indicator is set to 1 if a district is predicted to have over 10 dams within and but fewer than 10 upstream of the district. Finally, $\widehat{damsup}$ indicator is set to 1 if a district is predicted to have over 10 dams upstream but fewer than 10 within the district. Note that not all districts are covered by these indicators. Omitted are districts that are predicted to have both more than 10 dams within and upstream.[1]

In figure 3.2 and the accompanying table 3.1, observations are divided based on the indicator variables $\widehat{nodams}$, $\widehat{damsin}$ and $\widehat{damsup}$ described above. Each graph shows the distribution of the percentage change in nominal wage from one year to the next for observations in that group. The bunching of observations at 0% is suggestive evidence that nominal wage rigidity is affecting wages in all three types of districts. Note however that the proportion of observations at the bunching point differs between the categories of districts. Table 3.1 presents descriptive statistics regarding these distributions and the bunching. The mean percentage change in nominal wage is similar across these three types of districts at approximately 10%. While this may seem high, given that inflation during this time period is close to 7.5% this suggests an average real wage growth in the range of 2.5%. The standard deviation of these distributions is largest for $\widehat{damsin}$ districts while it is smallest for $\widehat{damsup}$ districts. Given the results in Duflo and Pande (2007) on how dams affect the variability of agricultural production, if nominal wage changes reflect changes in production this pattern is consistent with their findings.

To generate some descriptive information on bunching, I calculate the the percent of observations in each category where the percentage change in nominal wage was between 0 and 1% included. I also calculate the ratio of the actual number of observations falling in the 0-1% range over the expected number of observations. The expected number of observations is estimated by first taking the number of observations falling between -2% and 0 divided by 2 and the number of observations falling between 1% and 3% divided by two and then taking the average of these two values. The percent and ratio are reported in table 3.1. These descriptive statistics are consistent with the hypothesis that nominal wage rigidity will generate larger distortion effects in districts where production, and wages, are more volatile. The percent of observations falling at or just above 0% is proportionally larger in districts containing dams and smaller in districts downstream of dams. Indeed, districts with dams upstream appear to only have a small amount of bunching at 0% with a little over twice the

[1]This amounts to 4,459 district year observations out of 7,675.

number of observations falling in that range as might be expected. On the other hand, for dam containing districts, the 0% point has over 5 times the number of observations expected.

While this evidence is only suggestive, it is consistent with the hypothesis that areas subject to high production volatility will be more affected by downward wage rigidity.

Estimation

Given the suggestive evidence that dam presence may affect nominal wage distortions, I proceed to estimate equation 3.3. Table 3.2 presents my results. Columns 1 and 2 replicate the results from Kaur (2019) for the estimation specified in equation 3.2. Columns 3 and 4 apply estimation 3.3 to the data from 1971-1987 which overlaps with the availability of the dams instrument. Columns 5 and 6 add the interactions between rainfall shocks and instrumented dam presence as specified in equation 3.3.

Results from the estimation of equation 3.3 in columns 5 and 6 are imprecise and not statistically significant. The δ_{4k} and δ_{5k} coefficients on the interactions between dam presence and the sequences of rainfall shocks have large standard errors and are not statistically significant. I cannot reject that dam presence has no effect on how binding the effects of nominal wage rigidity are. Moreover, the δ_{1k} coefficients that test for nominal wage rigidity are substantially smaller in magnitude, as compared to the results in Kaur (2019), replicated in columns 1 and 2, and are no longer statistically significant at conventional levels.

The reduced magnitude of the δ_{1k} coefficients is better understood when I compare the estimates from columns 5 and 6 to the estimates in columns 3 and 4. Columns 3 and 4 apply the estimation strategy from equation 3.2 to the subset of data for which the dams instrument is available and that is used in columns 5 and 6. The coefficients reported in columns 3 and 4 are also smaller in magnitude and not statistically significant suggesting that nominal wage rigidity was not binding during that period. Indeed, a closer inspection of historical data suggests that the years from 1971-1987 coincide with a period of high inflation and price volatility in India, which would make the effects of nominal wage rigidity less pronounced and less detectable as nominal wage rigidity would be less binding. The role of inflation in generating the observed differences between columns 1 and 2 and columns 3 and 4 is further discussed in appendix C.4. The remainder of the paper investigates other potential confound that can explain the lack of results detected in the estimates of δ_{4k} and δ_{5k}

3.6 Confounding Effects

I investigate two implicit assumptions that potentially could be confounding my results in the analysis above. The first is that there are linear marginal returns to dams and the second is the assumption that the productivity impacts of rainfall are not persistent from one year to the next.

Marginal Returns to Dams

The estimation strategy detailed in equation 3.3 that incorporated the interactions of the rainfall shocks with the continuous variables for dams is implicitly estimating a linear effect for these variables. This would be appropriate if we believe that the marginal effect of the first ten dams will be comparable to the marginal effect of the the last ten dams constructed in a district. This seems unlikely. Indeed, intuition would suggest that policy makers would elect to build the first dams in areas where a dam would provide the highest returns, i.e. in locations with either low construction costs, and/or high impacts on agricultural productivity. It thus seems plausible that the marginal return to dams would not be linear, and may possibly exhibit diminishing returns. In this case, estimating a linear effect of dams would add significant noise to my estimation, confounding the results. Furthermore, because the instrumental variables strategy relies on early dam construction from 1957-1971 to instrument for later dam construction from 1971-1987, results in columns 5 and 6 of table 3.2 would be estimating the effects of the more marginal dams.

To test this I regress log production and log wages on predicted dams in district and then separately for dams upstream of a district, D_{dt} and D^U_{dt}, instrumented by $\widehat{D_{dt}}$ and $\widehat{D^U_{dt}}$, as well as the quadratic of D_{dt} and D^U_{dt}. I include the relevant instrumental variable controls, Z_{dt} and Z^U_{dt} as well as district and state by year fixed effects. Thus, separately for both D_{dt} and D^U_{dt}, I estimate

$$\begin{aligned} y_{dt} &= \pi_0 + \pi_1 D_{dt} + \pi_3 Z_{dt} + \nu_d + \mu_{st} + \omega_{dt} \\ y_{dt} &= \tilde{\pi}_0 + \tilde{\pi}_1 D_{dt} + \tilde{\pi}_2 D^2_{dt} + \tilde{\pi}_3 Z_{dt} + \nu_d + \mu_{st} + \omega_{dt}. \end{aligned} \tag{3.4}$$

Results for the estimation outlined in equation 3.4 are reported in table 3.3. The odd numbered columns estimate a linear effect of dams and upstream dams, as detailed in the first equation of specification 3.4, while the even columns incorporate the quadratic term as detailed in the second equation of specification 3.4. Estimates for dams in a district are highly imprecise with large standard errors suggestive of their ambiguous effects, as discussed in Duflo and Pande (2007). Estimates for upstream dams on log agricultural production in columns 5 and 6 are statistically significant and suggestive of diminishing returns to dams in the upstream district.

The difference in these coefficients would suggest that going from 0 to 10 upstream dams would increase production by 18.8% rather than 4% using the linear model. Regarding the changing marginal effects, a district going from 0 to 10 upstream dams would see estimated agricultural production gains of 18.8% whereas there would be no additional returns to upstream dams beyond 177 upstream dams. To test whether these diminishing returns might be affecting my estimates in table 3.2, I repeat the same estimations but on log nominal wages (columns 3, 4, 7 and 8). Once again, in district estimates are highly imprecise with large standard errors. Estimates for upstream dams, follow the expected pattern given the results in column 6, though the magnitude of the coefficients are much smaller and they are not statistically significant. Note though that when adjusting for non-linearity, the coefficient on the linear term changes signs. It is plausible that the non linearity of dam effects may be confounding the estimated effects of dams on nominal wages in table 3.2.

Lagged Rainfall Effects on Productivity

Kaur's approach requires the assumption that the productivity impacts of rainfall are not persistent and that rain induced productivity shocks are determined entirely by the current year's rain realization. By focusing on monsoon shocks and then relating her findings to inflation and employment Kaur can show that TFP is not driving her results. While this is a standard assumption in prior work on rain and agriculture (Paxson 1992, M. R. Rosenzweig and Wolpin 1993, Townsend 1994, Jayachandran 2006), other studies have found it necessary to correct for lagged production effects of rain (M. Rosenzweig and Udry (2013)). As irrigation dams act to smooth water availability between different time periods, it seems likely that the agricultural productivity of districts irrigated by dams would be affected by lagged rainfall shocks in addition to current rainfall shocks in a way that differs from districts unaffected by irrigation dams. Thus, given that dams are, by design, built to generate precisely these lag effects, the assumption that there are no lagged productivity effects of rainfall is inappropriate for the question under consideration here. Thus the interpretation of differences in coefficients from equation 3.3 for districts differently affected by dams becomes ambiguous.

Comparing the magnitude and direction of the interaction coefficients in table 3.2 will not allow me to distinguish between dams differential effects on nominal wages from differential productivity effects. Coefficient differences between downstream, dam containing and unaffected districts could reflect nominal wage rigidities or the effects of lagged rainfall on productivity. Take for instance, the positive coefficient on the Dams Upstream*[+;-] interaction. All else equal, if nominal wage rigidity is less binding in downstream districts we would expect this coefficient to be negative. However, having dams upstream would be particularly beneficial for contemporaneous production after the [+,-] rainfall realization as reservoirs would be filled in the first year and help avoid a production shock from droughts in the second. It becomes apparent that more work will be required to disentangle potential production effects of lagged shocks from evidence of nominal wage rigidity.

To attempt to disentangle lagged productivity effects from wage rigidity effects, I estimate the effects of positive and negative lagged rainfall shocks on nominal agricultural wages, agricultural production, weighted yields, cropped area, fertilizer use and the area planted with HYV varieties. I regress these outcomes on the same positive and negative rainfall shock indicators $S_{dt}^{+,-}$ and include historic shock indicators for the years $t-1$ and $t-2$ as well as district and year fixed effects, as detailed in the following specification,

$$y_{dt} = \rho_1 + \rho_2 S_{dt}^{+} + \rho_3 S_{d,t-1}^{+} + \rho_4 S_{d,t-2}^{+} + \rho_5 S_{dt}^{-} + \rho_6 S_{d,t-1}^{-} + \rho_4 S_{d,t-2}^{-} + \nu_d + \mu_t + \epsilon_{dt}. \quad (3.5)$$

In addition to estimating this specification on the full 1956-1987 sample, I also run this estimation of the 1971-1987 subset as well as on the $\widehat{nodams}$, $\widehat{damsin}$ and $\widehat{damsup}$ subsamples, to see how dams impact the lagged effects of rainfall shocks. Results are presented in tables 3.4, 3.5 and 3.6.

Table 3.4 presents the results of equation 3.5 on log nominal wages and log production. Consistent with the null results observed in table 3.2, past positive rainfall shocks positively impact current nominal wages, though primarily in the 1956-1987 sample, with no clear difference based on dam presence. Past negative rainfall shocks have little effect on current nominal wages, which is consistent with downward nominal wage rigidity. Regarding the impacts on production, negative shocks have contemporaneous and sustained lagged effects on production in all of the samples except the $\widehat{damsin}$ which experiences a positive productivity effect from past positive rainfall shocks. Positive increases from contemporaneous and lagged positive rainfall shocks appear for the 1956-1987 sample but not for the 1971-1987 sample. The magnitude of these effects is quite large and larger in the latter sample. In the 1971-1987 sample, an annual rain realization in the lowest quantile decreases production by 5% that year and further decreases production by 5% the subsequent year and 9% in the year after. These results suggests that negative rainfall realizations two years ago can actually have a substantial effect on current production levels.

The negative coefficients on production for districts downstream of dams are somewhat surprising. While the coefficient on lagged negative shocks may be explained by reservoir depletion, somewhat surprisingly these districts have a coefficient on contemporaneous negative shocks that is comparable, and possibly even larger, than districts unaffected by dams. While this seems difficult to reconcile with Duflo and Pande's rainfall results, it may follow from their findings that crops shift towards more water intensive crops in downstream districts. Since here we focus on exceptionally bad rainfall realizations, it may be that while dams help stabilize production in most years, changes in cropping patterns may still leave these districts vulnerable, or even increase their vulnerability, to extreme negative rain shocks.

While factors such as soil moisture may explain some of these effects (M. Rosenzweig and Udry (2013)), a significant part of of this effect is due to changes in input use following negative shocks. The results in table 3.5 explore the intensive and extensive margins of input use that could be driving the observed change in production. I consider whether the observed changes in production are driven by changes in yields or changes in the amount of area that is cultivated. Indeed, past rainfall realizations could influence soil moisture as well as fertilizer and seed usage leading to increased yields on existing farm plots. Past rainfall could also change household decisions on weather cultivating a plot in a given year would be profitable. Inspection of the coefficients reported in table 3.5 suggest that both of these mechanisms contribute to the observed reduction in productivity following negative rain shock years. Yields are lower in years that follow negative rainfall shocks. Similarly, the area under cultivation is also reduced in years that follow negative rainfall shocks. Interestingly, there do not appear to be positive effects following positive rainfall realizations on yields. There is some evidence of a positive effect of positive rainfall shocks on cultivated area in the 1956-

1987 sample. Interestingly, in areas that are unaffected by dams, there is a strong negative effect of about 2% for contemporaneous and past positive rainfall realizations on the area cultivated. This may be due to vulnerability of these regions to flooding in high rainfall years.

The impact of rainfall on yields is particularly interesting as it could be due to the direct effects of rainfall (water and soil moisture are important farming inputs), but could also be due to changes in the usage of other farm inputs, such as fertilizer usage or the use of high yield variety seeds. The use of these purchased inputs could depend on past rainfall realizations if farmers, in a credit constrained environment, use revenues generated by the previous year's crops to purchase inputs for the following year's crops.

Table 3.6 reports results for equation 3.5 on log fertilizer use and the area planted to high yield varieties. Fertilizer use appears to be strongly dependent on contemporaneous, and past rainfall shocks. Looking at results for the full 1956-1987 sample, contemporaneous and previous rainfall shocks result in about a 6% increase in fertilizer usage. However if we focus on the 1971-1987 sample, positive rainfall shocks have a smaller effect on fertilizer use and instead we see a strong response to negative rainfall realizations which lead to substantial decreases in fertilizer application of 7% for contemporaneous rainfall but that persist as past negative rainfall still reduces fertilizer application by 4.6% two years later. The difference between the results for the 1956-1987 and the 1971-1987 samples may be due to the timing of the widespread adoption of fertilizer in India. In the earlier period of the 1956-1987 sample, fertilizer use may not have been the norm and positive rainfall shocks could have induced the uptake of this new technology. In the later period from 1971-1987, fertilizer was more broadly adopted but its application might have been constrained if poor rainfall in previous seasons limited credit constrained producers ability to purchase fertilizer inputs. Regarding differential effects by dam presence, the broad patterns for all three types of districts are broadly similar to those found in the whole sample for 1971-1987 and I am underpowered in these smaller subsample to detect any clear differences in fertilizer usage.

Finally, the results looking at the area cultivated with high yield varieties are inconclusive. I cannot reject that contemporaneous and past rainfall realizations have no effect on the planting of high yield varieties.

To further distinguish the nominal wage responses that are not explained by productivity differences, table 3.7 presents estimates of the five rain shock effects on both nominal wages and production for the five samples as specified in equation 3.2. Table 3.8 reports $\beta_{shock}^{nom.wage} - \beta_{shock}^{prod.}$, the differences between the effect of a shock sequence on nominal wages from the effect the shock has on production.

Note that for the samples other than the 1956-1987 sample, none of the coefficients on log wages are statistically different from zero. Few of the coefficients on log production are statistically different from zero for these samples with the exception of the coefficients on $[(0,-);-]$ which are negative and statistically significant for four of these subsamples. Once again, the evidence suggests that, counter intuitively, production in districts affected by dams, both within and upstream, is particularly vulnerable to negative shocks, especially two consecutive negative shocks. On the other hand, and also somewhat counter intuitively, contemporaneous positive shocks appear to matter less for production in the later sample.

Regarding evidence of wage rigidities, the differences in the effect of a shock sequence on nominal wages from the effect the shock on production as reported in table 3.8 is consistent with nominal wage rigidity for the 1956-1987 sample. Changes in nominal wages appear to reflect changes in contemporaneous production, but not after the $[+;-]$ shock sequence when nominal wages are higher than expected due to ratcheting.[2] While we do not see the same pattern of upward ratcheting appear for the 1971-1987 sample, the failure of nominal wages to downward adjust during $[(0,-);-]$ shocks, which have particularly severe negative effects on production in this period is consistent with downward wage rigidity. Regarding how dams may interact with the relationship between production and nominal wages, estimates on the smaller samples reported in columns 3, 4 and 5 as well as in 8,9 and 10 of table 3.7 are quite imprecise with large standard errors. The differences between these coefficients reported in columns 3 ,4 and 5 of table 3.8 are thus difficult to interpret and inconclusive.

3.7 Conclusion

While there is suggestive evidence that dam presence makes nominal wage rigidities more or less binding, several of the assumptions required for identification proved problematic. First, by imposing a linear functional form on dam productivity effects, simple interactions with dam presence masks interesting differences in the marginal effects of dams, potentially confounding some results. Secondly, given the nature of dam technology, assuming rain shocks do not generate lagged effects on productivity seems too strong of an assumption for this research question.

Investigations into lag effects of rain shocks reveal further identifications challenges. First, Duflo and Pande's findings that upstream dams reduce sensitivity to rainfall may hide a much more complicated relationship. Results here suggest production in districts downstream of dams is actually very sensitive to negative rain shocks in the lowest quantile, likely as a result of input and crop responses to dam presence. Thus it is possible that while upstream dams may reduce production variability in response to most rainfall realizations, leading farmers to use more water dependent crops and inputs, they offer insufficient protection against the worst rainfall realizations. When paired with increased water dependent farming this could generate the results observed here for severe rain shocks and reconcile these findings with those of Duflo and Pande. While this complicates my identification strategy, that uses these extreme rainfall realizations, the idea that producers may be optimizing around a technology that reduces smaller, more frequent losses by selecting production strategies that increase average productivity but expose them to large but infrequent losses is interesting. Secondly, dam presence appears to affect input selection. If these inputs, such as fertilizer, are themselves complementary to rainfall, the relationship between rainfall shocks, production and wages may be changing in a way that is independent of the wage rigidity effects studied

[2]Note that evidence of ratcheting following a $[+;0]$ does not hold using this approach.

here, thus confounding this analysis. This is even further complicated by the fact that input use also responds to lag rain shocks.

It is clear that dams have changed the exposure of producers and workers to risk in a fundamental way. Though I cannot currently make a definitive statement on how nominal wage rigidity specifically was affected by this change, the confounding forces uncovered while trying to identify this effect are interesting in and of themselves and hold promise for avenues of future research.

Tables

Table 3.1: Distributions of the Yearly Percent Change in Nominal Wage by Dam Presence

	All Districts	Without Dams	Dams Upstream	Dams Within
Mean	0.08	0.10	0.10	0.10
Standard Deviation	0.15	0.12	0.11	0.14
Percent of Observations at Zero	9.24	3.25	2.22	4.76
Ratio at Zero to Expected	0.18	0.08	0.05	0.10
N	7675	2199	570	447

Note: Observations are divided based on the predicted number of dams within and upstream of a district. If a district is predicted to have fewer than 10 dams within and upstream it is classified as being without dams. If a district is predicted to have over 10 dams upstream but fewer than 10 within the district is classified as having upstream dams. If a district is predicted to have over 10 dams within and but fewer than 10 upstream it is classified as having dams within. Districts that are predicted to have both more than 10 dams within and upstream are omitted. Percent of observations at zero is the percent of observations where the percentage change in nominal wage was between 0 and 1% included. The number of observations expected to be at zero is calculated as taking the number of observations falling between -2% and 0 divided by 2 and the number of observations falling between 1% and 3% divided by two and then taking the average of these two values.

Table 3.2: Dam Impacts on Nominal Wage Rigidity

	Log Nominal Daily Wages					
	(56-87)	(56-87)	(71-87)	(71-87)	(71-87)	(71-87)
Last Year: (0,-,+); This Year: (+)	0.026^{***}	0.042^{***}	0.003	0.010	0.017	0.012
	(0.009)	(0.010)	(0.009)	(0.010)	(0.013)	(0.015)
Last Year: (0,-); This Year: (-)	-0.011	-0.014	-0.019	-0.020	-0.00001	0.001
	(0.010)	(0.010)	(0.013)	(0.013)	(0.015)	(0.015)
Last Year: (+); This Year: (-)	0.035^{*}	0.052^{**}	-0.004	0.003	-0.018	-0.024
	(0.020)	(0.021)	(0.020)	(0.020)	(0.020)	(0.021)
Last Year: (+); This Year: (0)	0.020^{**}	0.037^{***}	0.0005	0.008	0.015	0.010
	(0.010)	(0.011)	(0.010)	(0.011)	(0.019)	(0.021)
Dams in District					0.940^{*}	0.939^{*}
					(0.480)	(0.480)
Dams Upstream					-0.042	-0.040
					(0.066)	(0.066)
Dams in Dist.*[(+,0,-);+]					-0.042	-0.036
					(0.124)	(0.126)
Dams Upstream*[(+,0,-);+]					-0.081	-0.080
					(0.056)	(0.056)
Dams in Dist*[(0,-);-]					-0.147	-0.142
					(0.145)	(0.145)
Dams Upstream*[(0,-);-]					0.055	0.052
					(0.057)	(0.057)
Dams in Dist*[+;-]					-0.094	-0.085
					(0.322)	(0.323)
Dams Upstream*[+;-]					0.135	0.135
					(0.148)	(0.148)
Dams in Dist*[+,0]					0.003	0.015
					(0.247)	(0.252)
Dams Upstream*[+;0]					-0.071	-0.071
					(0.080)	(0.080)
Prior Shock History Controls?	No	Yes	No	Yes	No	Yes
District Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	-	-
State by Year Fixed Effects	No	No	No	No	Yes	Yes
Observations	7,675	7,675	4,080	4,080	4,063	4,063

Note: Standard errors in parentheses clustered by region by year. $^{*}p < 0.10,^{**}\ p < 0.05,\ ^{***}\ p < 0.01$. Shocks are grouped by common predictions. Indicator variables are set to 1 when a shock sequence in the group is realized with $[(-, 0); 0]$ as the omitted category. The first shock listed is the shock in $t-1$, the second is the contemporaneous shock. Columns 5 and 6 include results using the 2SLS strategy for dams. These regressions also include predicted dams*gradient interactions, geography controls and the same set of variables for upstream districts as well as an indicator for whether there is an upstream district. The number of dams is divided by 100 making coefficients multiplied by 100.

Table 3.3: Changing Marginal Returns to Dams

	Dependent variable:							
	Log Production		Log Wages		Log Production		Log Wages	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dams in District	0.471 (1.187)	−1.135 (7.829)	0.924 (1.081)	2.675 (6.906)				
Dams in District Squared		1.926 (8.236)		−2.100 (7.306)				
Dams Upstream					0.396** (0.195)	1.991** (0.774)	−0.068 (0.141)	0.042 (0.382)
Dams Upstream Squared						−1.123* (0.627)		−0.078 (0.275)
District Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State by Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4,063	4,063	4,063	4,063	4,063	4,063	4,063	4,063

Note: Standard errors in parentheses clustered by district. $^*p < 0.10$,** $p < 0.05$, *** $p < 0.01$. All regressions use the 2SLS strategy for dams and include district and state*year fixed effects. These regressions include predicted dams*gradient interactions, geography controls and the same set of variables for upstream districts as well as an indicator for whether there is an upstream district. The number of dams is divided by 100 making coefficients multiplied by 100.

Table 3.4: Lagged Rainfall Shocks on Agricultural Wages and Productivity

	Log Nominal Wages					Log Production				
	56-87	71-87	$\widehat{nodams}$	$\widehat{damsup}$	$\widehat{damsin}$	56-87	71-87	$\widehat{nodams}$	$\widehat{damsup}$	$\widehat{damsin}$
This Year (+)	0.025***	0.003	0.003	0.017	0.023	0.023***	−0.005	−0.036***	−0.032	0.041
	(0.006)	(0.006)	(0.009)	(0.017)	(0.024)	(0.008)	(0.010)	(0.012)	(0.028)	(0.037)
Last Year (+)	0.022***	0.006	0.0002	0.003	0.031	0.027***	0.005	−0.007	−0.071**	0.114***
	(0.006)	(0.006)	(0.009)	(0.017)	(0.024)	(0.008)	(0.010)	(0.012)	(0.028)	(0.037)
Year-2 (+)	0.029***	0.017***	0.009	0.013	0.016	0.030***	0.004	−0.019*	0.019	0.084**
	(0.006)	(0.006)	(0.009)	(0.016)	(0.024)	(0.008)	(0.010)	(0.012)	(0.027)	(0.037)
This Year (-)	−0.004	−0.015*	−0.020*	0.011	−0.059*	−0.027***	−0.052***	−0.036**	−0.093**	−0.091*
	(0.007)	(0.008)	(0.011)	(0.022)	(0.033)	(0.009)	(0.013)	(0.015)	(0.037)	(0.051)
Last Year (-)	0.007	−0.002	−0.004	0.008	−0.008	−0.022**	−0.047***	−0.034**	−0.079**	0.014
	(0.007)	(0.009)	(0.012)	(0.023)	(0.034)	(0.010)	(0.013)	(0.016)	(0.039)	(0.053)
Year-2 (-)	0.004	−0.002	0.0004	0.034	−0.030	−0.029***	−0.091***	−0.083***	−0.139***	−0.049
	(0.007)	(0.008)	(0.011)	(0.021)	(0.035)	(0.010)	(0.013)	(0.015)	(0.036)	(0.055)
District Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	7,436	4,080	2,199	570	447	7,409	4,063	2,199	570	447

Note: Standard errors in parentheses clustered by district. $^{*}p < 0.10$,** $p < 0.05$, *** $p < 0.01$. Each regression contains district and year fixed effects. In columns 3-5 and 8-10, regressions are estimated on subsets, determined based on the predicted number of dams within and upstream of a district. If a district is predicted to have fewer than 10 dams within and upstream, it is classified as being without dams. If a district is predicted to have over 10 dams upstream but fewer than 10 within the district is classified as having upstream dams. If a district is predicted to have over 10 dams within but fewer than 10 upstream it is classified as having dams within.

Table 3.5: Lagged Rainfall Shocks on Agricultural Yields and Cultivated Area

	Log Yields					Log Cultivated Area				
	56-87	71-87	$\widehat{nodams}$	$\widehat{damsup}$	$\widehat{damsin}$	56-87	71-87	$\widehat{nodams}$	$\widehat{damsup}$	$\widehat{damsin}$
This Year (+)	0.0002	−0.005	−0.017*	−0.026	0.004	0.011***	0.0002	−0.017***	0.009	0.006
	(0.007)	(0.008)	(0.009)	(0.022)	(0.033)	(0.003)	(0.005)	(0.006)	(0.013)	(0.014)
Last Year (+)	0.004	0.010	0.010	−0.048**	0.087***	0.007**	−0.003	−0.018***	0.001	0.016
	(0.007)	(0.008)	(0.009)	(0.022)	(0.032)	(0.003)	(0.005)	(0.006)	(0.013)	(0.014)
Year-2 (+)	0.009	0.006	0.002	0.003	0.047	0.003	−0.005	−0.022***	0.003	0.025*
	(0.007)	(0.008)	(0.009)	(0.022)	(0.032)	(0.003)	(0.005)	(0.006)	(0.013)	(0.014)
This Year (-)	−0.014*	−0.018*	−0.001	−0.051*	−0.087*	−0.015***	−0.029***	−0.031***	−0.023	−0.027
	(0.008)	(0.010)	(0.011)	(0.029)	(0.045)	(0.004)	(0.006)	(0.008)	(0.017)	(0.019)
Last Year (-)	−0.016**	−0.027***	−0.004	−0.038	−0.007	−0.006	−0.021***	−0.027***	−0.017	0.007
	(0.008)	(0.011)	(0.012)	(0.031)	(0.046)	(0.004)	(0.006)	(0.009)	(0.018)	(0.020)
Year-2 (-)	−0.023***	−0.063***	−0.044***	−0.097***	−0.053	−0.012***	−0.032***	−0.041***	−0.025	0.001
	(0.008)	(0.010)	(0.012)	(0.028)	(0.048)	(0.004)	(0.006)	(0.008)	(0.017)	(0.020)
District Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	7,409	4,063	2,199	570	447	7,391	4,046	2,187	570	442

Note: Standard errors in parentheses clustered by district. $*p < 0.10$,$** \; p < 0.05$, $*** \; p < 0.01$. Each regression contains district and year fixed effects. In columns 3-5 and 8-10, regressions are estimated on subsets, determined based on the predicted number of dams within and upstream of a district. If a district is predicted to have fewer than 10 dams within and upstream, it is classified as being without dams. If a district is predicted to have over 10 dams upstream but fewer than 10 within the district is classified as having upstream dams. If a district is predicted to have over 10 dams within but fewer than 10 upstream it is classified as having dams within.

Table 3.6: Lagged Rainfall Shocks on the use of Agricultural Inputs

	Log Fertilizer					HYV Area				
	56-87	71-87	$\widehat{nodams}$	$\widehat{damsup}$	$\widehat{damsin}$	56-87	71-87	$\widehat{nodams}$	$\widehat{damsup}$	$\widehat{damsin}$
This Year (+)	0.055***	0.016	0.027*	−0.013	−0.051	−0.337	−0.183	−2.649	−5.025	1.206
	(0.016)	(0.012)	(0.016)	(0.033)	(0.044)	(1.910)	(1.807)	(2.415)	(4.494)	(2.960)
Last Year (+)	0.060***	0.028**	0.019	−0.043	0.064	0.465	−0.225	−1.509	−6.091	5.389*
	(0.016)	(0.012)	(0.015)	(0.032)	(0.044)	(1.944)	(1.785)	(2.346)	(4.420)	(2.947)
Year-2 (+)	0.064***	0.029**	0.027*	0.014	0.028	−0.980	−1.919	−3.602	0.197	3.388
	(0.016)	(0.012)	(0.015)	(0.032)	(0.044)	(1.970)	(1.785)	(2.359)	(4.354)	(2.951)
This Year (-)	0.005	−0.072***	−0.078***	−0.117***	−0.077	−0.678	−1.511	−1.441	1.796	−9.069**
	(0.018)	(0.016)	(0.020)	(0.043)	(0.060)	(2.313)	(2.329)	(3.061)	(5.941)	(4.067)
Last Year (-)	0.005	−0.069***	−0.078***	−0.079*	−0.088	−1.253	−1.423	−2.720	−4.202	−5.135
	(0.019)	(0.016)	(0.021)	(0.045)	(0.062)	(2.295)	(2.428)	(3.231)	(6.194)	(4.206)
Year-2 (-)	0.040**	−0.046***	−0.019	−0.058	−0.063	−4.198*	−3.272	−4.215	−3.136	−2.218
	(0.018)	(0.016)	(0.020)	(0.042)	(0.064)	(2.237)	(2.341)	(3.084)	(5.725)	(4.337)
District Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	7,298	4,061	2,197	570	447	5,206	4,042	2,180	570	447

Note: Standard errors in parentheses clustered by district. *$p < 0.10$,** $p < 0.05$, *** $p < 0.01$. Each regression contains district and year fixed effects. In columns 3-5 and 8-10, regressions are estimated on subsets, determined based on the predicted number of dams within and upstream of a district. If a district is predicted to have fewer than 10 dams within and upstream, it is classified as being without dams. If a district is predicted to have over 10 dams upstream but fewer than 10 within the district is classified as having upstream dams. If a district is predicted to have over 10 dams within but fewer than 10 upstream it is classified as having dams within.

Table 3.7: Annual Rainfall Shock Sequences on Nominal Wages and Agricultural Production

	Log Nominal Wages					Log Production				
	56-87	71-87	$\widehat{nodams}$	$\widehat{damsup}$	$\widehat{damsin}$	56-87	71-87	$\widehat{nodams}$	$\widehat{damsup}$	$\widehat{damsin}$
Last Year: (0,-,+); This Year: (+)	0.026***	0.003	0.003	0.017	0.026	0.035***	−0.003	−0.033*	−0.036	0.057
	(0.009)	(0.009)	(0.013)	(0.018)	(0.023)	(0.011)	(0.014)	(0.017)	(0.036)	(0.041)
Last Year: (0,-); This Year: (-)	−0.011	−0.019	−0.016	−0.019	−0.061	−0.020	−0.054**	−0.033	−0.115**	−0.113*
	(0.010)	(0.013)	(0.018)	(0.027)	(0.037)	(0.015)	(0.022)	(0.029)	(0.049)	(0.060)
Last Year: (+); This Year: (-)	0.035*	−0.004	−0.022	0.042	−0.014	−0.002	−0.016	−0.026	−0.062	0.131
	(0.020)	(0.020)	(0.022)	(0.053)	(0.087)	(0.027)	(0.035)	(0.040)	(0.071)	(0.089)
Last Year: (+); This Year: (0)	0.020**	0.0005	0.004	−0.001	0.012	0.035***	0.011	0.004	−0.050	0.058
	(0.010)	(0.010)	(0.015)	(0.021)	(0.032)	(0.013)	(0.016)	(0.021)	(0.036)	(0.039)
District Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	7,675	4,080	2,199	570	447	7,648	4,063	2,199	570	447

Note: Standard errors in parentheses clustered by district. $^{*}p < 0.10$,** $p < 0.05$, *** $p < 0.01$. Each regression contains district and year fixed effects. Positive (Negative) shocks are defined as rainfall realizations falling in the upper (lower) quartile for that district. Shocks are grouped by common predictions. Indicator variables are set to 1 when a shock sequence in the group is realized with $[(-,0);0]$ as the omitted category. The first shock listed is the shock in $t-1$, the second is the contemporaneous shock. In columns 3-5 and 8-10, regressions are estimated on subsets, determined based on the predicted number of dams within and upstream of a district. If a district is predicted to have fewer than 10 dams within and upstream, it is classified as being without dams. If a district is predicted to have over 10 dams upstream but fewer than 10 within the district is classified as having upstream dams. If a district is predicted to have over 10 dams within but fewer than 10 upstream it is classified as having dams within.

Table 3.8: $\beta^{nom.wage}_{shock} - \beta^{prod.}_{shock}$

	56-87	71-87	$\widehat{nodams}$	$\widehat{damsup}$	$\widehat{damsin}$
Last Year: (0,-,+); This Year: (+)	-0.01	0.01	0.04	0.05	-0.03
Last Year: (0,-); This Year: (-)	0.01	0.03	0.02	0.1	0.05
Last Year: (+); This Year: (-)	0.04	0.01	0	0.1	-0.14
Last Year: (+); This Year: (0)	-0.01	-0.01	0	0.05	-0.05

Note: This table reports the coefficients on log nominal wages less the coefficient on log production from table 3.7.

Figures

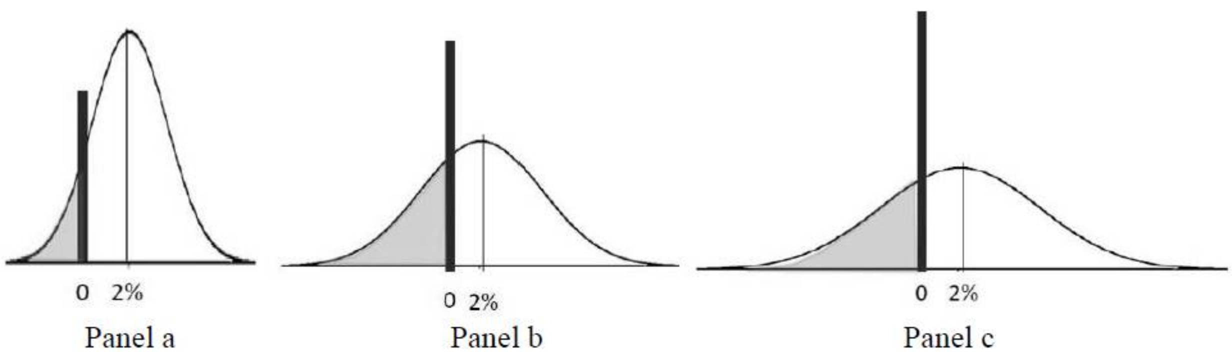


Figure 3.1: Nominal Wage Rigidity Under Different Wage Change Distributions

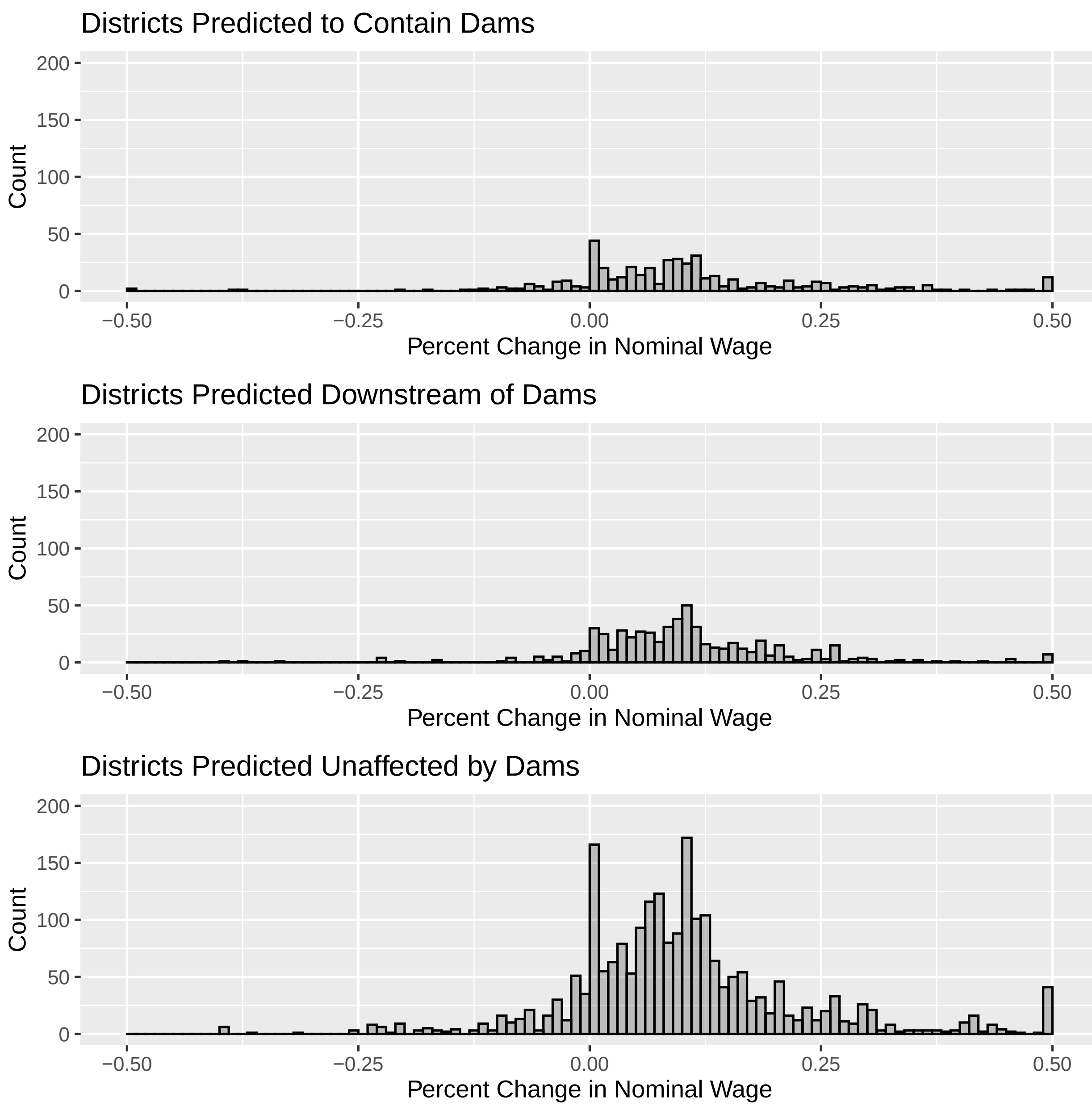


Figure 3.2: Distribution of the Yearly Percent Change in Nominal Wage by Dam Presence

Bibliography

Abdulkadiroglu, Atila and Tayfun Sonmez (2003). “School choice: A mechanism design approach”. In: *American Economic Review* 93.3, pp. 729–747.

Anderson, Michael L, Justin Gallagher, and Elizabeth Ramirez Ritchie (2018). “How the Quality of School Lunch Affects Students’ Academic Performance”. In: *The Education Digest* 83.6, pp. 61–64.

Beegle, Kathleen and Luc Christiaensen (2019). *Accelerating poverty reduction in Africa*. World Bank Publications.

Beegle, Kathleen, Luc Christiaensen, et al. (2016). *Poverty in a rising Africa*. The World Bank.

Carvalho, Leandro S, Stephan Meier, and Stephanie W Wang (2016). “Poverty and economic decision-making: Evidence from changes in financial resources at payday”. In: *American Economic Review* 106.2, pp. 260–84.

Chafuwa, Chiyembekezo (2017). *Priorities for irrigation investment in Malawi*. Policy Note 28. International Food Policy Research Institute.

Collier, Paul (2008). “The politics of hunger: How illusion and greed fan the food crisis”. In: *Foreign Affairs* Nov-Dec, pp. 67–79.

Collier, Paul and Stefan Dercon (2014). “African agriculture in 50 years: smallholders in a rapidly changing world?” In: *World Development* 63, pp. 92–101.

Danaher, Kelly and Christian S Crandall (2008). “Stereotype threat in applied settings reexamined”. In: *Journal of Applied Social Psychology* 38.6, pp. 1639–1655.

Dercon, Stefan and Douglas Gollin (2014). “Agriculture in African development: theories and strategies”. In: *Annual Review of Resource Economics* 6.1, pp. 471–492.

Duflo, Esther and Rohini Pande (2007). “Dams”. In: *The Quarterly Journal of Economics* 122.2, pp. 601–646.

Dustan, Andrew, Alain De Janvry, and Elisabeth Sadoulet (2017). “Flourish or fail? The risky reward of elite high school admission in Mexico City”. In: *Journal of Human Resources* 52.3, pp. 756–799.

Editors (2010). “SYMPOSIUM: Bias in the SAT? Continuing the Debate”. In: *Harvard Educational Review* 80.3, pp. 391–394.

ETS (2014). *ETS Standards for Quality and Fairness*. Tech. rep. Educational Testing Service.

Evidence Action (2014). *Seasonal Migration Feasibility Analysis: Malawi and Zambia*. Tech. rep.

Freedle, Roy (2003). "Correcting the SAT's ethnic and social-class bias: A method for reestimating SAT scores". In: *Harvard Educational Review* 73.1, pp. 1–43.

— (2010). "On replicating ethnic test bias effects: The Santelices and Wilson study". In: *Harvard Educational Review* 80.3, pp. 394–404.

Fryer Jr, Roland G, Steven D Levitt, and John A List (2008). "Exploring the impact of financial incentives on stereotype threat: Evidence from a pilot study". In: *American Economic Review* 98.2, pp. 370–75.

Gassman-Pines, Anna and Laura Bellows (2018). "Food instability and academic achievement: a quasi-experiment using SNAP benefit timing". In: *American Educational Research Journal* 55.5, pp. 897–927.

Gollin, Douglas, David Lagakos, and Michael E Waugh (2013). "The agricultural productivity gap". In: *The Quarterly Journal of Economics* 129.2, pp. 939–993.

Haushofer, Johannes and Ernst Fehr (2014). "On the psychology of poverty". In: *Science* 344.6186, pp. 862–867.

Hoyos, Rafael E. de (2014). "La evaluación que no evalúa". In: *Nexos*.

IFAD (2016). *Rural Development Report 2016. Fostering Inclusive Rural Transformation.*

Jayachandran, Seema (2006). "Selling labor low: Wage responses to productivity shocks in developing countries". In: *Journal of political Economy* 114.3, pp. 538–575.

Kaur, Supreet (2019). "Nominal wage rigidity in village labor markets". In: *American Economic Review* 109.10, pp. 3585–3616.

Kaur, Supreet et al. (2019). *Does Financial Strain Lower Productivity?* Tech. rep. Working Paper.

Lagakos, David, Ahmed Mushfiq Mobarak, and Michael E Waugh (2018). *The Welfare Effects of Encouraging Rural-Urban Migration.* Working Paper w24193. National Bureau of Economic Research.

LaRoche, Sylvie, Marc Joncas, and Pierre Foy (2016). "Sample Design in TIMSS 2015". In: *Methods and Procedures in TIMSS 2015*, pp. 3.1–3.37.

Lea, Nicholas and Lucia Hanmer (2009). *Constraints to growth in Malawi.* The World Bank.

Lewis, W Arthur (1954). "Economic development with unlimited supplies of labour". In: *The Manchester School* 22.2, pp. 139–191.

Mani, Anandi et al. (2013). "Poverty impedes cognitive function". In: *Science* 341.6149, pp. 976–980.

McCullough, Ellen B (2017). "Labor productivity and employment gaps in Sub-Saharan Africa". In: *Food Policy* 67, pp. 133–152.

McMillan, Margaret and Dani Rodrik (2011). "Globalization, structural change and productivity growth". In: *Making Globalization Socially Sustainable.* Ed. by M Bacchetta and M. Jense.

McMillan, Margaret, Dani Rodrik, and Íñigo Verduzco-Gallo (2014). "Globalization, Structural Change, and Productivity Growth, with an Update on Africa". In: *World Development* 63, pp. 11–32.

Munshi, Kaivan (2003). "Networks in the modern economy: Mexican migrants in the US labor market". In: *The Quarterly Journal of Economics* 118.2, pp. 549–599.

Orr, Alastair (2000). "'Green Gold'?: Burley tobacco, smallholder agriculture, and poverty alleviation in Malawi". In: *World Development* 28.2, pp. 347–363.
Ozler, Berk, Gaurav Datt, and Martin Ravallion (1996). "A database on poverty and growth in India". In: *World Bank Development Research Group, Washington DC*.
Paxson, Christina H (1992). "Using weather variability to estimate the response of savings to transitory income in Thailand". In: *The American Economic Review*, pp. 15–33.
Ricker-Gilbert, Jacob (2014). "Wage and employment effects of Malawi's fertilizer subsidy program". In: *Agricultural Economics* 45.3, pp. 337–353.
Rosenzweig, Mark and Christopher R Udry (2013). *Forecasting profitability*. Tech. rep. National Bureau of Economic Research.
Rosenzweig, Mark R and Kenneth I Wolpin (1993). "Credit market constraints, consumption smoothing, and the accumulation of durable production assets in low-income countries: Investments in bullocks in India". In: *Journal of political economy* 101.2, pp. 223–244.
Sanghi, Apurva, KS Kavi Kumar, and James W McKinsey Jr (1998). "India agriculture and climate data set". In: *World Bank*.
Santelices, Maria Veronica and Mark Wilson (2010). "Unfair treatment? The case of Freedle, the SAT, and the standardization approach to differential item functioning". In: *Harvard Educational Review* 80.1, pp. 106–134.
Sarsons, Heather (2015). "Rainfall and conflict: A cautionary tale". In: *Journal of development Economics* 115, pp. 62–72.
Shah, Anuj K, Sendhil Mullainathan, and Eldar Shafir (2012). "Some consequences of having too little". In: *Science* 338.6107, pp. 682–685.
— (2018). "An exercise in self-replication: Replicating Shah, Mullainathan, and Shafir (2012)". In: *Journal of Economic Psychology*.
Spears, Dean (2011). "Economic decision-making in poverty depletes behavioral control". In: *The BE Journal of Economic Analysis & Policy* 11.1.
Spencer, Steven J, Christine Logel, and Paul G Davies (2016). "Stereotype threat". In: *Annual Review of Psychology* 67, pp. 415–437.
Steele, Claude M and Joshua Aronson (1995). "Stereotype threat and the intellectual test performance of African Americans." In: *Journal of Personality and Social Psychology* 69.5, p. 797.
Stricker, Lawrence J and William C Ward (2004). "Stereotype Threat, Inquiring About Test Takers' Ethnicity and Gender, and Standardized Test Performance 1". In: *Journal of Applied Social Psychology* 34.4, pp. 665–693.
Tomm, Brandon M and Jiaying Zhao (2016). "Scarcity captures attention and induces neglect: Eyetracking and behavioral evidence". In: *Proceedings of the 38th Annual Conference of the Cognitive Science Society*, pp. 1199–1204.
Townsend, Robert M (1994). "Risk and insurance in village India". In: *Econometrica: Journal of the Econometric Society*, pp. 539–591.
Vivanco, E (2013). "Evaluating the effects of school and teacher incentives: Quasi-Experimental Evidence from Mexico". In: *Manuscrito inédito*.

Wei, Thomas E (2012). “Sticks, stones, words, and broken bones: New field and lab evidence on stereotype threat”. In: *Educational Evaluation and Policy Analysis* 34.4, pp. 465–488.
Wodon, Quentin and Kathleen Beegle (2006). “Labor shortages despite underemployment? Seasonality in time use in Malawi”. In: *Gender, Time Use, and Poverty in Sub-Saharan Africa*. Ed. by Quentin Wodon and C Mark Blackden. The World Bank, pp. 97–116.
Xia, Fang and Klaus Deininger (2019). “Spillover Effects of Tobacco Farms on the Labor Supply, Education, and Health of Children: Evidence from Malawi”. In: *American Journal of Agricultural Economics* 101.4, pp. 1181–1202.

Appendix A

Appendix to Chapter 1

Table A.1: Missing Parental Education

	Prop Mon Q. in Booklet	Standardized Score			
	(1)	(2)	(3)	(4)	(5)
Missing Parental Education	-0.000239	-0.297***	-0.232***	-0.283***	-0.219***
	(0.000292)	(0.00367)	(0.00354)	(0.00491)	(0.00457)
Missing Par Edu. x Prop Mon Q.				-0.233***	-0.211***
				(0.0530)	(0.0468)
FE: Year	Yes	.	.	.	.
FE: Country	.	Yes	.	Yes	.
FE: Class	Yes	No	Yes	No	Yes
FE: Booklet x Year	No	Yes	Yes	Yes	Yes
N	469697	469849	469697	469849	469697

Note: Standard errors in parentheses. *$p < 0.10$,** $p < 0.05$, *** $p < 0.01$. Observations are at the student by examination level with a student level SES indicator: parental education. Omitted category is students with reported parental education levels. The proportion of monetary questions in a booklet is a value from 0 to 1.

Table A.2: TIMSS Main Results by Occupation

	Standardized Score	
	(1)	(2)
Small Business	-0.163***	-0.114***
	(0.00658)	(0.00597)
Clerical	-0.249***	-0.131***
	(0.00506)	(0.00465)
Skilled Labor	-0.373***	-0.220***
	(0.00672)	(0.00622)
General Labor	-0.518***	-0.292***
	(0.0100)	(0.00932)
Never Wk. for Pay	-0.438***	-0.217***
	(0.0106)	(0.00979)
Small Business x Prop Mon Q.	0.0142	-0.0646
	(0.0777)	(0.0700)
Clerical x Prop Mon Q.	-0.0179	-0.0374
	(0.0591)	(0.0532)
Skilled Labor x Prop Mon Q.	-0.0968	-0.0815
	(0.0784)	(0.0707)
General Labor x Prop Mon Q.	-0.182	-0.215**
	(0.114)	(0.103)
Never Wk. for Pay x Prop Mon Q.	-0.161	-0.212*
	(0.123)	(0.111)
Constant	0.148***	0.0857***
	(0.00192)	(0.00176)
FE: Booklet x Year	Yes	Yes
FE: Country	Yes	.
FE: Class	No	Yes
N	379468	379160

Note: Standard errors in parentheses. $^{*}p < 0.10$,** $p < 0.05$, *** $p < 0.01$. Observations are at the student by examination level with a student level SES indicator: parental occupation. Omitted categories are students with professional parental occupations. The proportion of monetary questions in a booklet is a value from 0 to 1.

Table A.3: TIMSS Question Fixed Effects

	Question Answered Correctly (=100)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Below Nat. Median x Mon Q.	-0.885***	-1.207***	-0.694***	-1.573***	-1.399***	-0.484***	-0.478***	-1.208***	-1.436***	-0.992***
	(0.123)	(0.131)	(0.124)	(0.126)	(0.127)	(0.128)	(0.128)	(0.131)	(0.128)	(0.131)
Below Nat. Median x 4 Post	-0.680***	-0.891***	-0.716***	-0.990***	-1.024***	-0.642***	-0.674***	-0.888***	-0.858***	-1.078***
	(0.0880)	(0.0969)	(0.0893)	(0.0957)	(0.0969)	(0.0886)	(0.0903)	(0.0954)	(0.0969)	(0.0970)
FE: Student	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FE: Question	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FE: Below Med. x Diff.	No	Yes	No	Yes	Yes	No	No	Yes	Yes	Yes
FE: Below Med. x Seq.	No	Yes	Yes	No	Yes	No	Yes	No	Yes	Yes
FE: Below Med. x QType x Country	No	Yes	No	No	No	Yes	Yes	Yes	Yes	No
FE: Below Med. x QTopic x Country	No	Yes	No	No	No	Yes	Yes	Yes	No	Yes
Dep. Variable Mean	49.93	49.93	49.93	49.93	49.93	49.93	49.93	49.93	49.93	49.93
Dep. Variable SD	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00
N	9564201	9564201	9564201	9564201	9564201	9564201	9564201	9564201	9564201	9564201

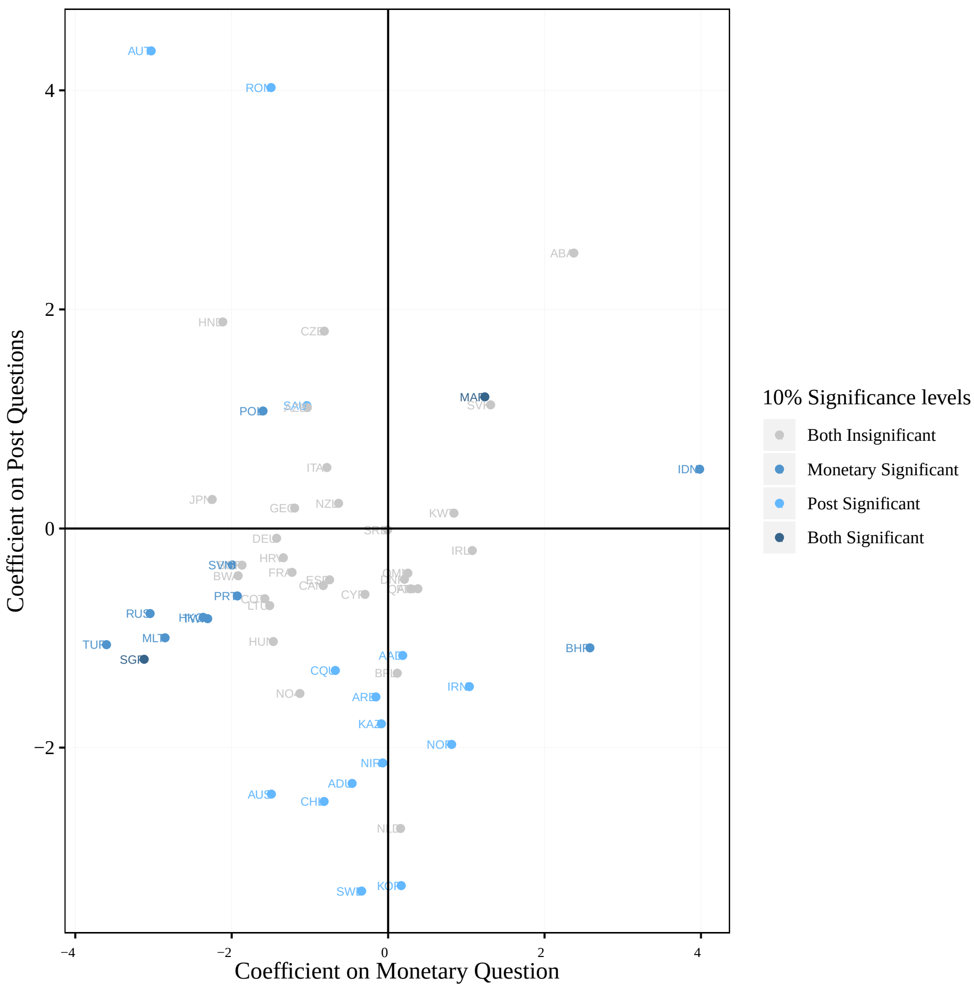
Note: Standard errors in parentheses clustered at the student level. $^{*}p < 0.10$,** $p < 0.05$, *** $p < 0.01$. Observations are at the question by student level with a student level SES indicator: parental education relative to the national median. When a question is answered correctly the indicator is set to 100, 0 otherwise. Omitted categories are students with parental education at or above the national median. Difficulty is a 20 bin binned indicator based on the performance on a question by students with parental education above the national median. Sequence is a 5 bin binned indicator based on the the position of a question within the exam booklet. Question type indicates whether a question is multiple choice or completed response. Question topic indicates categorized questions based on the topics listed in panel b of figure 1.8.

Table A.4: TIMSS Unanswered Questions

	Question Left Unanswered (=1)					
	(1)	(2)	(3)	(4)	(5)	(6)
Below Nat. Median x Mon Q.	-0.00378***		-0.00362***		-0.00169**	
	(0.000691)		(0.000723)		(0.000803)	
Post Sec. x Mon Q.		-0.000789		-0.00153*		-0.000661
		(0.000848)		(0.000888)		(0.000927)
Upper Sec. x Mon Q.		0.00120		-0.00182**		-0.00294***
		(0.000813)		(0.000847)		(0.000942)
Lower Sec. x Mon Q.		-0.00120		-0.00264**		-0.00266*
		(0.00127)		(0.00133)		(0.00146)
Primairy/No x Mon Q.		-0.00732***		-0.00842***		-0.00873***
		(0.00144)		(0.00151)		(0.00182)
Below Nat. Median x 4 Post	-0.00147***		-0.00296***		-0.00232***	
	(0.000545)		(0.000604)		(0.000666)	
Post Sec. x 4 Post		0.00157**		-0.00148**		-0.00166**
		(0.000671)		(0.000742)		(0.000780)
Upper Sec. x 4 Post		0.00416***		-0.000794		-0.00301***
		(0.000641)		(0.000705)		(0.000784)
Lower Sec. x 4 Post		0.00351***		-0.00151		-0.00324***
		(0.000994)		(0.00110)		(0.00121)
Primairy/No x 4 Post		0.00178		-0.00164		-0.00674***
		(0.00115)		(0.00129)		(0.00152)
FE: Student	Yes	Yes	Yes	Yes	Yes	Yes
FE: Question	Yes	Yes	Yes	Yes	Yes	Yes
FE: Below Med. x Diff.	No	.	Yes	.	Yes	.
FE: Below Med. x Seq.	No	.	Yes	.	Yes	.
FE: Below Med. x QType x Country	No	.	Yes	.	Yes	.
FE: Below Med. x QTopic x Country	No	.	Yes	.	Yes	.
FE: Par. Edu. x Diff.	.	No	.	Yes	.	Yes
FE: Par. Edu. x Seq.	.	No	.	Yes	.	Yes
FE: Par. Edu. x QType x Country	.	No	.	Yes	.	Yes
FE: Par. Edu. x QTopic x Country	.	No	.	Yes	.	Yes
FE: Class x Mon Q.	No	No	No	No	Yes	Yes
FE: Class x 4 Post	No	No	No	No	Yes	Yes
Dep. Variable Mean	0.0598	0.0598	0.0598	0.0598	0.0598	0.0598
Dep. Variable SD	0.237	0.237	0.237	0.237	0.237	0.237
N	9564201	9564201	9564201	9564201	9563918	9563918

Note: Standard errors in parentheses clustered at the student level. $^{*}p < 0.10$,** $p < 0.05$, *** $p < 0.01$. Observations are at the question by student level with a student level SES indicator: parental education relative to the national median. Omitted categories are students with parental education at or above the national median for columns 1, 3 and 5 and university educated parents for columns 2, 4 and 6. Difficulty is a 20 bin binned indicator based on the performance on a question by students with university educated parents. Sequence is a 5 bin binned indicator based on the the position of a question within the exam booklet. Question type indicates whether a question is multiple choice or completed response. Question topic indicates categorized questions based on the topics listed in panel b of figure 1.8.

Figure A.1: Country Estimates for Below National Median Students



Note: Estimating equation includes Student, Question, Below Med. x Diff., Below Med. x Seq., Below Med. x QType x Country, Below Med. x QTopic x Country fixed effects.

Table A.5: COMIPEMS Simulation using School Indicators

		Ineligible		Eligible			
		Score under 31 points		Not Assigned		Assigned	
Total	Actual	8,373	1.67%	84,513	16.80%	410,057	81.53%
	Simulated	8,353	1.66%	84,518	16.81%	410,072	81.53%
Missing	Actual	392	2.39%	3,312	20.15%	12,733	77.47%
	Simulated	392	2.39%	3,314	20.16%	12,731	77.45%
Very Advantaged	Actual	6,245	1.60%	69,815	17.84%	315,189	80.56%
	Simulated	6,245	1.60%	69,844	17.85%	315,160	80.55%
Advantaged	Actual	1,615	1.79%	10,717	11.89%	77,780	86.32%
	Simulated	1,600	1.78%	10,695	11.87%	77,817	86.36%
Middle	Actual	74	2.15%	389	11.29%	2,982	86.56%
	Simulated	71	2.06%	386	11.21%	2,988	86.73%
Disadvantaged	Actual	45	2.70%	274	16.45%	1,347	80.85%
	Simulated	43	2.58%	273	16.39%	1,350	81.03%
Very Disadvantaged	Actual	2	5.88%	6	17.65%	26	76.47%
	Simulated	2	5.88%	6	17.65%	26	76.47%

Table A.6: COMIPEMS Simulation using School Indicators: Movement Detail

	Remain Ineligible or Unassigned	Become Assigned	More Prefered Assignment	Unchanged Assignment	Less Prefered Assignment	Become Unassigned	Change in Mean Preference Rank*
Total	92,807	79	216	409,548	229	64	-.00007
	18.45%	0.02%	0.04%	81.43%	0.05%	0.01%	
Missing	3,703	1	1	12,725	4	3	-.00047
	22.53%	0.01%	0.01%	77.42%	0.02%	0.02%	
Very Advantaged	76,041	19	31	314,945	165	48	-.00082
	19.44%	0.00%	0.01%	80.50%	0.04%	0.01%	
Advantaged	12,282	50	157	77,550	60	13	.00256
	13.63%	0.06%	0.17%	86.06%	0.07%	0.01%	
Middle	457	6	21	2,961	0	0	.00905
	13.27%	0.17%	0.61%	85.95%	0%	0%	
Disdvantaged	316	3	6	1,341	0	0	.00594
	18.97%	0.18%	0.36%	80.49%	0%	0%	
Very Disadvantaged	8	0	0	26	0	0	0
	23.53%	0%	0%	76.47%	0%	0%	

Note: For students who are assigned in both the actual and simulated data.

Figure A.2: Example Page from 4th Grade ENLACE Mathematics

ENLACE.10_4°

17. Cuatro amigos leen un libro. La cantidad que cada uno ha leído se muestra en la tabla.

Amigo	Cantidad leída
Daniel	$\frac{2}{3}$
Fernando	$\frac{1}{5}$
Manuel	$\frac{3}{4}$
Guillermo	$\frac{5}{6}$

¿Qué amigo ha leído menos?

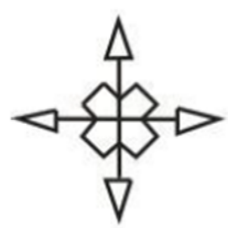
A) Fernando.
B) Daniel.
C) Manuel.
D) Guillermo.

18. María va a realizar el pago de los siguientes recibos: teléfono $209.40, luz $198.50 y agua $100.30. ¿Cuánto pagará en total?

A) $507.02
B) $507.12
C) $508.02
D) $508.20

19. ¿Cuál de las siguientes figuras tiene menos ejes de simetría?

A) 

B) 

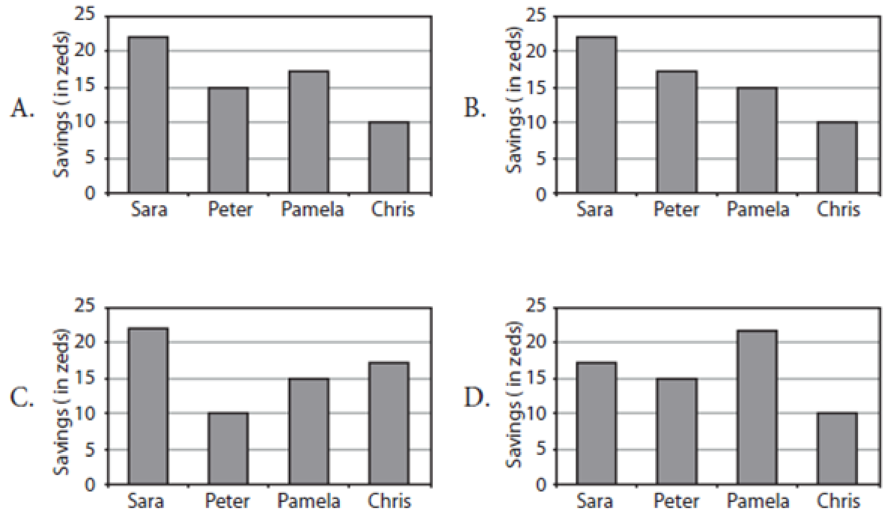
C) 

D) 

Figure A.3: Example Monetary Questions from the 2011 TIMSS

John was given the following table by his teacher and was asked to identify the graph that correctly displays the data. Which graph below should he choose?

Name	Savings
Sara	22 zeds
Peter	15 zeds
Pamela	17 zeds
Chris	10 zeds



Appendix B

Appendix to Chapter 2

Table B.1: Cropping Patterns

Mean acres	All	North	Central	South
Total	2.381	1.971	3.473	1.486
Maize	1.172	0.956	2.040	0.431
Maize-Beans	0.056	0.092	0.076	0.028
Maize-Pigeonpeas	0.135	0.001	0.006	0.290
Groundnuts	0.193	0.116	0.406	0.017
Tobacco	0.294	0.187	0.615	0.026
Other	0.180	0.271	0.150	0.183
Other-Maize	0.351	0.348	0.180	0.510
Observations	10,100	1,696	3,575	4,829

Note: Sample consists of all households reporting at least one cultivated plot. This includes 851 urban households. Land area is calculated using GPS measures of plot area.

Table B.2: Mean Labor Hours per Acre, by Crop

	Maize (MZ)	MZ-Beans	MZ-Pigeon Pea	MZ-Other	Groundnuts	Tobacco	Other
Total	418	441	471	410	484	592	496
....Planting	183	203	216	185	188	201	196
....Other	150	167	168	155	150	158	160
....Harvest	73	67	75	63	127	181	117
Observations	3,846	300	1,190	2,100	786	693	1,240

Note: Sample consists of all reported plots farmed using household labor only. Labor hours per acre are first winsorized at .05.

Table B.3: Descriptive Statistics by Year for Households Engaged in Agriculture

		2004	2010	2016
Cultivated area in acres	Mean	2.29	1.80	1.38
	Median	2	1.50	1
Total household labor hours in past week	Mean	59.19	41.00	31.73
	Median	50	30	21
Labor hours in past week in peak season (Dec-Jan)	Mean	72.20	58.91	45.57
	Median	63	51	36
Household size	Mean	4.77	4.72	4.43
	Median	5	5	4
Household working-age individuals not in school	Mean	2.02	1.91	1.79
	Median	2	2	2
Observations		9,798	10,096	9,470

Note: Sample consists of all households reporting at least one cultivated plot. For consistency across years, land area is calculated using self-reported plot size winsorized at 5pct.

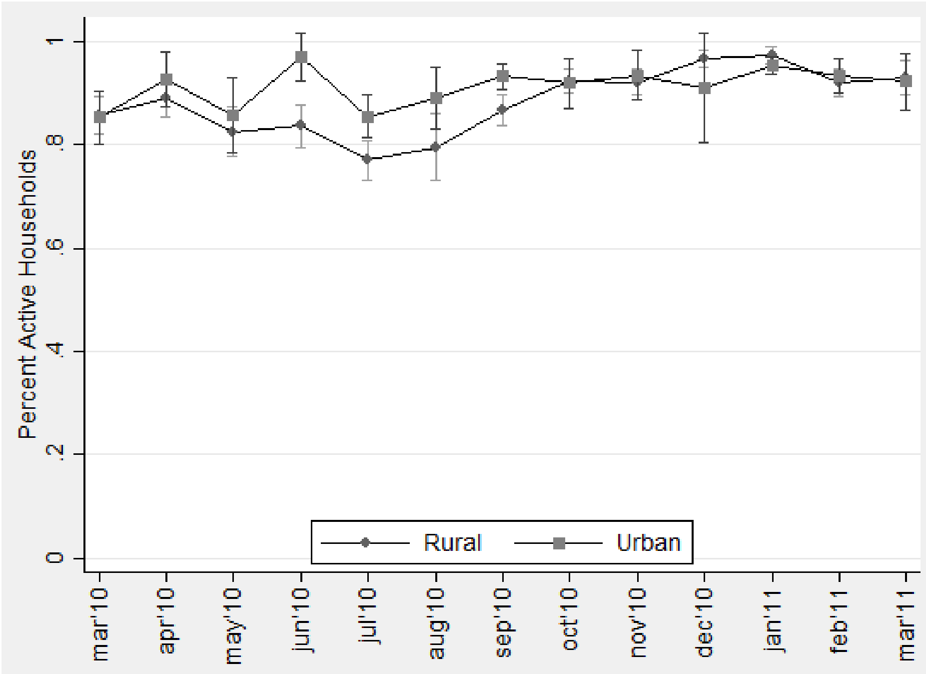
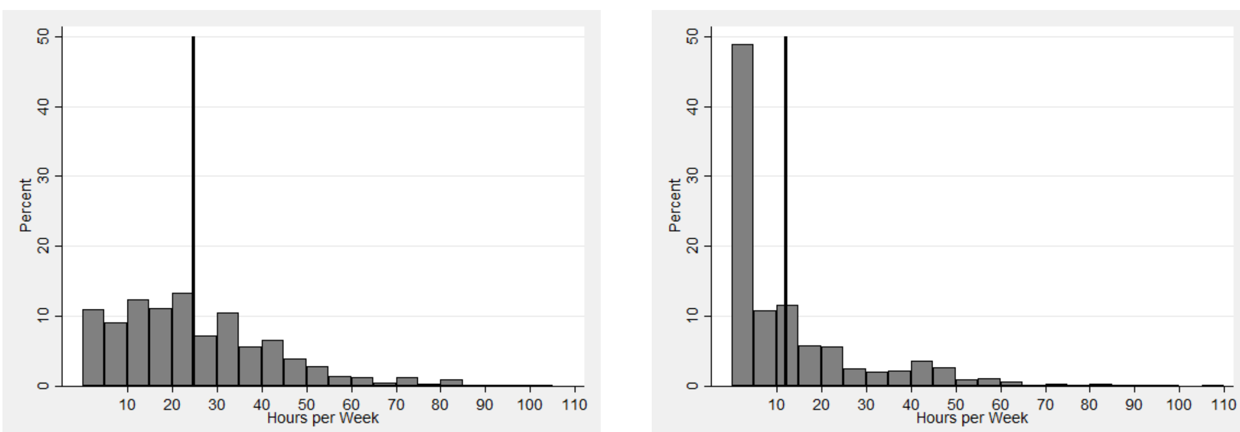


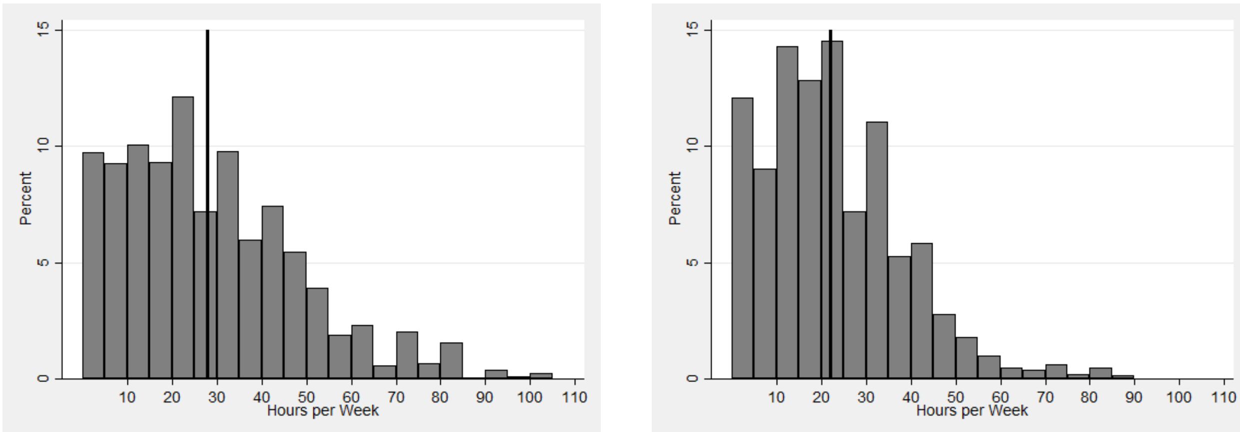
Figure B.1: Percent of active households last week



(a) High season (Dec-Jan)

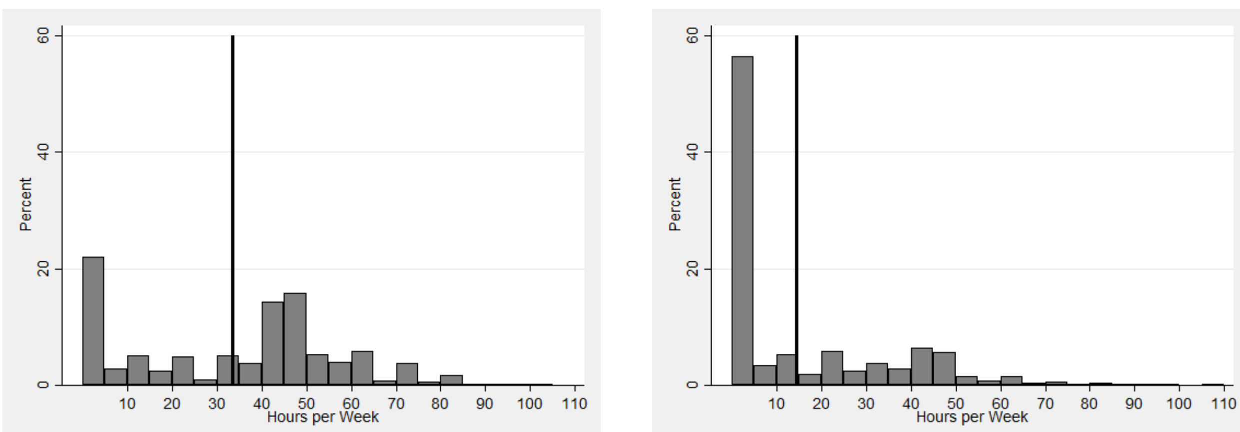
(b) Low season (Jul-Aug)

Figure B.2: Distribution of weekly hours reported by rural individuals by season



(a) Rural men (Dec-Jan) (b) Rural women (Dec-Jan)

Figure B.3: Distribution of weekly hours reported by rural individuals in the high season by gender



(a) Urban men (b) Urban women

Figure B.4: Distribution of weekly hours reported by urban individuals by gender

Table B.4: Labor Supplied by Households, Rural vs. Urban: 2004, 2010, and 2016

	Contrast	Obs	Total annual hrs	High weekly hrs	Low weekly hrs	Standard deviation	Coeff. of variation (%)
Rural vs. urban, 2010	2010 Rural	10,037	2,065	56.93	29.23	9.58	24.26
	2010 Urban	2,229	2,863	58.21	51.38	5.62	10.26
	2010 Rural/urban		0.72***	0.98	0.57***	1.70	2.36
Rural vs. urban, 2004	2004 Rural	9,840	3,088	70.98	48.41	8.75	14.82
	2004 Urban	1,440	3,266	66.67	61.58	6.06	9.72
	2004 Rural/urban		0.95	1.06	0.79***	1.44	1.52
Rural vs. urban, 2016	2016 Rural	10,175	1,488	37.12	26	7.97	25.67
	2016 Urban	2,272	2,277	54.39	43.99	11.31	23.76
	2016 Rural/urban		0.65***	0.68***	0.59***	0.70	1.08
Rural vs. urban, 04-10-16	Pooled Rural	30,052	1,858	49.37	27.06	8.33	23.45
	Pooled Urban	5,941	2,651	58.70	49.30	4.15	8.18
	Pooled Rural/urban		0.70***	0.84**	0.55***	2.01	2.87

Note: High season is December and January, low season is July and August.

Table B.5: Labor Engagement of Individuals, Rural vs. Urban: 2004, 2010, and 2016

	Contrast	Obs	Mean % active	High % active	Low % active	Standard deviation	Coeff. of variation (%)
Rural vs. urban, 2010	2010 Rural	18,699	0.79	0.93	0.64	0.10	13.08
	2010 Urban	4,625	0.64	0.67	0.61	0.05	8.58
	2010 Rural/urban		1.23***	1.39***	1.05	2	1.52
Rural vs. urban, 2004	2004 Rural	19,674	0.88	0.95	0.79	0.05	6.09
	2004 Urban	3,114	0.67	0.72	0.62	0.05	7.48
	2004 Rural/urban		1.31***	1.32***	1.27***	1	0.81
Rural vs. urban, 2016	2016 Rural	18,039	0.70	0.78	0.65	0.10	14.20
	2016 Urban	4,424	0.64	0.69	0.67	0.06	9.58
	2016 Rural/urban		1.09***	1.13***	0.97	1.67	1.48
Rural vs. urban, 04-10-16	Pooled rural	56,412	0.75	0.87	0.62	0.09	12.42
	Pooled urban	12,163	0.64	0.67	0.63	0.04	6.08
	Pooled rural/urban		1.17***	1.30***	0.98	2.25	2.04

Note: Sample consists of working age individuals who are not in school. High season is December and January, low season is July and August.

Appendix C

Appendix to Chapter 3

C.1 Instrumenting for Dams

As described, I instrument for irrigation dams using the approach developed in Duflo and Pande (2007). In the equation below, D_{dst} is the number of dams in district d, state s at time t and RGr_{kd} is the fraction of district d's river gradient falling in category k (four gradient categories are used). These are interacted with $\overline{D_{st}}$, predicted dam incidence in the state, which is that year's total dam construction in India multiplied by that state's fraction of Indian dams in 1970. M_d is a vector of geographic controls including district elevation, gradients, river length and area. RGr_{kd} is also interacted with year dummies T_t to account for national time varying characteristics. Finally district and state by year fixed effects are included. I estimate the following,

$$D_{dst} = \alpha_1 + \sum_{k=2}^{4} \alpha_{2k}(RGr_{kd} * \overline{D_{st}}) + \alpha_3(M_d * \overline{D_{st}}) + \sum_{k=2}^{4} \alpha_{4k}(RGr_{kd} * T_t) + \nu_d + \mu_{st} + \omega_{dst}. \tag{C.1}$$

The parameters estimated using equation C.1 are then used to generate a set of predicted values $\widehat{D_{dst}}$ and $\widehat{D^U_{dst}}$.

Dam construction decisions in India are decided at the federal and state level with the final location heavily influenced by engineering considerations that determine the cost of construction. This instrument is based on the identifying assumption that without dam construction, the evolution of outcomes between districts in the same state but with different river gradients would not have systematically differed across states with more dams in 1970 and states with fewer dams in 1970 (Duflo and Pande (2007)). Thus identification uses three sources of variation: differences in dam construction across years, across states and across districts within a state. There is significant variation across years and states as dam construction grew rapidly in India between 1970 and 1999 in some states but not in others. The average number of dams in a district increased from about 3 to 12 in my sample with a little less than half of district having no dams by 1999. Though dams are a count variable, OLS is used as there is a fairly continuous level of variation and the use of fixed effects for identification would be problematic if using a non-linear estimator.

To ensure the measure is exogenous to the number of dams in a district, the instrument uses predicted dam incidence $\overline{D_{st}}$. This predicted value is calculated using dam construction between 1956 and 1970 to estimate predicted construction by state after 1970, the period when dam construction rapidly accelerated (Duflo and Pande (2007)). As dam construction prior to 1970 is used to construct the instrument, predicted values for dams based on the instrument are only available starting in 1971; thus much of my analysis is limited to post-1971 data. While it would be possible to increase the sample by generating predicted values using an earlier year, this would decrease the predictive power of the instrument. As such, I elect to use the instrument as designed by Duflo and Pande.

Table C.1 presents the regression of geographical variables on dams. The coefficients highlight the importance of physical geography and engineering consideration in determining dam placement. Having a river gradient between 1.5 and 3% increases the likelihood of dam construction. River gradients greater then 6% also increases dam construction, though this is likely mostly dams whose primary purpose is hydroelectric rather than irrigation. The coefficients match those in Duflo and Pande and the R-square of 0.959 shows that the model effectively predicts districts likely to contain dams. Having successfully replicated the predicting equation and identified the set of controls Z_{ist}^{U} and Z_{ist}, the paper uses the predicted values generated by these coefficients and available in the dataset.

Tables and Figures for Appendix C.1

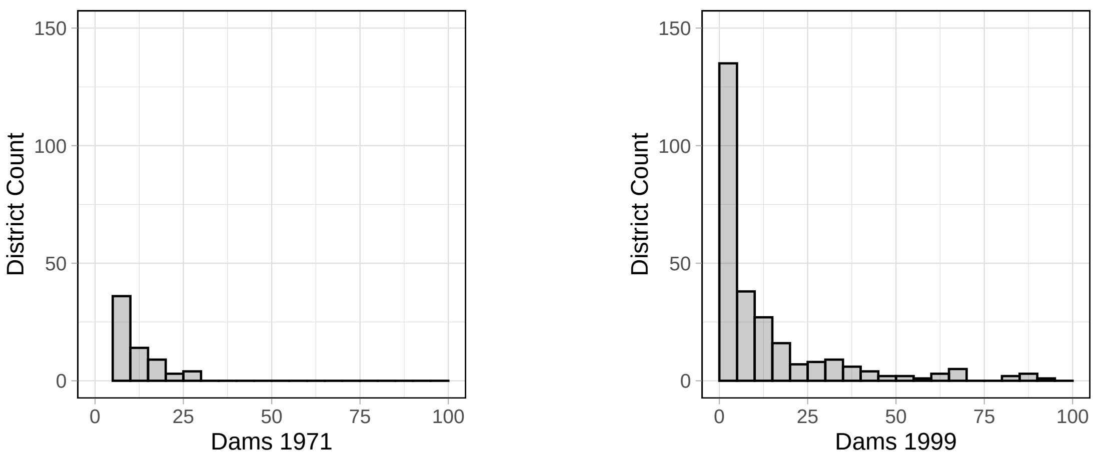


Figure C.1: Distribution of Number of Dams in District 1971 and 1999

Table C.1: Geography and Dam Construction 1971-1999

	Number of Dams
Fraction of river gradient 1.5-3pct	0.176*
	(0.098)
Fraction of river gradient 3-6pct	−0.219
	(0.134)
Fraction of river gradient above 6pct	0.097**
	(0.045)
Fixed effects	District
Observations	7,743
R^2	0.959

Note: Standard errors in parentheses clustered by district. *$p < 0.10$,** $p < 0.05$, *** $p < 0.01$. Controls include geography controls for river length, district area, elevation and district gradients as well as gradient *year interactions, stat*year interactions and district fixed effects.

C.2 Duflo and Pande replication

Duflo and Pande 2007 estimate

$$y_{dst} = \delta_1 + \delta_2 D_{dst} + \delta_3 D^U_{dst} + \delta_4 Z_{dst} + \delta_5 Z^U_{dst} + \nu_d + \mu_{st} + \omega_{dst} \tag{C.2}$$

where Z_{dst} and Z^U_{dst} are the right hand side variables in equation C.1 (interactions $RGr_{kd} * \overline{D_{st}}$ excepted) and the number of dams D_{dst} and upstream dams D^U_{dst} are instrumented for with the predicted values $\widehat{D_{dst}}$ and $\widehat{D^U_{dst}}$.

Following this approach, I replicate their results with respect to agricultural production, dams and rainfall, as these results play an important role in my hypothesis. I estimate the following two equations were y_{dst} is agricultural production and FAR is the Fractional Annual Rainfall[1]:

[1]In this analysis I use the same rainfall variable as used by Dulfo and Pande, the fractional deviation of the district's rainfall from the district mean (computed over 1971-1999).

$$y_{dst} = \delta_1 + \delta_2 D_{dst} + \delta_3 D^U_{dst} + \delta_4 Z_{dst} + \delta_5 Z^U_{dst} + \nu_i + \mu_{st} + \omega_{dst}, \tag{C.3}$$

$$\begin{aligned} y_{dst} = \delta_1 &+ \delta_2 D_{dst} + \delta_3 D_{dst} * FAR_{dst} + \delta_4 D^U_{dst} \\ &+ \delta_5 D^U_{dst} * FAR_{dst} + \delta_6 Z_{dst} + \delta_7 Z^U_{dst} + \nu_d + \mu_{st} + \omega_{dst}. \end{aligned} \tag{C.4}$$

Their results are replicated in the 71-99 columns of table C.2. The 71-92 column restricts the sample to years in which nominal wage data appears to be of good quality. I exclude the years 93-99 as the wage data displays some irregularities. The 71-87 sample covers years in which the full detail of agricultural data is available.

The three samples reveal similar effects to those noted in Duflo and Pande (2007), though the smaller samples mechanically loose some statistical significance. The large standard errors on the effect on agricultural production of having a dam within the district suggest an ambiguous effect on the dam containing district. This is likely in part due to data limitations. It is not possible to distinguish between districts where the dam is just within the upstream border, and therefore benefiting from the irrigation benefits versus districts where the dam is just within the downstream border and therefore mostly affected as a catchment area. Having dams upstream however has a positive effect on agricultural production though the coefficient is substantially smaller in the 71-92 and 71-87 samples. More critical to my hypothesis is dam effects on the sensitivity to rain shocks. The coefficient on the interaction terms indicates how dams affect the sensitivity of agricultural production to rainfall. As noted in Duflo and Pande (2007), dams within the district amplify the effect of rainfall shocks as the coefficient on the interaction term is positive like the rainfall coefficient. Dams in upstream districts dampen the effect of rainfall shocks, the coefficient on that interaction term being negative. Note that while the sign of these interaction terms holds, neither of these interactions stay statistically significant in the smaller samples.

Tables and Figures for Appendix C.2

Table C.2: Dams, Rainfall and Agricultural Production

	Log Agricultural Production					
	71-99	71-99	71-92	71-92	71-87	71-87
	(1)	(2)	(3)	(4)	(5)	(6)
Fractional Annual Rainfall	0.065**	0.008	0.139***	0.126***	0.100***	0.081***
	(0.032)	(0.047)	(0.023)	(0.033)	(0.023)	(0.031)
Dams in District	−0.011	0.109	0.495	0.519	−0.088	−0.045
	(1.301)	(1.301)	(1.182)	(1.182)	(1.111)	(1.129)
Dams in Dist.*Fractional Annual Rainfall	0.722***	0.734***	0.374*	0.365*	0.321	0.319
	(0.209)	(0.207)	(0.191)	(0.189)	(0.219)	(0.219)
Upstream Dams		0.898**		0.435		0.305
		(0.386)		(0.351)		(0.305)
Upstream Dams*Fractional Annual Rainfall		−0.184*		−0.126		−0.036
		(0.097)		(0.134)		(0.118)
Observations	7,078	7,078	4,063	4,063	5,013	5,013

Note: Standard errors in parentheses, clustered by district. $^{*}p < 0.10$,** $p < 0.05$, *** $p < 0.01$. All results are from 2SLS regressions. Regressions include district fixed effects, state*year interactions, predicted dams*gradient interactions, geography controls and the same set of variables for upstream districts as well as an indicator for whether there is and upstream district. The number of dams is divided by 100 making coefficients multiplied by 100.

C.3 Kaur Replication

I replicate the results in Kaur (2019) that use the data from the 1956-1987 World Bank Agriculture and Climate dataset. Like her, I estimate the following regression where she regesses lnw_{dt}, the log of nominal wages in district d in year t on dummy indicators for positive (Pos_{dt}) and negative (Neg_{dt}) monsoon rainfall realizations in district d in year t, controlling for district (ν_d) and year (ρ_t) fixed effects,[2]

$$lnw_{dt} = \alpha_0 + \alpha_1 Pos_{dt} + \alpha_2 Neg_{dt} + \nu_d + \rho_t + \epsilon_{dt}. \tag{C.5}$$

Table C.3 reports the regression of the rainfall shock indicators on log nominal wages for two periods, the 1956-1987 sample used in Kaur (2019) and the 1971-1987 sample that

[2]Positive and negative shocks are defined as rainfall realizations for the first month of the monsoon that fall above the 80th percentile or below the 20th percentile of a district's typical rainfall realization.

overlaps with the years in which the Duflo and Pande (2007) instrument for dams is available. Focusing first on the first three columns that replicate the results in C.3, estimates in column 1 show that positive rainfall shocks result in an upward wage adjustment. Column 2 shows that positive shocks in the previous year leads to a persistently higher wages in the following year, even when that year has no shock or even a negative shock, as seen by the positive coefficient in column 3. Note also that there is no clear effect of past negative shocks in any of the specifications, which is consistent with downward nominal wage rigidity.

In contrast, when considering the 1971-1987 period of interest, there is no evidence that this pattern holds during this period. We cannot reject that a positive shock this year or last year has no effect on current wages. Furthermore, the negative and marginally significant coefficient in column 6 suggests that negative shocks in past years may actually lower wages.

The contrast in results between these two time periods, hints at a more fundamental issue that complicates identification of nominal wage rigidity in the 1971-1987 sample. The average inflation rate for the 1956-1971 sample was 6.7% while for the 1971-1987 sample it was 7.5%, almost a full percentage point higher. As nominal wage rigidity models predict, downward wage rigidity is less likely to cause distortionary effects in high inflation environments and will thus be more difficult to detect in the 1971-1987 data. Given that the magnitude of Kaur's estimated effects on log nominal wages are in the 2 percentage point range, the higher average inflation in the usable sub-sample years could absorb the effects measured by Kaur making downward wage rigidity less binding and more difficult to observe.

Following C.3 methodology, I proceed to apply the full test defined by the following specification, where $NonPos_{d,t-1}$ is an indicator for a non-positive shock last year and $\sum_{k=2}^{K} \phi_k Pos_{d,t-k}$ controls for positive shocks 2 and 3 years ago,

$$
\begin{aligned}
lnw_{dt} = \beta_0 &+ \beta_1 Pos_{dt} + \beta_2 NonPos_{d,t-1} Neg_{dt} + \beta_3 Pos_{d,t-1} Neg_{dt} + \beta_4 Pos_{d,t-1} Zero_{dt} \\
&+ \sum_{k=2}^{K} \phi_k Pos_{d,t-k} + \nu_d + \rho_t + \epsilon_{dt}.
\end{aligned}
\tag{C.6}
$$

Results are displayed in table C.4. Looking at the full 1956-1987 sample used in columns 1 and 2, the coefficients estimated match those found in Kaur (2019) and are consistent with downward nominal wage rigidity. β_1 and β_2 suggest that wages adjust upwards in positive shock years but not downwards following negative shocks. Furthermore, β_3 and β_4 display the ratcheting pattern consistent with downwards wage rigidity: Wages are higher in district that experiences recent past positive rainfall shocks, even if current rainfall realizations are not particularly good, because of downward wage rigidity. Note however that when looking at the 1971-1987 sub-sample these patterns do not hold, as might be expected given the results from table C.3. There is no evidence of asymmetric wage adjustments to current positive or negative rainfall shocks (if anything, there is more evidence for downward wage adjustments), and no evidence of wage ratcheting for this time period. Again, the difference in underlying inflation rates which were of 7.5% between 1971 and 1987 and only 5.6%

between 1956 and 1970 could be part of what explained the lack of observed effects in the later part of the sample, the mechanics of which are discussed in appendix C.4.

Tables and Figures for Appendix C.3

Table C.3: Effect of Rainfall Shocks on Wages

	1956-1987			*1971-1987*		
	All Obs.	All Obs.	No Shock this Year	All Obs.	All Obs.	No Shock this Year
	(1)	(2)	(3)	(4)	(5)	(6)
Positive Shock this Year	0.021**			0.003		
	(0.009)			(0.009)		
Negative Shock this Year	−0.004			−0.014		
	(0.010)			(0.012)		
Positive Shock last Year		0.017**	0.021**		0.003	−0.003
		(0.009)	(0.010)		(0.008)	(0.011)
Negative Shock last Year		0.007	−0.001		−0.001	−0.023*
		(0.009)	(0.011)		(0.012)	(0.013)
District by Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	7,680	7,680	4,806	4,080	4,080	2,532

Note: Standard errors in parenthesis are clustered by region-year. *$p < 0.10$,** $p < 0.05$, *** $p < 0.01$. All regressions include district and year fixed effects. Positive (Negative) shocks are defined as rainfall realizations in the first month of the monsoon that fall in the upper (lower) quantile of rainfall realizations for that district. Column 3 restricts the analysis to observations where there was a negative or no shock this year.

Table C.4: Test for Wage Adjustment

	Log Nominal Wage			
	1956-1987		*1971-1987*	
	(1)	(2)	(3)	(4)
Last Year: (0,-,+); This Year: (+)	0.026***	0.043***	0.003	0.010
	(0.009)	(0.009)	(0.009)	(0.010)
Last Year: (0,-); This Year: (-)	−0.011	−0.014	−0.019	−0.020
	(0.010)	(0.010)	(0.013)	(0.013)
Last Year: (+); This Year: (-)	0.035*	0.052**	−0.004	0.003
	(0.020)	(0.021)	(0.020)	(0.020)
Last Year: (+); This Year: (0)	0.020**	0.037***	0.0005	0.008
	(0.010)	(0.011)	(0.010)	(0.011)
Prior Shock History Controls	No	Yes	No	Yes
District by Year Fixed Effects	Yes	Yes	Yes	Yes
Observations	7,680	7,680	4,080	4,080

Note: Standard errors in parenthesis are clustered by region-year. $^{*}p < 0.10$,** $p < 0.05$, *** $p < 0.01$. All regressions include district and year fixed effects. Positive (Negative) shocks are defined as rainfall realizations in the first month of the monsoon that fall in the upper (lower) quantile of rainfall realizations for that district.The shock variables are indicators set to 1 if the district experienced the sequence of shocks described. The omitted category it "Last Year: (0,-); This Year: (0)". Columns 2 and 4 include controls for positive shocks two and three years ago. Columns 3 and 4 reestimate the specification on the subset of data used to test the effect of dams.

C.4 Nominal Wage Rigidity and Inflation

Nominal wage rigidity behaviors will be less binding in high inflation environments. Figure C.2 shows the inflation rates prevalent in the dataset alongside the number of districts where the percentage change in the nominal wage is within 0% and 1% included as well as the nominal wage growth rate. It is clear that the period from 1971 to 1987 is characterized by higher inflation rates and significant volatility in prices as compared to the period from 1956 to 1970. Figure C.3 plots the distribution of the percentage change in year on year wages. The top two figures that use data from 1956 to 1987 replicate the figures in Kaur

(2019). The bottom two figures show the distributions when limiting the data to the years 1971-1987. These histogram suggest that the exclusion of the earliest years in the world bank dataset, years in which inflation was particularly low in India, likely explains the difference in the magnitude of the coefficients in table C.4 for the 1956-1987 time period versus the 1971-1987 time period. Figure C.3 illustrates the differences between the 1956-87 and 1971-1987 samples. While there is still some bunching visible at 0% nominal wage change is in the 1971-1987 sample, it is much less pronounced and thus more difficult to detect.

I replicate Kaur's results examining the impact of inflation on measured nominal wage rigidities. I follow her approach and estimate the specification from equation C.6, interacting the shock categories with I_t, a measure of inflation as detailed below,

$$\begin{aligned} lnw_{dt} = \gamma_0 &+ \gamma_1 Pos_{dt} + \gamma_2 NonPos_{d,t-1} Neg_{dt} \\ &+ \gamma_3 Pos_{d,t-1} Neg_{dt} + \gamma_4 Pos_{d,t-1} Zero_{dt} \\ &+ \psi_1 Pos_{dt} \times I_t + \psi_2 NonPos_{d,t-1} Neg_{dt} \times I_t \\ &+ \psi_3 Pos_{d,t-1} Neg_{dt} \times I_t + \psi_4 Pos_{d,t-1} Zero_{dt} \times I_t \\ &+ \sum_{k=2}^{K} \phi_K Pos_{d,t-k} + \delta_d + \rho_t + \epsilon_{dt} \end{aligned} \tag{C.7}$$

Table C.5 reports the results of the replication on the 1956-1987 sample in columns 1 and 4. Results using the the 1971-1987 sample are reported in columns 2 and 5. To capture the component of inflation that is nationally determined, inflation is measured as the average inflation in other states over the course of the year in columns 1, 2 and 3 and an indicator for if this value was over 6% in columns 4, 5 and 6.

When evaluating the results using the full 1956-1971 sample used in Kaur (2019), we see that the ψ coefficients in the specification above demonstrate a pattern consistent with nominal wage rigidity. The coefficients in row 2 show that inflation does not affect nominal wages when a district experiences a contemporaneous positive shock, but when facing a contemporaneous drought, districts with high inflation experience lower wages that the omitted category as seen in row 4. Similarly, inflation reduces the effect of ratcheting in rows 6 an 8, as districts that experienced a positive shock in the previous year that have high inflation see less upward ratcheting of current year wages.

The ψ coefficient estimates using the 1971-1987 sample demonstrate a broadly similar pattern though the magnitude of the coefficients is somewhat smaller and they are not statisitcally significant. Furthermore, the estimates of the γ coefficients remain small and insignificant as in table C.4.

To consider the effects on inflation on the interaction of dams and rainfall shocks, I augment equation C.7, adding the relevant interactions with the dams instrument. I estimate the following,

$$
\begin{aligned}
lnw_{dt}) = \gamma_0 &+ \sum_{k=2}^{5} \gamma_{2k} S_{dt} + \sum_{k=2}^{5} \psi_{2k} S_{dt} \times I_t \\
&+ \delta_1 D_{dt} + \delta_2 D_{dt}^U + \pi_1 D_{dt} \times I_t + \pi_2 D_{dt}^U \times I_t \\
&+ \sum_{k=2}^{5} \lambda_{2k} S_{dt} \times D_{dt} + \sum_{k=2}^{5} \lambda_{3k} S_{dt} \times D_{dt}^U \\
&+ \sum_{k=2}^{5} \pi_{2k} S_{dt} \times D_{dt} \times I_t + \sum_{k=2}^{5} \pi_{3k} S_{dt} \times D_{dt}^U \times I_t \\
&+ \sum_{k=2}^{K} \phi_K Pos_{d,t-k} + \delta_4 Z_{dt} + \delta_5 Z_{dt}^U + \delta_d + \nu_{st} + \epsilon_{dt}
\end{aligned}
\tag{C.8}
$$

Results are reported in columns 3 and 6 of table C.5. Given the lack of significant results when looking at the effects of nominal wage rigidity in the 1971-1987 time frame, it is not surprising that adding the inflation interactions does not yield any valuable insight or significant results.

Tables and Figures for Appendix C.4

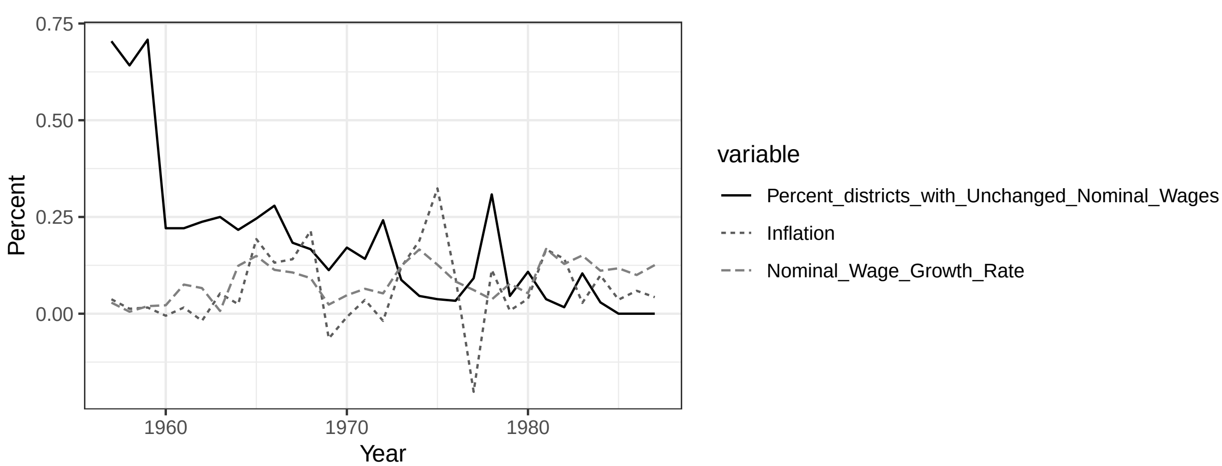


Figure C.2: Wage Rigidity, Nominal Wage Growth and Inflation

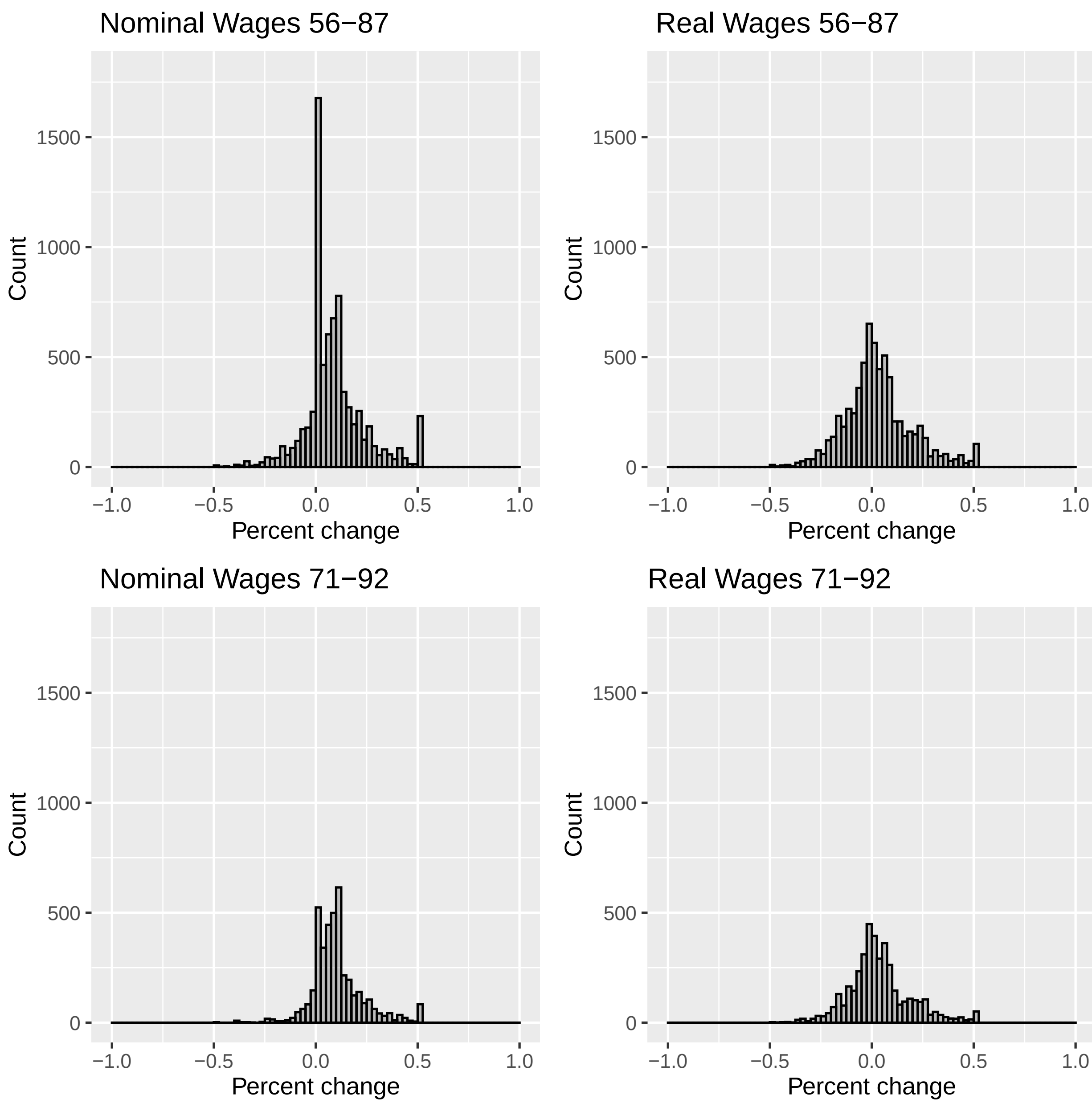


Figure C.3: Distributions of the Year-on-Year Percent Changes in Wages

Table C.5: Inflation, Dams and Log Nominal Agricultural Wages

	Log Nominal Daily Wages					
	Average Inflation Rate			High Inflation Rate		
	56-87	71-87	71-87	56-87	71-87	71-87
Last Year: (0,-,+); This Year: (+)	0.042*** (0.010)	0.009 (0.010)	0.005 (0.015)	0.047*** (0.011)	−0.001 (0.012)	−0.016 (0.016)
Inflation Measure*[(0,-,+);+]	0.010 (0.094)	0.033 (0.078)	0.094 (0.152)	−0.013 (0.019)	0.025 (0.018)	0.047* (0.026)
Last Year: (0,-); This Year: (-)	0.0001 (0.012)	−0.008 (0.014)	0.004 (0.015)	0.001 (0.013)	−0.019 (0.016)	−0.009 (0.018)
Inflation Measure*[(0,-);-]	−0.186* (0.104)	−0.165 (0.137)	−0.003 (0.187)	−0.032 (0.020)	−0.001 (0.026)	0.024 (0.033)
Last Year: (+); This Year: (-)	0.083*** (0.025)	0.022 (0.022)	−0.012 (0.024)	0.085*** (0.029)	0.023 (0.027)	−0.019 (0.029)
Inflation Measure*[+;-]	−0.479** (0.204)	−0.283 (0.232)	−0.058 (0.274)	−0.082** (0.037)	−0.044 (0.047)	−0.002 (0.046)
Last Year: (+); This Year: (0)	0.056*** (0.014)	0.020 (0.017)	0.038 (0.027)	0.057*** (0.015)	0.020 (0.018)	0.016 (0.023)
Inflation Measure*[+,0]	−0.247** (0.097)	−0.125 (0.106)	−0.611 (0.400)	−0.045** (0.020)	−0.019 (0.021)	−0.021 (0.037)
Dams in District			1.267** (0.630)			0.997** (0.505)
Dams Upstream			−0.119 (0.088)			−0.050 (0.072)
Dams in Dist.*Inf. Measure			−0.998 (1.615)			0.176 (0.185)
Dams Upstream*Inf.Measure			0.882 (0.691)			0.015 (0.045)
Dams in Dist.*[(+,0,-);+]			−0.161 (0.229)			−0.087 (0.153)
Dams Upstream*[(+,0,-);+]			−0.010 (0.082)			−0.0004 (0.076)
Dams in Dist*[(0,-);-]			0.474 (0.390)			0.025 (0.189)
Dams Upstream*[(0,-);-]			−0.052 (0.155)			0.050 (0.097)
Dams in Dist*[+;-]			0.308 (0.807)			0.018 (0.484)
Dams Upstream*[+;-]			0.129 (0.269)			0.101 (0.225)
Dams in Dist*[+,0]			−0.884* (0.478)			−0.450 (0.314)
Dams Upstream*[+;0]			0.113 (0.134)			0.039 (0.109)
Dams in Dist.*[(+,0,-);+]*Inf. Measure			1.789 (2.656)			0.259 (0.276)
Dams Upstream*[(+,0,-);+]*Inf. Measure			−1.062 (0.796)			−0.177* (0.103)
Dams in Dist*[(0,-);-]*Inf. Measure			−10.010 (6.535)			−0.402 (0.345)
Dams Upstream*[(0,-);-]*Inf. Measure			1.539 (2.179)			0.049 (0.136)
Dams in Dist*[+;-]*Inf. Measure			−10.010 (12.680)			−0.540 (0.626)
Dams Upstream*[+;-]*Inf. Measure			0.071 (2.887)			0.094 (0.274)
Dams in Dist*[+,0]*Inf. Measure			14.830* (7.853)			1.290** (0.547)
Dams Upstream*[+;0]*Inf. Measure			−2.260 (1.397)			−0.296* (0.171)
Prior Shock History Controls?	Yes	Yes	Yes	Yes	Yes	Yes
District Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	-	Yes	Yes	-
State by Year Fixed Effects	No	No	Yes	No	No	Yes
Observations	7,675	4,080	4,063	7,675	4,080	4,063

Note: Standard errors in parenthesis, clustered at the region by year level.