

**From the holdings of the Department of Special Collections , Princeton University Library,  
Princeton, NJ**

These documents are either in the public domain, or can only be used for educational and research purposes (Fair use) as per U.S. Copyright law (text below). By accessing this file, all users agree that their use falls in the public domain or within fair use as defined by the copyright law. We do not charge any permission or usage fees for the publication of images of material in our collections, including those provided by Princeton University Library, via our website, catalog, or directly from staff. For more information regarding copyright, credit, and citation please review the information on the Copyright, Credit and Citation Guidelines page of our website.

<https://library.princeton.edu/special-collections/policies/copyright-credit-and-citation-guidelines>. If you have any questions, including how to order higher-quality images from the collection, contact Special Collections Public Services via our Ask Us! Form. <https://library.princeton.edu/special-collections/ask-us>

---

Doctoral Dissertations, 2011-2026

Economics

---

2023

# Essays on Preference Formation, Norms Transmission, and Healthcare in Development Economics

Agte, Patrick

Princeton, NJ : Princeton University

---

<http://arks.princeton.edu/ark:/88435/dsp010k225f31v>

*U.S. Copyright law: The copyright law of the United States (Title 17, §108, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be used for any purpose other than private study, scholarship, or research. If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of fair use that user may be liable for copyright infringement. This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve a violation of copyright law. <https://www.copyright.gov/title17/title17.pdf>*

ESSAYS ON PREFERENCE FORMATION, NORMS  
TRANSMISSION, AND HEALTHCARE IN  
DEVELOPMENT ECONOMICS

PATRICK AGTE

A DISSERTATION  
PRESENTED TO THE FACULTY  
OF PRINCETON UNIVERSITY  
IN CANDIDACY FOR THE DEGREE  
OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE BY  
THE DEPARTMENT OF  
ECONOMICS

ADVISERS: JANET CURRIE AND THOMAS FUJIWARA

MAY 2023

© Copyright by PATRICK AGTE , 2023.

All rights reserved.

# Abstract

This dissertation investigates the determinants of preference formation, norms transmission, and healthcare utilization.

Chapter 1, joint with Mariel Bedoya, studies whether temporary work assignments to poor areas affect worker preferences, beliefs, and career choices. We provide evidence on this question using random variation in the assignment of psychologists within a one-year mandatory rural service program in Peru. Psychologists that completed the program in poorer places are later 15% more likely to work for the public sector and 86% more likely to work in the poorest districts in the country. We provide survey evidence that points to increased prosociality as an important mechanism.

Chapter 2, joint with Arielle Bernhardt, studies adherence to caste norms. This paper shows that caste norms are weakened when Hindus live alongside Adivasis, an indigenous minority outside of the caste system. Using a number of estimation strategies, we show that having more Adivasi neighbors decreases Hindus' adherence to a wide range of caste rules. Hindu women in Adivasi-majority villages are 50% more likely to work and villages are more likely to be integrated. We argue that Hindus adhere to caste norms as an investment in status within the caste system, and that this investment is less valuable when Adivasis form a larger share of the village population.

Chapter 3 studies supply-side constraints to healthcare utilization. This paper exploits variation in the timing and location of new public primary healthcare facilities in Peru to analyze the effects of increased access to primary healthcare on birth outcomes. I find that an additional healthcare facility in a rural area reduces the incidence of low birth weight by 1.3 percentage points after an initial adjustment period. This effect seems to be driven by higher quality prenatal care and changes in fertility behavior. Overall, these results suggest that increasing access to primary healthcare can have large health benefits.

# Acknowledgements

I would like to express my sincere gratitude to all those who have supported me throughout my academic journey.

First and foremost, I am deeply indebted to Janet Currie, Thomas Fujiwara, and Rohini Pande for their invaluable encouragement, advice, and patience throughout the past years. I also would like to thank Nicholas Buchholz, Valentina Duque, Erica Field, Kate Ho, Allan Hsiao, Seema Jayachandran, Adam Kapor, Jakub Kastl, Ilyana Kuziemko, Dominic Leggett, Alessandro Lizzeri, Matt Lowe, Christopher Neilson, Natalia Rigol, and Maria Micaela Sviatschi, for helpful advice and comments.

I have been lucky to have two incredible co-authors, Mariel Bedoya and Arielle Bernhardt. Our collaborations pushed the work forward and deepened our friendships.

Two projects relied on intensive fieldwork and I want to thank the field teams for their tireless work, especially Prachi Das, Prasanta Panda, Santosh Sahoo, Zaragoza Limachi, and Jorge Luis Tipisman. The projects were made possible with funding from the Princeton Research Program in Development Economics, the Princeton Industrial Relations Section, the Harvard South Asia Institute, and the Griffin Fund. I am grateful to Winston V. Loyola Balarezo at the Ministry of Health in Peru and to Ram Singh Nayak at the Sambalpur Settlement Office for their time and assistance. I also want to thank Princeton's administrative team, including Valerie Ching, Justine Conoline, Laura Hedden, Kristen Hubbard, Lori Mitrano, and Laura Sciarrotta for the kind support they provided over the years.

My experience in graduate school greatly benefited from the amazing people I met during my time here at Princeton. I would like to thank my friends Dorisz Albrecht, Juanma Castro-Vincenzi, Mateus Dias, Stephanie Hao, Anastasia Karpova, Benny Kleinman, Kwok-Hao Lee, Don Noh, and Christiane Szerman for all the wonderful moments we shared together.

I would like to extend my heartfelt gratitude to my family, particularly my parents, for their unwavering love and encouragement. I thank my fiancé Maria Victoria de la Torre

for enduring countless discussions about economics and making the past years the most incredible time of my life. I am very happy we found each other and I cannot wait for our future together.

For mum, dad, Celine, Torsten, and Vicky.

# Contents

|                                                                                                               |           |
|---------------------------------------------------------------------------------------------------------------|-----------|
| Abstract . . . . .                                                                                            | 3         |
| Acknowledgements . . . . .                                                                                    | 4         |
| List of Tables . . . . .                                                                                      | 10        |
| List of Figures . . . . .                                                                                     | 13        |
| <br>                                                                                                          |           |
| <b>1 The Making of a Public Sector Worker: The Causal Effects of Temporary Work Assignments to Poor Areas</b> | <b>15</b> |
| 1.1 Introduction . . . . .                                                                                    | 15        |
| 1.2 Background . . . . .                                                                                      | 20        |
| 1.2.1 Context . . . . .                                                                                       | 20        |
| 1.2.2 Conceptual Framework . . . . .                                                                          | 24        |
| 1.3 Study Design . . . . .                                                                                    | 26        |
| 1.4 Results . . . . .                                                                                         | 30        |
| 1.5 Conclusion . . . . .                                                                                      | 38        |
| 1.6 Tables and Figures . . . . .                                                                              | 39        |
| 1.7 Appendix . . . . .                                                                                        | 50        |
| <br>                                                                                                          |           |
| <b>2 The Economics of Caste Norms: Purity, Status, and Women’s Work in India</b>                              | <b>71</b> |
| 2.1 Introduction . . . . .                                                                                    | 71        |
| 2.2 Historical Background . . . . .                                                                           | 80        |
| 2.2.1 Social Hierarchy and Cultural Practices in Adivasi-Hindu Villages . .                                   | 81        |

|       |                                                                  |     |
|-------|------------------------------------------------------------------|-----|
| 2.2.2 | Conceptual Framework . . . . .                                   | 85  |
| 2.3   | Data and Measurement . . . . .                                   | 90  |
| 2.4   | Social Hierarchy and the Take-up of Caste Purity Norms . . . . . | 93  |
| 2.5   | OLS Estimates . . . . .                                          | 95  |
| 2.5.1 | Empirical Specification . . . . .                                | 95  |
| 2.5.2 | OLS Results . . . . .                                            | 96  |
| 2.5.3 | Robustness and Validity Checks . . . . .                         | 99  |
| 2.6   | Spatial Regression Discontinuity Estimates . . . . .             | 101 |
| 2.6.1 | The Railroad and In-Migration to Sambalpur, Odisha . . . . .     | 102 |
| 2.6.2 | Regression Discontinuity Empirical Specification . . . . .       | 106 |
| 2.6.3 | Regression Discontinuity Results . . . . .                       | 109 |
| 2.7   | Mechanisms . . . . .                                             | 114 |
| 2.8   | Discussion . . . . .                                             | 117 |
| 2.9   | Tables and Figures . . . . .                                     | 118 |
| 2.10  | Appendix . . . . .                                               | 133 |

### **3 The Impact of Public Primary Healthcare Facilities on Birth Outcomes:**

|       |                                                    |            |
|-------|----------------------------------------------------|------------|
|       | <b>Evidence from Peru</b>                          | <b>154</b> |
| 3.1   | Introduction . . . . .                             | 154        |
| 3.2   | Context . . . . .                                  | 157        |
| 3.3   | Data and Estimation . . . . .                      | 159        |
| 3.3.1 | Data . . . . .                                     | 159        |
| 3.3.2 | Empirical Strategy . . . . .                       | 161        |
| 3.4   | Results . . . . .                                  | 164        |
| 3.4.1 | Effect of Facilities on Low Birth Weight . . . . . | 164        |
| 3.4.2 | Robustness Checks . . . . .                        | 165        |
| 3.4.3 | Mechanisms . . . . .                               | 168        |

|       |                                 |            |
|-------|---------------------------------|------------|
| 3.4.4 | Other Health Outcomes . . . . . | 171        |
| 3.5   | Conclusion . . . . .            | 171        |
| 3.6   | Tables and Figures . . . . .    | 173        |
| 3.7   | Appendix . . . . .              | 184        |
|       | <b>Bibliography</b>             | <b>209</b> |

# List of Tables

|        |                                                                            |    |
|--------|----------------------------------------------------------------------------|----|
| 1.1    | Internal Validity Checks . . . . .                                         | 41 |
| 1.2    | Effect of Lottery Rank on SERUMS Location . . . . .                        | 42 |
| 1.3    | Effect of Lottery Rank on SERUMS Location . . . . .                        | 43 |
| 1.4    | Effect of Lottery Rank on Employment Outcomes . . . . .                    | 44 |
| 1.5    | Effect of Lottery Rank on Additional Employment Outcomes . . . . .         | 45 |
| 1.6    | Effect of Lottery Rank on Prosociality . . . . .                           | 46 |
| 1.7    | Additional Evidence for Supply-Side Channels . . . . .                     | 47 |
| 1.8    | Effect of Lottery Rank on Skill and Network Formation . . . . .            | 48 |
| 1.9    | Effect of Lottery Rank on Job Availability . . . . .                       | 49 |
| 1.A.1  | Bonus Point Structure . . . . .                                            | 58 |
| 1.A.2  | Effect of Lottery Rank on SERUMS Location for Nutritionists . . . . .      | 59 |
| 1.A.3  | Attrition Check . . . . .                                                  | 60 |
| 1.A.4  | Balance Check Conditional on Survey Completion . . . . .                   | 61 |
| 1.A.5  | Relationship Between Lottery Rank and SERUMS Location Attributes . .       | 62 |
| 1.A.6  | Effect of Lottery Rank on Donation Subindex Components . . . . .           | 63 |
| 1.A.7  | Effect of Lottery Rank on Additional Donation Outcomes . . . . .           | 64 |
| 1.A.8  | Effect of Lottery Rank on Beliefs . . . . .                                | 65 |
| 1.A.9  | Effect of Lottery Rank on Perceptions of Jobs in Public Sector . . . . .   | 66 |
| 1.A.10 | Effect of Lottery Rank on Perceptions of Jobs in Poorest Districts . . . . | 67 |
| 1.A.11 | Effect of Lottery Rank on Additional Job Preferences . . . . .             | 68 |
| 1.A.12 | Effect of Lottery Rank on Main Outcomes by Birthplace . . . . .            | 69 |

|        |                                                                                                                              |     |
|--------|------------------------------------------------------------------------------------------------------------------------------|-----|
| 1.A.13 | Randomization Inference . . . . .                                                                                            | 70  |
| 2.1    | Average Purity Norms by Caste Group . . . . .                                                                                | 124 |
| 2.2    | Work IHDS . . . . .                                                                                                          | 125 |
| 2.3    | Purity Norms Indexes . . . . .                                                                                               | 126 |
| 2.4    | RDD Balance Checks . . . . .                                                                                                 | 127 |
| 2.5    | First Stage, Work Outcomes, and Beliefs . . . . .                                                                            | 128 |
| 2.6    | Intercaste Purity Norms and Food Taboos . . . . .                                                                            | 129 |
| 2.7    | Social Interactions . . . . .                                                                                                | 130 |
| 2.8    | Balance Check: Comparison of Villages Where Adivasis Historically Prac-<br>ticed Plough vs. Non-Plough Agriculture . . . . . | 131 |
| 2.9    | Impacts of Historical Adivasi Agricultural Practices on Current Gender Norms                                                 | 132 |
| 2.A.1  | Village Differences by Adivasi Share . . . . .                                                                               | 139 |
| 2.A.2  | Political Participation . . . . .                                                                                            | 140 |
| 2.A.3  | Mobility Index Components IHDS . . . . .                                                                                     | 141 |
| 2.A.4  | Marriage and Intercaste Relations Restrictions Components IHDS . . . . .                                                     | 142 |
| 2.A.5  | Food Taboos Components IHDS . . . . .                                                                                        | 143 |
| 2.A.6  | Work IHDS with Village Fixed Effects . . . . .                                                                               | 144 |
| 2.A.7  | Purity Norms Indexes with Village Fixed Effects . . . . .                                                                    | 145 |
| 2.A.8  | Impact of Plough and Non-Plough Adivasis . . . . .                                                                           | 146 |
| 2.A.9  | RDD Results with Conley Standard Errors . . . . .                                                                            | 147 |
| 2.A.10 | Randomization Inference . . . . .                                                                                            | 148 |
| 2.A.11 | Alternative Labor Demand and Supply Explanations . . . . .                                                                   | 149 |
| 2.A.12 | Geographic Balance Checks for Watchmen Analysis . . . . .                                                                    | 150 |
| 2.A.13 | Purity Norms Using Watchman Block Assignment . . . . .                                                                       | 151 |
| 2.A.14 | Effect of Density on Hindus . . . . .                                                                                        | 152 |
| 2.A.15 | Effect of Adivasi Share on Hindus and on Adivasis . . . . .                                                                  | 153 |

|        |                                                                                 |     |
|--------|---------------------------------------------------------------------------------|-----|
| 3.1    | Summary Statistics by Group . . . . .                                           | 182 |
| 3.2    | The Effect of an Additional I-2 Facility on Low Birth Weight in Rural Districts | 183 |
| 3.A.1  | Count of Districts by Number of I-2 Facilities in 2000 And 2018 . . . . .       | 190 |
| 3.A.2  | Robustness Check: Low Birth Weight (DHS) . . . . .                              | 191 |
| 3.A.3  | Robustness Check: Place of delivery - respondent's home . . . . .               | 192 |
| 3.A.4  | Robustness Check: Place of delivery - private clinic . . . . .                  | 193 |
| 3.A.5  | Robustness Check: Place of delivery - hospital minsa . . . . .                  | 194 |
| 3.A.6  | Robustness Check: Place of delivery - health post minsa . . . . .               | 195 |
| 3.A.7  | Robustness Check: Underweight . . . . .                                         | 196 |
| 3.A.8  | Robustness Check: Anemia . . . . .                                              | 197 |
| 3.A.9  | Robustness Check: Number of Prenatal Visits . . . . .                           | 198 |
| 3.A.10 | Robustness Check: Above-Median Number of Prenatal Visits . . . . .              | 199 |
| 3.A.11 | Robustness Check: Prenatal Care Conducted by Doctor . . . . .                   | 200 |
| 3.A.12 | Robustness Check: Iron Supplements . . . . .                                    | 201 |
| 3.A.13 | Robustness Check: Tetanus Injections . . . . .                                  | 202 |
| 3.A.14 | Robustness Check: Vitamin Doses . . . . .                                       | 203 |
| 3.A.15 | Robustness Check: Number of Births . . . . .                                    | 204 |
| 3.A.16 | Robustness Check: Wanted Pregnancy . . . . .                                    | 205 |
| 3.A.17 | Robustness Check: Condom Usage . . . . .                                        | 206 |
| 3.A.18 | Robustness Check: Neonatal Mortality Rate . . . . .                             | 207 |
| 3.A.19 | Robustness Check: Infant Mortality Rate . . . . .                               | 208 |

# List of Figures

|       |                                                                              |     |
|-------|------------------------------------------------------------------------------|-----|
| 1.1   | Healthcare Utilization by Household Wealth . . . . .                         | 39  |
| 1.2   | Timeline of Data Collection . . . . .                                        | 40  |
| 1.A.1 | Psychologist Density and Mental Healthcare Utilization by District Poverty   | 50  |
| 1.A.2 | Share of Health Facilities Managed by Ministry of Health by District Poverty | 51  |
| 1.A.3 | Distribution of Government Poverty Index Across SERUMS Positions . . .       | 52  |
| 1.A.4 | Map of SERUMS Position . . . . .                                             | 53  |
| 1.A.5 | Robustness Checks: Public Sector Employment . . . . .                        | 54  |
| 1.A.6 | Robustness Checks: Ministry of Health Employment . . . . .                   | 55  |
| 1.A.7 | Robustness Checks: Employment in Poorest Districts . . . . .                 | 56  |
| 1.A.8 | Robustness Checks: Prosociality Index . . . . .                              | 57  |
| 2.1   | Sambalpur District in 1906 . . . . .                                         | 118 |
| 2.2   | Adivasi Population Share and Women’s Work Outside the Home . . . . .         | 119 |
| 2.3   | Purity Norms Adoption in the 20th and 21st Centuries by Caste . . . . .      | 120 |
| 2.4   | IHDS Adoption of Purity Norms by Caste Group . . . . .                       | 121 |
| 2.5   | RDD Plots . . . . .                                                          | 122 |
| 2.A.1 | IHDS Distribution of Adivasi Share . . . . .                                 | 133 |
| 2.A.2 | Non-Plough and Plough Adivasi Share Effects on Caste Women’s Work . .        | 134 |
| 2.A.3 | Robustness to RD Bandwidth . . . . .                                         | 135 |
| 2.A.4 | Additional Robustness Checks . . . . .                                       | 137 |
| 3.1   | Number of New I-2 Facilities over Time . . . . .                             | 173 |

|       |                                                                                                                                                      |     |
|-------|------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 3.2   | The Effect of an Additional I-2 Facility on Low Birth Weight . . . . .                                                                               | 174 |
| 3.3   | Robustness Checks with DHS Data (Rural Districts) . . . . .                                                                                          | 175 |
| 3.4   | The Effect of an Additional I-2 Facility on Place of Delivery in Rural Districts                                                                     | 176 |
| 3.5   | The Effect of an Additional I-2 Facility on Respondent's Health in Rural<br>Districts . . . . .                                                      | 177 |
| 3.6   | The Effect of an Additional I-2 Facility on Prenatal Visits in Rural Districts                                                                       | 178 |
| 3.7   | The Effect of an Additional I-2 Facility on Prenatal Care in Rural Districts                                                                         | 179 |
| 3.8   | The Effect of an Additional I-2 Facility on Family Planning in Rural Districts                                                                       | 180 |
| 3.9   | The Effect of an Additional I-2 Facility on Mortality Rates in Rural Districts                                                                       | 181 |
| 3.A.1 | Map of Facility Openings . . . . .                                                                                                                   | 184 |
| 3.A.2 | Effect of Facility Openings on Low Birth Weight in Rural Areas (Without<br>Districts that Received a New I-2 Facility Between 2000 and 2002) . . . . | 185 |
| 3.A.3 | The Effect of an Additional I-2 Facility on Delivering at Home by Wealth<br>Status . . . . .                                                         | 186 |
| 3.A.4 | The Effect of an Additional I-2 Facility on Delivering at a Public Hospital<br>by Wealth Status . . . . .                                            | 187 |
| 3.A.5 | The Effect of an Additional I-2 Facility on Prenatal Care in Rural Districts                                                                         | 188 |
| 3.A.6 | The Effect of an Additional I-2 Facility on Family Planning in Rural Districts                                                                       | 189 |

# Chapter 1

## The Making of a Public Sector Worker: The Causal Effects of Temporary Work Assignments to Poor Areas

### 1.1 Introduction

A growing body of research shows that beliefs and preferences are malleable (Kosse et al., 2020; Cappelen et al., 2020; Abeler et al., 2021; Rao, 2019; Alan et al., 2022). While most of the existing work is based on low-stakes decisions in lab-in-the-field experiments, these insights have the potential to address broader economic behavior and inform the optimal designs of markets. Policies that aim to harness the endogeneity of preferences and beliefs could thus supplement existing approaches like monetary incentives to induce desired behaviors. For example, making individuals more patient could increase educational investments, while

higher levels of prosociality could reduce conflict, decrease shirking, and affect occupational choices.

We focus on a particular policy problem – the unequal distribution of health workers in urban versus rural areas in many developed and developing countries – and ask whether temporary assignments of health workers to poorer areas can affect subsequent career decisions. The question is motivated by previous research showing that contact with disadvantaged communities increases prosocial behavior and reduces stereotypes (Rao, 2019; Lowe, 2021; Mousa, 2020) and that public sector workers, who are often the only providers in remote areas, tend to be more prosocial (Cowley and Smith, 2014; Banuri and Keefer, 2013). To study this question, we use random variation in the location assignment of psychologists within a mandatory 12-month rural service program (SERUMS) in Peru.<sup>1</sup> As part of the program, health workers are sent to remote public primary healthcare facilities to provide services to local communities after the completion of their medical degree. Participation in this program is mandatory for all health workers who ever want to work in the public sector. Similar policies are used by more than 70 countries (Frehywot et al., 2010).

Our econometric strategy exploits exogenous variation in psychologists’ location choice sets: psychologists are given their choice over facilities based on a randomized ranking (the individual who is drawn first gets first pick over locations, the individual who is drawn second gets their pick among the remaining locations, etc.). We collaborated with the Ministry of Health in Peru to access administrative data on health worker placements and conduct an online survey with 559 psychologists who completed the rural service program 3-33 months prior. As part of the survey, we collected information on employment outcomes, the respondent’s rank during the assignment progress, preferences, beliefs, and the development of skills and networks during the rural service program. We create a prosociality index based on

---

<sup>1</sup>The acronym stands for the *Servicio Rural y Urbano Marginal de Salud*.

stated job preferences, redistributive preferences, and (past, hypothetical, and incentivized) donation decisions.

We start by documenting that psychologists who chose later during the lottery process completed the rural service program in poorer and more remote locations. We then show that this experience affects subsequent career outcomes. Relative to psychologists who were given first choice over locations, respondents in the bottom rank tercile are 8.8 percentage points, or 16%, more likely to work for the public sector after the program. These effects are concentrated in facilities that provide healthcare services to poor households. Besides changes in the type of employer, we also observe a change in respondents' later career location. Psychologists in the bottom tercile of lottery ranks are 8.7 percentage points, or 86%, more likely to subsequently work in the poorest districts in the country. We show that these effects are not driven by inertia since less than 5% of respondents continue to stay at the facility in which they completed the rural service program. Importantly, we do not observe a difference in monthly income based on the respondent's rank. Instead, psychologists seem to face a trade-off between public sector work in remote areas with fewer amenities and private sector jobs in cities.

What is driving these career effects? Guided by a simple framework of two-sided labor market matching, we consider four primary explanations: a) increased prosociality (e.g., stronger preferences for helping the poor), b) changes in beliefs about job attributes, c) changes in other job preferences, and d) differences in hireability due to differential skill and network formation during the rural service program. We discuss each explanation in turn. We first show that psychologists in the bottom tercile of lottery ranks score 0.17 standard deviations higher on a prosociality index than psychologists in the top tercile. They are more likely to say that opportunities to help the poor are very important for their job choice, more likely to support that the government should increase aid to the poor, and more likely to donate. We show that respondents in the bottom rank tercile also seem to update

their beliefs on the effectiveness of the public sector relative to the private sector, but find no differences in the perceived attributes of jobs in rich versus poor areas. We do not observe changes in other job preferences.

An alternative explanation for the differences in employment outcomes could be differences in the hireability of psychologists. For example, psychologists who complete the mandatory service program in poorer areas might receive worse training, which could make them less attractive to the private sector. Similarly, workers in poorer areas might have fewer opportunities to develop their professional networks, which could make it harder to find a private sector job. We address these concerns in four ways. First, we show that respondents in the bottom rank tercile are more likely to say that the rural service program increased their willingness to work for the Ministry of Health, providing additional evidence that the program location led to a change in the supply side. Further, consistent with an increase in prosociality, 54% of respondents say that the rural service program made them more willing to work for the Ministry of Health because it became more important for them to have a job in which they can help the poorest. Second, we show that the rank tercile did not affect the perceptions of how the rural service program impacted skill formation and professional networks. We also do not find that respondents that chose later give different answers about how the changes in skills and networks helped them with getting their current job. Overall, only 20% of respondents say that someone they met through the rural service program helped them to get their current job, emphasizing that network formation is not a core feature of the program. Finally, we show that the lottery rank does not affect the perceived availability of job options in the public and private sectors.

We argue that contact with rural poverty increased prosociality, thereby increasing the willingness to work for the public sector in poor areas. We provide qualitative evidence that documents how the experience affected the perspective of respondents. Furthermore,

consistent with this explanation, we show suggestive evidence that the effects are larger for respondents who had less contact with rural poverty prior to the program.

This paper relates to three bodies of research. We show that temporary assignments to poor areas make health workers more likely to subsequently work in the public sector and in poor areas. These findings supplement existing work on how to address imbalances in geographic distribution of health workers (Dal Bó et al., 2013; Bobba et al., 2021; Costa et al., 2021). We also contribute to a nascent literature on the endogeneity of preferences and beliefs (Kremer et al., 2019; Kosse et al., 2020). Previous studies have shown that contact with disadvantaged communities can reduce stereotypes, change redistributive preferences, and increase prosocial behavior (Lowe, 2021; Rao, 2019; Londoño-Vélez, 2022; Mo and Conn, 2018; Mousa, 2020). We add to this literature by showing that contact with the poor through temporary work assignments can also affect high-stake decisions like career outcomes. In related work, Dobbie and Fryer Jr (2015) show that participation in the volunteer program “Teach for America” make participants more likely to work in the education sector after the program. We extend these findings by studying a public policy that is implemented by more than 70 countries and examining a broader range of employment outcomes. We also focus on variation in assignments within such a program instead of studying the extensive margin effect of program participation, allowing us to rule out that the effects are driven by other program components like training content. Finally, our paper speaks to existing work on the long-run effects of early career experiences (Oreopoulos et al., 2012; Fadlon et al., 2022; Angrist and Chen, 2011). Our paper is related to Fadlon et al. (2022) who find that female physicians in Denmark who received unfavorable internship positions are more likely to sort into less desirable local labor markets in the long run, partly due to worse professional networks. Our setting differs from theirs since network effects are much less important for rural service programs than for internships during residency. Instead, we show that our career effects are driven by changes in prosociality.

The rest of the paper is organized as follows. Section 2 describes the institutional setting and the conceptual framework. In Section 3, we describe the study design. Section 4 reports our results. Section 5 concludes.

## 1.2 Background

We begin by describing the healthcare sector and the mandatory rural service program in Peru. We then describe a basic framework of how temporary assignments to poorer locations can affect career choices.

### 1.2.1 Context

**The health workforce in Peru.** Similar to many other countries, Peru faces a geographical imbalance of health workers. The health worker density is almost twice as high in urban than in rural areas (PAHO, 2017). Appendix Figure 1.A.1 shows the relationship between the number of psychologists in the public sector per 100,000 people and a district-level poverty index.<sup>2</sup> The index is created by FONCODES, a government agency, based on poverty indicators in the 2007 census.<sup>3</sup> The index ranges from 0 to 1 and indicates the percentile rank of the district relative to the rest of the country. The government uses this ranking to classify districts into poverty quintiles such that each quintile has roughly the same population.<sup>4</sup> We use the poverty index and the poverty quintiles throughout the paper as our preferred measure of location classifications. The blue line in Appendix Figure 1.A.1 shows that the richest districts have around 22 psychologists per 100,000 people, whereas the poorest districts only have a psychology density of around 13. At the same time, the need for better

---

<sup>2</sup>Districts are the third-level administrative unit in Peru and have a median population of 4,366 people according to the 2007 census.

<sup>3</sup>The index is based on the following indicators: the share of homes without drinking water, the share of homes without proper sanitation, the share of homes without electricity, the share of households with at least one malnourished member, the illiteracy rate of women aged 15 and over, and the population share of children aged 0-12 years.

<sup>4</sup>Since poorer districts tend to have a smaller population, the poorest quintile consists of more districts.

access to healthcare services is often larger in poorer areas. For example, the share of people with mental health conditions who seek care is only half as large in the poorest districts as the in the richest districts in the country (orange line). We find a similar pattern when looking at a proxy of unfilled vacancies at public facilities. Whereas 51% of public primary facilities that are supposed to provide mental healthcare services do not have a psychologist in districts in the poorest quintile, this share is only 17% in districts in the richest quintile.

**The mandatory rural service program.** To address the shortage of healthcare professionals in rural areas, the government launched the *Servicio Rural y Urbano Marginal de Salud* (SERUMS) program in 1972. The law requires that healthcare professionals have to work for one year in a rural or semi-urban area if they ever want to apply to the public sector.<sup>5</sup> Psychologists can apply to SERUMS after the completion of a 3-year bachelor program. According to anecdotal evidence, many psychologists still apply to the program even if they do not intend to immediately work for the public sector to maximize future career options. A back-of-the-envelope calculation suggests that roughly 70% of psychology graduates apply to SERUMS.<sup>6</sup>

Assignments are either paid or unpaid. Paid positions are only available in districts that are in the three poorest quintiles. Unpaid positions are usually offered in richer locations and are assigned separately. The tasks of psychologists during the rural service program range from the diagnosis and treatment of mental health conditions, such as depression, anxiety, schizophrenia and alcoholism, to providing support to victims of domestic violence.

While all positions are located in rural or semi-urban areas, there is a substantial variation in the poverty level of SERUMS locations. Appendix Figure 1.A.3 shows the distribution of

---

<sup>5</sup>Some professions, including physicians, are also required to complete SERUMS to apply for residency programs.

<sup>6</sup>The calculation is based on the following statistics: (i) university records document that there were 2,537 psychology graduates in 2015, (ii) there were 1,723 SERUMS positions for 3,505 unique psychology applicants across both lottery rounds in 2019, (iii) the average respondent in our survey applied to SERUMS three times. If we assume that the numbers are roughly constant across years and that psychologists stop applying after five times, we would predict that  $\sum_{i=0}^4 2537 * 0.51^i = 4998$  psychologists should apply in a given year. Comparing this to the actual number of applicants in 2019 leads to an application rate of 70%.

the poverty level of SERUMS locations within our sample. Positions are available across the country (see Appendix Figure 1.A.4 for a map). All positions are paid the same (nominal) salary except for approximately 20% of positions that are located in emergency or border areas and receive around 30% higher salaries. In addition, the government offers bonus points to health workers according to the classification of districts into poverty quintiles. Psychologists can use the points when applying for public sector jobs (see Appendix Table 1.A.1 for details). However, since 87% of positions are located in the two poorest quintiles, the variation in bonus points is relatively small. Most of the psychologists in our pilot activities were not aware that the bonus points even applied to them and only 26% of our survey respondents say that the bonus points were important for them when they chose their SERUMS location.<sup>7</sup>

**The assignment process.** The assignment process to healthcare facilities differs across professions. Professions with a nationwide exam like physicians and nurses choose facilities based on merit, while professions without a nationwide exam choose based on a randomized order. We focus on psychologists in our study since they are the largest profession for which the assignment is based on a lottery. Assignments occur twice per year. The government publishes a list of vacancies one month before the lottery process. The list only contains information on the geographical location of the vacancy, the institution that manages the facility, the facility size, the poverty quintile, and whether the position is located in an emergency or border area. The poverty quintile of the position is the most salient characteristic since applicants use it as a proxy for the remoteness and safety level of the location. Applicants often also obtain additional information through personal networks and social media groups.

Based on the location of the university, applicants are either assigned to the central lottery site in Lima or to one of eight regional lottery sites. During the registration process,

---

<sup>7</sup>The bonus points are more important for other professions like physicians and nurses since they are also relevant for residency program applications.

applicants then self-select into one of the regions that are available at their lottery site and decide to which institution they want to apply.<sup>8</sup> With the region-institution combination, applicants are called upon in a random order to select one of the remaining facilities on the list.<sup>9</sup> The government conducts the randomization digitally through the website random.org. Applicants also have the option to withdraw from the lottery when it is their turn. In that case, they can either enter a separate lottery for unpaid positions that takes place in another week or they can wait six months until the next lottery cycle. However, active withdrawal for paid places is rare in practice. A more common occurrence is that applicants are absent when called (10%) or make mistakes (5%). The latter includes applicants who forget to bring the correct documents or who accidentally choose facilities that have already been selected. As we discuss later, we do not find significant differences in acceptance rates based on the lottery position. The process stops once all facilities have been assigned. For psychologists, the program is heavily oversubscribed. In the first round of 2021, 3,886 psychologists applied but only 336 paid positions and 717 unpaid positions were available. Applicants whose names are not drawn have to wait until the next lottery cycle.

**Career choices after SERUMS.** After the mandatory service program, psychologists can make job choices without further restrictions. Less than 5% of respondents continue to stay at the facility in which they completed SERUMS. Psychologists have various career options. According to our survey, 68% work in the health sector after SERUMS. Another 18% work in the education sector, while the rest works in a variety of areas, including public administration, marketing, and human resource management. Within the healthcare sector, most psychologists either work for the Ministry of Health, ESSALUD, or private facilities. The facilities of the Ministry of Health provide healthcare to beneficiaries of the

---

<sup>8</sup>87% of SERUMS positions for psychologists are located at Ministry of Health facilities. Regional lottery sites usually only offer positions within the same region, whereas the central lottery offers positions across eleven regions. 52% of applicants are assigned through the central lottery site.

<sup>9</sup>Exceptions are special cases (pregnant women, women with infants, and applicants with a disability) that are allowed to choose facilities first.

public social protection scheme, *Seguro Integral de Salud* (SIS), consisting of poor households and informal workers. ESSALUD is a public contribution-based social security system for salaried workers with a separate network of facilities. Private facilities consist of individual practices or larger hospitals for which patients either have to pay out-of-pocket or receive coverage through private insurance schemes. As shown in Figure 1.1, the large majority of poor households receive healthcare from Ministry of Health facilities. Ministry of Health facilities are usually also the only formal healthcare providers in poorer districts (Appendix Figure 1.A.2). For psychologists, the salaries across the different institutions are similar. Instead, the main difference between the institutions is that vacancies at the Ministry of Health tend to be located in more remote areas. In our sample, 58% of private sector jobs are located in the two richest district categories, whereas only 24% of public sector jobs are located in the same areas.

Multiple studies have shown that health workers overall have strong preferences to live in cities (Miranda et al., 2012). Incorrect perceptions about life in rural areas are also likely to contribute to location decisions since health workers tend to come from better socio-economic backgrounds. Relative to the general population, psychologists in our sample come from more urban areas (78% vs 70%), have more educated parents (67% vs 47% have mothers that completed secondary school), and are less likely to speak an indigenous language as their mother tongue (7% vs 21%). Those who still decide to work in poor areas tend to be more prosocial. We find that psychologists who work in the poorest districts score 0.19 standard deviations ( $p = 0.003$ ) higher on a prosociality index than other psychologists.

### 1.2.2 Conceptual Framework

We consider a basic framework of labor market matches to understand how SERUMS assignments to poor areas can affect later labor market outcomes. We start with considering the decision of a worker to accept a public sector job in a rural area or a private sector job in

an urban area. The worker will accept the public sector job if the utility he gets from the job is higher than the utility from the private sector job ( $u_{public}(\beta\tilde{X}, \alpha\tilde{Y}) > u_{private}$ ). We assume that the utility function consists of two main components, the private component ( $\beta\tilde{X}$ ) and the prosocial component ( $\alpha\tilde{Y}$ ), where  $\beta$  represents the worker’s preferences,  $\tilde{X}$  stands for the worker’s expectation over the job attributes,  $\alpha$  represents altruism, and  $\tilde{Y}$  stands for the worker’s expectation over the change in community welfare if he accepts the job.

Inspired by previous work that shows how contact with disadvantaged communities can increase the prosocial behavior of individuals (Rao, 2019), we hypothesize that assignments to poor areas can increase a worker’s willingness to work for the public sector in remote areas by making them more prosocial. This could either occur because they start to care more about helping others ( $\alpha$ ) or because they update their beliefs about how much community welfare would change if they accept public sector jobs ( $\tilde{Y}$ ). These beliefs consist of several factors, including the perceived effectiveness of the public sector, the perceived size of the beneficiary pool, and the perceived causes that underlie poverty.

Assignments to poor areas could further affect other preferences ( $\beta$ ), e.g. how much the worker cares about salaries, or beliefs about other job characteristics ( $\tilde{X}$ ), e.g. the perceived difference in available amenities between rural and urban areas. More broadly, labor market outcomes are based on two-sided matches and assignments to poor areas could also affect labor demand. Positions in poorer areas might provide worse training, e.g. due to worse equipment or lack of medicines, which could make workers who went to poor areas less attractive for the private sector. Similarly, workers in poorer areas might have fewer opportunities to develop their professional networks since there might be fewer or less qualified peers in poor areas. This could make it harder to find a job in the private sector after the initial assignment.

We will use this basic framework to guide our analysis and distinguish between different mechanisms. Overall, the relationship between the poverty level of the SERUMS location and

subsequent public sector employment is ambiguous ex ante. For example, while contact with disadvantaged communities might make the worker more altruistic, a lack of infrastructure in remote facilities could also make the worker disillusioned about the effectiveness of public sector work.

## 1.3 Study Design

Our primary data source is an online survey that we conducted in 2022. Our main sample consists of 1,106 psychologists who have completed a paid SERUMS assignment at a Ministry of Health facility between November 2019 and May 2022.<sup>10</sup> For this sample, we obtained the name, email address, phone number, university, SERUMS placement, and SERUMS lottery details from the Ministry of Health.

Figure 1.2 documents the data collection process. All respondents received an initial email with the link to the online survey on August 17th, 2022. They were informed that the survey is conducted by an independent research team to analyze potential improvements to the SERUMS program. Respondents were offered an incentive payment of 19 soles ( $\approx$  10 USD PPP). After the initial email, we used a call center to make follow-up calls and remind respondents to fill out the survey. Respondents could also request to be sent a WhatsApp message with the survey link. In addition to the follow-up calls, we sent two reminder emails over the course of 1.5 weeks. In the final stage of the data collection, we conducted an abbreviated phone survey with respondents who had not filled out the online survey by that time. We further sent the online surveys to respondents who performed the SERUMS program at the time of the survey to understand at what point the effects emerge. These respondents received a shorter survey and were not offered an incentive payment.

---

<sup>10</sup>We also sent the survey to nutritionists, the second-largest profession that chooses SERUMS facilities in a randomized order. However, we do not find that their lottery rank affects the poverty level of the SERUMS location (Appendix Table 1.A.2). This is partly driven by less variation in the poverty status of available vacancies. We thus exclude them from our analysis.

63% of the main sample either completed the online or the phone survey, which is high compared to similar studies (Dobbie and Fryer Jr, 2015; Mo and Conn, 2018). When we restrict the sample to online survey responses, our completion rate is 53%. Another 4.2% of respondents started but did not complete the online survey. In our preferred specification, we always pool phone survey responses, partial online survey responses, and completed online survey responses. We show in the appendix that the results also hold when we restrict the sample to respondents who completed the online survey.

Unfortunately, the administrative data do not contain information on lottery rank. Our main measure of an applicant’s position in the lottery thus comes from survey data. However, we managed to obtain Zoom video recordings of 18 lotteries that cover a total of 349 applicants and were done between 2020 and 2022. We use this subsample to conduct internal validity checks. Columns 1-2 in Table 1.1 show that an applicant’s rank is not correlated with gender or type of university. We also do not observe that acceptance rates vary significantly based on lottery ranks (columns 3-6). Appendix Table 1.A.3 further shows that we do not observe differential survey attrition ranks based on an applicant’s lottery rank. Appendix Table 2.9 further reports a balance check among survey respondents. Psychologists that chose later do not differ along time-invariant characteristics like university status, age, gender, and birthplace. We also cannot reject a joint test of the hypotheses that each of these differences is zero ( $p = 0.52$ ). We further included a retrospective question on their career plans before doing the SERUMS program and do not find any significant differences. We next describe our survey components.

*Lottery Rank:* Our main independent variable is based on the self-reported rank during the lottery assignment. We ask respondents about the number of positions that were available, their position in the lottery, and the number of places left when it was their turn to choose. If the answers were internally inconsistent or the reported total number of positions was inconsistent with administrative data, we tried to call back the respondents to clarify

their answers. We use the actual rank from the zoom recordings whenever possible. Reassuringly, we find that the self-reported and observed lottery rank for the video recordings subsample is highly correlated (the Pearson correlation coefficient is equal to 0.77). Overall, 76% of respondents say that they are certain or very certain about their answers regarding the lottery assignment.

*Employment:* Our employment module asked whether the respondent worked for pay in the past week and, if they worked, in which sector, institution, and district their current job is located. The survey also asked about secondary jobs at other institutions as well as the respondent’s monthly salary for each job (based on six salary categories).

*Measures of Prosociality:* We collect measures of prosociality across three dimensions. First, we ask respondents about how important it is for them to have a job that provides opportunities to help the poor on a scale from 1 to 7. Second, we ask respondents whether they think that the government should increase aid to the poor. Third, we collect four different measures of donation behavior. We ask respondents whether they donated time or money to an organization that supports those in need in the past 30 days. We also adopt a measure from the Global Preference Survey that asks how much the respondents would donate to a good cause if they were to receive 300 soles ( $\approx 158$  USD PPP) unexpectedly (Falk et al., 2018). Previous research has shown that this is correlated with incentivized decisions in a dictator game (Falk et al., 2016). In addition, we conducted our own dictator game where we asked respondents whether they want to donate part of the incentive payment (19 soles) to two local nonprofit organizations. Respondents could decide to donate to one or both organizations. We selected two unknown NGOs to avoid concerns that the location of the SERUMS assignment could affect the respondent’s knowledge about larger nonprofit organizations. The first organization provides vitamin A supplements to children in rural areas in Peru. The script informed respondents that 12% of Peruvian children suffered from vitamin A deficiency and provided estimates of the cost-effectiveness of vitamin A

supplementation based on calculations of GiveWell, a nonprofit organization that assesses charities. We provided this information to hold the knowledge of participants fixed and focus on how much they care about helping the beneficiaries of the organization. The second nonprofit organization organizes arts and cultural programs in urban Peru and was intended to differentiate between changes in general generosity and generosity towards the rural poor.

*Job preferences and beliefs about job attributes:* We asked respondents to rate six additional job attributes from 1 to 7 according to their importance. These measures were salary, work-life balance, intellectual satisfaction, compatibility with a partner, work environment, and local infrastructure. We further asked respondents to compare positions at the Ministry of Health and in the private sector according to their salary, their opportunities to help the poor, and their work-life balance on a 5-point scale. Similarly, we also ask them to compare jobs at the Ministry of Health in the poorest and richest districts according to their local infrastructure, their opportunities to help the poor, their safety levels, and their compatibility with a partner. We further collected the respondent’s opinions on whether they agree that the SERUMS experience increased their willingness to work for the Ministry of Health and to work in rural areas.

*Skills and network formation:* We obtained information on how much the respondents believe that the SERUMS experience improved their theoretical knowledge, clinical skills, language abilities, and professional networks. We ask whether these changes in skills directly helped the respondents to get their current jobs. For networks, we included the following question: “did any member of your professional network that you met through SERUMS help you find or get hired at your current job?”. We also obtained information on the perceived availability of jobs by asking whether respondents agree that they could get a job at the Ministry of Health, at ESSALUD, and in the private sector if they wanted.

## 1.4 Results

Our empirical strategy exploits that psychologists choose facilities for the SERUMS program in randomized order, creating exogenous variation in choice sets. We restrict our sample to Ministry of Health lotteries with at least six facilities to ensure sufficient variation in the percentile rank. We exclude respondents that were classified as special cases and were allowed to choose first. We estimate the following specification for psychologist  $i$ :

$$y_i = \alpha + \gamma f(\textit{PercentileRank}_i) + \delta_i + X_i + \epsilon_i. \quad (1.1)$$

$f(\textit{PercentileRank}_i)$  is a flexible function of the respondent's rank. We show results based on tercile dummies as well as based on a linear specification of the respondent's rank. Our preferred specification is the tercile dummies since they do not assume a linear relationship between the respondent's rank and the outcome of interest. The tercile dummies are also less likely to be affected by measurement error since respondents might not recall their exact position in the lottery but remember whether they chose positions at the beginning or at the end of the assignment process.  $y_i$  is the outcome for respondent  $i$ ,  $\delta_i$  are lottery fixed effects, and  $X_i$  is a vector of controls that consists of age, gender, a dummy for whether the respondent's university was public, a dummy for whether Spanish is the respondent's native language, birth region fixed effects, and a dummy variable that indicates whether the survey was done over the phone. We always report robust standard errors.

Table 1.3 shows that respondents who chose later during the lottery process went to poorer places during the SERUMS program (Panel A). Relative to psychologists whose rank was in the top tercile, psychologists in the bottom tercile go to districts that rank 6.7 percentile points ( $p = 0.008$ ) higher in the government poverty index (Panel B). The share of the population that is rural in the district also increases by 6.4 percentage points ( $p = 0.013$ ).

For the rest of the paper, we assume that less desirable positions are equivalent to poorer and more remote locations.<sup>11</sup>

We next examine the effects on subsequent career outcomes in Table 1.4. We first show that the lottery rank does not predict whether a respondent worked for pay in the past week.<sup>12</sup> We then investigate changes in the type of employment. Respondents in the bottom rank tercile are 8.8 percentage points ( $p = 0.087$ ) more likely to work for the public sector (column 2). These effects are driven by an increase of 12.2 percentage points ( $p = 0.016$ ) in Ministry of Health employment (column 3), the institution that provides healthcare services to the poor. Columns 4-5 show that the gains in public sector employment come at the expense of private sector employment (including private sector work at secondary jobs). We find positive but insignificant effects on whether the jobs continue to be in the health sector (column 6).

After documenting differences in the type of employer, we also examine changes in location choices. We first show that psychologists rarely stay in the same district in which they did their SERUMS assignment and that this also does not differ by lottery ranks (column 1, Table 1.5), ruling out inertia as a potential driver for results. We then use the government’s classification of districts into poverty quintiles to characterize location choices. Column 6 shows that the share of respondents that work in the poorest districts is 8.7 percentage points ( $p = 0.021$ ), or 86%, higher for psychologists in the bottom rank tercile. This increase seems to come in part from psychologists who would otherwise work in medium or poor (quintile 2

---

<sup>11</sup>In Appendix Table 1.A.5, we regress a respondent’s lottery rank on various SERUMS position characteristics. We find that the poverty index is the main characteristic that is significantly correlated with the percentile ranking. Psychologists who chose later also tend to go to smaller facilities. We find no significant correlation between a respondent’s rank and net salaries or whether the district is located in the same department (aka state) as the respondent’s birthplace. While these are the official position characteristics that the government provides as part of the lottery process, we cannot rule out that positions that were chosen later might also differ along other dimensions.

<sup>12</sup>23% of respondents report that did not have a paid job at the time of the survey. While some of the respondents were full-time students or took care of family members, 73% of unemployed respondents said that they are currently looking for a job. The high unemployment rate is in part driven by psychologists who had just completed SERUMS 3 months prior to the survey since they just entered the job market recently. For this subsample, only 63% of respondents worked for pay in the past week.

and 3) districts. As a benchmark, Miranda et al. (2012) finds in a discrete choice experiment with physicians in Peru that introducing an allowance that would increase the salary for rural jobs by 50% would increase the share of physicians willing to work in rural areas by 86%. The similarity in the relative magnitude of the effects emphasizes the policy importance of the change in employment outcomes due to differences in SERUMS assignments. Column 7 shows that we do not observe changes in the monthly income of respondents. As described in Section 2, psychologists seem to face instead a decision between public sector jobs in remote areas (with fewer amenities) and private sector jobs in urban areas.

Taken together, these results document a substantial change in the employment status of respondents based on their rank in the SERUMS lottery. Respondents in the bottom tercile rank are more likely to work for the public sector and in poorer areas. What is driving these results? We use our conceptual framework to guide our analysis and differentiate between changes in preferences, beliefs, skill development, and network formation.

We examine whether changes in prosociality can explain the results in Table 1.6. Our preferred outcome is a prosociality index that captures job preferences, redistributive preferences, and donating behavior. Relative to psychologists in the top lottery tercile, respondents in the bottom lottery tercile score 0.17 standard deviations ( $p = 0.004$ ) in the prosociality index (column 1). We also observe significant increases in each of the index components. Respondents in the bottom tercile are 9.9 percentage points ( $p = 0.079$ ) more likely to say that opportunities to help the poor are very important for choosing a job, 8.1 percentage points ( $p = 0.070$ ) more likely to say that the government should increase to the poor, and score 0.09 standard deviations ( $p = 0.164$ ) higher in a donation subindex (columns 2-4). Appendix Table 1.A.6 shows the results for each component of the subindex. A potential concern is that the relationship between public sector work and prosociality is reversed and that it is public sector work that makes individuals more prosocial instead of the other way around. To rule this out and understand the timing of the effects, we replicate the analysis

for respondents who either did SERUMS at the time of the survey or completed the program three months prior (columns 5-8). We find similar effects for this subgroup which suggests that it is the increase in prosociality that leads to an increased willingness to work for the public sector.

We interpret the changes in the prosociality index as an increase in the overall prosociality component ( $\alpha\tilde{Y}$ ). Separating between a change in altruism and beliefs is difficult since e.g. opinions on government aid are affected by how much respondents care about helping and by how much they think government aid will help the beneficiaries. The higher share of respondents in the bottom tercile who answer that opportunities to help the poor are very important for choosing a job already point to a change in preferences. As an additional test, we also examine the amount that was donated to the Vitamin A NGO. We provided detailed information on the cost-effectiveness of the intervention as well as the share of children with vitamin A deficiency in Peru to keep beliefs fixed. The comparison between donations to the Vitamin A NGO and the urban culture NGO was further intended to differentiate changes in general prosociality and prosociality towards the rural poor. Appendix Table 1.A.7 shows that we find noisy and insignificant effects for both outcomes.. However, we caution against the interpretation of the results, since it might be the case that more prosocial respondents already donate to other organizations or that increased public sector work in poor areas crowds out alternative ways of supporting the poor. Consistent with this, we find suggestive evidence that respondents in the bottom rank tercile tend to be more likely to say that they did not donate in the dictator game because they are already supporting other organizations or because they are already helping the poor through their current work (Appendix Table 1.A.7, columns 5-6).

We also collected direct information on beliefs. We consider four main beliefs that could limit interest in public sector work: (i) respondents might think that people are poor because of a lack of individual effort so any contributions might have limited effect, (ii) respondents

might underestimate the share of patients who benefit from the public sector, (iii) respondents might have limited trust in the public sector and (iv) respondents might perceive the private sector provides more opportunities to help the poor than the public sector. Column 1 in Table 1.A.8 examines whether the locations of the SERUMS assignments change what respondents perceive to be the underlying reasons for poverty. Across both specifications, we find positive (but noisy and insignificant) effects for whether respondents who chose later are more likely to think that circumstances beyond a person’s control, instead of lack of individual effort, are the main reason for poverty. We do not find differences in government trust (column 2) and even slightly negative effects when looking at the perceived share of the population that is covered by the government’s social protection scheme that provides healthcare to poor households and informal workers. The last result stands in contrast to Londoño-Vélez (2022) who finds that contact with poor students in the classrooms affects perceptions about the income distribution of a country. A potential explanation is that respondents already seem to have accurate beliefs about the government’s social protection scheme: the outcome mean of the omitted group (63%) is very close to the actual share of the population that is covered by the scheme (66%).

However, we do find positive effects when considering perceptions about the relative effectiveness of the public sector. In this exercise, we asked respondents to compare job attributes in the public and private sectors based on a 5-point scale, where 5 corresponds to better attributes in the public sector. Column 4 shows that respondents who chose later in the lottery rate tend to have a more favorable view of the public sector in terms of the opportunities it provides to help the poor. This pattern also remains when we restrict the sample to respondents who are either currently doing the SERUMS program or who completed the program just three months before our survey. We do not observe similar effects when we asked respondents to compare jobs in the public and private sectors along other attributes (Appendix Table 1.A.9). An exception is that respondents in the bottom

tercile also seem to have a more favorable view of the public sector in terms of salaries. However, these differences are smaller and noisier when restricting the sample to current and recent SERUMS participants. In a similar exercise, we also asked respondents to compare jobs at the Ministry of Health in the poorest and richest districts. Appendix Table 1.A.10 documents that we do not observe differences along this dimension. Appendix Table 1.A.11 further shows that respondents have similar answers when asked about other job preferences.

Finally, we consider whether changes in the hireability of respondents due to differential changes in network or skill formation can explain the results. The concern is that physicians who completed the SERUMS program in poorer areas had fewer opportunities to develop their professional skills. This could make them less attractive to the private sector, so changes in demand and not changes in beliefs or preferences could drive the increase in public sector employment. Similarly, physicians that went to poor areas might be exposed to fewer or less-qualified peers which could make it harder to find a job in the private sector after SERUMS.

We start by showing additional evidence for a change in the supply side. We directly asked respondents whether they would agree that the SERUMS experience increased their willingness to work for the Ministry of Health or to work in the poorest districts in the country. Table 1.7 shows that psychologists in the bottom lottery tercile are 10.5 percentage points ( $p = 0.035$ ) more likely to strongly agree that SERUMS made them more willing to work for the Ministry of Health. When asked about why their willingness to work for the Ministry of Health increased, 54% of respondents say that SERUMS made it more important for them to have a job where they can help the poorest. We only asked the follow-up question in the phone survey but we consider it as additional evidence that an increase in prosociality is driving the results.

We next examine whether lottery rank affects the perceptions of how the rural service program improved three sets of skills: theoretical knowledge, clinical skills, and indigenous

language skills. Columns 1-3 in Table 1.8 show that we do not find any evidence for differential skill gains based on the lottery rank. Respondents in the bottom lottery tercile are also not less likely to say that the improvements in these attributes helped them to find their current job. We then also asked similar questions about network formation. Columns 4 and 8 in Table 1.8 report that we also do not find changes in outcomes based on a respondent’s lottery rank along this margin. We also note that only 20% of all psychologists say that someone they met through SERUMS helped them to find or get hired at their current job. This emphasizes that a rural service program like SERUMS differs from residency programs for physicians that have been examined in previous work (Fadlon et al., 2022) and play a much more pivotal role in the network formation of participants. Finally, we also asked respondents whether they could get a job at the Ministry of Health, at ESSALUD, and at private facilities if they wanted. Table 1.9 shows that the lottery rank does not affect the perceived availability of job options.<sup>13</sup> These results are also consistent with qualitative interviews with former SERUMS participants who told us that employers care about whether an applicant participated in SERUMS but not where the assignment was completed.

Overall, these results suggest that the career effects are driven by changes in the pro-sociality of respondents. Qualitative evidence adds to how the SERUMS experiences affected the perspective of psychologists. One of our respondents told us that through SERUMS she “learned that it is not only enough to put all your effort to be able to study ([she] also worked and studied since [she] was a child) but the distance from schools, the lack of public services in their villages, the type of upbringing, the training received and the example to follow greatly influence many to reject studying and/or progressing”. Similarly, another respondent mentioned that the SERUMS program should be expanded so more people are “able to live that great experience in order to have a better identification with the other realities

---

<sup>13</sup>This further helps us to address the potential concern that the career effects are caused by the additional bonus points that psychologists in poorer SERUMS locations receive. The average psychologists in the bottom lottery tercile only receive 0.65 additional bonus percentage points relative to psychologists in the top lottery tercile, which makes it unlikely that the differences in bonus points can explain the career effects.

of our Peru.” These quotes emphasize how the SERUMS assignments were a significant experience in respondents’ lives and led them to change their perspectives. Since healthcare professionals come from better socio-economic backgrounds (as discussed in Section 2), many of them come in contact with rural poverty through SERUMS for the first time. Consistent with this explanation, we also find suggestive evidence that the effects on our main outcomes are concentrated in respondents with less contact with rural poverty prior to the program, measured as being born in Lima (Appendix Table 1.A.12).<sup>14</sup>

We conduct various robustness checks for our main results (Appendix Tables 1.A.5-1.A.8). Our findings hold if (i) we exclude controls, (ii) restrict the sample to respondents who at least partially completed the online survey, (iii) restrict the sample to respondents who completed the online survey, (iv) and exclude respondents that are uncertain or very uncertain about their lottery rank. The only exception is the effect on overall public sector employment, for which the size of the coefficients remains similar but the estimates become insignificant if we restrict the sample size. Another concern is that the Covid-19 pandemic affects the external validity of the findings. We show that we find similar but noisier results if we either restrict the sample to respondents who completed SERUMS before the pandemic but entered the labor market when the first wave of cases in Peru emerged or restrict the sample to respondents who completed SERUMS during the pandemic but entered the labor market when vaccines were already widely available.<sup>15</sup> Finally, we also report p-values based on randomization inference by reassigning the lottery rank to each respondent 2,000 times (Appendix Table 2.A.10). The adjusted p-value on the coefficient for the bottom rank tercile increases to 0.162 for overall public sector employment, but remains below 0.06 for the other main outcomes.

---

<sup>14</sup>Due to small sample size, we pool the medium and bottom rank tercile in this specification.

<sup>15</sup>An exception is the likelihood to work in the poorest districts for respondents who entered the labor market during the pandemic. However, the results are very noisy and we cannot rule out that the effect size is the same.

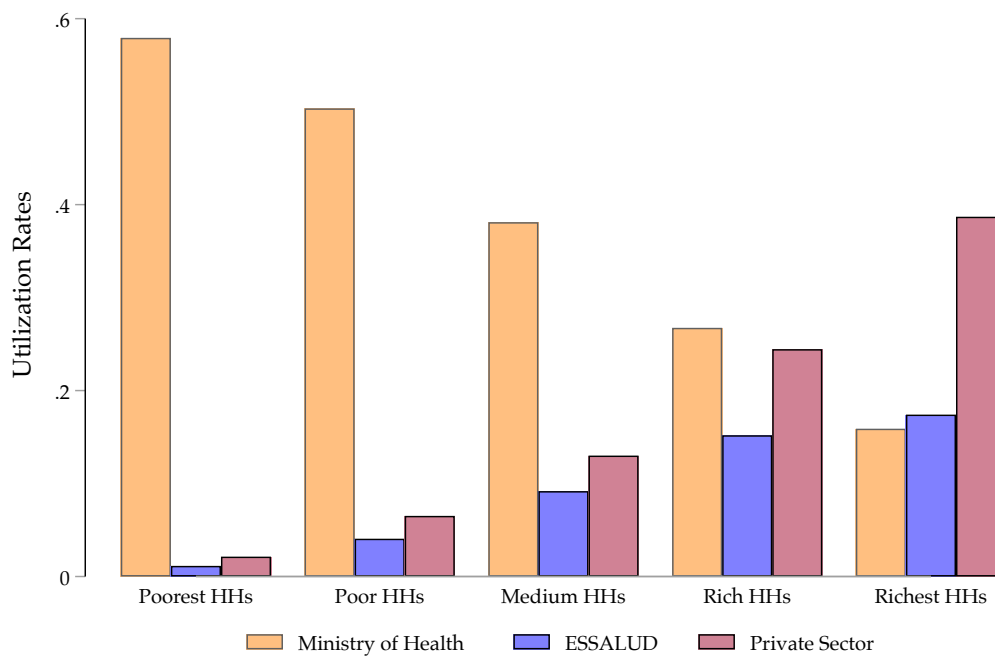
## 1.5 Conclusion

This paper uses exogenous variation in choice sets within a mandatory rural service program in Peru to demonstrate that temporary work assignments to poor areas can increase prosociality and affect career decisions. It adds to a nascent literature that shows how contact with disadvantaged communities in a variety of settings can increase prosocial behavior (Rao, 2019). Our findings document that these changes in prosociality can affect influence important policy outcomes like public sector employment and emphasize the importance of taking into account the endogeneity of preferences and beliefs in economic theory and the design of public policies.

The insights speak to the optimal design of recruitment and posting policies for public sector workers. Compensation schemes should take into account that preferences and beliefs might change because of the assignment itself, so it could be efficient to e.g. offer high lump-sum payments upfront to encourage workers to take up work in poor areas for at least some time. The findings further provide justification for a variety of social service programs, including mandatory rural service programs like SERUMS, voluntary schemes like ‘Teach for America’, and corporate efforts to encourage social work among employees. Taking into account the endogeneity of preferences and beliefs could further resolve a potential trade-off between hiring intrinsically and extrinsically motivated candidates (Ashraf et al., 2020), as the work itself might affect a worker’s intrinsic motivation.

## 1.6 Tables and Figures

Figure 1.1: Healthcare Utilization by Household Wealth



Notes: The data consists of the Peru Demographic Health Surveys from 2003 until 2017. The x-axis shows DHS wealth quintiles. The y-axis shows healthcare utilization rates for under 5-year old children who suffered from fever in the past two weeks. Omitted outcomes include visits to private pharmacies, healthcare from friends and family, and no treatment at all.

Figure 1.2: Timeline of Data Collection

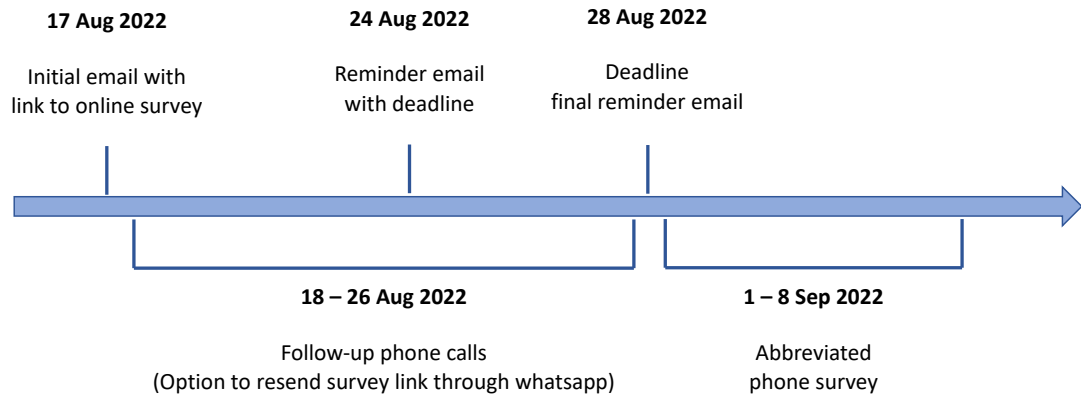


Table 1.1: Internal Validity Checks

|                                          | Balance           |                             | Assignment Decision |                   |                   |                   |
|------------------------------------------|-------------------|-----------------------------|---------------------|-------------------|-------------------|-------------------|
|                                          | Female<br>(1)     | Public<br>University<br>(2) | Accepted<br>(3)     | Rejected<br>(4)   | Mistake<br>(5)    | Absent<br>(6)     |
| <i>Panel A: Linear Specification</i>     |                   |                             |                     |                   |                   |                   |
| Percentile Rank                          | -0.072<br>(0.078) | -0.025<br>(0.045)           | -0.069<br>(0.078)   | 0.030<br>(0.040)  | 0.022<br>(0.041)  | 0.018<br>(0.063)  |
| <i>Panel B: Discrete Specification</i>   |                   |                             |                     |                   |                   |                   |
| Medium Rank Tercile                      | -0.011<br>(0.050) | 0.037<br>(0.035)            | 0.017<br>(0.050)    | -0.007<br>(0.023) | 0.046*<br>(0.026) | -0.056<br>(0.039) |
| Bottom Rank Tercile                      | -0.059<br>(0.052) | -0.007<br>(0.031)           | -0.044<br>(0.053)   | 0.017<br>(0.027)  | 0.029<br>(0.026)  | -0.002<br>(0.043) |
| pval: Medium Tercile<br>= Bottom Tercile | 0.367             | 0.201                       | 0.226               | 0.330             | 0.576             | 0.148             |
| Outcome Mean                             | 0.853             | 0.059                       | 0.816               | 0.039             | 0.019             | 0.126             |
| Observations                             | 325               | 325                         | 330                 | 330               | 330               | 330               |

Notes: Robust standard errors appear in parentheses. All regressions include lottery stratification fixed effects.

Table 1.2: Effect of Lottery Rank on SERUMS Location

|                                          | Government<br>Poverty index<br>(1) | Rural Share<br>(2) |
|------------------------------------------|------------------------------------|--------------------|
| <i>Panel A: Linear Specification</i>     |                                    |                    |
| Percentile Rank                          | 0.094***<br>(0.035)                | 0.089**<br>(0.036) |
| <i>Panel B: Discrete Specification</i>   |                                    |                    |
| Medium Rank Tercile                      | 0.035<br>(0.026)                   | 0.037<br>(0.026)   |
| Bottom Rank Tercile                      | 0.067***<br>(0.025)                | 0.064**<br>(0.026) |
| pval: Medium Tercile<br>= Bottom Tercile | 0.253                              | 0.333              |
| Outcome Mean                             | 0.549                              | 0.563              |
| Included in Phone Survey                 | Yes                                | Yes                |
| Observations                             | 557                                | 557                |

Notes: Robust standard errors appear in parentheses. All regressions include lottery stratification fixed effects and a vector of controls that consists of age, gender, a dummy for whether the respondent's university was public, a dummy for whether Spanish is the respondent's native language, birth region fixed effects, and a dummy variable that indicates whether the survey was done over the phone.

Table 1.3: Effect of Lottery Rank on SERUMS Location

|                                          | Government<br>Poverty index<br>(1) | Rural Share<br>(2) |
|------------------------------------------|------------------------------------|--------------------|
| <i>Panel A: Linear Specification</i>     |                                    |                    |
| Percentile Rank                          | 0.094***<br>(0.035)                | 0.089**<br>(0.036) |
| <i>Panel B: Discrete Specification</i>   |                                    |                    |
| Medium Rank Tercile                      | 0.035<br>(0.026)                   | 0.037<br>(0.026)   |
| Bottom Rank Tercile                      | 0.067***<br>(0.025)                | 0.064**<br>(0.026) |
| pval: Medium Tercile<br>= Bottom Tercile | 0.253                              | 0.333              |
| Outcome Mean                             | 0.549                              | 0.563              |
| Included in Phone Survey                 | Yes                                | Yes                |
| Observations                             | 557                                | 557                |

Notes: Robust standard errors appear in parentheses. All regressions include lottery stratification fixed effects and a vector of controls that consists of age, gender, a dummy for whether the respondent's university was public, a dummy for whether Spanish is the respondent's native language, birth region fixed effects, and a dummy variable that indicates whether the survey was done over the phone.

Table 1.4: Effect of Lottery Rank on Employment Outcomes

|                                          | Any Work<br>for Pay<br>(1) | Works in<br>Public<br>Sector<br>(2) | Works for<br>Ministry of<br>Health<br>(3) | Works for<br>ESSALUD<br>(4) | Works in<br>Private<br>Sector<br>(5) | Works in<br>Health<br>Sector<br>(6) |
|------------------------------------------|----------------------------|-------------------------------------|-------------------------------------------|-----------------------------|--------------------------------------|-------------------------------------|
| <i>Panel A: Linear Specification</i>     |                            |                                     |                                           |                             |                                      |                                     |
| Percentile Rank                          | 0.028<br>(0.065)           | 0.119*<br>(0.071)                   | 0.187***<br>(0.071)                       | -0.008<br>(0.014)           | -0.134**<br>(0.062)                  | 0.086<br>(0.076)                    |
| <i>Panel B: Discrete Specification</i>   |                            |                                     |                                           |                             |                                      |                                     |
| Medium Rank Tercile                      | -0.097**<br>(0.046)        | -0.032<br>(0.054)                   | 0.034<br>(0.052)                          | -0.012*<br>(0.007)          | -0.088*<br>(0.045)                   | -0.025<br>(0.054)                   |
| Bottom Rank Tercile                      | 0.018<br>(0.045)           | 0.088*<br>(0.052)                   | 0.122**<br>(0.051)                        | 0.002<br>(0.009)            | -0.091**<br>(0.046)                  | 0.067<br>(0.054)                    |
| pval: Medium Tercile<br>= Bottom Tercile | 0.022                      | 0.026                               | 0.107                                     | 0.133                       | 0.952                                | 0.117                               |
| Outcome Mean                             | 0.808                      | 0.579                               | 0.332                                     | 0.009                       | 0.271                                | 0.519                               |
| Observations                             | 559                        | 559                                 | 559                                       | 559                         | 559                                  | 552                                 |

Notes: Robust standard errors appear in parentheses. All regressions include lottery stratification fixed effects and a vector of controls that consists of age, gender, and a dummy for whether the respondent's university was public. Columns 1-6 were also asked as part of the phone survey and include further a dummy variable for whether the survey was done by phone.

Table 1.5: Effect of Lottery Rank on Additional Employment Outcomes

|                                          | Works in...                   |                      |                   |                     |                   |                      | Monthly<br>Salary      |
|------------------------------------------|-------------------------------|----------------------|-------------------|---------------------|-------------------|----------------------|------------------------|
|                                          | Same<br>District as<br>SERUMS | Richest<br>Districts | Rich<br>Districts | Medium<br>Districts | Poor<br>Districts | Poorest<br>Districts |                        |
|                                          | (1)                           | (2)                  | (3)               | (4)                 | (5)               | (6)                  |                        |
|                                          | (1)                           | (2)                  | (3)               | (4)                 | (5)               | (6)                  |                        |
| <i>Panel A: Linear Specification</i>     |                               |                      |                   |                     |                   |                      |                        |
| Percentile Rank                          | 0.053<br>(0.032)              | -0.006<br>(0.048)    | 0.012<br>(0.048)  | -0.031<br>(0.058)   | -0.049<br>(0.065) | 0.110**<br>(0.052)   | 190.424<br>(210.830)   |
| <i>Panel B: Discrete Specification</i>   |                               |                      |                   |                     |                   |                      |                        |
| Medium Rank Tercile                      | 0.018<br>(0.024)              | -0.044<br>(0.038)    | -0.024<br>(0.036) | -0.030<br>(0.038)   | -0.045<br>(0.046) | 0.041<br>(0.037)     | -275.440*<br>(162.203) |
| Bottom Rank Tercile                      | 0.032<br>(0.024)              | -0.036<br>(0.036)    | 0.004<br>(0.034)  | -0.014<br>(0.041)   | -0.018<br>(0.045) | 0.087**<br>(0.037)   | 167.275<br>(153.625)   |
| pval: Medium Tercile<br>= Bottom Tercile | 0.589                         | 0.843                | 0.446             | 0.692               | 0.546             | 0.282                | 0.006                  |
| Outcome Mean                             | 0.034                         | 0.150                | 0.145             | 0.159               | 0.246             | 0.101                | 1957.447               |
| Observations                             | 542                           | 542                  | 542               | 542                 | 542               | 542                  | 486                    |

Notes: Robust standard errors appear in parentheses. All regressions include lottery stratification fixed effects and a vector of controls that consists of age, gender, a dummy for whether the respondent's university was public, a dummy for whether Spanish is the respondent's native language, and birth region fixed effects. Columns 1-6 were also asked as part of the phone survey and include further a dummy variable for whether the survey was done by phone.

Table 1.6: Effect of Lottery Rank on Prosociality

|                                          | Main Sample           |                                                            |                                                     |                      | Current And Recent Participants |                                                            |                                                     |                      |
|------------------------------------------|-----------------------|------------------------------------------------------------|-----------------------------------------------------|----------------------|---------------------------------|------------------------------------------------------------|-----------------------------------------------------|----------------------|
|                                          | Prosociality<br>Index | Index Components                                           |                                                     |                      | Prosociality<br>Index           | Index Components                                           |                                                     |                      |
|                                          |                       | Opportunities<br>to Help the<br>Poor are Very<br>Important | Government<br>Should<br>Increase Aid<br>to the Poor | Donation<br>Subindex |                                 | Opportunities<br>to Help the<br>Poor are Very<br>Important | Government<br>Should<br>Increase Aid<br>to the Poor | Donation<br>Subindex |
|                                          | (1)                   | (2)                                                        | (3)                                                 | (4)                  | (5)                             | (6)                                                        | (7)                                                 | (8)                  |
| <i>Panel A: Linear Specification</i>     |                       |                                                            |                                                     |                      |                                 |                                                            |                                                     |                      |
| Percentile Rank                          | 0.233***<br>(0.085)   | 0.136*<br>(0.079)                                          | 0.129**<br>(0.064)                                  | 0.094<br>(0.095)     | 0.357***<br>(0.108)             | 0.170<br>(0.107)                                           | 0.275***<br>(0.090)                                 | 0.095<br>(0.129)     |
| <i>Panel B: Discrete Specification</i>   |                       |                                                            |                                                     |                      |                                 |                                                            |                                                     |                      |
| Medium Rank Tercile                      | 0.076<br>(0.064)      | 0.075<br>(0.057)                                           | -0.003<br>(0.050)                                   | 0.037<br>(0.068)     | -0.013<br>(0.082)               | -0.049<br>(0.075)                                          | 0.051<br>(0.068)                                    | -0.045<br>(0.095)    |
| Bottom Rank Tercile                      | 0.169***<br>(0.059)   | 0.099*<br>(0.056)                                          | 0.081*<br>(0.045)                                   | 0.094<br>(0.067)     | 0.240***<br>(0.077)             | 0.107<br>(0.077)                                           | 0.182***<br>(0.067)                                 | 0.083<br>(0.095)     |
| pval: Medium Tercile<br>= Bottom Tercile | 0.142                 | 0.680                                                      | 0.092                                               | 0.426                | 0.001                           | 0.043                                                      | 0.049                                               | 0.218                |
| Outcome Mean                             | -0.018                | 0.434                                                      | 0.756                                               | 0.020                | -0.004                          | 0.504                                                      | 0.700                                               | 0.017                |
| Included in Phone Survey                 | Yes                   | Yes                                                        | Yes                                                 | Yes                  | Yes                             | Yes                                                        | Yes                                                 | Yes                  |
| Observations                             | 555                   | 553                                                        | 549                                                 | 532                  | 317                             | 313                                                        | 317                                                 | 310                  |

Notes: Robust standard errors appear in parentheses. All regressions include lottery stratification fixed effects and a vector of controls that consists of age, gender, a dummy for whether the respondent's university was public, a dummy for whether Spanish is the respondent's native language, birth region fixed effects, and a dummy variable that indicates whether the survey was done over the phone. The sample in columns 1-4 consists of psychologists who completed SERUMS at the point of the survey. The sample in columns 5-8 consists of psychologists who were currently doing SERUMS or who completed the program three months prior.

Table 1.7: Additional Evidence for Supply-Side Channels

|                                          | SERUMS Increased Willingness to Work |                       |                    |                       |
|------------------------------------------|--------------------------------------|-----------------------|--------------------|-----------------------|
|                                          | for Ministry of Health               |                       | in Poor Areas      |                       |
|                                          | Agree<br>(1)                         | Strongly Agree<br>(2) | Agree<br>(3)       | Strongly Agree<br>(4) |
| <i>Panel A: Linear Specification</i>     |                                      |                       |                    |                       |
| Percentile Rank                          | 0.129**<br>(0.062)                   | 0.182***<br>(0.068)   | 0.059<br>(0.070)   | -0.038<br>(0.075)     |
| <i>Panel B: Discrete Specification</i>   |                                      |                       |                    |                       |
| Medium Rank Tercile                      | 0.156***<br>(0.042)                  | 0.120**<br>(0.052)    | 0.097**<br>(0.048) | 0.082<br>(0.058)      |
| Bottom Rank Tercile                      | 0.069<br>(0.046)                     | 0.105**<br>(0.049)    | 0.026<br>(0.051)   | -0.048<br>(0.052)     |
| pval: Medium Tercile<br>= Bottom Tercile | 0.050                                | 0.772                 | 0.159              | 0.026                 |
| Outcome Mean                             | 0.751                                | 0.215                 | 0.764              | 0.230                 |
| Included in Phone Survey                 | Yes                                  | Yes                   | No                 | No                    |
| Observations                             | 552                                  | 552                   | 483                | 483                   |

Notes: Robust standard errors appear in parentheses. All regressions include lottery stratification fixed effects and a vector of controls that consists of age, gender, a dummy for whether the respondent's university was public, a dummy for whether Spanish is the respondent's native language, and birth region fixed effects. Columns 1-2 were also asked as part of the phone survey and include further a dummy variable for whether the survey was done by phone.

Table 1.8: Effect of Lottery Rank on Skill and Network Formation

|                                          | Sample: Current Workers  |                           |                           |                   |                                             |                           |                           |                  |
|------------------------------------------|--------------------------|---------------------------|---------------------------|-------------------|---------------------------------------------|---------------------------|---------------------------|------------------|
|                                          | SERUMS Helped to Improve |                           |                           |                   | Improvements in ... Helped with Getting Job |                           |                           |                  |
|                                          | Knowledge<br>(1)         | Clinical<br>Skills<br>(2) | Language<br>Skills<br>(3) | Networks<br>(4)   | Knowledge<br>(5)                            | Clinical<br>Skills<br>(6) | Language<br>Skills<br>(7) | Networks<br>(8)  |
| <i>Panel A: Linear Specification</i>     |                          |                           |                           |                   |                                             |                           |                           |                  |
| Percentile Rank                          | -0.128<br>(0.171)        | -0.042<br>(0.142)         | -0.244<br>(0.205)         | -0.210<br>(0.195) | -0.031<br>(0.098)                           | 0.152<br>(0.095)          | 0.079<br>(0.099)          | 0.015<br>(0.084) |
| <i>Panel B: Discrete Specification</i>   |                          |                           |                           |                   |                                             |                           |                           |                  |
| Medium Rank Tercile                      | 0.219*<br>(0.126)        | 0.038<br>(0.105)          | 0.206<br>(0.156)          | 0.020<br>(0.129)  | 0.099<br>(0.072)                            | 0.191***<br>(0.069)       | 0.093<br>(0.070)          | 0.015<br>(0.067) |
| Bottom Rank Tercile                      | -0.014<br>(0.123)        | -0.022<br>(0.102)         | -0.125<br>(0.150)         | -0.102<br>(0.139) | -0.002<br>(0.069)                           | 0.102<br>(0.067)          | 0.029<br>(0.069)          | 0.033<br>(0.057) |
| pval: Medium Tercile<br>= Bottom Tercile | 0.072                    | 0.607                     | 0.058                     | 0.361             | 0.185                                       | 0.199                     | 0.394                     | 0.800            |
| Outcome Mean                             | 3.301                    | 3.549                     | 2.294                     | 3.157             | 0.644                                       | 0.676                     | 0.203                     | 0.201            |
| Included in Phone Survey                 | No                       | No                        | No                        | No                | No                                          | No                        | No                        | No               |
| Observations                             | 361                      | 360                       | 360                       | 360               | 354                                         | 350                       | 347                       | 353              |

Notes: Robust standard errors appear in parentheses. All regressions include lottery stratification fixed effects and a vector of controls that consists of age, gender, a dummy for whether the respondent's university was public, a dummy for whether Spanish is the respondent's native language, and birth region fixed effects. The sample is restricted to respondents who were currently working for pay at the point of the survey.

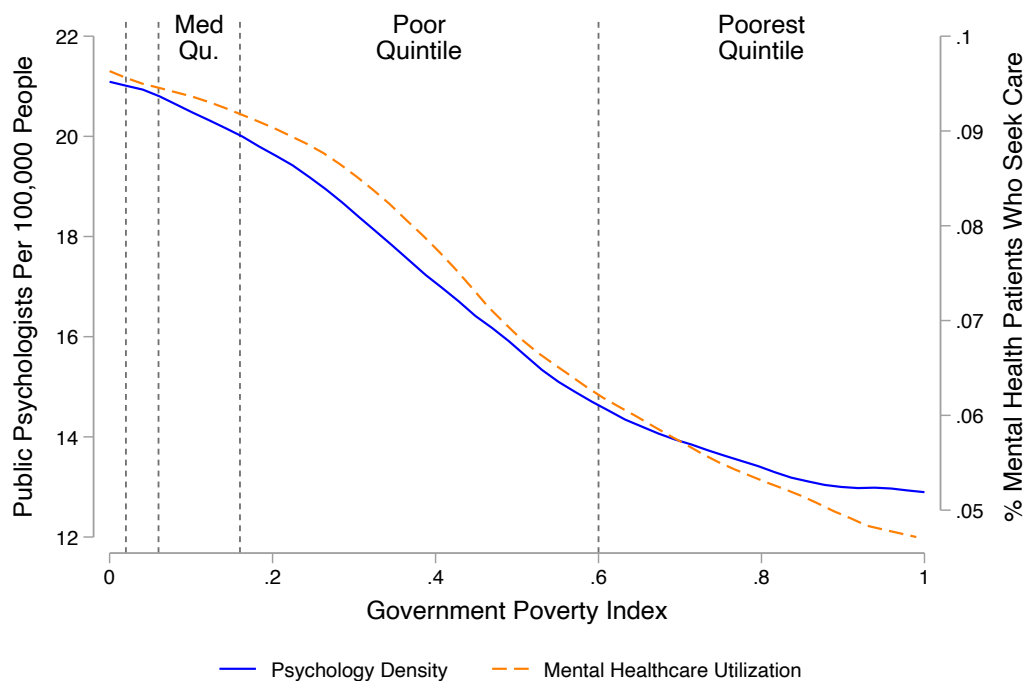
Table 1.9: Effect of Lottery Rank on Job Availability

|                                          | Respondent Could Get a Job |                          |                   |                          |                   |                          |
|------------------------------------------|----------------------------|--------------------------|-------------------|--------------------------|-------------------|--------------------------|
|                                          | at Ministry of Health      |                          | at ESSALUD        |                          | in Private Sector |                          |
|                                          | Agree<br>(1)               | Strongly<br>Agree<br>(2) | Agree<br>(3)      | Strongly<br>Agree<br>(4) | Agree<br>(5)      | Strongly<br>Agree<br>(6) |
| <i>Panel A: Linear Specification</i>     |                            |                          |                   |                          |                   |                          |
| Percentile Rank                          | -0.035<br>(0.082)          | 0.043<br>(0.070)         | -0.052<br>(0.084) | 0.015<br>(0.058)         | -0.011<br>(0.073) | 0.013<br>(0.060)         |
| <i>Panel B: Discrete Specification</i>   |                            |                          |                   |                          |                   |                          |
| Medium Rank Tercile                      | 0.017<br>(0.058)           | 0.039<br>(0.049)         | 0.012<br>(0.062)  | 0.010<br>(0.041)         | -0.021<br>(0.055) | 0.009<br>(0.043)         |
| Bottom Rank Tercile                      | 0.015<br>(0.058)           | 0.048<br>(0.052)         | 0.003<br>(0.060)  | 0.010<br>(0.041)         | -0.001<br>(0.051) | 0.005<br>(0.043)         |
| pval: Medium Tercile<br>= Bottom Tercile | 0.971                      | 0.863                    | 0.886             | 0.995                    | 0.717             | 0.922                    |
| Outcome Mean                             | 0.681                      | 0.186                    | 0.511             | 0.128                    | 0.777             | 0.160                    |
| Included in Phone Survey                 | No                         | No                       | No                | No                       | No                | No                       |
| Observations                             | 478                        | 478                      | 478               | 478                      | 478               | 478                      |

Notes: Robust standard errors appear in parentheses. All regressions include lottery stratification fixed effects and a vector of controls that consists of age, gender, a dummy for whether the respondent's university was public, a dummy for whether Spanish is the respondent's native language, and birth region fixed effects.

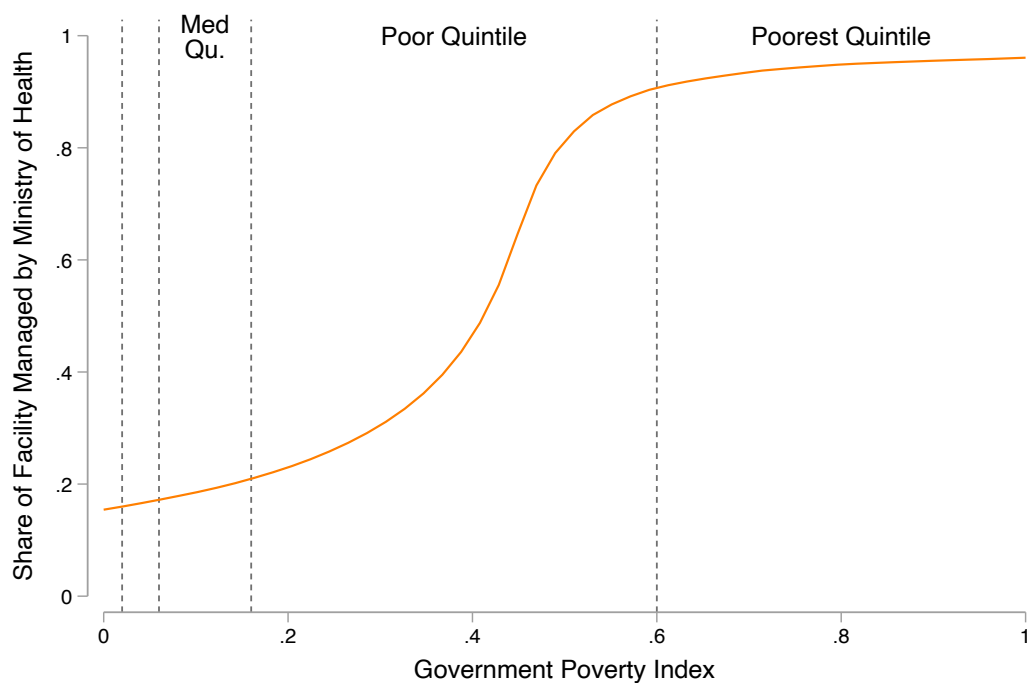
## 1.7 Appendix

Figure 1.A.1: Psychologist Density and Mental Healthcare Utilization by District Poverty



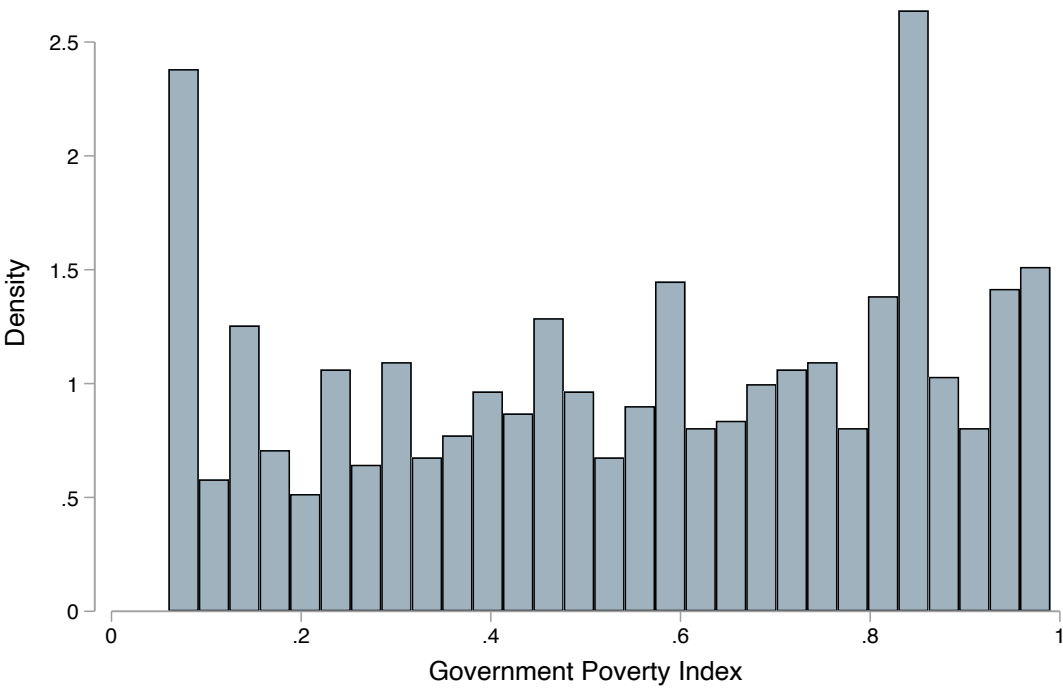
Notes: The blue line shows psychology density and corresponds to the left y-axis. The orange line shows mental healthcare utilization and corresponds to the right y-axis. Psychology density is based on administrative data from the Ministry of Health and mental healthcare utilization is based on the National Household Survey (ENAH). The x-axis shows the district-level government poverty index from FONCODES. The vertical lines indicate the poverty quintile cutoffs from FONCODES.

Figure 1.A.2: Share of Health Facilities Managed by Ministry of Health by District Poverty



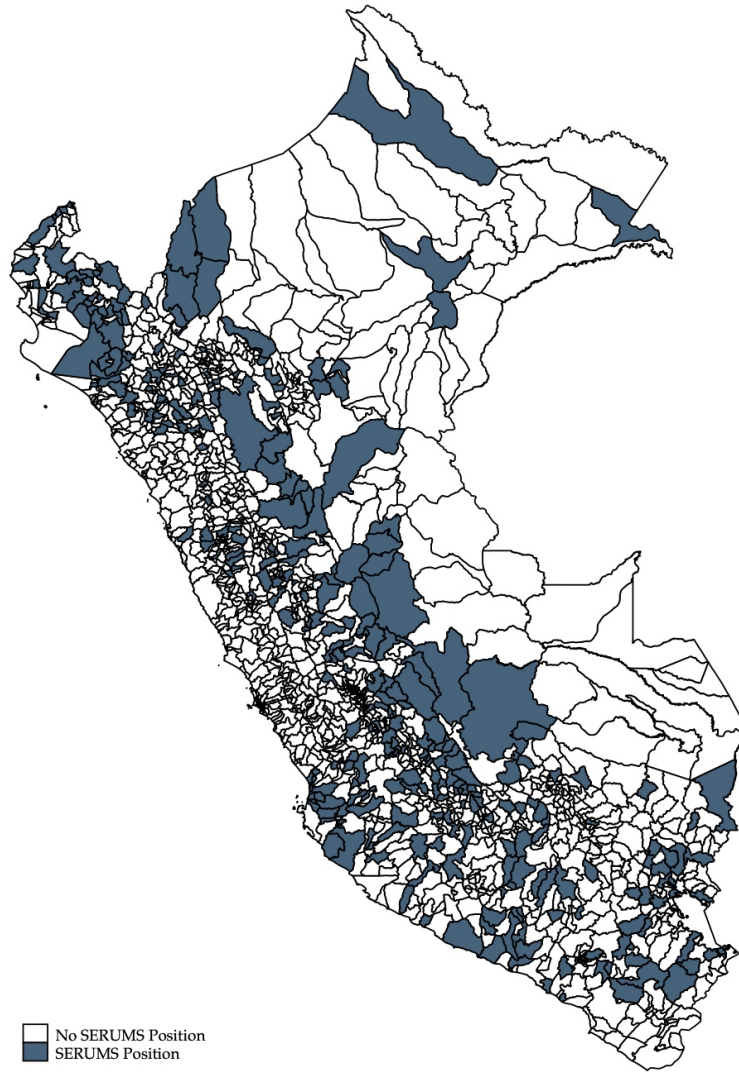
Notes: The y-axis shows the district-level share of facilities managed by the Ministry of Health according to the National Registry of Health Establishments. The x-axis shows the district-level government poverty index from FONCODES. The vertical lines indicate the poverty quintile cutoffs from FONCODES.

Figure 1.A.3: Distribution of Government Poverty Index Across SERUMS Positions



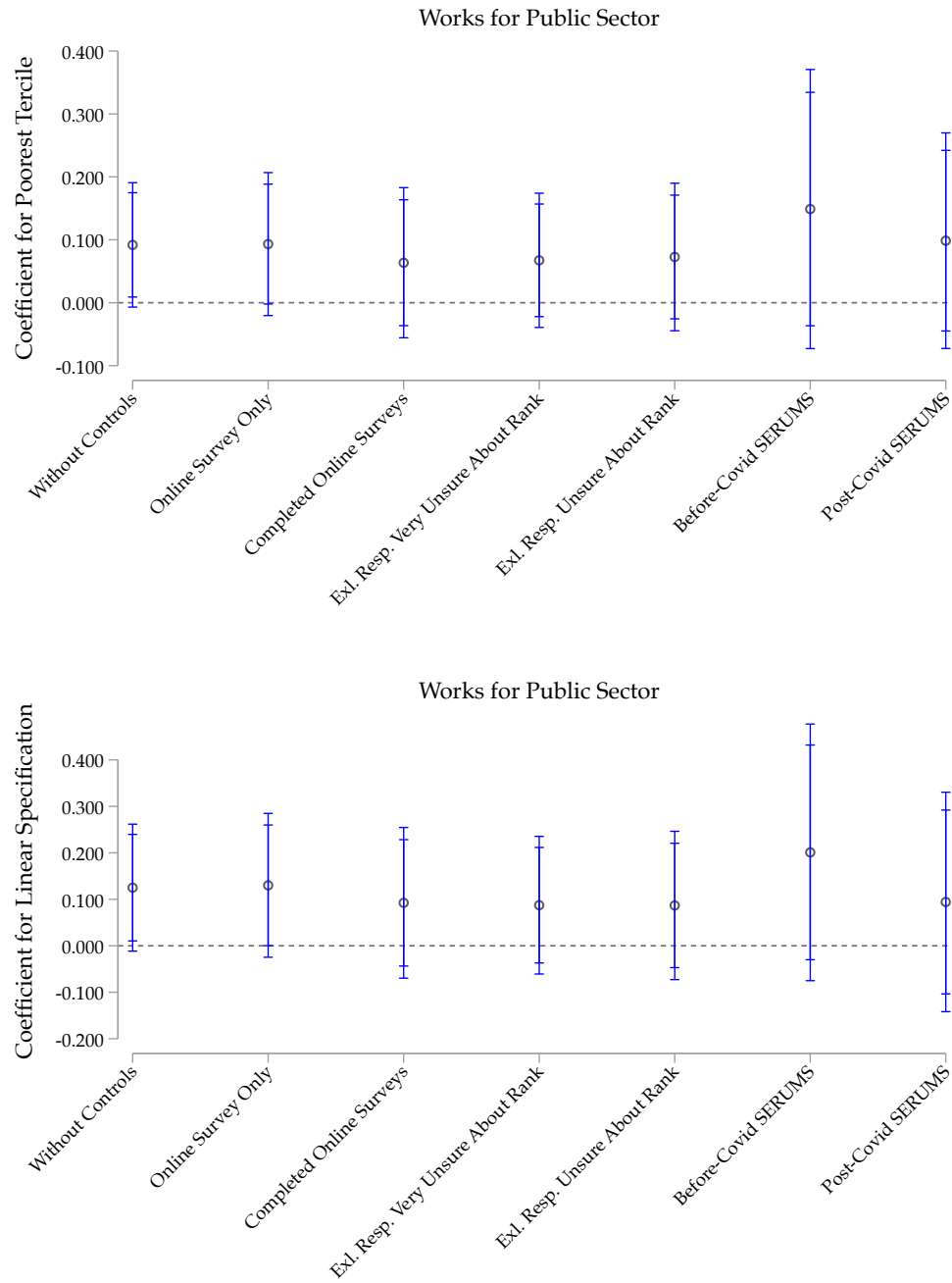
Notes: The figure plots the distribution of the district-level government poverty index from FONCODES.

Figure 1.A.4: Map of SERUMS Position



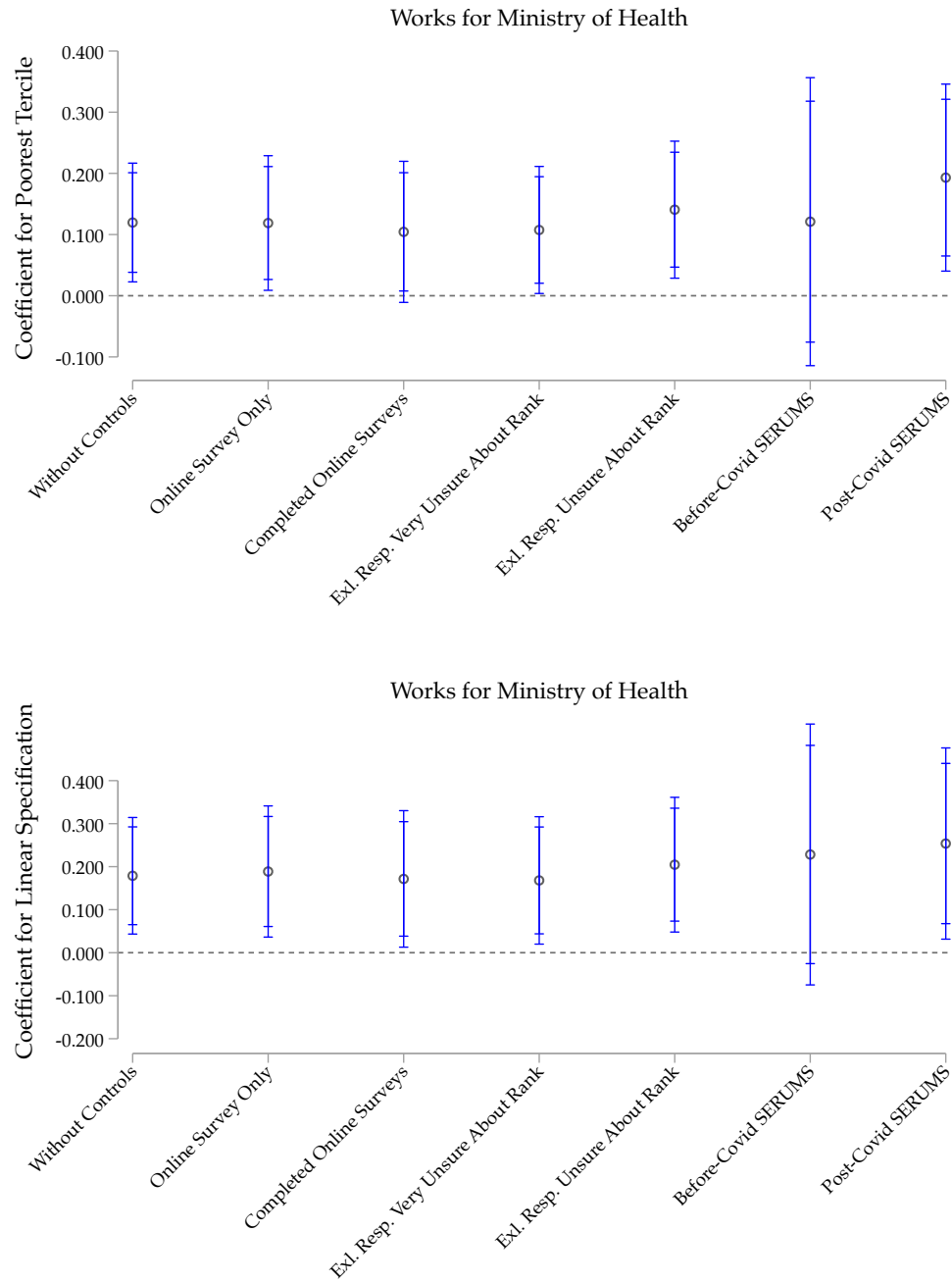
Notes: Each polgyon corresponds to a district in Peru. Areas in dark blue received a SERUMS psychol-  
ogists during our sample period.

Figure 1.A.5: Robustness Checks: Public Sector Employment



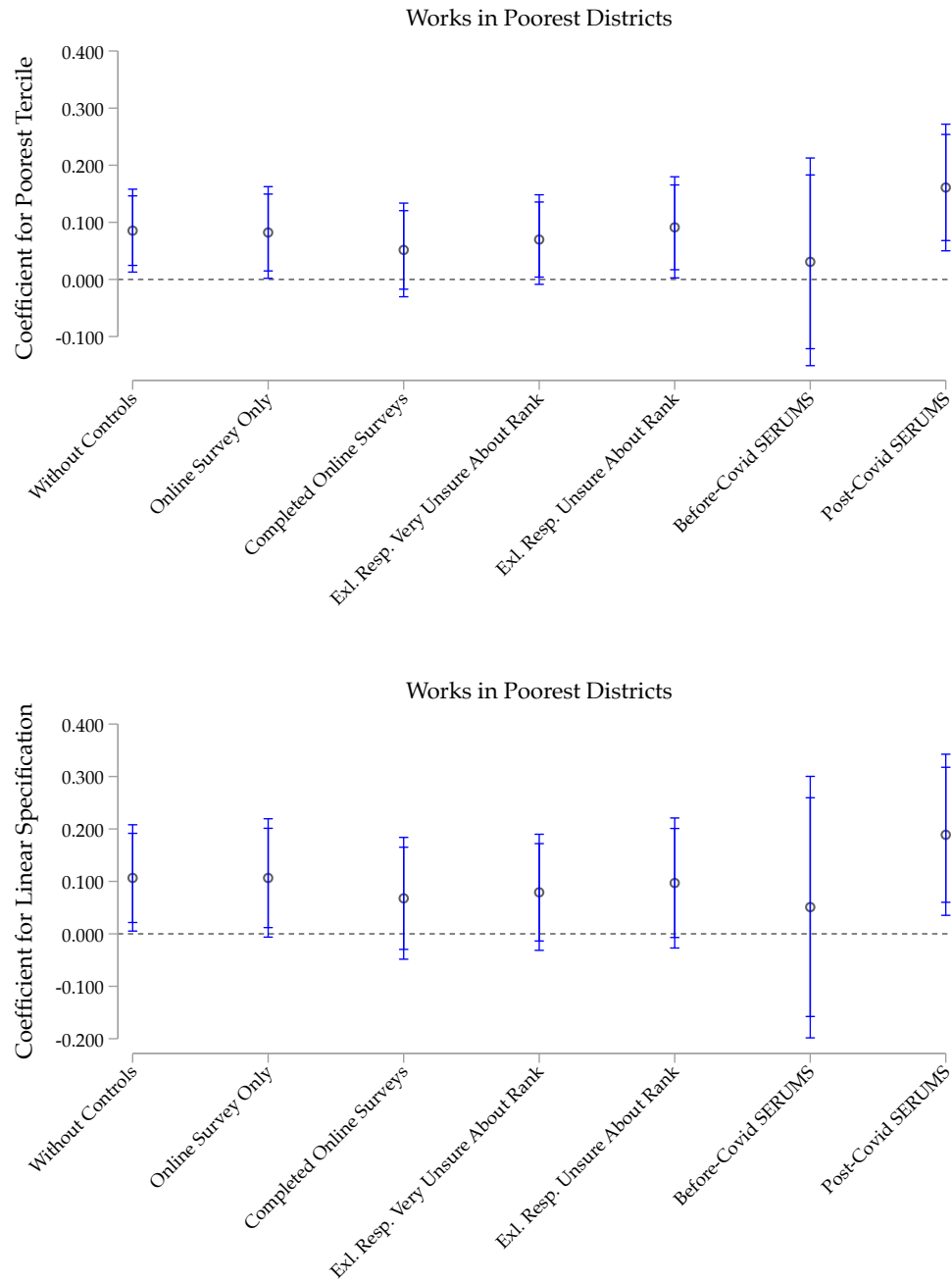
Notes: The first figure plots the regression coefficients for the poorest poverty tercile dummy and the second figure plots the regression coefficient for the linear poverty score variable. The figures show 90 and 95 percent confidence intervals.

Figure 1.A.6: Robustness Checks: Ministry of Health Employment



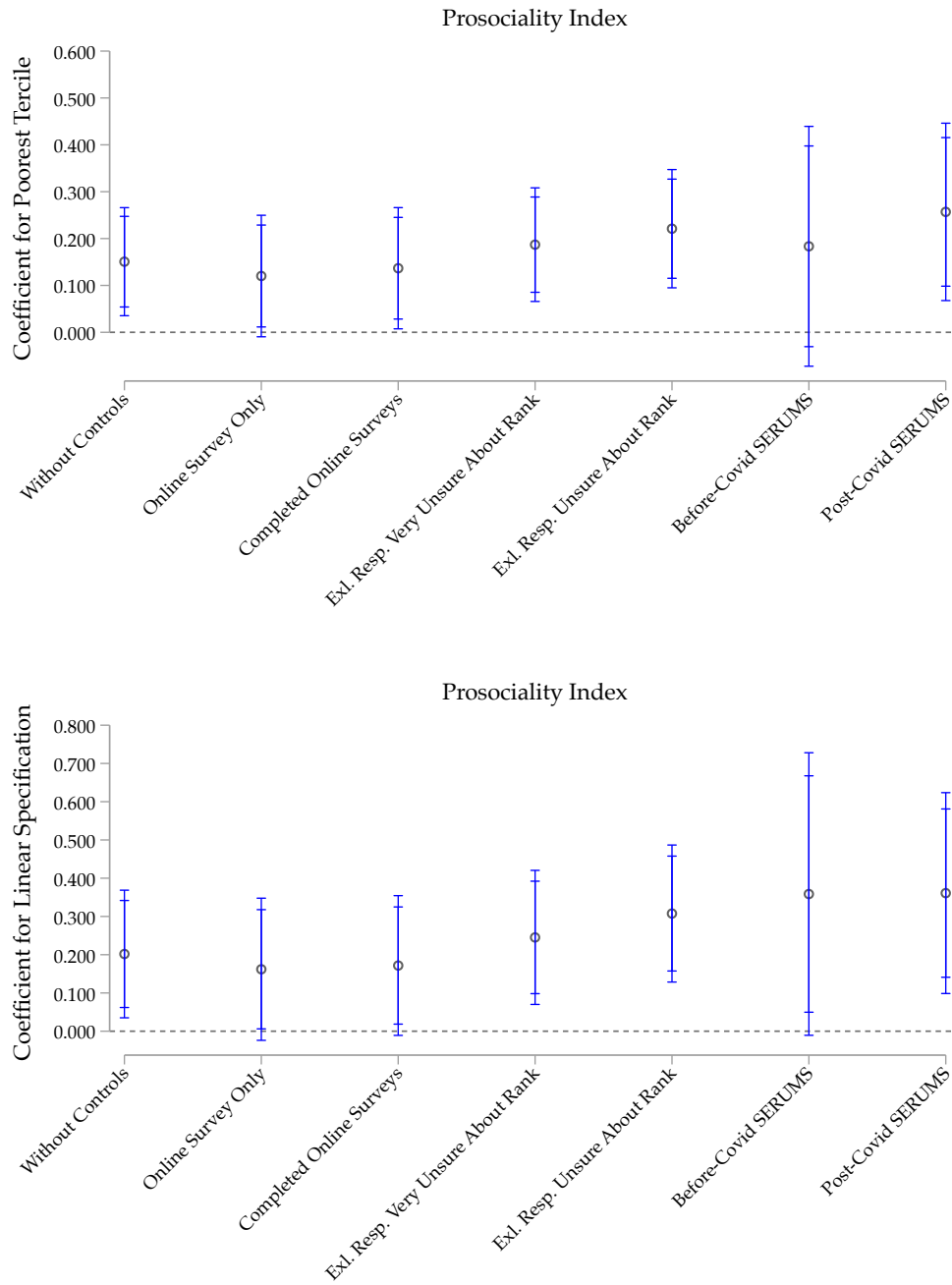
Notes: The first figure plots the regression coefficients for the poorest poverty tercile dummy and the second figure plots the regression coefficient for the linear poverty score variable. The figures show 90 and 95 percent confidence intervals.

Figure 1.A.7: Robustness Checks: Employment in Poorest Districts



Notes: The first figure plots the regression coefficients for the poorest poverty tercile dummy and the second figure plots the regression coefficient for the linear poverty score variable. The figures show 90 and 95 percent confidence intervals.

Figure 1.A.8: Robustness Checks: Prosociality Index



Notes: The first figure plots the regression coefficients for the poorest poverty tercile dummy and the second figure plots the regression coefficient for the linear poverty score variable. The figures show 90 and 95 percent confidence intervals.

Table 1.A.1: Bonus Point Structure

|            | Poverty Index | Public Sector |
|------------|---------------|---------------|
| Poverty Q1 | 0.6 – 1.0     | 15%           |
| Poverty Q2 | 0.15 – 0.59   | 10%           |
| Poverty Q3 | 0.06 – 0.14   | 5%            |
| Poverty Q4 | 0.03 – 0.05   | 2%            |
| Poverty Q5 | 0 – 0.02      | 0%            |

Notes: The second column shows the cutoff points for the poverty quintile definitions. The third column shows the bonus points that are given to psychologists for public sector application based on the poverty quintile of the SERUMS location.

Table 1.A.2: Effect of Lottery Rank on SERUMS Location for Nutritionists

|                                          | Government<br>Poverty index<br>(1) | Rural Share<br>(2) |
|------------------------------------------|------------------------------------|--------------------|
| <i>Panel A: Linear Specification</i>     |                                    |                    |
| Percentile Rank                          | 0.026<br>(0.059)                   | -0.029<br>(0.057)  |
| <i>Panel B: Discrete Specification</i>   |                                    |                    |
| Medium Rank Tercile                      | 0.009<br>(0.048)                   | 0.026<br>(0.046)   |
| Bottom Rank Tercile                      | 0.016<br>(0.042)                   | -0.020<br>(0.039)  |
| pval: Medium Tercile<br>= Bottom Tercile | 0.891                              | 0.350              |
| Outcome Mean                             | 0.454                              | 0.480              |
| Included in Phone Survey                 | Yes                                | Yes                |
| Observations                             | 185                                | 185                |

Notes: Robust standard errors appear in parentheses. All regressions include lottery stratification fixed effects and a vector of controls that consists of age, gender, a dummy for whether the respondent's university was public, a dummy for whether Spanish is the respondent's native language, birth region fixed effects, and a dummy variable that indicates whether the survey was done over the phone.

Table 1.A.3: Attrition Check

|                                          | Started<br>Survey<br>(1) | Finished<br>Survey<br>(2) |
|------------------------------------------|--------------------------|---------------------------|
| <i>Panel A: Linear Specification</i>     |                          |                           |
| Percentile Rank                          | 0.039<br>(0.129)         | 0.033<br>(0.132)          |
| <i>Panel B: Discrete Specification</i>   |                          |                           |
| Medium Rank Tercile                      | 0.019<br>(0.100)         | -0.033<br>(0.102)         |
| Bottom Rank Tercile                      | 0.045<br>(0.098)         | 0.040<br>(0.099)          |
| pval: Medium Tercile<br>= Bottom Tercile | 0.741                    | 0.387                     |
| Outcome Mean                             | 0.604                    | 0.583                     |
| Observations                             | 192                      | 192                       |

Notes: Robust standard errors appear in parentheses. All regressions include lottery stratification fixed effects.

Table 1.A.4: Balance Check Conditional on Survey Completion

|                                         | Public<br>University | Age at<br>SERUMS<br>Start | Born in<br>Lima   | Native<br>Language<br>is Spanish | Female            | Wanted to<br>Work for<br>Public<br>Sector | Wanted to<br>Work for<br>Ministry of<br>Health |
|-----------------------------------------|----------------------|---------------------------|-------------------|----------------------------------|-------------------|-------------------------------------------|------------------------------------------------|
|                                         | (1)                  | (2)                       | (3)               | (4)                              | (5)               | (6)                                       | (7)                                            |
| <i>Panel A: Linear Specification</i>    |                      |                           |                   |                                  |                   |                                           |                                                |
| Percentile Rank                         | -0.014<br>(0.034)    | -0.451<br>(1.303)         | -0.071<br>(0.055) | -0.025<br>(0.047)                | -0.008<br>(0.066) | 0.014<br>(0.063)                          | -0.112<br>(0.082)                              |
| <i>Panel B: Discrete Specification</i>  |                      |                           |                   |                                  |                   |                                           |                                                |
| Medium Rank Tercile                     | 0.000<br>(0.021)     | -0.591<br>(0.957)         | -0.029<br>(0.039) | -0.022<br>(0.034)                | -0.004<br>(0.047) | 0.098**<br>(0.042)                        | -0.002<br>(0.060)                              |
| Bottom Rank Tercile                     | -0.007<br>(0.023)    | -0.173<br>(0.918)         | -0.041<br>(0.040) | -0.009<br>(0.034)                | -0.007<br>(0.048) | -0.000<br>(0.046)                         | -0.103*<br>(0.059)                             |
| pval: Medium Tercile<br>= Worst Tercile | 0.750                | 0.687                     | 0.773             | 0.733                            | 0.959             | 0.029                                     | 0.114                                          |
| Outcome Mean                            | 0.173                | 27.069                    | 0.201             | 0.883                            | 0.706             | 0.828                                     | 0.478                                          |
| Included in Phone Survey                | Yes                  | Yes                       | Yes               | Yes                              | Yes               | No                                        | No                                             |
| Observations                            | 611                  | 611                       | 557               | 611                              | 611               | 475                                       | 475                                            |

Notes: Robust standard errors appear in parentheses. All regressions include lottery stratification fixed effects. The sample consists of survey respondents. Column 1 comes from administrative data. Columns 2-8 come from survey data.

Table 1.A.5: Relationship Between Lottery Rank and SERUMS Location Attributes

|                                       | Bottom<br>Rank<br>Tercile |
|---------------------------------------|---------------------------|
|                                       | (1)                       |
| Government Poverty Index              | 0.307**<br>(0.149)        |
| Mid-sized Primary Healthcare Facility | -0.256*<br>(0.153)        |
| Large Primary Healthcare Facility     | -0.140<br>(0.160)         |
| Hospital                              | -0.426**<br>(0.192)       |
| Same Department as Birthplace         | -0.119<br>(0.117)         |
| Monthly Salary (in 1,000 Soles)       | 0.013<br>(0.073)          |
| Outcome Mean                          | 0.445                     |
| Observations                          | 384                       |

Notes: Robust standard errors appear in parentheses. All regressions include lottery stratification fixed effects.

Table 1.A.6: Effect of Lottery Rank on Donation Subindex Components

|                                          | Main Sample         |                      |                              |                      | Current And Recent Participants |                      |                              |
|------------------------------------------|---------------------|----------------------|------------------------------|----------------------|---------------------------------|----------------------|------------------------------|
|                                          | Past 30 Days        |                      | Donates Everything           |                      | Past 30 Days                    |                      | Donates Everything           |
|                                          | Donated Time<br>(1) | Donated Money<br>(2) | Hypothetical Scenario<br>(3) | Dictator Game<br>(4) | Donated Time<br>(5)             | Donated Money<br>(6) | Hypothetical Scenario<br>(7) |
| <i>Panel A: Linear Specification</i>     |                     |                      |                              |                      |                                 |                      |                              |
| Percentile Rank                          | 0.127<br>(0.077)    | -0.003<br>(0.077)    | 0.027<br>(0.054)             | 0.014<br>(0.071)     | 0.057<br>(0.104)                | -0.079<br>(0.097)    | 0.108<br>(0.075)             |
| <i>Panel B: Discrete Specification</i>   |                     |                      |                              |                      |                                 |                      |                              |
| Medium Rank Tercile                      | 0.046<br>(0.060)    | -0.044<br>(0.058)    | 0.018<br>(0.040)             | 0.003<br>(0.048)     | -0.075<br>(0.081)               | -0.032<br>(0.072)    | 0.026<br>(0.051)             |
| Bottom Rank Tercile                      | 0.077<br>(0.055)    | 0.019<br>(0.055)     | 0.047<br>(0.038)             | -0.015<br>(0.050)    | 0.039<br>(0.077)                | -0.054<br>(0.073)    | 0.091*<br>(0.053)            |
| pval: Medium Tercile<br>= Bottom Tercile | 0.600               | 0.287                | 0.494                        | 0.729                | 0.158                           | 0.774                | 0.242                        |
| Outcome Mean                             | 0.567               | 0.383                | 0.134                        | 0.190                | 0.462                           | 0.605                | 0.126                        |
| Included in Phone Survey                 | Yes                 | Yes                  | Yes                          | No                   | Yes                             | Yes                  | Yes                          |
| Observations                             | 528                 | 527                  | 529                          | 457                  | 310                             | 310                  | 310                          |

Notes: Robust standard errors appear in parentheses. All regressions include lottery stratification fixed effects and a vector of controls that consists of age, gender, a dummy for whether the respondent's university was public, a dummy for whether Spanish is the respondent's native language, and birth region fixed effects. Columns 1-3 and 5-7 were also asked as part of the phone survey and include further a dummy variable for whether the survey was done by phone. The sample in columns 1-4 consists of psychologists who completed SERUMS at the point of the survey. The sample in columns 5-7 consists of psychologists who were currently doing SERUMS or who completed the program three months prior.

Table 1.A.7: Effect of Lottery Rank on Additional Donation Outcomes

|                                         | Total Donation to       |                                | Reason for Not Donating   |                   |                           |                                 |                                    |                       |
|-----------------------------------------|-------------------------|--------------------------------|---------------------------|-------------------|---------------------------|---------------------------------|------------------------------------|-----------------------|
|                                         | Vitamin A<br>NGO<br>(1) | Urban<br>Culture<br>NGO<br>(2) | Needs More<br>Info<br>(3) | Mistrust<br>(4)   | Already<br>Donates<br>(5) | Helps<br>Through<br>Work<br>(6) | Does Not<br>Support<br>NGOs<br>(7) | Needs<br>Money<br>(8) |
| <i>Panel A: Linear Specification</i>    |                         |                                |                           |                   |                           |                                 |                                    |                       |
| Percentile Rank                         | -0.060*<br>(0.036)      | 0.028<br>(0.018)               | 0.005<br>(0.163)          | -0.007<br>(0.168) | 0.128<br>(0.129)          | 0.123<br>(0.150)                | -0.063<br>(0.097)                  | -0.184<br>(0.138)     |
| <i>Panel B: Discrete Specification</i>  |                         |                                |                           |                   |                           |                                 |                                    |                       |
| Medium Rank Tercile                     | -0.499<br>(0.572)       | 0.756<br>(0.523)               | 0.106<br>(0.111)          | -0.053<br>(0.114) | -0.146*<br>(0.085)        | -0.072<br>(0.115)               | -0.053<br>(0.053)                  | -0.008<br>(0.106)     |
| Bottom Rank Tercile                     | -0.508<br>(0.636)       | 0.459<br>(0.498)               | 0.045<br>(0.119)          | -0.012<br>(0.123) | 0.058<br>(0.109)          | 0.077<br>(0.119)                | -0.047<br>(0.068)                  | -0.165*<br>(0.096)    |
| pval: Medium Tercile<br>= Worst Tercile | 0.987                   | 0.604                          | 0.604                     | 0.750             | 0.026                     | 0.210                           | 0.905                              | 0.084                 |
| Outcome Mean                            | 2.503                   | 1.536                          | 0.200                     | 0.217             | 0.167                     | 0.300                           | 0.050                              | 0.200                 |
| Included in Phone Survey                | No                      | No                             | No                        | No                | No                        | No                              | No                                 | No                    |
| Observations                            | 457                     | 457                            | 153                       | 153               | 153                       | 153                             | 153                                | 153                   |

Notes: Robust standard errors appear in parentheses. All regressions include lottery stratification fixed effects and a vector of controls that consists of age, gender, a dummy for whether the respondent's university was public, a dummy for whether Spanish is the respondent's native language, and birth region fixed effects. Columns 3-8 is restricted to respondents who did not donate anything as part of the dictator game.

Table 1.A.8: Effect of Lottery Rank on Beliefs

|                                          | Main Sample                                                 |                             |                                                                                  |                                                                  | Current And Recent Participants                             |                             |                                                                                  |                                                                  |
|------------------------------------------|-------------------------------------------------------------|-----------------------------|----------------------------------------------------------------------------------|------------------------------------------------------------------|-------------------------------------------------------------|-----------------------------|----------------------------------------------------------------------------------|------------------------------------------------------------------|
|                                          | External<br>Circum-<br>stances Are<br>Reason for<br>Poverty | Trust in<br>Govern-<br>ment | Perceived<br>Population<br>Share<br>Covered by<br>Social<br>Protection<br>Scheme | Public vs.<br>Private:<br>Opportuni-<br>ties to Help<br>the Poor | External<br>Circum-<br>stances Are<br>Reason for<br>Poverty | Trust in<br>Govern-<br>ment | Perceived<br>Population<br>Share<br>Covered by<br>Social<br>Protection<br>Scheme | Public vs.<br>Private:<br>Opportuni-<br>ties to Help<br>the Poor |
|                                          | (1)                                                         | (2)                         | (3)                                                                              | (4)                                                              | (5)                                                         | (6)                         | (7)                                                                              | (8)                                                              |
| <i>Panel A: Linear Specification</i>     |                                                             |                             |                                                                                  |                                                                  |                                                             |                             |                                                                                  |                                                                  |
| Percentile Rank                          | 0.071<br>(0.074)                                            | 0.040<br>(0.113)            | -6.191**<br>(2.888)                                                              | 0.687***<br>(0.192)                                              | 0.019<br>(0.107)                                            | -0.148<br>(0.158)           | -2.278<br>(4.393)                                                                | 0.600**<br>(0.282)                                               |
| <i>Panel B: Discrete Specification</i>   |                                                             |                             |                                                                                  |                                                                  |                                                             |                             |                                                                                  |                                                                  |
| Medium Rank Tercile                      | 0.052<br>(0.058)                                            | -0.009<br>(0.086)           | -3.316<br>(2.186)                                                                | 0.043<br>(0.153)                                                 | 0.040<br>(0.077)                                            | -0.153<br>(0.127)           | 1.045<br>(2.973)                                                                 | 0.285<br>(0.210)                                                 |
| Bottom Rank Tercile                      | 0.079<br>(0.053)                                            | 0.045<br>(0.080)            | -3.453*<br>(2.070)                                                               | 0.372***<br>(0.134)                                              | -0.006<br>(0.078)                                           | -0.070<br>(0.115)           | -0.454<br>(3.146)                                                                | 0.364*<br>(0.195)                                                |
| pval: Medium Tercile<br>= Bottom Tercile | 0.632                                                       | 0.553                       | 0.954                                                                            | 0.037                                                            | 0.558                                                       | 0.476                       | 0.610                                                                            | 0.716                                                            |
| Outcome Mean                             | 0.560                                                       | 3.129                       | 63.387                                                                           | 3.630                                                            | 0.658                                                       | 3.104                       | 62.304                                                                           | 3.424                                                            |
| Included in Phone Survey                 | Yes                                                         | No                          | No                                                                               | No                                                               | Yes                                                         | No                          | No                                                                               | No                                                               |
| Observations                             | 549                                                         | 475                         | 475                                                                              | 481                                                              | 317                                                         | 298                         | 298                                                                              | 302                                                              |

Notes: Robust standard errors appear in parentheses. All regressions include lottery stratification fixed effects and a vector of controls that consists of age, gender, a dummy for whether the respondent's university was public, a dummy for whether Spanish is the respondent's native language, and birth region fixed effects. Column 1 was also asked as part of the phone survey and includes further a dummy variable for whether the survey was done by phone.

Table 1.A.9: Effect of Lottery Rank on Perceptions of Jobs in Public Sector

|                                          | Private vs. Public Sector   |                     |                                          |                                 |                   |                                          |
|------------------------------------------|-----------------------------|---------------------|------------------------------------------|---------------------------------|-------------------|------------------------------------------|
|                                          | Main Sample                 |                     |                                          | Current And Recent Participants |                   |                                          |
|                                          | Work-Life<br>Balance<br>(1) | Salary<br>(2)       | Intellectual<br>Satisfac-<br>tion<br>(3) | Work-Life<br>Balance<br>(4)     | Salary<br>(5)     | Intellectual<br>Satisfac-<br>tion<br>(6) |
| <i>Panel A: Linear Specification</i>     |                             |                     |                                          |                                 |                   |                                          |
| Percentile Rank                          | 0.238<br>(0.194)            | 0.629***<br>(0.204) | 0.123<br>(0.186)                         | 0.144<br>(0.268)                | 0.485*<br>(0.286) | 0.186<br>(0.266)                         |
| <i>Panel B: Discrete Specification</i>   |                             |                     |                                          |                                 |                   |                                          |
| Medium Rank Tercile                      | 0.046<br>(0.142)            | -0.031<br>(0.148)   | 0.017<br>(0.139)                         | 0.009<br>(0.177)                | 0.053<br>(0.209)  | -0.086<br>(0.179)                        |
| Bottom Rank Tercile                      | 0.092<br>(0.140)            | 0.359**<br>(0.152)  | 0.127<br>(0.133)                         | 0.058<br>(0.197)                | 0.291<br>(0.207)  | 0.202<br>(0.188)                         |
| pval: Medium Tercile<br>= Bottom Tercile | 0.770                       | 0.014               | 0.456                                    | 0.817                           | 0.270             | 0.137                                    |
| Outcome Mean                             | 3.016                       | 2.953               | 3.197                                    | 2.898                           | 2.992             | 3.154                                    |
| Included in Phone Survey                 | No                          | No                  | No                                       | No                              | No                | No                                       |
| Observations                             | 480                         | 482                 | 479                                      | 302                             | 302               | 301                                      |

Notes: Robust standard errors appear in parentheses. All regressions include lottery stratification fixed effects and a vector of controls that consists of age, gender, a dummy for whether the respondent's university was public, a dummy for whether Spanish is the respondent's native language, birth region fixed effects, and a dummy variable that indicates whether the survey was done over the phone.

Table 1.A.10: Effect of Lottery Rank on Perceptions of Jobs in Poorest Districts

|                                          | Richest vs. Poorest Districts               |                                     |                            |                                      |                                             |                                     |                            |                                      |
|------------------------------------------|---------------------------------------------|-------------------------------------|----------------------------|--------------------------------------|---------------------------------------------|-------------------------------------|----------------------------|--------------------------------------|
|                                          | Main Sample                                 |                                     |                            |                                      | Current And Recent Participants             |                                     |                            |                                      |
|                                          | Opportunities<br>to Help the<br>Poor<br>(1) | Local<br>Infrastruc-<br>ture<br>(2) | Safety<br>Situation<br>(3) | Compability<br>with<br>Spouse<br>(4) | Opportunities<br>to Help the<br>Poor<br>(5) | Local<br>Infrastruc-<br>ture<br>(6) | Safety<br>Situation<br>(7) | Compability<br>with<br>Spouse<br>(8) |
| <i>Panel A: Linear Specification</i>     |                                             |                                     |                            |                                      |                                             |                                     |                            |                                      |
| Percentile Rank                          | 0.057<br>(0.220)                            | -0.199<br>(0.220)                   | 0.108<br>(0.210)           | -0.005<br>(0.179)                    | 0.021<br>(0.310)                            | -0.088<br>(0.302)                   | -0.185<br>(0.282)          | -0.143<br>(0.257)                    |
| <i>Panel B: Discrete Specification</i>   |                                             |                                     |                            |                                      |                                             |                                     |                            |                                      |
| Medium Rank Tercile                      | -0.218<br>(0.172)                           | 0.058<br>(0.166)                    | 0.114<br>(0.158)           | -0.148<br>(0.142)                    | -0.032<br>(0.210)                           | -0.044<br>(0.210)                   | -0.168<br>(0.195)          | -0.325*<br>(0.180)                   |
| Bottom Rank Tercile                      | -0.052<br>(0.157)                           | -0.139<br>(0.158)                   | 0.051<br>(0.152)           | -0.057<br>(0.129)                    | -0.014<br>(0.226)                           | -0.136<br>(0.221)                   | -0.163<br>(0.207)          | -0.150<br>(0.194)                    |
| pval: Medium Tercile<br>= Bottom Tercile | 0.353                                       | 0.254                               | 0.706                      | 0.528                                | 0.944                                       | 0.677                               | 0.983                      | 0.398                                |
| Outcome Mean                             | 3.580                                       | 2.314                               | 2.883                      | 2.697                                | 3.440                                       | 2.427                               | 2.957                      | 2.774                                |
| Included in Phone Survey                 | No                                          | No                                  | No                         | No                                   | No                                          | No                                  | No                         | No                                   |
| Observations                             | 479                                         | 479                                 | 479                        | 478                                  | 300                                         | 301                                 | 299                        | 298                                  |

Notes: Robust standard errors appear in parentheses. All regressions include lottery stratification fixed effects and a vector of controls that consists of age, gender, a dummy for whether the respondent's university was public, a dummy for whether Spanish is the respondent's native language, and birth region fixed effects. The sample in columns 1-4 consists of psychologists who completed SERUMS at the point of the survey. The sample in columns 5-8 consists of psychologists who were currently doing SERUMS or who completed the program three months prior.

Table 1.A.11: Effect of Lottery Rank on Additional Job Preferences

|                                          | ... Is Very Important for Choosing a Job |                   |                   |                         |                            |                   |                       |
|------------------------------------------|------------------------------------------|-------------------|-------------------|-------------------------|----------------------------|-------------------|-----------------------|
|                                          | Salary                                   | Work-Life Balance | Safety            | Compability with Spouse | Intellectual Satisfac-tion | Work Envi-ronment | Local Infrastruc-ture |
|                                          | (1)                                      | (2)               | (3)               | (4)                     | (5)                        | (6)               | (7)                   |
| <i>Panel A: Linear Specification</i>     |                                          |                   |                   |                         |                            |                   |                       |
| Percentile Rank                          | -0.018<br>(0.187)                        | 0.290<br>(0.177)  | -0.088<br>(0.207) | -0.111<br>(0.275)       | 0.169<br>(0.174)           | 0.153<br>(0.230)  | 0.057<br>(0.251)      |
| <i>Panel B: Discrete Specification</i>   |                                          |                   |                   |                         |                            |                   |                       |
| Medium Rank Tercile                      | -0.008<br>(0.056)                        | 0.059<br>(0.058)  | -0.033<br>(0.059) | -0.043<br>(0.045)       | 0.044<br>(0.058)           | 0.038<br>(0.059)  | -0.027<br>(0.048)     |
| Bottom Rank Tercile                      | 0.005<br>(0.052)                         | 0.088<br>(0.054)  | 0.004<br>(0.054)  | 0.017<br>(0.041)        | 0.095*<br>(0.056)          | 0.088<br>(0.055)  | 0.033<br>(0.047)      |
| pval: Medium Tercile<br>= Bottom Tercile | 0.818                                    | 0.618             | 0.531             | 0.198                   | 0.399                      | 0.402             | 0.227                 |
| Outcome Mean                             | 0.384                                    | 0.434             | 0.397             | 0.206                   | 0.410                      | 0.483             | 0.249                 |
| Included in Phone Survey                 | Yes                                      | Yes               | Yes               | Yes                     | Yes                        | Yes               | Yes                   |
| Observations                             | 551                                      | 553               | 550               | 535                     | 552                        | 547               | 549                   |

Notes: Robust standard errors appear in parentheses. All regressions include lottery stratification fixed effects and a vector of controls that consists of age, gender, a dummy for whether the respondent's university was public, a dummy for whether Spanish is the respondent's native language, birth region fixed effects, and a dummy variable that indicates whether the survey was done over the phone.

Table 1.A.12: Effect of Lottery Rank on Main Outcomes by Birthplace

|                                                     | Outcome Index       | Index Components             |                   |                              |                                    |
|-----------------------------------------------------|---------------------|------------------------------|-------------------|------------------------------|------------------------------------|
|                                                     |                     | Works for Ministry of Health | Altruism Index    | More Willing to Work for MoH | More Willing to Work in Poor Areas |
|                                                     | (1)                 | (2)                          | (3)               | (4)                          | (5)                                |
| Medium or Bottom Rank Tercile $\times$ Born in Lima | 0.234*<br>(0.124)   | 0.096<br>(0.105)             | 0.078<br>(0.058)  | 0.131<br>(0.094)             | 0.171<br>(0.115)                   |
| Medium or Bottom Rank Tercile                       | 0.076<br>(0.058)    | 0.069<br>(0.048)             | 0.028<br>(0.027)  | 0.081*<br>(0.048)            | -0.028<br>(0.050)                  |
| Born in Lima                                        | -0.200**<br>(0.097) | -0.114<br>(0.086)            | -0.074<br>(0.048) | -0.083<br>(0.078)            | -0.118<br>(0.089)                  |
| Outcome Mean                                        | -0.043              | 0.332                        | 0.503             | 0.215                        | 0.230                              |
| Observations                                        | 557                 | 557                          | 553               | 550                          | 481                                |

Notes: Robust standard errors appear in parentheses. All regressions include lottery stratification fixed effects and a vector of controls that consists of age, gender, a dummy for whether the respondent's university was public, and a dummy for whether Spanish is the respondent's native language.

Table 1.A.13: Randomization Inference

| Variable                     | p-value<br>(Bottom Rank Tercile) | p-value<br>(Percentile Rank) |
|------------------------------|----------------------------------|------------------------------|
| Works For Public Sector      | 0.162                            | 0.186                        |
| Works for Ministry of Health | 0.035                            | 0.024                        |
| Works in Poorest Districts   | 0.053                            | 0.092                        |
| Prosociality Index           | 0.023                            | 0.027                        |

Notes: The exercise randomly re-assigns the lottery rank to each respondent 2,000 times. For each iteration, we then repeat our analysis. The p-value is calculated as the share of the placebo coefficients that are larger in magnitude than the actual coefficient (in absolute terms).

# Chapter 2

## The Economics of Caste Norms: Purity, Status, and Women's Work in India

### 2.1 Introduction

An estimated 100 million Indian Hindu women are disallowed from working.<sup>1</sup> Constraints on Hindu women's work are tied to the practice of female seclusion required by caste norms, the set of religious and social laws that govern behavior within the Hindu caste system (Chen, 1995; Field et al., 2010; Luke and Munshi, 2011; Jayachandran, 2015). Caste norms are guided by the idea that certain people, activities, and foods are "impure" and that (higher caste-rank) individuals should seek to preserve own purity by avoiding contact with these polluting aspects of society.<sup>2</sup> Because of this, women are instructed to stay secluded within the home, which severely limits their ability to work and participate in the community. Caste

---

<sup>1</sup>Estimate is based on responses to the 2011 India Human Development Survey, a nationally representative survey. Adult women are asked, "If you found a suitable job, would you be allowed to work?"

<sup>2</sup>All Hindus are divided into castes, which are hereditary, exogamous ethnic groups. There are approximately 3,000 distinct castes across India and, together, these ethnic groups account for 80% of India's population (Munshi, 2019).

purity norms also restrict inter-group interactions and promote discriminatory practices towards low caste-rank individuals, imposing substantial equity and efficiency costs on social and economic behavior in Indian villages.<sup>3</sup>

Caste norms are generally assumed to reflect Hindus’ religious beliefs and their presence is understood to be a fixed, long-standing feature of Hindu society (Munshi, 2019). This paper asks whether caste norms are weakened when Hindus live alongside an ethnic group that is outside of the caste system. Specifically, we study the integration of Hindus with Adivasis, an indigenous minority that makes up just under a fifth of the rural population of central India. Since Adivasis are traditionally non-Hindu, they are not subject to caste purity rules.<sup>4,5</sup> On average, Adivasi women are 1.5 times as likely to work as Hindu women and they score higher along a range of empowerment indicators.<sup>6</sup> Similarly, Adivasi men and women do not adhere to caste purity restrictions regarding the “untouchability” of low caste-rank individuals or to caste food taboos.

There is limited causal evidence on the effect of integration with an out-group on own group cultural identity (Giuliano, 2021; Bisin and Verdier, 2022). Theoretically, the impact is ambiguous: on the one hand, assimilation and desire for conformity may lead to convergence in culture while, on the other hand, desire to maintain separation may lead individuals to increase investment in their own group’s cultural practices (Bisin and Verdier, 2010). The latter may be an important consideration in our setting: though they are outside of the caste system, Adivasis are often considered to be less pure—and therefore lower social

---

<sup>3</sup>Previous empirical studies have documented wide-ranging impacts of caste purity norms, including on the efficiency of local governance (Munshi and Rosenzweig, 2015); inter-caste trade (Anderson, 2011); occupational decisions and discrimination in the labor market (Attewell and Madheswaran, 2007; Oh, 2021; Cassan et al., 2022); health and education outcomes (Munshi and Rosenzweig, 2006; Jayachandran and Pande, 2017; Spears and Thorat, 2019); and consumption behavior (Atkin et al., 2021).

<sup>4</sup>Today, a majority of Adivasis report that they practice Hinduism in addition to or instead of their traditional religious practices (IHDS 2011). However, for clarity of exposition, we will use the term “Hindu” throughout the paper to refer to ethnic groups who have traditionally been part of the Hindu caste system.

<sup>5</sup>Adivasis’ self-identification highlights that a key feature of their society is the “absence of division of labour and caste system” and also the “relative freedom of women” (UNWGIP 1994).

<sup>6</sup>For instance, Adivasi women have more intra-household bargaining power and are more likely to participate in local governance (IHDS 2011).

status—than most Hindus. At the same time, adherence to caste purity norms is costly; anthropologists argue that Hindus are willing to incur these costs primarily as a means of gaining or maintaining status within the caste system (Srinivas, 1966). If increased Adivasi population share is associated with reduced value to caste status, then we should expect Adivasi presence to be associated with a reduction in adherence to caste norms.

To shed light on this question, we turn to central India where 64% of the total Adivasi population in India is based. This area corresponds to the former Gondwana Kingdom, a region settled and ruled by the Gonds, an Adivasi tribe. Hindus’ in-migration to central India occurred primarily during the 18th to 20th centuries and village composition has changed little since that time.<sup>7</sup> Using multiple empirical strategies, including cross-sectional analysis for the entire region and a historical natural experiment in the state of Odisha, we provide causal evidence that caste norms are weaker when Hindus in central India live in villages with a higher share of Adivasi neighbors.

Our analysis is guided by the *Manusmriti*, a 200 BCE religious text. Caste norms practiced today are derived primarily from the Manusmriti’s over 2,000 prescriptions for appropriate behavior concerning: (a) women’s seclusion; (b) marriage rites; (c) inter-caste interactions; and (d) food taboos (Chakravarti, 2018; Piketty, 2021).<sup>8,9</sup> Across all settings, we find that Hindu women in Adivasi-majority villages are less likely to practice *purdah* (female veiling) and report more mobility. Consistent with this, they are substantially more likely to work and have higher yearly earnings. Similarly, reduced adherence to female seclusion norms improves Hindu women’s participation in local governance. We also find that Hindus in Adivasi-majority villages are less likely to report that their caste community ad-

---

<sup>7</sup>Central India includes the following states: Andhra Pradesh, Chhattisgarh, Gujarat, Jharkhand, Madhya Pradesh, Maharashtra, Odisha, Rajasthan, and Telangana. Almost every Hindu and Adivasi household in our sample reports that their family has been located in their current village for over 90 years.

<sup>8</sup>The British colonial government used the Manusmriti as the basis for Hindu Law and it continues to be cited in Indian court cases and government policy (Agnes, 1999).

<sup>9</sup>The Manusmriti also provides guidelines on appropriate occupations for members of different castes. However, these guidelines are less relevant today, especially in rural areas, where a majority of the population works in agriculture (Cassan et al., 2022).

heres to Manusmriti rules on food taboos and on marriage, including the practice of dowry and the prohibition of widow remarriage. When less constrained by rules imposed by the caste system, Hindus have different social and economic relationships with one another. We find, for instance, that villages are more likely to be integrated and that Hindus are less likely to practice untouchability towards *Dalits* (a group of castes formerly referred to as “untouchables”). This is a striking finding since the practice of untouchability against Dalits is considered to be the one of the most harmful and most deeply rooted caste purity norms (Dirks, 2011).

The key econometric challenge to our analysis is that Hindus and Adivasis settled non-randomly, raising the concern that selection or unobservable features of Adivasi villages influence our results. Almost all households in both settings report that their families have been located in their current village for over 90 years, reducing the concern that contemporaneous migration drives the observed patterns. That said, we show that results are robust to the inclusion of caste fixed effects, which helps to address the concern that certain castes (e.g. less purity-norm-adherent castes) were more likely to migrate in to villages with Adivasis. In an extension to our analysis, we leverage a prediction from our conceptual framework that Adivasi-share impacts should be heterogeneous by caste rank. For this analysis, we include village fixed effects and show that results remain robust, which helps to address the concern that unobservable village-level confounders drive our estimates.

Despite the robustness of our cross-sectional findings, there remains the concern that omitted characteristics of Adivasi villages influence our analysis. We therefore exploit a historical natural experiment that led to quasi-random variation in Adivasi population share across villages in western Odisha (one of the states in our central India sample). In 1894, railroad expansion to the district of Sambalpur in Odisha expanded opportunities for trade and led to an influx of Hindu cultivators from neighboring districts into a previously Adivasi-majority area. The new railroad stop was located directly on the eastern bank of the Ma-

hanadi River, a 1 mile wide river whose single crossing point in the region was at the railroad station itself (Figure 1). Extensive historical evidence documents that Hindu migrants came primarily from districts to the west and south-west of the new railway station and were constrained by the river to settle on its western banks. Using the Mahanadi River boundary within a regression discontinuity design, we find that villages directly to the east of the river (where fewer Hindus settled) have roughly 20 percentage points higher Adivasi share today.

The validity of our identification strategy relies on the assumption that the villages on either side of the river crossing were comparable prior to the introduction of the railroad and also that the railroad did not differentially impact market access and profits for rice farmers on the eastern versus western side of the river. We collect and digitize rich archival data to validate these assumptions: using village records and individual land-ownership data for 1915-1955, we show evidence that villages were settled by the same set of castes and tribes on both side of the river and that village ethnic composition has not meaningfully changed in the past century. Villages on either side of the river are not distinguishable in terms of land value, tax incidence, and the price of rice paid to farmers in the decades following the railroad construction. Our regression discontinuity results are also highly robust to the selection of bandwidth and RD functional form.

Having established that Adivasis' presence shapes the practice of caste norms among their Hindu neighbors, we turn to mechanisms. Our conceptual approach is motivated by an extensive ethnographic and historical literature on caste norms, which argues that adoption of purity norms is seen as a means of upward social mobility for caste groups, especially those that are in the middle of the caste hierarchy (versus those at the bottom, who have a harder time escaping their low inherited social status).<sup>10</sup> Qualitative researchers emphasize that widespread take-up of purity practices occurred relatively recently, when British colonial

---

<sup>10</sup>For discussion of *Sanskritization* (also termed *Brahmanisation*), the social process by which castes adopt purity norms as a means of increasing their social status, see for instance Srinivas (1952); Ghurye (1969); Bayly (2001); Chakravarti (2018).

interventions and other features of the socio-economic environment made it more valuable to assimilate with Brahmins, the highest-ranking caste and one with a long history of practicing purity norms (Bayly, 2001; Dirks, 2011; Piketty, 2021).

With these historical observations in mind, we hypothesize that increased Adivasi share weakens caste norms because it is associated with reduced value to assimilating with Brahmins and, conversely, increased value to maintaining social and economic ties with Adivasis. Since caste norms directly limit interactions with Adivasis and promote discriminatory behavior towards them, maintaining ties is easier when adherence is lower. To provide evidence for this hypothesis, we turn first to historical data on castes' cultural practices. For each of the main castes in our RD sample, we collect and digitize data on their practice of purity norms at the beginning of the 20th century and compare it to estimates from our primary survey data collection today (Figure 3). We find that, while Brahmins have consistently adhered to purity norms regarding female seclusion, widow remarriage, and meat taboos, for other castes in our sample take-up of these norms was non-existent then but prevalent now. Similarly, Hindus' relationship with Adivasis has also changed: data from 1906 shows that, at that time, even high-caste Hindus in our sample districts did not consider Adivasis to be "impure." Now, over half our sample practice untouchability towards them.<sup>11</sup> Strikingly, though, in higher Adivasi share villages (east of river), the proportion of respondents who practice Adivasi untouchability drops by 50 percent. Thus, we find that discrimination against Adivasis decreases when they are dominant in the village. In line with this, we find that Hindus have much stronger social and financial ties with Adivasis when they are larger in group size. Moreover, we show that the village social hierarchy is malleable and adjusts to ethnic composition: we conduct a survey with elders in our RD sample villages, asking

---

<sup>11</sup>The practice of untouchability has important implications for Adivasis' ability to access public resources and participate in community events in the village. Following the anthropological literature, we measure untouchability by asking respondents whether they would touch water or food that has been touched by an Adivasi (Xaxa, 2008; Skoda et al., 2013). The rate of homicides against Adivasis and low-rank castes has also been linked to adherence to purity norms against sharing water with members of these ethnicities (Bros and Couttenier, 2015).

them to rank each caste and tribe in their village in terms of social status. In high Adivasi share villages, Adivasis are twice as likely to be ranked as either equal to or higher than a middle-rank caste group. These findings also help to make sense of a surprising feature of our results: we observe that cultural practices converge to those exercised by a socially marginalized group. Within the context of our conceptual framework, though, Adivasis' low social status can be understood at least in part as an *outcome*, rather than a fixed feature of the social context.

An alternative explanation for our findings is that when Hindus are exposed to a larger population of Adivasis they are more likely to learn about the costs and benefits of adhering to purity norms (Jarotschkin and Zhuravskaya, 2019; Schmitz et al., 2021; Boelmann et al., 2022). Or, another possibility is that Hindus emulate Adivasi cultural practices because they wish to assimilate to the most commonly observed behavior among their peers (Bisin and Verdier, 2010; Bisin et al., 2016). Both of these mechanisms may contribute to the patterns we observe though we argue that they are unlikely to be a primary driver of our results. Without additional assumptions, these mechanisms would generally imply convergence in behavior *among* Hindus and the latter mechanism would also imply that Adivasis should emulate Hindu caste norms when they are in the minority. Yet we find no impact of Hindu population share on Adivasi behavior and considerable heterogeneity in Adivasi-share impacts on Hindus by their caste rank. On average, Brahmins and other high-rank castes have strongest adherence to purity norms and adherence declines with declining caste rank. In line with our conceptual framework, increased Adivasi group size has the largest impact for middle-rank castes: intuitively, these individuals are high enough in the caste hierarchy to have found it worthwhile to invest in purity norms in the first place, but close enough in rank to Adivasis to be impacted by their presence. When Adivasis are in the majority, middle-rank Hindus have strong social network ties with them and their adherence to pu-

rity norms diminishes to the extent that, along many dimensions (including women’s work), middle-rank Hindus are not distinguishable from Adivasis.

In a final exercise, we provide further evidence for the hypothesis that adherence to purity norms will be lower when returns to interacting with Adivasis is higher. We turn to a distinct setting where British colonial policy increased Adivasis’ political and economic power, independent of their group size. In certain districts of the current states of Chhattisgarh and Odisha, the British government assigned the position of “village watchman” to an existing community leader: the village’s Adivasi priest (*jhankar*). The historical record shows that it was standard in villages across central India for there to be both an Adivasi priest and a Brahmin priest and that both served Adivasi and Hindu communities. The village watchman position was generally assigned based on merit but, as we discuss in Section 8 and Appendix B, the need to quickly put in place an administrative governing structure upon the death of the previous ruler led the British to assign this position to an existing person of respect in the community. Watchmen received tax-free land and were paid yearly tributes by all other village households. We show that blocks where the watchman position was assigned to Adivasis are not different in terms of Adivasi population share or geographic characteristics; however we caution that our sample of blocks is small and that unobservable characteristics may be confounders. With that caveat, we find that Hindus in blocks where Adivasis were assigned the watchman position are less likely to practice untouchability towards Adivasis today. They are also less likely to report adherence to female seclusion norms and caste food taboos.

This paper contributes to three bodies of work in economics. First, our findings shed light on the process of inter-group cultural transmission. There is a small but growing number of empirical studies that evaluate the impact of individual immigrants on shaping a specific cultural trait of the native majority (Jarotschkin and Zhuravskaya, 2019; Schmitz et al., 2021; Boelmann et al., 2022). Our results, which document cultural convergence across a

broad set of a practices, broaden our understanding of the ways which cultural traits are linked and jointly evolve (Bisin and Verdier, 2010). Additionally, our findings expand our understanding of cultural transmission by documenting a case where cultural practices of a higher status group converge to those of a lower status group. While the emulation of practices associated with high-status groups has been documented across several studies (La Ferrara et al., 2012; Bisin et al., 2016; Fouka, 2020), the question of whether cultural practices are shaped from below has received far less attention.

Separately, our results contribute to the growing empirical literature on the role of gender attitudes in constraining women’s work (see Jayachandran (2015), Giuliano (2021), and Jayachandran (2021) for reviews). A number of studies document the deep roots of gender attitudes and the strong persistence of these attitudes over time (Fernández and Fogli, 2009; Alesina et al., 2013; Carranza, 2014; Teso, 2019; Becker, 2021). We add to these findings by documenting how attitudes evolve in the medium-run, as the result of the integration of two groups with different religious and ethnic backgrounds. As such, our findings also relate to a set of papers that document the process of cultural change and the malleability of norms and beliefs (Fernández et al., 2004; Fernández, 2013; Bursztyn et al., 2020; Field et al., 2021; Dhar et al., 2022).

Finally, we add to our understanding of the economic and social implications of caste. Previous work has documented the influence of caste norms on occupational preferences Oh (2021); Cassan et al. (2022); health and education outcomes (Munshi and Rosenzweig, 2006; Jayachandran and Pande, 2017; Spears and Thorat, 2019); and on the efficiency of inter-caste trade (Anderson, 2011) and local governance Banerjee and Pande (2007). Several papers have studied the economic implications of caste networks (see Munshi (2019) for a review), though much of this work has focused on within-caste ties; we add to these findings by demonstrating the importance of inter-caste networks and relationships. In demonstrating that Hindus’ adherence to purity norms responds to Adivasi presence, our findings connect

to recent studies that document the endogeneity of caste identity to social and economic incentives Cassan and Vandewalle (2017); Atkin et al. (2021).

The rest of this paper proceeds as follows: Section 2 discusses the historical background and conceptual framework; Section 3 gives an overview of our data; Section 4 presents descriptive statistics on the take-up of purity norms across caste and Adivasi groups; Section 5 presents our OLS estimates of the impact of Adivasi population share on Hindus’ adherence to purity norms; Section 6 extends our analysis of this same question to our spatial regression discontinuity setting; Section 7 discusses mechanisms; and Section 8 concludes.

## 2.2 Historical Background

Our study setting is Central India, a region once ruled by the Gonds, an Adivasi tribe.<sup>12</sup> This wooded and hilly region was populated primarily by Gonds and other Adivasi tribes for most of the fourteenth through eighteenth centuries, isolated from the “Hindu empires of the plains” to the north and south (Deloche, 1993). Yet by the beginning of the 20th century several waves of Hindu migrants had arrived and villages in Central India had become, and continue to be, mixed Adivasi-Hindu societies. Hindus’ in-migration was due in part to the construction of the railroad: Between 1853 and 1930, the British government in India built an extensive railroad network, connecting inland districts in Central India to one another and to the rest of the sub-continent (Donaldson, 2018). The railroad greatly improved communication and brought new opportunities for trade to Central India, drawing in Hindu cultivators from surrounding regions (Marten, 1912). One of our estimation strategies exploits village-level variation in Hindu in-migration to Central India after the 1894 construction of a railroad stop in the state of Odisha. We further describe the historical context as it relates to the railroad construction and our identification strategy in Section

---

<sup>12</sup>The Gond Kingdom, called Gondwana, was conquered by the Marathas in the 18th century and most of the region was taken over by the British by the middle of the 19th century.

2.6. In this section, we offer historical context on the integration of Adivasi and Hindu societies and on the Hindu caste system. We then bring these historical insights together to motivate a conceptual framework for understanding how the presence of Adivasis would impact Hindus' adoption of caste purity norms.

### 2.2.1 Social Hierarchy and Cultural Practices in Adivasi-Hindu Villages

Central India is now home to at least 600 Hindu and Adivasi groups and, on average, there are 12 distinct ethnicities (tribes and castes) represented in a village.<sup>13</sup> Historians and anthropologists emphasize two key determinants of social capital and social relations within Indian villages: (1) caste hierarchy and (2) groups' relative size, political power, and economic resources. Below we discuss each of these as well as their relation to caste purity norms.

**Social Hierarchy within the Caste System** The caste system is a system of social stratification. Each Hindu person is associated with a caste (also termed *jati*), which is a hereditary, endogamous ethnic group. Traditionally, each caste was also associated with a specific occupation. Going back at least as far as 1500–500 BCE, there is evidence of a broad ranking of these castes based on concepts of purity and pollution. The justification of this ranking, as well the implications of it for activities of daily life, are summarized in the Manusmriti, a 200 BCE text. The Manusmriti is a set of social laws written by Brahmin scholars—Brahmins are the priest caste and everywhere in the Central India the highest ranking group—that was later codified as religious law for all Hindus by the British colonial government (Agnes, 1999).

---

<sup>13</sup>The Government of India collects but does not release information on the distribution of *jatis* across regions. Our estimate is based on responses to the 1992-1993 National Family and Health Survey (NFHS), which collected *jati* information for each household and is representative at the district level. Estimate of number of castes in each village is from the 2011-2011 IHDS survey.

Manu describes a society divided into four classes, or *varnas*. Brahmins sit at the top while Dalits (a term that refers to a group of castes) are considered to be so impure that they are excluded from this structure entirely. There are, however, over 3,000 castes across India and 70% of individuals in Central India belong to castes that sit between the Brahmins and the Dalits in the social hierarchy.<sup>14</sup> The British colonial government, in their 1901 Census of India, later expanded upon the four-fold *varna* classification and introduced a written ranking of every caste by region. Historians (and later census officers) now recognize that the 1901 ranking did not accurately reflect the multitude of localized rankings across villages and regions at the time (Ghurye, 1969; Bayly, 2001; Dirks, 2011).<sup>15</sup> Regardless, and potentially due to the very fact that the census ranking was created,<sup>16</sup> we find that the 1901 census rankings closely match the current caste hierarchy: in our RDD sample of 160 villages, we ask elders from different castes to rank all groups in their village and find that the rankings are almost identical to those from the 1901 census.

**Sanskritization and the Adoption of Purity Norms** The above discussion of caste rank highlights the fact that, to a large extent, an individual’s perceived purity is fixed at birth. That being said, there is also some scope for upwards or downwards social mobility within the caste system; this mobility is a function of an individual’s (and their fellow caste members’) adherence to caste purity norms. Manu, over more than 2,000 verses, gives clear prescriptions for behavior required to maintain purity. These prescriptions can be categorized into the following areas: (a) women’s role in society and the home, (b) marriage rites, (c) food taboos, (d) inter-caste interactions, and (e) occupational guidelines Chakravarti (2018).

---

<sup>14</sup>The *Kshatriyas*, or so-called “warrior” castes, are ranked second in the Manusmriti and similarly tend to be ranked immediately after the Brahmins in local caste hierarchies.

<sup>15</sup>For instance, J. H. Hutton, the 1931 census commissioner of India, wrote that “All subsequent census officers in India must have cursed the day when it occurred to Sir Herbert Risley [the 1901 census commissioner], no doubt in order to test his admirable theory of the relative nasal index, to attempt to draw up a list of castes according to their rank in society. He failed, but the results of his attempt are almost as troublesome as if he had succeeded” (Hutton 1931).

<sup>16</sup>See Dirks (2011) for a detailed discussion of the social, political, and economic impacts of the creation the caste ranking in the 1901 census.

A majority of the prescriptions are directed specifically towards Brahmins; Manu makes a distinction between how Brahmins should behave and how lower-caste individuals should behave. For instance, while widow remarriage is forbidden for Brahmins in the Manusmriti (a woman is seen as impure upon the death of her husband), it is encouraged for lower castes. Similarly, castes aside from the Brahmins historically did not practice taboos against meat or alcohol.

The fact that Brahmins, the highest ranking caste, have traditionally held different cultural practices and norms than other castes gives individuals lower on the hierarchy a clear path for seeking higher caste status: emulating Brahmin purity norms. This practice, which is termed *Sanskritization* or *Brahmanisation*, is described by the anthropologist M.N. Srinivas (1952) in the following way: “A lower caste [is] able to rise to a higher position in the hierarchy by adopting vegetarianism and teetotalism [etc.]... in short, it [takes] over as far as possible, the customs, rites and beliefs of the Brahmins.”

High caste status has historically offered clear economic advantages. The British colonial government assigned government jobs, decided court cases, and allocated resources in ways that preferenced higher ranked castes (Bayly, 2001; Dirks, 2011; Piketty, 2021). Notwithstanding the current affirmative action policies for lower ranked castes, caste-based discrimination continues to offer material benefits to higher ranked castes today (Munshi, 2019). Ethnographic evidence also points to the fact that individuals enjoy the feeling of respect or superiority that comes from being high status (Srinivas, 1966). Likely related to these two points, Brahmins and other higher ranked castes did not historically respond well to the adoption of purity norms on the part of castes ranked below them; for instance, Chakravarti (2018) describes a series of riots in the Tamil Nadu region in 1980 when lower rank caste women attempted to dress in a way that was reserved (prescribed by the Manusmriti) for higher rank women.

Bringing the these features of the caste hierarchy together, D.F. Pocock (1957) summarizes the caste system as a system of “inclusion and exclusion... [in which] each caste tries to include itself with the higher castes and, at the same time, dissociate itself from the lower castes. This process being widespread, the castes above refuse to accept the inclusion just as the castes below try to elbow their way in.”

**Adivasis and the Locally Dominant Caste** Where do Adivasis fit in to the above system of social stratification? Adivasis are heterogeneous in terms of their lineage systems, kinship structures, marriage customs, and other social institutions; in many ways, the key characteristic that unites Adivasi tribes is the *absence* of caste. Consistent with a broad ethnographic literature, our own qualitative data collection shows that Adivasis often consider themselves to be outside the caste system and therefore not subject to purity norms.<sup>17,18</sup> At the same time, as part of society within Indian villages, Adivasis are generally assigned a purity rank. This contradiction is exemplified by Adivasis’ duplicate entry in the 1901 census: the British placed Adivasis within the caste hierarchy—below almost all castes but above Dalits—but simultaneously they also listed Adivasis elsewhere as an “unranked” group outside the caste system (alongside Muslims and other religious minorities).

Qualitative scholars also emphasize that for Adivasis and for other lower-status groups, actual respect and status within a village greatly depends on their dominance, where dominance is measured in terms of land ownership or population share (Srinivas 1966; Dumont 1970). For example, Skoda et al. (2013) in an ethnography of the relationship between Adivasis and Hindus in villages in northern Sambalpur, writes that higher status Hindus “concede” to accept water from groups ranked below them when that group holds a majority of the land. Accepting water from Adivasis and, for Brahmins, acting as spiritual

---

<sup>17</sup>See Xaxa (2008) for a review of ethnographic studies on Adivasi identity and, in particular, on Adivasi attitudes towards gender.

<sup>18</sup>This point is also highlighted in Adivasis’ self-identification submitted to a UN Commission on Indigenous Peoples (1994), which includes the following descriptors: “Relative freedom of women within the society; Absence of division of labour and caste system, [and] Lack of food taboos.”

leaders for Adivasis, strongly contradicts the prescriptions on inter-caste interactions laid out in the Manusmriti. Yet the 1911 census commissioner for Central India observes that “Many Brahmans are willing to serve an aboriginal *of good standing and wealth...* The fact that the tribes are not looked upon by Hindus as impure aids the process” (emphasis our own) (Marten, 1912).

Due to the fact that Adivasis were the first settlers of Central India, Adivasis historically held both considerable land and, in some areas, local political power. We digitize British land records and find that, in 1891, Adivasis served as village headmen in almost a third of villages in Central India. We also find that Adivasi population share continues to be strongly correlated with land ownership share today. It is also predictive of local political power, as measured by the share of local elected officials in unreserved positions who are Adivasi.

### 2.2.2 Conceptual Framework

Taken together, the historical and ethnographic evidence point to a tension in the caste system between, on the one hand, establishing superior purity over other groups and, on the other, aiming to maintain ties with those who have political or economic power (who may or may not be high in purity rank). This section formalizes these insights in a simple conceptual framework of culture choice. Our approach is modeled on Akerlof and Kranton (2000)’s and Shayo (2020)’s endogenous social identity framework, where social groups are defined based on their differences and where the key means by which individuals gain access to a group is through conforming to that group’s norms and behavior. Our framework delivers comparative statics which we use to distinguish between our approach and competing explanations for inter-group cultural transmission.

Consider a village with multiple ethnic groups (e.g. tribes and castes), indexed by  $j = 1, \dots, J$ . Ethnic groups in a village can be ranked along a social hierarchy, where the ordering is fixed, exogenously given, and known by all. In our setting, each ethnic group  $j$ ’s norms and

behavior are summarized by group members' average level of adherence to an index of purity norms, which we denote by  $\bar{x}_j$  and where  $x \in [0, 1]$ . Higher  $\bar{x}_j$  corresponds to higher average adherence to the purity prescriptions described in Section 2.2.1 including, for instance, the seclusion of women inside the home. Membership in an ethnic group comes with material payoffs,  $\pi_j$ . These material payoffs are the result of social and economic interactions with group members and are increasing in a group's size, political power, or economic resources. Because of historical and current caste-based discrimination, higher rank groups have, on average, greater political power and economic resources. Reflecting the fact that adherence to purity norms is costly (for instance, households who adhere to the female seclusion norm lose out on women's potential income from working outside the home), material payoffs are decreasing in  $\bar{x}_j$ . Individuals may also derive intrinsic utility from having a high social hierarchy rank, which is denoted by  $r_j$  (Tajfel and Turner 1979).

Each individual is endowed with a fixed ethnicity, exogenously given at birth. Our starting assumption is that ethnic groups can be ordered in terms of their social distance from Brahmins, the priest caste and highest status Hindu ethnic group. However, there is some scope for social mobility: individuals can seek to assimilate with a group above or below them on the social hierarchy. Assimilation allows individuals to access some of the material or status payoffs associated with a different ethnic group. The degree to which an individual from ethnic group  $i$  gains from assimilation with ethnic group  $j$  will depend on the social distance between  $i$  and  $j$  (Akerlof, 1997; Shayo, 2020). Social distance  $d_{ij}$  is increasing in the distance between own adherence to purity norms and group  $j$ 's prescribed behavior,  $x_{ij} = |x_i - \bar{x}_j|$ .<sup>19</sup> This reflects the basic idea that individuals gain access to a social group by conforming to that group's norms and behavior.  $d_{ij}$  is also increasing in  $r_{ij} = |r_i - r_j|$ , the rank difference between  $i$  and  $j$ . This reflects the inherited, or exogenously given, component of social distance. The idea that individuals can assimilate with other

---

<sup>19</sup>For simplicity of exposition, we assume here that all members of group  $i$  have the same level of adherence to purity norms though this assumption is not required.

social groups by emulating their behavior—but also that assimilation ability is constrained by fixed characteristics of the individual—is consistent with models of assimilation proposed in other settings, such as in the United States where skin color imposes important constraints Fouka et al. (2021).

Bringing these features of culture choice and ethnic identity together, we can represent the utility of an individual from an ethnic group  $i$  who chooses to assimilate with the ethnic group  $j$  in the following way:

$$U_{ij}(x) = U(\pi_j(x), r_j, d_{ij}(x, \bar{x}_j, r_{ij})),$$

where  $U$  is increasing in  $\pi_j$  and  $r_j$  and decreasing in  $d_{ij}$ . Normalize  $U(0) = 0$ .

Let  $x_i^*$  denote the optimal level of purity norm adoption for an individual from ethnic group  $i$ , given their rank  $r_i$  as well as the payoffs and purity norms associated with other ethnic groups in the village. In line with the prescriptions for appropriate behavior put forward by the Manusmriti (Section 2.2.1), the practice of purity norms is traditionally associated with Brahmins, the priest caste and also the highest ranking caste (i.e.  $r_B = 1$ ). In any given village  $v$ , Brahmins’ traditional level of adherence to purity norms can be denoted by  $\bar{x}_{B,v} \in [0, 1]$  and it is known by all. It then follows that adoption of purity norms by individuals from other ethnic groups in that village can be described in the following way:

$$x_{i+1,v}^* \leq x_{i,v}^* \leq \bar{x}_{B,v},$$

where  $x_{i+1,v}^*$  denotes the optimal choice of adherence to purity norms by someone in the ethnic group that is one rank below  $i$  in village  $v$  (where “below” means lower status rank). This follows from the assumption that benefits to assimilation are decreasing in rank distance from a given group. In Section 2.4, we provide evidence in support of this statement.

The framework generates two comparative statistics relevant to our empirical context:

1. *Average level of adherence to purity norms decreases as  $\pi_A$ , the material payoffs associated with an Adivasi ethnic identity, increases.* Adivasis have a lower rank than most Hindu caste groups; an increase in  $\pi_A$  will lead Hindus ranked above them to assimilate towards Adivasis which, following the basic result above, will lead them to lower their adherence to purity norms. We can consider multiple channels through which  $\pi_A$  might increase:

- An increase in Adivasi share of the population. This could lead to higher material payoffs from assimilating with Adivasis because, for instance, it increases the value of informal insurance from Adivasi network ties. As discussed in Section 2.4, Adivasi population is also strongly correlated with Adivasi land ownership share and Adivasi political power. Population share is the primary measure of  $\pi_A$  employed in our empirical analysis.
- A direct increase in Adivasi political power or economic resources, independent of population share. In Section 7, we extend our analysis by testing this channel.
- An increase in the opportunity cost of purity norms adherence due to market- or price-based effects. By this we mean, for instance, that the presence of Adivasis could be associated with a decrease in the cost of meat or an increase in the going wage for women's work. In Section 6, we discuss negative evidence for this channel.

2. *The impact of an increase in  $\pi_A$  on Hindus' adherence to purity norms will have an inverted U-shape with respect to rank.* Hindus of low rank already have low take-up of purity norms so the magnitude of decrease in adherence for this group will be smaller than it would be for those higher up the social hierarchy. At the same time, caste groups with higher ranks have greater rank distance from Adivasis and so have less incentive to change their behavior in response to an increase in  $\pi_A$ .

**Alternative Mechanisms** Our framework assumes that Hindus adopt purity norms as a means of upward social mobility. It is useful to understand how predictions regarding the presence of Adivasis would differ under alternate assumptions of what drives Hindus’ choice of cultural practices. One view proposed by both anthropologists and within the economics literature is that adoption of purity norms reflects deeply held religious or moral beliefs (Dumont, 1970; Jayachandran, 2021). Under this assumption, we would not necessarily expect the presence of Adivasis to shape purity norm practice. A second view, and one that would be consistent with canonical models of norms transmission (Bisin and Verdier, 2010, 2022), is that Hindus and Adivasis select their cultural practices by imitating the most commonly observed behavior amongst their peers. This could explain average convergence to Adivasi cultural practices in Adivasi majority villages, though it would not explain why Adivasis don’t pick up Hindu norms when they are in the minority.<sup>20</sup> Additionally, without imposing further assumptions, it would not predict either the baseline heterogeneity in take-up of purity norms by rank or the heterogeneity in Adivasi share impact by rank.

Finally, we note a distinction between our approach and others in the literature where individuals choose their cultural identity out of a desire to establish difference but where, unlike in our setting, an increase in the population share of a lower-status group is assumed to lead to backlash; or, in other words, to the increased take-up of own-group cultural practices (Bisin et al., 2016). These approaches do not generally take into account the economic costs of maintaining social distance from other groups (Advani and Reich, 2015). In the final section of this paper, we further test this important assumption of our model by turning to a setting where Adivasi share stays constant but where the economic returns to interacting with Adivasis increases.

---

<sup>20</sup>Bisin and Verdier (2010)’s canonical model of horizontal norms transmission assumes that we select behavior based on imitating peers or role models, rather than persons against whom we express animosity. Given the baseline level of discrimination against Adivasis, a prediction of convergence towards Adivasi norms would first require that Adivasis’ presence leads to reduced prejudice against their group (Bisin et al., 2016). “Contact hypothesis” models could deliver this prediction (Allport et al., 1954).

## 2.3 Data and Measurement

Our analysis brings together data from several sources:

**Occupational data from the Socio-Economic and Caste Census** Our primary result, which is that Hindu caste women are substantially more likely to work for pay when they live in villages with Adivasi women, comes from individual-level data from the Government of India’s 2011 Socio-Economic and Caste Census (SECC). Our dataset includes all individuals residing in the states of Madhya Pradesh and Odisha, which together include a population of 133 million people. The SECC survey asks individuals to list their primary occupation, which is recorded as a free-text field. We clean this field and classify women as working if they list their primary occupation as one which involves work outside the home and which is typically paid. In our main analysis, we classify women who report that their primary occupation is “cultivation on own field” as not working, since the social norm which restricts women’s female labor force participation centers on work outside of the home. We restrict our analysis to adult women between the ages of 15 and 65 who live in rural areas.

**Caste purity norms data for Central India from the IHDS Survey** We use the 2011 round of the Indian Household Development Survey (IHDS) to complement the SECC data and generate our main reduced form evidence on Hindu caste households’ take-up of a broad set of purity norms. The IHDS survey includes rich data on households’ cultural practices, occupational outcomes, and other demographic information. We classify cultural practices as purity norms by directly matching the IHDS survey questions to the code of behavior contained in the Manusmriti. As discussed in Section 2.2.1, caste purity norms follow directly from the Manusmriti, a 200 BCE text of religious codes which the British colonial government later adapted and implemented as the basis for legal precedents in India. The Manusmriti contains over 2,000 rules on personal behavior, which are classified according to the following

categories: (a) women’s seclusion; (b) marriage rites; (d) inter-caste interactions; (d) food taboos; and, (e) permitted occupations. We review all survey questions in the IHDS and match these to codes from each of categories (a)-(d). For instance, the Manusmriti directly prohibits widow remarriage among Brahmins: *By twice-born men [Brahmins] a widow must not be appointed to (cohabit with) any other (than her husband); for they who appoint (her) to another (man), will violate the eternal law* (Chapter IX, Verse 64). We match this rule to the IHDS household survey question, “In your community (jati), for a family like yours, is it permissible for a widow to remarry?” We exclude norms pertaining to traditional occupations (category e) from our analysis since these rules are no longer closely followed, especially in rural areas where the majority of households practice agriculture.<sup>21</sup> We also aggregate variables into purity norms categories (female decision-making, female mobility, food taboos, marriage rules, and inter-caste interactions) constructed using standardized indexes per (Kling et al., 2007).

Our conceptual framework predicts that the impact of Adivasi share will differ based on jatis’ initial distance from Brahmins (where distance is proxied by jatis’ status ranking in the 1901 Census). The IHDS survey includes a free-text jati field for all caste and Adivasi respondents. We clean and harmonize jati names,<sup>22</sup> and then hand-match these names to the Caste Rankings from the 1901 Census of India. Using this match, we categorize caste jatis according to whether they were located above or below Adivasi tribes in the rankings.<sup>23</sup>

As discussed in Section 2.2, we focus our analysis on Central India (12,046 survey sample households), which corresponds to region occupied by the former Gondwana Kingdom. Cen-

---

<sup>21</sup>Cassan et al. (2022) find that Hindus are, on average, only 2% more likely than the general population to be employed in the occupation traditionally associated with their caste.

<sup>22</sup>Many castes go by several names and, additionally, some responses to the jati question included a household’s clan (rather than jati) name. We harmonize the list of jatis by referencing the Anthropological Survey of India’s “List of Main Communities” (Singh, 1996), which includes synonyms and clan names for 2,205 jatis across India. We thank Dan Keniston for pointing us to this invaluable resource.

<sup>23</sup>Caste rankings were completed separately for each region; we match districts in our sample to their corresponding 1901 British Province (where there is an imperfect match, we classify districts according to the province which overlaps with the largest portion of district land).

tral India includes the following states: Andhra Pradesh; Chhattisgarh; Gujarat; Jharkhand; Madhya Pradesh; Maharashtra; Odisha; Rajasthan; and, Telengana.<sup>24</sup> We additionally show results in the Appendix for the nationwide sample (27,579 households).

**Spatial Regression Discontinuity Sample Household Survey** We conduct a survey of 874 Hindu men across 143 villages that are located within 20 kilometers distance to either side of the the Mahanadi River (our RDD sample). We interview men about their wife’s labor force participation; their personal beliefs about women’s work and their perception of community beliefs about women’s work; as well as their practice of additional caste purity norms, including food taboos and the untouchability of Adivasis. To shed light on mechanisms, we include a social network module. Motivated by our conceptual framework, we restrict our survey sample to men from “middle-rank” castes: those that rank above Dalits, or “untouchables” and below Brahmins in the caste hierarchy.

Separately, we also interview at least two male elders in each sample village. In that survey, we elicit from elders the purity rank of every ethnic group (caste or tribe) in their village.

**Historical Land and Village Records** In order to validate key identifying assumptions for our spatial regression discontinuity design, we collect and digitize individual land records and village records for Sambalpur district for the period 1915-1955. These land and village records contain detailed information on (1) village inhabitants, including name and ethnic group of every household head; (2) village governance structure, including ethnic group of village headman; (3) features of the agricultural labor and trade markets, including wages and price received by cultivators for their rice. The 1915 records were the first written

---

<sup>24</sup>We focus on Central India because (a) it is home to 64% of the total Adivasi population in the country; (b) many of the primary castes and Adivasi tribes in Central India are located throughout the region (and are not located in other parts of the country); and (c) there is strong overlap between our study area and the region which formed the British Central Provinces, thus local governance and other colonial institutions were fairly homogenous across our sample.

records of land ownership created by the British colonial government for this district. See Appendix D for a sample record.

## 2.4 Social Hierarchy and the Take-up of Caste Purity Norms

Before turning to our main empirical analysis, it is instructive to look at the raw data on average adherence to purity norms among Hindus and Adivasis. Our conceptual framework states that take-up of purity norms among Hindus will be decreasing in caste rank distance to Brahmins, the highest ranking caste. This is because we assume that low-rank caste groups are less likely to be able to successfully assimilate with Brahmins, even given the same investment in purity norms, as groups ranked above them. Thus it is not worthwhile for these lower-rank individuals to invest in costly purity norm behavior. Survey data from the IHDS supports this framework assumption.

In Table 2.1, we show average adherence to purity norms among Adivasis and, separately, among high-, middle-, and low-rank castes. Low-rank refers to Dalits, the group of castes formerly referred to as “untouchables.” High-rank refers to Brahmins and Rajputs, who are everywhere at the top of the caste hierarchy.<sup>25</sup> Middle-rank are all castes in between these two groups; they consist of 75% of the Hindu population in Central India.<sup>26</sup> As explained in Section 2.3, norms questions from the IHDS are matched to specific verses from the Manusmriti and are classified into four broad norms categories that appear in the text: (1) women’s role in society; (2) marriage rules; (3) food taboos; and (4) inter-caste interactions. For category (1), the Manusmriti instructs women to stay secluded within the

---

<sup>25</sup>We include Rajputs, the “king” or “warrior” caste, alongside the Brahmins because the Manusmriti often groups them together under the category of “twice-born” castes.

<sup>26</sup>The Indian government categorizes Hindu castes into the following categories: General, Other Backward Castes, and Scheduled Castes. These rankings are used for affirmative action programs and other administrative purposes, but do not correspond to purity rankings or social relations between castes (Munshi, 2019).

home and states that women should always remain dependent on male family members; thus, we examine IHDS survey questions related to women’s mobility and decision-making. Category (2) includes whether widow remarriage is prohibited, whether dowry is practice, and whether the women had a say in her choice of spouse. For food taboos, category (3), we include whether the household consumes meat, consumes alcohol, or whether the men in the household eat meals before women do (the Manusmriti explicitly instructs men to do this). Lastly, under (4) inter-caste interactions, we analyze whether intercaste marriage is prohibited and whether the household practices untouchability towards Dalits.

In column 1 of Table 2.1, we present the mean of each outcome among Adivasis. In columns 2 through 4, we show the difference in average purity norm adherence between Adivasis and low-rank castes, middle-rank castes, and high-rank castes, respectively. Overall, the summary statistics shown in Table 2.1 highlight the *Sanskritization* that has occurred among all ethnic groups: even among Adivasis, who are outside of the caste system, almost a third report that their community disallows widow remarriage and almost two-thirds report that they practice female veiling (*purdah*). Among Dalits, the castes within the low-rank group, purity norm adherence is similar to that of Adivasis: with the exception of the practice of dowry, the differences between these groups are small in magnitude and most of them are not statistically significant. This is consistent with the hypothesis put forward by our conceptual framework: since Dalits have low inherited rank, and therefore low return to investments in purity norms, their average take-up is low.

Conversely, middle-rank caste groups—who, as we discuss in Section 2.2.1, did not historically practice most purity norms shown in Table 2.1—have today significantly (economically and statistically) higher take-up of purity norms than Adivasis. Middle-rank castes are less likely to report that women have ever worked, are allowed to work, make the decision to work, or are mobile outside the home. These caste groups are also significantly more likely to practice food and alcohol taboos, prohibit widow remarriage, practice dowry, and prac-

tice untouchability against Dalits. The point estimates on high-rank castes (column 4) are almost always larger than those in column 3, consistent with the historical evidence.

We standardize norms variables to the Adivasi mean and standard deviation and present a set of summary indexes in Figure 2.4. This figure shows visually adherence to purity norms is decreasing as Hindus are further in social distance from Brahmins and Rajputs. Taken together, these descriptive statistics are consistent with the idea that Hindus use adherence to purity rules as a means of upwards social mobility—in other words, to assimilate with Brahmins or Rajputs. To understand how the decision to invest in purity norms is influenced by the local social and economic environment, we now turn to our analysis of the impact of Adivasi share on purity norms adherence.

## 2.5 OLS Estimates

Hindu and Adivasi integration in Central India occurred between the 18th and 20th centuries, when Hindus in-migrated to the previously Adivasi-dominated region (see Section 2.2.1). The ethnic composition of villages has remained almost unchanged since that time: 93% of households in the IHDS rural sample for Central India report that their family has been living in their village for over 90 years.<sup>27</sup> We begin our analysis by examining the cross-sectional relationship between Adivasi population share in these villages and Hindus' adherence to purity norms.

### 2.5.1 Empirical Specification

Our empirical analysis of the IHDS survey data estimates two main regression specifications:

$$y_i = \alpha + \phi_1 AdivasiShare + \gamma X_i + \delta Z_v + \theta_s + \epsilon_i \quad (2.1)$$

---

<sup>27</sup>This is true both for Hindu and Adivasi households and does not vary with Adivasi population share.

where  $y_i$  is an individual or household-level outcome for Hindu households; *AdivasiShare* is the Adivasi population share in household  $i$ 's PSU (village);  $X_i$  is a vector of controls for the age and marital status of the respondent as well as the age of the male and female household heads;  $Z_v$  is a vector of village-level controls for distance to urban centers; and  $\theta_s$  is a state fixed effect. Standard errors are clustered at the PSU level.

In Figure 2.A.1 we plot the distribution of *AdivasiShare* across the 648 PSUs that make up our IHDS sample. Adivasis make up 12% of the average village population though over half of villages have no Adivasi presence. Conditional on at least one Adivasi in the village, the average Adivasi share of the village is 30%.

We report results of Specification 2.1 in Panel A of Tables 2.2–2.3 and 2.A.3–2.A.5. In Panel B of these tables, we separately estimate the relationship between *AdivasiShare* and the outcome variable for high-rank castes ( $\gamma_2$ ), middle-rank castes ( $\gamma_1$ ), and low-rank castes ( $\gamma_3$ ) using the following specification:

$$y_i = \alpha + \gamma_1 Middle * AdivasiShare + \gamma_2 High * AdivasiShare + \gamma_3 Low * AdivasiShare + \gamma_4 High + \gamma_5 Low + \beta X_i + \delta Z_v + \theta_s + \epsilon_i \quad (2.2)$$

where *High*, *Middle*, and *Low* are binary indicators for the caste rank group to which the respondent belongs.

## 2.5.2 OLS Results

**Hindu women's work and political participation** In Table 2.2, we present results on work outcomes and norms. On average, 46% of Hindu women who live in a village with no Adivasis report ever having worked. In Panel A of column 1, we see that Hindu women are nearly 50% more likely to have worked as the share of Adivasis in her village increases from

zero to one. In columns 2 and 3, we see that the probability that Adivasi share is positively correlated with whether a woman reports being allowed to work and whether she makes the decision of whether she works or not. In column 4, we see that Adivasi share has substantial consequence for the earnings that a woman reports. When Adivasi share in a village is 1, women report earnings 71% higher earnings.

In Panel B of Table 2.2, we show results by the caste rank of the respondent. Consistent with our model predictions, we see that Middle Rank castes are the most impacted by Adivasi share across a range of work outcomes. In column 1, High status caste women report ever working only 21% of the time when there are no Adivasis in the village and their probability of working is not affected by Adivasi share. Dalit women in no-Adivasi villages are nearly four times as likely as high status caste women to have ever worked, and similarly their probability of working is not impacted by the presence of Adivasis. Lastly, we see that 43% of middle status caste women report ever working. Shifting Adivasi share to 1, increases the probability of work for these women by 27 percentage points, or 63%. Middle status caste women converge to the rate of work of Dalit women in no-Adivasi share villages.

We next examine changes in the political participation of Hindu women (Table 2.A.2). We start by considering the share of women among elected village representatives. We restrict our analysis to seats that have not been reserved for quotas for women or underrepresented groups (scheduled castes and scheduled tribes/Adivasi) in the most recent election. Column 1 shows that the female share among elected village representatives increases with Adivasi share. While this effect might come in part from higher political participation rates of Adivasi women, we also observe an increase in the female share if we restrict the sample to Hindu politicians (column 2). In column 3, we further show that the likelihood that Hindu women discuss politics with their husbands increase by 16% as the Adivasi share in the village increased from zero to one. The results are noisy but significant at the 10% level.

**Purity Norms Outcomes** In Table 2.3, we report the relationship between Adivasi share and a broad set of norms outcomes. As discussed in Section 2.3, the broad norms categories match the four norms categories outlined in the Manusmriti (women’s seclusion, food taboos, marriage restrictions, and restrictions on intercaste relations) and questions from the IHDS are matched to Manusmriti verses. Each question within the index is normalized and aggregated to generate each broad norms index. In Panel A, we normalize the indexes to the outcomes of no-Adivasi share villages. In Panel B, we normalize to the outcomes of each caste group separately.

Looking at Panel A, we see that across all outcomes, a higher Adivasi share is associated with lower norms compliance by caste households, all significant at at least the 5% level. In column 1, we see that a shift from a no Adivasi PSU to one with an Adivasi share of 1 is associated with a 27 percentage points decline in the practice of purdah (female veiling). Hindu caste women are also more mobile and more likely to be involved in household decision-making in villages with higher Adivasi shares. The components of the food taboos index are presented in Appendix Table 2.A.5 and the components of the marriage and intercaste restrictions indexes are in Appendix Table 2.A.4. Adivasi share of 1 is associated with a nearly 50% drop in the likelihood that the household is vegetarian, a 20% decrease in the likelihood that the household consumes alcohol, and a 61% decrease in the probability that the men in the household eat before the women. All estimates are significant at at least the 5% level. Turning to marriage and intercaste relations restrictions, we see that the coefficient on Adivasi share is always large and negative, but is statistically significant only for whether widow remarriage is prohibited in the household’s caste in the village, in the likelihood that the woman was not consulted in the selection of her spouse, and importantly in untouchability of Adivasis. When Adivasi share is 1, caste households are 0.1 percentage point less likely to say they do not accept water from Adivasis, compared to a mean of 0.14 in villages with 0 Adivasi share. So shifting a caste household from a no-Adivasi share village

to the average village with at least one Adivasi is associated with caste households being 21% less likely to practice untouchability.

In Panel B, we disaggregate the effect of Adivasi share by caste grouping. Consistent with Table 2.2, middle status castes are significantly less likely to practice purity norms in the presence of higher Adivasi PSU share. For female veiling and the (inverted) mobility index, the coefficients are large and negative and we can reject at the 1% level that the effects for high and middle status castes are the same. In the case of Dalits, the effect of Adivasi share is always negative on the norms indexes and significant for all outcomes except the intercase relations index. For food, the coefficients are half as large as for the middle status group, but we cannot reject equality for any of the outcomes. Overall, we take the results to be suggestively consistent with the predictions of the model.

### 2.5.3 Robustness and Validity Checks

Hindus and Adivasis did not settle randomly and, as such, there is the concern that the Adivasi-share effects may be endogenous to other features of the local social or economic environment. First, there is the possibility that differential selection in- or out- of Adivasi villages drives our results. However, we believe this to be unlikely: 93% of households in the IHDS rural sample for Central India report that their family has been in their current village for at least 90 years (the maximum possible response). For most castes, adoption of purity norms regarding marriage rules, women's role in society, and food taboos has only occurred within the past 100 years (Figure 3). Given this, there is not a strong reason to believe that these caste groups would have selectively migrated in to Adivasi villages based on their own purity norm preferences.

A second potential concern is that unobservable factors common to Adivasi villages, such as differential access to markets, may have direct effects on Hindus' practice of purity norms. In order for such an unobservable factor to explain our results, it would have to be something

that differentially affects middle-rank Hindus versus Adivasis and high- and low-rank Hindus. Standard economic explanations (such as remoteness or differential infrastructure) could not easily explain, for instance, why middle-rank Hindu FLFP increases with Adivasi share but that Adivasi FLFP and high- and low-rank FLFP are unaffected. Moreover, we find impacts of Adivasi presence across a wide range of Hindu cultural practices. This makes it less likely than an unobservable feature of the economic environment drives our findings. In Appendix Tables 2.A.6 and 2.A.7, we further address this concern by including village fixed effects in our estimation of the impact of Adivasi share on Hindu adherence to purity norms. Though effects are noisily estimated, we continue to find impacts on middle caste-rank women’s ability to work and on middle caste-rank households’ adherence to marriage restrictions, food taboos, and limits on inter-caste interactions. This suggests that unobservable factors common to Adivasi villages could not fully explain the impacts we observe and provides further evidence for the argument that Adivasi presence has a direct impact on Hindus’ adherence to purity norms.

Finally, we provide evidence on the role of norms transmission by exploiting heterogeneity in gender attitudes *among* Adivasi groups. An important distinction among Adivasi groups is that while some groups historically practiced plough agriculture, other groups historically practiced shifting agriculture.<sup>28</sup> Ethnographic and empirical evidence shows that societies which formerly practiced shifting agriculture — a method of cultivation which employs more female labor — have higher rates of women’s work and more equal gender norms today (Boserup, 1970; Alesina et al., 2013). We therefore hypothesize that Adivasi groups which historically practiced shifting agriculture should have higher rates of female labor force participation; and, as such, norms transmission to caste groups should be stronger in these villages.

---

<sup>28</sup>All Adivasi groups in Sambalpur now exclusively practice plough agriculture.

The IHDS asks households to list the name of their caste or tribe. We match these tribe names to a dataset which contains information on each tribe’s historical agricultural practices, which we created using ethnographic data on Adivasi groups. In Table 2.8 we compare village-level characteristics of villages in our sample. We regress the village level characteristic on the share of plough Adivasis in the village and on the share of non-plough Adivasis in the village. while villages with Adivasis are different than non-Adivasi villages, we can’t reject equality of coefficients on non-plough and plough Adivasi share.

Table 2.A.8 shows the results of the analysis on female labor force participation. Consistent with the hypothesis outlined above, we find that Adivasi women who belong to Adivasi groups which used to practice shifting agriculture are 11.5 percentage points more likely to work outside the home for pay today (Panel A). Panel B shows the result of a regression of Hindu women’s work participation on the share of the village population that belongs to a shifting or plough agriculture Adivasi group. We find that the increase in caste women’s likelihood of working is driven entirely by historically shifting agriculture Adivasi groups. The sample for Table 5 includes Adivasi women (Panel A) and Hindu women (Panel B) across Central India. This finding is also replicated in Figure 2.A.2, which shows a local polynomial regression of Hindu FLFP on Adivasi share, separating Adivasi groups into those that practiced plough versus shifting agriculture.

## 2.6 Spatial Regression Discontinuity Estimates

Despite the robustness of our OLS estimates and the fact that we observe Adivasi-share impacts across a range of purity norm outcomes, there remains the concern that there are unobservables that may be biasing our estimates. For example, remoteness to urban centers may differentially affect middle caste rank Hindus. Given this possibility, we also undertake an alternative estimation strategy: we exploit a historical natural experiment that led to variation in village-level Adivasi share among villages within a narrow geographic area. In

this section, we first discuss the relevant historical context; then our estimation strategy and identification assumptions; finally, we present results of our RD analysis.

### **2.6.1 The Railroad and In-Migration to Sambalpur, Odisha**

Our causal analysis exploits variation in Adivasi share due to the in-migration of Hindu households following the 1894 construction of a railroad stop in Sambalpur, Odisha.<sup>29</sup> The district of Sambalpur was taken over by the British colonial government in 1849 upon the death of the former Raj.<sup>30</sup> Sambalpur, which sits on the eastern edge of the former Gondwana Kingdom, is situated over 300 kilometers from Cuttack, the coastal city and capital of British Orissa at that time.<sup>31</sup> Prior to the opening of the railroad, the district was isolated with minimal trade or in- or out-migration to neighboring areas (Nethersole 1874; 1885). The existing trade route was the Mahanadi River, which bisects Sambalpur from north to south and flows to Cuttack and other coastal cities. Rice was and continues to be the dominant crop in Sambalpur.

The population of Sambalpur even prior to the railroad construction was multi-ethnic: among the 19 largest landowners in 1889, 12 were Gond, 2 belonged to a different Adivasi tribe, and the remaining 5 belonged to various Hindu cultivator castes (Russell 1889). The railroad construction in 1894 led to an additional surge of Hindu migrants from neighboring districts: In the words of the British Settlement Officer at the time, the construction of the railroad stop in Sambalpur “doubled the price of rice and changed the run of trade [from downstream the Mahanadi to northwards along the railroad]” which brought “an invasion of Hindu cultivators” to what had previously been a predominantly tribal area (Dewar, 1908).

---

<sup>29</sup>We use the term Sambalpur district to refer to the region of that name as it was defined in the British Colonial period. This region was later subdivided and now includes Bargarh, Jharsuguda, Nuapada and Sambalpur districts in the state of Odisha and parts of Mahasamund and Raigarh districts in the neighboring state of Chhattisgarh.

<sup>30</sup>The British colonial government at the time had a policy of annexing native (non-British controlled) states if the ruler died without an heir. See Iyer (2010) for further discussion.

<sup>31</sup>The former state of Orissa was renamed as Odisha in 2011.

The Sambalpur stop was the terminus of a branch line from the main Bengal-Nagpur line (Figure 2). While the districts directly to the north, northwest, and northeast of Sambalpur were served by the main line, the station in Sambalpur City became the nearest railroad access point for not only Sambalpur district, but also the states directly to the southwest (Baud, Patna, and Sonpur), and southeast of Sambalpur (Athmallik and Rairakhol). These states (shaded in gray in Figure 2 below) became the main sending areas for Hindu migrants to the areas around the Sambalpur railroad station.

**The Mahanadi River Boundary** Our treatment area of interest is the set of villages close to the Sambalpur railroad station, located on either side of the Mahanadi River. This key feature of Sambalpur geography restricted the settlement locations of new Hindu migrants: the river, which is almost a mile wide in parts, constrained population movement between the east and west sides.<sup>32</sup> This is because, at the time of the railroad station construction, the river only had one crossing in Sambalpur district: a pontoon bridge located at Sambalpur City, directly at the site of the new railroad station (Dewar, 1908; Deloche, 1993). Migrants, who arrived in the area after walking with their families and belongings from areas up to 200 kilometers away, settled in the plains surrounding the railroad station where they found land available (Dewar, 1908; O'Malley, 1909). Cultivable land was abundant; in 1889, only 56% of total land in the district was cultivated or under other use (Russell 1889). In general, cultivators did not bring their rice paddy directly to the railroad station; rather, they brought their paddy on head-loads down village paths to traders stationed along the primary district road, which ran from west to east through the district (Figure 2). Thus the difference in market accessibility for a cultivator located, for instance, along the district road 10 kilometers from the river on the east versus the west side was marginal: the only difference was the

---

<sup>32</sup>The description of population settlement after the introduction of the railroad draws heavily from the 1906 Sambalpur British Land Settlement Report, which includes a detailed accounting of population movement and market access during the prior decade.

need for a trader who purchased on the west side to additionally transport the paddy across the pontoon bridge.

That migrants tended to settle on the side of the river from which they had come is also borne out in the historical land records data. We digitize individual land records from 1915 for a sample of 204 villages in our RD sample area in Sambalpur. These land records contain the name and ethnicity of every household in a village at that time. We use ethnicity names to determine whether a household was a likely migrant from the period after the introduction of the railroad in 1894. Most Hindu caste ethnicities present in Sambalpur prior to the railroad were also present in the neighboring states; however, we identify two Hindu caste groups *not* present in Sambalpur prior to the railroad but that were present in neighboring states. The *Bhulia* caste is described as being a dominant group in Sonpur state, to the southwest of Sambalpur while the *Chasa* caste is described as a primary ethnic group in Rairakhol state, to the southeast. Our digitization of the 1915 land records shows that Bhulia caste individuals held land in villages to the west but not east of the Mahanadi River in 1915; the reverse is true for individuals from the Chasa caste.

Our empirical analysis will exploit the discontinuous change in Adivasi population share at the Mahanadi River boundary due to higher levels of in-migration of Hindus on the west versus east sides of the river. The historical record documents several causes of this difference in the magnitude of migration. The first and primary reason is differences in geography: to the east and southeast of Sambalpur, at the border with Rairakhol state, there is a hill range covered in dense forests. The hill range is described at the time as being “almost impassable,”<sup>33</sup> and, beyond it, the states of Rairakhol and Athmallik were both very isolated and sparsely populated. At the time of the railroad construction, Athmallik was home to only 56 persons per square mile and Rairakhol to 32 persons per square mile. They were two of the three least densely populated states in Orissa (Cobden-Ramsay, 1910). In contrast, the

---

<sup>33</sup>Orissa Historical Research Journal, vol I., 1952.

fertile and open plains that characterize the villages on either side of the Mahanadi River near to the railroad station continued westwards and southwards into Baud, Patna, and Sonpur states. These three states had population densities of 70, 116, and 188 persons per square mile, respectively. In other words, there was both a larger population of would-be migrants in areas to the west and southwest and the travel required for these migrants was less costly.<sup>34</sup> The second reason for the larger number of migrants from the west and southwest is that there was poor weather and a resulting famine in the area in 1899-1900. Patna and Sonpur states were especially badly affected, leading to an additional surge of migrants towards Sambalpur (which was less badly hit) in that year. The British Settlement Officer described the situation thusly: “crowds of aimless wanderers [from Patna and Sonpur]...flocked to the relative Eden that was Sambalpur” (Dewar, 1908). Though village-level census population data is not available for this period, the 1906 British Land Settlement Report for Sambalpur notes that only four villages in the district grew sufficiently over the decade from 1891-1901 to surpass a population of 2,000 individuals: these four villages are all located to the west of the Mahanadi River (Dewar, 1908).

Prior to the 1894 influx of Hindu migrants, the British colonial government had mapped the full Sambalpur region and divided it into village administrative units, each with their own hereditary headman. Thus new migrants to the Sambalpur area settled into existing villages, expanding the area under cultivation. In addition to the village headman, who collected taxes from individual households and acted as a liaison with the British government, each village had a watchman who was in charge of resolving disputes and reporting crime to the British. This structure of local governance was common to villages on the east and west side of the river (Dewar 1906; Hamid 1921).

---

<sup>34</sup>The historical record and our own qualitative interviews with village elders in our RD sample areas confirm that some migrants from Rairakhol and Athmallik did, in fact, traverse the forested areas to settle in Sambalpur post-construction of the railroad. Our aim here is only to highlight the difference in relative magnitude of migrants.

### 2.6.2 Regression Discontinuity Empirical Specification

We exploit the discontinuous change in Adivasi share of the population on either side of the Mahanadi River boundary to estimate a causal effect of Adivasi share on our outcomes of interest. Our empirical specification is the following:

$$y_{i,v} = \alpha + \gamma East + f(\text{location}_v) + \beta X_{i,v} + \epsilon_{i,v} \quad (2.3)$$

where  $y_{i,v}$  is the outcome of interest for individual  $i$  in village  $v$  and  $East$  is an indicator variable equal to 1 if the village is on the eastern side of the Mahanadi River boundary and zero otherwise.  $f(\text{location}_v)$  is the RD polynomial, which controls for smooth functions of geographic location for village  $v$ .  $X_i$  is a vector of covariates for individual  $i$ , which include age, survey date fixed effects, and enumerator fixed effects.

Following Calonico et al. (2014) and Gelman and Imbens (2019), our baseline specification is a local linear polynomial in distance to the river estimated separately on the west and east sides. Our preferred specification will use a fixed bandwidth of 20 km to the left and right of the river and a triangular kernel. We also calculate the optimal bandwidth using Calonico et al. (2020) mean-squared-error minimizing procedure. Our results are highly robust to using a variety of alternate bandwidths (Appendix Figure 2.A.3) and kernels (Appendix Figure 2.A.4). We further report Conley standard errors with a cut-off window of 10 kilometers to account for spatial auto-correlation (Conley, 1999) as well as p-values computed using a randomization exercise in which we randomly re-assign distance to the river (Appendix Tables 2.A.9 and 2.A.10).

Our coefficient of interest is  $\gamma$ , which can be interpreted as the impact on a Hindu household of residing in a village just to the east side of the Mahanadi River. The key identifying assumption for the regression discontinuity approach presented in Specification 2.3 is that all relevant factors besides treatment vary smoothly at the river boundary. This

assumption is needed for Hindu individuals located just across the river on the (lower Adivasi share) western side to be appropriate counterfactuals for Hindu individuals on the (higher Adivasi share) eastern side. The intuition behind our specification is that, absent the influx of Hindu migrants in 1894-1900 to the western side of the river, villages directly to the east and west of the river would have similar levels of Adivasi share of the population. They should also be comparable in terms of geography, historical governance institutions, and traditional norms of caste households. This allows us to identify the impact of the average difference in Adivasi share between villages east and west of the river on Hindu women's labor force participation and purity norms among Hindu households.

To assess the plausibility of our identifying assumption, Panel A of Table 2.9 examines a variety of geographic characteristics. Agriculture is the primary source of employment for both men and women in villages in our sample area and rice is the primary crop. Table 2.9 shows that villages on either side of the river boundary are balanced in terms of elevation, flow accumulation, rice suitability, and forest share. We do, however, find that villages to the east side of the river have steeper slope, which may indirectly affect crop output. In Appendix Figure 2.A.4, we replicate our main analysis controlling for geographic characteristics and find that estimates are consistent.

In Section 2.6.1, we show historical evidence that villages to the east and west side of the river were similar in terms of ethnic composition and other relevant factors prior to the introduction of the railroad. To further validate this assumption, we collect and digitize individual land records from 1887 for households in our regression discontinuity sample. These land records contain information on each household's caste or tribe, as well as their total landholdings. Columns 1 and 2 of Panel B in Table 2.9 show that villages are balanced in terms of Adivasi population share and Adivasi land-ownership share prior to the introduction of the railroad.<sup>35</sup> In Column 3 of Panel B in Table 2.9, we further utilize the land records

---

<sup>35</sup>We are able to find complete land records for 117 of our sample of 142 villages. The share of missing records is not differential on the east versus west side of the river.

to show that villages on either side of the river were balanced on an important measure of Adivasi political and economic power prior to the 1894 in-migration shock: we find that 15.4% of villages had Adivasi village headman, and that this share does not differ in villages east versus west of the river.

A second possible threat to identification is that the construction of the railroad could have differentially impacted households on either side of the river. For instance, villages on the eastern side, where the railroad is located, may have better access to agricultural markets. However, as we discuss in Section 2.6.1, the railroad station is directly at the river boundary. Thus except for the distance of the actual crossing of the river (there was a pontoon bridge at the railroad station crossing), households on either side of the river were equidistant to the railway station. To further validate the assumption of equal access to markets, we are also in the process of digitizing 1961 village notes completed by government officials that contain information on price of rice paid to cultivators in each village at that time, along with other measures of local agricultural markets. A preliminary analysis of records for 40 sample villages shows that the historical price of rice was similar on both sides of the river. As a last point of evidence for this assumption, we examine whether agricultural output and prices vary across the river in the present day. Column 2 of Table 2.A.11 shows that the revenue received by rice farmers in villages east of the river is not significantly different from the price of rice received by farmers west of the river. Farmers on both sides of the river are equally likely to hire women (Column 1 of Table 2.A.11) and, consistent with the fact that female labor supply is higher in villages east of the river, we find that the wage for female labor is 19% lower in east-of-river villages (Column 4). We find that farmers east of the river report 17% higher output per acre, potentially as a result of an increase in the intensive margin of female labor usage or because cost savings on labor are put towards other inputs.

Our third identification assumption is no selective sorting of incoming caste migrants across the river boundary. In Section 2.6.1, we provide historical evidence that migrants

arrived primarily from regions to the south-west of our study area and that the presence of the river made it costly for them to settle in the east: there was only one river crossing in the district (at Sambalpur City) and crossing was via a pontoon bridge or ferry, both of which had fees. However, it is possible that certain caste migrants, such as those with a greater affinity for Adivasi people or those who had more equal gender norms, were more likely to be willing to pay the higher effort and monetary cost to settle on the eastern side of the river. We cannot rule out this possibility, but we note that this would have been an unlikely scenario: at the time of caste households' migration, adherence to purity norms among non-Brahmin castes was low.

### 2.6.3 Regression Discontinuity Results

We now turn to the results of our regression discontinuity analysis. Table 2.5 uses the Socio-Economic and Caste Census from 2011 and limits the analysis to our sample villages. In column 1, we show the village-level Adivasi share and in column 2 we show caste women's labor force participation. Consistent with the historical evidence discussed in Section 2.6.1, villages just east of the river have roughly 25 percentage points higher Adivasi share. On average, villages west of the river have 25% Adivasi share, so being east of the river is associated with a 100% increase in the fraction of Adivasis in the average village. In column 2 we see that caste women in villages directly east of the river are 10 percentage points more likely to report that their primary occupation is one which involves work outside the home for pay.

Figure 2.5 presents standard RD plots for Adivasi share and Hindu FLFP, with distance to the river as the running variable and a local linear trend to each side of the discontinuity. We observe there is a clear discontinuity at the river point in terms of Adivasi share. With one exception, the average Adivasi share in each distance bin is always lower on the west side of the river. In the second panel, we see that on both sides of the river, being closer to the

river is associated with higher caste FLFP. Again, we observe the discontinuous jump at the river boundary and, within each distance bin, the share of women who work on the east side of the river is always higher than on the west side. In Table 2.A.3 we show results varying the bandwidth and in Table 2.A.4 we show results using alternate kernels (e.g. uniform and Epanechnikov instead of triangular); results remain consistent with those shown in Table 2.5.

### **Evidence on the Transmission of Gender Norms from Adivasi to Hindu Households**

We now turn to our survey data to investigate impacts on norms. In columns 3-8 in Table 2.5, we present results on women's labor force participation as well as norms related to women's work. We asked respondents to report the primary occupation of their wives. On average, Hindu women's primary occupation on the west side of the river is worker 34% of the time. This is about 9 percentage points higher than the estimate in the SECC. In column 3 of Table 2.5, we see a 0.14 percentage point (or a 43%) reduction in the probability that a Hindu woman is a housewife. The point estimate is statistically indistinguishable from the point estimate in column 2 using the SECC data. In column 4 we also see that caste women are almost twice as likely to have worked outside the home for pay in the previous 30 days.

We asked husbands to report their own and community beliefs about the appropriateness and desirability of women's work. In columns 5 - 6, the outcomes are own beliefs. In columns 7 - 8 we ask respondents to provide second-order beliefs - what they believe other households in their caste in their village think. In columns 5 and 7, we see that respondents on the east side are more likely to report whether they think it is appropriate for women to work even if even the household does not have a financial need and that the community believes it is appropriate for women to work, although only the latter is significant (at the 10% level). To understand the role of working women in the community, we presented respondents with

a vignette that asked them to imagine hypothetically choosing one of two wives for their marriage-age son. Both women come from the same caste as the respondent and have exactly the same financial circumstances as each other, but one wife prefers to stay at home and only look after domestic duties while the other prefers to work outside the home for pay. Respondents on the east side of the river are 35% less likely to prefer a daughter-in-law who wants to be a housewife. Consistently they report that other households in their caste in their village are less likely to prefer the housewife to the working woman.

We further observe a meaningful upending of other purity norms, including restrictions regarding caste interactions and adherence to food taboos, on the east side of the river. In column 1 of Table 2.6, we note that nearly two-thirds of caste households west of the river would not touch water that an Adivasi has touched - a practice akin to perceiving an Adivasi as “less than human”.<sup>36</sup> The practice was officially abolished in the Indian Constitution and the Indian Parliament has passed a variety of laws since independence (the latest in 2018) to define forbidden practices as well as punishments for violations. Nonetheless, between 1991 and 2020, there were over 700,000 instances of violence due to untouchability, including the rape of 38,000 women.<sup>37</sup> Strikingly, caste households east of the river are 46% less likely to practice untouchability towards Adivasis and 7% less likely to practice untouchability towards Dalits (columns 1 and 2).

Perhaps the most surprising result indicates a shift not only in the personal practice of purity norms, but in the conception of social stratification itself. We gave respondents a list of all the castes and Adivasi groups that were represented in their village and asked them to rank the list in terms of social status in their community. In column 3 we show a dummy variable that indicates that the respondent ranks Adivasis as highly or more highly than a

---

<sup>36</sup>National Geographic, 2003

<sup>37</sup>Statistics for Dalits. <https://indianexpress.com/article/opinion/columns/in-the-death-of-a-dalit-student-betrayal-of-the-vision-of-freedom-8093920/>

middle status caste.<sup>38</sup> Caste respondents east of the river are almost three times as likely to rank Adivasis at least as highly in the social hierarchy as a middle status group. Consistent with an upending of purity norms, we further observe that caste households east of the river adhere less to food taboos (column 4).

Our results are robust to various changes to the specification, including alternative bandwidths (Appendix Figure 2.A.3), alternative kernels, second-order RD polynomials, the inclusion of geographic controls (Appendix Figure 2.A.4), and accounting for spatial correlation (Appendix Table 2.A.9). We also calculate standard errors for each of our main outcomes based on randomization inference by randomly re-assigning the distance to the river for each sample village 1,000 times (Dell et al., 2018). For each iteration, we then repeat our analysis based on the re-assigned indicator for whether the village is located on the east side of the river. The p-value is calculated as the share of the placebo coefficients that are larger in magnitude than the actual coefficient on the east side dummy (in absolute terms). Appendix Table 2.A.10 shows that our results also hold in this exercise.

### **Alternative Explanations for the Relationship between Adivasi Share and Hindu**

**Women’s Work** We argue that higher Adivasi share leads to an increase in caste women’s supply of labor through decreasing the social stigma surrounding women’s work for caste households. We now evaluate alternative labor demand- or supply-based explanations for the observed relationship:

The presence of Adivasis in a village community may increase landowners’ demand for female agricultural labor through increasing either the real or perceived returns to hiring women. For instance, fixed costs to hiring women may lead landowners to only hire male laborers when the pool of female laborers is small. If this were the case, we should also see that the rate of Adivasi women’s work also increases with Adivasi share. But as shown in

---

<sup>38</sup>We define the variable to specifically mean that an Adivasi group ranks as highly or higher than an Other Backwards Castes (OBC) caste group.

Figure 2.2, Adivasi female labor force participation does not change with their population share in the community. The presence of Adivasis in their village community may instead increase labor demand through changing some landowners' perception of the benefits of hiring female labor; in other words, Adivasi share may lead to a shift in attitudes among owners, not laborers. This is unlikely since agricultural labor has traditionally been and continues to be highly gender-segregated in our setting: certain tasks, such as transplanting, are typically done by women while other tasks, such as ploughing, are typically done by men. In Table 2.A.11, we cannot reject that landowners on the east and west side are equally likely to hire women (column 1) — the coefficient is noisy, but implies a lower likelihood of hiring on the east side. Moreover, nearly all landowners on both sides of the river express no preference in the caste of the workers that they hire.

Our norms-based explanation for the observed increase in Hindu women's work is one of several possible reasons why Hindu women's labor supply might increase as a result of living in a community with Adivasi women. First, Hindu women may be more willing to work outside the home because, through exposure to Adivasi women, they learn and update their beliefs about the costs or benefits of working. A majority of women in our study sample area perform agricultural work on their own land; thus, learning would have to be associated with gains specific to agricultural work on someone else's land. In our survey, we measure possible learning through questions on the following topics: beliefs regarding safety of women working outside the home; knowledge of the daily wage for female labor; and, knowledge of landowners in the village who hire female agricultural workers.

Second, Hindu FLFP may increase with Adivasi share because of complementarities in caste women's labor supply function: for instance, Hindu women may have a stronger preference for safety than Adivasi women. If Hindu women are only willing to work as laborers when there is a sufficient number of other women working, then we would expect their labor supply to increase with Adivasi share. To assess these possible complementarities,

we ask respondents from our RD sample survey about safety and preferences for number of women with whom they would work, as well as availability of childcare.

In Table 2.A.11, we present a variety of measures related to these two potential labor supply stories. In columns 3 and 4, we see that households east of the river are just as well informed as those west of the river about paid agricultural work opportunities for women as well as current female agricultural daily wages; nearly all respondents in the sample report being informed about both of these outcomes. In column 5, we see that women are equally likely to have access to childcare across both sides of the river. Nearly all respondents on both the west and east side express no preference for the caste of the hiring landowner, the caste of workers for women, and believe that it is safe for women to engage in paid agricultural opportunities in their village.

## 2.7 Mechanisms

Our conceptual framework argues that Hindus will be less likely to adopt purity norms when it is costly to exclude Adivasis from economic and social interactions. In Table 2.7, we analyze the quantity and quality of Hindu-Adivasi interactions in high Adivasi share villages (east of river) within our RD sample.

We asked respondents to tell us about their social interactions with members of their own caste, with Adivasis, and with Brahmins and members of other middle status castes. Specifically, by each of these three subgroups, we ask how many total friends they have, how many they socialize with frequently, how many they ask for personal or business advice, how many could they ask for a loan in case of an emergency, and how many would provide “manpower assistance” (assistance in, for example, setting up for a wedding or another type of event). In Table 2.7, we analyze each of the questions by group: Panel A shows for own caste, Panel B for Adivasis, and Panel C shows Brahmins or other middle and upper rank castes.

In columns 1 - 3 of Table 2.7, we show standardized indexes that aggregate outcomes by reference caste group. First, in column 1 we see that Hindu respondents east of the river have very similar social interactions within their own caste group. Looking at Panel A of Table 2.7, we see no large or any significant differences in the number of friends listed in each category. We do, however, see that respondents east of the river have 0.65 standard deviation unit more interactions with Adivasis than respondents west of the river. In Panel B of Table 2.7, we see that respondents east of the river report at least double the number of Adivasi friends in total, as well as those they socialize with, ask advice from, and could ask for financial or manpower assistant. Additional social interactions with Adivasis appear to substitute interactions with Brahmins or members of other middle status castes. Across all outcomes in Table 2.7, we cannot reject that coefficients in Panel B are the same magnitude as those in Panel C.

We conclude our analysis by turning to a distinct setting, where Adivasis are assigned more political and economic power, independent of their population share. In three districts (Feudatory States) of Odisha, the British colonial government assigned the position of village watchman to the village's Adivasi priest (*Jhankar*). In other districts, the status quo was instead to assign position based on merit. The historical record suggests that the decision was ad-hoc, and a response to needing to quickly put in place an administrative structure after death of the former Raj (Russell 1885; Maddox 1901; O'Malley 1909; Hamid 1921; Pati 1986). Importantly, the watchman position carried significant economic benefits: the watchman was given land on which he did not have to pay taxes and was also privy to tax income from all other village households (Central Provinces Land Act, 1882). In addition, the watchman was in charge of resolving village disputes and relaying information back to the British colonial government. Several sources document the political power and social capital of the *jhankar* in watchmen districts (Dewar 1906; O'Malley 1909; Cobden-Ramsey 1911; Hamid 1921). The watchman position was abolished post-independence, but the 1962 Orissa

Land Reform Act allowed *jhankar* households to keep the land that had been historically allocated (Board of Revenue 1962). We therefore argue that the political and economic power of Adivasis was shifted in watchman districts.

To analyze how this shift in Adivasi power affected Hindu households, we create a sample of the blocks (administrative unit below district but above Panchayat) that border one another on either side of the district border. In Table A.13, we compare these blocks on a variety of geographic characteristics. Overall, we do not find significant differences between watchman and non-watchman bordering blocks. Importantly, the Adivasi population share in watchman blocks is statistically equivalent to that across the block border. That being said, we caution against a causal interpretation since there may be unobservable confounders.

For analysis of the impact of increased Adivasi power, we estimate the following specification:

$$y_i = \alpha + \phi_1 WatchmanBlock + \theta_c + \epsilon_i \quad (2.4)$$

where  $y_i$  is an individual or household-level outcome for Hindu caste households. *WatchmanBlock* is an indicator for whether the block was a watchman block.  $\theta_c$  is a caste fixed effect.

The results are presented in Table A14. We see that first watchman blocks are, in present day, 30% more likely to have an Adivasi in the position of *jhankhar*. So despite the abolition of the watchman position after independence, there appears to be persistence in the political empowerment of Adivasis. In columns 2 - 5 we then show a variety of purity norms practices by Hindu caste households. Hindu caste households are 17 percentage points more likely to have had an Adivasi priest officiate their wedding and are 7 percentage points less likely to practice untouchability. They are 8% more likely to state that it is appropriate for women to work and almost 20% less likely to practice meat taboos.

Overall, results are consistent with the argument in our conceptual framework that, in settings where out-groups have substantive economic or political power, Hindus caste groups are less likely to adopt norms that create social separations from the out-group.

## 2.8 Discussion

In this paper, we examine how the presence of an out-group in their community impacts Hindus' adherence to caste purity norms. Using a number of estimation strategies, including a historical natural experiment, we show that caste norms weaken when Hindus in central India live in villages with Adivasis, an indigenous group outside the caste system. We find that increased Adivasi population share is associated with decreased adherence to caste purity norms regarding women's seclusion and role in society, food taboos, and inter-caste interactions. When they have more Adivasi neighbors, Hindu women are more likely to work and have substantially higher earnings. They are also more empowered along a number of dimensions. When less constrained by caste norms, Hindus have stronger social and economic relationships both with Adivasis and with Hindus of low caste rank. We conclude from these findings that the presence of Adivasis leads to reduced social stratification within mixed Adivasi-Hindu communities.

One implication of our results is that policies that give additional political or economic power to Adivasis or other marginalized groups, such as Dalits, could not only lead to reduced discrimination against those groups, but may also change caste norms among the broader Hindu population. More broadly, by demonstrating how adoption of caste purity norms that are costly to women is at least partly an economic and social choice, as well as a religious one, we highlight how reductions in India's female labor force participation can, potentially, be corrected by policy.

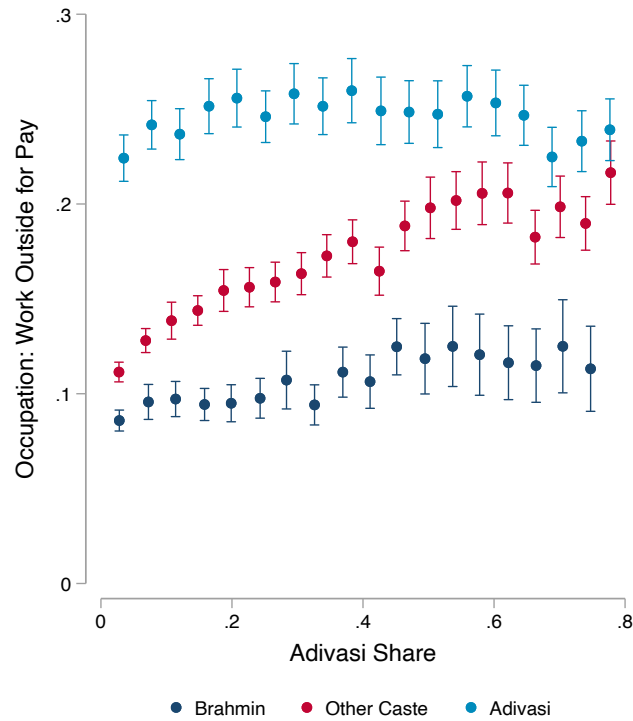
## 2.9 Tables and Figures

Figure 2.1: Sambalpur District in 1906



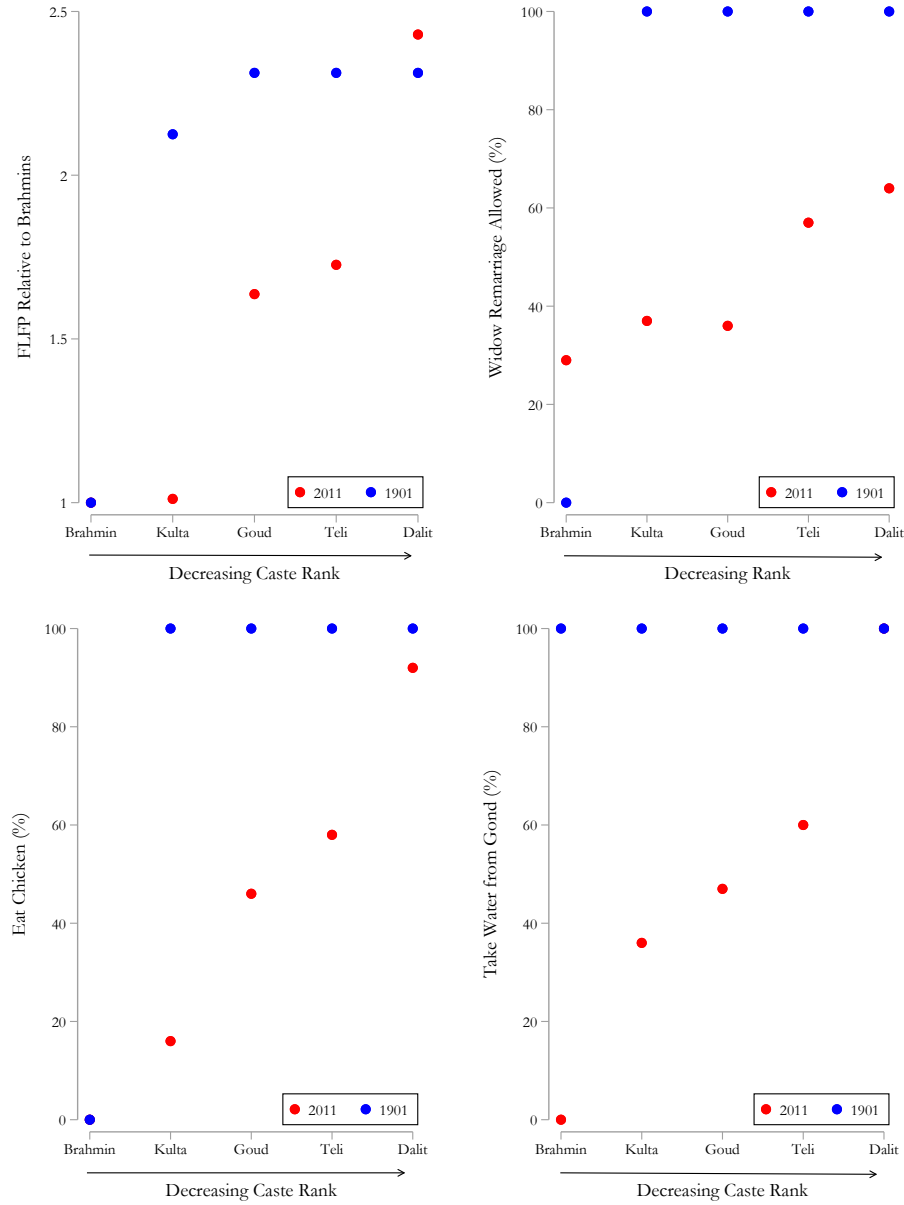
Notes: The Mahanadi River is shaded in blue and the red arrows represent the direction of migration.

Figure 2.2: Adivasi Population Share and Women's Work Outside the Home



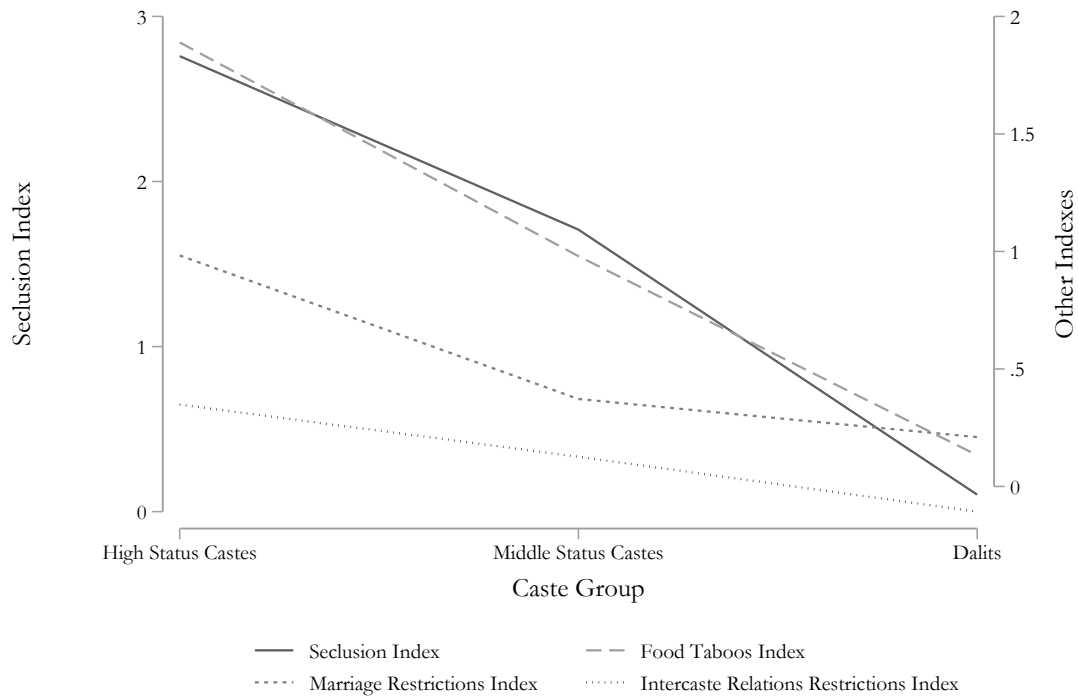
Notes: The figure is constructed using individual-level data on caste women aged 25-64 years from the 2011 Socio-Economic and Caste Census. The outcome on the y-axis indicates whether the Hindu woman worked based on the classification of a free-text occupation question (not including work on own farm). The x-axis represents the share of individuals who are classified as scheduled tribes in the village based on the 2011 population census. The figure plots binned means separately for Brahmins, other caste households, and Adivasis.

Figure 2.3: Purity Norms Adoption in the 20th and 21st Centuries by Caste



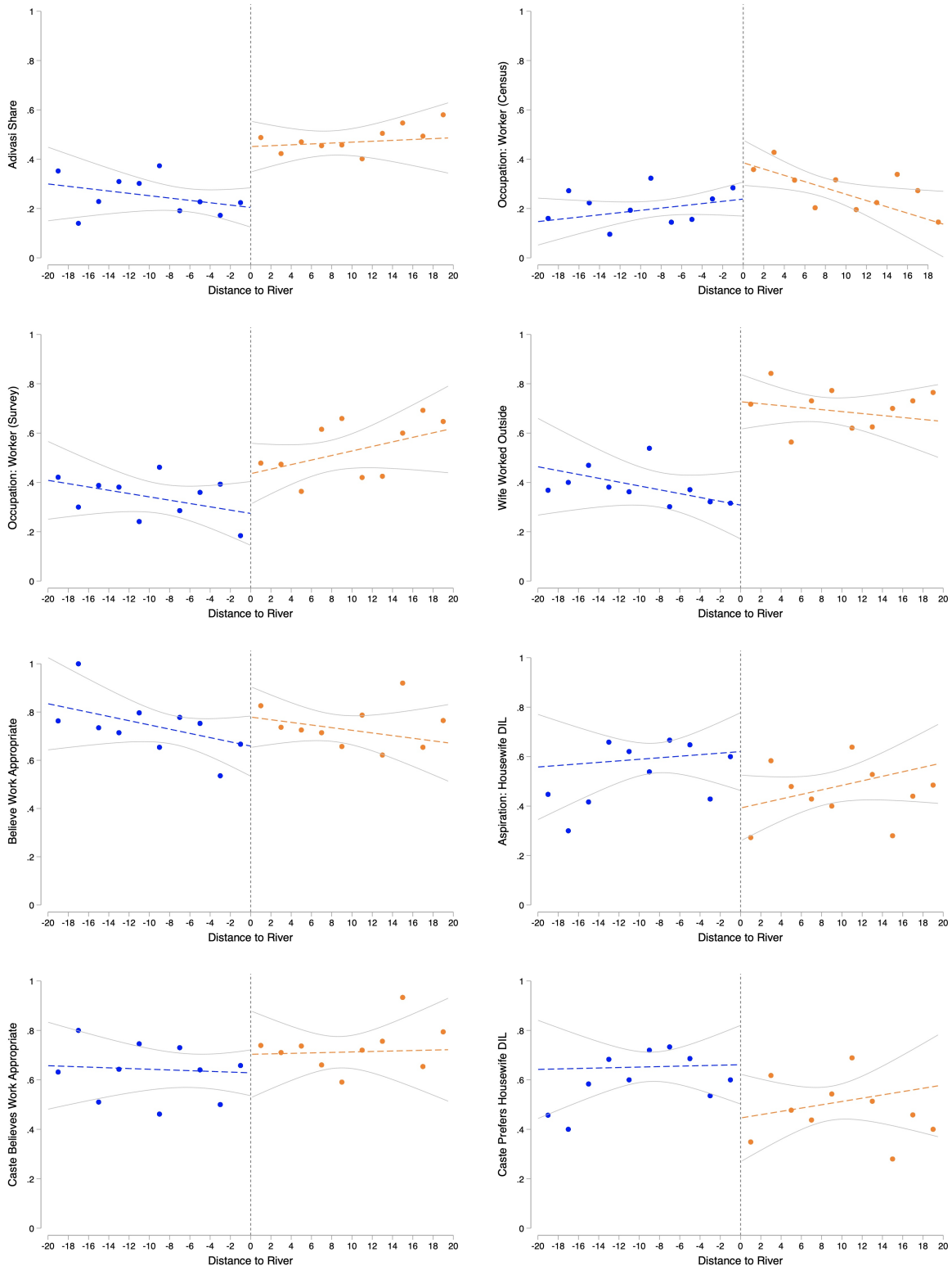
Notes: The figure plots the adoption of purity norms against caste rank over time. The caste ranking is from the 1901 British Census of India. The female labor force participation data for 1901 is obtained from the 1901 Census of India and for 2011 is obtained from the 2011 Socio-Economic Caste Census. The historical purity norms data is obtained from: The 1885-1889, 1906, and 1921 Reports on the Land Settlement of the Sambalpur District; Correspondence on the Revised Land Settlement Report for the Years 1900-1910 for the Sambalpur District; and The 1909 and 1915 Imperial Gazetteers for the Sambalpur District. Historical data is validated against contemporaneous ethnographic evidence, see Data Appendix for details. The present-day purity norms outcomes are obtained from primary survey data collection in Odisha (see Section 4).

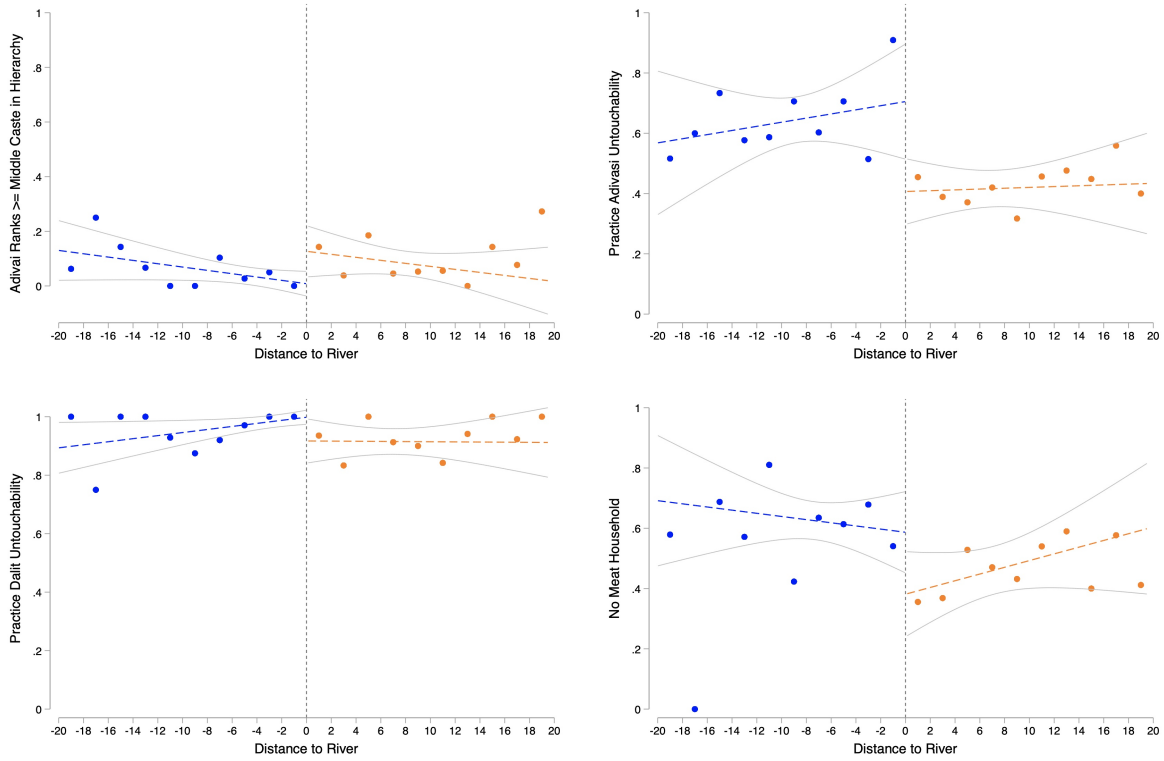
Figure 2.4: IHDS Adoption of Purity Norms by Caste Group



Notes: The figure is constructed using data from IHDS (2011). The sample includes all caste households in central states: Andhra Pradesh, Chhattisgarh, Gujarat, Jharkhand, Madhya Pradesh, Maharashtra, Odisha, Rajasthan, and Telengana. High rank castes consist of Brahmins and Rajputs, and middle rank castes consist of all other castes besides Dalits. See Appendix Tables 2.A.3, 2.A.4, and 2.A.5 for index components. The seclusion index combines the practice of purdah with the inverted decision-making index.

Figure 2.5: RDD Plots





Notes: The figures show RD plots for our main outcomes. Each dot represents the outcome mean within a 2km bin. The local linear trends are estimated separately on each side of the river using a triangular kernel. Standard errors are clustered at the village level. The range plots correspond to 95 percent confidence intervals. The sample consists of 143 villages on either side of the Mahanadi River. The first figure shows village-level Adivasi share based on the 2011 population census. The second figure shows female labor force participation rates based on individual-level data from the Government of India's 2011 Socio-Economic and Caste Census. The remaining outcomes are obtained from our own surveys with Hindu caste men and village elders. See Tables 2.5 and 2.6 for outcome definitions.

Table 2.1: Average Purity Norms by Caste Group

|                                      | (1)<br>Adivasi<br>Mean | (2)<br>Low Rank<br>Difference | (3)<br>Middle Rank<br>Difference | (4)<br>High Rank<br>Difference |
|--------------------------------------|------------------------|-------------------------------|----------------------------------|--------------------------------|
| <b>Panel A: Seclusion</b>            |                        |                               |                                  |                                |
| Ever Worked                          | 0.76<br>[0.42]         | -0.03*<br>(0.02)              | -0.29***<br>(0.01)               | -0.52***<br>(0.02)             |
| Allowed to Work                      | 0.87<br>[0.34]         | 0.00<br>(0.01)                | -0.16***<br>(0.01)               | -0.18***<br>(0.02)             |
| Own Decision to Work                 | 0.55<br>[0.50]         | -0.06***<br>(0.02)            | -0.14***<br>(0.01)               | -0.17***<br>(0.03)             |
| Earnings in Past Year                | 5885.67<br>[12955.90]  | 1785.65***<br>(434.88)        | -2232.75***<br>(236.56)          | -4234.69***<br>(536.08)        |
| Practice Purdah                      | 0.63<br>[0.48]         | 0.05***<br>(0.02)             | 0.11***<br>(0.01)                | 0.30***<br>(0.02)              |
| Can Attend Panchayat Meeting         | 0.09<br>[0.28]         | 0.05***<br>(0.01)             | -0.00<br>(0.01)                  | -0.04***<br>(0.01)             |
| Can Visit Friend Alone               | 0.85<br>[0.35]         | -0.06***<br>(0.01)            | -0.07***<br>(0.01)               | -0.10***<br>(0.02)             |
| Can Take Public Transport Alone      | 0.48<br>[0.50]         | -0.02<br>(0.02)               | -0.01<br>(0.01)                  | 0.00<br>(0.02)                 |
| <b>Panel B: Food</b>                 |                        |                               |                                  |                                |
| Men Eat First                        | 0.31<br>[0.46]         | -0.02<br>(0.02)               | 0.02<br>(0.01)                   | 0.12***<br>(0.02)              |
| Alcohol Taboo                        | 0.47<br>[0.47]         | 0.04*<br>(0.02)               | 0.23***<br>(0.01)                | 0.29***<br>(0.03)              |
| Meat Taboo                           | 0.13<br>[0.34]         | 0.05***<br>(0.01)             | 0.26***<br>(0.01)                | 0.51***<br>(0.02)              |
| <b>Panel C: Marriage</b>             |                        |                               |                                  |                                |
| Widow Remarriage Prohibited          | 0.32<br>[0.46]         | -0.03**<br>(0.02)             | 0.05***<br>(0.01)                | 0.18***<br>(0.02)              |
| Practice Dowry                       | 0.64<br>[0.48]         | 0.13***<br>(0.02)             | 0.15***<br>(0.01)                | 0.11***<br>(0.02)              |
| No Choice in Spouse                  | 0.56<br>[0.50]         | 0.02<br>(0.02)                | -0.01<br>(0.01)                  | 0.11***<br>(0.02)              |
| <b>Panel D: Intercaste Relations</b> |                        |                               |                                  |                                |
| Intercaste Marriage Prohibited       | 0.74<br>[0.44]         | 0.00<br>(0.02)                | 0.02*<br>(0.01)                  | 0.12***<br>(0.02)              |
| Practice Untouchability              | 0.08<br>[0.28]         | -0.04***<br>(0.01)            | 0.06***<br>(0.01)                | 0.23***<br>(0.02)              |
| N                                    | 1468                   | 1552                          | 8087                             | 752                            |

Notes: The data is obtained from IHDS (2011). The sample includes all caste households in central states: Andhra Pradesh, Chhattisgarh, Gujarat, Jharkhand, Madhya Pradesh, Maharashtra, Odisha, Rajasthan, and Telengana. Low rank castes consist of dalits, high rank castes consist of Brahmins and Rajputs, and middle rank castes consist of all other castes. Columns 2-4 report coefficients from a joint regression. Robust standard errors are shown in parentheses. \* Significant at the 10 percent level, \*\* Significant at the 5 percent level, \*\*\* Significant at the 1 percent level.

Table 2.2: Work IHDS

|                                          | (1)<br>Ever<br>Worked | (2)<br>Allowed<br>to Work | (3)<br>Own Decision<br>to Work | (4)<br>Earnings in<br>Past Year |
|------------------------------------------|-----------------------|---------------------------|--------------------------------|---------------------------------|
| <b>Panel A: Pooled</b>                   |                       |                           |                                |                                 |
| Adivasi Share                            | 0.220***<br>(0.059)   | 0.147***<br>(0.053)       | 0.283***<br>(0.077)            | 2687.388***<br>(672.717)        |
| Mean if Adivasi Share=0                  | 0.46<br>[0.50]        | 0.71<br>[0.45]            | 0.38<br>[0.49]                 | 3764.53<br>[7887.74]            |
| <b>Panel B: By Caste Rank</b>            |                       |                           |                                |                                 |
| $\gamma_1$ : Adivasi Share * Middle Rank | 0.268***<br>(0.062)   | 0.194***<br>(0.054)       | 0.342***<br>(0.081)            | 3096.969***<br>(721.990)        |
| $\gamma_2$ : Adivasi Share * High Rank   | 0.045<br>(0.140)      | -0.360*<br>(0.214)        | 0.030<br>(0.185)               | -508.798<br>(1252.949)          |
| $\gamma_3$ : Adivasi Share * Low Rank    | 0.056<br>(0.101)      | 0.118<br>(0.089)          | 0.069<br>(0.158)               | 1715.324<br>(1729.018)          |
| <i>P-value from F-Test</i>               |                       |                           |                                |                                 |
| $\gamma_1 = \gamma_2$                    | 0.129                 | 0.011                     | 0.104                          | 0.012                           |
| $\gamma_1 = \gamma_3$                    | 0.044                 | 0.389                     | 0.084                          | 0.449                           |
| Mean if Adivasi Share=0 and:             |                       |                           |                                |                                 |
| Middle Rank                              | 0.43<br>[0.49]        | 0.68<br>[0.47]            | 0.36<br>[0.48]                 | 3321.57<br>[7679.22]            |
| High Rank                                | 0.21<br>[0.41]        | 0.69<br>[0.46]            | 0.38<br>[0.49]                 | 1417.08<br>[5367.77]            |
| Low Rank                                 | 0.74<br>[0.44]        | 0.87<br>[0.34]            | 0.49<br>[0.50]                 | 7315.65<br>[8940.29]            |
| N                                        | 8774                  | 8422                      | 7753                           | 8785                            |
| PSU                                      | 648                   | 648                       | 648                            | 648                             |

Notes: The data is obtained from IHDS (2011) The sample consists of Hindu caste women in central states: Andhra Pradesh, Chhattisgarh, Gujarat, Jharkhand, Madhya Pradesh, Maharashtra, Odisha, Rajasthan, and Telengana. Low rank castes consist of dalits, high rank castes consist of Brahmins and Rajputs, and middle rank castes consist of all other castes. All regressions include state fixed effects and control for the distance to the nearest town. Panel B reports separate coefficients for each caste group. Standard errors clustered at the PSU (village) level. \* Significant at the 10 percent level, \*\* Significant at the 5 percent level, \*\*\* Significant at the 1 percent level.

Table 2.3: Purity Norms Indexes

|                                          | (1)                  | (2)                                      | (3)                           | (4)                    | (5)                               | (6)                              |
|------------------------------------------|----------------------|------------------------------------------|-------------------------------|------------------------|-----------------------------------|----------------------------------|
|                                          | Practices<br>Purdah  | Inverted<br>Decision-<br>Making<br>Index | Inverted<br>Mobility<br>Index | Food<br>Taboo<br>Index | Marriage<br>Restrictions<br>Index | Intercaste<br>Relations<br>Index |
| <b>Panel A: Pooled</b>                   |                      |                                          |                               |                        |                                   |                                  |
| Adivasi Share                            | -0.265***<br>(0.075) | -2.126**<br>(0.878)                      | -0.823***<br>(0.225)          | -0.921***<br>(0.176)   | -0.799***<br>(0.243)              | -0.213**<br>(0.093)              |
| <b>Panel B: By Caste Rank</b>            |                      |                                          |                               |                        |                                   |                                  |
| $\gamma_1$ : Adivasi Share * Middle Rank | -0.275***<br>(0.081) | -2.223**<br>(0.930)                      | -1.033***<br>(0.235)          | -1.070***<br>(0.198)   | -0.915***<br>(0.264)              | -0.225**<br>(0.105)              |
| $\gamma_2$ : Adivasi Share * High Rank   | 0.115<br>(0.080)     | -0.654<br>(2.772)                        | 1.402**<br>(0.619)            | -0.179<br>(0.691)      | 0.737<br>(0.487)                  | -0.024<br>(0.200)                |
| $\gamma_3$ : Adivasi Share * Low Rank    | -0.403**<br>(0.166)  | -2.437*<br>(1.261)                       | -0.742*<br>(0.397)            | -0.570*<br>(0.315)     | -1.143***<br>(0.407)              | -0.186<br>(0.214)                |
| <i>P-value from F-Test</i>               |                      |                                          |                               |                        |                                   |                                  |
| $\gamma_1 = \gamma_2$                    | 0.001                | 0.571                                    | 0.000                         | 0.210                  | 0.002                             | 0.362                            |
| $\gamma_1 = \gamma_3$                    | 0.442                | 0.873                                    | 0.478                         | 0.162                  | 0.592                             | 0.859                            |
| N                                        | 8782                 | 8779                                     | 8784                          | 10391                  | 10391                             | 10391                            |
| PSU                                      | 648                  | 648                                      | 648                           | 648                    | 648                               | 648                              |

Notes: The data is obtained from IHDS (2011). The sample includes all caste households in central states: Andhra Pradesh, Chhattisgarh, Gujarat, Jharkhand, Madhya Pradesh, Maharashtra, Odisha, Rajasthan, and Telengana. Low rank castes consist of dalits, high rank castes consist of Brahmins and Rajputs, and middle rank castes consist of all other castes. All regressions include state fixed effects. The outcome in column 1 indicates whether a Hindu caste woman practices purdah (female veiling). See Appendix Tables 2.A.3, 2.A.4, and 2.A.5 for treatment effects on index components. Panel B reports separate coefficients for each caste group. Standard errors are clustered at the PSU (village) level. \* Significant at the 10 percent level, \*\* Significant at the 5 percent level, \*\*\* Significant at the 1 percent level.

Table 2.4: RDD Balance Checks

**Panel A: Geographic Characteristics**

|                        | (1)<br>Elevation    | (2)<br>Slope       | (3)<br>Log Flow<br>Accumulation | (4)<br>Rice<br>Suitability | (5)<br>Forest<br>Share |
|------------------------|---------------------|--------------------|---------------------------------|----------------------------|------------------------|
| East                   | 4.366<br>(3.369)    | 0.445**<br>(0.177) | 0.914<br>(0.686)                | 0.344<br>(0.286)           | 0.026<br>(0.028)       |
| Mean for West of River | 178.693<br>[41.582] | 0.589<br>[1.106]   | 2.063<br>[1.408]                | 1.448<br>[0.455]           | 0.042<br>[0.069]       |
| N                      | 142                 | 142                | 142                             | 142                        | 142                    |
| Villages               | 142                 | 142                | 142                             | 142                        | 142                    |

**Panel B: Distribution of Hindu Households by Caste Rank**

|                        | (1)<br>High Rank<br>Caste<br>Share | (2)<br>Middle Rank<br>Caste<br>Share | (3)<br>Low Rank<br>Caste<br>Share |
|------------------------|------------------------------------|--------------------------------------|-----------------------------------|
| East                   | 0.020<br>(0.038)                   | -0.084<br>(0.062)                    | 0.064<br>(0.066)                  |
| Mean for West of River | 0.070<br>[0.077]                   | 0.515<br>[0.204]                     | 0.415<br>[0.198]                  |
| N                      | 141                                | 141                                  | 141                               |
| Villages               | 141                                | 141                                  | 141                               |

**Panel C: Historical Village Characteristics**

|                        | (1)<br>1887<br>Adivasi<br>Population Share | (2)<br>1887<br>Adivasi<br>Land Share | (3)<br>1887<br>Village<br>Headman<br>is Adivasi |
|------------------------|--------------------------------------------|--------------------------------------|-------------------------------------------------|
| East                   | 0.001<br>(0.046)                           | 0.025<br>(0.054)                     | -0.030<br>(0.062)                               |
| Mean for West of River | 0.454<br>[0.250]                           | 0.371<br>[0.246]                     | 0.154<br>[0.364]                                |
| Villages               | 117                                        | 113                                  | 125                                             |

Notes: All regressions are based on a local linear specification estimated separately on each side of the river boundary with a triangular kernel and 20km bandwidth. Standard errors clustered at the village level. \* Significant at the 10 percent level, \*\* Significant at the 5 percent level, \*\*\* Significant at the 1 percent level.

Table 2.5: First Stage, Work Outcomes, and Beliefs

|                        | Census Data         |                       | Survey Data           |                           |                                |                                 |                                          |                                      |
|------------------------|---------------------|-----------------------|-----------------------|---------------------------|--------------------------------|---------------------------------|------------------------------------------|--------------------------------------|
|                        |                     |                       | FLFP                  |                           | Own Beliefs                    |                                 | Community Beliefs                        |                                      |
|                        | (1)                 | (2)                   | (3)                   | (4)                       | (5)                            | (6)                             | (7)                                      | (8)                                  |
|                        | Adivasi<br>Share    | Occupation:<br>Worker | Occupation:<br>Worker | Wife<br>Worked<br>Outside | Believe<br>Work<br>Appropriate | Aspiration:<br>Housewife<br>DIL | Caste<br>Believes<br>Work<br>Appropriate | Caste<br>Prefers<br>Housewife<br>DIL |
| East                   | 0.248***<br>(0.065) | 0.098*<br>(0.052)     | 0.143**<br>(0.072)    | 0.375***<br>(0.082)       | 0.092<br>(0.068)               | -0.199***<br>(0.075)            | 0.111*<br>(0.059)                        | -0.166**<br>(0.078)                  |
| Mean for West of River | 0.250<br>[0.178]    | 0.205<br>[0.404]      | 0.336<br>[0.473]      | 0.374<br>[0.484]          | 0.739<br>[0.440]               | 0.571<br>[0.495]                | 0.638<br>[0.481]                         | 0.630<br>[0.483]                     |
| N                      | 142                 | 26,283                | 856                   | 856                       | 813                            | 798                             | 861                                      | 773                                  |
| Villages               | 142                 | 141                   | 142                   | 142                       | 142                            | 142                             | 142                                      | 142                                  |

Notes: All regressions are based on a local linear specification estimated separately on each side of the river boundary with a triangular kernel and 20km bandwidth. Column 1 uses village-level data from the 2011 population census. Column 2 uses individual-level data on caste women aged 25-64 years from the 2011 Socio-Economic and Caste Census and controls for age, marital status, and whether the household is classified as scheduled caste. Columns 3-8 use our own survey data with Hindu caste men and controls for age, enumerator fixed effects, and survey time fixed effects. The outcome in column 1 is the share of individuals who are classified as scheduled tribes in the village. The outcome in column 2 indicates whether the Hindu woman worked based on the classification of a free-text occupation question (not including work on own farm). The outcome in column 3 is an indicator variable that is equal to one if the primary occupation of the respondent's wife is not housewife. The outcome in column 4 indicates whether the respondent's wife did one of the following activities at least once in the past year: agricultural work for pay on someone else's land, self-employment, non-agricultural daily labor, or salaried work. The outcome in column 5 indicates whether the respondent believes that it is appropriate for a Hindu woman to work outside, even if not financially constrained. The outcome in column 6 indicates whether the respondent replied 'wife who wants to work for pay' to the following vignette: 'assume you had a son of marriageable age and you could choose between two wives for your son. Both wives are from your jati and have the same education and same financial status. However, only one of them wants to work outside for pay. Which wife would you prefer for your son?'. The outcomes in columns 7-8 are equivalent to the outcomes in columns 5-6, but instead of asking about the respondent's own beliefs, we ask the respondent about what other households in the village believe. Appendix Figure 2.A.3 shows results for different bandwidths and Appendix Figure 2.A.4 shows results for optimal bandwidths, different kernels, second-order RD polynomials, and geographic controls. Appendix Table 2.A.9 reports Conley standard errors that adjust for spatial correlation and Appendix Table 2.A.10 reports p-values based on randomization inference. Standard errors clustered at the village level. \* Significant at the 10 percent level, \*\* Significant at the 5 percent level, \*\*\* Significant at the 1 percent level.

Table 2.6: Inter caste Purity Norms and Food Taboos

|                        | Untouchability                               |                                            |                                                             | Food Taboo                  |
|------------------------|----------------------------------------------|--------------------------------------------|-------------------------------------------------------------|-----------------------------|
|                        | (1)<br>Practice<br>Adivasi<br>Untouchability | (2)<br>Practice<br>Dalit<br>Untouchability | (3)<br>Adivasi Ranks $\geq$<br>Middle Caste<br>in Hierarchy | (4)<br>No Meat<br>Household |
| East                   | -0.298***<br>(0.110)                         | -0.081**<br>(0.040)                        | 0.118**<br>(0.052)                                          | -0.186**<br>(0.091)         |
| Mean for West of River | 0.643<br>[0.480]                             | 0.959<br>[0.199]                           | 0.061<br>[0.240]                                            | 0.616<br>[0.487]            |
| N                      | 835                                          | 361                                        | 375                                                         | 848                         |
| Villages               | 143                                          | 143                                        | 143                                                         | 143                         |

Notes: East of River corresponds to higher Adivasi population share. All regressions are based on a local linear specification estimated separately on each side of the river boundary with a triangular kernel and 20km bandwidth. Column 1 pools data from our own survey data with village elders and with Hindu caste men. Columns 2-3 use our own survey data with village elders. Column 4 uses our own survey data with Hindu caste men and controls for age, enumerator fixed effects, and survey time fixed effects. The outcome in column 1 indicates whether the respondent accepts water from at least one Adivasi group in the village. The outcome in column 2 indicates whether the respondent accepts water from Dalits in the village. The outcome in column 3 indicates whether there is at least one Adivasi group in the village that ranks at least as high as the lowest middle (OBC) caste group in the village. The outcome in column 4 indicates whether the respondent's wife eats chicken. Appendix Figure 2.A.3 shows results for different bandwidths and Appendix Figure 2.A.4 shows results for optimal bandwidths, different kernels, second-order RD polynomials, and geographic controls. Appendix Table 2.A.9 reports Conley standard errors that adjust for spatial correlation and Appendix Table 2.A.10 reports p-values based on randomization inference. Standard errors clustered at the village level. \* Significant at the 10 percent level, \*\* Significant at the 5 percent level, \*\*\* Significant at the 1 percent level.

Table 2.7: Social Interactions

|                                              | (1)                 | (2)                  | (3)                  | (4)                      | (5)                     |
|----------------------------------------------|---------------------|----------------------|----------------------|--------------------------|-------------------------|
|                                              | Total Friends       | Any Socialize        | Any Advice           | Any Financial Assistance | Any Manpower Assistance |
| <b>Panel A: Own Caste</b>                    |                     |                      |                      |                          |                         |
| East                                         | -0.106<br>(0.376)   | -0.026<br>(0.039)    | -0.072<br>(0.050)    | 0.048<br>(0.079)         | -0.026<br>(0.046)       |
| Mean for West of River                       | 2.756<br>[1.879]    | 0.963<br>[0.189]     | 0.942<br>[0.235]     | 0.455<br>[0.499]         | 0.963<br>[0.189]        |
| <b>Panel B: Adivasi</b>                      |                     |                      |                      |                          |                         |
| East                                         | 0.829***<br>(0.236) | 0.360***<br>(0.117)  | 0.363***<br>(0.105)  | 0.152***<br>(0.058)      | 0.361***<br>(0.116)     |
| Mean for West of River                       | 0.602<br>[0.860]    | 0.366<br>[0.482]     | 0.255<br>[0.436]     | 0.064<br>[0.244]         | 0.342<br>[0.475]        |
| <b>Panel C: Upper or Other Middle Castes</b> |                     |                      |                      |                          |                         |
| East                                         | -0.557**<br>(0.273) | -0.388***<br>(0.113) | -0.362***<br>(0.110) | -0.116<br>(0.073)        | -0.385***<br>(0.101)    |
| Mean for West of River                       | 1.432<br>[1.192]    | 0.557<br>[0.497]     | 0.493<br>[0.501]     | 0.146<br>[0.353]         | 0.576<br>[0.495]        |
| N                                            | 712                 | 712                  | 712                  | 712                      | 712                     |
| Villages                                     | 143                 | 143                  | 143                  | 143                      | 143                     |

Notes: East of River corresponds to higher Adivasi population share. All regressions are based on a local linear specification estimated separately on each side of the river boundary with a triangular kernel and 20km bandwidth. All columns use our own survey data with Hindu caste men and controls for age, enumerator fixed effects, and survey time fixed effects. The outcomes are obtained from a social network exercise based on the following questions: ‘How many families are there in this village whose house you visit frequently or with whom you socialize frequently?’, ‘How many families are there in this village who give you important advice, such as advice on farming, health, or financial issues?’, ‘In your village, how many families are there who you could borrow from without interest in case of a medical emergency?’, and ‘If you needed manpower assistance such as, for instance, if there was a wedding in your household, from how many families in this village could you ask for unpaid assistance?’. For each person listed in the module, we then elicited the name and caste of the person. Column 1 reports the total number of unique friends. The outcomes in columns 2-4 indicate whether the respondent has at least one friend in the respective categories. Panel A restricts the responses to friends from the respondent’s own caste. Panel B restricts the responses to Adivasi friends. Panel C restricts the responses to friends other upper and middle rank castes. Standard errors clustered at the village level. \* Significant at the 10 percent level, \*\* Significant at the 5 percent level, \*\*\* Significant at the 1 percent level.

Table 2.8: Balance Check: Comparison of Villages Where Adivasis Historically Practiced Plough vs. Non-Plough Agriculture

|                                                                         | Distance to<br>Nearest<br>Town<br>(1) | Distance to<br>District<br>HQ<br>(2) | Any<br>Private<br>Primary<br>School<br>(3) | No<br>Drainage<br>System<br>(4) | Share HHs<br>Open<br>Defecate<br>(5) | Share HHs<br>Electricity<br>(6) | Share HHs<br>Mobile<br>Phone<br>(7) |
|-------------------------------------------------------------------------|---------------------------------------|--------------------------------------|--------------------------------------------|---------------------------------|--------------------------------------|---------------------------------|-------------------------------------|
| Traditionally Non-Plough Adivasi Share                                  | 4.493<br>(4.948)                      | 8.954<br>(13.054)                    | -0.565***<br>(0.189)                       | 0.071<br>(0.200)                | -0.037<br>(0.097)                    | -0.135*<br>(0.081)              | 0.018<br>(0.167)                    |
| Traditionally Plough Adivasi Share                                      | 0.223<br>(8.949)                      | 7.616<br>(23.706)                    | -0.226<br>(0.370)                          | 0.082<br>(0.359)                | 0.124<br>(0.134)                     | -0.170<br>(0.126)               | -0.301<br>(0.206)                   |
| p-value: Trad. Non-Plough Adivasi Share =<br>Trad. Plough Adivasi Share | 0.618                                 | 0.953                                | 0.346                                      | 0.978                           | 0.358                                | 0.823                           | 0.307                               |
| Outcome Mean for Adivasi Share = 0                                      | 14.938                                | 51.124                               | 0.403                                      | 0.496                           | 0.812                                | 0.781                           | 1.311                               |
| Observations                                                            | 198                                   | 198                                  | 198                                        | 198                             | 198                                  | 198                             | 198                                 |

Notes: Sample includes all villages in central states: Orissa, Madhya Pradesh, Chhattisgarh, Jharkhand, Maharashtra, Gujarat, Rajasthan, Andhra Pradesh, Telengana. All regressions include state fixed effects and standard errors clustered at PSU (village) level.

Table 2.9: Impacts of Historical Adivasi Agricultural Practices on Current Gender Norms

**Panel A: Impacts on Adivasi Women**

|                                               | Ever<br>Worked     | (Inverted)<br>Seclusion<br>Index | (Inverted)<br>Marriage<br>Restrictions Index |
|-----------------------------------------------|--------------------|----------------------------------|----------------------------------------------|
|                                               | (1)                | (2)                              | (3)                                          |
| Woman is Traditionally Non-Plough Adivasi     | 0.107**<br>(0.045) | 0.317**<br>(0.157)               | 0.304*<br>(0.158)                            |
| Outcome Mean for Traditionally Plough Adivasi | 0.802              | 0.247                            | -0.017                                       |
| N                                             | 483                | 483                              | 483                                          |
| Villages                                      | 82                 | 82                               | 82                                           |

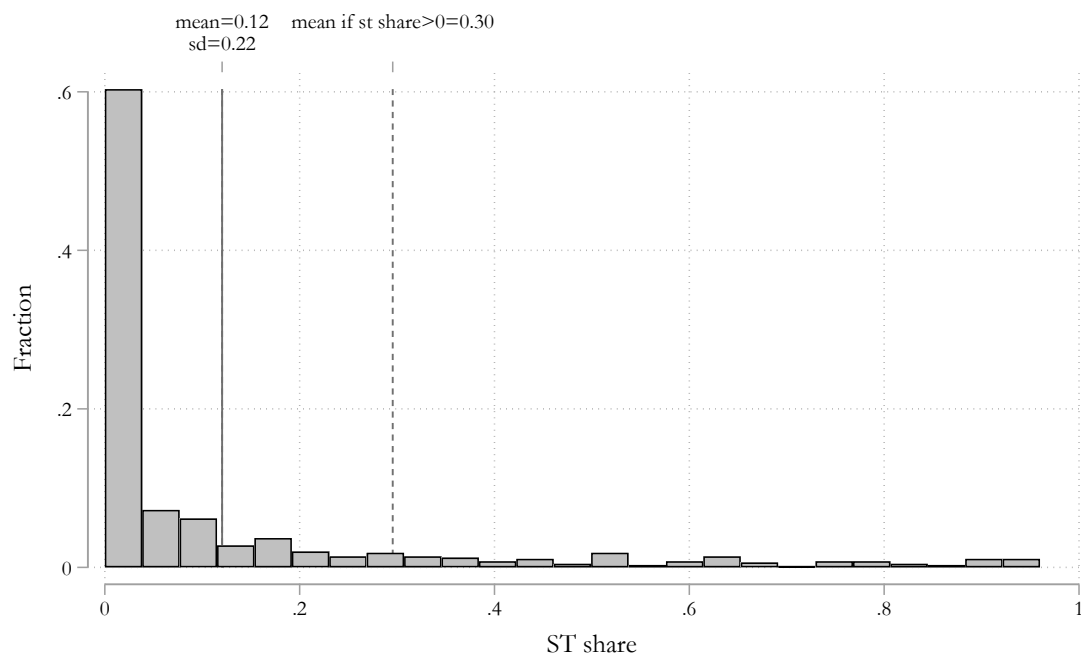
**Panel B: Impacts on Hindu Women**

|                                                                         | Ever<br>Worked      | (Inverted)<br>Seclusion<br>Index | (Inverted)<br>Marriage<br>Restrictions Index | HH<br>Earnings<br>from Agri-<br>cultural<br>Labor<br>(asinh) |
|-------------------------------------------------------------------------|---------------------|----------------------------------|----------------------------------------------|--------------------------------------------------------------|
|                                                                         | (1)                 | (2)                              | (3)                                          | (4)                                                          |
| Traditionally Non-Plough Adivasi Share                                  | 0.562***<br>(0.083) | 0.828***<br>(0.202)              | 0.790***<br>(0.144)                          | 3.904***<br>(0.835)                                          |
| Traditionally Plough Adivasi Share                                      | 0.175<br>(0.209)    | 0.295<br>(0.304)                 | -0.154<br>(0.405)                            | 0.207<br>(1.748)                                             |
| p-value: Trad. Non-Plough Adivasi Share =<br>Trad. Plough Adivasi Share | 0.064               | 0.130                            | 0.032                                        | 0.049                                                        |
| Outcome Mean for Adivasi Share = 0                                      | 0.495               | -0.001                           | 0.002                                        | 3.504                                                        |
| N                                                                       | 1,835               | 1,836                            | 1,836                                        | 1,658                                                        |
| Villages                                                                | 197                 | 198                              | 198                                          | 197                                                          |

Notes: Standard errors clustered at the village level. \* Significant at the 10 percent level, \*\* Significant at the 5 percent level, \*\*\* Significant at the 1 percent level.

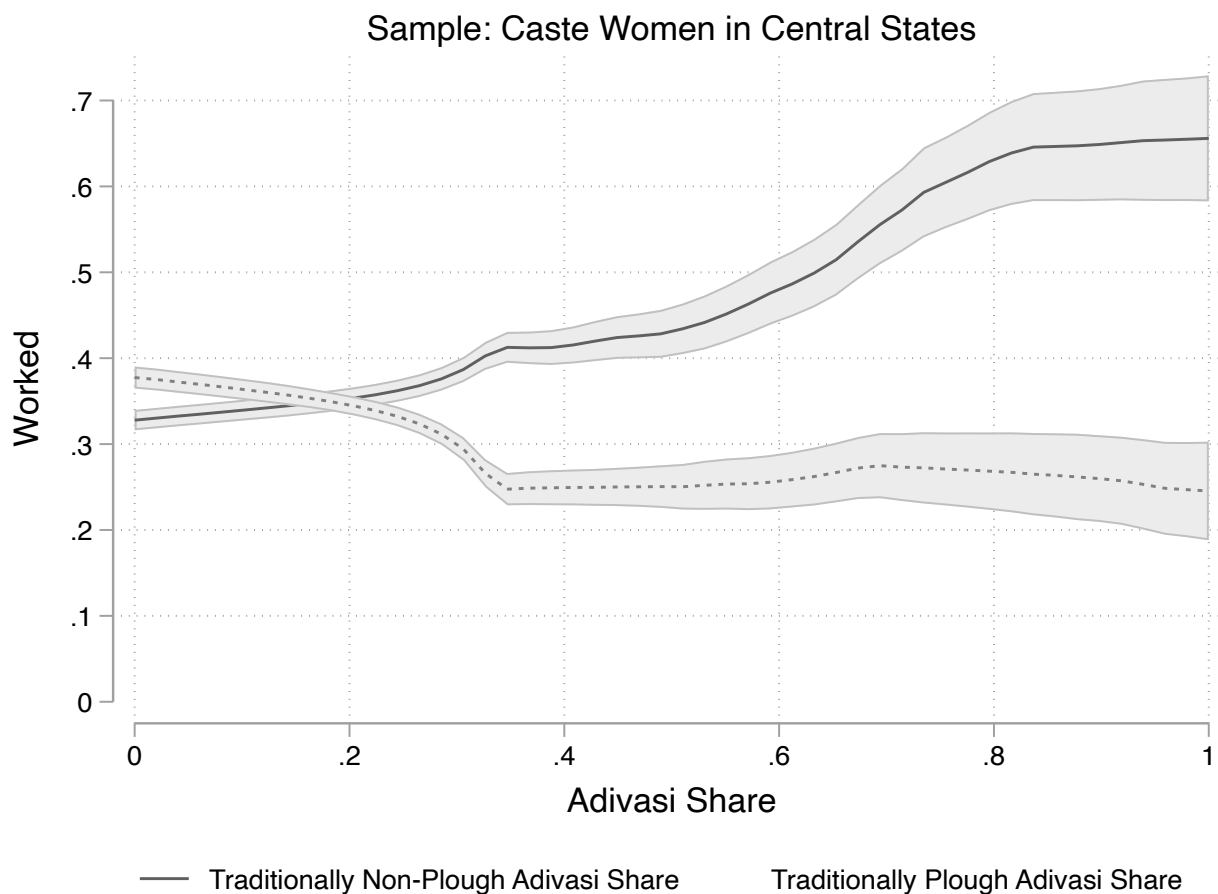
## 2.10 Appendix

Figure 2.A.1: IHDS Distribution of Adivasi Share



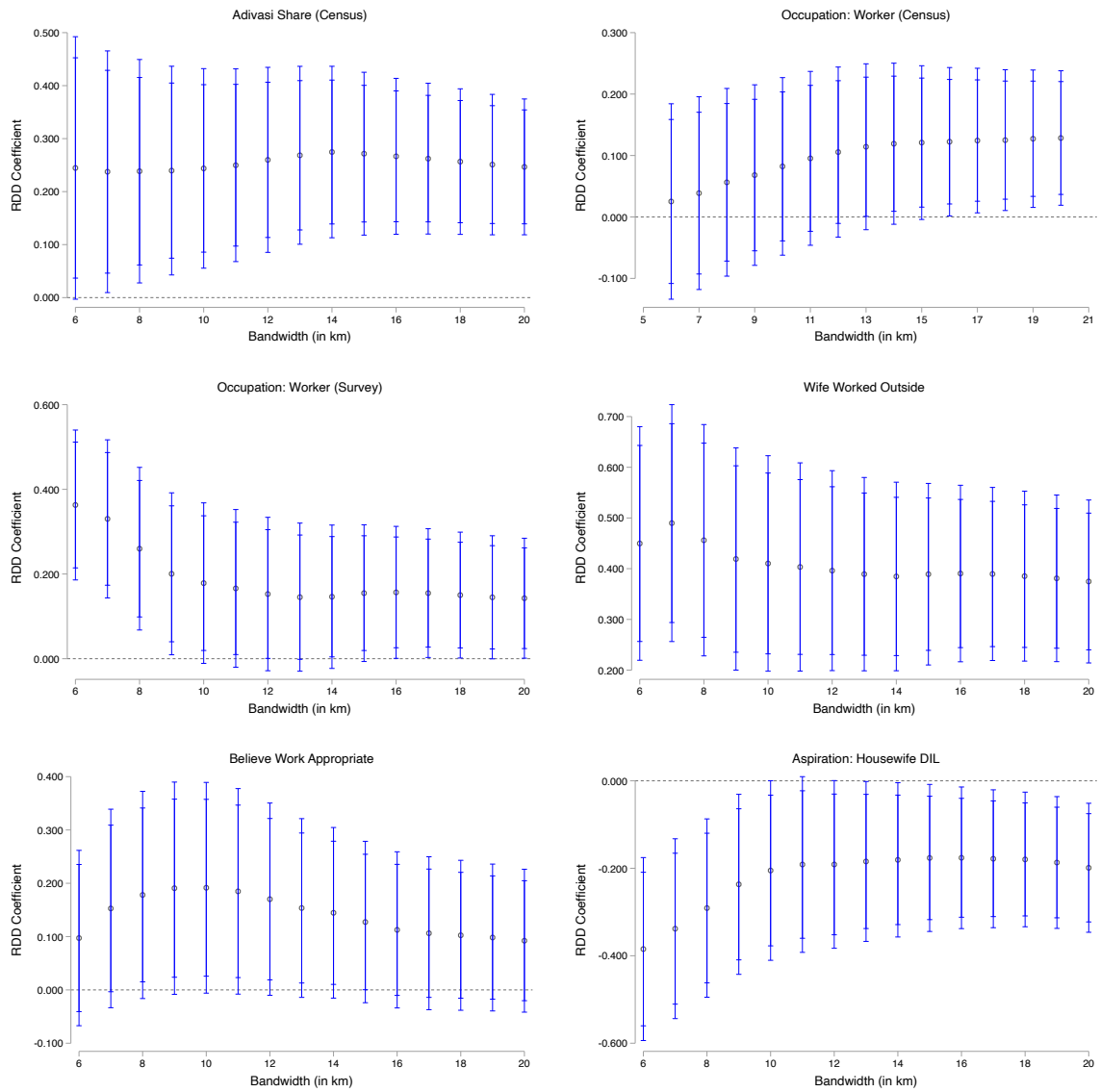
Notes: Figure constructed using data from IHDS (2011). The sample includes all caste households in central states: Andhra Pradesh, Chhattisgarh, Gujarat, Jharkhand, Madhya Pradesh, Maharashtra, Odisha, Rajasthan, and Telengana.

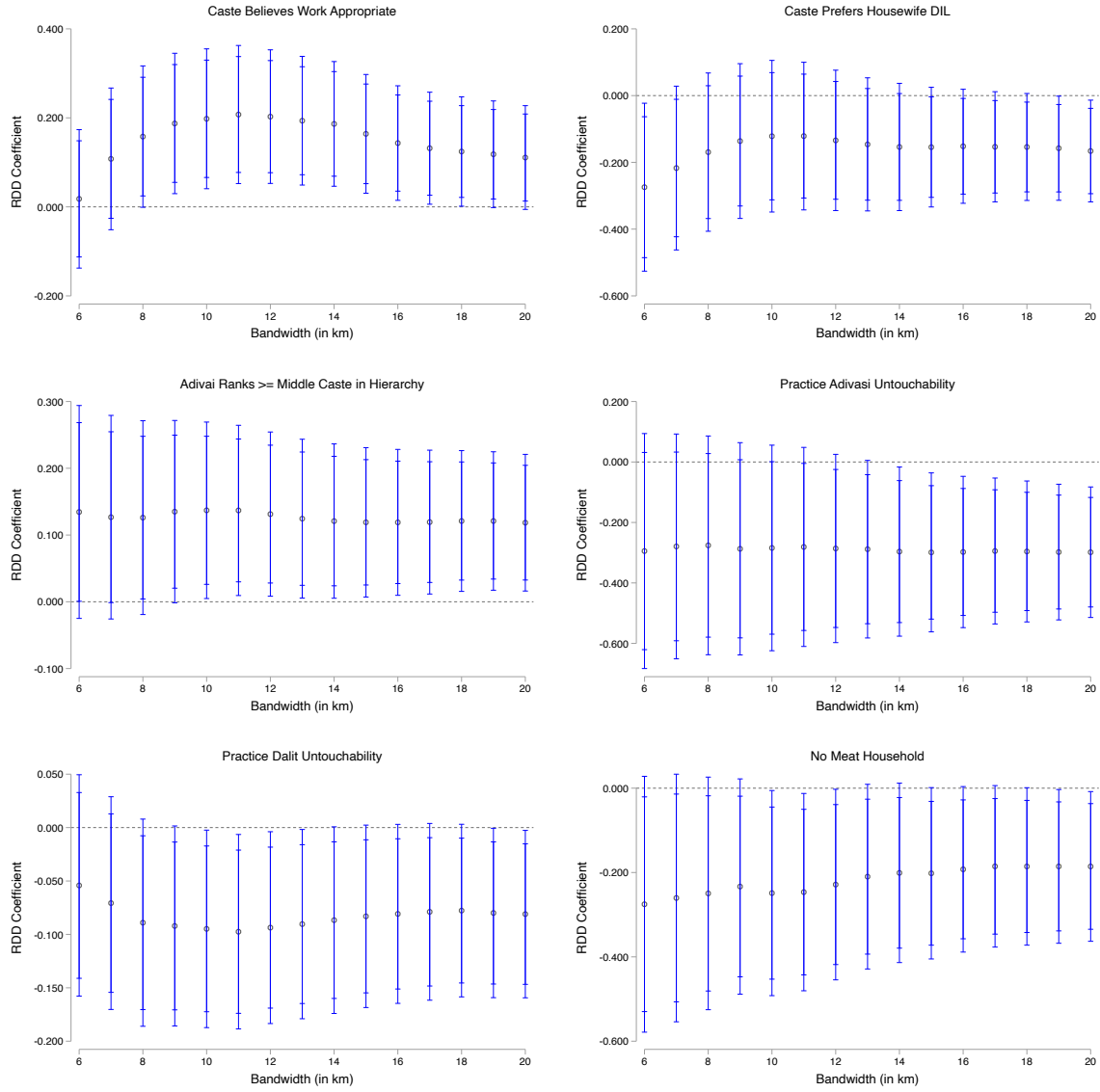
Figure 2.A.2: Non-Plough and Plough Adivasi Share Effects on Caste Women's Work



Notes: Figure constructed using data from IHDS (2011). The sample includes all Hindu caste women aged 15-65 years in rural areas in central states: Andhra Pradesh, Chhattisgarh, Gujarat, Jharkhand, Madhya Pradesh, Maharashtra, Odisha, Rajasthan, and Telengana. We use the name of the tribe in IHDS to classify Adivasis into traditionally plough and non-plough using ethnographic data on Adivasi groups. The outcome variable indicates whether a Hindu caste woman worked for pay in the past 12 months (not including work on own farm).

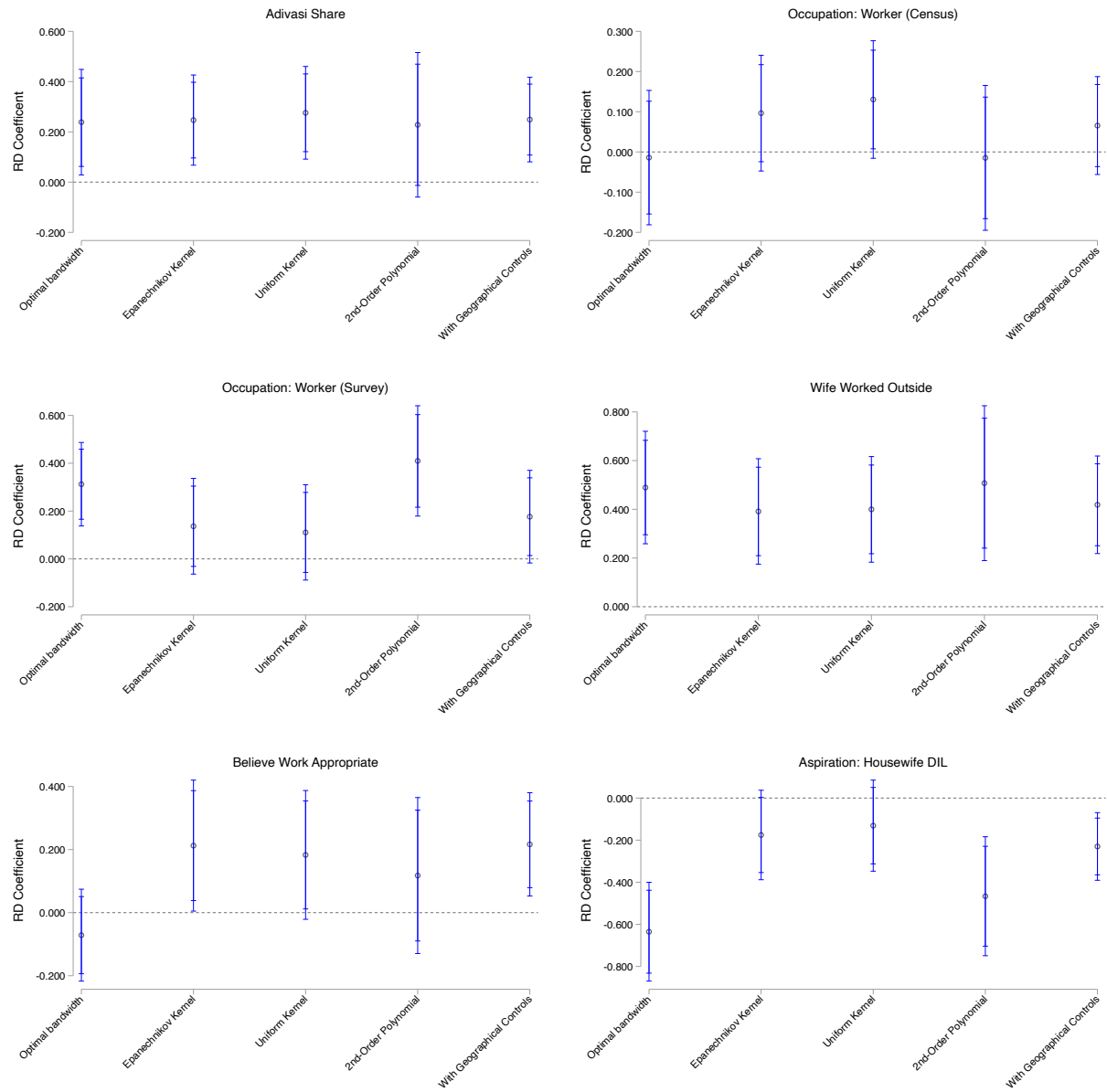
Figure 2.A.3: Robustness to RD Bandwidth

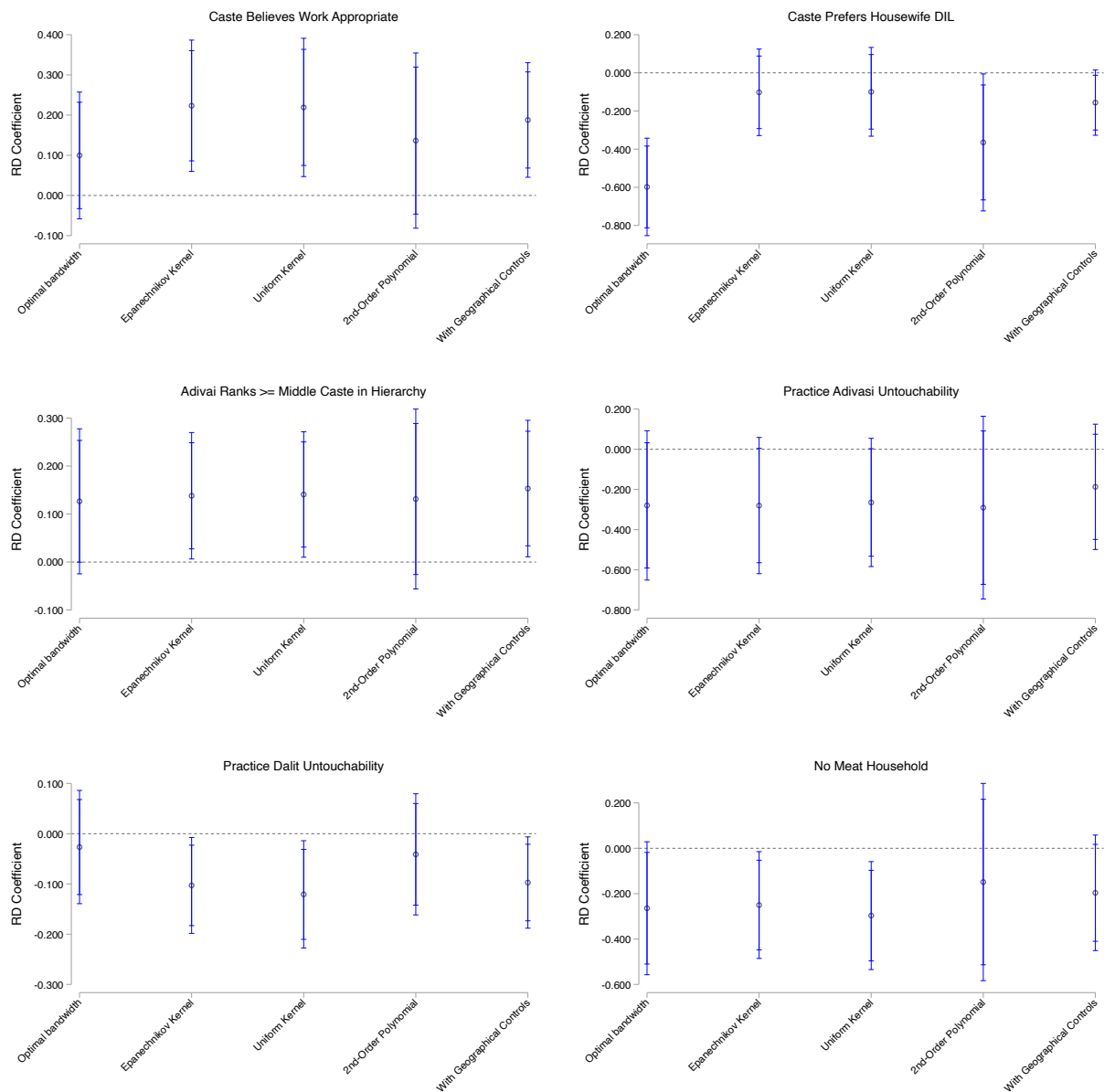




Notes: The figures show the RD coefficients for our main outcomes across different bandwidth definitions. The local linear trends are estimated separately on each side of the river using a triangular kernel. Standard errors are clustered at the village level. The range plots correspond to 90 and 95 percent confidence intervals. The first figure shows village-level Adivasi share based on the 2011 population census. The second figure shows female labor force participation rates based on individual-level data from the Government of India's 2011 Socio-Economic and Caste Census. The remaining outcomes are obtained from our own surveys with Hindu caste men and village elders. See Tables 2.5 and 2.6 for outcome definitions.

Figure 2.A.4: Additional Robustness Checks





Notes: The figures show the RD coefficients for our main outcomes across different specifications. The first coefficients come from regressions using mean-squared-error optimal bandwidths (Calonico et al., 2020). The second and third coefficients come from regressions that use epanechnikov and uniform kernels respectively. The fourth coefficients come from regressions that use 2nd-order RD polynomials. The fifth coefficients come from regressions that include the geographic controls listed in Appendix Table 2.9. The range plots correspond to 90 and 95 percent confidence intervals. The first figure shows village-level Adivasi share based on the 2011 population census. The second figure shows female labor force participation rates based on individual-level data from the Government of India's 2011 Socio-Economic and Caste Census. The remaining outcomes are obtained from our own surveys with Hindu caste men and village elders. See Tables 2.5 and 2.6 for outcome definitions.

Table 2.A.1: Village Differences by Adivasi Share

|                                                     | (1)<br>Advasi Share = 0<br>Mean | (2)<br>Adivasi Share<br>Coefficient |
|-----------------------------------------------------|---------------------------------|-------------------------------------|
| <b>Panel A: Population Numbers and Distribution</b> |                                 |                                     |
| Number of Households                                | 547.53<br>[590.47]              | -190.49***<br>(69.17)               |
| Dalit Share                                         | 0.16<br>[0.20]                  | 0.06<br>(0.04)                      |
| Middle Status Caste Share                           | 0.77<br>[0.23]                  | -0.03<br>(0.05)                     |
| High Status Caste Share                             | 0.08<br>[0.15]                  | -0.03<br>(0.03)                     |
| <b>Panel B: Remoteness</b>                          |                                 |                                     |
| Distance to Nearest Town                            | 13.94<br>[9.71]                 | 4.99***<br>(1.50)                   |
| Distance to District HQ                             | 51.09<br>[30.34]                | 16.82***<br>(4.33)                  |
| <b>Panel C: Village Infrastructure</b>              |                                 |                                     |
| Share HHs Electricity                               | 0.77<br>[0.27]                  | -0.10***<br>(0.03)                  |
| Share HHs Mobile Phone                              | 0.83<br>[0.17]                  | -0.28***<br>(0.03)                  |
| Share HHs Open Defecate                             | 0.66<br>[0.30]                  | 0.17***<br>(0.04)                   |
| Has Drainage System                                 | 0.68<br>[0.47]                  | -0.07<br>(0.06)                     |
| Village Has Paved Road                              | 0.87<br>[0.34]                  | -0.19***<br>(0.05)                  |
| Villages                                            | 380                             | 683                                 |

Notes: Sample includes villages in central states: Orissa, Madhya Pradesh, Chhattisgarh, Jharkhand, Maharashtra, Gujarat, Rajasthan, Andhra Pradesh, West Bengal. All regressions include state fixed effects and standard errors clustered at PSU (village) level.

Table 2.A.2: Political Participation

|                                          | (1)                                             | (2)                                                      | (3)                                   |
|------------------------------------------|-------------------------------------------------|----------------------------------------------------------|---------------------------------------|
|                                          | Female Share<br>Unreserved<br>Local Politicians | Hindu<br>Female Share<br>Unreserved<br>Local Politicians | Discusses<br>Politics<br>with Husband |
| <b>Panel A: Pooled</b>                   |                                                 |                                                          |                                       |
| Adivasi Share                            | 0.075*<br>(0.041)                               | 0.115*<br>(0.061)                                        | 0.110*<br>(0.059)                     |
| Mean if Adivasi Share=0                  | 0.12<br>[0.17]                                  | 0.12<br>[0.17]                                           | 0.70<br>[0.46]                        |
| <b>Panel B: By Caste Rank</b>            |                                                 |                                                          |                                       |
| $\gamma_1$ : Adivasi Share * Middle Rank |                                                 |                                                          | 0.120*<br>(0.063)                     |
| $\gamma_2$ : Adivasi Share * High Rank   |                                                 |                                                          | -0.221<br>(0.168)                     |
| $\gamma_3$ : Adivasi Share * Low Rank    |                                                 |                                                          | 0.219*<br>(0.129)                     |
| <i>P-value from F-Test</i>               |                                                 |                                                          |                                       |
| $\gamma_1 = \gamma_2$                    |                                                 |                                                          | 0.049                                 |
| $\gamma_1 = \gamma_3$                    |                                                 |                                                          | 0.448                                 |
| Mean if Adivasi Share=0 and:             |                                                 |                                                          |                                       |
| Middle Rank                              |                                                 |                                                          | 0.68<br>[0.47]                        |
| High Rank                                |                                                 |                                                          | 0.71<br>[0.45]                        |
| Low Rank                                 |                                                 |                                                          | 0.72<br>[0.45]                        |
| N                                        | 504                                             | 504                                                      | 8489                                  |
| PSU                                      | 504                                             | 504                                                      | 648                                   |

Notes: The data is obtained from IHDS (2011). All regressions include state fixed effects and control for the distance to the nearest town. The village-level outcome in column 1 is the share of women among elected village representatives, restricted to seats that have not been reserved for quotas for women or underrepresented groups (scheduled castes and scheduled tribes/Adivasi) in the most recent election. The outcome in column 2 replicates the variable in column 1 but restricts the sample to Hindu village representatives. The individual-level outcome in column 3 indicates whether a Hindu caste woman responds that she discusses politics with her husband at least sometimes. Panel B reports separate coefficients for each caste group. Standard errors clustered at the PSU (village) level. \* Significant at the 10 percent level, \*\* Significant at the 5 percent level, \*\*\* Significant at the 1 percent level.

Table 2.A.3: Mobility Index Components IHDS

|                                          | (1)<br>Cannot Attend<br>Panchayat Meeting | (2)<br>Cannot Travel<br>Alone to<br>Friend House | (3)<br>Cannot Take<br>Public Transport<br>Alone |
|------------------------------------------|-------------------------------------------|--------------------------------------------------|-------------------------------------------------|
| <b>Panel A: Pooled</b>                   |                                           |                                                  |                                                 |
| Adivasi Share                            | -0.077**<br>(0.031)                       | -0.112**<br>(0.046)                              | -0.147**<br>(0.066)                             |
| Mean if Adivasi Share=0                  | 0.91<br>[0.29]                            | 0.24<br>[0.43]                                   | 0.54<br>[0.50]                                  |
| <b>Panel B: By Caste Rank</b>            |                                           |                                                  |                                                 |
| $\gamma_1$ : Adivasi Share * Middle Rank | -0.088**<br>(0.035)                       | -0.160***<br>(0.048)                             | -0.179***<br>(0.067)                            |
| $\gamma_2$ : Adivasi Share * High Rank   | -0.105<br>(0.091)                         | 0.292*<br>(0.165)                                | 0.550***<br>(0.140)                             |
| $\gamma_3$ : Adivasi Share * Low Rank    | -0.002<br>(0.054)                         | -0.045<br>(0.077)                                | -0.309**<br>(0.131)                             |
| <i>P-value from F-Test</i>               |                                           |                                                  |                                                 |
| $\gamma_1 = \gamma_2$                    | 0.855                                     | 0.007                                            | 0.000                                           |
| $\gamma_1 = \gamma_3$                    | 0.166                                     | 0.177                                            | 0.335                                           |
| Mean if Adivasi Share=0 and:             |                                           |                                                  |                                                 |
| Middle Rank                              | 0.91<br>[0.28]                            | 0.25<br>[0.43]                                   | 0.55<br>[0.50]                                  |
| High Rank                                | 0.96<br>[0.20]                            | 0.25<br>[0.43]                                   | 0.51<br>[0.50]                                  |
| Low Rank                                 | 0.86<br>[0.35]                            | 0.21<br>[0.41]                                   | 0.54<br>[0.50]                                  |
| N                                        | 8776                                      | 8687                                             | 8729                                            |
| PSU                                      | 648                                       | 648                                              | 648                                             |

Notes: Sample includes all caste households in central states: Orissa, Madhya Pradesh, Chhattisgarh, Jharkhand, Maharashtra, Gujarat, Rajasthan, Andhra Pradesh, West Bengal. All regressions include state fixed effects and standard errors clustered at PSU (village) level.

Table 2.A.4: Marriage and Intercaste Relations Restrictions Components IHDS

|                                          | (1)<br>Widow<br>Remarriage<br>Prohibited | (2)<br>Practice<br>Dowry | (3)<br>No Choice<br>in Spouse | (4)<br>Intercaste<br>Marriage<br>Prohibited | (5)<br>Practice<br>Untouchability |
|------------------------------------------|------------------------------------------|--------------------------|-------------------------------|---------------------------------------------|-----------------------------------|
| <b>Panel A: Pooled</b>                   |                                          |                          |                               |                                             |                                   |
| Adivasi Share                            | -0.181**<br>(0.077)                      | -0.065<br>(0.053)        | -0.142**<br>(0.063)           | -0.036<br>(0.053)                           | -0.098*<br>(0.052)                |
| Mean if Adivasi Share=0                  | 0.37<br>[0.48]                           | 0.78<br>[0.41]           | 0.57<br>[0.49]                | 0.77<br>[0.42]                              | 0.14<br>[0.35]                    |
| <b>Panel B: By Caste Rank</b>            |                                          |                          |                               |                                             |                                   |
| $\gamma_1$ : Adivasi Share * Middle Rank | -0.207***<br>(0.077)                     | -0.094*<br>(0.056)       | -0.147**<br>(0.069)           | -0.033<br>(0.055)                           | -0.101*<br>(0.061)                |
| $\gamma_2$ : Adivasi Share * High Rank   | 0.419**<br>(0.196)                       | 0.080<br>(0.172)         | -0.298**<br>(0.148)           | 0.257**<br>(0.131)                          | -0.074<br>(0.226)                 |
| $\gamma_3$ : Adivasi Share * Low Rank    | -0.370***<br>(0.098)                     | 0.023<br>(0.117)         | -0.038<br>(0.103)             | -0.197<br>(0.125)                           | -0.065<br>(0.069)                 |
| <i>P-value from F-Test</i>               |                                          |                          |                               |                                             |                                   |
| $\gamma_1 = \gamma_2$                    | 0.002                                    | 0.326                    | 0.344                         | 0.038                                       | 0.909                             |
| $\gamma_1 = \gamma_3$                    | 0.102                                    | 0.305                    | 0.313                         | 0.190                                       | 0.668                             |
| Mean if Adivasi Share=0 and:             |                                          |                          |                               |                                             |                                   |
| Middle Rank                              | 0.38<br>[0.49]                           | 0.80<br>[0.40]           | 0.56<br>[0.50]                | 0.76<br>[0.43]                              | 0.15<br>[0.36]                    |
| High Rank                                | 0.47<br>[0.50]                           | 0.71<br>[0.45]           | 0.73<br>[0.44]                | 0.83<br>[0.38]                              | 0.33<br>[0.47]                    |
| Low Rank                                 | 0.29<br>[0.45]                           | 0.74<br>[0.44]           | 0.59<br>[0.49]                | 0.78<br>[0.42]                              | 0.05<br>[0.22]                    |
| N                                        | 9150                                     | 9241                     | 8780                          | 8884                                        | 6350                              |
| PSU                                      | 648                                      | 648                      | 648                           | 648                                         | 648                               |

Notes: Sample includes all caste households in central states: Orissa, Madhya Pradesh, Chhattisgarh, Jharkhand, Maharashtra, Gujarat, Rajasthan, Andhra Pradesh, West Bengal. All regressions include state fixed effects and standard errors clustered at PSU (village) level.

Table 2.A.5: Food Taboos Components IHDS

|                                          | (1)<br>Meat<br>Taboo | (2)<br>Alcohol<br>Taboo | (3)<br>Men Eat<br>First |
|------------------------------------------|----------------------|-------------------------|-------------------------|
| <b>Panel A: Pooled</b>                   |                      |                         |                         |
| Adivasi Share                            | -0.177***<br>(0.052) | -0.135**<br>(0.052)     | -0.214***<br>(0.061)    |
| Mean if Adivasi Share=0                  | 0.39<br>[0.49]       | 0.67<br>[0.45]          | 0.35<br>[0.48]          |
| <b>Panel B: By Caste Rank</b>            |                      |                         |                         |
| $\gamma_1$ : Adivasi Share * Middle Rank | -0.236***<br>(0.060) | -0.141**<br>(0.058)     | -0.217***<br>(0.065)    |
| $\gamma_2$ : Adivasi Share * High Rank   | 0.018<br>(0.220)     | 0.127<br>(0.143)        | -0.170<br>(0.180)       |
| $\gamma_3$ : Adivasi Share * Low Rank    | 0.041<br>(0.089)     | -0.225**<br>(0.104)     | -0.222<br>(0.147)       |
| <i>P-value from F-Test</i>               |                      |                         |                         |
| $\gamma_1 = \gamma_2$                    | 0.265                | 0.083                   | 0.803                   |
| $\gamma_1 = \gamma_3$                    | 0.007                | 0.453                   | 0.974                   |
| Mean if Adivasi Share=0 and:             |                      |                         |                         |
| Middle Rank                              | 0.41<br>[0.49]       | 0.71<br>[0.44]          | 0.35<br>[0.48]          |
| High Rank                                | 0.63<br>[0.48]       | 0.73<br>[0.43]          | 0.48<br>[0.50]          |
| Low Rank                                 | 0.20<br>[0.40]       | 0.48<br>[0.48]          | 0.31<br>[0.46]          |
| N                                        | 9258                 | 5544                    | 8768                    |
| PSU                                      | 648                  | 648                     | 648                     |

Notes: Sample includes all caste households in central states: Orissa, Madhya Pradesh, Chhattisgarh, Jharkhand, Maharashtra, Gujarat, Rajasthan, Andhra Pradesh, West Bengal. All regressions include state fixed effects and standard errors clustered at PSU (village) level.

Table 2.A.6: Work IHDS with Village Fixed Effects

|                                          | Ever<br>Worked   | Allowed<br>to Work | Own Decision<br>to Work | Earnings in<br>Past Year |
|------------------------------------------|------------------|--------------------|-------------------------|--------------------------|
| $\gamma_1$ : Adivasi Share * Middle Rank | 0.108<br>(0.115) | 0.210**<br>(0.106) | 0.192*<br>(0.112)       | 996.878<br>(2067.342)    |
| N                                        | 8763             | 8408               | 7732                    | 8774                     |
| PSU                                      | 648              | 648                | 648                     | 648                      |

Notes: Sample includes all caste households in central states: Orissa, Madhya Pradesh, Chhattisgarh, Jharkhand, Maharashtra, Gujarat, Rajasthan, Andhra Pradesh, West Bengal. All regressions include PSU (village) fixed effects.

Table 2.A.7: Purity Norms Indexes with Village Fixed Effects

|                                          | Practices<br>Purdah | Inverted<br>Decision-<br>Making<br>Index | Inverted<br>Mobility<br>Index | Food<br>Taboo<br>Index | Marriage<br>Restrictions<br>Index | Intercaste<br>Relations<br>Index |
|------------------------------------------|---------------------|------------------------------------------|-------------------------------|------------------------|-----------------------------------|----------------------------------|
| $\gamma_1$ : Adivasi Share * Middle Rank | -0.058<br>(0.076)   | -0.256<br>(1.272)                        | -0.640<br>(0.473)             | -0.844**<br>(0.339)    | -0.099<br>(0.401)                 | -0.264<br>(0.164)                |
| N                                        | 8771                | 8768                                     | 8773                          | 8774                   | 8774                              | 8774                             |
| PSU                                      | 648                 | 648                                      | 648                           | 648                    | 648                               | 648                              |

Notes: Sample includes all caste households in central states: Orissa, Madhya Pradesh, Chhattisgarh, Jharkhand, Maharashtra, Gujarat, Rajasthan, Andhra Pradesh, West Bengal. All regressions include PSU (village) fixed effects.

Table 2.A.8: Impact of Plough and Non-Plough Adivasis

|                                                                         | Worked<br>(1)       |
|-------------------------------------------------------------------------|---------------------|
| <b>Panel A: Adivasi Woman Worked</b>                                    |                     |
| Woman is Traditionally Non-Plough Adivasi                               | 0.115***<br>(0.036) |
| Outcome Mean for Traditionally Plough Adivasi                           | 0.372               |
| Village Controls                                                        | Yes                 |
| State FEs                                                               | Yes                 |
| Observations                                                            | 4,954               |
| <b>Panel B: Caste Woman Worked</b>                                      |                     |
| Traditionally Non-Plough Adivasi Share                                  | 0.292***<br>(0.068) |
| Traditionally Plough Adivasi Share                                      | -0.033<br>(0.079)   |
| p-value: Trad. Non-Plough Adivasi Share =<br>Trad. Plough Adivasi Share | 0.000               |
| Outcome Mean for Adivasi Share < 0.1                                    | 0.298               |
| Village Controls                                                        | Yes                 |
| State FEs                                                               | Yes                 |
| Observations                                                            | 2,449               |

Notes: Sample includes all caste households in central states: Orissa, Madhya Pradesh, Chhattisgarh, Jharkhand, Maharashtra, Gujarat, Rajasthan, Andhra Pradesh, Telengana. All regressions include state fixed effects and standard errors clustered at PSU (village) level.

Table 2.A.9: RDD Results with Conley Standard Errors

| Variable                                      | Estimate | S.E.    | p-value |
|-----------------------------------------------|----------|---------|---------|
| Adivasi Share                                 | 0.074    | (0.074) | 0.074   |
| Occupation: Worker (Census)                   | 0.042    | (0.042) | 0.042   |
| Occupation: Worker (Survey)                   | 0.054    | (0.054) | 0.054   |
| Wife Worked Outside                           | 0.064    | (0.064) | 0.064   |
| Believe Work Appropriate                      | 0.045    | (0.045) | 0.045   |
| Aspiration: Housewife DIL                     | 0.073    | (0.073) | 0.073   |
| Caste Believes Work Appropriate               | 0.048    | (0.048) | 0.048   |
| Caste Prefers Housewife DIL                   | 0.082    | (0.082) | 0.082   |
| No Meat Household                             | 0.081    | (0.081) | 0.081   |
| Adivai Ranks $\geq$ Middle Caste in Hierarchy | 0.046    | (0.046) | 0.046   |
| Practice Dalit Untouchability                 | 0.030    | (0.030) | 0.030   |
| Practice Adivasi Untouchability               | 0.084    | (0.084) | 0.084   |

Notes: All regressions are based on a local linear specification estimated separately on each side of the river boundary with a uniform kernel and 20km bandwidth. The table reports Conley standard errors with a cut-off window of 10 kilometres to account for spatial auto-correlation (Conley, 1999).

Table 2.A.10: Randomization Inference

| Variable                                       | p-value |
|------------------------------------------------|---------|
| Adivasi Share                                  | 0.002   |
| Occupation: Worker (Census)                    | 0.016   |
| Occupation: Worker (Survey)                    | 0.098   |
| Wife Worked Outside                            | 0.001   |
| Believe Work Appropriate                       | 0.258   |
| Aspiration: Housewife DIL                      | 0.020   |
| Caste Believes Work Appropriate                | 0.127   |
| Caste Prefers Housewife DIL                    | 0.056   |
| No Meat Household                              | 0.056   |
| Adivasi Ranks $\geq$ Middle Caste in Hierarchy | 0.084   |
| Practice Dalit Untouchability                  | 0.146   |
| Practice Adivasi Untouchability                | 0.009   |

Notes: All regressions are based on a local linear specification estimated separately on each side of the river boundary with a triangular kernel and 20km bandwidth. The exercise randomly re-assigns the distance to the river for each sample village 1,000 times. For each iteration, we then repeat our analysis based on the re-assigned indicator for whether the village is located on the east side of the river. The p-value is calculated as the share of the placebo coefficients that are larger in magnitude than the actual coefficient on the east side dummy (in absolute terms).

Table 2.A.11: Alternative Labor Demand and Supply Explanations

|                        | Labor Demand      |                                       |                                 | Labor Supply                   |                      |                                        |                        |
|------------------------|-------------------|---------------------------------------|---------------------------------|--------------------------------|----------------------|----------------------------------------|------------------------|
|                        | (1)               | (2)                                   | (3)                             | (4)                            | (5)                  | (6)                                    | (7)                    |
|                        | Hire<br>Women     | Price<br>Received<br>Rice<br>(Rs./Kg) | Rice Output<br>Per Acre<br>(Kg) | Women's<br>Daily Wage<br>(Rs.) | Know Current<br>Wage | Know of<br>Agri Labor<br>Opportunities | Childcare<br>Available |
| East                   | -0.115<br>(0.070) | -0.945<br>(0.778)                     | 260.147**<br>(131.107)          | -54.276***<br>(14.184)         | 0.018<br>(0.027)     | -0.064<br>(0.045)                      | -0.073<br>(0.061)      |
| Mean for West of River | 0.848<br>[0.360]  | 17.021<br>[3.033]                     | 1572.402<br>[652.842]           | 279.176<br>[44.658]            | 0.964<br>[0.188]     | 0.953<br>[0.212]                       | 0.607<br>[0.489]       |
| N                      | 532               | 307                                   | 507                             | 357                            | 373                  | 331                                    | 855                    |
| Villages               | 143               | 143                                   | 143                             | 143                            | 143                  | 143                                    | 143                    |

Notes: All regressions are based on a local linear specification estimated separately on each side of the river boundary with a triangular kernel and 20km bandwidth. Outcomes come from our own survey data with Hindu caste men and controls for age, enumerator fixed effects, and survey time fixed effects. Standard errors clustered at the village level. \* Significant at the 10 percent level, \*\* Significant at the 5 percent level, \*\*\* Significant at the 1 percent level.

Table 2.A.12: Geographic Balance Checks for Watchmen Analysis

|                             | (1)              | (2)                  | (3)               | (4)                      | (5)                 | (6)               |
|-----------------------------|------------------|----------------------|-------------------|--------------------------|---------------------|-------------------|
|                             | Adivasi Share    | Elevation            | Slope             | Log Flow<br>Accumulation | Rice<br>Suitability | Forest<br>Share   |
| Watchman Block              | 0.009<br>(0.015) | 18.854***<br>(1.566) | -0.040<br>(0.030) | -0.040<br>(0.232)        | 0.006<br>(0.019)    | -0.009<br>(0.008) |
| Mean for Not Watchman Block | 0.159<br>[0.150] | 148.004<br>[12.977]  | 0.312<br>[0.302]  | 2.777<br>[2.363]         | 1.294<br>[0.208]    | 0.037<br>[0.086]  |
| Villages                    | 443              | 443                  | 443               | 443                      | 443                 | 443               |

Notes: The outcomes in columns 2 and 4 come from the Hydroshed database. The outcome in column 3 comes from the SHRUG database (?). The outcome in column 5 comes from the Global Agro-Ecological Zoning (GAEZ) models under the assumption of intermediate input usage. The outcomes in columns 1 and 5 come from the 2011 Population Census. Standard errors clustered at the village level. \* Significant at the 10 percent level, \*\* Significant at the 5 percent level, \*\*\* Significant at the 1 percent level.

Table 2.A.13: Purity Norms Using Watchman Block Assignment

|                             | (1)                            | (2)                           | (3)                        | (4)                            | (5)                  |
|-----------------------------|--------------------------------|-------------------------------|----------------------------|--------------------------------|----------------------|
|                             | Village Has<br>Adivasi Jhankar | Adivasi Priest<br>Own Wedding | Practice<br>Untouchability | Believe<br>Work<br>Appropriate | No Meat<br>Household |
| Watchman Block              | 0.179***<br>(0.036)            | 0.169***<br>(0.026)           | -0.072***<br>(0.017)       | 0.044*<br>(0.025)              | -0.104***<br>(0.030) |
| Mean for Not Watchman Block | 0.525<br>[0.500]               | 0.690<br>[0.463]              | 0.870<br>[0.336]           | 0.662<br>[0.473]               | 0.649<br>[0.477]     |
| N                           | 1,780                          | 1,780                         | 1,780                      | 1,780                          | 1,780                |
| Villages                    | 791                            | 791                           | 791                        | 791                            | 791                  |

Notes: The outcomes are obtained from phone surveys with elected village representatives. Standard errors clustered at the village level. \* Significant at the 10 percent level, \*\* Significant at the 5 percent level, \*\*\* Significant at the 1 percent level.

Table 2.A.14: Effect of Density on Hindus

|                                      | (1)<br>Mean      | (2)<br>Population Density |
|--------------------------------------|------------------|---------------------------|
| <b>Panel A: Seclusion</b>            |                  |                           |
| Ever Worked                          | 0.493<br>[0.500] | -0.011<br>(0.008)         |
| Allowed to Work                      | 0.735<br>[0.441] | 0.004<br>(0.012)          |
| Own Decision to Work                 | 0.421<br>[0.494] | -0.012<br>(0.011)         |
| Practice Purdah                      | 0.741<br>[0.438] | -0.021**<br>(0.010)       |
| Can Attend Panchayat Meeting         | 0.091<br>[0.287] | 0.001<br>(0.004)          |
| Can Visit Health Center Alone        | 0.640<br>[0.480] | -0.005<br>(0.008)         |
| Can Visit Friend Alone               | 0.780<br>[0.414] | -0.015**<br>(0.007)       |
| Can Take Public Transport Alone      | 0.475<br>[0.499] | -0.009<br>(0.008)         |
| <b>Panel B: Food</b>                 |                  |                           |
| Men Eat First                        | 0.335<br>[0.472] | 0.012<br>(0.011)          |
| Alcohol Taboo                        | 0.670<br>[0.454] | 0.010<br>(0.010)          |
| Meat Taboo                           | 0.375<br>[0.484] | 0.003<br>(0.006)          |
| <b>Panel C: Marriage</b>             |                  |                           |
| Widow Remarriage Prohibited          | 0.360<br>[0.480] | 0.006<br>(0.011)          |
| Practice Dowry                       | 0.777<br>[0.416] | 0.002<br>(0.007)          |
| No Choice in Spouse                  | 0.558<br>[0.497] | -0.037***<br>(0.008)      |
| <b>Panel D: Intercaste Relations</b> |                  |                           |
| Intercaste Marriage Prohibited       | 0.763<br>[0.425] | -0.006<br>(0.009)         |
| Practice Untouchability              | 0.133<br>[0.339] | 0.006<br>(0.007)          |

uses population density and controls for area

Table 2.A.15: Effect of Adivasi Share on Hindus and on Adivasis

|                                      | Hindus                                |                                     | Adivasis         |                                     |
|--------------------------------------|---------------------------------------|-------------------------------------|------------------|-------------------------------------|
|                                      | (1)<br>Mean<br>if Adivasi<br>Share =0 | (2)<br>Adivasi Share<br>Coefficient | (3)<br>Mean      | (4)<br>Adivasi Share<br>Coefficient |
| <b>Panel A: Seclusion</b>            |                                       |                                     |                  |                                     |
| Ever Worked                          | 0.493<br>[0.500]                      | 0.219***<br>(0.059)                 | 0.541<br>[0.498] | 0.020<br>(0.050)                    |
| Allowed to Work                      | 0.735<br>[0.441]                      | 0.147***<br>(0.053)                 | 0.758<br>[0.428] | 0.000<br>(0.042)                    |
| Own Decision to Work                 | 0.421<br>[0.494]                      | 0.287***<br>(0.077)                 | 0.445<br>[0.497] | -0.019<br>(0.076)                   |
| Practice Purdah                      | 0.741<br>[0.438]                      | -0.262***<br>(0.075)                | 0.721<br>[0.448] | -0.070<br>(0.081)                   |
| Can Attend Panchayat Meeting         | 0.091<br>[0.287]                      | 0.078**<br>(0.031)                  | 0.090<br>[0.287] | 0.016<br>(0.027)                    |
| Can Visit Health Center Alone        | 0.640<br>[0.480]                      | 0.144**<br>(0.071)                  | 0.644<br>[0.479] | 0.057<br>(0.071)                    |
| Can Visit Friend Alone               | 0.780<br>[0.414]                      | 0.115**<br>(0.046)                  | 0.793<br>[0.405] | 0.085*<br>(0.045)                   |
| Can Take Public Transport Alone      | 0.475<br>[0.499]                      | 0.149**<br>(0.066)                  | 0.476<br>[0.499] | 0.065<br>(0.060)                    |
| <b>Panel B: Food</b>                 |                                       |                                     |                  |                                     |
| Men Eat First                        | 0.335<br>[0.472]                      | -0.215***<br>(0.061)                | 0.331<br>[0.471] | -0.034<br>(0.055)                   |
| Alcohol Taboo                        | 0.670<br>[0.454]                      | -0.134**<br>(0.052)                 | 0.628<br>[0.464] | -0.047<br>(0.058)                   |
| Meat Taboo                           | 0.375<br>[0.484]                      | -0.176***<br>(0.052)                | 0.331<br>[0.471] | -0.063<br>(0.039)                   |
| <b>Panel C: Marriage</b>             |                                       |                                     |                  |                                     |
| Widow Remarriage Prohibited          | 0.360<br>[0.480]                      | -0.182**<br>(0.077)                 | 0.352<br>[0.478] | 0.130*<br>(0.071)                   |
| Practice Dowry                       | 0.777<br>[0.416]                      | -0.064<br>(0.054)                   | 0.752<br>[0.432] | -0.283***<br>(0.068)                |
| No Choice in Spouse                  | 0.558<br>[0.497]                      | -0.141**<br>(0.062)                 | 0.558<br>[0.497] | -0.082<br>(0.062)                   |
| <b>Panel D: Intercaste Relations</b> |                                       |                                     |                  |                                     |
| Intercaste Marriage Prohibited       | 0.763<br>[0.425]                      | -0.037<br>(0.053)                   | 0.759<br>[0.428] | 0.012<br>(0.057)                    |
| Practice Untouchability              | 0.133<br>[0.339]                      | -0.098*<br>(0.052)                  | 0.123<br>[0.329] | -0.034<br>(0.035)                   |

# Chapter 3

## The Impact of Public Primary Healthcare Facilities on Birth Outcomes: Evidence from Peru

### 3.1 Introduction

Primary healthcare is often seen as a cornerstone for achieving universal health coverage (Binagwaho and Adhanom Ghebreyesus, 2019). Primary healthcare facilities act as the first point of contact for the sick and promote preventive healthcare through community interventions. Its importance is emphasized by recent estimates that suggest that governments in low- and middle-income countries spend on average as much as 1.3 percent of GDP on primary healthcare (Xu et al., 2018). Yet despite its prominent role in policy, there is only limited rigorous evidence on the impact of primary healthcare facilities on health outcomes, especially from developing countries. The scarcity of studies on developing countries matters: while the potential for improving health outcomes is larger in developing countries due to a higher burden of disease, primary healthcare facilities are also often plagued by high

rates of absenteeism and low quality of care due to limited state capacity (Chaudhury et al., 2006).

This paper examines the impact of primary healthcare facilities on birth outcomes by exploiting plausibly exogenous variation in the timing and location of new public primary healthcare facilities in Peru. In particular, I focus on the openings of I-2 facilities, the smallest public healthcare facility with a doctor in Peru. These facilities act as the first point of contact for local communities and provide basic outpatient services, including prenatal care and treatments of common diseases. Data from the national birth registry allow me to obtain information on an objective health outcome, low birth weight, on a local level with high frequency. Due to data limitations, I restrict the main analysis to the 234 facilities that opened between 2000 and 2018. If the government behaved strategically and prioritized facility constructions according to their potential effect on health outcomes, the estimates would provide a lower bound of the effect of an average I-2 facility. At the same time, I argue that measuring the effect of recent facilities is of academic and policy relevance by itself, as it emphasizes the benefits of last-mile provision of primary healthcare.

To measure the effect of increased access to primary healthcare, I adopt an event-study framework to compare the outcomes of treated and untreated districts before and after a new facility opened. This identification strategy assumes that treated and untreated districts would have been on the same trend in the absence of the additional I-2 facility. I argue that this assumption seems reasonable in this setting since health investments are decided by each department separately and since constructions often experience significant delays due to political and logistical reasons. The event-study framework provides further support for this assumption by showing that treated and untreated districts were on a similar trend before a new facility opened.

My main results document that an additional I-2 facility in a district leads, on average, to a decline in low birth weight by 1.3 percentage points after an initial two-year adjustment

period. Consistent with previous studies (Gujral and Basu, 2019) and a higher availability of alternative healthcare providers in cities, I find no change in birth outcomes for urban areas. These results remain robust to the inclusion of public infrastructure controls and interactions between linear time trends and various district characteristics. Results from the National Health and Family Survey also suggest that the findings are not driven by changes in the selection into the birth registry.

Having documented a decline in low birth weight, I then study three potential mechanisms that could explain how increased access to primary healthcare could affect birth outcomes. I begin by ruling out that changes in general maternal health or the number of prenatal visits led to a reduction in low birth weight. I then show that an additional healthcare facility seems to have increased the quality of prenatal care. While I do not have information on many intermediate outcomes that should directly affect the incidence of low birth weight (like the likelihood of having a preterm birth), I find improvements in various proxies for quality of prenatal care, including the likelihood of receiving a tetanus vaccine during pregnancy. Additionally, I provide suggestive evidence for changes in fertility behavior through increased access to family planning methods.

My findings contribute to the growing literature on the effects of increased access to healthcare on health outcomes. While several studies examine the effect of increased access to primary healthcare (Cesur et al., 2017; Bailey and Goodman-Bacon, 2015; Bailey, 2012; Currie et al., 1995; Olkkola, 2019) and other healthcare infrastructure (Gujral and Basu, 2019; Buchmueller et al., 2006) in developed countries like the US and Finland, there is only limited evidence from developing countries (Carrillo and Feres, 2019; Rocha and Soares, 2010; Hsiao, 2019). The paper closest to mine is a study of the impact of community health centers which were constructed as part of the War on Poverty program in the US in the 1960s (Bailey and Goodman-Bacon, 2015). Using a similar identification strategy, the authors find a large decline in elder mortality rates but no change in infant mortality

rates in treated counties. Similarly, Cesur et al. (2017) show that providing free healthcare through state-employed family physicians decreased mortality rates among several age groups in Turkey. Evidence from a developing country comes from Hsiao (2019), who examines the effect of constructing new primary healthcare centers in Indonesia. However, since the study focuses on the measurement of misallocation, the author only examines changes to healthcare utilization and not health outcomes. Two related studies show that adding a doctor to existing primary healthcare facilities reduced mortality rates in Nigeria (Okeke, 2023) but that an increase in the supply of doctors in rural areas in Brazil did not affect birth outcomes (Carrillo and Feres, 2019). A reason for the scarcity of studies on the impact of increased access to primary healthcare in developing countries is the need for reliable, granular, and frequent data on health outcomes. The availability of detailed information from the national birth registry and annual household surveys make Peru an ideal place to study the question. The findings also contribute to the wider literature on the determinants of child health outcomes (Almond et al., 2012; Bharadwaj et al., 2014; Currie and Gruber, 1996; Currie and Schwandt, 2016). Finally, the paper is related to the literature on the impact of infrastructure projects in developing countries (Jensen, 2007; Jensen and Miller, 2018; Asher and Novosad, 2018; Hjort and Poulsen, 2019; Duflo and Pande, 2007; Duflo, 2001).

The rest of the paper is organized as follows. Section 2 provides information on the healthcare sector in Peru, including details on the services provided by I-2 facilities. Section 3 describes the data and discusses the empirical strategy. Section 4 presents the main results and provides suggestive evidence for different mechanisms. Section 5 concludes.

## 3.2 Context

Peru experienced rapid economic growth and large improvements in health outcomes in the past two decades. This includes a reduction in the infant mortality rate from 55 deaths per

1,000 live births in 1992 to 17 in 2013 and a drop in the stunting rate for children under five from 36.5 percent in 1992 to 13.1 in 2013 (Vermeersch et al., 2014).<sup>1</sup> These changes were driven to a large extent by significant investments in public healthcare services and a conditional cash transfer that required mothers to regularly take their children to growth monitoring and promotion check-ups at public health centers (Marini et al., 2014). The government also implemented a national health insurance program (SIS) in 2005 to provide free healthcare to the poor at public facilities. As of 2018, 39% of the population were covered under the scheme (Vermeersch et al., 2014). Despite these improvements, Peru also had many problems in the healthcare sector. The country suffered from a shortage of doctors and an audit study found that 25 percent of public healthcare providers in Peru were absent at a given point of time (Chaudhury et al., 2006).

Public primary healthcare facilities in Peru are divided into four categories. The smallest (I-1) facilities are usually staffed by a nurse and provide maternal and child care in remote locations. The second-smallest (I-2) facilities are located across the country and provide basic outpatient services in addition to child and maternal care. They are open for twelve hours throughout the week and are usually staffed by a doctor and two nurses. Larger primary healthcare facilities (I-3 and I-4) are located in urban areas and also provide inpatient care.

This paper focuses on I-2 facilities due to sufficient variation in the location and timing of new facilities and because their volume and catchment area are large enough to affect district-level outcomes. According to provider-patient interaction data from the Health Information System in March 2019, the most popular services provided by these facilities are dental surgery, nursing, general medicine, obstetrics, and growth and development. The health workers also conduct weekly community visits to provide nutritional supplements and increase awareness of health-promoting behaviors.

---

<sup>1</sup>As a reference point, the infant mortality rate in the US in 1960 was 26 deaths per 1,000 live births (Bailey and Goodman-Bacon, 2015) and the average infant mortality rate in lower middle-income countries in 2018 was 37 deaths per 1,000 live births according to World Bank estimates.

Health policy is the responsibility of the departments, the first-level administrative subdivision of Peru. Departments set their own priorities and can request funding from the central level. This is important for my identification strategy since the lack of central planning leads to plausibly exogenous variation in the timing of new facilities across departments. Delays in the construction and opening dates of new facilities due to logistical and political reasons, including land ownership disputes and separate negotiations with different international development agencies, further create plausibly exogenous variation within departments.

While similar studies usually exploit specific construction programs (e.g. Bailey and Goodman-Bacon (2015)), my variation comes from the continuous openings of new facilities over time, as shown in Figure 3.1. The vertical lines demarcate the years for which information from the birth registry is available (2005 - 2016). This data limitation is an important caveat of my analysis since this implies that I can only estimate the effect of recent facility openings on birth outcomes. While this would not matter if the impact of an additional facility was constant, it is more reasonable to assume that the effect of an additional facility depends on local circumstances. If the government behaved strategically and prioritized the construction of new facilities in areas, in which it expected to see the largest effects on health outcomes, my estimates would provide a lower bound for the effect of an average I-2 facility. At the same time, measuring the effect of recent facilities might be of academic and policy relevance by itself, if we are interested in the benefits of last-mile provision of primary healthcare.

## **3.3 Data and Estimation**

### **3.3.1 Data**

Data on the classification, location, and opening dates of the healthcare facilities come from the National Registry of Health Facilities (RENAES). The registry was started in 2009 and

covers all operational public healthcare facilities as of January 2019.<sup>2</sup> I supplement this data with information from the National Birth Registry that is available from 2005 until 2016. The birth registry contains information on the district of delivery, the district of residence of the parents, and the birth weight of the child. I use the district of residence to merge the data with the facility locations, since I assume that the local availability of healthcare services during and before pregnancy is more important for birth outcomes than the healthcare infrastructure in the district of delivery.<sup>3</sup> Data on the number of postnatal and infant deaths are obtained from the Ministry of Health.

To differentiate between mechanisms, I use data from the National Family Health Survey (DHS/ENDES). Peru was the first country to implement a continuous survey with the DHS program in 2005. This means that the surveys are not conducted at the typical five-year interval but in five annual cycles. The sample in each cycle is representative at the national level and is based on a random selection of clusters. The survey contains district identifiers as well as information on various biomarkers, the usage of family planning methods, and the birth history of each respondent, including prenatal care, birth weight, and place of delivery. Since the birth history information covers all births of the respondents, it allows me to create a district-year level dataset retrospectively even if a respondent was not interviewed in a particular year. The sample size increased over the years and ranges from 13,201 households in 2005 to 37,486 in 2018.

District-level characteristics are obtained from the 2007 population census. I aggregate individual-level information to calculate the total population, the urban share, the SIS coverage rate, the average wealth index, and the secondary school enrollment rate in each district.

---

<sup>2</sup>While the registry officially also includes private healthcare facilities, the coverage of the private sector is relatively incomplete. Another caveat is that I do not have information on closures. However, they are relatively rare in this setting.

<sup>3</sup>For 50 percent of the births, the district of residence is the same as the district of delivery. The main results are robust to using the district of delivery instead. A caveat of both variables is that they are sometimes incorrectly entered: 9% of births could not be matched to any of the census districts. Reassuringly, unmatched births are not significantly more or less likely to suffer from low birth weight.

Based on the urban share, I divide districts into rural and urban areas. I use the socioeconomic district characteristics to control for differential trends in these characteristics in the event-study regressions.

### 3.3.2 Empirical Strategy

My empirical strategy exploits variation in when and where public healthcare facilities opened to examine their impact on birth outcomes. The identification assumption is that districts with and without a new facility would have been on the same trend in the absence of a new facility. To examine how the impact of new facilities evolved over time, I use a flexible event-study design:

$$Y_{dt} = \theta_d + \delta_t + X_{dt}\gamma + \sum_j^{J_d} \sum_{y=-5}^5 \beta_y 1(t - e_j^d = y) + \epsilon_{dt}. \quad (3.1)$$

$Y_{dt}$  is the outcome in district  $d$  in period  $t$ .  $\theta_d$  is a set of district fixed effects, which absorb time-invariant differences in observable and unobservable district characteristics.  $\delta_t$  is a set of year-quarter or year fixed effects, which capture time-varying changes on the national level.  $X_{dt}$  includes a constant, and, depending on the specification, counts of public healthcare facilities in each category and interactions of 2007 census characteristics, region fixed effects, and department fixed effects with linear time trends.  $1(t - T^* = y)$  is an indicator variable that is equal to one if the year of observation is  $y$  quarters from  $e_j^d$ , the date when a new healthcare facility was opened. Events outside of sample period are captured by dummies,  $1(t - e_j^d = y) \leq -5$  and  $1(t - e_j^d = y) \geq 5$ .

Since a district can receive more than one event during my data period, I follow the recommendation from Sandler and Sandler (2014) and adopt a multiple-indicators approach. This implies that I sum over all  $J_d$  facilities that a district receives over my data period. For example, if a new facility opened in a district in the second and third quarter of 2010,

the event variables  $t - 1$  and  $t - 2$  would be equal to one for the first quarter of 2010. Similarly, an event variable can also have a value of more than 1 if several facilities started their operations within the same quarter. Appendix Table 1 shows the number of districts based on the number of I-2 facilities that were open in 2000 and 2018. I exclude districts that ended up with more than eight districts in 2018 from my analysis due to the concern that an additional facility has probably only a small effect on birth outcomes if several I-2 facilities are already operational in the district. Appendix Figure 1 shows that facilities opened across the country during my data period. Following standard conventions, I use the quarter just before the opening date of a facility as my reference period by excluding the  $t - 1$  event variable. I cluster my standard errors at the district level to allow for correlations in my outcomes within a district.

If the outcome is obtained from the birth registry, a district is the unit of observation. Districts are weighted by the number of births unless otherwise specified. If the outcome is obtained from the DHS survey, the unit of observation is either a respondent or a birth from the birth history module. Individual-level regressions allow me to include household controls as part of the robustness checks. Without household controls, this is equivalent to weighting districts based on the number of respondents or the number of births. I obtain similar results for the respondent-level outcomes if I use the 2007 census population as sample weights instead (not shown).

I present results using an unbalanced set of districts to include as many events as possible. An unbalanced panel allows me to use facilities that opened four years before my data period (2000-2004) or two years after my data period (2017-2018) in my estimation of the event variables. As part of the robustness checks, I show that my results remain similar if I exclude districts that received a new facility between 2000 and 2002.<sup>4</sup>

---

<sup>4</sup>Unfortunately, the sample size is too small to show results for a fully balanced panel.

Table 3.1 provides summary statistics for districts that never received a facility (control), for districts that received at least one facility in or after 2000 (event), and for districts that received at least one facility before 2000 but not afterward (past event), separately for urban and rural areas. Including the pure control group and past-event group addresses the identification problem discussed in Borusyak and Jaravel (2018) since they contribute to the identification of the time fixed effects. For ease of reading, I refer to the control and past-event group as untreated districts in the rest of the paper.

Comparing urban and rural districts, the table documents that the number of alternative healthcare providers (private facilities and public hospitals) is larger in urban areas than in rural areas. The results in the first three columns further suggest that the government did not seem to have placed the facilities randomly across the country. As we would predict, the government focused instead on districts with a higher demand for additional access to healthcare services. In particular, event districts have a larger population and a higher incidence of low birth weight in 2005. They also have a lower secondary school enrollment rate and a higher SIS coverage rate.

The estimation strategy allows for these differences through the inclusion of district fixed effects as long as the trends would have been the same between treated and untreated districts in the absence of the treatment. To support the plausibility of the parallel trend assumption, I check for pre-trends in the event-study regression. If differences in the trends only appear after the opening of a facility, it provides strong evidence that these changes are driven by the new healthcare facility. To address the concern that poorer areas might have been on a different time trend through additional support from the government, I also report results from a specification that includes interactions of 2007 census characteristics with linear time trends.

## 3.4 Results

### 3.4.1 Effect of Facilities on Low Birth Weight

I begin by examining the impact of I-2 facilities on the incidence of low birth weight in rural areas. Following WHO conventions, I define low birth weight as weighing less than 2,500 grams at birth. Figure 2a shows the results from an event-study regression with district and quarter-year fixed effects. Each circle corresponds to the coefficient of the event variable plus the 2005 outcome mean in the control group. The grey lines indicate 95% confidence intervals. While the pre-period estimates are noisy, they provide no clear evidence for differential trends between treated and untreated districts. This supports the identification assumption that treated and untreated districts would have been on a similar trend in the absence of the opening of a new I-2 facility. At first, the outcome also remains relatively similar across treated and untreated areas after the opening of a new facility. It is only after around 2.5 years that we observe a significant drop in the incidence of low birth weight in treated districts. Potential explanations for the delay are that it takes time to change healthcare utilization patterns, for awareness campaigns to affect the behavior of the people in the community, and for facilities to be fully functional.

A concern is that this result is driven by the nature of the unbalanced panel. Since I only observe birth outcomes from 2005 onwards but include facilities that opened before 2005 in my estimation of the event variables, the decline could be driven by systematic differences in the 2000 - 2002 facilities relative to the other facilities in my sample. To address this concern, I rerun the event-study regression without the 47 districts that received a new facility between 2000 and 2002, as these facilities only contribute to the identification of the event variables for  $t = 9$  and higher. Appendix Figure 2 shows that, while the results become noisier, the drop after ten quarters remains if I exclude these districts.

Figure 2b shows I do not find similar effects for urban areas. Treated and untreated districts remain on similar paths even after the opening of a new facility. This is consistent with findings from Gujral and Basu (2019), who only observe a decline in inpatient mortality after hospital closures in rural areas in California. The intuition behind this result is that transportation costs are lower in urban areas and that more alternative healthcare providers are available (as shown in Table 1). Thus, it is easier for households in urban areas to seek healthcare in the absence of an additional health facility than for households in rural areas.

### 3.4.2 Robustness Checks

#### Alternative Specifications

Table 3.2 reports results from a pooled regression across different specifications. Column (1) replicates the basic specification with district and quarter-year fixed effects. The estimated coefficient implies that an additional health facility leads to a decrease in low birth weight by 1.3 percentage points eleven to nineteen quarters after the opening date. This result means a drop by 15 percent relative to the 2005 mean in the event sample of 0.087 (Table 3.1). This estimate is robust to controlling for additional public healthcare infrastructure, interactions between linear time trends and district characteristics, and interactions between linear time trends and geographic regions (coast, mountains, and jungle).<sup>5</sup> The exception is Column (6), in which I focus on the within-department variation by including interactions between linear time trends and department fixed effects. The estimate remains negative but drops by half and is no longer significant. This suggests that departments that opened new facilities might have been on different trends than other departments during this period. However, since this specification absorbs most of the across-department variation, it is also more sensitive to measurement error. Hence, it is unclear which specification is preferable. Reassuringly,

---

<sup>5</sup>My estimates also remain robust to including interactions between the different controls and quarter-year fixed effects instead.

none of the pre-trend coefficients are significant in any of the specifications, supporting the assumption that treated and untreated districts would have been on the same trend in the absence of the treatment.

## Addressing Selection

A main concern is that the previous results are driven by selection into the birth registry. Comparing the number of infants in the 2007 census to the number of registered births in 2007, I find that the number of infants in the census is 54 percent higher than the number of births in the birth registry in the same year. Thus, it might be the case that a new facility did not lead to actual improvements in birth outcomes but just affected the number of registered births. For this argument to hold, we need that the marginal births that are registered due to the new facility have better birth outcomes than the average birth in the birth registry.

To address this concern, I first use data from the birth history module of the National Family Health Survey (DHS). The advantage of this dataset is that it collects information on all births of the respondent.<sup>6</sup> While this helps me to overcome a part of the selection problem, it does not solve it completely. First, among children that were delivered at home, birth weight is only available for 59 percent of births since many children were not weighed at the time of birth. Second, even if birth weight is available for children that were delivered at home, it is only based on the recall of the mother.

Figure 3a shows the results from the event-study regression based on birth weight from the DHS birth history module. While the point estimates suggest a decline in low birth weight immediately after the opening of an additional facility, the results are overall very noisy and not statistically significant. Note that the results from Table 2 are within the confidence intervals in Appendix Table 2. The DHS data also document a much higher

---

<sup>6</sup>Since information is collected retrospectively, the DHS data further allow me to cover a larger time period.

share of children with low birth weight in the control group in 2005 (0.079 instead of 0.05). However, if I restrict the sample to the children for which birth weight was obtained from a birth card in the DHS survey, the control group mean in 2005 goes down to 0.049. Without further data, it remains unclear whether this is driven by recall bias or selection. Since the results remain inconclusive, I continue to address the selection concern by examining the place of delivery in more detail instead.

Figure 3b documents that the distributions of birth weight for births delivered at home and for births delivered at the public hospital look surprisingly similar. While there is clear bunching for births delivered at home, suggesting that respondents round the birth weight in their recall towards the nearest decimal, it seems that both groups come from the same underlying distribution. This is consistent with a story in which different selection channels offset each other. On the one hand, we would expect that wealthier households are more likely to deliver at a public hospital since they can pay the transportation costs and might be more aware of the benefits of delivering at a public hospital. On the other hand, births with complications are more likely to be delivered at a public hospital as well, lowering the average birth weight at a public hospital. If the average child thus switches the place of delivery, we should not expect a change in low birth weight. Instead, a decrease in low birth weight requires that the healthier part of the distribution switches the place of delivery. This could occur if wealthier households live closer to the new healthcare center, making it easier for them to access the new facility.

Figures 4a-d provide suggestive evidence that adding a healthcare facility indeed changes the place of delivery: mothers seem to switch from their home to the public hospital. While I-1 and I-2 facilities are only rarely used as places of delivery themselves, their presence could affect delivery choices through referrals or awareness campaigns. Heterogeneity analysis further shows that these effects are concentrated among households below median wealth (Appendix Figures 3 and 4). This goes against the concern that the main findings are driven

by selection since this suggests that the switchers have actually lower birth weight than the average registered birth.

### **3.4.3 Mechanisms**

In this subsection, I discuss three potential mechanisms of how an additional I-2 facility could affect the birth weight of a child. These include improvements in the health of the mother, increased access to high-quality prenatal care, and changes to fertility.

#### **Improved Maternal Health**

I begin by discussing the effects on maternal health. Birth outcomes could improve if mothers are healthier and less malnourished through increased access to primary healthcare. In an ideal dataset, I would thus examine changes to body weight and other health outcomes among pregnant women. However, since I do not have this information for a large sample, I consider health outcomes for women of reproductive age instead. I focus on two objectively measured health outcomes, the body mass index and the prevalence of anemia, which are correlated with low birth weight according to previous studies (Rahmati et al., 2017; Gondwe et al., 2018). Figure 5 shows that I find no significant effects for neither of these outcomes, providing suggestive evidence against this channel.

#### **Improved prenatal care**

An alternative mechanism is that an additional I-2 facility improves access to high-quality prenatal care. Previous research suggests that prenatal care could lower the incidence of low birth weight through anti-smoking awareness programs, nutritional support, and a decreased risk of preterm births through increased medical support (Alexander and Korenbrot, 1995). Ideally, I would examine the effects on these outcomes explicitly. This includes the prevalence of smoking among pregnant women, the likelihood of receiving nutritional supplements, and

the likelihood of having a preterm birth. Unfortunately, except for information on iron supplements, none of this information is available in the datasets I have. Instead, I consider proxies for the quality of prenatal care like tetanus injections and nutritional support after pregnancy.

I find no changes in the number of prenatal visits (Figure 6) but an increased likelihood that a mother received prenatal care by a doctor (Figure 7a). Appendix Figure 5 suggests that this effect is mostly driven by substituting nurses with doctors. This change by itself might not affect birth outcomes if the quality of prenatal care across healthcare professionals is similar. For example, a previous study in Brazil found that a government policy that led to a substitution of nurses for doctors had no effect on low birth weight or mortality rates (Carrillo and Feres, 2019). To address this concern, I consider changes in three proxies for the quality of prenatal care: iron supplements before birth, tetanus injections before birth, and vitamin doses after pregnancy. The latter two outcomes should not have a direct effect on birth weight, but they could indicate the general quality of care that a mother receives. While I see no changes for iron supplements (Figures 7b), I find improvements for tetanus injections and vitamin doses after the opening of a new I-2 facility. Overall, the results provide suggestive evidence that access to higher quality prenatal care could drive the decline in low birth weight.

## **Family Planning**

A final mechanism that I consider is a change in fertility behavior. Primary healthcare facilities could affect fertility through awareness programs and increased access to family planning methods. According to DHS data, 38 percent of respondents reported that they receive condoms at an I-1 or I-2 facility, emphasizing the importance of primary healthcare for family planning in this setting. To examine the effect on fertility, I run a regression with the number of births in the birth registry as the outcome. Since the number of births

is an aggregate and not an individual-level outcome and since I do not want to weigh by the outcome itself, I remove the sample weights in this specification. Figure 8a shows that the number of births is drastically declining after the opening of a new facility. Eleven to nineteen quarters after the opening of a new I-2 facility, treated districts have on average seven fewer births than untreated districts. The timing is consistent with the drop in low birth weight after around 2-3 years. The actual effect might be even larger when we take into account the selection through increased deliveries at public hospitals. I also find a noisy increase in the likelihood that a mother report that the pregnancy was wanted after the opening of a new facility (Figure 8b). The low outcome mean in the control group of 33 percent documents the latent demand for family planning. However, this outcome is reported retrospectively and might be subject to reporting bias. Figure 8c provides further suggestive evidence for this mechanism by showing a noisy increase in condom usage. Together, these results are consistent with studies from the US that find a persistent decline in fertility after the introduction of family planning programs (Bailey, 2012).

If additional facilities increased access to family planning, we would also expect that we see changes in maternal age and birth order. However, Appendix Figure 6 suggests that mothers are not less likely to be very young or very old and that births are not less likely to be of higher birth order. An explanation for this result could be that the selection bias goes in the opposite direction. Controlling for household characteristics, higher birth order children are more likely to be delivered at home (not shown here). This might be the case because more experienced mothers do not see the need to go to the hospital. If an additional I-2 facility encourages such mothers to deliver at a public facility, this selection effect could offset any changes that are driven by increased access to family planning. A similar argument could also explain why I find no changes in maternal age. In an ideal setting, I would further examine effects on maternal socioeconomic status, e.g. income or education, but this data is not available in the birth registry for most years.

### 3.4.4 Other Health Outcomes

In this subsection, I examine changes in mortality rates as an alternative health outcome. As shown in Figure 9, I find no decrease in neonatal and child mortality rates after the opening of a new facility. However, the concern with mortality rates is that this information is even more prone to selection bias than birth weight since a large fraction of deaths in rural areas in Peru is not registered. Thus, the (noisy) increase in infant mortality could be explained by improvements in the monitoring and tracking capacity of the government through the presence of an additional primary healthcare facility in the area.

Related studies often use census cohorts to examine changes in mortality rates (e.g. Lleras-Muney (2005)). By tracking the number of individuals within birth cohorts across censuses based on the place of birth, this method allows the analysis of long-term differences in mortality rates. Unfortunately, I cannot adopt this approach since I only have information from the 2007 census and since the dataset does not include information on the place of birth. Based on the cross-sectional data alone, I am not able to differentiate between mortality and fertility effects across cohorts.

Due to these limitations, I do not want to put too much emphasis on the mortality results. Further research with better datasets is necessary to rigorously examine the effect of increased access to primary healthcare facilities on mortality.

## 3.5 Conclusion

This paper provides new evidence on the impact of primary healthcare facilities on birth outcomes. It contributes to the debate on the benefits of primary healthcare and shows that primary healthcare can make a difference even in a developing country like Peru. My empirical strategy exploits variation in the timing and location of new facilities to quantify the effects of an additional I-2 facility in a district on low birth weight. My results imply that

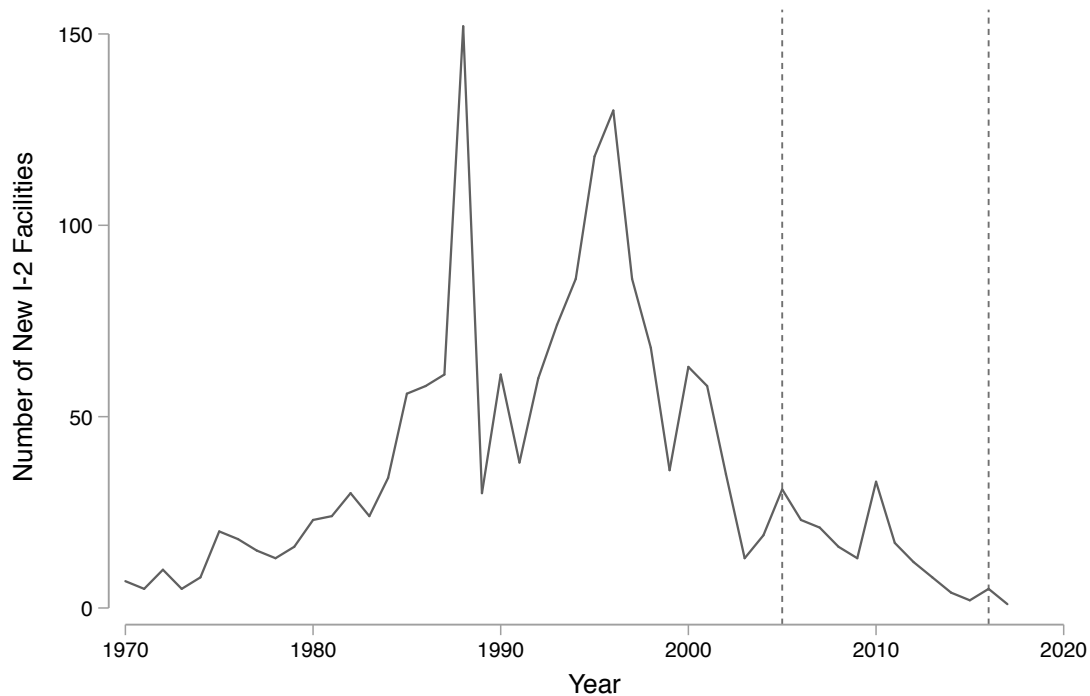
increased access to primary healthcare led to a significant drop in low birth weight in the long run. Further evidence suggests that these changes are driven by higher quality prenatal care and increased access to family planning.

There are important caveats of the study that I would like to emphasize. Due to data limitations, I can only estimate the effect of recent facility openings. The impact of a last-mile facility would be lower than the impact of an average facility if the government behaved strategically and prioritized constructions in areas in which it expected to see the largest effects. Another concern is the limited coverage of the birth registry. While I provide evidence against the argument that my results are driven by selection, I cannot completely rule out that the estimates are affected by more births being registered after the opening of a new facility.

Further research is necessary to convincingly examine the effect of an additional facility on mortality rates. The estimates might also cover up important heterogeneity across facilities based on the local disease environment and government capacity. Studies that exploit larger construction programs could help to shed light on these issues. Finally, cost estimates would be helpful to conduct a cost-benefit analysis and contribute to the debate on how to achieve universal health coverage most effectively.

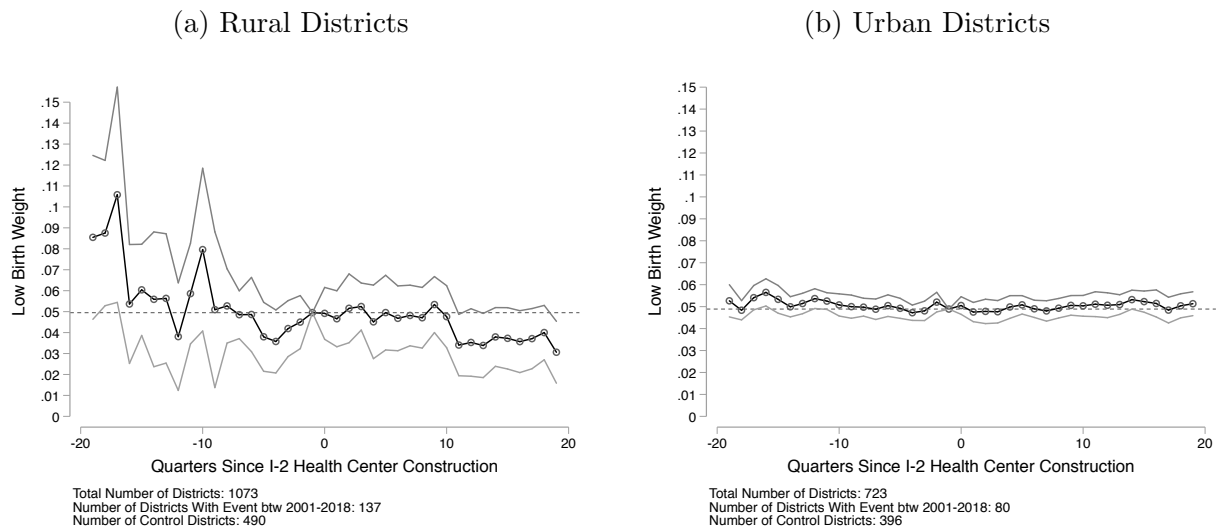
## 3.6 Tables and Figures

Figure 3.1: Number of New I-2 Facilities over Time



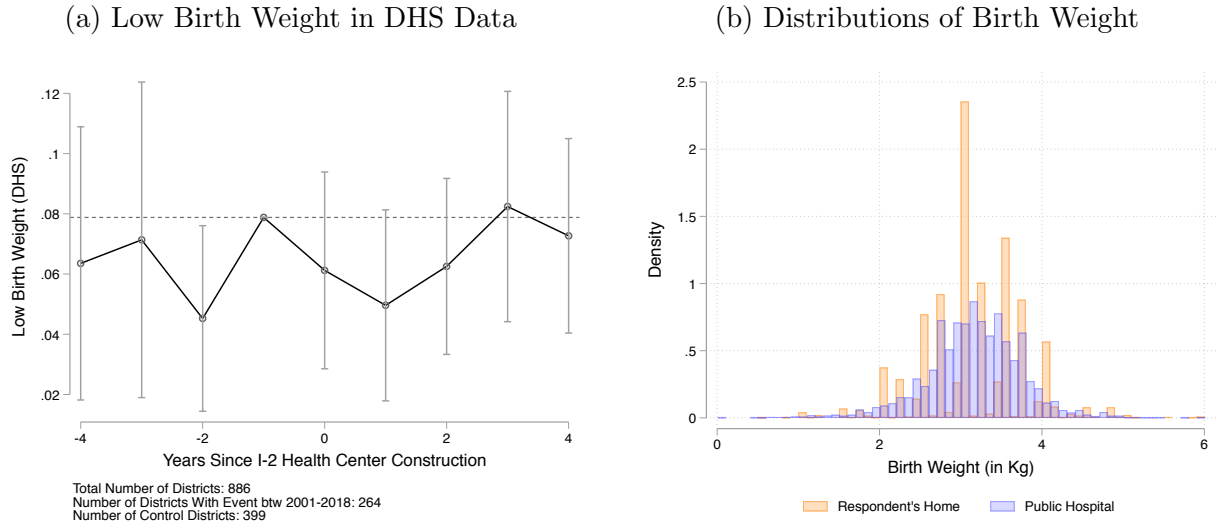
Notes: Information on opening dates are obtained from the National Registry of Health Facilities. The vertical dashed lines demarcate the period for which data from the birth registry are available (2005 - 2016).

Figure 3.2: The Effect of an Additional I-2 Facility on Low Birth Weight



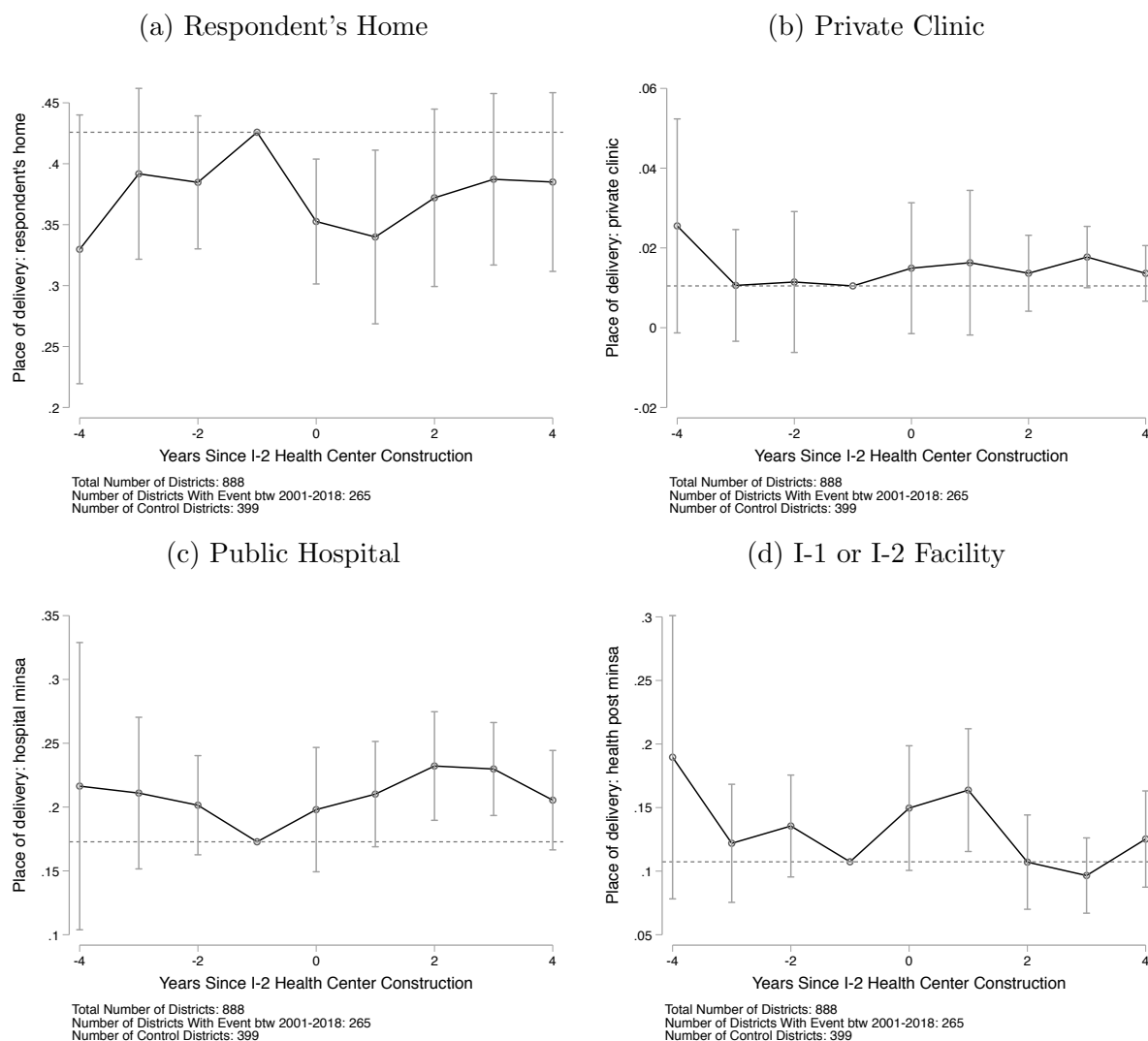
Notes: The series plots results from an event-study regression with district and quarter-year fixed effects. The omitted category is -1. Districts are weighted by the number of births. The horizontal line shows the outcome mean in 2005 in untreated districts. Each circle shows the estimated coefficient plus the outcome mean in 2005 in untreated districts. The grey solid lines represent the 95 percent confidence intervals, where standard errors are clustered at the district level. The event variables for  $t = 20$  or more and  $t = -20$  or less are included in the regressions but not shown in the figure. Districts are defined as rural if 50 percent or less of the population lives in an urban area according to the 2007 population census. Low birth weight is defined as weighing less than 2,500 grams at birth and comes from the birth registry.

Figure 3.3: Robustness Checks with DHS Data (Rural Districts)



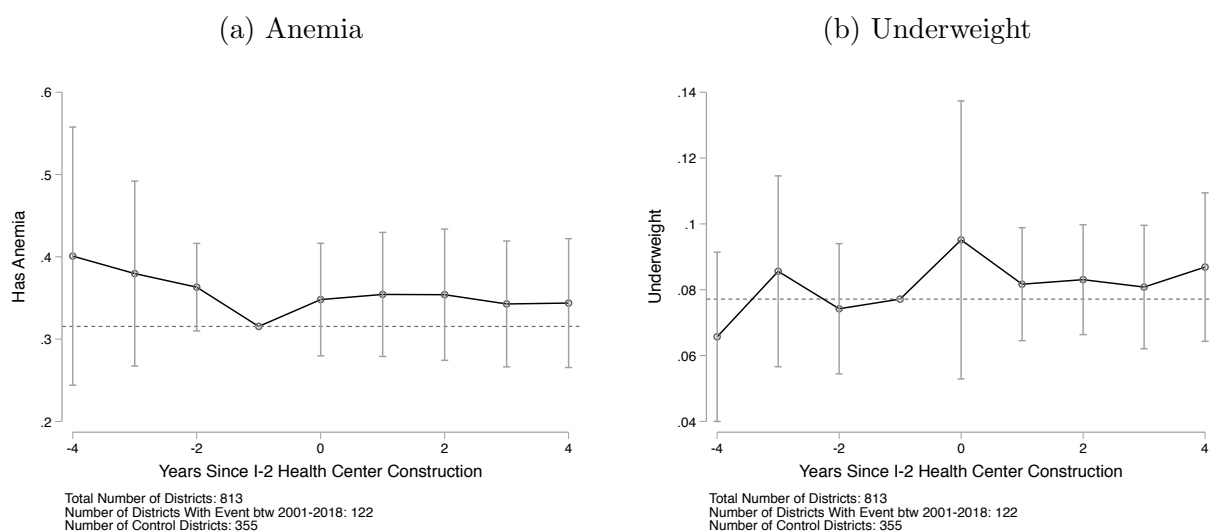
Notes: Figure 3a plots results from an event-study regression with district and year fixed effects. The omitted category is -1. Districts are weighted by the number of births. The horizontal line shows the outcome mean in 2005 in untreated districts. Each circle shows the estimated coefficient plus the outcome mean in 2005 in untreated districts. The grey solid lines represent the 95 percent confidence intervals, where standard errors are clustered at the district level. The event variables for  $t = 5$  or more and  $t = -5$  or less are included in the regressions but not shown in the figure. Figure 3b plots the distribution of birth weight by place of delivery. The sample in both figures is restricted to rural districts. Outcomes are obtained from the birth history module in the DHS survey. The year of birth ranges from 1999 to 2018. Robustness checks are shown in Appendix Table 2.

Figure 3.4: The Effect of an Additional I-2 Facility on Place of Delivery in Rural Districts



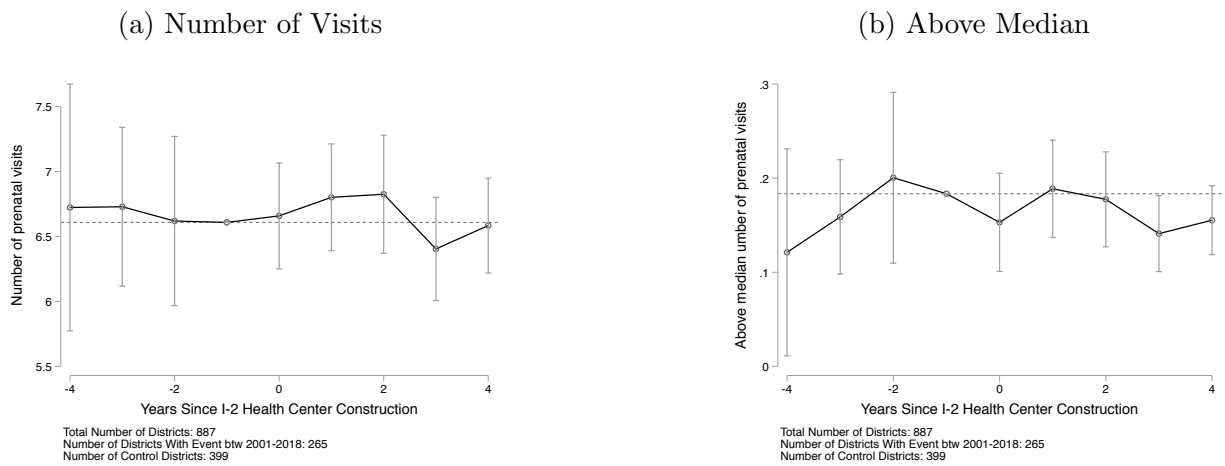
Notes: Figures plot results from an event-study regression with district and year fixed effects. The omitted category is -1. Districts are weighted by the number of births. The horizontal line shows the outcome mean in 2005 in untreated districts. Each circle shows the estimated coefficient plus the outcome mean in 2005 in untreated districts. The grey solid lines represent the 95 percent confidence intervals, where standard errors are clustered at the district level. The event variables for  $t = 5$  or more and  $t = -5$  or less are included in the regressions but not shown in the figure. The sample is restricted to rural districts. Outcomes are obtained from the birth history module in the DHS survey. The year of birth ranges from 1999 to 2018. Minsa health posts include I-1 and I-2 facilities. Robustness checks are shown in Appendix Tables 3-6.

Figure 3.5: The Effect of an Additional I-2 Facility on Respondent's Health in Rural Districts



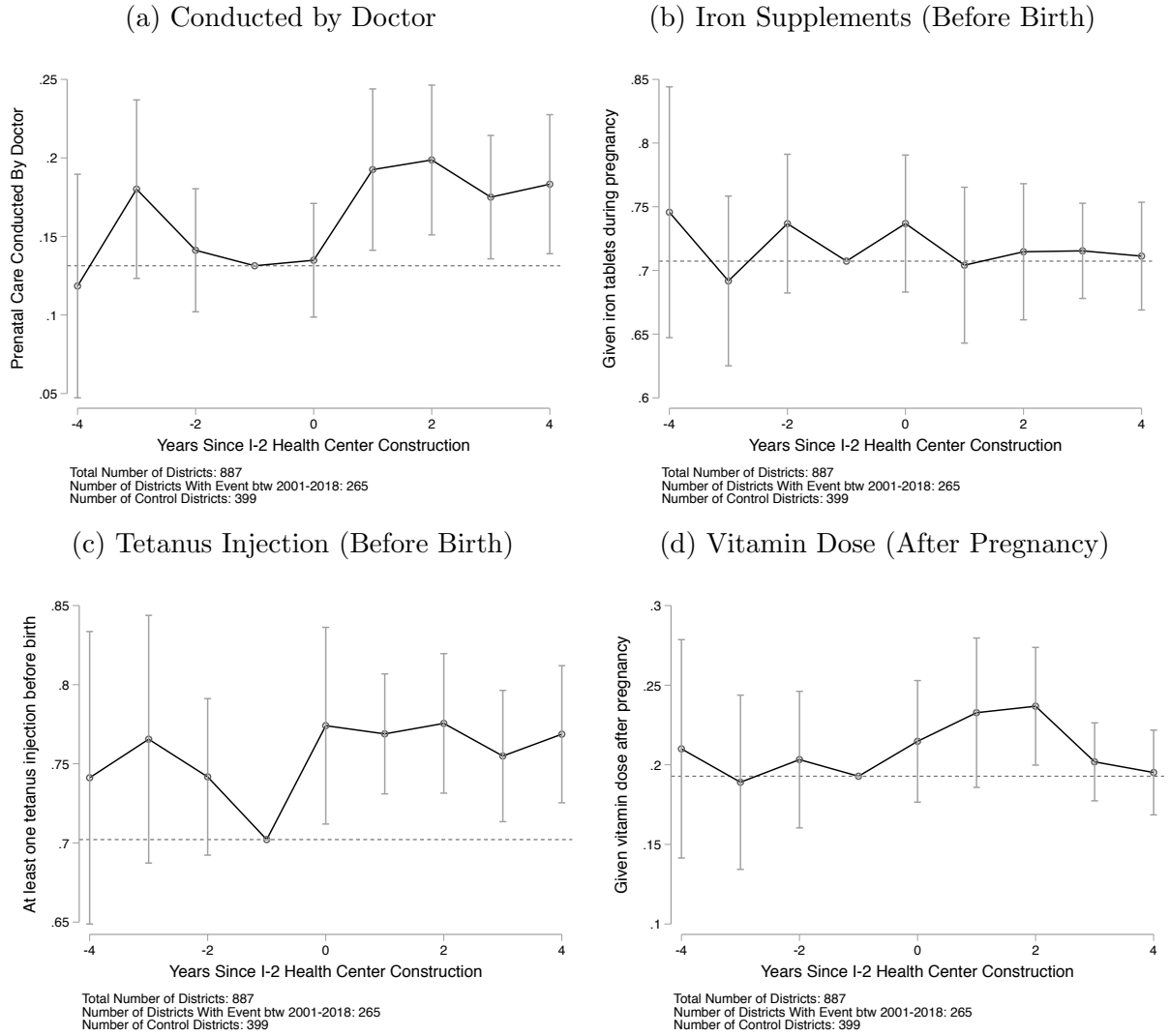
Notes: Figures plot results from an event-study regression with district and year fixed effects. The omitted category is -1. The unit of observation is a respondent. The horizontal line shows the outcome mean in 2005 in untreated districts. Each circle shows the estimated coefficient plus the outcome mean in 2005 in untreated districts. The grey solid lines represent the 95 percent confidence intervals, where standard errors are clustered at the district level. The event variables for  $t = 5 \text{ or more}$  and  $t = -5 \text{ or less}$  are included in the regressions but not shown in the figure. The sample is restricted to rural districts. Underweight is defined as having a BMI of under 18.5 and a respondent is classified as anemic if she has at least mild anemia. Outcomes are obtained from the biomarker module in the DHS 2005 - 2018 survey rounds. Robustness checks are shown in Appendix Tables 7-8.

Figure 3.6: The Effect of an Additional I-2 Facility on Prenatal Visits in Rural Districts



Notes: Figures plot results from an event-study regression with district and year fixed effects. The omitted category is -1. The unit of observation is a birth and the outcomes are obtained from the birth history in the DHS data. The median number of prenatal visits is 9. The horizontal line shows the outcome mean in 2005 in untreated districts. Each circle shows the estimated coefficient plus the outcome mean in 2005 in untreated districts. The grey solid lines represent the 95 percent confidence intervals, where standard errors are clustered at the district level. The event variables for  $t = 5$  or more and  $t = -5$  or less are included in the regressions but not shown in the figure. The sample is restricted to rural districts. The year of birth ranges from 1999 to 2018. Robustness checks are shown in Appendix Tables 9-10.

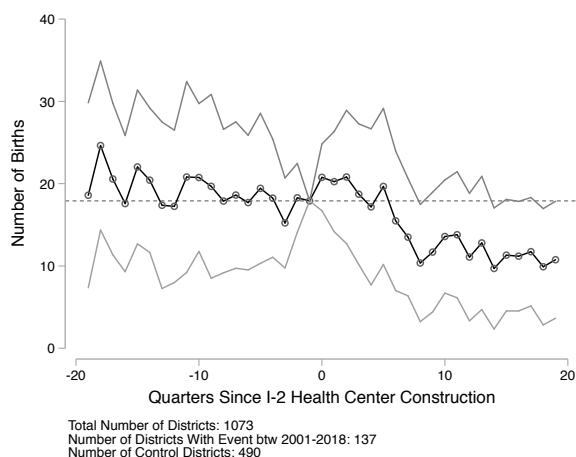
Figure 3.7: The Effect of an Additional I-2 Facility on Prenatal Care in Rural Districts



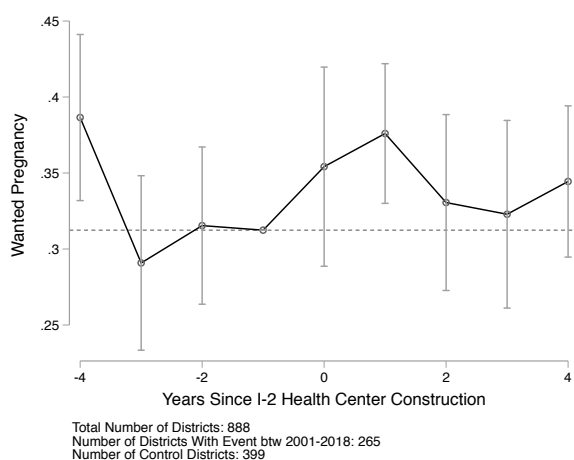
Notes: Figures plot results from an event-study regression with district and year fixed effects. The omitted category is -1. The unit of observation is a birth and the outcomes are obtained from the birth history in the DHS data. The horizontal line shows the outcome mean in 2005 in untreated districts. Each circle shows the estimated coefficient plus the outcome mean in 2005 in untreated districts. The grey solid lines represent the 95 percent confidence intervals, where standard errors are clustered at the district level. The event variables for  $t = 5$  or more and  $t = -5$  or less are included in the regressions but not shown in the figure. The sample is restricted to rural districts. The year of birth ranges from 1999 to 2018. Robustness checks are shown in Appendix Tables 11-14.

Figure 3.8: The Effect of an Additional I-2 Facility on Family Planning in Rural Districts

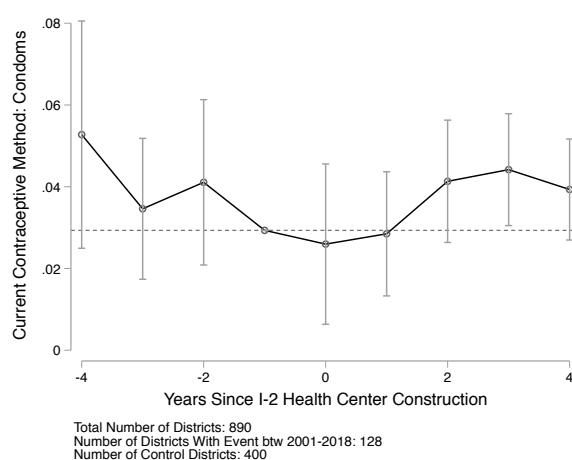
(a) Number of Births



(b) Wanted Pregnancy



(c) Condom Usage



Notes: Figures plot results from an event-study regression with district and quarter-year fixed effects (a) or district and year fixed effects (b and c). The omitted category is -1. Districts are weighted by the number of births in Figure b. The unit of observation is a district in Figures a and b and a respondent in Figure c. The horizontal line shows the outcome mean in 2005 in untreated districts. Each circle shows the estimated coefficient plus the outcome mean in 2005 in untreated districts. The grey solid lines represent the 95 percent confidence intervals, where standard errors are clustered at the district level. The event variables for  $t = 20$  or more and  $t = -20$  or less (a) or  $t = 5$  or more and  $t = -5$  or less (b and c) are included in the regressions but not shown in the figure. The sample is restricted to rural districts. The number of births is obtained from the birth registry. Information on *Wanted Pregnancy* comes from the birth history in the DHS data and information on condom usage comes from the family planning module in the DHS data. Robustness checks are shown in Appendix Tables 15-17.

Figure 3.9: The Effect of an Additional I-2 Facility on Mortality Rates in Rural Districts



Notes: Figures plot results from an event-study regression with district and year fixed effects. The omitted category is -1. Districts are weighted by the number of births. The horizontal line shows the outcome mean in 2005 in untreated districts. Each circle shows the estimated coefficient plus the outcome mean in 2005 in untreated districts. The grey solid lines represent the 95 percent confidence intervals, where standard errors are clustered at the district level. The event variables for  $t = 5$  or more and  $t = -5$  or less are included in the regressions but not shown in the figure. The number of neonatal and infant deaths comes from the Ministry of Health and the number of births comes from the birth registry. Neonatal mortality is the number of deaths of children under one month of age per 1000 live births and infant mortality is the number of deaths of children under one year of age per 1000 live births. Robustness checks are shown in Appendix Tables 18-19.

Table 3.1: Summary Statistics by Group

|                                             | Rural Districts |           |            | Urban Districts |            |            |
|---------------------------------------------|-----------------|-----------|------------|-----------------|------------|------------|
|                                             | Control         | Event     | Past Event | Control         | Event      | Past Event |
| <i>Census Variables (2007)</i>              |                 |           |            |                 |            |            |
| Total Population                            | 4,728           | 10,352*** | 6,993***   | 10,700          | 101,026*** | 29,008***  |
| Urban Share                                 | 0.273           | 0.215***  | 0.230***   | 0.740           | 0.866***   | 0.789***   |
| SIS Coverage Rate                           | 0.365           | 0.416***  | 0.380      | 0.247           | 0.174***   | 0.212***   |
| Average Wealth Index                        | -0.669          | -0.664    | -0.678     | -0.379          | 0.010***   | -0.273***  |
| Sec. School Rate                            | 0.323           | 0.285***  | 0.318      | 0.481           | 0.570***   | 0.520***   |
| <i>Infrastructure Variables (2001 - Q1)</i> |                 |           |            |                 |            |            |
| N I-2 Facilities                            | 0.000           | 1.247***  | 1.453***   | 0.000           | 2.287***   | 1.615***   |
| N Private Facilities                        | 0.022           | 0.021     | 0.020      | 1.065           | 2.851**    | 1.393      |
| N Public Hospitals                          | 0.014           | 0.048*    | 0.025      | 0.111           | 0.494***   | 0.171      |
| <i>Other Variables (2005)</i>               |                 |           |            |                 |            |            |
| Low Birth Weight Share                      | 0.052           | 0.089***  | 0.052      | 0.056           | 0.044*     | 0.046*     |
| Number of Districts                         | 490             | 146       | 448        | 398             | 87         | 252        |

Notes: Control districts are districts that do not have any I-2 facility by the fourth quarter in 2018. Event districts are districts in which at least one new I-2 facility opened between 2000 and 2018. Past event districts are districts that had at least one I-2 facility before 2000 and did not receive an additional facility between 2000 and 2018. Districts are defined as rural if 50 percent or less of the population lives in an urban area according to the 2007 population census. Census variables come from the 2007 population census. Infrastructure variables refer to the number of public healthcare facilities in the first quarter from 2000 and are obtained from the National Registry of Health Facilities. The share of births with low birth weight is obtained from the birth registry. Each value is the mean of the variable in the corresponding subsample. Asterisks are obtained from regressing the variable against indicators for being in the event or past event sample and denote significance: \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$ .

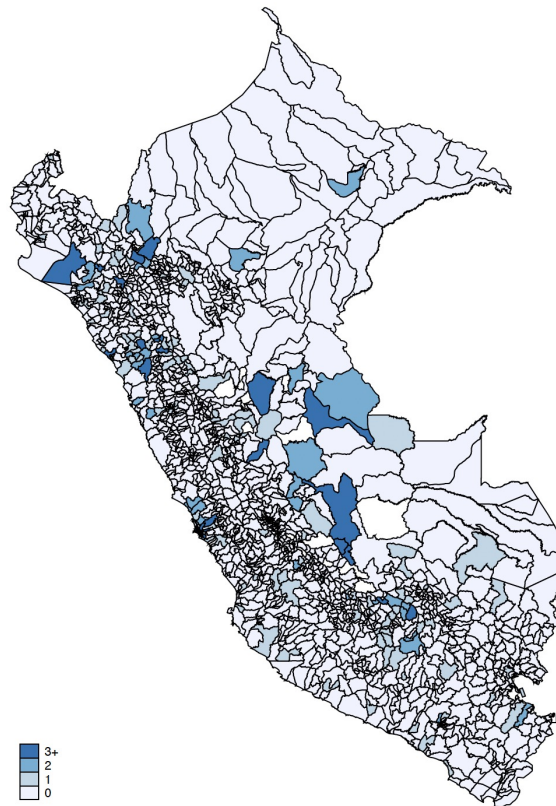
Table 3.2: The Effect of an Additional I-2 Facility on Low Birth Weight in Rural Districts

|                                       | Low Birth Weight    |                     |                     |                     |                   |
|---------------------------------------|---------------------|---------------------|---------------------|---------------------|-------------------|
|                                       | (1)                 | (2)                 | (3)                 | (4)                 | (5)               |
| 11-19 Quarters Before                 | 0.016<br>[0.010]    | 0.015<br>[0.010]    | 0.015<br>[0.010]    | 0.014<br>[0.010]    | 0.009<br>[0.008]  |
| 2-10 Quarters Before                  | -0.001<br>[0.006]   | -0.001<br>[0.006]   | -0.001<br>[0.006]   | -0.001<br>[0.006]   | 0.000<br>[0.006]  |
| 0-10 Quarters After                   | -0.001<br>[0.005]   | -0.002<br>[0.005]   | -0.002<br>[0.005]   | -0.001<br>[0.005]   | 0.001<br>[0.005]  |
| 11-19 Quarters After                  | -0.013**<br>[0.006] | -0.015**<br>[0.006] | -0.014**<br>[0.006] | -0.012**<br>[0.006] | -0.006<br>[0.006] |
| Sample                                | Rural               | Rural               | Rural               | Rural               | Rural             |
| Infrastructure Controls               | No                  | Yes                 | Yes                 | Yes                 | Yes               |
| <i>Time Trends Interacted With...</i> |                     |                     |                     |                     |                   |
| ...2007 Census Characteristics        | No                  | No                  | Yes                 | Yes                 | Yes               |
| ...Region FEs                         | No                  | No                  | No                  | Yes                 | No                |
| ...Department FEs                     | No                  | No                  | No                  | No                  | Yes               |
| 2005 Control Mean                     | 0.050               | 0.050               | 0.050               | 0.050               | 0.050             |
| R-Squared                             | 0.124               | 0.125               | 0.128               | 0.129               | 0.139             |
| Observations                          | 42,300              | 42,300              | 42,300              | 42,300              | 42,300            |

Notes: Each column comes from a separate event-study regression. All specifications include district and quarter-year fixed effects. The omitted category is -1. Districts are weighted by the number of births. Infrastructure controls include the number of I-1, I-3, I-4, II, and IIII facilities and the number of Essalud facilities in the district. The census characteristics are urban share, SIS coverage rate, average wealth index, and secondary school completion rate. Regions fixed effects indicate whether the district is located in the coastal area, in the jungle area, or in the mountain area. Standard errors are clustered at the district level. The event variables for  $t = 20$  or more and  $t = -20$  or less are included in the regression but not shown in the table. Districts are defined as rural if 50 percent or less of the population lives in an urban area according to the 2007 population census. Low birth weight is defined as weighing less than 2,500 grams at birth and comes from the birth registry.

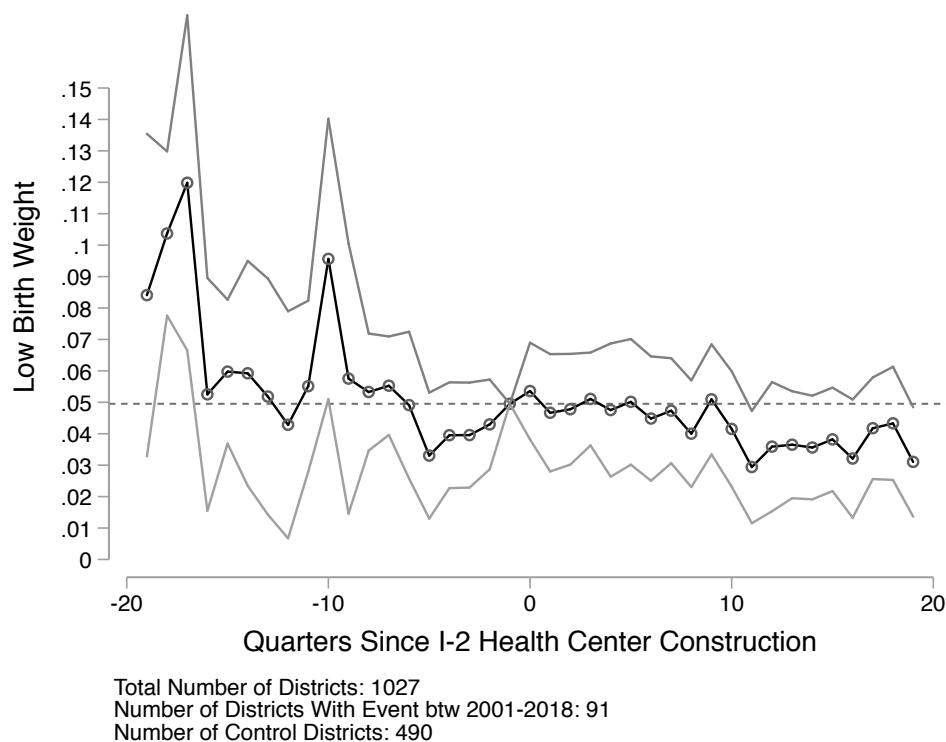
## 3.7 Appendix

Figure 3.A.1: Map of Facility Openings



Notes: The map displays each district in Peru. The color of the district indicates how many I-2 facilities opened between 2001 and 2018 in the corresponding district. The number of new facilities is top-coded at 3.

Figure 3.A.2: Effect of Facility Openings on Low Birth Weight in Rural Areas  
(Without Districts that Received a New I-2 Facility Between 2000 and 2002)



Notes: The series plots results from an event-study regression with district and quarter-year fixed effects. The omitted category is -1. Districts are weighted by the number of births. The horizontal line shows the outcome mean in 2005 in untreated districts. Each circle shows the estimated coefficient plus the outcome mean in 2005 in untreated districts. The grey solid lines represent the 95 percent confidence intervals, where standard errors are clustered at the district level. The event variables for  $t = 20$  or more and  $t = -20$  or less are included in the regressions but not shown in the figure. Districts are defined as rural if 50 percent or less of the population lives in an urban area according to the 2007 population census. Low birth weight is defined as weighing less than 2,500 grams at birth and comes from the birth registry. Districts that received a new I-2 facility between 2000 and 2002 are excluded from the sample.

Figure 3.A.3: The Effect of an Additional I-2 Facility on Delivering at Home by Wealth Status



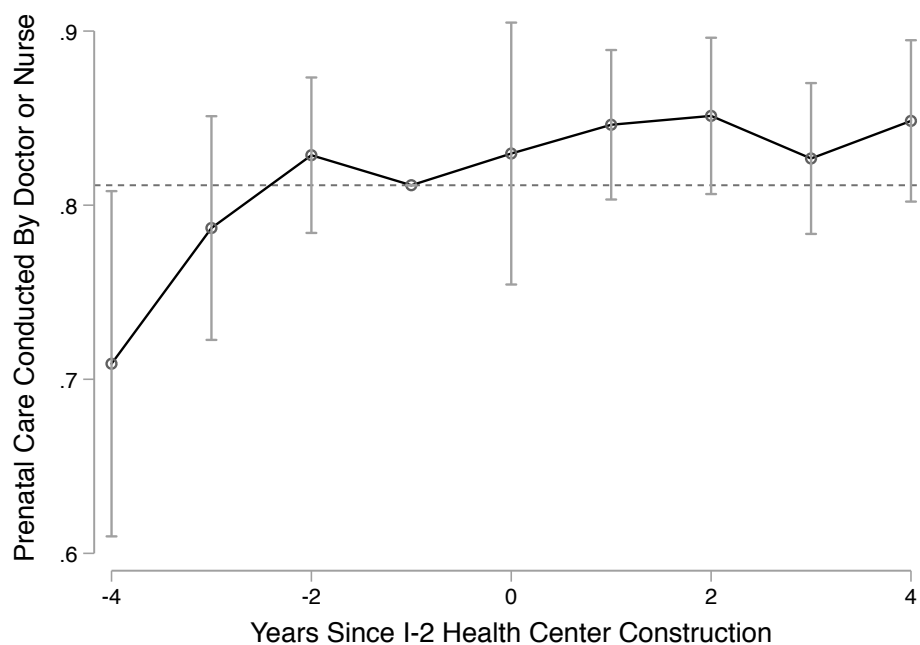
Notes: Figures plot results from an event-study regression with district and year fixed effects. The omitted category is -1. The unit of observation is a birth and the outcomes are obtained from the birth history in the DHS data. The horizontal line shows the outcome mean in 2005 in untreated districts. Each circle shows the estimated coefficient plus the outcome mean in 2005 in untreated districts. The grey solid lines represent the 95 percent confidence intervals, where standard errors are clustered at the district level. The event variables for  $t = 5$  or more and  $t = -5$  or less are included in the regressions but not shown in the figure. The sample is restricted to rural districts.

Figure 3.A.4: The Effect of an Additional I-2 Facility on Delivering at a Public Hospital by Wealth Status



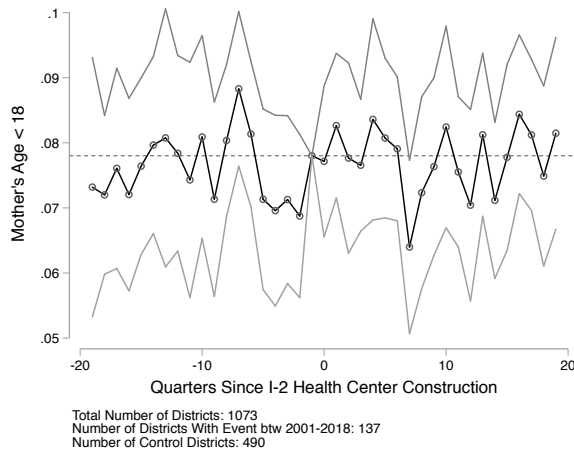
Notes: Figures plot results from an event-study regression with district and year fixed effects. The omitted category is -1. The unit of observation is a birth and the outcomes are obtained from the birth history in the DHS data. The horizontal line shows the outcome mean in 2005 in untreated districts. Each circle shows the estimated coefficient plus the outcome mean in 2005 in untreated districts. The grey solid lines represent the 95 percent confidence intervals, where standard errors are clustered at the district level. The event variables for  $t = 5$  or more and  $t = -5$  or less are included in the regressions but not shown in the figure. The sample is restricted to rural districts.

Figure 3.A.5: The Effect of an Additional I-2 Facility on Prenatal Care in Rural Districts

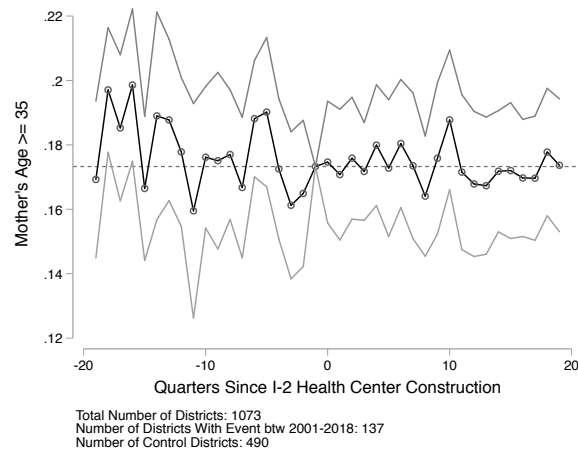


Total Number of Districts: 887  
Number of Districts With Event btw 2001-2018: 265  
Number of Control Districts: 399

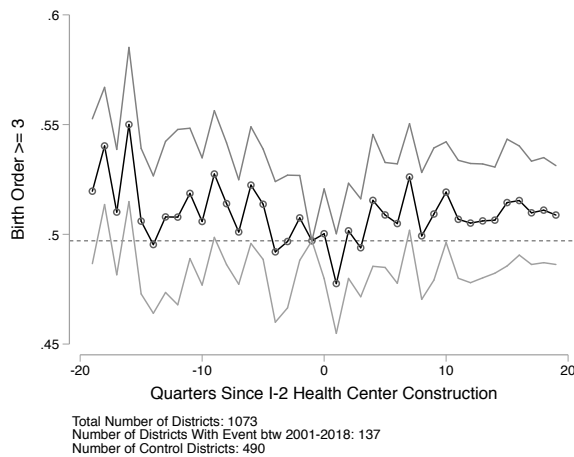
Notes: Figures plot results from an event-study regression with district and year fixed effects. The omitted category is -1. The unit of observation is a birth and the outcomes are obtained from the birth history in the DHS data. The horizontal line shows the outcome mean in 2005 in untreated districts. Each circle shows the estimated coefficient plus the outcome mean in 2005 in untreated districts. The grey solid lines represent the 95 percent confidence intervals, where standard errors are clustered at the district level. The event variables for  $t = 5$  or more and  $t = -5$  or less are included in the regressions but not shown in the figure. The sample is restricted to rural districts. The year of birth ranges from 1999 to 2018.



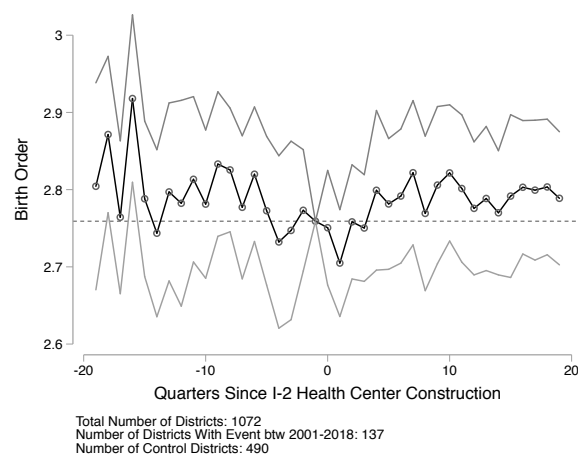
(a) Mother's Age < 18



(b) Mother's Age  $\geq 35$



(c) Birth Order  $\geq 3$



(d) Average Birth Order

Figure 3.A.6: The Effect of an Additional I-2 Facility on Family Planning in Rural Districts  
Notes: Figures plot results from an event-study regression with quarter-year fixed effects. The omitted category is -1. Districts are weighted by the number of births. The horizontal line shows the outcome mean in 2005 in untreated districts. Each circle shows the estimated coefficient plus the outcome mean in 2005 in untreated districts. The grey solid lines represent the 95 percent confidence intervals, where standard errors are clustered at the district level. The event variables for  $t = 20$  or more and  $t = -20$  or less are included in the regressions but not shown in the figure. The sample is restricted to rural districts.

Table 3.A.1: Count of Districts by Number of I-2 Facilities in 2000 And 2018

| 2001<br>- Q1 | 2018 - Q4 |     |     |    |    |    |    |   |   |   |    |    |    |    |    |    |    |    | Total |
|--------------|-----------|-----|-----|----|----|----|----|---|---|---|----|----|----|----|----|----|----|----|-------|
|              | 0         | 1   | 2   | 3  | 4  | 5  | 6  | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 18 | 19 | 22 |       |
| 0            | 890       | 58  | 7   | 2  | 1  | 0  | 0  | 0 | 0 | 0 | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 958   |
| 1            | 0         | 508 | 50  | 3  | 6  | 2  | 0  | 0 | 0 | 0 | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 569   |
| 2            | 0         | 0   | 158 | 14 | 8  | 1  | 0  | 0 | 0 | 0 | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 181   |
| 3            | 0         | 0   | 0   | 47 | 9  | 4  | 0  | 0 | 0 | 0 | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 60    |
| 4            | 0         | 0   | 0   | 0  | 12 | 1  | 4  | 1 | 0 | 0 | 0  | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 19    |
| 5            | 0         | 0   | 0   | 0  | 0  | 5  | 0  | 1 | 0 | 0 | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 6     |
| 6            | 0         | 0   | 0   | 0  | 0  | 0  | 8  | 1 | 2 | 2 | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 14    |
| 7            | 0         | 0   | 0   | 0  | 0  | 0  | 0  | 2 | 1 | 0 | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 4     |
| 8            | 0         | 0   | 0   | 0  | 0  | 0  | 0  | 0 | 2 | 3 | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 5     |
| 9            | 0         | 0   | 0   | 0  | 0  | 0  | 0  | 0 | 0 | 2 | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 3     |
| 10           | 0         | 0   | 0   | 0  | 0  | 0  | 0  | 0 | 0 | 0 | 2  | 0  | 1  | 0  | 0  | 0  | 0  | 0  | 3     |
| 11           | 0         | 0   | 0   | 0  | 0  | 0  | 0  | 0 | 0 | 0 | 0  | 1  | 0  | 1  | 0  | 0  | 0  | 0  | 2     |
| 12           | 0         | 0   | 0   | 0  | 0  | 0  | 0  | 0 | 0 | 0 | 0  | 0  | 0  | 0  | 1  | 0  | 0  | 0  | 1     |
| 13           | 0         | 0   | 0   | 0  | 0  | 0  | 0  | 0 | 0 | 0 | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 0  | 1     |
| 14           | 0         | 0   | 0   | 0  | 0  | 0  | 0  | 0 | 0 | 0 | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 1     |
| 18           | 0         | 0   | 0   | 0  | 0  | 0  | 0  | 0 | 0 | 0 | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 1     |
| Total        | 890       | 566 | 215 | 66 | 36 | 13 | 12 | 5 | 5 | 7 | 5  | 2  | 1  | 1  | 1  | 1  | 1  | 1  | 1,828 |

Notes: Each value is the number of districts in the corresponding bin. The first column indicates the number of I-2 facilities the district had in the first quarter of 2001. The first row indicates the number of I-2 facilities the district had in the fourt quarter of 2018. Information on opening dates are obtained from the National Registry of Health Facilities.

Table 3.A.2: Robustness Check: Low Birth Weight (DHS)

|                                       | Low Birth Weight (DHS) |                   |                   |                   |                   |                   |
|---------------------------------------|------------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
|                                       | (1)                    | (2)               | (3)               | (4)               | (5)               | (6)               |
| 2-4 Years Before                      | -0.021<br>[0.018]      | -0.021<br>[0.017] | -0.021<br>[0.017] | -0.020<br>[0.017] | -0.020<br>[0.017] | -0.020<br>[0.018] |
| 0-2 Years After                       | -0.022<br>[0.013]      | -0.020<br>[0.013] | -0.020<br>[0.014] | -0.021<br>[0.014] | -0.020<br>[0.014] | -0.021<br>[0.015] |
| 3-4 Years After                       | -0.003<br>[0.016]      | -0.002<br>[0.016] | -0.003<br>[0.018] | -0.003<br>[0.018] | -0.004<br>[0.018] | -0.004<br>[0.018] |
| Sample                                | Rural                  | Rural             | Rural             | Rural             | Rural             | Rural             |
| Household Controls                    | No                     | Yes               | Yes               | Yes               | Yes               | Yes               |
| Infrastructure Controls               | No                     | No                | Yes               | Yes               | Yes               | Yes               |
| <i>Time Trends Interacted With...</i> |                        |                   |                   |                   |                   |                   |
| ...2007 Census Characteristics        | No                     | No                | No                | Yes               | Yes               | Yes               |
| ...Region FEs                         | No                     | No                | No                | No                | Yes               | No                |
| ...Department FEs                     | No                     | No                | No                | No                | No                | Yes               |
| 2005 Control Mean                     | 0.079                  | 0.079             | 0.079             | 0.079             | 0.079             | 0.079             |
| R-Squared                             | 0.039                  | 0.040             | 0.040             | 0.040             | 0.040             | 0.041             |
| Observations                          | 48,343                 | 48,154            | 48,154            | 48,154            | 48,154            | 48,154            |

Notes: Each column comes from a separate event-study regression. All specifications include district and year fixed effects. The omitted category is -1. The unit of observation is a birth in the birth history module in the DHS survey. The household controls consist of the years of education of the head of the household, gender of the head of the household, age of the head of the household, urban/rural indicator, and wealth index. Infrastructure controls include the number of I-1, I-3, I-4, II, and III facilities and the number of Essalud facilities in the district. The census characteristics are urban share, SIS coverage rate, average wealth index, and secondary school completion rate. Regions fixed effects indicate whether the district is located in the coastal area, in the jungle area, or in the mountain area. Standard errors are clustered at the district level. The event variables for  $t = 5$  or more and  $t = -25$  or less are included in the regression but not shown in the table. Districts are defined as rural if 50 percent or less of the population lives in an urban area according to the 2007 population census. Low birth weight is defined as weighing less than 2,500 grams at birth.

Table 3.A.3: Robustness Check: Place of delivery - respondent's home

|                                       | Place of delivery: respondent's home |                     |                     |                     |                     |                     |
|---------------------------------------|--------------------------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
|                                       | (1)                                  | (2)                 | (3)                 | (4)                 | (5)                 | (6)                 |
| 2-4 Years Before                      | -0.047*<br>[0.027]                   | -0.047*<br>[0.025]  | -0.049*<br>[0.025]  | -0.051**<br>[0.023] | -0.051**<br>[0.023] | -0.053**<br>[0.022] |
| 0-2 Years After                       | -0.068**<br>[0.031]                  | -0.061**<br>[0.031] | -0.064**<br>[0.030] | -0.068**<br>[0.029] | -0.068**<br>[0.029] | -0.071**<br>[0.028] |
| 3-4 Years After                       | -0.038<br>[0.034]                    | -0.034<br>[0.031]   | -0.036<br>[0.030]   | -0.046<br>[0.029]   | -0.048*<br>[0.029]  | -0.052*<br>[0.029]  |
| Sample                                | Rural                                | Rural               | Rural               | Rural               | Rural               | Rural               |
| Household Controls                    | No                                   | Yes                 | Yes                 | Yes                 | Yes                 | Yes                 |
| Infrastructure Controls               | No                                   | No                  | Yes                 | Yes                 | Yes                 | Yes                 |
| <i>Time Trends Interacted With...</i> |                                      |                     |                     |                     |                     |                     |
| ...2007 Census Characteristics        | No                                   | No                  | No                  | Yes                 | Yes                 | Yes                 |
| ...Region FEs                         | No                                   | No                  | No                  | No                  | Yes                 | No                  |
| ...Department FEs                     | No                                   | No                  | No                  | No                  | No                  | Yes                 |
| 2005 Control Mean                     | 0.426                                | 0.426               | 0.426               | 0.426               | 0.426               | 0.426               |
| R-Squared                             | 0.348                                | 0.369               | 0.370               | 0.372               | 0.373               | 0.378               |
| Observations                          | 55,486                               | 55,259              | 55,259              | 55,259              | 55,259              | 55,259              |

Notes: Each column comes from a separate event-study regression. All specifications include district and year fixed effects. The omitted category is -1. The unit of observation is a birth in the birth history module in the DHS survey. The household controls consist of the years of education of the head of the household, gender of the head of the household, age of the head of the household, urban/rural indicator, and wealth index. Infrastructure controls include the number of I-1, I-3, I-4, II, and III facilities and the number of Essalud facilities in the district. The census characteristics are urban share, SIS coverage rate, average wealth index, and secondary school completion rate. Regions fixed effects indicate whether the district is located in the coastal area, in the jungle area, or in the mountain area. Standard errors are clustered at the district level. The event variables for  $t = 5$  or more and  $t = -25$  or less are included in the regression but not shown in the table. Districts are defined as rural if 50 percent or less of the population lives in an urban area according to the 2007 population census.

Table 3.A.4: Robustness Check: Place of delivery - private clinic

|                                       | Place of delivery: private clinic |                  |                   |                   |                   |                  |
|---------------------------------------|-----------------------------------|------------------|-------------------|-------------------|-------------------|------------------|
|                                       | (1)                               | (2)              | (3)               | (4)               | (5)               | (6)              |
| 2-4 Years Before                      | 0.003<br>[0.008]                  | 0.003<br>[0.009] | 0.001<br>[0.009]  | 0.001<br>[0.009]  | 0.001<br>[0.009]  | 0.001<br>[0.009] |
| 0-2 Years After                       | 0.004<br>[0.006]                  | 0.003<br>[0.006] | 0.000<br>[0.009]  | -0.001<br>[0.009] | -0.001<br>[0.009] | 0.001<br>[0.009] |
| 3-4 Years After                       | 0.004<br>[0.003]                  | 0.004<br>[0.003] | -0.001<br>[0.007] | -0.003<br>[0.007] | -0.003<br>[0.007] | 0.001<br>[0.007] |
| Sample                                | Rural                             | Rural            | Rural             | Rural             | Rural             | Rural            |
| Household Controls                    | No                                | Yes              | Yes               | Yes               | Yes               | Yes              |
| Infrastructure Controls               | No                                | No               | Yes               | Yes               | Yes               | Yes              |
| <i>Time Trends Interacted With...</i> |                                   |                  |                   |                   |                   |                  |
| ...2007 Census Characteristics        | No                                | No               | No                | Yes               | Yes               | Yes              |
| ...Region FEs                         | No                                | No               | No                | No                | Yes               | No               |
| ...Department FEs                     | No                                | No               | No                | No                | No                | Yes              |
| 2005 Control Mean                     | 0.010                             | 0.010            | 0.010             | 0.010             | 0.010             | 0.010            |
| R-Squared                             | 0.079                             | 0.087            | 0.087             | 0.089             | 0.089             | 0.091            |
| Observations                          | 55,486                            | 55,259           | 55,259            | 55,259            | 55,259            | 55,259           |

Notes: Each column comes from a separate event-study regression. All specifications include district and year fixed effects. The omitted category is -1. The unit of observation is a birth in the birth history module in the DHS survey. The household controls consist of the years of education of the head of the household, gender of the head of the household, age of the head of the household, urban/rural indicator, and wealth index. Infrastructure controls include the number of I-1, I-3, I-4, II, and III facilities and the number of Essalud facilities in the district. The census characteristics are urban share, SIS coverage rate, average wealth index, and secondary school completion rate. Regions fixed effects indicate whether the district is located in the coastal area, in the jungle area, or in the mountain area. Standard errors are clustered at the district level. The event variables for  $t = 5$  or more and  $t = -25$  or less are included in the regression but not shown in the table. Districts are defined as rural if 50 percent or less of the population lives in an urban area according to the 2007 population census.

Table 3.A.5: Robustness Check: Place of delivery - hospital minsa

|                                       | Place of delivery: hospital minsa |                    |                    |                    |                    |                    |
|---------------------------------------|-----------------------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
|                                       | (1)                               | (2)                | (3)                | (4)                | (5)                | (6)                |
| 2-4 Years Before                      | 0.033*<br>[0.018]                 | 0.031*<br>[0.018]  | 0.031*<br>[0.018]  | 0.029<br>[0.018]   | 0.029<br>[0.018]   | 0.030<br>[0.019]   |
| 0-2 Years After                       | 0.042**<br>[0.020]                | 0.039**<br>[0.019] | 0.047**<br>[0.024] | 0.049**<br>[0.024] | 0.050**<br>[0.023] | 0.047**<br>[0.023] |
| 3-4 Years After                       | 0.041**<br>[0.019]                | 0.039**<br>[0.018] | 0.052**<br>[0.026] | 0.056**<br>[0.026] | 0.060**<br>[0.025] | 0.054**<br>[0.025] |
| Sample                                | Rural                             | Rural              | Rural              | Rural              | Rural              | Rural              |
| Household Controls                    | No                                | Yes                | Yes                | Yes                | Yes                | Yes                |
| Infrastructure Controls               | No                                | No                 | Yes                | Yes                | Yes                | Yes                |
| <i>Time Trends Interacted With...</i> |                                   |                    |                    |                    |                    |                    |
| ...2007 Census Characteristics        | No                                | No                 | No                 | Yes                | Yes                | Yes                |
| ...Region FEs                         | No                                | No                 | No                 | No                 | Yes                | No                 |
| ...Department FEs                     | No                                | No                 | No                 | No                 | No                 | Yes                |
| 2005 Control Mean                     | 0.173                             | 0.173              | 0.173              | 0.173              | 0.173              | 0.173              |
| R-Squared                             | 0.211                             | 0.217              | 0.218              | 0.218              | 0.220              | 0.222              |
| Observations                          | 55,486                            | 55,259             | 55,259             | 55,259             | 55,259             | 55,259             |

Notes: Each column comes from a separate event-study regression. All specifications include district and year fixed effects. The omitted category is -1. The unit of observation is a birth in the birth history module in the DHS survey. The household controls consist of the years of education of the head of the household, gender of the head of the household, age of the head of the household, urban/rural indicator, and wealth index. Infrastructure controls include the number of I-1, I-3, I-4, II, and III facilities and the number of Essalud facilities in the district. The census characteristics are urban share, SIS coverage rate, average wealth index, and secondary school completion rate. Regions fixed effects indicate whether the district is located in the coastal area, in the jungle area, or in the mountain area. Standard errors are clustered at the district level. The event variables for  $t = 5$  or more and  $t = -25$  or less are included in the regression but not shown in the table. Districts are defined as rural if 50 percent or less of the population lives in an urban area according to the 2007 population census.

Table 3.A.6: Robustness Check: Place of delivery - health post minsa

|                                       | Place of delivery: health post minsa |                   |                   |                   |                   |                   |
|---------------------------------------|--------------------------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
|                                       | (1)                                  | (2)               | (3)               | (4)               | (5)               | (6)               |
| 2-4 Years Before                      | 0.032*<br>[0.018]                    | 0.031*<br>[0.018] | 0.030*<br>[0.018] | 0.029*<br>[0.018] | 0.030*<br>[0.018] | 0.031*<br>[0.018] |
| 0-2 Years After                       | 0.029<br>[0.019]                     | 0.030<br>[0.019]  | 0.029<br>[0.021]  | 0.029<br>[0.021]  | 0.027<br>[0.021]  | 0.033<br>[0.020]  |
| 3-4 Years After                       | 0.004<br>[0.016]                     | 0.005<br>[0.016]  | 0.003<br>[0.020]  | 0.002<br>[0.020]  | -0.001<br>[0.020] | 0.006<br>[0.019]  |
| Sample                                | Rural                                | Rural             | Rural             | Rural             | Rural             | Rural             |
| Household Controls                    | No                                   | Yes               | Yes               | Yes               | Yes               | Yes               |
| Infrastructure Controls               | No                                   | No                | Yes               | Yes               | Yes               | Yes               |
| <i>Time Trends Interacted With...</i> |                                      |                   |                   |                   |                   |                   |
| ...2007 Census Characteristics        | No                                   | No                | No                | Yes               | Yes               | Yes               |
| ...Region FEs                         | No                                   | No                | No                | No                | Yes               | No                |
| ...Department FEs                     | No                                   | No                | No                | No                | No                | Yes               |
| 2005 Control Mean                     | 0.107                                | 0.107             | 0.107             | 0.107             | 0.107             | 0.107             |
| R-Squared                             | 0.151                                | 0.154             | 0.154             | 0.155             | 0.157             | 0.163             |
| Observations                          | 55,486                               | 55,259            | 55,259            | 55,259            | 55,259            | 55,259            |

Notes: Each column comes from a separate event-study regression. Minsa health posts include I-1 and I-2 facilities. All specifications include district and year fixed effects. The omitted category is -1. The unit of observation is a birth in the birth history module in the DHS survey. The household controls consist of the years of education of the head of the household, gender of the head of the household, age of the head of the household, urban/rural indicator, and wealth index. Infrastructure controls include the number of I-1, I-3, I-4, II, and III facilities and the number of Essalud facilities in the district. The census characteristics are urban share, SIS coverage rate, average wealth index, and secondary school completion rate. Regions fixed effects indicate whether the district is located in the coastal area, in the jungle area, or in the mountain area. Standard errors are clustered at the district level. The event variables for  $t = 5$  or more and  $t = -25$  or less are included in the regression but not shown in the table. Districts are defined as rural if 50 percent or less of the population lives in an urban area according to the 2007 population census.

Table 3.A.7: Robustness Check: Underweight

|                                       | Underweight       |                   |                    |                    |                    |                     |
|---------------------------------------|-------------------|-------------------|--------------------|--------------------|--------------------|---------------------|
|                                       | (1)               | (2)               | (3)                | (4)                | (5)                | (6)                 |
| 2-4 Years Before                      | -0.000<br>[0.010] | -0.009<br>[0.011] | -0.010<br>[0.011]  | -0.009<br>[0.010]  | -0.008<br>[0.010]  | -0.006<br>[0.013]   |
| 0-2 Years After                       | 0.009<br>[0.010]  | 0.012<br>[0.012]  | 0.025<br>[0.016]   | 0.026*<br>[0.015]  | 0.027*<br>[0.015]  | 0.034**<br>[0.014]  |
| 3-4 Years After                       | 0.007<br>[0.010]  | 0.025<br>[0.017]  | 0.054**<br>[0.026] | 0.058**<br>[0.025] | 0.057**<br>[0.024] | 0.062***<br>[0.021] |
| Sample                                | Rural             | Rural             | Rural              | Rural              | Rural              | Rural               |
| Household Controls                    | No                | Yes               | Yes                | Yes                | Yes                | Yes                 |
| Infrastructure Controls               | No                | No                | Yes                | Yes                | Yes                | Yes                 |
| <i>Time Trends Interacted With...</i> |                   |                   |                    |                    |                    |                     |
| ...2007 Census Characteristics        | No                | No                | No                 | Yes                | Yes                | Yes                 |
| ...Region FEs                         | No                | No                | No                 | No                 | Yes                | No                  |
| ...Department FEs                     | No                | No                | No                 | No                 | No                 | Yes                 |
| 2005 Control Mean                     | 0.077             | 0.077             | 0.077              | 0.077              | 0.077              | 0.077               |
| R-Squared                             | 0.046             | 0.062             | 0.063              | 0.063              | 0.063              | 0.065               |
| Observations                          | 73,555            | 28,559            | 28,559             | 28,559             | 28,559             | 28,559              |

Notes: Each column comes from a separate event-study regression. All specifications include district and year fixed effects. The omitted category is -1. The unit of observation is a respondent in the biomarker module in the DHS survey. The household controls consist of the years of education of the head of the household, gender of the head of the household, age of the head of the household, urban/rural indicator, and wealth index. Infrastructure controls include the number of I-1, I-3, I-4, II, and III facilities and the number of Essalud facilities in the district. The census characteristics are urban share, SIS coverage rate, average wealth index, and secondary school completion rate. Regions fixed effects indicate whether the district is located in the coastal area, in the jungle area, or in the mountain area. Standard errors are clustered at the district level. The event variables for  $t = 5$  or more and  $t = -25$  or less are included in the regression but not shown in the table. Districts are defined as rural if 50 percent or less of the population lives in an urban area according to the 2007 population census.

Table 3.A.8: Robustness Check: Anemia

|                                       | Has Anemia       |                  |                   |                   |                   |                  |
|---------------------------------------|------------------|------------------|-------------------|-------------------|-------------------|------------------|
|                                       | (1)              | (2)              | (3)               | (4)               | (5)               | (6)              |
| 2-4 Years Before                      | 0.058<br>[0.041] | 0.068<br>[0.045] | 0.069<br>[0.044]  | 0.074<br>[0.046]  | 0.070<br>[0.044]  | 0.066<br>[0.043] |
| 0-2 Years After                       | 0.035<br>[0.035] | 0.094<br>[0.058] | 0.088*<br>[0.052] | 0.093*<br>[0.052] | 0.091*<br>[0.051] | 0.071<br>[0.050] |
| 3-4 Years After                       | 0.025<br>[0.036] | 0.100<br>[0.066] | 0.049<br>[0.068]  | 0.064<br>[0.068]  | 0.063<br>[0.065]  | 0.035<br>[0.058] |
| Sample                                | Rural            | Rural            | Rural             | Rural             | Rural             | Rural            |
| Household Controls                    | No               | Yes              | Yes               | Yes               | Yes               | Yes              |
| Infrastructure Controls               | No               | No               | Yes               | Yes               | Yes               | Yes              |
| <i>Time Trends Interacted With...</i> |                  |                  |                   |                   |                   |                  |
| ...2007 Census Characteristics        | No               | No               | No                | Yes               | Yes               | Yes              |
| ...Region FEs                         | No               | No               | No                | No                | Yes               | No               |
| ...Department FEs                     | No               | No               | No                | No                | No                | Yes              |
| 2005 Control Mean                     | 0.315            | 0.315            | 0.315             | 0.315             | 0.315             | 0.315            |
| R-Squared                             | 0.063            | 0.101            | 0.102             | 0.102             | 0.102             | 0.105            |
| Observations                          | 69,779           | 25,186           | 25,186            | 25,186            | 25,186            | 25,186           |

Notes: Each column comes from a separate event-study regression. All specifications include district and year fixed effects. The omitted category is -1. The unit of observation is a respondent in the biomarker module in the DHS survey. The household controls consist of the years of education of the head of the household, gender of the head of the household, age of the head of the household, urban/rural indicator, and wealth index. Infrastructure controls include the number of I-1, I-3, I-4, II, and III facilities and the number of Essalud facilities in the district. The census characteristics are urban share, SIS coverage rate, average wealth index, and secondary school completion rate. Regions fixed effects indicate whether the district is located in the coastal area, in the jungle area, or in the mountain area. Standard errors are clustered at the district level. The event variables for  $t = 5$  or more and  $t = -25$  or less are included in the regression but not shown in the table. Districts are defined as rural if 50 percent or less of the population lives in an urban area according to the 2007 population census.

Table 3.A.9: Robustness Check: Number of Prenatal Visits

|                                       | Number of prenatal visits |                   |                   |                   |                   |                   |
|---------------------------------------|---------------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
|                                       | (1)                       | (2)               | (3)               | (4)               | (5)               | (6)               |
| 2-4 Years Before                      | 0.055<br>[0.312]          | 0.060<br>[0.302]  | 0.049<br>[0.297]  | 0.063<br>[0.285]  | 0.070<br>[0.284]  | 0.059<br>[0.290]  |
| 0-2 Years After                       | 0.156<br>[0.193]          | 0.095<br>[0.190]  | 0.093<br>[0.201]  | 0.105<br>[0.195]  | 0.098<br>[0.197]  | 0.093<br>[0.198]  |
| 3-4 Years After                       | -0.110<br>[0.186]         | -0.143<br>[0.179] | -0.152<br>[0.198] | -0.113<br>[0.192] | -0.133<br>[0.195] | -0.155<br>[0.198] |
| Sample                                | Rural                     | Rural             | Rural             | Rural             | Rural             | Rural             |
| Household Controls                    | No                        | Yes               | Yes               | Yes               | Yes               | Yes               |
| Infrastructure Controls               | No                        | No                | Yes               | Yes               | Yes               | Yes               |
| <i>Time Trends Interacted With...</i> |                           |                   |                   |                   |                   |                   |
| ...2007 Census Characteristics        | No                        | No                | No                | Yes               | Yes               | Yes               |
| ...Region FEs                         | No                        | No                | No                | No                | Yes               | No                |
| ...Department FEs                     | No                        | No                | No                | No                | No                | Yes               |
| 2005 Control Mean                     | 6.608                     | 6.608             | 6.608             | 6.608             | 6.608             | 6.608             |
| R-Squared                             | 0.185                     | 0.203             | 0.203             | 0.205             | 0.205             | 0.208             |
| Observations                          | 46,956                    | 46,769            | 46,769            | 46,769            | 46,769            | 46,769            |

Notes: Each column comes from a separate event-study regression. All specifications include district and year fixed effects. The omitted category is -1. The unit of observation is a birth in the birth history module in the DHS survey. The household controls consist of the years of education of the head of the household, gender of the head of the household, age of the head of the household, urban/rural indicator, and wealth index. Infrastructure controls include the number of I-1, I-3, I-4, II, and III facilities and the number of Essalud facilities in the district. The census characteristics are urban share, SIS coverage rate, average wealth index, and secondary school completion rate. Regions fixed effects indicate whether the district is located in the coastal area, in the jungle area, or in the mountain area. Standard errors are clustered at the district level. The event variables for  $t = 5$  or more and  $t = -25$  or less are included in the regression but not shown in the table. Districts are defined as rural if 50 percent or less of the population lives in an urban area according to the 2007 population census.

Table 3.A.10: Robustness Check: Above-Median Number of Prenatal Visits

|                                       | Above median umber of prenatal visits |                    |                   |                   |                   |                   |
|---------------------------------------|---------------------------------------|--------------------|-------------------|-------------------|-------------------|-------------------|
|                                       | (1)                                   | (2)                | (3)               | (4)               | (5)               | (6)               |
| 2-4 Years Before                      | -0.009<br>[0.038]                     | -0.008<br>[0.037]  | -0.009<br>[0.036] | -0.007<br>[0.035] | -0.005<br>[0.034] | -0.008<br>[0.034] |
| 0-2 Years After                       | -0.007<br>[0.023]                     | -0.012<br>[0.023]  | 0.003<br>[0.024]  | 0.002<br>[0.024]  | 0.000<br>[0.024]  | 0.000<br>[0.023]  |
| 3-4 Years After                       | -0.031*<br>[0.019]                    | -0.034*<br>[0.018] | -0.012<br>[0.023] | -0.011<br>[0.023] | -0.015<br>[0.023] | -0.017<br>[0.022] |
| Sample                                | Rural                                 | Rural              | Rural             | Rural             | Rural             | Rural             |
| Household Controls                    | No                                    | Yes                | Yes               | Yes               | Yes               | Yes               |
| Infrastructure Controls               | No                                    | No                 | Yes               | Yes               | Yes               | Yes               |
| <i>Time Trends Interacted With...</i> |                                       |                    |                   |                   |                   |                   |
| ...2007 Census Characteristics        | No                                    | No                 | No                | Yes               | Yes               | Yes               |
| ...Region FEs                         | No                                    | No                 | No                | No                | Yes               | No                |
| ...Department FEs                     | No                                    | No                 | No                | No                | No                | Yes               |
| 2005 Control Mean                     | 0.183                                 | 0.183              | 0.183             | 0.183             | 0.183             | 0.183             |
| R-Squared                             | 0.100                                 | 0.111              | 0.111             | 0.112             | 0.112             | 0.116             |
| Observations                          | 46,956                                | 46,769             | 46,769            | 46,769            | 46,769            | 46,769            |

Notes: Each column comes from a separate event-study regression. The median number of prenatal visits is 9. All specifications include district and year fixed effects. The omitted category is -1. The unit of observation is a birth in the birth history module in the DHS survey. The household controls consist of the years of education of the head of the household, gender of the head of the household, age of the head of the household, urban/rural indicator, and wealth index. Infrastructure controls include the number of I-1, I-3, I-4, II, and III facilities and the number of Essalud facilities in the district. The census characteristics are urban share, SIS coverage rate, average wealth index, and secondary school completion rate. Regions fixed effects indicate whether the district is located in the coastal area, in the jungle area, or in the mountain area. Standard errors are clustered at the district level. The event variables for  $t = 5$  or more and  $t = -25$  or less are included in the regression but not shown in the table. Districts are defined as rural if 50 percent or less of the population lives in an urban area according to the 2007 population census.

Table 3.A.11: Robustness Check: Prenatal Care Conducted by Doctor

|                                       | Prenatal Care Conducted By Doctor |                    |                    |                    |                    |                    |
|---------------------------------------|-----------------------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
|                                       | (1)                               | (2)                | (3)                | (4)                | (5)                | (6)                |
| 2-4 Years Before                      | 0.016<br>[0.021]                  | 0.016<br>[0.021]   | 0.016<br>[0.021]   | 0.017<br>[0.021]   | 0.018<br>[0.021]   | 0.017<br>[0.021]   |
| 0-2 Years After                       | 0.046**<br>[0.018]                | 0.044**<br>[0.017] | 0.033**<br>[0.015] | 0.032**<br>[0.015] | 0.031**<br>[0.015] | 0.036**<br>[0.015] |
| 3-4 Years After                       | 0.047**<br>[0.021]                | 0.046**<br>[0.019] | 0.028*<br>[0.017]  | 0.027<br>[0.017]   | 0.025<br>[0.017]   | 0.034*<br>[0.018]  |
| Sample                                | Rural                             | Rural              | Rural              | Rural              | Rural              | Rural              |
| Household Controls                    | No                                | Yes                | Yes                | Yes                | Yes                | Yes                |
| Infrastructure Controls               | No                                | No                 | Yes                | Yes                | Yes                | Yes                |
| <i>Time Trends Interacted With...</i> |                                   |                    |                    |                    |                    |                    |
| ...2007 Census Characteristics        | No                                | No                 | No                 | Yes                | Yes                | Yes                |
| ...Region FEs                         | No                                | No                 | No                 | No                 | Yes                | No                 |
| ...Department FEs                     | No                                | No                 | No                 | No                 | No                 | Yes                |
| 2005 Control Mean                     | 0.131                             | 0.131              | 0.131              | 0.131              | 0.131              | 0.131              |
| R-Squared                             | 0.084                             | 0.097              | 0.098              | 0.098              | 0.098              | 0.100              |
| Observations                          | 47,085                            | 46,896             | 46,896             | 46,896             | 46,896             | 46,896             |

Notes: Each column comes from a separate event-study regression. All specifications include district and year fixed effects. The omitted category is -1. The unit of observation is a birth in the birth history module in the DHS survey. The household controls consist of the years of education of the head of the household, gender of the head of the household, age of the head of the household, urban/rural indicator, and wealth index. Infrastructure controls include the number of I-1, I-3, I-4, II, and III facilities and the number of Essalud facilities in the district. The census characteristics are urban share, SIS coverage rate, average wealth index, and secondary school completion rate. Regions fixed effects indicate whether the district is located in the coastal area, in the jungle area, or in the mountain area. Standard errors are clustered at the district level. The event variables for  $t = 5$  or more and  $t = -25$  or less are included in the regression but not shown in the table. Districts are defined as rural if 50 percent or less of the population lives in an urban area according to the 2007 population census.

Table 3.A.12: Robustness Check: Iron Supplements

|                                       | Given iron tablets during pregnancy |                  |                  |                  |                  |                  |
|---------------------------------------|-------------------------------------|------------------|------------------|------------------|------------------|------------------|
|                                       | (1)                                 | (2)              | (3)              | (4)              | (5)              | (6)              |
| 2-4 Years Before                      | 0.018<br>[0.028]                    | 0.019<br>[0.028] | 0.019<br>[0.028] | 0.020<br>[0.027] | 0.020<br>[0.027] | 0.018<br>[0.027] |
| 0-2 Years After                       | 0.011<br>[0.022]                    | 0.009<br>[0.022] | 0.025<br>[0.023] | 0.028<br>[0.023] | 0.027<br>[0.023] | 0.028<br>[0.023] |
| 3-4 Years After                       | 0.007<br>[0.019]                    | 0.008<br>[0.021] | 0.032<br>[0.024] | 0.036<br>[0.024] | 0.035<br>[0.024] | 0.034<br>[0.023] |
| Sample                                | Rural                               | Rural            | Rural            | Rural            | Rural            | Rural            |
| Household Controls                    | No                                  | Yes              | Yes              | Yes              | Yes              | Yes              |
| Infrastructure Controls               | No                                  | No               | Yes              | Yes              | Yes              | Yes              |
| <i>Time Trends Interacted With...</i> |                                     |                  |                  |                  |                  |                  |
| ...2007 Census Characteristics        | No                                  | No               | No               | Yes              | Yes              | Yes              |
| ...Region FEs                         | No                                  | No               | No               | No               | Yes              | No               |
| ...Department FEs                     | No                                  | No               | No               | No               | No               | Yes              |
| 2005 Control Mean                     | 0.707                               | 0.707            | 0.707            | 0.707            | 0.707            | 0.707            |
| R-Squared                             | 0.157                               | 0.160            | 0.161            | 0.162            | 0.163            | 0.165            |
| Observations                          | 47,055                              | 46,866           | 46,866           | 46,866           | 46,866           | 46,866           |

Notes: Each column comes from a separate event-study regression. All specifications include district and year fixed effects. The omitted category is -1. The unit of observation is a birth in the birth history module in the DHS survey. The household controls consist of the years of education of the head of the household, gender of the head of the household, age of the head of the household, urban/rural indicator, and wealth index. Infrastructure controls include the number of I-1, I-3, I-4, II, and III facilities and the number of Essalud facilities in the district. The census characteristics are urban share, SIS coverage rate, average wealth index, and secondary school completion rate. Regions fixed effects indicate whether the district is located in the coastal area, in the jungle area, or in the mountain area. Standard errors are clustered at the district level. The event variables for  $t = 5$  or more and  $t = -25$  or less are included in the regression but not shown in the table. Districts are defined as rural if 50 percent or less of the population lives in an urban area according to the 2007 population census.

Table 3.A.13: Robustness Check: Tetanus Injections

|                                       | At least one tetanus injection before birth |                     |                     |                     |                     |                     |
|---------------------------------------|---------------------------------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
|                                       | (1)                                         | (2)                 | (3)                 | (4)                 | (5)                 | (6)                 |
| 2-4 Years Before                      | 0.047*<br>[0.025]                           | 0.048*<br>[0.025]   | 0.048*<br>[0.025]   | 0.049*<br>[0.025]   | 0.048*<br>[0.025]   | 0.047**<br>[0.024]  |
| 0-2 Years After                       | 0.071***<br>[0.019]                         | 0.073***<br>[0.019] | 0.074***<br>[0.021] | 0.072***<br>[0.021] | 0.072***<br>[0.021] | 0.075***<br>[0.020] |
| 3-4 Years After                       | 0.061***<br>[0.018]                         | 0.062***<br>[0.018] | 0.065***<br>[0.021] | 0.062***<br>[0.021] | 0.062***<br>[0.021] | 0.064***<br>[0.020] |
| Sample                                | Rural                                       | Rural               | Rural               | Rural               | Rural               | Rural               |
| Household Controls                    | No                                          | Yes                 | Yes                 | Yes                 | Yes                 | Yes                 |
| Infrastructure Controls               | No                                          | No                  | Yes                 | Yes                 | Yes                 | Yes                 |
| <i>Time Trends Interacted With...</i> |                                             |                     |                     |                     |                     |                     |
| ...2007 Census Characteristics        | No                                          | No                  | No                  | Yes                 | Yes                 | Yes                 |
| ...Region FEs                         | No                                          | No                  | No                  | No                  | Yes                 | No                  |
| ...Department FEs                     | No                                          | No                  | No                  | No                  | No                  | Yes                 |
| 2005 Control Mean                     | 0.702                                       | 0.702               | 0.702               | 0.702               | 0.702               | 0.702               |
| R-Squared                             | 0.079                                       | 0.080               | 0.080               | 0.081               | 0.081               | 0.084               |
| Observations                          | 46,656                                      | 46,468              | 46,468              | 46,468              | 46,468              | 46,468              |

Notes: Each column comes from a separate event-study regression. All specifications include district and year fixed effects. The omitted category is -1. The unit of observation is a birth in the birth history module in the DHS survey. The household controls consist of the years of education of the head of the household, gender of the head of the household, age of the head of the household, urban/rural indicator, and wealth index. Infrastructure controls include the number of I-1, I-3, I-4, II, and III facilities and the number of Essalud facilities in the district. The census characteristics are urban share, SIS coverage rate, average wealth index, and secondary school completion rate. Regions fixed effects indicate whether the district is located in the coastal area, in the jungle area, or in the mountain area. Standard errors are clustered at the district level. The event variables for  $t = 5$  or more and  $t = -25$  or less are included in the regression but not shown in the table. Districts are defined as rural if 50 percent or less of the population lives in an urban area according to the 2007 population census.

Table 3.A.14: Robustness Check: Vitamin Doses

|                                       | Given vitamin dose after pregnancy |                    |                   |                   |                   |                    |
|---------------------------------------|------------------------------------|--------------------|-------------------|-------------------|-------------------|--------------------|
|                                       | (1)                                | (2)                | (3)               | (4)               | (5)               | (6)                |
| 2-4 Years Before                      | 0.006<br>[0.018]                   | 0.004<br>[0.018]   | 0.006<br>[0.017]  | 0.007<br>[0.017]  | 0.007<br>[0.017]  | 0.006<br>[0.018]   |
| 0-2 Years After                       | 0.036**<br>[0.014]                 | 0.035**<br>[0.014] | 0.026*<br>[0.013] | 0.026*<br>[0.014] | 0.026*<br>[0.014] | 0.028**<br>[0.014] |
| 3-4 Years After                       | 0.004<br>[0.011]                   | 0.003<br>[0.011]   | -0.012<br>[0.012] | -0.009<br>[0.012] | -0.009<br>[0.012] | -0.008<br>[0.013]  |
| Sample                                | Rural                              | Rural              | Rural             | Rural             | Rural             | Rural              |
| Household Controls                    | No                                 | Yes                | Yes               | Yes               | Yes               | Yes                |
| Infrastructure Controls               | No                                 | No                 | Yes               | Yes               | Yes               | Yes                |
| <i>Time Trends Interacted With...</i> |                                    |                    |                   |                   |                   |                    |
| ...2007 Census Characteristics        | No                                 | No                 | No                | Yes               | Yes               | Yes                |
| ...Region FEs                         | No                                 | No                 | No                | No                | Yes               | No                 |
| ...Department FEs                     | No                                 | No                 | No                | No                | No                | Yes                |
| 2005 Control Mean                     | 0.193                              | 0.193              | 0.193             | 0.193             | 0.193             | 0.193              |
| R-Squared                             | 0.074                              | 0.075              | 0.075             | 0.076             | 0.076             | 0.079              |
| Observations                          | 47,069                             | 46,880             | 46,880            | 46,880            | 46,880            | 46,880             |

Notes: Each column comes from a separate event-study regression. All specifications include district and year fixed effects. The omitted category is -1. The unit of observation is a birth in the birth history module in the DHS survey. The household controls consist of the years of education of the head of the household, gender of the head of the household, age of the head of the household, urban/rural indicator, and wealth index. Infrastructure controls include the number of I-1, I-3, I-4, II, and III facilities and the number of Essalud facilities in the district. The census characteristics are urban share, SIS coverage rate, average wealth index, and secondary school completion rate. Regions fixed effects indicate whether the district is located in the coastal area, in the jungle area, or in the mountain area. Standard errors are clustered at the district level. The event variables for  $t = 5$  or more and  $t = -25$  or less are included in the regression but not shown in the table. Districts are defined as rural if 50 percent or less of the population lives in an urban area according to the 2007 population census.

Table 3.A.15: Robustness Check: Number of Births

|                                       | Number of Births    |                     |                     |                    |                   |
|---------------------------------------|---------------------|---------------------|---------------------|--------------------|-------------------|
|                                       | (1)                 | (2)                 | (3)                 | (4)                | (5)               |
| 11-19 Quarters Before                 | 1.784<br>[4.137]    | 3.720<br>[3.811]    | 3.174<br>[3.708]    | 2.989<br>[3.679]   | 1.113<br>[3.589]  |
| 2-10 Quarters Before                  | 0.155<br>[3.202]    | 0.722<br>[3.140]    | 0.586<br>[3.112]    | 0.578<br>[3.123]   | -0.122<br>[3.065] |
| 0-10 Quarters After                   | -1.769<br>[3.430]   | -2.723<br>[3.467]   | -2.437<br>[3.442]   | -2.415<br>[3.447]  | -1.779<br>[3.482] |
| 11-19 Quarters After                  | -6.963**<br>[3.332] | -7.188**<br>[3.212] | -6.672**<br>[3.238] | -6.075*<br>[3.305] | -3.831<br>[3.329] |
| Sample                                | Rural               | Rural               | Rural               | Rural              | Rural             |
| Infrastructure Controls               | No                  | Yes                 | Yes                 | Yes                | Yes               |
| <i>Time Trends Interacted With...</i> |                     |                     |                     |                    |                   |
| ...2007 Census Characteristics        | No                  | No                  | Yes                 | Yes                | Yes               |
| ...Region FEs                         | No                  | No                  | No                  | Yes                | No                |
| ...Department FEs                     | No                  | No                  | No                  | No                 | Yes               |
| 2005 Control Mean                     | 17.910              | 17.910              | 17.910              | 17.910             | 17.910            |
| R-Squared                             | 0.592               | 0.593               | 0.598               | 0.600              | 0.611             |
| Observations                          | 42,422              | 42,422              | 42,422              | 42,422             | 42,422            |

Notes: Each column comes from a separate event-study regression. All specifications include district and year fixed effects. The omitted category is -1. The unit of observation is a district. Infrastructure controls include the number of I-1, I-3, I-4, II, and III facilities and the number of Essalud facilities in the district. The census characteristics are urban share, SIS coverage rate, average wealth index, and secondary school completion rate. Regions fixed effects indicate whether the district is located in the coastal area, in the jungle area, or in the mountain area. Standard errors are clustered at the district level. The event variables for  $t = 5$  or more and  $t = -25$  or less are included in the regression but not shown in the table. Districts are defined as rural if 50 percent or less of the population lives in an urban area according to the 2007 population census.

Table 3.A.16: Robustness Check: Wanted Pregnancy

|                                       | Wanted Pregnancy |                  |                  |                  |                  |                  |
|---------------------------------------|------------------|------------------|------------------|------------------|------------------|------------------|
|                                       | (1)              | (2)              | (3)              | (4)              | (5)              | (6)              |
| 2-4 Years Before                      | 0.011<br>[0.021] | 0.011<br>[0.020] | 0.014<br>[0.021] | 0.013<br>[0.021] | 0.013<br>[0.021] | 0.013<br>[0.021] |
| 0-2 Years After                       | 0.039<br>[0.026] | 0.036<br>[0.027] | 0.030<br>[0.025] | 0.030<br>[0.025] | 0.030<br>[0.025] | 0.029<br>[0.025] |
| 3-4 Years After                       | 0.021<br>[0.027] | 0.021<br>[0.026] | 0.012<br>[0.024] | 0.014<br>[0.024] | 0.014<br>[0.025] | 0.014<br>[0.025] |
| Sample                                | Rural            | Rural            | Rural            | Rural            | Rural            | Rural            |
| Household Controls                    | No               | Yes              | Yes              | Yes              | Yes              | Yes              |
| Infrastructure Controls               | No               | No               | Yes              | Yes              | Yes              | Yes              |
| <i>Time Trends Interacted With...</i> |                  |                  |                  |                  |                  |                  |
| ...2007 Census Characteristics        | No               | No               | No               | Yes              | Yes              | Yes              |
| ...Region FEs                         | No               | No               | No               | No               | Yes              | No               |
| ...Department FEs                     | No               | No               | No               | No               | No               | Yes              |
| 2005 Control Mean                     | 0.312            | 0.312            | 0.312            | 0.312            | 0.312            | 0.312            |
| R-Squared                             | 0.054            | 0.063            | 0.063            | 0.063            | 0.063            | 0.064            |
| Observations                          | 58,982           | 58,738           | 58,738           | 58,738           | 58,738           | 58,738           |

Notes: Each column comes from a separate event-study regression. All specifications include district and year fixed effects. The omitted category is -1. The unit of observation is a birth in the birth history module in the DHS survey. The household controls consist of the years of education of the head of the household, gender of the head of the household, age of the head of the household, urban/rural indicator, and wealth index. Infrastructure controls include the number of I-1, I-3, I-4, II, and III facilities and the number of Essalud facilities in the district. The census characteristics are urban share, SIS coverage rate, average wealth index, and secondary school completion rate. Regions fixed effects indicate whether the district is located in the coastal area, in the jungle area, or in the mountain area. Standard errors are clustered at the district level. The event variables for  $t = 5$  or more and  $t = -25$  or less are included in the regression but not shown in the table. Districts are defined as rural if 50 percent or less of the population lives in an urban area according to the 2007 population census.

Table 3.A.17: Robustness Check: Condom Usage

|                                       | Current Contraceptive Method: Condoms |                  |                    |                    |                     |                   |
|---------------------------------------|---------------------------------------|------------------|--------------------|--------------------|---------------------|-------------------|
|                                       | (1)                                   | (2)              | (3)                | (4)                | (5)                 | (6)               |
| 2-4 Years Before                      | 0.010<br>[0.009]                      | 0.008<br>[0.011] | 0.009<br>[0.012]   | 0.008<br>[0.011]   | 0.007<br>[0.011]    | 0.009<br>[0.011]  |
| 0-2 Years After                       | 0.004<br>[0.007]                      | 0.005<br>[0.008] | 0.010<br>[0.008]   | 0.009<br>[0.008]   | 0.010<br>[0.008]    | 0.007<br>[0.008]  |
| 3-4 Years After                       | 0.011*<br>[0.006]                     | 0.010<br>[0.007] | 0.016**<br>[0.007] | 0.015**<br>[0.007] | 0.016***<br>[0.006] | 0.012*<br>[0.007] |
| Sample                                | Rural                                 | Rural            | Rural              | Rural              | Rural               | Rural             |
| Household Controls                    | No                                    | Yes              | Yes                | Yes                | Yes                 | Yes               |
| Infrastructure Controls               | No                                    | No               | Yes                | Yes                | Yes                 | Yes               |
| <i>Time Trends Interacted With...</i> |                                       |                  |                    |                    |                     |                   |
| ...2007 Census Characteristics        | No                                    | No               | No                 | Yes                | Yes                 | Yes               |
| ...Region FEs                         | No                                    | No               | No                 | No                 | Yes                 | No                |
| ...Department FEs                     | No                                    | No               | No                 | No                 | No                  | Yes               |
| 2005 Control Mean                     | 0.029                                 | 0.029            | 0.029              | 0.029              | 0.029               | 0.029             |
| R-Squared                             | 0.032                                 | 0.043            | 0.043              | 0.043              | 0.043               | 0.044             |
| Observations                          | 99,773                                | 99,395           | 99,395             | 99,395             | 99,395              | 99,395            |

Notes: Each column comes from a separate event-study regression. All specifications include district and year fixed effects. The omitted category is -1. The unit of observation is a birth in the birth history module in the DHS survey. The household controls consist of the years of education of the head of the household, gender of the head of the household, age of the head of the household, urban/rural indicator, and wealth index. Infrastructure controls include the number of I-1, I-3, I-4, II, and III facilities and the number of Essalud facilities in the district. The census characteristics are urban share, SIS coverage rate, average wealth index, and secondary school completion rate. Regions fixed effects indicate whether the district is located in the coastal area, in the jungle area, or in the mountain area. Standard errors are clustered at the district level. The event variables for  $t = 5$  or more and  $t = -25$  or less are included in the regression but not shown in the table. Districts are defined as rural if 50 percent or less of the population lives in an urban area according to the 2007 population census.

Table 3.A.18: Robustness Check: Neonatal Mortality Rate

|                                       | Neonatal Mortality Rate |                   |                   |                   |                   |
|---------------------------------------|-------------------------|-------------------|-------------------|-------------------|-------------------|
|                                       | (1)                     | (2)               | (3)               | (4)               | (5)               |
| 2-4 Years Before                      | 0.454<br>[1.864]        | 0.378<br>[1.865]  | 0.582<br>[1.828]  | 0.956<br>[1.826]  | 1.529<br>[1.815]  |
| 0-2 Years After                       | -0.379<br>[1.757]       | -0.541<br>[1.747] | -0.682<br>[1.711] | -0.825<br>[1.657] | -1.127<br>[1.545] |
| 3-4 Years After                       | -0.690<br>[1.593]       | -0.563<br>[1.604] | -0.572<br>[1.583] | -1.055<br>[1.522] | -1.766<br>[1.449] |
| Sample                                | Rural                   | Rural             | Rural             | Rural             | Rural             |
| Infrastructure Controls               | No                      | Yes               | Yes               | Yes               | Yes               |
| <i>Time Trends Interacted With...</i> |                         |                   |                   |                   |                   |
| ...2007 Census Characteristics        | No                      | No                | Yes               | Yes               | Yes               |
| ...Region FEs                         | No                      | No                | No                | Yes               | No                |
| ...Department FEs                     | No                      | No                | No                | No                | Yes               |
| 2005 Control Mean                     | 11.124                  | 11.124            | 11.124            | 11.124            | 11.124            |
| R-Squared                             | 0.100                   | 0.101             | 0.102             | 0.102             | 0.105             |
| Observations                          | 6,116                   | 6,116             | 6,116             | 6,116             | 6,116             |

Notes: Each column comes from a separate event-study regression. All specifications include district and year fixed effects. The omitted category is -1. Districts are weighted by the number of births. Infrastructure controls include the number of I-1, I-3, I-4, II, and IIII facilities and the number of Essalud facilities in the district. The census characteristics are urban share, SIS coverage rate, average wealth index, and secondary school completion rate. Regions fixed effects indicate whether the district is located in the coastal area, in the jungle area, or in the mountain area. Standard errors are clustered at the district level. The event variables for  $t = 5$  or more and  $t = -25$  or less are included in the regression but not shown in the table. Districts are defined as rural if 50 percent or less of the population lives in an urban area according to the 2007 population census.

Table 3.A.19: Robustness Check: Infant Mortality Rate

|                                       | Infant Mortality Rate |                  |                   |                   |                   |
|---------------------------------------|-----------------------|------------------|-------------------|-------------------|-------------------|
|                                       | (1)                   | (2)              | (3)               | (4)               | (5)               |
| 2-4 Years Before                      | 1.346<br>[3.162]      | 1.407<br>[3.180] | 2.015<br>[3.135]  | 2.907<br>[3.114]  | 4.666<br>[3.033]  |
| 0-2 Years After                       | 0.207<br>[3.064]      | 0.117<br>[3.029] | -0.386<br>[2.913] | -0.716<br>[2.763] | -1.078<br>[2.404] |
| 3-4 Years After                       | 2.849<br>[3.045]      | 2.550<br>[3.088] | 2.023<br>[2.967]  | 0.757<br>[2.745]  | -0.437<br>[2.373] |
| Sample                                | Rural                 | Rural            | Rural             | Rural             | Rural             |
| Infrastructure Controls               | No                    | Yes              | Yes               | Yes               | Yes               |
| <i>Time Trends Interacted With...</i> |                       |                  |                   |                   |                   |
| ...2007 Census Characteristics        | No                    | No               | Yes               | Yes               | Yes               |
| ...Region FEs                         | No                    | No               | No                | Yes               | No                |
| ...Department FEs                     | No                    | No               | No                | No                | Yes               |
| 2005 Control Mean                     | 28.063                | 28.063           | 28.063            | 28.063            | 28.063            |
| R-Squared                             | 0.064                 | 0.064            | 0.065             | 0.066             | 0.068             |
| Observations                          | 6,116                 | 6,116            | 6,116             | 6,116             | 6,116             |

Notes: Each column comes from a separate event-study regression. All specifications include district and year fixed effects. The omitted category is -1. Districts are weighted by the number of births. Infrastructure controls include the number of I-1, I-3, I-4, II, and III facilities and the number of Essalud facilities in the district. The census characteristics are urban share, SIS coverage rate, average wealth index, and secondary school completion rate. Regions fixed effects indicate whether the district is located in the coastal area, in the jungle area, or in the mountain area. Standard errors are clustered at the district level. The event variables for  $t = 5$  or more and  $t = -25$  or less are included in the regression but not shown in the table. Districts are defined as rural if 50 percent or less of the population lives in an urban area according to the 2007 population census.

# Bibliography

- Abeler, Johannes, Armin Falk, and Fabian Kosse**, “Malleability of preferences for honesty,” 2021.
- Advani, Arun and Bryony Reich**, “Melting pot or salad bowl: The formation of heterogeneous communities,” Working Paper W15/30, Institute for Fiscal Studies (IFS) September 2015.
- Agnes, Flavia**, *Law and Gender Inequality: The Politics of Women’s Rights in India*, Oxford University Press, 1999.
- Akerlof, George A.**, “Social Distance and Social Decisions,” *Econometrica*, 1997, 65 (5), 1005–1027.
- Akerlof, George and Rachel Kranton**, “Economics and Identity,” *The Quarterly Journal of Economics*, 2000, 115 (3), 715–753.
- Alan, Sule, Gozde Corekcioglu, and Matthias Sutter**, “Improving Workplace Climate in Large Corporations: A Clustered Randomized Intervention\*,” *The Quarterly Journal of Economics*, 09 2022. qjac034.
- Alesina, Alberto, Paola Giuliano, and Nathan Nunn**, “On the origins of gender roles: Women and the plough,” *The quarterly journal of economics*, 2013, 128 (2), 469–530.
- Alexander, Greg R and Carol C Korenbrot**, “The role of prenatal care in preventing low birth weight,” *The future of children*, 1995, pp. 103–120.
- Allport, Gordon Willard, Kenneth Clark, and Thomas Pettigrew**, “The nature of prejudice,” 1954.
- Almond, Douglas, Janet Currie, and Mariesa Herrmann**, “From infant to mother: Early disease environment and future maternal health,” *Labour Economics*, 2012, 19 (4), 475–483.
- Anderson, Siwan**, “Caste as an Impediment to Trade,” *American Economic Journal: Applied Economics*, January 2011, 3 (1), 239–63.
- Angrist, Joshua D and Stacey H Chen**, “Schooling and the Vietnam-era GI Bill: Evidence from the draft lottery,” *American Economic Journal: Applied Economics*, 2011, 3 (2), 96–118.

- Asher, Sam and Paul Novosad**, “Rural Roads and Local Economic Development,” *American Economic Review* (Forthcoming) 2018.
- Ashraf, Nava, Oriana Bandiera, Edward Davenport, and Scott S Lee**, “Losing prosociality in the quest for talent? Sorting, selection, and productivity in the delivery of public services,” *American Economic Review*, 2020, 110 (5), 1355–94.
- Atkin, David, Eve Colson-Sihra, and Moses Shayo**, “How Do We Choose Our Identity? A Revealed Preference Approach Using Food Consumption,” *Journal of Political Economy*, 2021, 129 (4), 1193–1251.
- Attewell, Paul and S. Madheswaran**, “Caste Discrimination in the Indian Urban Labour Market: Evidence from the National Sample Survey,” *Economic and Political Weekly*, 2007, 42 (41), 4146–53.
- Bailey, Martha J**, “Reexamining the impact of family planning programs on US fertility: evidence from the War on Poverty and the early years of Title X,” *American Economic Journal: Applied Economics*, 2012, 4 (2), 62–97.
- **and Andrew Goodman-Bacon**, “The War on Poverty’s experiment in public medicine: Community health centers and the mortality of older Americans,” *American Economic Review*, 2015, 105 (3), 1067–1104.
- Banerjee, Abhijit V and Rohini Pande**, “Parochial politics: Ethnic preferences and politician corruption,” 2007.
- Banuri, Sheheryar and Philip Keefer**, “Intrinsic motivation, effort and the call to public service,” *World Bank Policy research working paper*, 2013, (6729).
- Bayly, Susan**, *Caste, society and politics in India from the eighteenth century to the modern age*, Vol. 3, Cambridge University Press, 2001.
- Becker, Anke**, “On the Economic Origins of Restricting Women’s Promiscuity,” *The Review of Economic Studies*, 2021.
- Bharadwaj, Prashant, Julian V Johnsen, and Katrine V Løken**, “Smoking bans, maternal smoking and birth outcomes,” *Journal of Public economics*, 2014, 115, 72–93.
- Binagwaho, Agnes and Tedros Adhanom Ghebreyesus**, “Primary healthcare is cornerstone of universal health coverage,” *BMJ*, 2019, 365.
- Bisin, A and T Verdier**, “The economics of cultural transmission and socialization,” in J Benhabib, A Bisin, and M Jackson, eds., *The Handbook of Social Economics*, Elsevier, 2010.
- Bisin, Alberto and Thierry Verdier**, “Advances in the Economic Theory of Cultural Transmission,” Working Paper 30466, National Bureau of Economic Research September 2022.

- , **Eleonora Patacchini, Thierry Verdier, and Yves Zenou**, “Bend it like Beckham: Ethnic identity and integration,” *European Economic Review*, 2016, *90*, 146–164.
- Bó, Ernesto Dal, Frederico Finan, and Martín A Rossi**, “Strengthening state capabilities: The role of financial incentives in the call to public service,” *The Quarterly Journal of Economics*, 2013, *128* (3), 1169–1218.
- Bobba, Matteo, Tim Ederer, Gianmarco Leon-Ciliotta, Christopher Neilson, and Marco G Nieddu**, “Teacher compensation and structural inequality: Evidence from centralized teacher school choice in Perú,” Technical Report, National Bureau of Economic Research 2021.
- Boelmann, Barbara, Anna Raute, and Uta Schönberg**, “Wind of Change? Cultural Determinants of Maternal Labor Supply,” Working Paper 2022.
- Borusyak, Kirill and Xavier Jaravel**, “Revisiting event study designs,” *Available at SSRN 2826228*, 2018.
- Boserup, E.**, *Woman’s Role in Economic Development*, Allen & Unwin, 1970.
- Bros, Catherine and Mathieu Couttenier**, “Untouchability, homicides and water access,” *Journal of Comparative Economics*, 2015, *43* (3), 549–558.
- Buchmueller, Thomas C, Mireille Jacobson, and Cheryl Wold**, “How far to the hospital? The effect of hospital closures on access to care,” *Journal of Health Economics*, 2006, *25* (4), 740–761.
- Bursztyn, Leonardo, Alessandra L González, and David Yanagizawa-Drott**, “Misperceived social norms: Women working outside the home in saudi arabia,” *American economic review*, 2020, *110* (10), 2997–3029.
- Calonico, Sebastian, Matias D Cattaneo, and Max H Farrell**, “Optimal bandwidth choice for robust bias-corrected inference in regression discontinuity designs,” *The Econometrics Journal*, 2020, *23* (2), 192–210.
- , – , and **Rocio Titiunik**, “Robust nonparametric confidence intervals for regression-discontinuity designs,” *Econometrica*, 2014, *82* (6), 2295–2326.
- Cappelen, Alexander, John List, Anya Samek, and Bertil Tungodden**, “The effect of early-childhood education on social preferences,” *Journal of Political Economy*, 2020, *128* (7), 2739–2758.
- Carranza, Eliana**, “Soil Endowments, Female Labor Force Participation, and the Demographic Deficit of Women in India,” *American Economic Journal: Applied Economics*, October 2014, *6* (4), 197–225.

- Carrillo, Bladimir and Jose Feres**, “Provider supply, utilization, and infant health: evidence from a physician distribution policy,” *American Economic Journal: Economic Policy*, 2019, 11 (3), 156–96.
- Cassan, Guilhem and Lore Vandewalle**, “Identities and Public Policies: Unintended Effects of Political Reservations for Women in India,” Working Paper HEIDWP18-2017, Graduate Institute Geneva 2017.
- , **Daniel Keniston, and Tatjana Kleineberg**, “A Division of Laborers: Identity and Efficiency in India,” Working Paper 28462, National Bureau of Economic Research February 2022.
- Cesur, Resul, Pınar Mine Güneş, Erdal Tekin, and Aydogan Ulker**, “The value of socialized medicine: The impact of universal primary healthcare provision on mortality rates in Turkey,” *Journal of Public Economics*, 2017, 150, 75–93.
- Chakravarti, Uma**, *Gendering caste: Through a feminist lens*, Sage Publications Pvt. Limited, 2018.
- Chaudhury, Nazmul, Jeffrey Hammer, Michael Kremer, Karthik Muralidharan, and F Halsey Rogers**, “Missing in action: teacher and health worker absence in developing countries,” *Journal of Economic Perspectives*, 2006, 20 (1), 91–116.
- Chen, Martha**, “A Matter of Survival: Women’s Right to Employment in India and Bangladesh,” in Martha C. Nussbaum and Jonathan Glover, eds., *Women, Culture, and Development: A Study of Human Capabilities*, Oxford Academic Press, 1995.
- Cobden-Ramsay, L. E. B.**, *Feudatory States of Orissa. Report number Bengal District Gazetteers XXI* 1910.
- Conley, Timothy G**, “GMM estimation with cross sectional dependence,” *Journal of econometrics*, 1999, 92 (1), 1–45.
- Costa, Francisco, Letícia Nunes, and Fabio Miessi Sanches**, “How to attract physicians to underserved areas? Policy recommendations from a structural model,” *The Review of Economics and Statistics*, 2021, pp. 1–45.
- Cowley, Edd and Sarah Smith**, “Motivation and mission in the public sector: Evidence from the World Values Survey,” *Theory and decision*, 2014, 76 (2), 241–263.
- Currie, Janet and Hannes Schwandt**, “The 9/11 dust cloud and pregnancy outcomes: A reconsideration,” *Journal of Human Resources*, 2016, 51 (4), 805–831.
- **and Jonathan Gruber**, “Saving Babies: The Efficacy and Cost of Recent Changes in the Medicaid Eligibility of Pregnant Women,” *Journal of Political Economy*, 1996, 104 (6), 1263–1296.

- , – , and **Michael Fischer**, “Physician Payments and Infant Mortality: Evidence from Medicaid Fee Policy,” *American Economic Review*, May 1995, 85 (2), 106–111.
- Dell, Melissa, Nathan Lane, and Pablo Querubin**, “The historical state, local collective action, and economic development in Vietnam,” *Econometrica*, 2018, 86 (6), 2083–2121.
- Deloche, J.**, *Transport and Communications in India Prior to Steam Locomotion: Water transport* French studies in South Asian culture and society, Oxford University Press, 1993.
- Dewar, Frances**, *Report of the Land Revenue Settlement of the Sambalpur District, 1906*, Bengal Secretariat Book Depot, 1908.
- Dhar, Diva, Tarun Jain, and Seema Jayachandran**, “Reshaping adolescents’ gender attitudes: Evidence from a school-based experiment in India,” *American Economic Review*, 2022, 112 (3), 899–927.
- Dirks, Nicholas B.**, “Castes of mind,” in “Castes of Mind,” Princeton University Press, 2011.
- Dobbie, Will and Roland G Fryer Jr.**, “The impact of voluntary youth service on future outcomes: Evidence from Teach For America,” *The BE Journal of Economic Analysis & Policy*, 2015, 15 (3), 1031–1065.
- Donaldson, Dave**, “Railroads of the Raj: Estimating the impact of transportation infrastructure,” *American Economic Review*, 2018, 108 (4-5), 899–934.
- Duflo, Esther**, “Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment,” *American Economic Review*, 2001, 91 (4), 795–813.
- and **Rohini Pande**, “Dams,” *Quarterly Journal of Economics*, 2007, 122 (2), 601–646.
- Dumont, Louis**, *Homo Hierarchicus: The Caste System and Its Implications*, Chicago: University of Chicago Press, 1970.
- Fadlon, Itzik, Frederik Plesner Lyngse, and Torben Heien Nielsen**, “Early Career, Life-Cycle Choices, and Gender,” Technical Report, National Bureau of Economic Research 2022.
- Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde**, “Global evidence on economic preferences,” *The Quarterly Journal of Economics*, 2018, 133 (4), 1645–1692.
- , – , **Thomas J Dohmen, David Huffman, and Uwe Sunde**, “The preference survey module: A validated instrument for measuring risk, time, and social preferences,” 2016.
- Fernández, Raquel**, “Cultural change as learning: The evolution of female labor force participation over a century,” *American Economic Review*, 2013, 103 (1), 472–500.

- , **Alessandra Fogli**, and **Claudia Olivetti**, “Mothers and sons: Preference formation and female labor force dynamics,” *The Quarterly Journal of Economics*, 2004, *119* (4), 1249–1299.
- **and** – , “Culture: An empirical investigation of beliefs, work, and fertility,” *American economic journal: Macroeconomics*, 2009, *1* (1), 146–77.
- Ferrara, Eliana La, Alberto Chong, and Suzanne Duryea**, “Soap operas and fertility: Evidence from Brazil,” *American Economic Journal: Applied Economics*, 2012, *4* (4), 1–31.
- Field, Erica, Rohini Pande, Natalia Rigol, Simone Schaner, and Charity Troyer Moore**, “On her own account: How strengthening women’s financial control impacts labor supply and gender norms,” *American Economic Review*, 2021, *111* (7), 2342–75.
- , **Seema Jayachandran**, and **Rohini Pande**, “Do Traditional Institutions Constrain Female Entrepreneurship? A Field Experiment on Business Training in India,” *American Economic Review*, May 2010, *100* (2), 125–29.
- Fouka, Vasiliki**, “Backlash: The unintended effects of language prohibition in US schools after World War I,” *The Review of Economic Studies*, 2020, *87* (1), 204–239.
- , **Soumyajit Mazumder**, and **Marco Tabellini**, “From Immigrants to Americans: Race and Assimilation during the Great Migration,” *The Review of Economic Studies*, 06 2021, *89* (2), 811–842.
- Frehywot, Seble, Fitzhugh Mullan, Perry W Payne, and Heather Ross**, “Compulsory service programmes for recruiting health workers in remote and rural areas: do they work?,” *Bulletin of the World Health Organization*, 2010, *88*, 364–370.
- Gelman, Andrew and Guido Imbens**, “Why high-order polynomials should not be used in regression discontinuity designs,” *Journal of Business & Economic Statistics*, 2019, *37* (3), 447–456.
- Ghurye, Govind Sadashiv**, *Caste and race in India*, Popular Prakashan, 1969.
- Giuliano, Paola**, “Gender and culture,” *Oxford Review of Economic Policy*, 2021, *36* (4), 944–961.
- Gondwe, Austrida, Per Ashorn, Ulla Ashorn, Kathryn G Dewey, Kenneth Maleta, Minyanga Nkhoma, John Mbotwa, and Josh M Jorgensen**, “Pre-pregnancy body mass index (BMI) and maternal gestational weight gain are positively associated with birth outcomes in rural Malawi,” *PloS one*, 2018, *13* (10), e0206035.
- Gujral, Kritee and Anirban Basu**, “Impact of Rural and Urban Hospital Closures on Inpatient Mortality,” Working Paper 26182, National Bureau of Economic Research August 2019.

- Hjort, Jonas and Jonas Poulsen**, “The Arrival of Fast Internet and Employment in Africa,” *American Economic Review*, March 2019, *109* (3), 1032–79.
- Hsiao, Allan**, “Misallocation and Infrastructure Investment: Evidence from Healthcare in Indonesia,” Working Paper 2019.
- Iyer, Lakshmi**, “Direct versus Indirect Colonial Rule in India: Long-Term Consequences,” *The Review of Economics and Statistics*, 11 2010, *92* (4), 693–713.
- Jarotschkin, Alexandra and Ekaterina Zhuravskaya**, “Diffusion of Gender Norms: Evidence from Stalin’s Ethnic Deportations,” 2019.
- Jayachandran, Seema**, “The Roots of Gender Inequality in Developing Countries,” *Annual Review of Economics*, 2015, *7* (1), 63–88.
- , “Social Norms as a Barrier to Women’s Employment in Developing Countries,” *IMF Economic Review*, September 2021, *69* (3), 576–595.
- **and Rohini Pande**, “Why Are Indian Children So Short? The Role of Birth Order and Son Preference,” *American Economic Review*, September 2017, *107* (9), 2600–2629.
- Jensen, Robert**, “The digital provide: Information (technology), market performance, and welfare in the South Indian fisheries sector,” *Quarterly Journal of Economics*, 2007, *122* (3), 879–924.
- **and Nolan H Miller**, “Market Integration, Demand, and the Growth of Firms: Evidence from a Natural Experiment in India,” *American Economic Review*, 2018, *108* (12), 3583–3625.
- Kling, Jeffrey R, Jeffrey B Liebman, and Lawrence F Katz**, “Experimental analysis of neighborhood effects,” *Econometrica*, 2007, *75* (1), 83–119.
- Kosse, Fabian, Thomas Deckers, Pia Pinger, Hannah Schildberg-Hörisch, and Armin Falk**, “The formation of prosociality: causal evidence on the role of social environment,” *Journal of Political Economy*, 2020, *128* (2), 434–467.
- Kremer, Michael, Gautam Rao, and Frank Schilbach**, “Behavioral development economics,” in “Handbook of Behavioral Economics: Applications and Foundations 1,” Vol. 2, Elsevier, 2019, pp. 345–458.
- Lleras-Muney, Adriana**, “The relationship between education and adult mortality in the United States,” *The Review of Economic Studies*, 2005, *72* (1), 189–221.
- Londoño-Vélez, Juliana**, “The impact of diversity on perceptions of income distribution and preferences for redistribution,” *Journal of Public Economics*, 2022, *214*, 104732.
- Lowe, Matt**, “Types of contact: A field experiment on collaborative and adversarial caste integration,” *American Economic Review*, 2021, *111* (6), 1807–44.

- Luke, Nancy and Kaivan Munshi**, “Women as agents of change: Female income and mobility in India,” *Journal of Development Economics*, 2011, *94* (1), 1–17.
- Marini, Alessandra, Claudia Rokx, and Paul Gallagher**, “Standing Tall: Peru’s Success in Overcoming its Stunting Crisis. Washington, D.C. : World Bank Group,” 2014.
- Marten, J.T.**, *Census of India 1911: Central Provinces and Berar. Report number v. 10*, pt. 1 1912.
- Miranda, J Jaime, Francisco Diez-Canseco, Claudia Lema, Andrés G Lescano, Mylene Lagarde, Duane Blaauw, and Luis Huicho**, “Stated preferences of doctors for choosing a job in rural areas of Peru: a discrete choice experiment,” *PloS one*, 2012, *7* (12), e50567.
- Mo, Cecilia Hyunjung and Katharine M Conn**, “When do the advantaged see the disadvantages of others? A quasi-experimental study of national service,” *American Political Science Review*, 2018, *112* (4), 721–741.
- Mousa, Salma**, “Building social cohesion between Christians and Muslims through soccer in post-ISIS Iraq,” *Science*, 2020, *369* (6505), 866–870.
- Munshi, Kaivan**, “Caste and the Indian Economy,” *Journal of Economic Literature*, December 2019, *57* (4), 781–834.
- **and Mark Rosenzweig**, “Traditional Institutions Meet the Modern World: Caste, Gender, and Schooling Choice in a Globalizing Economy,” *American Economic Review*, September 2006, *96* (4), 1225–1252.
- **and –**, “Insiders and Outsiders: Local Ethnic Politics and Public Goods Provision,” Working Paper 21720, National Bureau of Economic Research 2015.
- Oh, Suanna**, “Does identity affect labor supply?,” Working Paper 9487, CESifo December 2021.
- Okeke, Edward N**, “When a Doctor Falls from the Sky: The Impact of Easing Doctor Supply Constraints on Mortality,” *American Economic Review*, 2023, *113* (3), 585–627.
- Olkola, Maarit**, “Infant and Maternal Mortality Decline – The Role of Pre-natal and Well-child visits,” 2019. Working Paper.
- O’Malley, L. S. S.**, *Sambalpur number Bengal District Gazetteers* 1909.
- Oreopoulos, Philip, Till Von Wachter, and Andrew Heisz**, “The short-and long-term career effects of graduating in a recession,” *American Economic Journal: Applied Economics*, 2012, *4* (1), 1–29.
- Piketty, Thomas**, *Capital and Ideology*, Cambridge, MA and London, England: Harvard University Press, 2021.

- Rahmati, Shoboo, Ali Delpishe, Milad Azami, Mohammed Reza Hafezi Ahmadi, and Kurosh Sayehmiri**, “Maternal Anemia during pregnancy and infant low birth weight: A systematic review and Meta-analysis,” *International Journal of Reproductive BioMedicine*, 2017, 15 (3), 125.
- Rao, Gautam**, “Familiarity does not breed contempt: Generosity, discrimination, and diversity in Delhi schools,” *American Economic Review*, 2019, 109 (3), 774–809.
- Rocha, Romero and Rodrigo R Soares**, “Evaluating the impact of community-based health interventions: evidence from Brazil’s Family Health Program,” *Health Economics*, 2010, 19 (S1), 126–158.
- Sandler, Danielle H and Ryan Sandler**, “Multiple event studies in public finance and labor economics: A simulation study with applications,” *Journal of Economic and Social Measurement*, 2014, 39 (1, 2), 31–57.
- Schmitz, Sophia, Felix Weinhardt, and Jonas Jessen**, “Immigration and the Evolution of Local Cultural Norms,” Working Paper 2021.
- Shayo, Moses**, “Social Identity and Economic Policy,” *Annual Review of Economics*, 2020, 12 (1), 355–389.
- Skoda, Uwe, Kenneth Bo Nielsen, and Marianne Qvortrup Fibiger**, *Navigating social exclusion and inclusion in contemporary India and beyond : Structures, agents, practices*, New York, NY: Anthem Press: Anthem South Asian studies, 2013.
- Spears, Dean and Amit Thorat**, “The Puzzle of Open Defecation in Rural India: Evidence from a Novel Measure of Caste Attitudes in a Nationally Representative Survey,” *Economic Development and Cultural Change*, September 2019, 67 (4), 725–755.
- Srinivas, M.N.**, *Religion and society among the Coorgs of south India*, Bombay: Asia Publishing House, 1952.
- , *Social Change in Modern India* Rabindranath Tagore Memorial Lectureship Association for Asian Studies, University of California Press, 1966.
- Teso, Edoardo**, “The long-term effect of demographic shocks on the evolution of gender roles: Evidence from the transatlantic slave trade,” *Journal of the European Economic Association*, 2019, 17 (2), 497–534.
- Vermeersch, Christel, Andre C Medici, and Rory Narvaez**, “Universal Health Coverage for Inclusive and Sustainable Development: Country Summary Report for Peru,” 2014.
- Xaxa, V.**, *State, Society, and Tribes: Issues in Post-colonial India*, Dorling Kindersley (India), licencees of Pearson Education in South Asia, 2008.
- Xu, Ke, Agnes Soucat, Joseph Kutzin, and Callum Brindley**, “Public spending on health: a closer look at global trends,” Technical Report, World Health Organization 2018.