Essays in Public Finance and Education

by

Bhanu Gupta

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Economics)
in the University of Michigan
2020

Doctoral Committee:

Professor Hoyt Bleakley, Co-Chair
Professor Joel Slemrod, Co-Chair
Associate Professor Achyuta Adhvaryu
Professor James R. Hines

Bhanu Gupta

bhanug@umich.edu

ORCID iD: 0000-0001-8418-1328

ACKNOWLEDGEMENTS

This dissertation wouldn't have come to fruition without the constant support and guidance of the members of my doctoral committee. Working with Hoyt Bleakley benefited me in every dimension of research. Joel Slemrod was always generous with his time and providing feedback. I am grateful to James Hines for his guidance during tough times. It was always stimulating to talk to Achyuta Adhvaryu and discuss ideas regarding the dissertation. I am also grateful to my undergraduate mentors Abhiroop Mukhopadhyay, Nishith Prakash and Manisha Goel for investing in me so much.

I would forever be indebted to my mother, Vinod Gupta and father, Murli Manohar Gupta for their encouragement and support during my academic journey. None of this was possible without their unconditional love. I owe this PhD to SK Agarwal, SP Agarwal and other members of my extended family including Aashima, Niti, Ravinder, Shivani and Tushar.

Ann Arbor was a very different place from where I came from. I wouldn't have survived here for more than half a decade if it wasn't for the "Mezzanine crew". Thank you Thomas Brosy, Kimberly Conlon, Chiara Ferrero, Obeid Ur Rehman, Wenting Song, Merve Sariisik, Chih-Chan Tien, Robert Venyige and Huayu Xu for the heartfelt conversations and innumerable laughters. A special shout out is due to Nafisa Lohawala and the best roommate I have ever had, Nishaad Rao, for sitting through countless presentations of my job market paper. I believe Nishaad, when he says, that he can present my job market now without batting an eyelid.

Last, but not the least, I am grateful to a few people who have been the constants in my life. I thank Aditya Bisen, Sakshi Gupta, Siddharth Hari, Avinash Paliwal, Dimple Kukreja, Aparna Mudiam, Pratik Munjal and Anusha Ramakrishnan for putting up with my constant cribbing during this journey.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

LIST OF APPENDICES

Appendix

ABSTRACT

This dissertation studies the role of private agents in increasing tax compliance and documents the effects of fiscal and behavioral inefficiencies that distort human capital formation.

Chapter 1, joint with Keshav Choudhary, analyzes firms' response to changes in size-based exemptions from third-party audits and shows that third-party auditors help the government in raising tax compliance and government revenue. Using administrative panel data from India, we develop a novel empirical framework that considers dynamic responses of firm to such exemptions and conduct a difference-in-differences analysis. Our estimates suggest that firms remit 20 percent higher taxes and report 16 percent higher taxable income, once they are subject to third-party audits. Using these estimates, we conclude that the policy is cost effective and raises net social benefit.

Chapter 2, joint with Hoyt Bleakley, analyzes the fiscal costs and benefits related to schooling decisions. We use a Mincer-like model to characterize the fiscal externality associated with an additional year of schooling. Then we estimate the gap between fiscal benefits and costs by combining recently available data on school subsidies with the Nepalese household consumption survey. We find that within primary school, at a discount rate of 3 %, the fiscal benefits and costs, on the margin, are quite balanced, with subsidies closest to the present value of future taxes minus benefits. At higher levels of schooling, however, marginal fiscal benefits exceed costs by 5 percent of per-capita consumption.

Chapter 3 studies the effects of taste-based discrimination on the educational outcomes of the children in India. Using nationally-representative data, we find that there is a declining birth-order gradient in learning outcomes of both boys and girls. Conditional on birth-order, boys outperform girls which indicates general son-preferences. Moreover, after controlling for birth-order effects, boys with no elder brother perform better than boys who have an elder brother. This indicates preferences towards elder sons in the family. The gap widens in districts with higher elder-son preferences and results in a steeper birth-order gradient in those districts. Societal preferences affect learning through differential allocation of educational resources within a family.

CHAPTER I

Notching Really Matters: Effect of Third-party Audit on Tax Compliance of Firms

From a work with Keshav Choudhary

Abstract

Do third-party auditors act as watchdogs of tax administration or do they help firms misreport tax revenue? We answer this question by examining firm behaviour in response to audit notches—defined as discontinuities in the third-party audit requirement—and exogenous policy-induced changes in these notches over time. Using administrative panel data from India, we develop a novel empirical framework that considers dynamic responses of firms to a notch and conduct a difference-in-differences analysis. Our estimates suggest that firms remit 20 percent higher taxes and report 16 percent higher taxable income, once they are subject to third-party audits. Using these estimates, we conclude that the policy is cost effective and raises net social benefit.

JEL Codes: H26 , H32 , M42
Keywords:Third-party audit, Corporate tax, Notch, India

1.1 Introduction

Various government agencies rely on private auditors to complement their regulatory functions across wide-ranging spheres of governance. For instance, environment agencies use private auditors to report polluting firms. Financial regulators rely on the credit-rating agencies for maintaining financial stability. Tax administrations, the subject of inquiry of this study, also rely on private agents for various functions such as debt collection, customs, direct tax compliance, among others. However, in all these contexts, the effectiveness of these private auditors is not guaranteed because they are often hired and paid by the agents whom they are supposed to regulate–causing a conflict of interest. Conversely, factors such as effective regulations, reputational concerns, technical expertise, among others, can push these private auditors to overcome the conflict of interest and complement regulatory authorities.

While the role of private auditors has been analyzed in contexts such as environment (Duflo et al. (2013*a*)), credit ratings (Fracassi, Petry and Tate (2016), Griffin and Tang (2011)), social audit of global supply chains (Short, Toffel and Hugill (2016)), among others, this study is among the first that analyzes the role of private auditors in tax administration. We evaluate the Indian income tax department's use of third-party auditors to certify tax returns and report discrepancies. In the context of this policy, we investigate whether third-party auditors act as watchdogs on behalf of the tax department, or whether they instead help firms misreport their income to lower their tax liability. We find that these auditors are effective in increasing tax revenue despite the potential conflict of interest. The results are consequential for resource constrained tax administrations that might not have adequate financial, technical or human resources to effectively enforce tax law, and thus rely on private agents to augment their capacity.[1]

India provides an excellent setting to study the role of third-party auditors in tax administration for two reasons. First, we have rich administrative data that consists of anonymized income tax returns filed by Indian companies from FY 2009 to FY 2016. This data includes firm's revenue, taxable income, reported taxes and other firm-level characteristics. The advantage of using this data is that it provides information on small firms, both listed and unlisted on the stock exchange, that is not recorded in the conventional data sets used for analysis of firms in India[2]. Moreover, the panel structure of the data

[1] For instance, the Internal Revenue Service in USA has run several programs in the past decade, where private agencies are entrusted with collecting outstanding debts from the taxpayers. A list of functions outsourced across countries can be found in OECD (2019).

[2] Prowess database complied by CMIE, despite being a panel data set, is under-powered for analysis of small companies. Annual Survey of Industries has data on organized manufacturing firms. It contains repeated cross-section of 20 percent smaller manufacturing firms that may or may not be registered with tax

allows analysis of a firm's dynamic behavior in response to exogenous policy shocks. Additionally, we combine this data with other data sets to compute the differential effect of the third-party audit based on firm and industry-level characteristics.

Second, useful for our study is a provision in the Indian tax law that creates a discontinuity in the audit requirement – an audit notch. The tax law requires all companies whose gross revenue exceeds a specified threshold to undergo third-party audit. We can causally estimate the impact of third-party audit by exploiting a policy change that increases the audit notch over time, and thus provides quasi-experimental variation. Specifically, we exploit the change in 2012, when the threshold was increased from Rs.6 million to Rs.10 million,[3] providing some firms the opportunity to be exempted from third-party audit. We document the impact of exemption of third-party audit on a firm's tax payments.

The key empirical challenge is to estimate the impact of audit notch defined on a variable (revenue, in this case) other than the variable of interest (tax payments). A small but growing literature addresses this challenge by the using a combination of static bunching analysis, developed by Kleven and Waseem (2013), and other estimation strategies such as differences-in-differences (Hamilton (2018)) and regression discontinuity (Bachas and Soto (2018)). In this study, we use a difference-in-differences design after accounting for the bunching response of the firms. The first step is to find the treatment neighborhood in which firms respond to the audit notch by manipulating their revenue to bunch below the notch and escape audit. Static bunching analysis is the standard method to estimate this neighborhood.

While the conventional static analysis works well in a static one-shot setting, it may not work well if the firms have dynamic responses to a notch (Kleven (2016)). Specifically, the static analysis relies on the ocular method to determine the lower bound of the treatment neighborhood by assuming that all the bunchers originate from the right side of the threshold and bunch precisely at the threshold. The upper bound of the treatment neighborhood is found by constructing a counterfactual density so that the excess bunching mass below the threshold is equal to the missing mass above it. However, we consider the possibility that bunchers can originate from the left-side of the threshold in a dynamic setting – causing diffused excess mass. There can be several reasons for the firms to not bunch precisely at the threshold. For instance, if firms believe that reporting zero growth for multiple time-periods will increase the probability of getting caught, then they are likely to arrest their growth to bunch well below the threshold rather than precisely at it. In such a scenario,

authorities. Thus, it is not representative for smaller tax paying firms.

[3]The average exchange rate from 2009-16 was Rs.55.65 per dollar. This implies that the threshold increased from \$10,782 to \$17,969.

the excess mass will be diffused, causing researchers to miss some of the excess mass while constructing the counterfactual density using the static method and thus, underestimate the upper bound.

We develop a novel methodology that accounts for the dynamic responses of the firms to the notch. We estimate the size of the biggest firm that manipulates its revenue in response to the audit notch by leveraging the panel nature of the data and exogenous policy change. If the notch is introduced in year t, we estimate a firm's probability of reporting revenue just below the threshold in $t+1$—firms take at least a year to respond to the notch—based on its lagged revenue in $t-1$. However, a firm's natural growth path can also cause its revenue to be just below the threshold. We quantify this confounding effect by estimating the firm's probability of reporting revenue below the threshold in $t-1$, pre-notch period, based on lagged revenue in $t-3$. If the difference in probabilities is positive for a firm, then it is responding to the audit notch. A placebo test is used to rule out the possibility of macroeconomic trends driving the differences in probabilities across years. Using our improved methodology, we estimate that any firm having revenue up to Rs.15 million in 2011 has the opportunity to escape third-party audit by bunching below the threshold - defined at Rs.10 million. In comparison, the upper bound of the neighborhood in which the firms respond to the notch as calculated by the static analysis is considerably lower, at Rs.11.5 million.

Now, we can use a differences-in-differences design to calculate the intention-to-treat effect of the removal of audit requirement on tax payments. We use RST or restricted share transfer firms—defined as firms which have restrictions on the sale of their securities to non-shareholders—as the treated group if their revenue in the year before the policy change is in the treatment neighborhood. Firms without any restrictions on the sale of their securities, non-RST firms, who report revenue in the same neighborhood, form the comparison group.[4] Even though non-RST firms are also subject to third-party audit, the extra audit does not affect their tax payments. In contrast to the RST firms, the managers of non-RST firms are less likely to be shareholders[5] and as a result, their share of extra profits earned by under-reporting taxes is smaller. Thus, the managers of non-RST firms are less likely to evade taxes, with or without third-party reporting (Crocker and Slemrod

[4]We define firms that register as private firms under The Companies Act as RST firms and firms that register under the category of public firms as non-RST firms. Both kind of firms can be privately held or government owned. In contrast to RST firms, the non-RST firms have the option to list themselves on the stock exchange.

[5]The proportion of shareholders to managers is comparatively larger in non-RST firms than RST firms due to statutory and operational requirements (see appendix A.1.1). Thus, the probability of a manager being a shareholder is lower in a non-RST firm than a RST firm.

(2005)).[6]

We have three main findings. First, the firms respond to the audit-notch by bunching below the threshold to escape third-party audit. We also find evidence that the bunching of firms is closely related to the shifts in the thresholds during the sample period. Furthermore, firms that bunch below the threshold report lower fee paid to the auditor suggesting that these auditors are no longer doing tax audits of the firms.

Second, third-party auditors, despite the conflict of interest, raise tax compliance by increasing tax payments of the firms. We reach this conclusion by flipping the difference-in-differences result – RST firms reduce their tax payments by Rs.41,000 ($736) once they have an opportunity to escape third-party audit, as compared to non-RST firms. This represents a 20 percent reduction in taxes from their base-year mean tax liability. The decline in the taxable income is Rs.102,000 ($1,832), which is 16 percent of the base-year mean income.

Third, the audit effectiveness varies with firm and industry-level characteristics. Firms with fewer workers show greater reduction in their tax payments because their resource cost of evasion is lower. Additionally, the paper-trails generated by the industry in which the firm operates is positively correlated with the reduction in tax payments. One possible explanation could be that third-party auditors can easily verify traceable transactions and thus, firms with many such transactions have stronger incentives to escape third-party audit.

Finally, we conduct a cost-benefit analysis of the notched audit policy by using the estimated coefficients and a framework described in the literature (Almunia and Lopez-Rodriguez (2018), Slemrod (1994), Slemrod and Yitzhaki (1985)). The marginal cost is the fee charged by the external auditor. The marginal benefit is the reduction in expenditure incurred by firm to cover-up tax evasion. The amount of evasion is itself lowered due to the extra audit. Our calculations suggest that adding one firm under third-party audit increases both government revenue and net social benefit. Relatedly, the optimal audit threshold might be lower than the current statutory limit, under strong assumptions.

This study contributes to several strands of literature. Broadly, by evaluating the role of private agents in governance, it contributes to the literature on building state capability to achieve development goals (for an extensive discussion see Andrews, Woolcock and Pritchett (2017)). Specifically, we speak to the literature on third-party audits where both theoretical (Bolton, Freixas and Shapiro (2012)) and empirical literature (Duflo et al.

[6]Crocker and Slemrod (2005) show that penalties imposed on the tax manager are more effective in reducing evasion than those imposed on shareholders, even in the presence of contracts that incentivize managers to evade taxes.

(2013*a*); Jiang, Stanford and Xie (2012); Griffin and Tang (2011)) have concluded that conflict of interest overwhelms the effective functioning of private auditors. On the other hand, some work has shown that a few auditors do serve their clients effectively because of reputational concerns and demand for quality (Stolper (2009); Duflo et al. (2013*b*)). This paper's finding that private auditors can be effective in raising tax compliance, while being cost-effective, provides a scalable and sustainable policy option to augment state capacity.

The strength of this study lies in leveraging the administrative data to assess the impact of third-party audit on the key outcome variables – tax payments and taxable income. Measuring the responsiveness of tax payments to audit intensity, focus of this study, is necessary to design the optimal design of tax administration, and to evaluate the trade-off between improving tax administration versus changing the tax rates to raise government revenue (Keen and Slemrod (2017)). Additionally, this study complements to the literature of decreasing tax evasion by using various instruments such as government audits (Basri et al. (2019), Almunia and Lopez-Rodriguez (2018)); third-party information (Pomeranz (2015), Kleven et al. (2011)); performance pay (Khan, Khwaja and Olken (2016)), among others.

Our final contribution is methodological. Bunching analysis has become a preferred method to estimate the impact of a notched policy in several settings–optimal taxation (Kleven and Waseem (2013), Bachas and Soto (2018)); housing markets (Best and Kleven (2018)); fuel economy (Ito and Sallee (2018)), among others. There is a nascent literature on dynamic bunching that improves the static analysis by incorporating path-dependence (Marx (2018)) and predicting the counterfactual value of the running variable using non-parametric methods (Blomquist et al. (2018), Bertanha, McCallum and Seegert (2019)). We propose an alternate methodology to study the effect of a notch by incorporating strategic concerns.

The third-party audit policy experiment has been studied in a working paper by Tantri (2017) using Prowess data which, as discussed before, contains a non-representative sample of small firms. He finds that firms which are exempt from third-party audit pay higher taxes. The explanation is that under third-party audit firms pay a portion of their taxes to auditors as bribes, and once these firms are exempt from the third-party audit, they increase the tax payments as following the law of land increases their utility. The identification strategy uses firms above the notch, in the year before the policy change, as control group, while excluding firms below the notch from the treatment group. This can potentially bias the results because administrative data shows that the firms that were slightly above the audit notch in the year before the policy change start bunching below the threshold after the introduction of the notch. These firms, therefore, should be part of the treatment group. This paper, on the other hand, develops a theoretical framework to incorporate dynamic responses of the firms to a notch without any assumptions regarding honesty in the utility function. It also studies a distinct set of questions. We leverage data from the universe of taxpayers and use a novel identification strategy that combines dynamic bunching and differences-in-differences methods.

1.2 Context: Tax Law, Third-party Auditors and Companies

In this section we describe the salient features of corporate income tax law, the role of private auditors in implementing the law, and the differences between restricted share transfer (RST) and non-RST firms relevant for our analysis.

This study focuses on the firms that register as companies under the Companies Act and are recognized as a separate category of tax payers under the Indian income tax law.[7] The corporate income tax rate was 30 percent for all the domestic companies from 2009-15 and was reduced to 29 percent in 2016 for companies with revenue less than Rs.50 million in 2014.[8] In some cases, the company might have to pay a minimum alternate tax.[9]

All companies above a specified revenue threshold undergo third-party tax audit. This threshold was increased twice during our sample period. It was increased from Rs.4 million to Rs.6 million in 2010 and then to Rs.10 million in 2012. This audit is done by licensed auditors called Chartered Accountants (CA) and the audit reports must be filed along with the tax returns to the tax department.[10]. However, the tax department can officially audit any company independently. The third-party tax audit places significant compliance requirements on the company, which must now get minute tax-related details of its books of accounts certified by the CA. Examples of extra information include: (i) Details of all persons from whom loans have been taken or given during the year; (ii) De-

[7] There are seven category of taxpayers – Individuals, Hindu Undivided Families, Companies, Partnership Firms, Association of Persons, Local Authorities and Artificial Juridical Entities

[8] In addition to the tax, there is a surcharge for companies having net taxable income in excess of Rs.10 million and an education cess of 3 percent on the tax and surcharge paid. Less than 4 percent of the companies in our sample were liable to pay cess and surcharge in any year. More than 95 percent of the companies in our sample qualified for the tax rate cut in 2016.

[9] If the taxes paid by the company are less than 18.5 percent of the book profits then the company has to pay MAT. The extra tax paid by the company, over and above its normal tax liability is available to it as MAT credit. This credit can be carried forward for up to 15 years and adjusted against actual tax liability in the future.

[10] All firms must undergo another audit, known as statutory audit, which is filed to the Ministry of Corporate Affairs (MCA). This is also done by a CA which can be the same person conducting tax audit. According to our conversations with the tax officials, there is no exchange of information between the MCA and the Ministry of finance which is responsible for tax audits.

tails of all persons whose tax has been withheld; (iii) Quantitative description of stock; (iv) Depreciation schedule of every asset; (v) Details of payments to related parties. Besides information cost, the firms also have to pay an auditor's fee to the CA. The fee is determined according to several factors such as complexity of the firm's operations, location of the firms, among others. Non-compliance with respect to the tax audit report carries a penalty of 0.5% of the turnover of the company subject to a maximum penalty of Rs.150,000.

The Chartered Accountants are a class of professional private auditors in India who get their license after clearing a competitive exam. They are hired and paid by the firms to do the tax audit, which can create a potential conflict of interest. On the other hand, there are several checks to help them perform their duty independently. First, the appointment of the CA must be done by the board of directors. This means that the CAs cannot be ordinarily removed by the company once appointed unless there are valid grounds to do so which are required to be recorded. Second, they are regulated by the Chartered Accountants Act and the Institute of Chartered Accountants of India (ICAI), a self-regulating professional body. This body has drafted an ethical code which mandates that CAs cannot receive excessive fees, and it has the power to revoke the license in case of violations to the ethical code. Lastly, if the government agencies find inconsistencies in the report of the CA, they can recommend imposition of fines and even cancellation of their license to practice[11].

Firms under the Companies Act can register themselves as private or public firms, which affects their tax compliance. The latter have restrictions on the transfer of their shares to non-shareholders, which is why we call them restricted share transfer firms. Public or non-RST firms have no such restrictions. Additionally, the minimum number of shareholders of the company is higher for non-RST firms and the salary of their managers is capped. (Appendix A.1.1 describes the differences in more detail.) These differences make non-RST firms more tax compliant for two reasons. First, the managers of non-RST firms might have to split the gains of tax evasion with more shareholders, while facing similar punishment as a RST firm's manager, if they get caught. Second, if a non-RST firm lists itself on the stock exchange, then it has to release information of any tax dispute in the public domain which might decrease the valuation of it's stock. Because of these institutional reasons, the non-RST firms are more tax compliant with or without the third-party audit and make for a valid comparison group. We provide evidence of this claim in the Data and Results sections.

[11] According to Institute of Chartered Accountants of India Year Book (2018), there are 132,480 practicing CAs in 2018. Total disciplinary cases considered by ICAI were 598.

1.3 Theoretical Framework and Empirical Strategy

1.3.1 Strategic response of firms to a notch

We use differences-in-difference strategy to estimate the intention-to-treat effect of the third-party notch on tax payments. We exploit an exogenous policy change that increased the revenue threshold of the notch to estimate the treatment effect. In response to the introduction of audit notch, firms in its neighborhood manipulate their revenue to bunch below the threshold. Thus, we need to identify the revenue bandwidth in which firms are potentially affected by the change in the audit notch. The standard method to find this bandwidth is static bunching model, developed by Kleven and Waseem in 2013.

In the absence of dynamic behavior, the static model works well. It assumes that bunchers originate from the right-side of the threshold and bunch precisely below it. Hence, the researchers can visually determine the region where the excess mass of bunchers are located in response to the notch. The region from where the bunchers originated from is determined by estimating a counterfactual density after imposing the condition that the excess mass is equal to missing mass. If firms respond to the notch in a strategic way, then the treatment neighborhood estimated by the static model might be biased. For instance, some firms may slow down their growth resulting in diffused excess mass. The visual method of determining where the bunchers come from may miss some of this excess mass. Consequently, the upper bound of the region where the bunchers originated from would be under-estimated.

We now present a formal framework to develop the above intuition, and address the concerns in the empirical section.

1.3.1.1 Baseline Static Model of Homogeneous Firms

We modify the model presented in Almunia and Lopez-Rodriguez (2018), henceforth referred to as AL(2018), to analyze the behavior of firms to an audit notch. Consider a firm that uses inputs x and z to produce output y according to the production function $y = \psi f(x, z)$, where ψ is the productivity parameter and $f(.,.)$ is continuous, increasing and concave in both the arguments. The firms vary according to the exogenous parameter ψ which is distributed according to the density $d_0(\psi)$ over the base $[\underline{\psi}, \bar{\psi}]$. The prices of x and z are w and q respectively, while y is the numeraire good.

Let τ be the tax rate on reported profit, and x be the only tax-deductible input. The firm chooses to report $\bar{y}$ and under-report income y by $u \equiv y - \bar{y}$. There is a strictly increasing, continuous and convex resource cost of under-reporting given by $k(u)$. The probability of the firm getting caught, in any period, is given by $\delta = \phi h(u)$, where ϕ is the effective audit intensity faced by the firm. This includes audit by both the tax authorities and the CAs[12]. $h(.)$ is increasing and convex in u. If the firm get caught, it faces a penalty rate of θ on evaded taxes. The firm also pays an audit fee of c to the auditors.

The firm chooses x, z and u to maximize expected profits, given by: $E[\pi] = (1 - \tau)[\psi f(x,z) - u - wx] - c - qz - k(u) + u - \phi h(u)[\tau u + \theta\tau u]$. The FOCs are:

$$\psi f_x(x,z) = w$$

$$\psi f_z(x,z) = q/(1-\tau)$$

$$\tau[1 - \phi h(u)(1+\theta)] = k_u(u) + \tau u(1+\theta)\phi h_u(u) \tag{1.1}$$

Since all the firms are similar in terms of production technology and the functional forms of resource cost of under-reporting and audit intensity; $\exists$ density function of reported revenue $g_0(\bar{y})$ which is decreasing and convex in the domain $[\bar{y}(\underline{\psi}), \bar{y}(\bar{\psi})]$.

1.3.1.2 Multi-period Model of Heterogeneous firms with Strategic Mis-reporting

In this sub-section, we depart from the AL model by assuming a multi-period model where firms experience growth of revenue. We also model the change in the audit intensity by introducing the audit notch. A combination of heterogeneous resource cost of evasion and the belief that reporting zero growth for multiple time periods can increase the probability of caught causes diffused excess mass - contrary to what is presumed in the static analysis.

At t_0, the firms get a random productivity drawn from the underlying distribution, and then grow by a factor of γ in each period. For analytical simplicity, in the counterfactual baseline situation all firms are subject to third-party audit which makes them report their true income in all the time periods. The shift in the density from time period t_0 to t_1 is shown in Figure 1.1. Note that because firms differ in terms of resource cost of evasion, there now exists a a joint distribution of firms with density $\tilde{g}(\psi, h(u))$ on the domain $(\underline{\psi}, \bar{\psi}) \times (\underline{k}(u), \bar{k}(u))$.

Now, instead of all firms being subject to third-party audit, an audit-notch is introduced in t_1 at the income level y_ρ. Firms below this threshold are exempt from the third-party audit. The exemption decreases the audit intensity by $d\phi$. Firms that have lower cost of evasion can evade more, once they are exempt from third-party audit. Figure 1.2 depicts the change in the density due to introduction of the audit-notch. The green dashed line represents the density in presence of audit-notch. According to the behavior in the presence of notch, the firms can be classified into 4 categories:

[12] ϕ represents "monitoring effort parameter " in AL(2018)

1) *Small firms* – Let $\bar{y}_L$ be the income at t_0 where an income growth of γ results in income of y_ρ in t_1. Therefore, $\bar{y}_L$ is equal to $y_\rho/(1+\gamma)$. Firms with productivity draw in the range of $[\underline{\psi}, \psi^L)$ and income $[\bar{y}(\underline{\psi}), \bar{y}_L)$ get exempt from the third-party audit after the introduction of the notch. These firms don't undergo third-party audit even if they report their entire growth. On average, the small firms will under-report some portion of their growth resulting in downward shift of the density.

2) *Potential Left-origin bunchers* – Firms with income in the range $[\bar{y}_L, y_\rho)$ in t_0 will grow above the threshold limit in t_1. Firms with small resource cost will bunch below the audit-threshold. A proportion of bunchers will strategically mis-report to be well below the threshold rather than at the threshold, if they think that reporting zero growth can substantially increase their probability of getting caught[13] [See A.1.2 for a formal proof of this claim]. These firms will report small incremental growth in future time periods to remain below the threshold, and avoid the attention of tax authorities. As a result, the bunchers will not be concentrated precisely below the notch, but diffused in an area below the notch. This is shown as a plateau in Figure 1.2 for illustrative purposes.

Firms that have high resource cost will not react to notch, and report growth. Consequently, there would be diffused mass above the threshold instead of a hole in the density – which is also predicted in the static model.

3) *Potential Right-origin bunchers* - There would be a firm with income $\bar{y}_H$ in t_0 such that $\bar{y}_H > \bar{y}_\rho$, which will be indifferent between remaining above or bunching below the notch. For this firm, the $E[x, z, u \mid \phi, \psi^H, c] = E[x, z, u \mid (\phi - d\phi), \psi^H, 0]$. Among firms with reported income in the range $[y_\rho, \bar{y}_H)$ in t_0, a proportion will bunch below the threshold if their resource cost of evasion is not high. As compared to left-origin bunchers, smaller fraction of right-origin bunchers will strategically under-report far below the threshold. The reason is that they are mis-reporting more than the left-origin bunchers in terms of levels to get below the notch, which increases their resource cost of evasion.[14] The remaining firms with high resource cost will not bunch and continue to be under third-party audit.

4) *Big firms* - For all the firms with income above $\bar{y}_H$ in t_0, the cost of bunching (resource cost and increase in probability of getting caught) is strictly greater than the benefit. These firms will report the growth and continue to be under third-party audit. This implies that

[13] Not all the bunchers will strategically mis-report, because of heterogeneity in resource cost. There could be other sources of friction like uncertainties in returns on investment which translate into stochastic growth and firms are not able to reach the notch precisely. Such factors are similar to optimization frictions that result in diffused missing mass in the static model of Kleven and Waseem (2013).

[14] Also, the optimization frictions would be less for right-origin bunchers as compared to the left-origin bunchers, because they have to misreport on $\bar{y}_0$ to reach the notch. In contrast, the left-origin bunchers have to make investments with uncertain returns to grow from $\bar{y}_0$ and reach the threshold precisely.

in Figure 1.2, the density in t_1 after the income level of $\bar{y}_H + \gamma\bar{y}_H$ remains unaffected due to the introduction of the notch.

Bias in static analysis: In the empirical application of the static analysis, the lower bound (the minimum value of the running variable where the bunchers locate themselves) is determined visually to be the point where the density has positive slope. In figure 1.3, this would be $\bar{y}_L^S$. In our model, the excess mass of bunchers starts from $\bar{y}_L$ due to strategic mis-reporting and optimization frictions. Error in determining the lower bound will cause under-estimation of the actual mass of bunchers. The upper bound of bunching (maximum value of running variable where the bunchers come from) is estimated by fitting a counterfactual density so that the excess mass is equal to the missing mass. The upper bound in the Figure 1.3 is $\bar{y}_H^S$ which is an under-estimate of the true upper bound: $\bar{y}_H + \gamma\bar{y}_H$.

1.3.2 Empirical Strategy

Given the concern that the estimates from static analysis can be biased, we first describe a novel method to identify the treatment region where the firms manipulate their revenue in response to the introduction of the notch. Then, we explain our difference-in-differences model to estimate the effect of third-party audit on the firm's tax payments.

1.3.2.1 Estimation of the treatment neighborhood

Recall that the policy change relevant for our analysis is the change in audit notch from Rs.6 million to Rs.10 million. Hence, the lower bound of the treatment neighborhood is equal to Rs.6 million as all firms between Rs.6 million and Rs.10 million are now exempt from the third-party audit. To estimate the upper bound of the treatment neighborhood, we calculate the probability of being in the bunching region conditional on lagged income, where the bunching region is defined as the revenue bin just below the notch. A firm i is placed in bin b_t of width ω if its reported income in a year t, $\bar{y}_{it} \in [b, b+\omega)$. For all the firms in the bin b_{t-1}, the probability of being in the bunching region(BR) in the year t is given by:

$$\Pi_{t,b_{t-1}} = \sum_i [\mathbb{1}(\bar{y}_{it} \in BR) \mid (\bar{y}_{i,t-1} \in b_{t-1})] \Big/ \sum_i \mathbb{1}(\bar{y}_{i,t-1} \in b_{t-1}) \tag{1.2}$$

For our analysis, if the policy change happens in period t, then we calculate the probability of being in the bunching region one year later $(t+1)$ conditional on income in the year before the policy change $(t-1)$. This probability is calculated one year after the policy change as firms take some time to adjust their revenue levels in reaction to the policy change.

However, a firm can also be in the bunching region due to other reasons such as natural growth. We estimate the effect of these other reasons by calculating the probability of being in the bunching region before the introduction of the notch ($\Pi_{t-1,b_{t-3}}$) conditional on two-year lagged income. The difference between the two calculated probabilities will capture the incentive of firms to be in the bunching region in response to the notch. The upper bound of the treatment region is b^* such that,

$$\Pi_{t+1,b^*_{t-1}} - \Pi_{t-1,b^*_{t-3}} = 0 \quad \text{and} \quad b^* > y_\rho$$

where, y_ρ is the revenue-level of the notch. Any firm with $y_{i,t-1} \in [y_\rho, y_{b^*+\omega}]$ has the opportunity to modify its revenue, once the threshold moves to Rs.10 million. Firm with revenue above b^* in the year before the policy change is too big to react to the introduction of the notch. In the Results section, we also develop a formal method to test if the probability of being in the bunching region is significantly different for firms in the treatment region versus non-treatment region.

A potential concern with our method is that inter-temporal macroeconomic changes in the economy can lead to mechanical differences between the predicted probabilities of being in a particular bin across years. To alleviate such concerns, we calculate the probability of being in a placebo region that is unrelated to any notch. We expect the difference in probabilities to be insignificant for the placebo region, to rule out the effect of macroeconomic changes.

1.3.2.2 Difference-in-Differences Analysis

After identifying the treatment region, we can calculate the intention-to-treat (ITT) effect $\mathbb{E}[Z_i \mid T_i = 1] - \mathbb{E}[Z_i \mid T_i = 0]$, where Z_i is the outcome of interest of firm i and T_i is the treatment assignment. We consider RST firms in the treatment region to be affected by the notch ($T_i = 1$), whereas the non-RST firms in the treatment region remain unaffected by the notch($T_i = 0$). The institutional reason behind non-RST firms being more tax compliant are discussed in Section 1.2.

We use a difference-in-differences strategy to ensure that the differences in means between the RST and non-RST firms capture the effect of the audit-notch and not any other confounder. Taking advantage of the panel structure of the data, we estimate regressions of the form:

$$TaxDue_{ist} = \alpha_i + \beta treat_i \times after_t + \lambda_t + \gamma_s t + u_{ist} \tag{1.3}$$

where the dependent variable is tax paid by firm i operating in industry s at year t. The coefficient of interest is β which is interpreted as the difference in tax payments between the comparison and treatment firms, before and after the change in policy in 2012. Regressions include firm fixed effects to control for time-invariant differences across various firms. Year fixed effects capture any macro-economic changes that affect both the treatment and comparison group in any given year. Finally, we also include industry-specific time trends to control for any heterogeneous trend in the tax payments across different industries.

To validate the Difference in Difference estimates, we establish the absence of any pre-trends between the treatment and comparison group by conducting an event-study analysis. We obtain year-wise diff-in-diff coefficients from a regression of the following form:

$$TaxDue_{ist} = \alpha_i + \sum_j \beta_j \; treat_i \times \mathbb{1}(year = j)_t + \lambda_t + \gamma_s t + \epsilon_{ist} \tag{1.4}$$

where the coefficients of interest are β_j which are equal to the difference in tax payment between the treatment and comparison firms in year j, as compared to the base year. The base year is defined as the year before the policy change.

1.4 Data

Our primary data source is anonymized tax returns filed by small companies in India from 2009-16 which are obtained from the Ministry of Finance, India. It also includes the profit and loss account and balance sheet of the company. For our analysis, we use the reported revenue; details of expenses; auditor's fee; and profit before interest, taxes and depreciation(PBITD) from the profit and loss statement of the company. As discussed before, the firms hire auditors for both tax and statutory audit. The tax form records the total fee paid and doesn't contain audit-wise fee details. Information on taxable income and tax liability is obtained from the tax schedule of income tax form. The data also has information on the basic characteristics of the company like the sector of activity which we use for heterogeneity analysis.

Next, we describe the construction of the sample used in the analysis. Due to the large differences in the mean and median of the variables, except revenue (see Table 1.1), we remove the effect of the outliers by winsorizing the data above the 97th percentile to that level for each year. For variables that can have negative values, we censor the tales of the distribution at 1.5th and 98.5th percentile. There can also be extensive margin responses to the change in audit notch. For instance, a firm might split into two firms and both may report revenue just below the notch. This might affect the composition of the sample after the policy change. Therefore, we use a balanced sample of firms in our main analysis and test if the results are sensitive to this assumption. We also observe that some firms in the sample report zero revenue. These firms could be shell companies and might react very differently to third-party audit. Thus, we exclude them from our main analysis and only include them in the sensitivity analysis. Finally, some firms switch between RST and non-RST firm status. Less than 5 percent of the firms switch their status in any given year. Among the firms that do change their status, 58 percent firms change their status more than once in the sample period, which suggests errors in coding or mistakes while filling up the tax form. We include such firms only in the sensitivity tests. Our final sample consists of around 21,800 firms.

The summary statistics also validate the assumption of using non-RST firms as comparison group as they tend to more tax compliant than RST firms. For instance, before 2012, when both treatment and comparison firms were under third-party audit, the tax to revenue ratio of non-RST firms was, on average, 2.4 percent while for the RST firms it was 1.6 percent. A possible reason is that RST firm report lower business profits as a proportion of revenue.[15]. Business profits are considered to be most manipulable part of the tax return as the taxpayers do not need to submit any supporting documents to verify the level of business income. RST firms report only 11 percent of their revenue as profit, while the non-RST firms report around 20 percent of their revenue as profits.

Finally, non-parametric evidence indicates that RST firms react sharply to the changes in audit notches while non-RST firms are not that responsive. Appendix Figures A.1 to A.8 show that as the audit notch increases, the mass of RST firms disappears from below the old threshold and appears instead at the new threshold. On the other hand, the non-RST firms bunch at round numbers that may or may not be related to the audit thresholds. Appendix Figure A.9 shows the comparison in bunching behavior between RST and non-RST firms. The data is pooled for all the years after the introduction of the notch to increase the sample size of non-RST firms.

1.5 Results

1.5.1 Size of the treatment neighborhood

In this section, we estimate the upper bound of the treatment neighborhood using the difference in probabilities method described in Section 1.3.2.1. The lower bound is Rs.6

[15]Business profits are called PBITD in the tax returns, that is, Profits before interest, taxes and deductions

million where the previous notch was defined. First, we calculate the probability of a firm to report revenue in the bunching region after the introduction of the notch at Rs.10 million, conditional on its revenue before the introduction of the notch. The bunching region is defined as Rs.9 million to Rs.10 million.[16] The blue curve, in Figure 1.4, shows the probability of being in the bunching region in 2013, conditional on revenue in 2011. We calculate the probability in 2013 instead of 2012 (the year of the policy change) to allow firms to react to the policy change. However, the firms can in the bunching region for reasons other than the effect of the notch. We estimate the effect of these other reasons by calculating the probability of being in the bunching region in 2011, the year before the introduction of the notch, conditional on revenue in 2009 – shown by the red line in Figure 1.4. The revenue bandwidth in which there is a positive difference between the two probabilities is where the firms changed their revenue in response to the notch. Figure 1.4 shows that the firms with revenue between Rs.8 million and Rs.15 million are more likely to bunch below the notch after the introduction of the notch in 2012. Thus, we consider Rs.15 million to be the upper bound of the treatment neighborhood.

A concern is that the slowdown in the Indian economy during these years can mechanically drive the differences in probabilities, as firms grow slowly. To alleviate this concern, we conduct a placebo test where we graph the probability of being in revenue bins that are unrelated to the notch. We calculate the probabilities of a firm reporting revenue between Rs.18-19 million and Rs.24-25 million, before and after the policy change. Appendix Figures A.10 and A.11 show that there are no systematic differences in the probabilities.

Next, we use the difference-in-probabilities method to provide evidence that non-RST firms are a valid comparison group. Appendix Figure A.12 shows that only some non-RST firms on the left-side of the notch are more likely to report revenue in the bunching region after the policy change. If a small proportion of public firms do respond to the notch, this will cause the difference-in-differences estimates an under-estimate of the treatment effect.

Finally, we use a regression framework to test if the firms that were in the treatment neighborhood, before the policy change, are more likely to report revenue in the bunching region after the policy change as compared to firms that are were just outside the treatment neighborhood. The later type of firms are either too small or too big to react to the change in the audit notch. We run the following specification at the level of revenue bins:

[16] Any firm below the notch is exempt from undergoing third-party audit. So ideally, we should calculate the probability of reporting revenue less than Rs.10 million. We don't do this, because the notch was previously defined at Rs.6 million. Thus, the probability of being just above Rs 6 million will be lower in the pre-period which can inflate the differences in probabilities across time-periods.

$$\Pi_{t,b_{t-2}} = \alpha_b + \beta_1.after_t + \beta_2.\mathbb{1}(b_{t-2} \in [k_1, k_2]) \times after_t + \epsilon_{t,b_{t-2}} \tag{1.5}$$

where $\Pi_{t,b_{t-2}}$ is the probability of a firm reporting revenue in the bunching region in year t, conditional on reporting revenue in bin b in the year $t-2$. k_1 and k_2 represent the lower and upper bounds of the treatment neighborhood. From Figure 1.4, we infer these to be Rs.8 million and Rs.15 million, respectively. We expect β_2 to be positive and significant for RST firms when we correctly specify the bunching region. Appendix Table A.1 shows that if a RST firm was in the treatment neighborhood, then it was 2.18% more likely to be in the bunching region after the policy change as compared to a firm that was not in the treatment neighborhood (see column 1). On the other hand, for non-RST firms, the coefficient is statistically indistinguishable from 0 (column 4), suggesting that, in comparison to RST firms, the non-RST firms are not as responsive to the change in audit notch. We also conduct a placebo test by mis-specifying the bunching region and using the sample of RST firms. Columns 2 and 3 show that, as expected, there is no significant change in the probability after the policy change .

1.5.2 Effect of third-party audit on government revenue and firms

After identifying the treatment neighborhood, we calculate the ITT effect, where treatment is defined as getting exempt from third-party audit after the audit notch was moved from Rs.6 million to Rs.10 million in 2012. RST firms, which have revenue between Rs.6-15 million in 2011, form the treatment group, while non-RST firms in the same revenue bandwidth form the comparison group.

First, we calculate the effect of policy change on tax payments using the difference-in-differences estimation strategy given by equation 1.3. After the change in the policy, the RST firms reduce their tax payments by around Rs.41,000 as compared to the non-RST firms (see columns 1 and 3 of Table 1.2), which is equal to 20 percent of the average tax payment or approximately 5 percent of median business profits in 2011. Thus, exemption from third-party causes a significant drop in the tax payments of the firms. If we flip this result, it implies that third-party auditors are effective in increasing the government revenue, despite a potential conflict of interest. This result is also consistent with the bunching of firms observed just below the audit notch as the firms try to avoid paying extra taxes.

The identification strategy relies on the absence of pre-trends between the treatment and comparison groups before the change in the policy. We use an event analysis approach as described by equation 1.4 to test pre-trends. Columns 2 and 4 of Table 1.2 show that before the change in the policy in 2012, there are no significant differences between the tax payments of RST and non-RST firms. The differences becomes negative and even grows over time after the change in the audit notch (see Figure 1.7). This is expected because some of the treated firms might slow down their growth to remain below the audit notch, while the comparison group firms will continue to grow.

To decrease their tax payments, the firms might reduce their taxable income by under-reporting income. We find that there is a significant drop in the taxable income reported by RST firms after the policy change. The decline in the taxable income is approximately Rs.102,000, which is equal to 16 percent of the average taxable income in 2011.[17] Some sources of income are easier to manipulate than others. For instance, business profits can be mis-reported with relative ease since firms don't have to give any extra information on the tax form to prove the accuracy of their claim. Conversely, it is harder to mis-specify other sources of income such as rental income or claim tax-exemptions like area-based exemptions, which require documentary proof. Columns 1 & 2 of Table 1.3 show that RST firms, as compared to the non-RST firms, decrease their business profits after the policy change. There is an average decline of around Rs.700K (56 percent of the average profits in 2011) in the PBITD i.e. profit before interest, tax and depreciation. To rule out pre-trends, we estimate year-wise difference-in-difference coefficients, and plot them using Figure 1.8-1.10. Reassuringly, the difference between private and public firms for all the variables is insignificant before the policy change.

Finally, we check if there is a decline in fee paid by the company to the auditors as firms are exempt from the third-party audit. Table 1.3 also shows a significant decline in the fee paid to the auditors by the RST firms. The decline is modest in magnitude (9 percent of the average fee paid in 2011) as the firms report combined expenditure on both the statutory and tax audit in the tax form.

Sensitivity Analysis: Here, we address concerns regarding sample selection and describe placebo tests to show that our results are not driven by factors unrelated to the policy changes such as slowdown in the economy.

Appendix Table A.2 shows the difference-in-differences results as we relax the constraints imposed while constructing the sample. In our main analysis, we censored the variables at the 97th percentile to remove the effect of the outliers. Panels A and B show that as we include these outliers, the treatment effects become more salient. Thus, the

[17] The decrease in tax payments, found eariler, is not exactly equal to 30 percent of the decline in the tax base. The difference might be because of the minimum alternate tax (MAT) provision in the tax code. In the appendix table A.5, we show that firms that pay taxes under MAT show more decline in the tax payments than firms that pay taxes under the regular income tax schedule, although the coefficient is not significantly differently from 0, as only a few firms pay MAT.

magnitude of the treatment effects that we find in the main analysis can be viewed as conservative estimates of the true effects. To exclude the extensive margin responses to the policy change, we used a balanced panel of firms in the main analysis. If we use the unbalanced sample instead, the results remain statistically unchanged (see Panel C). In Panel D, we include firms that report nil revenue in the sample period. These firms were excluded in the main analysis because they could be shell companies. We find that the results are similar to the ones found in the main analysis. We now include firms that change their status between RST and non-RST firms during the sample period. Panel E shows that the magnitude of the coefficients decreases, however, they remain qualitatively similar to ones we found earlier. As discussed before, some of the observed switching might be due to errors in coding which can potentially introduce measurement errors in the values of other variables and bias the magnitude of the coefficients towards zero. Finally, in the last panel of the table , we run our main specifications on the entire sample without imposing any restrictions. The results remain qualitatively unchanged.

In our main analysis, we use the value of the old audit notch, Rs.6 million, as the lower bound of the treatment neighborhood. A potential concern is that firms just above Rs.6 million in 2011 had an opportunity to bunch as they were close to the threshold. Including these firms in the sample could cause selection bias as they chose to not bunch at the old threshold. To alleviate this concern, we restrict the treatment neighborhood to firms that are potentially too large to bunch at the old threshold in 2011. We now use a treatment neighborhood of Rs.10 million to Rs.15 million. Appendix table A.3 shows that the magnitude of the treatment effect is now larger than what we found earlier. This is expected because now we are only including the right-origin bunchers who, contrary to left-origin bunchers, would have to decrease their revenue to bunch below the threshold and get exempt from the third-party audit.

Finally, we conduct a placebo test by mis-specifying the treatment region that gets affected in response to the policy change in 2011. In theory, the big firms, as defined in the model, should not get affected by the notch because the cost of manipulating their revenue to bunch below the threshold is too high for them. Thus, we consider private firms with revenue between Rs.16-28 million in 2011 as the treatment group. All the public firms in the same revenue bandwidth form the comparison group. Appendix Table A.4 shows that the coefficients are insignificant, except for audit fee which is significant only at the 10% level. Moreover, the magnitude of the change in profits, taxable income and taxes is comparable to what we found in the main analysis, even though the average firm size is twice as large now.

1.5.3 Comparison with the static model

In this section, we analyze how the treatment effects change if we estimate the upper bound of the treatment neighborhood using the static bunching analysis. We construct a counterfactual distribution, representing the state of the world without audit notch, by fitting a polynomial function on the observed distribution after excluding observations in the bunching region. The lower bound of the region is determined visually as the point where the slope of the distribution becomes positive. Then, we guess the upper bound of the region as a point just above the notch and keep increasing it till the excess mass due to bunching is equal to the missing mass. Here excess (and reduced) mass is the difference between observed and counterfactual distribution. (For more details on static bunching analysis, please refer to Kleven and Waseem 2013.)

Coming to our context, we first graph the frequency distribution of the firms using combined data of all the years after the policy change in 2012. Figure 1.5 shows the distribution, where the firms are grouped in revenue bins of Rs.0.5 million. We visually determine the lower bound as Rs.8.5 million, and construct the counterfactual distribution by fitting a fourth-degree polynomial equation:

$$C_b = \sum_{i=0}^{4} \beta_i Y_b^i + \sum_{b=y^{lb}}^{y^{ub}} \delta_b \mathbb{1}(Y_b = b) + \eta_m + \epsilon_b \tag{1.6}$$

where C_b is the actual count of firms in the bin b. A firm is in bin b if its income $y_i \in [b, b + 0.5million)$. The β coefficients represent the polynomial terms, and δ coefficients are the dummies for bins in the omitted region and are interpreted as the difference between the actual and counterfactual density. We also control for potential round-number bunching by including dummy for whether the bin contains multiple of Rs.1 million. These are represented by η_m in the equation. The standard errors are estimated by using a bootstrapping procedure.

According to the static analysis, the upper bound of the treatment neighborhood is Rs.11.5 million [18]. The counterfactual density is shown in Figure 1.6. As discussed before, this might be an under-estimate of the actual upper bound. In Appendix Table A.7, we define the treatment neighborhood as spanning from Rs.6 million to Rs.11.5 million and estimate the treatment effects. All the coefficients are smaller in magnitude than what we found in our main analysis. This is expected because the new treatment neighborhood excludes larger firms that responded to the audit-notch. However, none of the coefficients

[18] We calculate alternative estimates of the upper bound by changing the width of the revenue bins in the appendix table A.6

are significantly different from the estimates we found earlier because the proportion of firms that get treated in the range of Rs.11.5-15 million is much smaller than the proportion of treated firms in the overlapping region across the two alternative treatment neighborhoods.

1.5.4 Heterogeneity in the effect of third-party audit

In this section, we explore if the effectiveness of the third-party audit changes as the audit intensity faced by the firm or the cost of under-reporting income changes. We find that, if the firms face high audit intensity or if they have low resource cost of under-reporting , then they report lower taxes, once they are exempt from the third-party audit.

The audit intensity faced by the firm under third-party audit is correlated with the paper-trails it creates during business transactions. Paper-trails help the auditor to catch discrepancies in the tax form. We use the place of the firm in the supply chain as a proxy for the paper-trails it generates. Under a value added tax(VAT) regime, firms which sell intermediate goods generate more receipts than the retail firms which sell final-consumption goods. The reason is that both the supplier and consumer of an intermediate goods producer demand receipts to claim tax credits under VAT. On the other hand, the final consumer of a retailer doesn't need a receipt as he can't claim any tax credit. [Pomeranz (2015), AL(2018)].

Since we do not have transaction-level data at the firm level, we use industry-level supply-use tables(SUTs)[19] that provide information on the final use of each industry's output. We match the industry description given in the SUTs to the industry codes reported by the firms in the income tax form.(Details of the matching are given in the appendix table A.8.) Each firm is given a score which is equal to the proportion of sales it makes to other industries instead of the final consumer. Higher score implies higher paper-trails and thus, more thorough audit by the private auditor.

We use the differences-in-differences framework used before, but now include a triple interaction term ($treat_i$ x $after_i$ x $UpstreamRatio$) in equation 1.3. Table 1.4 column 1 shows that as the proportion of intermediate sales increases, firms decrease their tax revenue. The bottom panel of the table shows that for firms operating at 90th percentile of upstream ratio, the treatment effect is around 34% higher than what we found earlier. On the other hand, firms that sell most of their products to final consumer do not find it worthwhile to decrease their revenue to bunch below the threshold, as they face low audit-intensity during the audit.

[19]We use the supply-use tables from 2011 which are compiled by the Ministry of Statistics and Program Implementation. These tables can be found using the link: http://mospi.nic.in/publication/supply-use-tables

Another factor that can affect the efficacy of third-party audit is the resource cost of under-reporting incurred by the firm. We know from previous studies that an increase in the number of agents who know about evasion decisions leads to an increase in the cost of under-reporting (Kleven, Kreiner and Saez (2016)). Firms find it difficult to sustain the collusive arrangements across several such agents who can whistleblow against the firm due to reasons like monetary incentives from the government or preferences towards honest behavior. These economic agents can be well-informed workers or incentivized final consumers (Naritomi (2019)).

The tax returns do not contain information on the number of workers. We use the ratio of expenditure on employees to the total expenditure of the firm to instrument for it. However, labor intensity may not be perfectly correlated with the total number of workers. Therefore, we also use the total labor expenditure of a firm as an alternative indicator for number of employees.[20]

We find that firms with more workers do not change their tax payments once they have an opportunity to get exempt from the third-party audit (see columns 2 & 3 of Table 1.4). On the other hand, firms that employ fewer workers decrease their tax payments by 54 percent more than the average effect we found earlier.

1.6 Cost-Benefit Analysis and Optimal Audit Threshold

In this section, we develop a framework to conduct a cost-benefit analysis of the third-party audit policy with a notch, and use the estimates from our empirical analysis to assess the net benefit of the Indian policy.

Building on the baseline model, the total welfare is equal to the after-tax profit of the firms and the net government revenue. The change in audit-intensity, due to third-party audit, changes the aggregate welfare only because of two factors - resource cost of evasion and administrative cost of audit. In the absence of these two factors, welfare is unaffected by the changes in audit intensity. Firms do not change their real output in response to shocks to audit intensity. They only change their reported income[21] which is simply a transfer between the firms and the government and doesn't affect total welfare. Conversely, if the two factors are present, then an increase in audit intensity, on one hand,

[20]One potential issue with using total labor expenditure is that firms with highly productive workers might have high labor cost but low number of workers. Therefore, we use both total and proportional labor expenditure as a proxy for the total number of workers. The results using both the measures are consistent with each other.

[21]This is because production and under-reporting decision are additively separable in our model. The FOCs for the use of inputs in equation 1.1 do not depend on the audit intensity. Keen and Slemrod (2017) and Basri et al. (2019) also use similar models.

increases welfare because firms incur lower resource cost of under-reporting as they report more income. On the other hand, increasing audit intensity causes reduction in welfare because the firms have to pay for the audit[22]. The net impact of the change in audit-intensity on welfare depends on these two countervailing effects.

We now modify the baseline model to develop the above intuition on the lines of AL (2018). Let W represent the aggregate welfare. We now assume that the cost of audit, borne by the firms, is a function of the audit-intensity and represented by $c(\phi)$. Further, $c(\phi)$ is increasing and convex in ϕ. The audit intensity is chosen by the government. The expected value of the welfare is given by the following equation:

$$E[W] = \int_{\bar{y}(\underline{\psi})}^{\bar{y}(\bar{\psi})} \{(1-\tau)[\psi f(x,z) - u - wx] - c(\phi) - qz - k(u) + u - \phi h(u)[\tau u + \theta\tau u]\} .g_0(\bar{y})\, d\bar{y}$$
$$+ \int_{\bar{y}(\underline{\psi})}^{\bar{y}(\bar{\psi})} \{\tau[\psi f(x,z) - u - wx] + \phi h(u)[\tau u + \theta\tau u]\} .g_0(\bar{y})\, d\bar{y}$$

The first term in the curly bracket represents the expected after-tax profit of the firm, while the second term represents the expected revenue of the government. Since the firms have already chosen the inputs and under-reported income optimally, we can use the envelope theorem while taking the derivative with respect to ϕ for the first term. The change in expected welfare due to the change in audit-intensity is given by :

$$\frac{dE[W]}{d\phi} = \int_{\bar{y}(\underline{\psi})}^{\bar{y}(\bar{\psi})} \left[-\tau\frac{du}{d\phi} + \phi\tau(1+\theta)\left(h(u)\frac{du}{d\phi} + u\frac{\partial h}{\partial u}\frac{du}{d\phi}\right) - c_\phi(\phi)\right] .g_0(\bar{y})\, d\bar{y}x$$

$$= \int_{\bar{y}(\underline{\psi})}^{\bar{y}(\bar{\psi})} \left[-k_u(u)\frac{du}{d\phi} - c_\phi(\phi)\right] .g_0(\bar{y})\, d\bar{y} \tag{1.7}$$

The second equality follows by substituting the first-order condition obtained by taking the derivative of the net profit of the firm with respect to the under-reported income in the baseline model (see equation 1.1). The net welfare change is a function of total change in resource cost of under-reporting and administrative cost.

To calculate the welfare change associated with the notch at Rs.10 million, we ask what would be the welfare gain/loss if one more firm comes under third-party audit as a result of lowering the notch by an epsilon rupee amount. By conducting the marginal analysis, we don't have to make assumptions across the entire distribution of firms using parameters that are estimated locally. The average taxable income in the dominated region[23] above

[22]Chetty (2009) develops analogous analysis in the context of personal taxation

[23]We estimated that firms under third-party audit pay Rs.2197 to the auditors. Data shows that firms

the notch is Rs.496,626. According to the difference-in-differences estimates, the reduction in taxes and reported taxable income due to the notch is Rs.41,285 and Rs.102,575 respectively. This implies that the firms that were unable to bunch had to pay at least Rs.41,285 extra, which is equal to 8.3 percent of the average reported taxable income. Thus, $k_u(u)$ is 0.083. If we assume that this is locally constant for the firms, then the total resource cost of evasion is 0.083 times the reduction in taxable income, which is equal to Rs.8,514. This is the marginal welfare gain of auditing one more firm. The marginal cost is the extra fee which firms pay to the auditors when they undergo tax-audit. In the empirical section, we estimated this cost to be Rs. 2,574. Thus, the total welfare gain of extending the third-party audit to one more firm is Rs.5,940 (US$107).

Finally, we use back-of-the-envelope calculations to obtain optimal audit threshold. If we assume that the marginal resource cost of evasion is same for all the firms in the entire distribution, then $k_u(u) = 0.083$. Let the average change in the reported tax base as a proportion of turnover ($\Delta u/\bar{y}$) be equal to $102,575/10,000,000 = 0.0102$. The increase in the administrative cost is Rs.2,574. Thus, the optimal threshold ($\bar{y}^*$) can be derived from the following formula: $[\Delta u/\bar{y}] \times \bar{y}^* \times k_u(u) = c_\phi(\phi)$. This gives us the optimal threshold to be Rs.3,040,396. In 2009, this would imply adding around 51,000 firms in the third-party audit regime[24]. Thus, under the above assumptions, it would be welfare enhancing to have more firms under the third-party audit by reducing the audit-threshold.[25]

1.7 Conclusion

In this study, we evaluate a notched policy implemented by the Indian tax department where, conditional on reported revenues being greater than a specified threshold, firms are required to undergo a third-party audit before filing their tax returns. Since the auditors are chosen and paid by the firms, they face a potential conflict of interest. Despite this, we find that the policy is effective in increasing government revenue as firms report higher taxable income. Firms, also try to escape the audit by bunching below the audit notch. The effect of the policy is heterogeneous. First, firms that generate substantial paper

having turnover between Rs.10 million and Rs.10.1 million have profit equal to 6 percent of the turnover. This implies that the no firm should be located in the region between Rs.10 million and Rs.10.036 million, as the auditor fee is higher than the profits the firms will get. This is the dominated region.

[24] We do this calculation for 2009 because the threshold at that time was Rs.4 million and the density from Rs. 7.4 to Rs.10 million is less likely to be affected by that threshold. The calculated number of firms is still an upper bound as some of the firms will bunch below the new threshold.

[25] If we use the estimates calculated from the treatment neighborhood found from static-bunching analysis, then the reduction in taxes and taxable income are Rs.40,462 and Rs.81.007 respectively. As a result, $k_u = 0.0815$ and $\Delta u/\bar{y} = 0.008$. The optimal threshold is Rs.2,516,871

trails report more taxes when their accounts are scrutinized by private auditors. Second, employing greater number of workers makes evasion difficult and such firms do not change their tax payments when subjected to extra audit. We also conduct a cost-benefit analysis and show that the net benefit of extending the policy to one more firm is substantial.

In future work, the findings of this paper can be extended in at least two directions. The change in tax income because of third-party audit can be due to reduction in economic activity by the firm or misreporting of profits or both. While there is a nascent literature that attempts to disentangle the changes in reported income into real and evasion responses (Almunia and Lopez-Rodriguez (2018), Velayudhan (2018)), it remains an area of active research. Second, the effects of third-party audits can vary according to characteristics of auditors such as reputation and quality. For this, data matching auditors characteristics to tax returns will be required.

It is becoming increasingly common to outsource the government's regulatory functions in both developed and developing countries, especially in sectors such as emissions control and food safety. With the enforcement budget of tax agencies like the IRS in USA on the decline[26], there is a case for privatizing more regulatory functions in the realm of tax administration. However, there are concerns that private auditor will not deliver desired results because of factors like conflict of interest, corruption etc. In this study we show that third-party audits were effective in increasing tax revenue in the context of India. Given that developing countries often suffer from low tax compliance and limited state capacity, our paper suggests that outsourcing this regulatory function in particular could be hugely beneficial.

1.8 Tables

[26]The percentage of Corporate Income tax returns examined by the IRS has fallen from 1.4 percent in 2013 to 0.9 percent in 2018 (IRS (2013-18))

Table 1.1: Summary Statistics

	Treatment Group							
	Before 2012				After 2012			
	Count	Mean	Median	Std.Deviation	Count	Mean	Median	Std.Deviation
Revenue	63,506	9,892,906	9,184,198	5,447,840	105,870	15,323,545	11,687,982	14,084,555
Total Cost	63,504	11,668,229	9,789,557	12,063,096	105,870	17,456,300	12,827,539	16,613,845
PBITD	63,506	1,173,221	679,802	1,854,479	105,870	1,741,391	845,072	3,142,611
Taxable Income	63,506	477,149	138,706	818,614	105,870	733,533	178,195	1,374,110
Tax Paid	63,506	159,867	54,191	259,771	105,870	242,605	69,127	436,786
Audit Fee	63,506	22,618	18,000	19,435	105,870	30,773	25,000	27,846
Wage Ratio	63,481	0.21	0.13	0.21	105,841	0.22	0.14	0.25
Wage	63,506	1,800,028	1,189,462	1,776,788	105,870	2,895,673	1,683,453	3,394,915
Upstream Ratio	51	0.36	0.33	0.34	51	0.36	0.33	0.34

	Comparison Group							
	Before 2012				After 2012			
	Count	Mean	Median	Std.Deviation	Count	Mean	Median	Std.Deviation
Revenue	1,896	11,064,946	9,633,706	6,985,863	3,160	16,884,050	11,707,338	17,099,600
Total Cost	1,896	14,952,564	10,890,315	19,367,045	3,160	20,180,162	12,890,495	21,565,627
PBITD	1,896	2,261,207	1,215,544	3,273,161	3,160	3,536,390	1,345,231	5,665,575
Taxable Income	1,896	789,058	114,441	1,207,152	3,160	1,149,721	98,330	1,969,816
Tax Paid	1,896	267,041	66,664	380,352	3,160	391,436	72,187	628,930
Audit Fee	1,896	31,136	21,852	26,201	3,160	41,914	29,207	37,260
Wage Ratio	1,891	0.22	0.13	0.21	3,155	0.24	0.15	0.23
Wage	1,896	1,965,551	1,190,394	2,019,811	3,160	2,881,183	1,703,390	3,405,506
Upstream Ratio	36	0.41	0.43	0.34	36	0.41	0.43	0.34

Source: All the data for this table is derived from Corporate Income Tax returns from 2009-16. The upstream ratio is calculated using data from the Supply and Use Tables of 2011-12 which are complied by Ministry of Statistics and Program Implementation and can be found using using the link http://mospi.nic.in/publication/supply-use-tables

Table 1.2: Effect of removal of third-party audit on tax payments

Variable	(1) Tax Paid	(2) Tax Paid	(3) Tax Paid	(4) Tax Paid
Treat x Post2012	-41,691***		-41,285***	
	(15,139)		(15,114)	
Treat x FY2009		-11,086		-11,493
		(12,436)		(12,458)
Treat x FY2010		-8,891		-9,805
		(11,108)		(11,104)
Treat x FY2012		-25,909**		-25,973**
		(11,642)		(11,624)
Treat x FY2013		-33,659**		-33,764**
		(14,618)		(14,579)
Treat x FY2014		-61,597***		-61,866***
		(19,841)		(19,852)
Treat x FY2015		-48,447**		-48,189**
		(21,863)		(21,846)
Treat x FY2016		-72,136***		-72,204***
		(25,978)		(25,980)
Sectoral Time trends	No	No	Yes	Yes
Observations	174,432	174,432	174,432	174,432
R-squared	0.602	0.602	0.603	0.603

Robust standard errors (clustered at firm level) in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Note–This tables shows the estimates from equations 1.3 and 1.4. The treatment group consists of restricted share transfer (RST) firms with revenue between Rs.6-15 million in 2011, a year before the audit threshold was moved from Rs.6 million to Rs.10 million. The comparison group consists of non-RST firms within the same revenue bandwidth in 2011. The treatment is removal of third-party audit requirement because of the change in the threshold. All the regressions in the table include firm fixed effects and year fixed effects. The data comes from Corporate Income Tax returns from 2009-16.

Table 1.3: Effect of removal of third-party audit on firm's behavior

VARIABLES	(1) PBITD	(2) PBITD	(3) Taxable Income	(4) Taxable Income	(5) Audit Fee	(6) Audit Fee
Treat x Post2012	-703,408***		-102,575**		-2,574***	
	(147,840)		(47,659)		(752.4)	
Treat x FY2009		-100,890		-15,892		-153.5
		(124,377)		(39,985)		(751.8)
Treat x FY2010		-24,412		-18,324		89.71
		(108,415)		(35,615)		(718.9)
Treat x FY2012		-248,640*		-67,561*		-451.1
		(127,024)		(38,190)		(641.6)
Treat x FY2013		-534,433***		-68,220		-1,326
		(137,117)		(45,563)		(827.8)
Treat x FY2014		-964,452***		-121,879**		-2,107**
		(187,319)		(58,000)		(947.3)
Treat x FY2015		-866,987***		-106,650		-4,412***
		(214,124)		(66,030)		(1,100)
Treat x FY2016		-1.113e+06***		-205,776**		-4,687***
		(258,201)		(82,736)		(1,254)
Observations	174,432	174,432	174,432	174,432	174,432	174,432
R-squared	0.581	0.581	0.603	0.603	0.726	0.726

Robust standard errors (clustered at firm level) in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Note–This tables shows the estimates from equations 1.3 and 1.4, where the dependent variable is mentioned in the column. PBITD refers to Profit before Interest, Taxes and Depreciation. The treatment group consists of restricted share transfer (RST) firms with revenue between Rs.6-15 million in 2011, a year before the audit threshold was moved from Rs.6 million to Rs.10 million. The comparison group consists of non-RST firms within the same revenue bandwidth in 2011. The treatment is removal of third-party audit requirement because of the change in the threshold. All the regressions include firm fixed effects, year fixed effects and sector-specific time trends. The data comes from Corporate Income Tax returns from 2009-16.

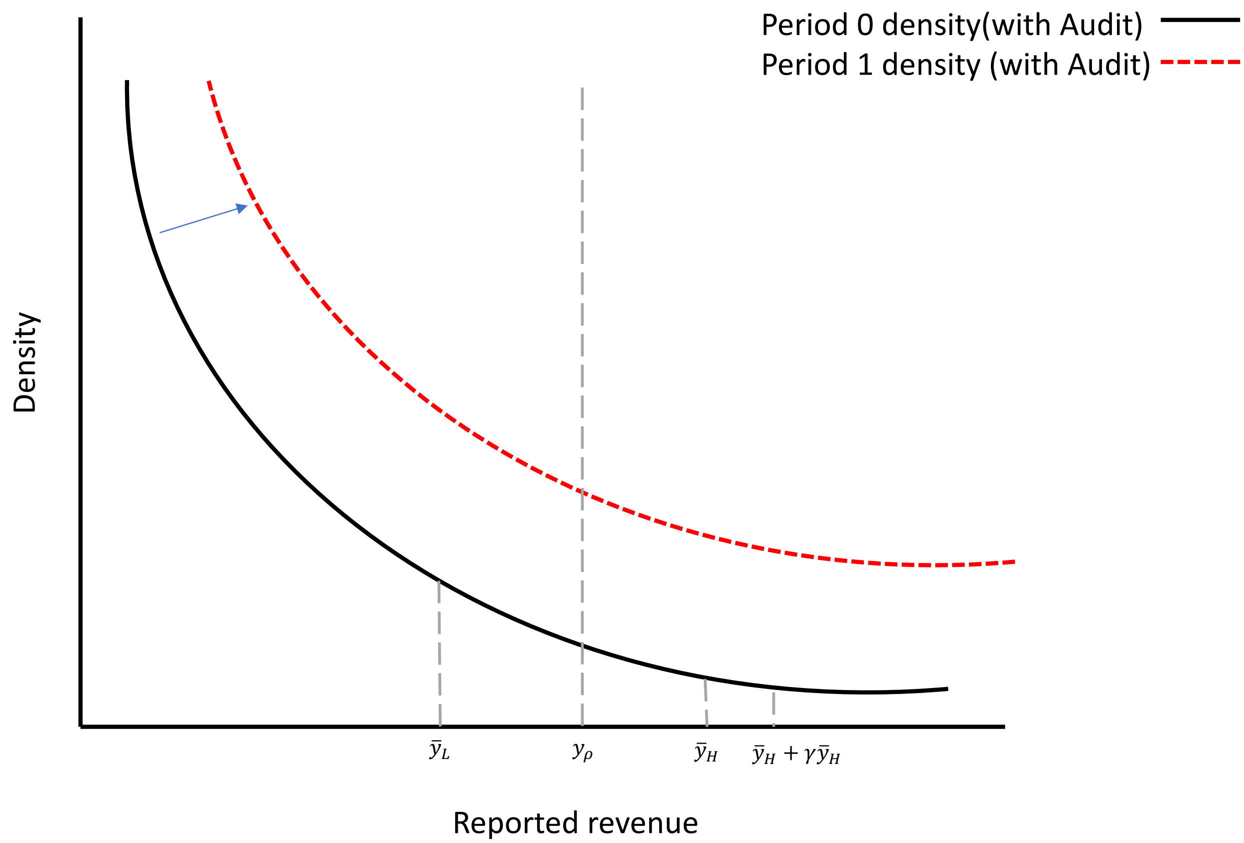
Table 1.4: Heterogeneity in the effect of third-party audit

VARIABLES	(1) Tax Paid	(2) Tax Paid	(3) Tax Paid
Treat x Post2012	-4,210	-72,574***	-74,956***
	(21,853)	(15,242)	(15,237)
Treat x Post2012 x UpstreamRatio	-60,179***		
	(8,392)		
Treat x Post2012 x WageRatio		144,423***	
		(11,766)	
Treat x Post2012 x Wage			0.0161***
			(0.00134)
Treat x Post2012 + Treat x Post2012 x p(25)	-5,968	-63,991***	-63,526***
	(21,808)	(15,177)	(15,137)
Treat x Post2012 + Treat x Post2012 x p(90)	-55,359 **	8,818	5,613
	(21,655)	(15,901)	(15,785)
Observations	77,855	174,392	174,432
R-squared	0.614	0.604	0.605

Robust standard errors (clustered at firm level) in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Note- This table shows the estimates from an augmented version of equation 1.3, which includes triple interaction term mentioned in the table. UpstreamRatio is defined as the proportion of sales that are used as intermediate inputs by other industries. This is calculated at the industry-level. Wage ratio is the proportion of total expenses of a firm spent on the employees one year before the policy change in 2011. Wage represents the total wage payments by the firm in 2011. Both the variables proxy for the total number of employees in a firm. The bottom panel calculates the treatment effect at the 25th and 90th percentile of the extra variable used in the triple interaction term. All the regressions include firm fixed effects, year fixed effects and sector-specific time trends. All the tax and expense data for this table is derived from Corporate Income Tax returns from 2009-16. To calculate the upstream ratio, we use the Supply and Use Tables of 2011-12, complied by Ministry of Statistics and Program Implementation.

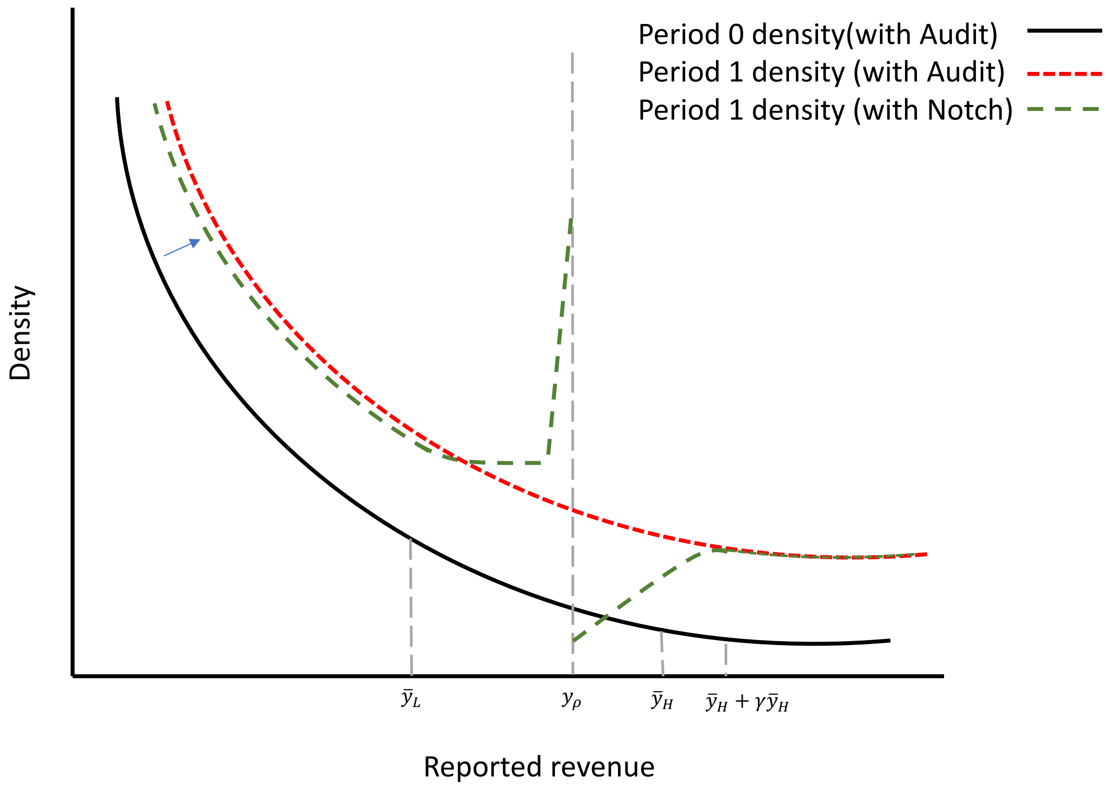
1.9 Figures

Figure 1.1: No Audit-Notch



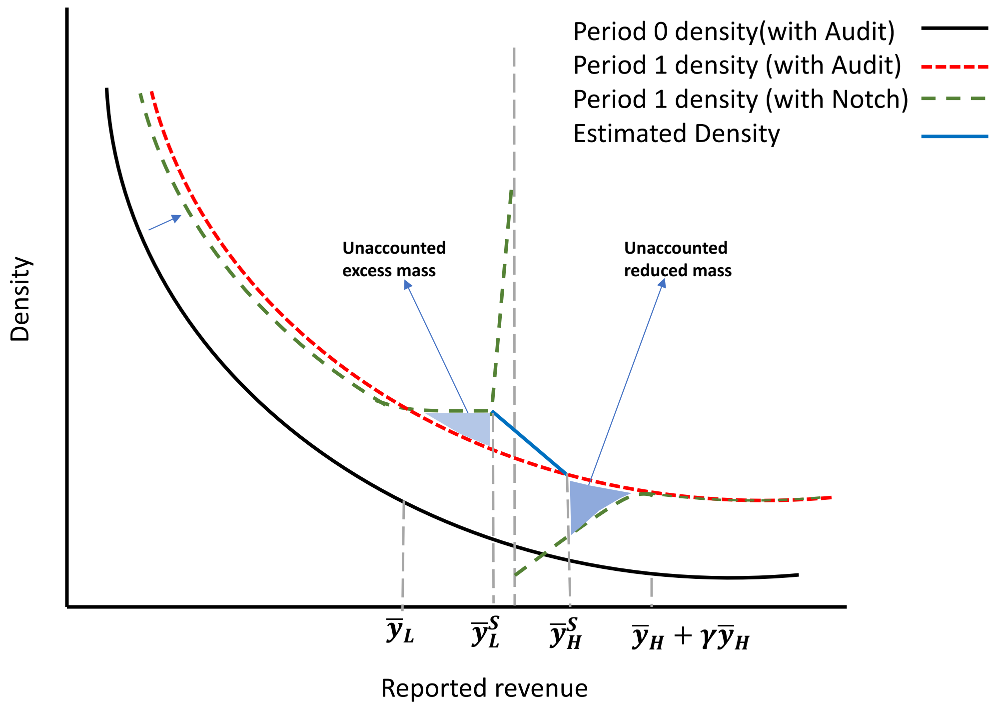
Note–This figure shows the distribution of firms without audit notch. The horizontal axis represents the running variable–reported revenue of the firms. The black and red lines represent the density in the time periods t_0 and t_1, respectively. At t_1 firms grow by a factor of γ. We assume that every firm undergoes third-party audit in both the time periods and therefore, report their true income.

Figure 1.2: With Audit Notch



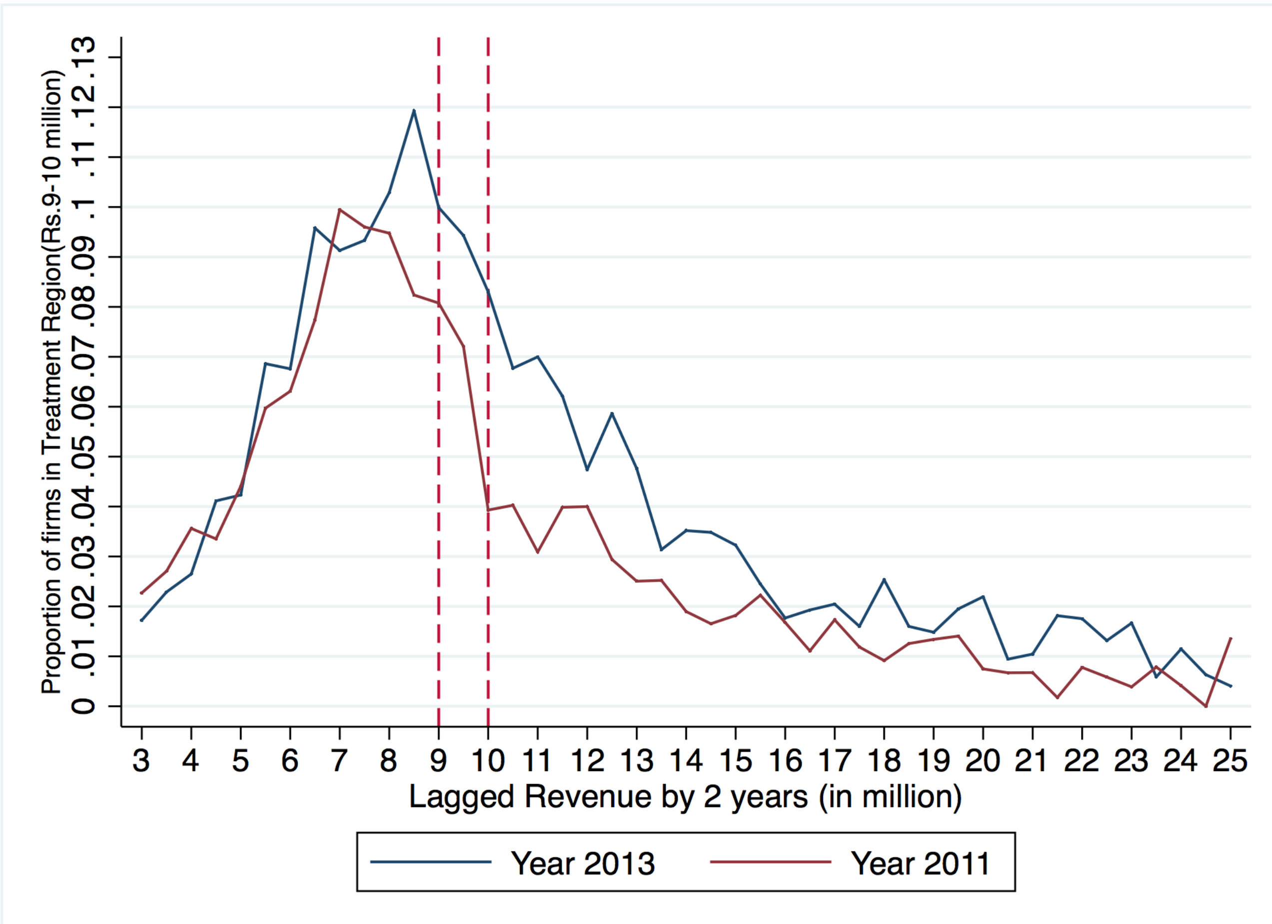
Note–In this figure, we show which firms react to the audit notch. An audit notch is introduced at t_1 and all the firms below the notch are exempt from third-party audit. The green dashed line, labeled as "with notch", represents the density when the audit-notch is defined at y_ρ. Let $\bar{y}_L$ be the income at t_0 where an income growth of γ results in income of y_ρ in t_1. Therefore, $\bar{y}_L$ is equal to $y_\rho/(1+\gamma)$. The left-origin bunchers will have $\bar{y_0} \in [\bar{y}_L, y_\rho)$. Consider a firm with income $\bar{y}_H$ in t_0 which is indifferent between reporting true income or bunching below the notch in t_1. Then, the right-origin bunchers will have $\bar{y_0} \in [y_\rho, \bar{y}_H)$. Note that some of the bunchers will strategically mis-report their revenue to be well below the notch, rather than bunch just below it, if they believe that reporting no growth in revenue for multiple time-periods will increase their probability of getting caught. This will result in diffused excess mass well below the notch, which is shown as a plateau in the density.

Figure 1.3: Bias in the Static Bunching Analysis



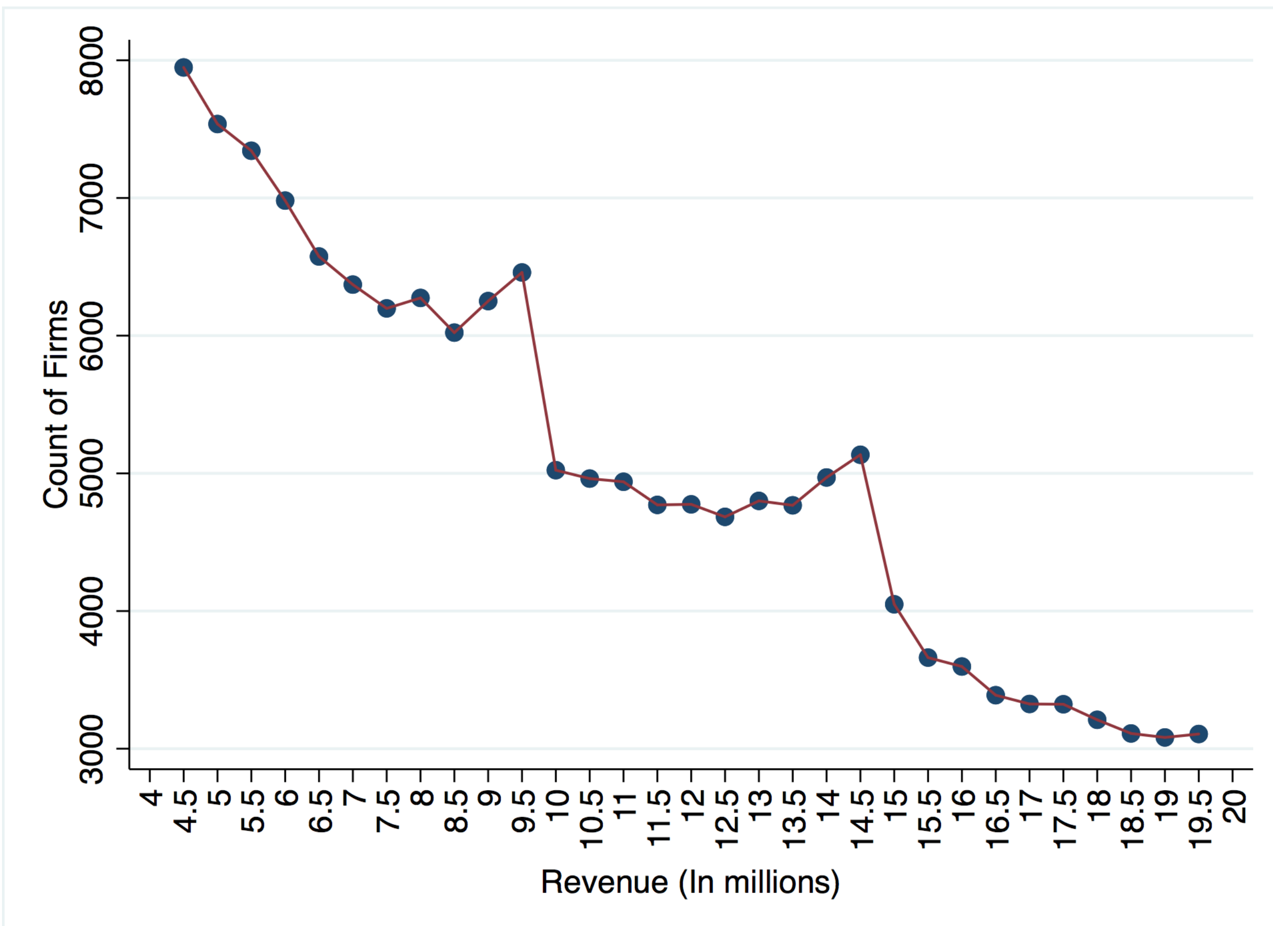
Note– The lower bound of the bunching region, according to the static analysis, is the point where the observed density slopes upward just below the audit-notch, that is, $\bar{y}_L^S$. Using this lower bound, if we fit a counterfactual density (solid blue line in the diagram) by equating excess mass to the missing mass, then the upper bound of the bunching region will be biased downward. This is because some of the excess mass is diffused and not accounted for while fitting the counterfactual density. In this figure, the estimated upper bound is $\bar{y}_H^S$, while the actual upper bound is $\bar{y}_H + \gamma\bar{y}_H$.

Figure 1.4: Estimating upper bound using the difference in probability method for RST firms



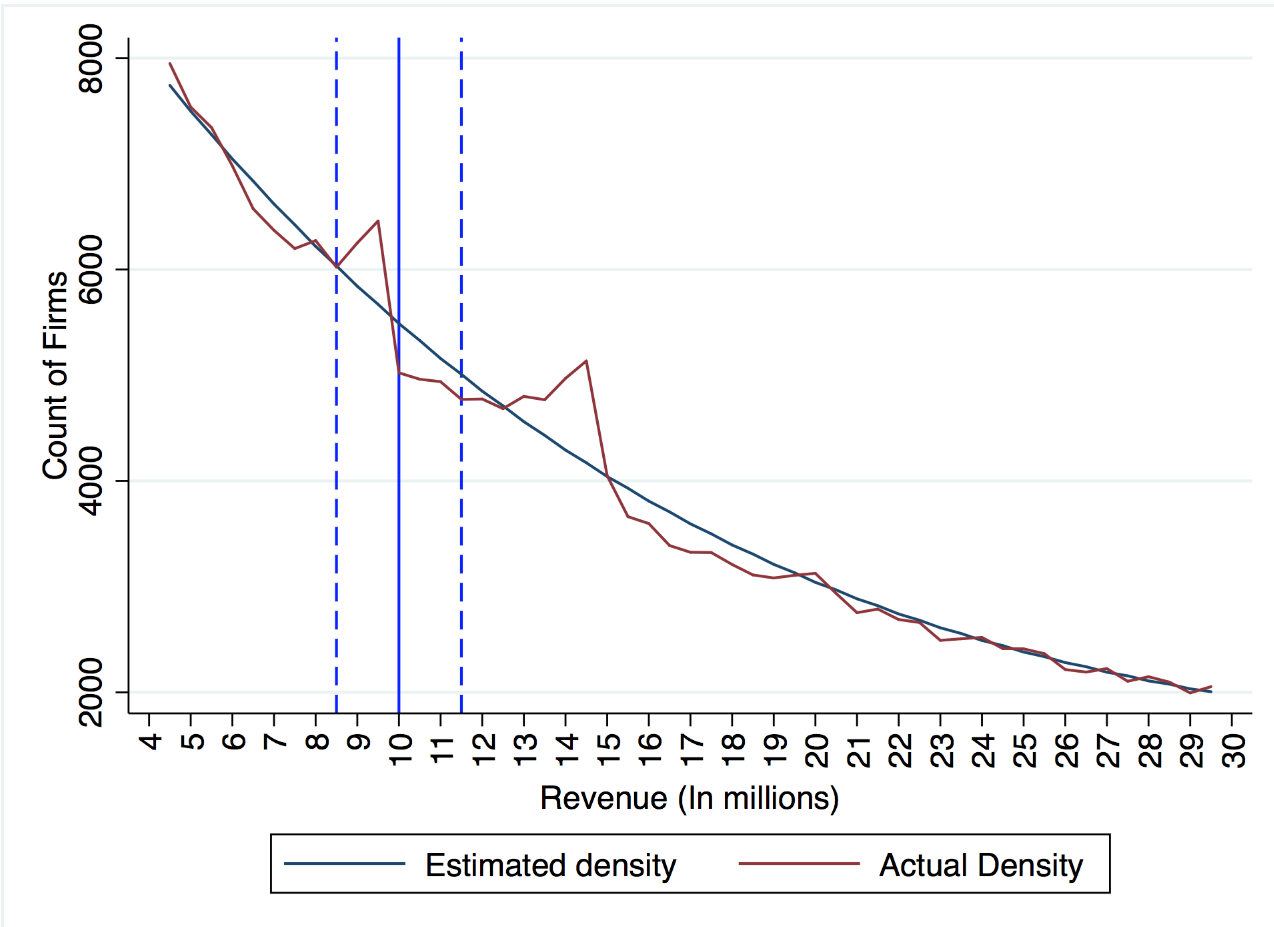
Note–In this graph, we plot the probability of being in the bunching region based on two-year lagged revenue. The notch was introduced at Rs.10 million in 2012. The blue line represents the probability of reporting revenue between Rs.9-10 million (bunching region) in 2013, conditional on revenue in 2011. Similarly, the red line represents the probability of reporting revenue in the same range in 2011, conditional on revenue in 2009 – both the years are before the change in policy. The difference between the two probabilities shows the effect of the notch on firm's bunching response. The upper bound estimated from this graph is Rs.15 million, which is larger than the upper bound estimated by static bunching analysis. The bin size used in this graph is Rs.0.5 million. All the data for this graph is derived from Corporate Income Tax returns from 2009-13.

Figure 1.5: Frequency distribution of RST firms from 2012-16



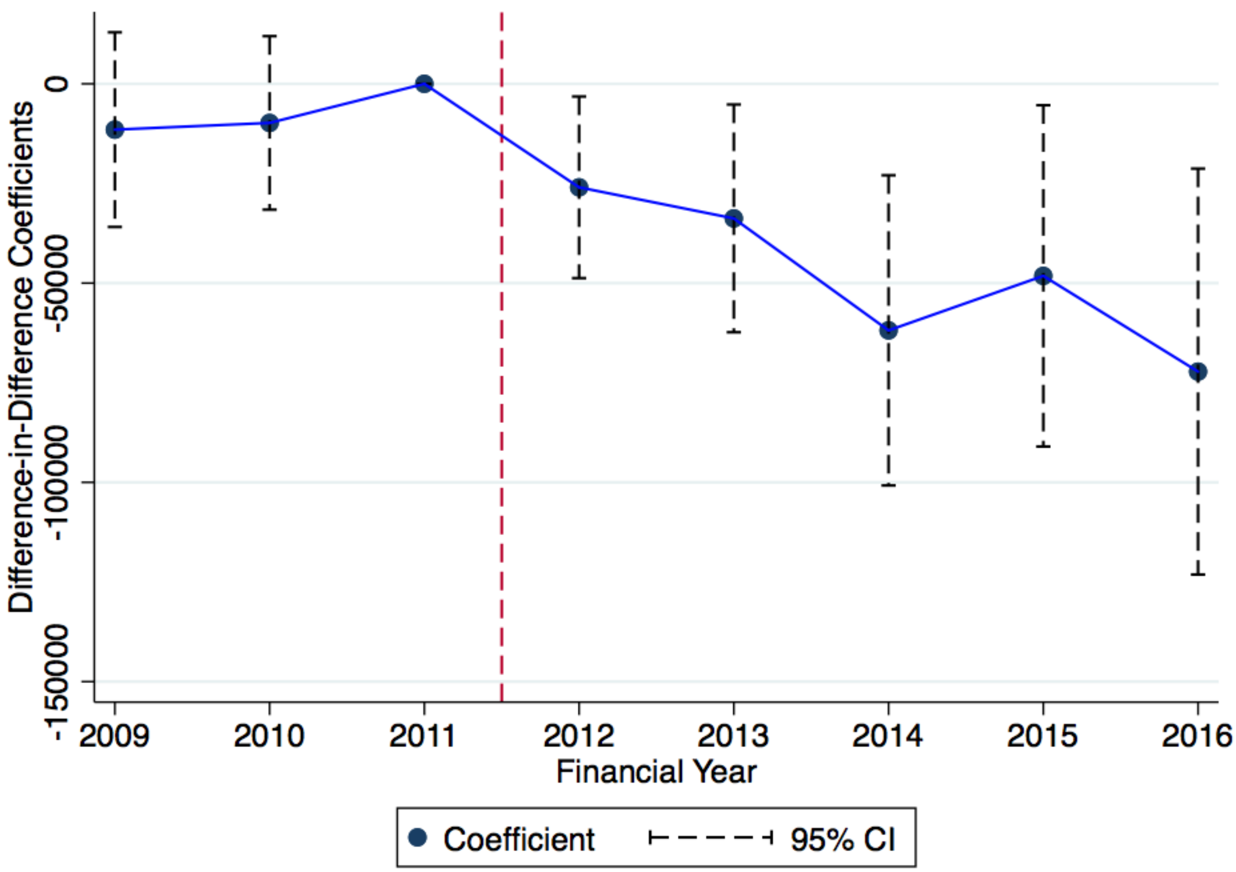
Note–We use the ocular method to estimate the lower bound of the bunching region for the static analysis. This is Rs. 8.5 million, where the slope of the distribution becomes positive. All the data for this graph is derived from Corporate Income Tax returns from 2012-16.

Figure 1.6: Estimating upper bound using the static bunching analysis



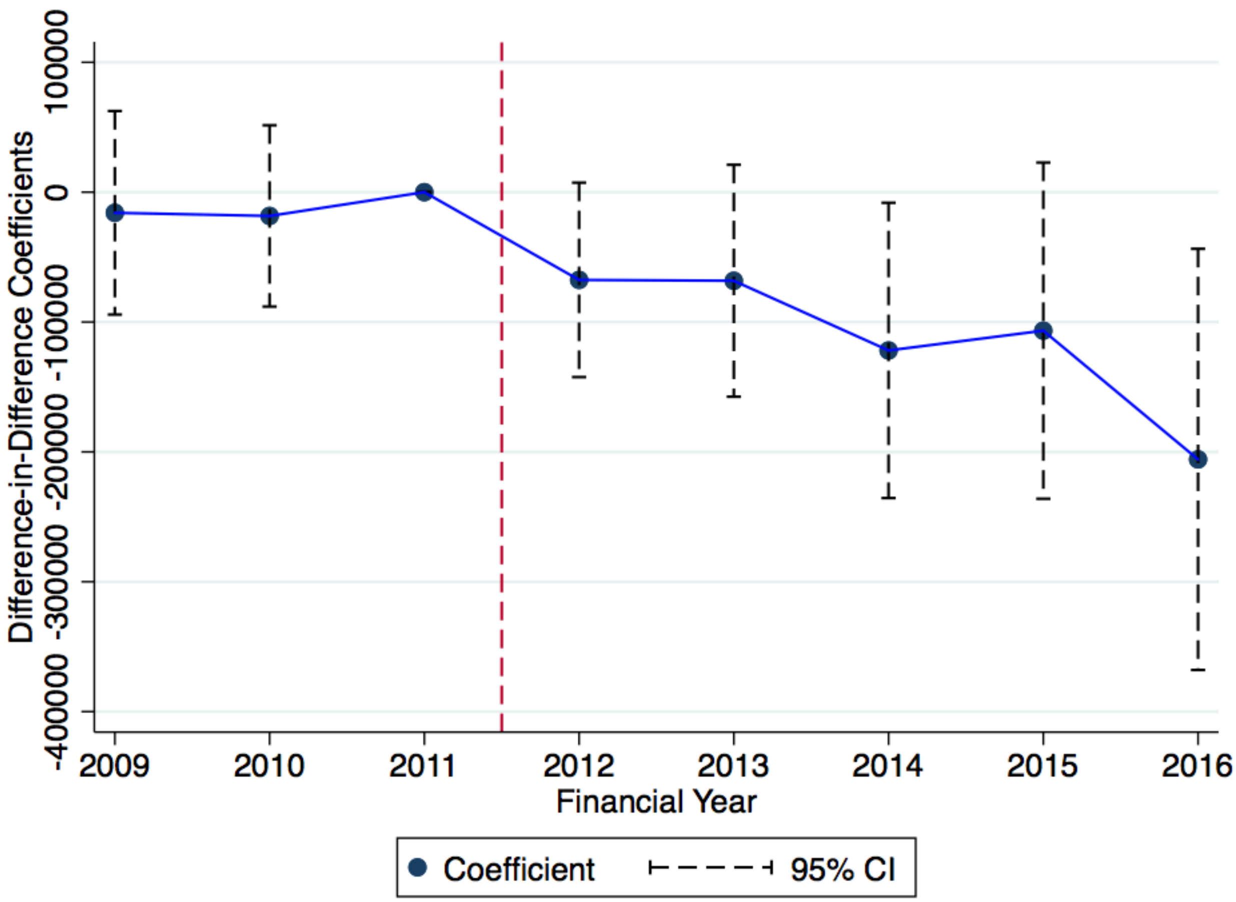
Note–This figure shows the actual and the counterfactual distribution estimated by using the static bunching analysis. The counterfactual distribution is estimated by using a fourth-degree polynomial. The lower bound is determined visually as Rs.8.5 million, while the upper bound is estimated as Rs.11.5 million by equating the excess mass to the missing mass. In the figure, the upper and lower bounds are represented by dashed lines, while the notch is represented by the solid line. The bin size is Rs 0.5 million. All the data for this graph is derived from Corporate Income Tax returns from 2012-16.

Figure 1.7: Effect of removal of third-party audit on tax payments



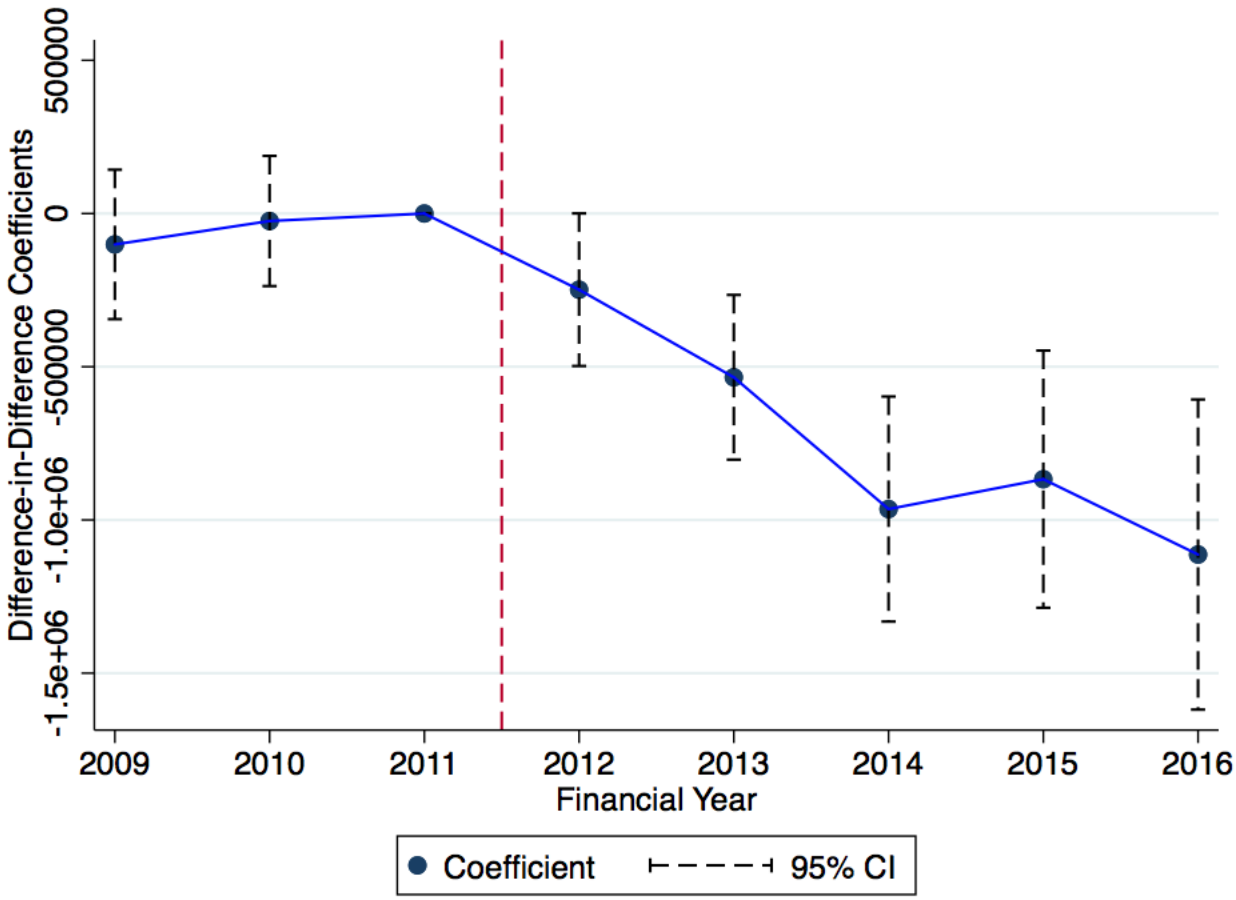
Note–In this graph, we plot the difference-in-differences coefficients and the 95% confidence intervals using the equation 1.4. The notch was introduced at Rs.10 million in 2012. The treatment group consists of RST firms that reported revenue between Rs.6-15 million in 2011, while the comparison group consists of non-RST firms in the same revenue bandwidth. The dependent variable is tax payments in INR. All the data for this graph is derived from Corporate Income Tax returns from 2009-16.

Figure 1.8: Effect of removal of third-party audit on taxable income



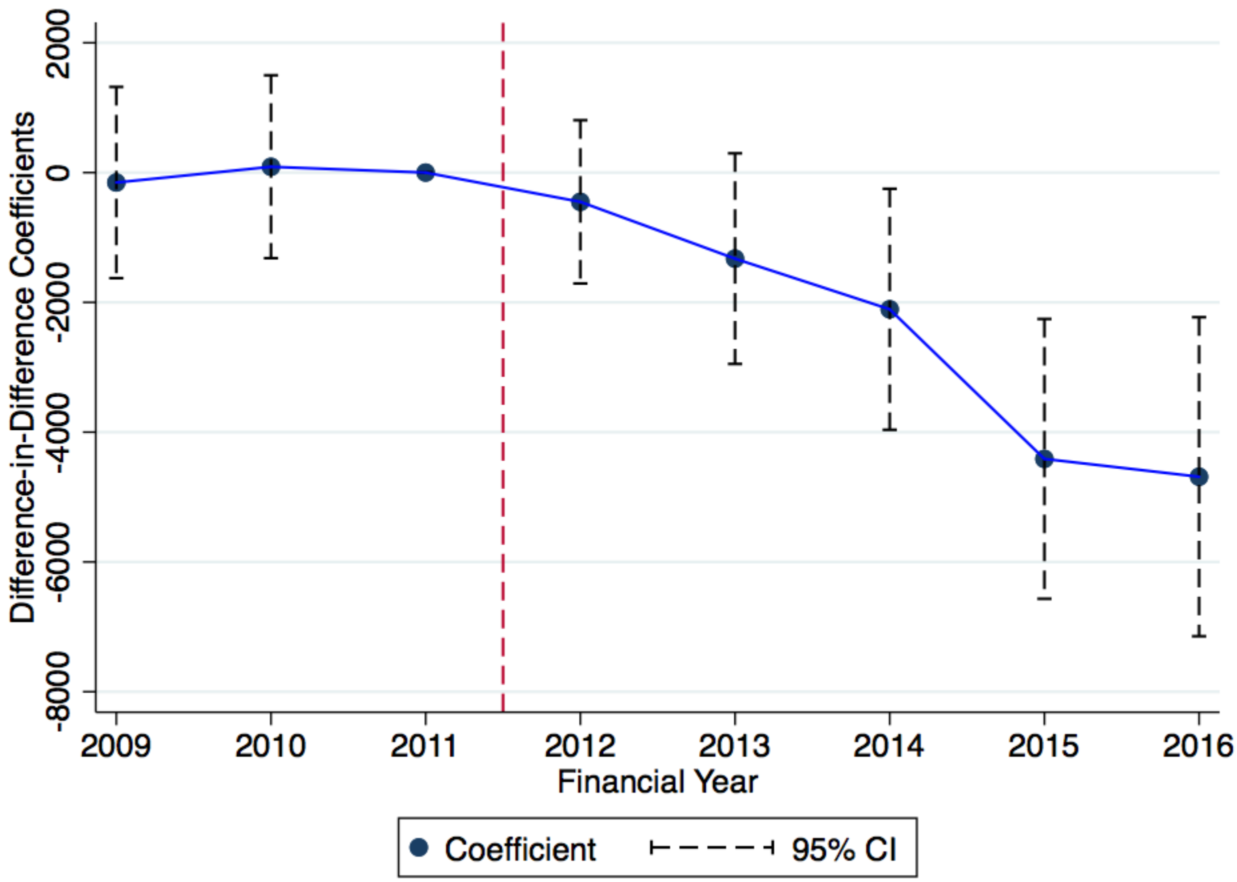
Note–In this graph, we plot the difference-in-differences coefficients and the 95% confidence intervals using the equation 1.4. The notch was introduced at Rs.10 million in 2012. The treatment group consists of RST firms that reported revenue between Rs.6-15 million in 2011, while the comparison group consists of non-RST firms in the same revenue bandwidth. The dependent variable is taxable income in INR. All the data for this graph is derived from Corporate Income Tax returns from 2009-16.

Figure 1.9: Effect of removal of third-party audit on PBITD



Note–In this graph, we plot the difference-in-differences coefficients and the 95% confidence intervals using the equation 1.4. The notch was introduced at Rs.10 million in 2012. The treatment group consists of RST firms that reported revenue between Rs.6-15 million in 2011, while the comparison group consists of non-RST firms in the same revenue bandwidth. The dependent variable is PBITD (profit before interest, tax and depreciation) measured in INR. All the data for this graph is derived from Corporate Income Tax returns from 2009-16.

Figure 1.10: Effect of removal of third-party audit on auditor's fee



Note–In this graph, we plot the difference-in-differences coefficients and the 95% confidence intervals using the equation 1.4. The notch was introduced at Rs.10 million in 2012. The treatment group consists of RST firms that reported revenue between Rs.6-15 million in 2011, while the comparison group consists of non-RST firms in the same revenue bandwidth. The dependent variable is audit fee measured in INR. All the data for this graph is derived from Corporate Income Tax returns from 2009-16.

CHAPTER II

Mind the Gap: Fiscal Externality, Informality and Schooling in Nepal

From a work with Hoyt Bleakley

Abstract

While increasing average years of school has been a development priority for decades, the associated fiscal costs and benefits have been less studied, in part because of the lack of appropriate data. Recently UNESCO organized a project measuring the extent of subsidies, by level of schooling, from all levels of government, in eight developing countries. One of these countries was Nepal, which also has a household budget survey that permits us to estimate the degree of formality, tax payment, and benefit receipt as a function of years of schooling. Using a simple Mincer-like model, we estimate the fiscal externality associated with an additional year of school. Using a discount rate of 3%, we find that within primary school, fiscal benefits and costs, on the margin, are quite balanced, with subsidies closest to the present value of future taxes minus benefits. At higher levels of schooling, however, marginal fiscal benefits exceed costs by 5 percent of per-capita consumption.

JEL Codes: I2, J2, H3
Keywords: Taxation, Subsidies, Schooling Decision, Nepal

2.1 Introduction

Raising levels of education has been a development priority for over a half-century. Around the world, governments subsidize education in part to encourage school attendance. Numerous studies (c.f., Psacharopoulos and Patrinos (2018)) show that children who get more years of schooling earn more as adults. Therefore the government's education spending might be thought of as an investment in expanding the future tax base. This combination describes a fiscal externality: 'fiscal' because the government's taxes and spending are involved, and 'externality' because the student (plus his family) do not pay the full cost of additional schooling nor receive the full benefit. Is there a gap between these benefits and costs, and, if so, how large is it?

The challenge in quantifying this gap has been the incomplete measurement of the extent of education subsidies in less developed countries. Recently, however, UNESCO organized a pilot project for "national education accounts" (NEA) in eight countries. The NEA reports measure public educational spending from all levels of government and decompose by all levels of schooling. Two of the eight countries (Nepal and Uganda) also had contemporaneous household budget surveys, which permit comparison of public and private spending.

We complement the subsidy data with estimated tax payments, made from a Nepalese budget survey. In these data, more educated workers have higher purchasing power, consistent with the literature. To this, we add a less-documented result: workers with more education also have substantially higher probabilities of being in the formal sector. Of workers with zero to two years of schooling, the percentage in the formal sector is in the low single digits. The fraction of formality rises to approximately half of workers with post-secondary education. Thus, more educated workers contribute more to the government's coffers because they spend more (higher VAT) and because their earnings are higher and more likely to be in the formal sector (higher income taxes). (Section 2.2 describes these data sets and summary statistics.)

We develop an empirical model in the spirit of Mincer (1958) to compare fiscal costs and benefits of an additional year of school. In the model, the student receives a marginal benefit in the form of higher future earnings, but pays contemporary marginal costs in the form of direct costs (tuition, books, transport, etc.) and of opportunity costs (foregone earnings). The government's problem is qualitatively similar, with higher future taxes (from higher future wages) being the marginal benefit and subsidies plus foregone taxes on foregone earnings being the marginal costs. With both linear and quantile regressions, we estimate the relationship between tax payments and years of schooling. These estimates combine with the subsidy data and an interest-rate to form the estimated gap, or fiscal externality. (We describe the model in Section 2.3 and present results in Section 2.4.)

We find that substantial fiscal costs and benefits associated with another year of schooling, although how the two balance varies by level. For finishing primary school, subsidies are close to the present value of future taxes, with a deficit of less than a few US$ per year, itself about 0.5% of household consumption per capita (among those with primary school only). For secondary school, the gap leans negative as well, especially at lower quantiles of distribution, although the mean effect is an even smaller fraction of household consumption than it is for primary school. For tertiary (university/higher) education, the opposite is seen, with positive gaps (benefits in excess of costs) for the mean and especially for the upper part of the distribution. This gap amount to five percent of household consumption, a significant surplus. The difference by level makes intuitive sense. The lower levels of education are highly subsidized and most of the workers with such education have low labor-market formality; therefore, the government recoups its investment mainly through the 13% VAT. In contrast, higher education, while more costly, is less proportionately subsidized, and its graduate earn more in formal employment and thus subject to the income tax. (This conclusion therefore relates to work by Johnson (2006) who argues that higher education subsidies are progressive if we consider the higher taxes paid by higher-income households whose children are disproportionately enrolled in higher education and that those children end up paying higher taxes as workers.)

We subject this result to sensitivity analysis in Section 2.5. We show that this result is not sensitive to alternative definitions of school levels, household, or formality. The results are also robust to alternate strategies for controlling for household characteristics, defining taxes, or accounting for emigration. In the main analysis, we use three percent as the default interest rate. We show that the interest rate would have to be drastically lower to convert all of the estimated gaps to positive. We then offer conclusions in Section 2.6.

2.2 Data and Descriptive Statistics

To quantify the fiscal balance at different levels of education, we require information on the taxes paid by household, their educational attainment and the non-household expenditure on education. The information on the latter is hard to come by, especially for developing countries, as there are various stakeholders involved in financing education and aggregating their expenses has been a challenge. Moreover, for our analysis we need information on the non-household expenditure according to different education levels. This issue has been addressed by the creation of National Education Accounts for some developing countries. We use this data along with a nationally representative household survey data and the Nepalese tax schedule to conduct our empirical analysis.

2.2.1 Education Subsidies

The National Education Accounts (NEA) of Nepal has information on expenditure by all the financial stakeholders –household and non-household– at different levels of education[1]. This information has been complied by UNESCO Institute for Statistics (UIS), International Institute for Educational Planning (IIEP) and Global Partnership for Education (GPE), not just for Nepal but for seven other developing countries (IIEP, UIS and IIEP Pôle de Dakar (2016*a*); IIEP, UIS and IIEP Pôle de Dakar (2016*b*)).[2]

Nepal's NEA data capture education expenditure per student across 7 levels of education. These are pre-primary, primary (1-5 grade), lower secondary (5-8), secondary (8-10), higher secondary(10-12), technical education, and higher education (college education). The main providers of educational resources are governments (central, state and local); households; NGO's; and schools. The expenditure per student varies depending on whether the student is studying in a public school or not. In our main analysis, we are interested in the total non-household education expenditure on a student which is defined as the expenditure by the government, NGOs and the school.

Figure 2.1 shows the ratio of non-household expenditure to total expenditure on each student by level. For public schools, this ratio is high. At the primary level, the non-household expenditure is NPR 7,209 per student which is 84 percent of the total expenditure. The government accounts for 95 percent of the total non-household expenditure per student-year. At higher education level, the proportion of non-household to total expenditure declines to 52 percent, but it terms of levels it increases to NPR 30,385. This pattern is markedly different for private schools, with households accounting for the bulk of the expenditure. Accordingly, at the primary level, the non-household expenditure is NPR 1,106, which is 6 percent of the total expenditure. For the higher-level private institutions, the non-household expenditure is NPR 14,337 per student which is 17 percent of the total expenditure. Comparing the distribution of subsidies in Nepal to Uganda, a country for which NEA has comparable data, the proportion of non-household expenditure to total expenditure is higher in Nepal across all the education categories.

[1] NEA data can be accessed using the link http://uis.unesco.org/en/news/national-education-accounts

[2] The seven other countries are – Côte d'Ivoire, Guinée, Lao PDR,Senegal, Uganda, Vietnam and Zimbabwe. Only Nepal and Uganda data distinguishes the per student government expenditure between public and private schools and also includes household expenditure.

2.2.2 Household data

We use the third round of Nepal Living Standards Measurement Survey (LSMS), which was conducted in 2010-11 by the Central Bureau of Statistics of Nepal and the World Bank to capture the demographic and consumption details of the households[3]. Useful for our analysis, the survey contains individual-level information on education, business and job characteristics, benefits received and migration status of the household members. It is a nationally representative survey and the cross-sectional data consists of 5,988 households and 28,760 individuals (CBS (2011)) [4]. We restrict our sample to households whose household-heads are aged between 20 and 80 years and whose education details are not missing. This reduces the number of households to 5,886.

We use the total household consumption data from the survey to impute the tax payments described in the next sub-section. The survey uses seven day recall period method to determine the consumption expenditure. There is a positive consumption gradient with education. The median per capita annual consumption of households whose head has finished primary education is around NPR 27,000, and this increases to NPR 65,000 if the household head has finished bachelors degree (Table 2.1).

Our main independent variable is the years of education completed by the household head, as reported in the LSMS data. For the grades 1 to 12, the number of years of education completed is equal to the school grade. The household heads with bachelors and masters degree are coded as having completed 15 and 17 years of education because the duration of bachelors and masters degree in Nepal is of three and two years respectively. Table 2.1 shows that around 65 percent of the household heads have highest education grade that lies within the primary education category, 18 percent have secondary education and the rest have some level of higher education. We also construct two additional measures to capture the education attainment of the rest of the household members -– first, total years of education of all the household members after excluding the household head and second, the maximum educational grade achieved by a member except the household head. We use these measures as controls in the empirical specifications.

[3]The details of the survey can be found at https://microdata.worldbank.org/index.php/catalog/1000/study-description

[4]We do not use the Nepal Labor Force Surveys (NLFS) which were conducted in the years 2007/08 and 2017/18 and do not overlap with the data in the NEAs which runs from 2009-15. Like most labor force surveys they do not record information on total consumption, the type of school attended by the individual and benefits received from the government. The national Census was conducted in 2011, but it doesn't record consumption or income. We choose the setting of the analysis to be Nepal over Uganda because the data from Uganda do not permit certain robustness checks.

2.2.3 Taxes, Formality, and Benefits

In this subsection, we briefly describe the methodology used to calculate the total tax payments of a household after adjusting for formality. There are two primary taxes paid by the households: income tax and value added tax (VAT). We impute the income tax payments of the household heads engaged in the formal sector by, first, constructing a tax table that delineates the income tax payments for each level of income according to the income tax schedule. Then, the household's consumption, c, equals $f(s) - T(s)$, where $f(s)$ is the taxable income, a function of years of schooling s, and $T(s)$ is the income tax payment. Next, we merge the income tax table with the LSMS data by matching the imputed consumption from the tax table to the reported consumption in the survey data. Thus, we can assign the taxable income and tax payment to each household in the survey data using the tax table. We calculate income tax payments only for household heads that are employed in the formal sector because those in the informal sector, in practice, do not pay income tax.[5]

To define tax-formality, we rely on the result that wage earners whose income is subject to third-party reporting are more likely to be tax-payers (Kleven et al. (2011), Slemrod (2007)). Therefore, anyone whose income is reported to the government by the employer or who benefits from social security net is considered to be in the formal sector (Azuara and Marinescu (2013); Camacho, Conover and Hoyos (2014)). According to our preferred definition, a wage earner is considered to be in the formal sector if any job that she does displays at least one of the following features - tax is deducted by the employer, employee contributes to the provident fund, pension on retirement, or subsidized medical care. Since self-employed are less likely to pay taxes, we consider only those businesses who are registered with the government as tax-payers (in an alternate definition, even these businesses are considered to be in the informal sector. These alternate definitions are discussed in the appendix). All other economic activity, in particular, agriculture is considered to be in the informal sector. Of the sample of household heads aged between 20 and 80 years, around 16 percent of 5886 household heads are in the formal sector.[6] No matter which definition of formality we use, the proportion of household heads in the formal sector increase as the years of education increase. For example, about 4 percent of people without school

[5] A more detailed explanation of the methodology used to calculate income tax and a description of Nepal's income tax schedule is given in the appendix.

[6] Of the 16 percent, 11.6 percent of the household heads have an income that is subject to non-trivial marginal income tax rate. (Unlike most countries, the marginal tax rate of lowest income bracket in Nepal is non-zero and equal to 1 percent.) This number compares favorably with the estimates of registered taxpayers in Nepal who are 10 percent of the total number of households (Inland Revenue Department Report (2015) and CBS (2012)).

education are employed in the formal sector, while 62 percent of people with a masters degree are engaged in the formal sector (Figure 2.2).

The value added tax (VAT) rate in Nepal is 13 percent for all the commodities. Some commodities like rice, pulses and other essential food items are zero-rated. In the main specification, we assume that the VAT is paid on the entire household consumption, although we exclude food consumption from the VAT tax base in the sensitivity analysis.

The total tax contribution of the households is the summation of Income tax and VAT[7] . Figure 2.3 shows a steep positive gradient of total taxes with education. The positive slope is due to the combined effect of increase in the tax base (income and consumption) and formality levels. People with higher education tend to be in the formal sector and thus, pay both income and value-added tax to the exchequer. Conversely, those with low education are mostly engaged in the informal sector and end up paying only VAT. For instance, people with no education on average pay NPR 18,757 in total taxes of which the VAT represents 95 percent. In contrast, people with higher education pay NPR 68,607 as taxes, of which VAT is only 60 percent.

From the government's standpoint, we also need to consider the transfers that the government makes as part of various welfare schemes. We can expect that more education leads to lower take-up and dependence on the welfare schemes. In the household survey we have information on the payments received by the households on seven major cash-transfer programs [8]. The magnitude of benefits provided under these schemes is much lower than the taxes remitted (Figure 2.3). Hence, we only include the benefits in the sensitivity analysis, and verify that our results remain robust.

2.3 Model

In this section, we analyze the choice of years of schooling in a stylized model based on the seminal work of Jacob Mincer (1958). We include taxes and subsidies in the analysis and use this framework to derive the fiscal externality (the gap) associated with an additional year of school. An individual starts school at $t = 0$ and faces an interest rate of r. The direct cost of schooling in year t is $c(t)$, which includes tuition, transport, uniforms, books, etc. The government contributes a subsidy of $\gamma(t)$, so the student only sees the net

[7]Contributory deductions from income, such as contributions towards provident fund, are not included in the total tax payments of an individual because the taxpayer gets back the return on these deductions over her lifetime. Thus, such contributions will be taken into account by an individual while deciding the level of schooling, and are therefore not a fiscal externality.

[8]These programs are - old age pension, widow pension, disability allowance, endangered ethnicity's pension, maternal incentive scheme, martyr's family benefits, and people's movement victim benefits

cost. The student leaves school at time s, which also represents the total years of schooling accumulated. At that time, the now former student starts earning $f(s)$, for $f'(s) > 0$. (We use primes to denote first derivatives.) Once working, the person has to pay a tax of $\tau(s)$. (Taxes are generally functions of earnings instead of schooling, so we could have written $\tau(f(s))$, e.g., but we treat it directly as a function of s for notational simplicity.)

The person's lifetime income, net of taxes and direct costs, is

$$\int_s^T e^{-rt} \left[f(s) - \tau(s)\right] dt - \int_0^s e^{-rt} \left[c(t) - \gamma(t)\right] dt$$

if the person stops working at T. From the individual's perspective, this object is the present value of his/her human capital. Let $\tilde{r} \equiv \dfrac{r}{1 - e^{-r(T-S)}}$, which we can think of as a required rate of return, adjusted for number of years a person works$(T - S)$.

What choice of years of school maximizes the individual's human capital? If we take the derivative with respect to s, we attain the following first-order condition (FOC) for optimality:

$$(f^{'} - \tau^{'})/\tilde{r} = (f - \tau) + (c - \gamma) \tag{2.1}$$

The left-hand side is the marginal benefit associated with additional time in school. This includes the increase in labor productivity (f'), but also the change in taxes (τ'). These changes apply to future flows, and the interest rate accounts for the accumulation of these flows over time. The marginal costs are found on the right-hand side of the equation and are grouped into two concepts. The first is the opportunity cost. While a person is in school, he is not working, but neither does he pay taxes on income that he does not earn. The second is the direct cost, net of the subsidy.

This equation has an intuitive interpretation. If dollar's worth of time invested today yields a future flow of payments greater than $\tilde{r}$, then the student should continue in school. When the flow payment per dollar drops below $\tilde{r}$, the student should leave school.

Government policies can shift the choice of schooling, although there are combinations of taxes and subsidies that deliver the socially optimal decision. If taxes and subsidies are both zero, the condition for optimality reduces to the following:

$$f^{'}/\tilde{r} = (f + c) \tag{2.2}$$

This condition defines the undistorted optimum for s. But other, nonzero combinations also leave this choice undistorted. If

$$\tau'/\tilde{r} = \tau + \gamma \tag{2.3}$$

then these terms drop out of equation 2.1, which leaves us with the undistorted optimum for s. A simple example of this would be a proportional income tax paired with school subsidy of exactly the same proportion.[9]

In general, however, there is a gap between the government's marginal benefits and costs. This represents a fiscal externality: changes in the private choice of years of schooling spill over onto the government's budget. The marginal benefit for the government is $\tau'/\tilde{r}$, the taxes received per unit increase in schooling. But an additional year of education costs the government $\tau + \gamma$, the fiscal opportunity and direct-subsidy costs. If this gap is positive ($MB > MC$), then the government receives more revenue per marginal year of school than it incurs in costs. Policies that raise schooling, such as compulsory attendance or higher subsidies, might well relax the government's budget constraint and bring schooling closer to optimum. In contrast, a negative gap has the opposite implication.

To estimate the gap for various years of schooling, we need to calibrate the tax (τ) and subsidy (γ) functions. The NEA data gives information about subsidies for different levels of school, both public and private. We use data from the LSMS expenditure survey to compute taxes and then calibrate $\tau(s)$.

2.4 Empirical Model and Results

We use a quantile regression to examine effect of education on tax payment at the median, as well as, at the 25th and 75th percentiles. We also use an ordinary least squares (OLS) regression to consider effects at the mean. These methods provide a convenient estimate of the level and (conditional) gradient of tax with respect to years of schooling for the various summary statistics. Let τ_h be the total tax payments of the household h. The primary dependent variable is the number of years spent in school, $Eduyear_{ih}$, by the household head i living in the household h. X_{ih} is a vector of other demographic characteristics of the head such as age. Z_h is a vector of household characteristics like the total years of education of other household members. Then, the θth quantile of the conditional distribution of τ_h, given the covariates, is a linear function,

$$Q_\theta(\tau_h | Eduyears_{ih}, X_{ih}, Z_h) = \alpha_{0\theta} + \beta_{1\theta} Eduyears_{ih} + \mathbf{X}'_{ih}\delta_{1\theta} + \mathbf{Z}'_h\delta_{2\theta} + u_{\theta ih} \tag{2.4}$$

Now, we can test how far is the fiscal gap in Nepal from the condition defined in equation 2.3 and hence, the undistorted optimal choice of schooling. The fiscal gap, defined as

[9] A proportional tax and subsidy of λ yields an individual FOC of $(1-\lambda)f'/\tilde{r} = (1-\lambda)(f+c)$, which reduces to equation 2.2 if $\lambda < 1$.

the difference between the marginal benefit and the marginal cost, is calculated by using the following formula:

$$\hat{\beta}_{1\theta} - r(\hat{\alpha}_{0\theta} + \hat{\beta}_{1\theta} * S + \bar{\mathbf{X}}'_{ih}\hat{\delta}_{1\theta} + \bar{\mathbf{Z}}'_{h}\hat{\delta_{2\theta}} + subsidy_{S\theta}), \tag{2.5}$$

where r is the discount rate and S is the point at which the fiscal balance is calculated. $\hat{\beta}_{1\theta}$ represents the marginal benefit (τ'), while the term in the parenthesis represents the fiscal opportunity and direct-subsidy cost ($\tau + \gamma$).[10] While calculating the subsidy cost, we use the θth quantile of subsidy at grade S.

For OLS regressions, we use the following statistical model:

$$\tau_h = \alpha_0 + \beta_1 Eduyears_{ih} + \mathbf{X}'_{ih}\delta_1 + \mathbf{Z}'_h\delta_2 + u_{ih} \tag{2.6}$$

Then, the fiscal gap is:

$$\hat{\beta}_1 - r(\hat{\alpha}_0 + \hat{\beta}_1 * S + \bar{\mathbf{X}}'_{ih}\hat{\delta}_1 + \bar{\mathbf{Z}}'_h\hat{\delta}_2 + subsidy_S), \tag{2.7}$$

where $subsidy_S$ is the mean subsidy at grade S.

Since the NEA dataset has subsidy information at aggregated education levels such as primary, secondary etc. and not at individual grades, we categorize household heads into three education categories: primary (0-5 grades), secondary (6-10 grade) and higher education (11-17 or masters grade). Subsequently, we conduct our empirical analysis within these education categories.

We find that an additional year of school is associated with substantial government spending and government revenue. For an interest rate of 3%, these fiscal costs and benefits are approximately balanced, though tilting negative, for primary and secondary school. For higher (tertiary) education, they instead tilt positive. Table 2.2 contains these results. For primary school, fiscal marginal benefits are generally less than fiscal marginal costs, although only by a small margin. Panel A of Table 2.2 shows these results. Consider first the median outcomes, which are shown in the first column. Another year of school is associated with a median tax payment that is higher by NPR 610 (approximately US$8.47 at the time of the survey). This fiscal benefit is akin to a dividend that is paid continually in the future. But there are two upfront costs. One is the opportunity cost: the foregone tax payments that are not made because the student is in school instead of working. We esti-

[10]In our analysis, we demean the controls included in the vectors X_{ih} and Z_h, so that the term in the parenthesis reduces to $\hat{\alpha}_{0\theta} + \hat{\beta}_{1\theta} * S + subsidy_{S\theta}$, where $\hat{\alpha}_{0\theta} + \hat{\beta}_{1\theta} * S$ is the fiscal opportunity cost. In other words, the intercept $\hat{\alpha}_{0\theta}$ can be interpreted as θ_{th} quantile of tax payments of the household whose head has zero years of schooling and mean value of other observables.

mate these to be almost NPR 18,750, which is the model's prediction for tax remittances by someone with five years of schooling and the mean of the other observables. The other fiscal cost is the school subsidy itself, which we compute as NPR 7209. We multiply these two costs by the 3% discount rate and subtracted from the benefit to obtain a gap between marginal benefits and costs of NPR 169 (US$2.35). This is not significantly different from zero at conventional levels of confidence. This amounts to 1/2 of a percent of household consumption per capita.

Given the simplicity of this calculation, we can take a moment to discuss the effect of a few small modifications. First, getting fiscal costs and benefits to exactly balance in this calculation would imply a break-even interest rate of 2.35%. This calculation is for an infinitely lived person, and therefore we require an even lower interest-rate to break-even, if mortality and retirement were taken into account. (Bleakley (2018), discusses incorporating death rates into interest-rate calculations for human capital. He argues that, for modern life tables in developing countries, interest rates typically need to modified by less than 100 basis points.) Another simple modification is to evaluate the fiscal gap for a student stopping at four rather than five years of school. By assumption, the marginal benefit and marginal subsidy cost would be the same, but the opportunity cost would be lower by NPR 610. This would only close the discounted gap, however, by NPR 18 (610 times .03), approximately a 10th of the total gap. (We discuss more complicated sensitivity analysis below.)

For other statistics of the distribution, we also find slightly negative balances of fiscal benefits and costs. These are found in the remaining columns of Panel A, where we consider the 25th percentile, 75th percentile, and mean as outcomes. In all cases, years of schooling predict higher tax payments, with the larger effects being at the higher percentiles and for the mean. As before, however, this flow of future benefits is arrayed against substantial costs in the beginning.[11] For a 3% discount rate, the net fiscal balance remains negative, however it is closer to zero than it was at the median. These numbers also reflect relatively small gaps when compared to household consumption per capita. Indeed, at the mean, this gap is less than one part in one thousand of household consumption.

Next we consider secondary education, for which fiscal balances turn somewhat more negative. These results are found in Panel B. Tax payments rise with education for all four

[11] Survey respondents report whether they attended a public (government) or private school. We use this information to impute different subsidy rates to each individual. Therefore, the implied subsidy for each column will depend on the mix in the data. For those reporting primary school as their highest level of education, nearly all of them report having attended government schools. So, the subsidy amount is the same across the three quartiles. Because a small number did report attending private schools, the subsidy rate at the mean is slightly lower. In the next table, we present some decompositions meant to better isolate the government support of schools.

of the statistics considered, however the strongest gains are at the 75th percentile and for the mean. Marginal fiscal costs come in between NPR 1700 and NPR 2200 (US$23 and US$31). These gaps are over 1% of the value of household consumption per person, except at the mean.

Finally, we turn to higher (tertiary) education, which starts at grade 11 in Nepal. These results are found in Panel C. An additional year of education is associated with substantial increases in tax payments for this group. This arises in part because of the effect of education on income, but what really distinguishes this group from the others is the much higher rate of formality by those with tertiary education. (Recall Figure 2.2.) As a result, this group pays more in direct taxes on the margin as its income rises. These gaps are substantial. At the 75th percentile, the gap between fiscal marginal benefits and costs is almost NPR 3737, which represents over 2.5% of household consumption per capita. At the mean, this is even larger: the gap is almost NPR 5400 (US $75), or over 5% of consumption. This represents a substantial fiscal benefit to encouraging higher education.This contrasts with the results from primary and secondary education, where the net fiscal impact is likely negative, albeit often difficult to distinguish from zero. Viewed in a different way, this represents a significant disincentive to attain higher education.

2.5 Sensitivity Analysis

The results above are qualitatively robust to a variety of alternate strategies for measurement and modelling, as we show in this section. The main set of robustness checks are found in Table 2.3, where we report estimates of the gap between fiscal marginal benefits and costs.

The first set of checks explore robustness to measuring funding levels. In row 1, we repeat the analysis from Table 2.2, but only with household heads who reported attending public schools. Estimates of the fiscal loss or gain are similar to the baseline, with the exception of primary school, where the losses appear worse, yet still less than 1% of consumption. In the second row, we continue our restriction to only public schools, but also consider only the government's contribution to the public-school subsidy. This moves the fiscal balance associated with your school in a positive direction, but only by a little bit. This is because non-government subsidies to schools are dwarfed by those from the government. In the third row, we treat higher education as extending all the way through to the Masters level. There appears to be a fiscal gain at this level as well, although this represents a trivial part of the sample. We turn next to the anomalously large administrative expenditure associated with grade 10. Nepal conducts national-level exam at grade

10 which is compulsory for all students to graduate to the next level. The cost incurred in conducting this exam can explain the jump in the subsidy reported in the National Education Accounts data. Removing the administrative cost component from the subsidy data reduces the gap at the secondary level and brings it closer to zero.

Next, we tweak our model for taxes on the household. The VAT in Nepal excludes certain food items, although the survey did not provide enough information to separately disaggregate covered versus non-covered expenditures. In row 5, we simply exclude all food consumption expenditures from the base use to compute that payments. This makes essentially no difference for the results. The next row reports the gap using the self-reported total expenditure on land, property, housing and income taxes in the LSMS data , instead of imputed income taxes. We do not use the self-reported tax variable in the main analysis because we suspect measurement error in the variable. For instance, professionals whose income is subject to third-party reporting, like government employees, tax officials etc., report less than 1% of their consumption in tax payments[12]. This seems very low because at the mean level of consumption, they should be paying 13% of their consumption in taxes, according to the tax schedule. Since the taxes of the formal sector workers are remitted by the employers, such workers may not perceive tax payments as expenditure and forget to report them in the survey – which can explain the low levels of self-reported taxes. The fiscal gap at the primary and secondary level remain unchanged because income tax is a small proportion of the total taxes at those educational levels. For higher education, the fiscal gap reduces because of lower level of self-reported tax as compared to imputed tax.

Another concern is that tax payment is only part of the fiscal story. If education reduces poverty, the first effect might not be to increase tax payment, but to reduce benefit receipt. Nepal has various cash and in-kind benefits.[13] The survey has information on these, and thus we can measure net tax payments (taxes minus benefits). While this sounds like it could possibly change the results, in practice the effect is quite small, and Figure 2.3 shows why. There we plot taxes paid and benefits received as a function of years of education. The slope of taxes with respect to education is evident in the graph. The relationship between benefits and education is, in fact, quite small and sometimes not even sloping upwards.[14] Accordingly, when we use a measure of taxes net benefits as the dependent

[12]Several studies like Kleven et al. (2011) show that tax evasion rate is close to zero for people whose income is subject to third-party reporting.

[13]Note that we are not treating the subsidized access to schooling as a benefit for the household here. This is because we examine the subsidy looking forward from the perspective of the student whose income and tax payment will be higher, not of the parents whose children might go to public school.

[14]On one hand, we expect the slope of benefits to increase with education because more educated people might be more aware of their legal entitlements. This is more important if the access to these programs is not

variable, the results are hardly different from baseline. This is seen in Row 6.

Our final modification to the model of taxes is to use a more conservative definition of formality which reduces the proportion of people employed in the formal sector at each level of education. Reduction in the formal sector employment causes a reduction in the imputed income tax payments, which in turn affects both the tax gradient with respect to education and the opportunity cost of spending an extra year in school. Row eight shows that for most of the specifications, the fiscal gap worsens due to decline in income tax payments.

In the next five rows, we consider a few alternative ways of accounting for differences in household composition. Our sample consists of households whose household-heads are aged between 20 and 80 years. This includes some household heads that currently pursuing higher education and still enrolled in school. In Row 9, we only include household heads that have finished schooling which reduces the sample in the higher education category from 893 to 808. The fiscal balance now becomes significant at the lower quantiles of the distribution and increases in magnitude for all the specifications. In Row 10, we restrict our analysis to male-headed households only. For primary and secondary education, the fiscal balance associated with an additional year of schooling looks worse, but looks slightly better for tertiary. None of the resulting changes are especially large, however, when considered as fractions of household consumption. Next, we consider alternatives to our default strategy of controlling for the education of other household members. There is a substantial correlation of education within a household, and it would not make sense to credit the household head's education with increases in expenditures that came from the increased earning power of a spouse, for example. Accordingly, the fiscal balances tend to look better when we drop the control for years of schooling held by other adults in the household (see Row 11), but this suffers from an omitted-variable problem. As an alternate approach to using the sum of years of schooling held by other household members, in Row 12, we use the maximum instead. These results are quite similar to the baseline. Finally, we test our strategy of using the education level of the household head in our analysis. We do so because we assume that household consumption is a function of her education. This may not be true. To understand the effect of this assumption, we redo our analysis by using the education level of a random working member of the household and

universal. In our sample, only 16 percent of the households report receiving money under any cash-transfer program. On the other hand, the slope can be negative if the benefits under the social protection programs do not form a significant proportion of consumption and there are pecuniary and non-pecuniary costs in getting access to entitlements. Average payments to the entire household under the old age pension program, for instance, are 3 percent of the mean per capita income of households whose heads have completed higher education. We find that the two effects mostly cancel each other – leading to an almost flat gradient of benefits with education.

control for the sum of education of the rest of members. Row 13 shows that the results remain statistically indistinguishable from the baseline .

In the next two rows, we modify the definition of education intervals for secondary and higher level. In Row 14, we change the starting grade of secondary education to grade 5 which is the last grade of primary education. Similarly, the beginning of higher education is taken as grade 10 instead of grade 11. Now, the tax gradient not only captures the effect of increasing education within secondary or higher level but also transitioning from previous level to the current level. This does not change the results very much, though the gap improves a little for all the specifications in the secondary level. This is possible if there are larger gains in income from transitioning from primary to secondary level than increasing years of education within secondary level. Lastly, we modify the definition of completing grade 10. The LSMS data distinguishes between people who have completed grade 10 versus those who have passed the national-level exam at the end of grade 10. The next education grade reported in the data is graduating from grade 12. Thus, anyone who drops out of grade 11 or 12 is coded as having passed the national-level 10th exam. Due to this peculiar feature of the data, we assume that anyone who passes the national-level grade 10 exam has 11 years of education. In Row 15, we use an alternate definition – people who passed national level education have only 10 years of education. This increases the sample size at the secondary level and reduces it at the higher level. Row 15 shows that the fiscal balance improves at the secondary level, but it worsens at the higher level, due to this modification.

Finally in Table 2.4, we employ alternative discount rates. When we lower the discount rate from 3%[15] to 1%, the fiscal gap improves as the discounted value of marginal cost falls (see equations 2.5 & 2.7). At higher level, the fiscal gap is now positive and significant at all the moments of the distribution. On the other hand, for primary and secondary level, the fiscal balance is quite balanced in majority of the specifications. Once, we increase the discount rate above the baseline, the fiscal balance, as expected, worsens (see Panel B and C of Table 2.4). At a discount rate of 5% the fiscal gap is negative for all the levels except for higher education, while at the rate of 9%, the marginal fiscal cost is higher than the marginal benefit for all the levels of education.

Next, we consider how migration would affect these calculations. (In the sample, around 32 percent of the households report having a member outside Nepal.) On one hand, migrants leave Nepal after their school years take within their human capital, which was built with a government subsidy to some degree. On the other hand, those migrants

[15]In the main analysis, we use the discount rate of 3% which is close to the average real interest rate of Nepal from 1975-2010 (Source: https://data.worldbank.org/indicator/FR.INR.RINR?locations=NP)

(many of whom left home in search of higher incomes) might very well send back remittances. These remittances would, if spent, expand the national tax base. We attempt to characterize the magnitude of these effects by using information on the migrant's education and remittances .

First, we calculate the probability of migration (P) for each household head based on her school grade [16]. This probability is equal to the proportion of people who migrated conditional on school grade. For instance, there are 7 international migrants out of 139 people, aged between 20 and 80 years, who have completed one year of education. Then, the probability of migration is 5 percent for household heads who have finished one year of schooling. Appendix Table B.3 contains these probabilities for each education grade. Next, we modify the tax payments using the migration probability. If the household head migrates, then she doesn't pay any taxes. In other words, we assume that the migrant leaves for work immediately after finishing school. This implies that there are no fiscal opportunity costs in terms of foregone tax payments on domestically earned income, because the government doesn't loose tax revenue if a future migrant remains in school for an additional year. However, the household members in Nepal must pay VAT when they consume the remittances, R, sent by the migrant. We assume that the household head, if she migrates, will send back money equal to the average remittance sent by a migrant of same education level. If the household head doesn't migrate then, there is no change in the tax payment. Thus, the migration-adjusted tax payments (M), are given by:

$$M = (1 - P)\tau(s) + 0.13 \times PR,$$

where VAT is 13 percent of remittances. Row 6 of the Appendix Table B.3 contains the average tax payments adjusted for migration at each education level. Table 2.5 shows that result do not change much if we use the migration-adjusted tax payments of the household heads. The fiscal balance improves a little for both higher and secondary levels in most of the specifications, while it worsens a bit for the primary level.

2.6 Conclusion

In this paper, we argue that government's investment in education can be analyzed in terms of expanding the tax base. This fact is not internalized by the individual as she

[16]In the LSMS survey, the household head, by definition, cannot be away from the household for more than 6 months in the last year and hence, is not classified as a migrant (CBS (2011)). Hence, we calculate the probability of migration of the household head by using the education information of the household members present at the time of survey and the migrants.

neither bears the full cost of education, because of subsidies, nor does she realize the entire benefits of education due to taxes. We find that, on average, these distortions–taxes and subsidies–do not cause deviations from the optimum choice of schooling at the primary and secondary level. However, at the tertiary level, the fiscal gap is positive and significant. This study provides a novel explanation for this: people with higher education are more likely to be in the formal sector and hence, pay income taxes. We subject our findings to a variety of sensitivity analyses, including the effect of emigration, and show that the results remain robust.

While the positive fiscal gap implies that the government can reap significant fiscal returns from investing in higher education, there can be spillover effects of such investment. On one hand, higher average education may cause a rise in transfers at lower levels of income (and education) which can worsen the fiscal balance at those levels. On the other hand, increasing higher education can speed-up formalization of the economy resulting in an increase in formal sector participation of people with primary and secondary education. This can improve the fiscal balance at lower levels of education. Measuring the magnitude of these effects is a subject of future research.

In future work, we would also like to match the subsidy data to administrative tax data and a survey that tracks individuals over their lifetime to improve our estimates. Lastly, this study would not have been possible without access to information on aggregate government spending per student at different levels of education. To the best of our knowledge, disaggregated information on subsidies is not available for a majority of developing countries. Since fiscal returns to education are an important policy parameter, it might be useful to construct such data for many more countries.

2.7 Tables

Table 2.1: Summary Statistics

	HH Consumption per capita (in NPRs)	Direct tax (in NPRs)	VAT (in NPRs)	Subsidy at end point of the interval (in NPRs)	Years of education of the household head	Proportion of households heads in the formal sector
			Panel A: Primary Education (Grades 0-5)			
Count	3,787	3,787	3,787	353	3,787	3,787
Median	26,768	0	15,296	7,209	0	0
25th percentile	18,292	0	10,710	7,209	0	0
75th percentile	40,345	0	22,692	7,209	2	0
Mean	33,859	1,061	18,833	7,192	1.09	0.05
Standard deviation	26,132	9,187	14,267	325	1.78	0.22
			Panel B: Secondary Education (Grades 6-10)			
Count	1,081	1,081	1,081	292	1,081	1,081
Median	38,760	0	22,877	30,510	8	0
25th percentile	25,950	0	15,331	30,510	7	0
75th percentile	59,864	0	33,747	30,510	10	0
Mean	49,303	7,766	28,653	29,816	8.16	0.21
Standard deviation	40,675	36,016	27,315	4,438	1.41	0.41
			Panel C: Higher Education (Grade 11 to Bachelor's Degree)			
Count	893	893	893	222	893	893
Median	64,898	0	32,969	30,385	12	0
25th percentile	43,100	0	21,547	30,385	11	0
75th percentile	97,275	16,105	49,276	30,385	15	1
Mean	80,091	23,718	40,877	27,277	12.29	0.46
Standard deviation	58,555	61,512	30,782	6,356	1.62	0.50

Note: This table presents the summary statistics of the main variables used in the analysis. The rest of the variables are described in the appendix. The primary data source is Nepal Living Standards Survey-2010. The subsidy data comes from National Education Accounts reports compiled by International Institute for Educational Planning (IIEP), UNESCO Institute for Statistics (UIS) and Global Partnership for Education [IIEP Reports 2016a 2016b]. This data can be accessed at http://uis.unesco.org/en/news/national-education-accounts

Table 2.2: Estimated Fiscal Benefits and Costs for a Year of Education

Panel A: Primary Education (Grades 0-5), N=3787

	Median	*25th %ile*	*75th %ile*	*Mean*
Years of Schooling	610***	511***	836***	870***
	(92)	(74)	(176)	(188)
Opportunity Cost at Grade 5	18747***	14199***	26069***	23294***
	(392)	(315)	(751)	(841)
Subsidy at Grade 5	7209	7209	7209	7192
Diff between MB and MC	-169**	-131**	-163	-44
	(81)	(65)	(156)	(165)
(MB-MC) / Consumption per capita	-0.005	-0.006	-0.003	-0.001

Panel B: Secondary Education (Grades 6-10), N=1081

Years of Schooling	547	220	1354	2221*
	(415)	(224)	(830)	(1172)
Opportunity Cost at Grade 10	26415***	18445***	41544***	40505***
	(959)	(516)	(1917)	(3023)
Subsidy at Grade 10	30510	30510	30510	29816
Diff between MB and MC	-1161***	-1249***	-808	111
	(393)	(211)	(785)	(1094)
(MB-MC) / Consumption per capita	-0.026	-0.039	-0.012	0.002

Panel C: Higher Education (Grade 11 to Bachelor's Degree), N=893

Years of Schooling	3737***	2000***	7623***	8875***
	(941)	(489)	(2398)	(1937)
Opportunity Cost at Grade 15	52675***	33080***	99141***	88647***
	(2945)	(1532)	(7508)	(6658)
Subsidy at Grade 15	30385	30385	30385	27277
Diff between MB and MC	1245	96	3737*	5398***
	(865)	(450)	(2206)	(1752)
(MB-MC) / Consumption per capita	0.014	0.002	0.027	0.05

Note:This table calculates the difference between MB and MC (fiscal gap) at the end-point of each level of education. The dependent variable is total tax payments - income and consumption tax (VAT). We assume that only people in the formal sector pay income tax. Everyone pays VAT. Main coefficient of interest is "years of schooling" of the household-head which is equal to the MB. "Opportunity cost" is the tax forgone due to an additional year of schooling at the end-point of the interval. "Subsidy" is non-household expenditure per student which includes central and local government expenditure, international and local NGO, external loans and grants, off-budget assistance and internally generated funds by the schools. Other controls include quadratic terms of the age of the household-head and the sum of education level of all the other family members. We demean the controls so that the marginal cost is the discounted value of the sum of opportunity cost and subsidy. We use a discount rate of three percent. In the final row of each panel, we take the average per-capita consumption of households whose head has education level equal to the end-point of the interval, and use it to standardize the fiscal gap. In 2010-11, the year of the analysis, 1 USD was equal to 72 Nepalese rupees. The primary data source is Nepal Living Standards Survey - 2010. The subsidy data comes from National Education Accounts reports compiled by International Institute for Educational Planning (IIEP), UNESCO Institute for Statistics (UIS) and Global Partnership for Education [IIEP Reports 2016a 2016b]. This data can be accessed at http://uis.unesco.org/en/news/national-education-accounts. Standard errors are shown in parentheses. Three stars denotes signifigance at the 1% level; two stars, 5%; and one star, 10%.

Table 2.3: Alternate Estimates of the Gap

		Primary				Secondary				Higher			
		Median	25th %ile	75th %ile	Mean	Median	25th %ile	75th %ile	Mean	Median	25th %ile	75th %ile	Mean
1	Sample restricted to public schools	-492**	-220	-320	-269	-1182***	-1241***	-805	155	1215	147	4116*	5686***
		(206)	(177)	(456)	(350)	(419)	(206)	(783)	(1112)	(953)	(469)	(2339)	(2022)
2	Only public schools and only government subsidy	-483**	-211	-311	-259	-1091***	-1151***	-715	245	1275	206	4175*	5745***
		(206)	(177)	(456)	(350)	(419)	(206)	(783)	(1112)	(953)	(469)	(2339)	(2022)
3	Top higher ed. master instead of bachelor									429	-465	2705*	4327***
										(658)	(359)	(1577)	(1265)
4	Omit admin expenditure for secondary					-553	-641***	-200	705				
						(393)	(211)	(785)	(1094)				
5	VAT only on non-food consumption	-76*	-152***	171*	147	-684**	-646***	-506	429	977	232	4741**	5567***
		(43)	(22)	(91)	(145)	(268)	(141)	(730)	(1032)	(650)	(374)	(2175)	(1673)
6	Dependent variable is self-reported tax instead of imputed tax	-185**	-129**	-205	-215*	-1116***	-1272***	-871*	-464	86	-587*	1556**	1182**
		(80)	(65)	(138)	(112)	(336)	(207)	(495)	(525)	(428)	(323)	(611)	(582)
7	MB is equal to taxes net of benefits	-169***	-131***	-163***	4	-1161***	-1274***	-808***	-9	1245***	93**	3737***	5510***
		(12)	(9)	(23)	(25)	(29)	(15)	(58)	(91)	(88)	(46)	(225)	(200)
8	Conservative definition of formality	-175**	-142**	-209	-79	-1102***	-1221***	-969	-48	738	-380	2074	3747**
		(81)	(65)	(153)	(161)	(368)	(207)	(718)	(1095)	(597)	(390)	(1749)	(1652)
9	Household heads who have completed schooling									2359**	881*	8236***	6943***
										(1025)	(503)	(2396)	(2122)
10	Male-headed households only	-309***	-310***	-265	-76	-1189**	-1213***	-1108	-20	1018	137	4864**	6076***
		(92)	(77)	(174)	(195)	(465)	(241)	(958)	(1332)	(932)	(491)	(2451)	(1897)
11	Drop control for other household members' education	122	72	224	519***	-188	-937***	-65	1907	3220***	807*	9031***	8797***
		(91)	(72)	(148)	(185)	(328)	(239)	(883)	(1166)	(795)	(485)	(2191)	(2073)

Table continues on the next page

Table 2.3: Alternate Estimates of the Gap (continued)

		Primary				Secondary				Higher			
		Median	25th %ile	75th %ile	Mean	Median	25th %ile	75th %ile	Mean	Median	25th %ile	75th %ile	Mean
12	Use max of other household members' education instead of sum	-203**	-87	-172	21	-819**	-1140***	-965	624	2050**	47	7077***	5650***
		(85)	(63)	(164)	(172)	(357)	(253)	(780)	(1141)	(857)	(460)	(2174)	(2059)
13	Using random working-age member instead of household head	-164*	-160**	-143	77	-849**	-970***	-657	186	853	289	2925*	5007***
		(88)	(68)	(174)	(208)	(406)	(257)	(679)	(676)	(597)	(387)	(1562)	(1710)
14	Each level starts at endpoint of previous level					-882***	-1049***	-571	410	1444**	274	4281***	5224***
						(256)	(134)	(449)	(622)	(575)	(346)	(1558)	(1452)
15	Alternative definition of Grade 10					-499	-939***	-70	548	-115	-447	1106	2736
						(373)	(193)	(757)	(820)	(1872)	(784)	(3865)	(2747)

Note:This table presents estimates of the gap between fiscal marginal benefits and marginal costs of education under different assumptions. See notes from Table 2.2 for specifications. Standard errors are shown in parentheses. Three stars denotes significance at the 1% level; two stars, 5%; and one star, 10%.

Table 2.4: Alternate Discount Rates

	Median	25th %ile	75th %ile	Mean
	Panel A: Discount of 1 percent			
Primary	350***	297***	503***	565***
	(88)	(71)	(169)	(180)
Secondary	-23	-270	633	1518
	(408)	(219)	(815)	(1146)
Higher	2906***	1365***	6328***	7716***
	(915)	(476)	(2333)	(1875)
	Panel B: Discount of 5 percent			
Primary	-688***	-559***	-828***	-654***
	(74)	(60)	(142)	(149)
Secondary	-2300***	-2228***	-2249***	-1295
	(378)	(204)	(756)	(1043)
Higher	-416	-1174***	1147	3079*
	(816)	(425)	(2082)	(1630)
	Panel C: Discount of 9 percent			
Primary	-1726***	-1416***	-2159***	-1873***
	(61)	(49)	(117)	(119)
Secondary	-4577***	-4186***	-5131***	-4108***
	(350)	(189)	(700)	(944)
Higher	-3738***	-3712***	-4034**	-1558
	(723)	(376)	(1845)	(1392)

Note: This table presents estimates of the gap between fiscal marginal benefits and marginal costs of education under different assumptions about alternative discount rates. (The discount rate used in the original analysis is 3 percent.) See notes from Table 2.2 for specifications.Standard errors are shown in parentheses. Three stars denotes significance at the 1% level; two stars, 5%; and one star, 10%

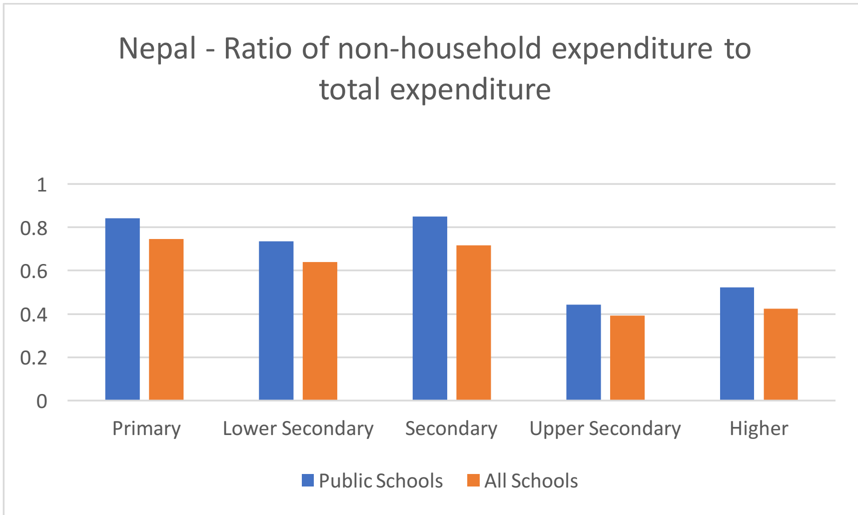
Table 2.5: Gap between MB and MC, adjusted for migration

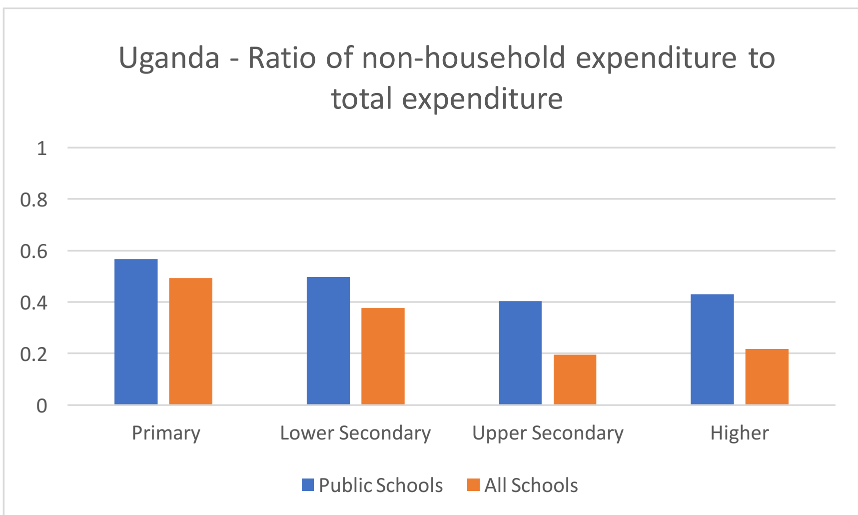
	Median	25th %ile	75th %ile	Mean
Primary	-381***	-224***	-503***	-443***
	(74)	(61)	(135)	(141)
Secondary	-866***	-869***	-700	31
	(301)	(162)	(612)	(850)
Higher	1437**	23	4306**	5506***
	(699)	(379)	(1817)	(1506)

Note: In this table we use the same specification as Table 2.2. The dependent variable is adjusted for migration by using the formula given in the appendix. Standard errors are shown in parentheses. Three stars denotes significance at the 1% level; two stars, 5%; and one star, 10%

2.8 Figures

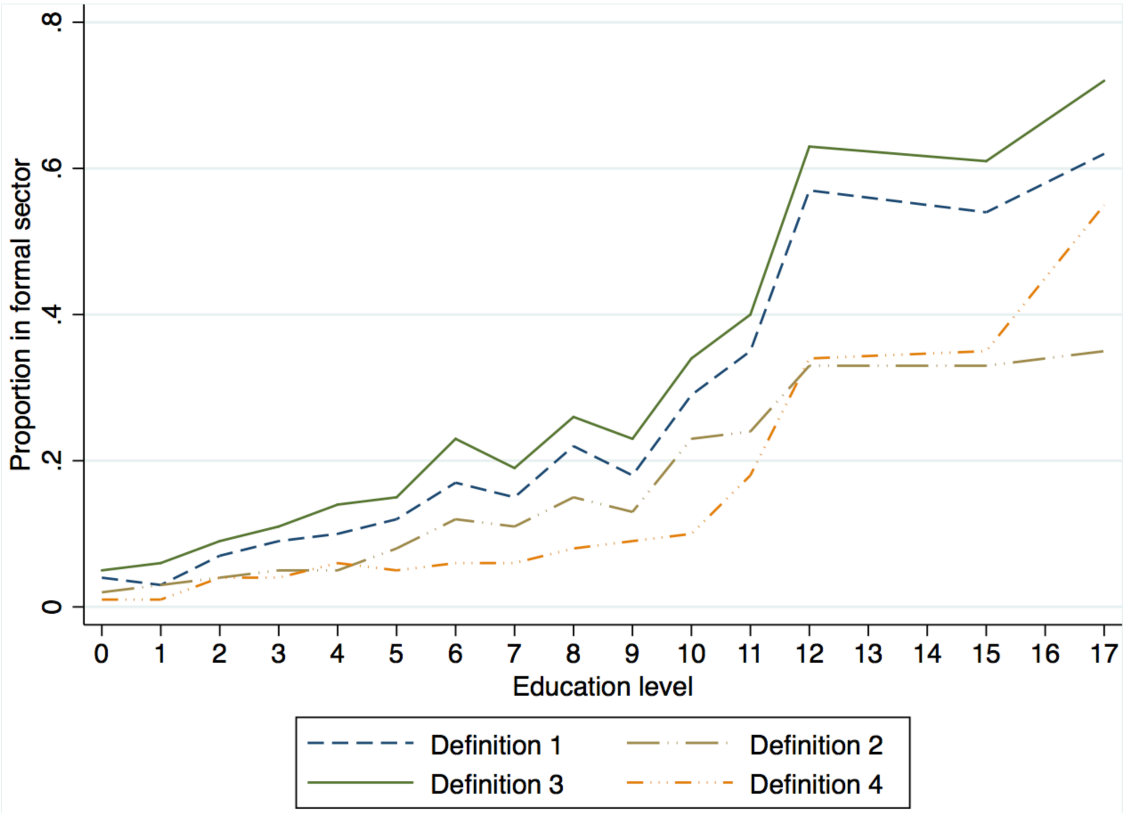
Figure 2.1: Education subsidies, by level, in Nepal and its comparison with Uganda





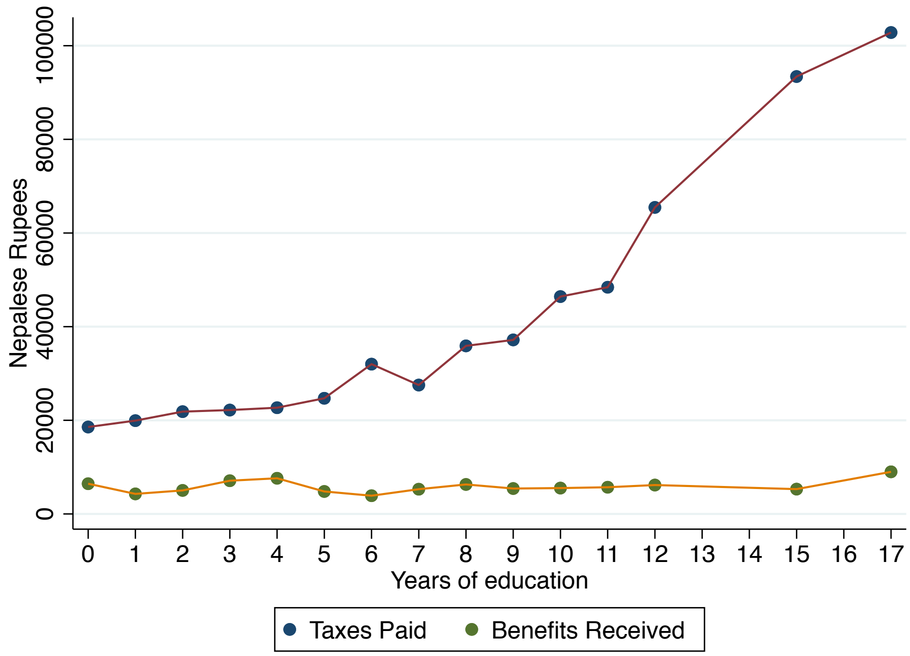
Note: The data for these graphs is sourced from National Education Accounts (NEA) reports. For Nepal, primary education goes from grades 1-5, lower secondary from 6-8, secondary from 9- 10, upper secondary from 11-12 and higher education's duration is of three years (Source: National Education Accounts In Nepal Expenditure for Education 2009- 2015. Kathmandu, 2016). For Uganda, primary education is from grades 1-7, lower secondary from 8-11, upper secondary from 11-13 and three years of higher education (Source: National Education Accounts Report. Kampala, 2016).

Figure 2.2: Formality and Education Level



Note: This figure graphs the change in proportion of formal labor force according to education level. For our analysis, formality is synonymous with being compliant with tax laws. Definition 1, which is used in the analysis, includes a wage earner in the formal sector if her income is reported to the government by the employer or if she receives benefits from the social security net. Additionally, businesses registered with the government are also included in the formal sector. Informal sector consists of rest of the workforce including workers engaged in agriculture. This graph is based on calculations using the data from Nepal Living Standards Survey 2010.

Figure 2.3: Taxes, Benefits and Education



Note: This figure plots the mean of benefits received under various cash-transfer schemes and taxes paid, according to the education level. Based on calculations using the income tax schedule of Nepal and data from Nepal Living Standards Survey 2010.

CHAPTER III

The Big-Brother Effect: Intra-household Determinants of Learning Crisis in India

Abstract

Does learning crises in developing countries affect all children within a family equally? Using nationally representative Indian data and exploiting within-family variation, this paper finds that elder children have a learning advantage over their younger siblings. Eldest sons have higher levels of learning than their younger brothers, and the gap widens in districts with stronger elder-son preferences as captured by district-level Sex Ratio at Last Birth. This results in a steeper birth-order gradient in those districts. Additionally, controlling for birth-order, boys outperform girls, pointing towards general son-preferences. Societal preferences affect learning through differential allocation of educational resources within a family.

JEL Codes: I21, J13, J16, O15, Z13
Keywords: Learning outcomes, Intra-household Allocation, Birth-order

3.1 Introduction

Despite substantial gains in school enrollment, a learning crisis has emerged in the developing world. The World Bank Development Report (2018) points out – "the average student in low-income countries performs worse than 95 percent of the students in high-income countries". The current literature has largely focused on exploring the reasons for low average levels of learning. These reasons include: inadequate school infrastructure (Duflo (2001)), poor teacher quality (Rivkin, Hanushek and Kain (2005), Duflo, Hanna and Ryan (2012)), low health investments in early childhood (Black et al. (2017), Almond and Currie (2011)), among others. In contrast, this study asks whether there are difference in learning among the children of the same household in an environment where the average learning levels are low. It is an important question because the heterogeneity in learning arising from taste-based factors may not be mitigated by addressing problems that cause poor overall learning outcomes .

Specifically, I ask the following questions : 1) Is there intra-household variation in learning outcomes with respect to birth-order, gender and sibling composition? 2) Can gender-preferences partially explain the intra-household variation? and 3) What are the mechanisms through which these preferences influence actual learning?

India provides an excellent setting for this study for three reasons. First, the country has made great strides in improving access to primary education, however the quality of education remains low (see figure 3.1). Second, like many other countries in Asia such as China and South Korea, Indian parents have a strong son-preference. Among sons, the eldest son has a higher place in the social hierarchy because he is expected to take care of parents in their old age, perform parents' last rites, and also helps his younger brothers (Das Gupta et al. (2003), Jayachandran (2015), Dyson and Moore (1983)). Third, there are significant regional differences in the gender preferences across the country which allows us to isolate the effects of these preferences on learning outcomes.(North India has stronger preference for sons as compared to south and north-east India. See figure 3.2)

The primary data source used in the empirical analysis is the Annual Status of Education Report (ASER) 2014; it is a nationally representative household-level survey conducted in all the rural districts of India. I also use this survey to calculate the sex ratio at last birth (SRLB) at the district level, which measure not only absolute son-preference but also preference for the eldest son.[1] I model the learning outcomes as a function of birth

[1]SRLB is defined as the number of boys per girl at the district level, by considering only the youngest children across families. For instance, if there are 10 families in a district and the youngest children across the ten families comprise of 8 boys and 2 girls, then the district-level SRLB takes a value of 4. SRLB increases before Sex ratio at Birth in settings like India where there is limited access to abortion (Yoo, Hayford and

order, gender and sibling composition. Then, I analyze how the effect of these factors change as the gender preferences change across Indian districts. I use several controls at the child and parent's level that influence learning outcomes such as such as school grade, parent's education and age, among others. To control for unobservable factors such as fertility decisions, I use two strategies. First, I include mother fixed effects in the model to control for time-invariant family-level unobservables. This strategy is similar to one used by Jayachandran and Pande (2017), henceforth referred to as JP(2017), in analyzing the height differences of children within the family. Second, I only include those families in the sample where the mother has completed fertility. Such mothers are identified using their age and the number of years since the birth of last child. I also include fixed effects for the year of birth and the district of residence as additional controls. Finally, I check if the results are robust to alternate model specifications and different definitions of learning outcomes.

There are three main findings of this study. First, there is a negative birth order gradient in learning outcomes for both the genders, with the first-borns having the highest learning outcomes. This result is consistent with Black, Devereux and Salvanes (2005) who find that increased family size affects the educational attainment of only the later born, and doesn't affect the average years of education of other siblings. Second, boys learning outcomes are higher than girls at each birth-order. The learning gap increases in districts with higher SLRB and indicates the role of absolute son-preferences in explaining the gap. Third, there is a significant effect of having an elder brother on the learning outcomes of boys. The eldest son is 2.1 percentage points more likely to attain the required level of learning according to the grade in which he is studying than a boy who has an elder brother – "the big-brother effect"[2]. This effect if positively correlated with the elder son preferences of the Indian parents. In fact, in the top decile of districts, ranked according to SRLB, the eldest son is 3.8 percentage point learning advantage over a boy who has an elder brother. For girls, the effect of having an elder brother is ambiguous. According to JP (2017), on one hand, if the girl has an elder brother, then the resources will be diverted towards him instead of the girl. They call it the "Sibling rivalry effect". On the other hand, if the girl doesn't have a brother, then the family might conserve resources in anticipation of a male child, which again adversely affects the education on the girl. This is called the "Fertility stopping effect". In India, the two effects seem to have an equal negative effect on the girl's education.

Agadjanian (2017)). The mean and standard deviation of the district-level SRLB is 1.41 boys per girl and 0.34, whereas the corresponding figures for district-level Sex Ratio of population from 0 to 6 years, according to 2011 census, is 1.08 boys per girl and 0.05

[2]For the entire sample, 38 percent of children reach the adequate leaning level

Next, I show that differential investment in education is an important channel through which birth order and sibling composition influence what children learn. Parents are more likely to send their elder children and the eldest sons to more expensive private schools and enroll them in private tuition. Thus, the parental preferences not only differentially affect the human capital outcomes of children at the post-natal stage (JP 2017), but they have an effect at the later stages of childhood too.

This study speaks to three strands of literature. First, it contributes to the literature of birth-order effects on education by proposing a novel explanation for such affects, that is, gender preferences. We know that there are several studies which document the effects of the birth order on outcomes such as education, earnings, height, among others (Black, Devereux and Salvanes (2005), Savage et al. (2013), JP (2017)). There is a nascent literature on the mechanisms through which birth order effects affects education such as time spent by parents on children (Price (2008), Pavan (2016)) and discipline imposed by parents (Hotz and Pantano (2015)). I show that elder son preference can cause a declining birth order gradient among boys and the desire of have at least one son can result in declining birth order among girls. By taking advantage of the large sample size of the ASER data and regional variation in gender preferences, this paper also shows the spatial heterogeneity in the birth order effects.

Second, I make a methodological contribution. Sibling composition is extensively used as an instrument for family size to estimate the effect of family size on education (Kugler and Kumar (2017), Kang (2011),Lee (2008)). Azam and Saing (2018) have questioned the exclusion restriction of using the sex of the first child as an instrument. By discovering an independent effect of sibling composition on learning outcomes, this paper lends support to their hypothesis.

Finally, there are studies which document gender differences in intra-household allocation of education resources like enrollment in secondary school (Azam and Kingdon (2013)) and private school enrollment (Sahoo (2017)). However, a less documented fact is the gender gap in learning outcomes, given that learning is influenced by several factors. We show that not only do such gaps exist across the two genders but also within a gender based on sibling composition. Eldest sons in the family have higher learning as compared to sons who do have elder brother. Relatedly, this result has implications for the substantial literature on calculating the returns to education (Psacharopoulos and Patrinos (2018)). Since, the correlation between years of education and learning varies by gender, sibling composition and birth order, there is a case for exploring the effect of these factors on returns to education.

3.2 Related Literature and Context

The quantity-quality trade-off as postulated by Becker and Lewis (1973) has been extended by recent literature which stresses on the importance of accounting for birth-order (Black, Devereux and Salvanes (2005), Azam and Saing (2018)). However, there is ambiguity regarding the exact effect of birth-order. Older children might be at an advantage if the younger siblings suffer from DNA impairment, mother's body has suffered physically from previous births, or parents are older and have less energy (Blake (1989)). Additionally, parents might have already chosen their favorite child for resource allocation putting the later-born at a disadvantageous position. Kantarevic and Mechoulan (2006) find that the firstborns have more years of education and higher earnings over a lifetime.

On the other hand, the later-borns might get an advantage if the parents are less credit constrained and well-settled in their careers (Parish and Willis (1993)). They can also benefit from elder sibling's education, particularly if the parents are not educated. Ejrnæs and Pörtner (2004) find positive birth order effects with youngest children having higher years of schooling. Coffey and Spears (2019) document an increased survival rate among the later-borns due to better health and social status of women. Thus, the net effect of birth-order on human capital formation in children is an open question.

A preferred strategy to estimate the quantity-quality trade-off in education is to use sibling composition as an instrument for variation in the family size [Kugler and Kumar (2017), Azam and Saing (2018)]. For instance, Conley and Glauber (2006) find that the families where all the children are of same sex are more likely to increase their family size than families with children of both genders. The crucial assumption is that "sibling sex mix affects educational measures only through an increase in sibship size?". This paper investigates and finds that the sibling composition independently affects the learning outcomes, and thus, casts doubt on the above assumption.

A few studies have estimated the effect of sibling composition on education in different contexts like Pakistan (Qureshi (2018)), USA (Butcher and Case (1994)), Turkey (Dayioğlu, Kirdar and Tansel (2009)) etc. However, the above studies have not explicitly looked at the influence of gender preferences on the effect of birth order and sibling composition on education. Moreover, the sample size is small which doesn't allow the authors to investigate spatial heterogeneity in the birth-order and sibling composition effects.

Preference for sons, more for elder sons.

In India, there is son-preference because of various reasons – son's perform last rites of parents, they "carry on one's family line", patrilineality (passing productive assets through the male line) and patrilocality (parents residing in son's house) (Das Gupta et al. (2003) and Jayachandran (2015)). Often, among the sons, it is the eldest son present who is entrusted to do these tasks. Anukriti (2018) while analyzing a financial incentive scheme in Haryana (a state in north India) finds that as "fertility decreases, the sex ratio at birth worsens as high son preference families are unwilling to forgo a son despite substantially higher benefits for a daughter". Jayachandran (2017) finds a "strong desire to have at least one son and a preference for gender balance thereafter" in the north Indian state of Haryana.

There is stronger son-preference as one moves from south to north India (Figure 3.2). Dyson and Moore (1983) argue that in the north, women have to migrate farther after marriage and parents can expect little help from them after marriage. Moreover, women have limited property rights in north as compared to south India. Eldest son has higher socio-economic status in the north because "males tend to cooperate with and receive help from other males to whom they are related by blood, frequently their adult brothers".

There is a vast literature documenting the existence of gender bias in allocation of various economic resources in India (Sen and Sengupta (1983), Behrman (1988), Jayachandran and Kuziemko (2011)). Within education, several studies document gender differences in intra-household allocation of education resources like enrollment in secondary school (Azam and Kingdon (2013)) and private school enrollment (Sahoo (2017)). This paper contributes to this literature by looking at an important outcome variable (learning), and investigating the impact of gender preferences on it.

3.3 Predictions

Intra-household variation in learning is a function of birth order, gender of the child and the gender of its siblings. The functional form can change as the underlying gender preferences change. If there is son-preference, in particular, elder son preference, then we should see the following patterns in the data:

1. The learning outcomes for boys should be higher than girls, conditional on birth order.

2. The slope of the birth order gradient would be more negative for boys in districts where there is a stronger elder son preference. The elder sons might get more resources for education and hence, have a significant advantage over their younger brothers.

3. Relatedly, boys who don't have an elder brother - they are first born or have only female siblings, will have higher learning outcomes as compared to boys who do have an elder brother. Again this effect should be larger in areas with stronger elder son preference.

4. For girls, having no elder brother versus having one is theoretically ambiguous. JP (2017) call this "sibling rivalry" versus "fertility stopping" effect. The girls who have an elder brother will suffer from the disproportionate allocation of resources towards the brother. On the other hand, if they have only elder sisters, then households might exceed their desired level of fertility to have at least one son.

5. The households who have only girls should be larger than households that have at least one son, as households might exceed their desired fertility to have at least one son (also pointing towards limited access/use of sex-selective abortion). This effect will be more pronounced in areas with stronger son-preference.

3.4 Data description

In the analysis, I use the Annual Status of Education Report (ASER) 2014 data [3] which captures the learning outcomes of children - the key variable of interest.[4] It is a household-level survey done in all the rural districts of India with 600 households surveyed in each district of the 31 states and union territories. Within each district, 30 villages are chosen and 20 households per village are surveyed. Information of all the children aged between 3-16 years is collected, with learning outcomes recorded for those in the age group of 5-16 years. Data is also collected for certain paternal and household characteristics. For this study, I am using the data from 2014 survey, the latest year for which it is available.

One way to gauge what children learn at school is to assess their math aptitude. In the data, five levels of math aptitude are recorded with the lowest ranging from no knowledge of arithmetic to the highest being able to divide a three digit number by a single digit.[5] I standardize the math level by creating a new binary variable - Standard Math Level (StdMath), which takes a value of 1 if the child has attained the requisite math level according to the grade in which she is studying, and 0 if she hasn't attained that level.[6]

[3]The report can be accessed using the link http://www.asercentre.org/Keywords/p/234.html. The data is available upon request

[4]IHDS II is another potential data source but is not entirely suitable as it only records learning outcomes of children aged between 8 to 11 years and has a much smaller sample size as compared to ASER data.

[5]Level 1 = No Arithmetic, 2 = Recognize numbers 1-9; 3 = Recognize numbers 11-99; 4 = two-digit subtraction; 5 = division

[6]Grade 1 and 2 children should be able to recognize numbers from 1 to 99; grade 3 students should be

The aforementioned method assigns the StdMath dummy a value of one to any child who can do division and is in grade 3 or above. Thus, children in higher classes will have a higher probability of having the value of dummy equal to one. The data shows that there is variation in the dummy for all school grades[7]. I also control for child age and school class in the regressions to alleviate the above concern. In the robustness checks, I also use non-standardized math scores with grade fixed effects, and the main results don't change. Finally, the standardized math learning outcome variable can be calculated only for children who go to school.

Additionally, the original sample had 642,911 child-level observations. I drop households that have either missing mother characteristics or child characteristics. This reduces the sample to 348,465 which is 52.8 percent of the original data. Thus, I have to assume that data is Missing Completely at Random (MCAR) to infer that the results are unbiased. Moreover, in the regressions with mother fixed effects, the missing data should not result in biased coefficients, as I drop the entire household whenever the data is missing even for a single child.

In the analysis, the birth order is defined according to the age of the child. Since the data exists only for children from 3-16 years, therefore, the assigned birth order may not be an accurate measure of the actual birth order as there can be children in the household who are younger than 3 years and not captured in the data. This is not a major concern as very young and very old children might not directly compete for resources with the school-going children. To define the birth order, I restrict it to three categories - firstborn (oldest), the second born and later born. I also drop twins from the data as it is difficult to assign birth order to them. Other child-level characteristics include child age and the class standard. Other important covariates include parent and household-level characteristics.[8]

Another variable of interest is Sex Ratio at Last Birth (SRLB) - which captures preference to have at least one son. I calculate this variable at the district-level by calculating the sex ratio using only the youngest children in the families where mothers might have completed their fertility. The survey doesn't ask if the woman has completed her fertility. I proxy for such mothers by restricting the sample to women who are at least 35 years old or who haven't had children for more than 4 years. Appendix Figure C.1 shows that the number of children born to a mother rises with the mother's age till the age of 35 years

able to do subtraction, and any child above grade 3 should be able to do division. The metric is decided using NCERT textbooks

[7]Even in grade 12, only 73 percent of the students are able to do division.

[8]Paternal characteristics - Parent's age and schooling. Household wealth is proxied by an infrastructure index created using principal component analysis. Infrastructure index uses dummies indicating if the house has electricity, toilet and television.

and after that, there is a slight negative trend as older children are not captured in the data. Appendix Figure C.2 shows that less than 14 percent of children in the restricted data (households with more than one child) have birth space between their closest sibling of more than 4 years. Figure C.2 shows heterogeneity across India in terms of SRLB with the northern states showing worse SRLB as compared to southern and north-eastern states. The top 10 percent of districts ranked according to the SRLB have more than 1.88 boys per girl.

The summary statistics in Table 3.1 show that 48 percent of all the children are girls. The average age of the child in the sample is 9.3 years. The average number of children per mother is 2.26 according to the sample, which is slightly lower than the national Total Fertility Rate (TFR) of 2.3 in 2013 [9]. The average age of mothers in the sample is 33 years with 58 percent of them having attended school. Additionally, 55 percent of the children are first born, 30 percent are second born, and the rest are later born.

3.5 Empirical Strategy and Results

3.5.1 Birth Order

To calculate the magnitude and direction of the birth order gradient on standardized math learning outcome, I use a multi-variate fixed effects regression framework. I create birth order dummies - with the oldest child being assigned the first child dummy. To control for child-level characteristics that can affect the learning outcomes - child age and the school standard in which she is studying are included. I also control for parental education and household-infrastructure which can influence the learning of the children. District Fixed Effects are included in some of the specifications to control for time-invariant district level factors that affect the educational attainment.

However, as JP(2017) point out - households can still vary in terms of unobservables like fertility decisions. The birth order dummies can pick up the effect of family size which will lead to biased coefficients. To control for the family size, I employ two strategies. First, the sample is restricted to mothers who may have completed their fertility[10], and then the total number of children in a family is included as a control variable. The other strategy is to use Mother FEs, which controls for fertility preferences of the families (this is similar to the estimation strategy of JP(2017)). I also include birth year fixed effects to account for differences in the environment in which the children were born. For example,

[9] Sample Registration System Statistical Report (2013)

[10] Women who are more than 35 years old or haven't had children for the last 4 years are considered to have completed their fertility (See Appendix Figures C.1 and C.2).

children born in the years with good monsoon, might get more nutrition due to an increase in their family income, and thus, have higher cognitive ability. Lastly, the standard errors are clustered at the village level. The underlying assumption is that the unobservables in the regression are correlated at the village level. The preferred specification takes the following form:

$$\begin{aligned}StdMath_{imt} = \beta_1 SecondChild_{imt} + \beta_2 LaterChild_{imt} + \theta_1 ChildAge_{imt} + \\ \theta_2 SchoolClass_{imt} + \eta_m + \nu_t + \epsilon_{imt}\end{aligned} \tag{3.1}$$

where $StdMath_{imt}$ represents the standardized math outcomes of child i born to mother m in year t. The first child is the reference category. β_1 and β_2 are the main coefficients of interest with η_m and ν_t denoting the mother and birth-year fixed effect respectively. ϵ_{imt} represents the error term. Household-level variables drop from the regression due to no variation at the mother level.

Table 3.2 column 1 shows the results from regression controlling for only child-level characteristics, whereas column 2 also includes household-level characteristics. Columns 3 and 4 include village FEs, with column 4 also restricting the sample to mothers with completed fertility and controlling for the total number of children. Column 5 includes Mother FEs, and thus can only include households who have more than one child.

The results point to a negative birth order gradient among the Indian children with the second and later-born children having a learning deficit as compared to the firstborn. As mentioned before, the advantage to firstborns can come from factors like greater paternal attention, the biological condition of the mother, etc. If we focus on results from column 5, as it controls for all the time-invariant mother-level characteristics, then the second born children are 5.5 percentage points less likely to attain the math level they are supposed to have, as compared to the firstborns. The magnitude of the learning gap for the later-borns increases to 11.9 percentage points. Since 43 percent of the firstborns reach the desired level of learning, the birth order effect reduces the learning outcome by 12.8 percent and 27.7 percent for the second and the later-born as compared to the firstborn. Moreover, the birth order gradient becomes steeper after the inclusion of Mother FEs implying that the omitted variables in other specifications were biasing the coefficients upwards. Reassuringly, the other covariates have expected signs (Columns 1-4) . Mother's age and parent's education have positive and statistically significant effect on learning. Household infrastructure also has a positive and significant result. Consistent with the result obtained by Black et al (2005), the coefficient of family size is very small in magnitude after controlling for birth order. Thus, even in the context of an overall low learning environment, there is a steep birth order gradient.

3.5.2 Birth Order and Gender

Now, I test if being a girl has a negative impact on learning (Prediction 1), and if the birth order gradient varies across gender. To do this, I add a dummy for being a girl child and its interaction with the birth order dummies in the Equation(1) :

$$\begin{aligned} StdMath_{imt} = {} & \beta_1 SecondChild_{imt} + \beta_2 LaterChild_{imt} + \theta_1 ChildAge_{imt} + \\ & \theta_2 SchoolClass_{imt} + \gamma_1 Girl_{imt} + \gamma_2 (Girl_{imt} * SecondChild_{imt}) + \\ & \gamma_3 (Girl_{imt} * LaterChild_{imt}) + \eta_m + \nu_t + \epsilon_{imt} \end{aligned} \tag{3.2}$$

Here, γ_1 captures the difference in learning outcomes between boys and girl - absolute son preference. γ_2 and γ_3 capture any difference in the birth order gradient between boys and girls. Table 3.3 column 3 shows the coefficients of interest for the above specification. Across various specifications, there is an absolute disadvantage to girls as compared to boys, conditional on birth order. The dummy for the girl child is negative and significant. The magnitude of the coefficient is, in fact, comparable to the second child dummy. This validates prediction 1 - there is an absolute son preference in India which translates into lower learning outcome for girls as compared to boys, conditional on birth order. Figure 3.3 plots the relative learning outcomes of both boys and girls with respect to the first-born boy.

Moreover, allowing for birth order to vary by gender leads to steeper birth order gradient for boys, that is, the coefficients on birth order dummies are more negative in table 3.3 than table 3.2. A potential reason could be the elder son preference, over and above general son preference, which is investigated in the next section.

3.5.3 Birth Order, Gender and Elder son preference

In this section, I explore if there is heterogeneity in the birth-order gradient according to changes in the district-level elder son preference. Moreover, I test if the elder son preference translates into elder-son advantage in learning which leads to changes in birth-order gradient.

To test prediction 2, I modify Equation 2 to allow the gender-specific birth order gradient to vary with SRLB - measure of elder son preference. A triple interaction of birth order, gender and SRLB is included in the regression. The modified equation takes the following form:

$$\begin{aligned}
StdMath_{imt} = {} & \beta_1 SecondChild_{imt} + \beta_2 LaterChild_{imt} + \theta_1 ChildAge_{imt} + \\
& \theta_2 SchoolClass_{imt} + \gamma_1 Girl_{imt} + \gamma_2 (Girl_{imt} * SecondChild_{imt}) + \\
& \gamma_3 (Girl_{imt} * LaterChild_{imt}) + \tau_1 (Girl_{imt} * SRLB) + \\
& \tau_2 (SecondChild_{imt} * SRLB) + \tau_3 (LaterChild_{imt} * SRLB) + \\
& \tau_4 (SecondChild_{imt} * Girl_{imt} * SRLB) + \\
& \tau_5 (LaterChild_{imt} * Girl_{imt} * SRLB) + \eta_m + \nu_t + \epsilon_{imt}
\end{aligned} \tag{3.3}$$

The γ coefficients have the same interpretation as before. τ_1 estimates if the absolute son preference changes with SRLB. τ_2 and τ_3 capture the change in the birth order gradient of boys, while τ_4 and τ_5 capture the relative change in girls' birth order gradient with respect to boys as the SRLB changes. Moreover, if the SRLB is 1, then the birth order gradient between the second and first boy is given by $\beta_1 + \tau_2$. The change in gradient between the second and the first born girl as compared to the similar gradient between boys, when SRLB is equal to 1, is given by $\gamma_2 + \tau_4$.

First, we would expect absolute discrimination against girls to increase with SRLB because stronger elder son preference also implies stronger general son preference. Table 3.4 shows that the interaction term between the girl dummy and SRLB (τ_1) is negative and significant across specifications. The first three rows of Appendix Table C.1 calculate the magnitude of the girl dummy at different levels of SRLB, and indeed the discrimination against girls increases as the elder son preference becomes stronger.[11]

Second, the effect of SRLB on the birth order gradient of boys is negative and significant - given by the coefficients for the interaction between the birth order dummies and the SRLB (τ_2 and τ_3) in Table 3.4 column 3 (the most preferred specification). The learning disadvantage for the second born as compared to the firstborn increases from 5 to 8 percentage points as SRLB increases from 1 to 1.88 boys per girl. For the later-borns, the disadvantage increases from 10 to 15 percentage points for a similar rise in SRLB (see Appendix Table C.1). Moreover, the relative difference in the birth order gradient between boys and girls is insignificant (the last six rows of Appendix Table C.1), except in column 3 where the birth order gradient between first and second girls is slightly flatter than boys. The more negative slope of birth order gradient for girls with an increase in SRLB cannot be explained by stronger elder son preference, which is a limitation of this paper. Figure 3.4 plots the relative learning outcomes of boys and girls with respect to the first born boy,

[11] I calculate the size of the coefficients at three points of the district-level SRLB – gender parity (SRLB=1 boy per girl), median level (SRLB = 1.36 boys per girl), and 90th percentile (SRLB = 1.88 boys per girl).

at different levels of SRLB.

Next, to establish the presence of eldest-son advantage in the learning, I construct a dummy for no-elder-brother (as done by JP (2017)). This takes a value of 1 if the boy is first born or has only female siblings and 0 otherwise. The no-elder-brother dummy is also interacted with the gender dummy to isolate the heterogeneous impact of having no elder brother on the two genders. I also control for the birth order effects and use mother Fixed effects to control for total fertility in the regression. The estimation equation takes the following form:

$$\begin{aligned} StdMath_{imt} = {} & \beta_1 SecondChild_{imt} + \beta_2 LaterChild_{imt} + \theta_1 ChildAge_{imt} + \\ & \theta_2 SchoolClass_{imt} + \gamma_1 Girl_{imt} + \delta_1 NoElderBro_{imt} + \\ & \delta_2 (Girl_{imt} * NoElderBro_{imt}) + \eta_m + \nu_t + \epsilon_{imt} \end{aligned} \tag{3.4}$$

Here the coefficient of interest is δ_1, which gives the magnitude of no-elder-brother effect for boys. For girls, the magnitude is given by $\delta_1 + \delta_2$. While we expect δ_1 to be positive, the sign of $\delta_1 + \delta_2$ is ambiguous due to the presence of two opposing effects - Sibling rivalry effect and fertility stopping rule (JP(2017)).

Across all the specifications in Table 3.5, the coefficient for no-elder-brother is positive and significant. The eldest son has 2.1 percentage point learning advantage over a boy who does have an elder brother (Table 3.5 column 3). For girls, the magnitude of the no-elder-brother effect is small and insignificant. Thus, the sibling rivalry and the fertility stopping rule seem to be equally strong for the entire sample.

Finally, I investigate if the no-elder-brother effect increases in areas with stronger son preference to explain the steeper birth order gradient. I modify equation 4 to include the interaction of no-elder-brother dummy with SRLB. The modified equation is:

$$\begin{aligned} StdMath_{imt} = {} & \beta_1 SecondChild_{imt} + \beta_2 LaterChild_{imt} + \theta_1 ChildAge_{imt} + \\ & \theta_2 SchoolClass_{imt} + \gamma_1 Girl_{imt} + \delta_1 NoElderBro_{imt} + \\ & \delta_2 (Girl_{imt} * NoElderBro_{imt}) + \lambda_1 (NoElderBro_{imt} * SRLB) + \\ & \lambda_2 (Girl_{imt} * NoElderBro_{imt} * SRLB) + \eta_m + \nu_t + \epsilon_{imt} \end{aligned} \tag{3.5}$$

According to Prediction 3, the sum of the coefficients $\delta_1 + \lambda_1$ should be positive in areas with high SRLB, and λ_1 itself should be greater than 0. Table 3.6, columns 1 to 3 show that the coefficient for the interaction term of no-elder-brother and SRLB is positive and significant across specifications. The bottom panel calculates the total no-elder-brother effect for boys $(\delta_1 + \lambda_1)$ at different levels of SRLB. At SRLB equal to 1 (gender parity)

- there is zero no-elder-brother effect, but at the 90th percentile of SRLB (strong son-preference) the magnitude of the effect increases to 3.8 percentage points.

Moreover, the effect of having an elder brother is muted for girls (see figure 3.5). The total no-elder-brother effect for girls in equation 5, is given by $\delta_1 + \delta_2 + \lambda_1 + \lambda_2$. Bottom panel for table 3.6 column 3, shows the sum of the coefficients to be insignificant at gender parity and the median value of SRLB. It is positive and significant at 90th percentile with the magnitude smaller than boys. For the rest of the specifications (columns 1 and 2), the coefficient is not statistically different from 0 at all levels of SRLB. The absence of no-elder-brother effect can be explained by the combined and equal effect of sibling rivalry and fertility stopping rule (Prediction 4).

3.6 Mechanisms and underlying assumptions

In this section, I identify potential mechanisms through which gender preferences influence the learning outcomes. JP (2017) show that "girls in India receive fewer postnatal resources if their family does not yet have a son." Lower investments in early childhood could lead to lower cognition later in life. Another potential channel can be differential investments directly into the education of the child. I measure educational investments in 2 ways - the probability of giving child out-of-school tuition by a private tutor, and the probability of enrolling the child in a private school. In India private schools tend to be more expensive than the public schools. Härmä (2009) estimates the full cost to parents of sending their kids to private schools in the range of \$22.42-26.44 per annum as compared to a cost of \$3 per annum in government schools.

Empirically, I test if the no-elder-brother effect exists in the probability of accessing private tuition or getting enrolled in a private school. I modify Equation 4 by having these two measures of school investment as the dependent variable. The results are reported in Table 3.7.[12]

Girls have a lower probability of getting enrolled for both private tuition and schooling. Moreover, the birth order gradient is downward sloping for educational investments - the coefficients of the birth order dummies are negative (except column 3 where they are insignificant). Finally, there is a positive and significant no-elder-brother effect for boys similar to the one in learning outcomes.[13] For girls, the effect of having an elder brother versus not ranges from null to negative. It seems that the negative effect of fertility stop-

[12] In the first three columns, the dependent variable is a dummy for the child getting private tuition and in the next three, it is a dummy for enrollment in private school.

[13] The no-elder-brother dummy is insignificant in column 6, after the inclusion of Mother fixed effects. However, it does become significant at the top decile of SRLB, which is not in the table.

ping rule dominates the negative effect of sibling rivalry effect for educational investment in girls, but the evidence is not conclusive.

Finally, the Sex Ratio at Last Birth (SRLB) indeed represents elder son preference if families exceed their desired fertility to have at least one son. This assumption can be tested by asking if the families whose two oldest children are girls are larger than the ones who have at least one son. I restrict the sample to households who have at least 2 children[14], and then create a dummy which takes a value of 1 if the oldest two children are both girls and 0 otherwise. I also interact the dummy with SRLB, to check if the desire of having at least one son is increasing with rising SRLB. The following regression in run at the household-level:

$$\begin{aligned} TotalChildren_{mv} = \beta_1 BothGirls_{mv} + \tau_1(BothGirls_{mv} * SRLB) + \\ \mathbf{X}'_{mv}\delta + \eta_v + \epsilon_{mv} \end{aligned} \tag{3.6}$$

$TotalChildren_{mv}$ represents the total number of children born to mother m in village v. $\mathbf{X}_{mv}$ represents vector of mother/household level controls and η_v indicates village FEs. We expect $\beta_1 + \tau_1$ to be jointly positive and τ_1 to be individually positive.

Table 3.8 documents the results. Column 2 includes village Fixed effects while Column 3 also restricts the sample to women who might have completed their fertility. The household size is significantly larger if the first two children are girls as compared to having at least one boy among the first two children. Moreover, the interaction term is positive and significant which implies that the family size becomes larger with higher SRLB.

3.7 Robustness

3.7.1 Alternate Specification

The dependent variable used is a dummy which takes a value of 1 if the child has reached the requisite level of learning according to the grade in which she is studying. However, the highest math ability captured in the ASER survey is doing division which every child above grade 3 should be able to do. As a result, for any child above grade 3, the probability of learning outcome dummy getting a value of 1 increases mechanically with the grade of the child. I control for that by including the child age and school grade as explanatory variables. To alleviate any concern, I also test the main specifications of

[14]The total fertility rate for India is 2.3 according to Sample Registration System Report (2013)

the analysis, by taking the raw scores in the data[15], and flexibly controlling for the school grade by introducing the school-grade fixed effects. This approach has been used by Shah and Steinberg (2017) who use ASER data to look at the impact of rainfall on human capital formation.

The results are reported in Appendix table C.2 which replicate the main specifications. The results are consistent with the primary analysis, although the coefficients are not comparable as raw scores instead of a dummy variable are used as the dependent variable. Girls have lower learning outcomes than boys, and there is a negative birth order gradient. The no-elder-brother dummy is positive and significant indicating the advantage of being the eldest son in the family. For girls, the sibling rivalry effect and fertility stopping rule effect tends to cancel out resulting in statistically insignificant no-elder-brother effect for girls.

3.7.2 Using Reading Scores

ASER survey also documents the reading skills of children which can be an alternate way to define the learning outcomes. I create a dummy for standardized learning outcome for reading, which takes a value of 1 if the child has attained the level which she should have according to the grade in which she is studying.[16]I replicate the main specifications by using the standardized reading level as the dependent variable. Table C.3 in the appendix documents the regression results. There is a birth order gradient in learning outcomes with the oldest child having the highest learning outcomes. In line with previous results, the no-elder-brother effect is positive and significant for boys. For girls, it is ambiguous and statistically similar to zero. The sign of the girl dummy is positive which is a bit surprising, but this advantage attenuates in areas with higher SRLB. Thus, the effect of birth order and sibling composition on learning outcomes are invariant to the dimension on which these outcomes are measured.

3.8 Conclusion and Policy Prescription

In conclusion, there is a declining birth order gradient in learning outcomes with the older children having learning advantage over their younger siblings. Gender and sibling composition also affect learning outcomes, which varies according to the change in the

[15]Level 1 = No Arithmetic, 2 = Recognize numbers 1-9; 3 = Recognize numbers 11-99; 4 = two-digit subtraction; 5 = division

[16]Original reading levels in the data are: level 1 = cannot read anything; level 2 = identify letters; level 3 = can read words; level 4 = can read Standard 1 text; level 5 = can read Standard 2 level text.

underlying preference for elder sons. For boys, having no elder brother improves their learning outcomes, and this advantage rises with an increase in district-level sex ratio at last birth (SRLB) – a proxy for elder son preference. It also results in steeper birth-order gradient for boys. The no-elder-brother effect is muted for girls due to two opposing effects – fertility stopping rule and sibling rivalry effect. Preferences affect learning through differential investment in schooling with the eldest son more likely to go to private school and get private tuition.

There are some policy implications of this analysis. This study proves that the assumption of common gender bias across all children in the household is flawed. In fact, even boys can be disadvantaged within a family if they have an elder brother. Such taste-based discrimination cannot be mitigated by policies like wage equality laws. Therefore, data can be collected on observables like sibling composition and birth order to specifically target disadvantaged children. For instance, private schools in Delhi (capital of India) collect data on birth order and siblings as part of their admission process. Such data can also be collected while implementing schemes that provide remedial classes to disadvantaged children.

There is evidence of son preference in developed countries like USA (Dahl and Moretti (2008)). Future research should investigate if the patterns of learning that emerge in developing countries, also exist in more developed countries. Moreover, gender preferences have been found to be persistent even in the presence of positive economic shocks. Almond, Edlund and Milligan (2013) have documented the presence of strong son-preference among the immigrants from South and East Asia to Canada, even though they tend to be much richer than the average inhabitants of their native countries. Do gender preferences affects the learning outcomes of the children in these immigrant populations? – is an open question. Thus, it is possible that the effect of intra-household discrimination on learning crisis, that is quantified in this study, may not be mitigated by the economic growth witnessed by the developing world.

3.9 Tables

Table 3.1: Summary Statistics

	Count	Mean	SD	Min	Max
School Class	259109	5.20	2.98	1	12
Standardized Math level	259109	0.38	0.49	0	1
Total No. of Children	339461	2.26	1.06	1	8
Girl	339461	0.48	0.50	0	1
Child Age	339461	9.27	3.87	3	16
Dummy of First Child	339461	0.55	0.50	0	1
Dummy of Second Child	339461	0.30	0.46	0	1
Dummy of Later Child	339461	0.14	0.35	0	1
Dummy of Girl x Second Child	339461	0.14	0.35	0	1
Dummy of Girl x Later Child	339461	0.07	0.25	0	1
Dummy of No Elder Brother	339461	0.76	0.43	0	1
Dummy of Girl x No Elder Brother	339461	0.37	0.48	0	1
Dummy of Mother School	339461	0.58	0.49	0	1
Mother Age	339461	33.06	6.89	17	80
Father Age	339461	38.22	7.73	17	85
Dummy of Father School	339461	0.77	0.42	0	1
HH Infrastructure Index	339461	0.00	1.35	-2.38	1.39
Dummy of Private Tuition	274960	0.22	0.41	0	1
Dummy of Private School	287145	0.33	0.47	0	1
Sex Ratio of Last Birth(District level)	577	1.41	0.34	0.73	2.79

[Source] ASER 2014

Table 3.2: Effect of Birth Order

Dependent Variable (Math Level)	(1) OLS	(2) OLS	(3) Village FE	(4) Village FE	(5) Mother FE
	Whole Sample	Whole Sample	Whole Sample	Completed fertility	# of children>1
Second Child	-0.0441***	-0.0383***	-0.0347***	-0.0272***	-0.0548***
	(0.00217)	(0.00217)	(0.00204)	(0.00259)	(0.00356)
Later Child	-0.120***	-0.0866***	-0.0829***	-0.0732***	-0.119***
	(0.00352)	(0.00345)	(0.00315)	(0.00441)	(0.00651)
Child Age	0.0231***	0.0268***	0.0253***	0.0197***	0.0228***
	(0.00108)	(0.00105)	(0.000944)	(0.00122)	(0.00158)
School Class	0.0153***	0.00579***	0.00958***	0.0155***	-0.00186
	(0.00111)	(0.00107)	(0.000946)	(0.00105)	(0.00140)
Dummy Mother School		0.103***	0.0958***	0.0976***	
		(0.00290)	(0.00263)	(0.00310)	
Mother Age		0.00614***	0.00149***	0.000628*	
		(0.000342)	(0.000301)	(0.000340)	
Father Age		-0.00198***	0.000209	-0.000381	
		(0.000297)	(0.000267)	(0.000304)	
Dummy Father School		0.0775***	0.0626***	0.0659***	
		(0.00302)	(0.00270)	(0.00316)	
HH Infra		0.0529***	0.0459***	0.0449***	
		(0.00121)	(0.00115)	(0.00136)	
Total No. of Children				-0.00716***	
				(0.00150)	
Constant	0.0686***	-0.170***	-0.0853***	0.0197	0.197***
	(0.00764)	(0.00949)	(0.00932)	(0.0161)	(0.0150)
Hh Controls	NO	YES	YES	YES	NO
Observations	259,109	259,109	259,109	198,963	202,793

*** p<0.01, ** p<0.05, * p<0.1
Robust standard errors (clustered at village level) in parentheses. Average learning outcome of first born child (omitted category) in the data is 0.43. The coefficients should be interpreted as change in the probability of attaining the minimum level of learning. Column 4 restricts the sample to mothers who might have completed their fertility and thus, allows to control for total number of children. Column 5 uses mother Fixed effects to control for mother-level time invariant factors. All the regressions also include birth year Fixed effects.
[Source] All the data for this table is derived from ASER-2014.

Table 3.3: Effect of Birth Order and Gender

Dependent Variable (Math Level)	(1) Village FE	(2) Village FE	(3) Mother FE
	Whole Sample	Completed fertility	# of children>1
Girl	-0.0467***	-0.0490***	-0.0551***
	(0.00242)	(0.00282)	(0.00389)
Second Child	-0.0369***	-0.0330***	-0.0657***
	(0.00277)	(0.00332)	(0.00443)
Later Child	-0.0828***	-0.0789***	-0.127***
	(0.00401)	(0.00527)	(0.00726)
Girl x Second Child	0.00273	0.00499	0.0198***
	(0.00383)	(0.00442)	(0.00520)
Girl x Later Child	-0.00421	-0.00452	0.00847
	(0.00507)	(0.00597)	(0.00634)
Constant	-0.0587***	0.0368**	0.225***
	(0.00940)	(0.0161)	(0.0151)
Hh Controls	YES	YES	NO
Observations	259,109	198,963	202,793

*** $p<0.01$, ** $p<0.05$, * $p<0.1$
Robust standard errors (clustered at village level) in parentheses. Average learning outcome of first born male child (omitted category) in the data is 0.46. The coefficients should be interpreted as change in the probability of attaining the minimum level of learning. Column (2) restricts the sample to mothers who might have completed their fertility and thus, allows to control for the total number of children. All the regressions also include birth year Fixed effects. Child level controls include the age and the class in which she is studying. Household level controls include the parents' education, parents' age and index for household infrastructure.
[Source] All the data for this table is derived from ASER-2014.

Table 3.4: Heterogeneity in effect of Birth Order and Gender

Dependent Variable (Math Level)	(1) Village FE	(2) Village FE	(3) Mother FE
	Whole Sample	Completed Fertility	# of children>1
Girl	-0.0154	-0.0156	-0.0241
	(0.0105)	(0.0122)	(0.0177)
Second Child	-0.0222*	-0.0163	-0.0140
	(0.0119)	(0.0132)	(0.0159)
Later Child	-0.0376**	-0.0316	-0.0361
	(0.0185)	(0.0216)	(0.0242)
Girl x Second Child	0.0115	0.00459	0.00113
	(0.0173)	(0.0197)	(0.0240)
Girl x Later Child	0.0239	0.0280	-0.00506
	(0.0256)	(0.0299)	(0.0314)
Girl x Dist. SRLB	-0.0219***	-0.0234***	-0.0217*
	(0.00720)	(0.00841)	(0.0120)
Second child x Dist. SRLB	-0.0103	-0.0118	-0.0353***
	(0.00797)	(0.00880)	(0.0104)
Later child x Dist. SRLB	-0.0313**	-0.0331**	-0.0620***
	(0.0125)	(0.0144)	(0.0159)
Girl x Second child x Dist. SRLB	-0.00633	.00006	0.0125
	(0.0118)	(0.0135)	(0.0163)
Girl x Later child x Dist. SRLB	-0.0199	-0.0233	0.00889
	(0.0175)	(0.0205)	(0.0213)
Constant	-0.0585***	0.0356**	0.222***
	(0.00940)	(0.0161)	(0.0151)
Hh Controls	YES	YES	NO
Observations	259,109	198,963	202,793

*** p<0.01, ** p<0.05, * p<0.1
Robust standard errors (clustered at village level) in parentheses. Sex Ratio at Last Birth (SRLB) is number of boys per girl in a district, conditional on being the last born. This ratio is calculated using families where mothers might have completed their fertility. Average learning outcome of first born male child (omitted category) in the data is 0.46. The coefficients should be interpreted as change in the probability of attaining the minimum level of learning. Column (2) restricts the sample to mothers who might have completed their fertility and thus, allows to control for total number of children. All the regressions also include birth year Fixed effects. Child level controls include the age and the class in which she is studying. Household level controls include the parents' education, parents' age and index for household infrastructure.
[Source] All the data for this table is derived from ASER-2014.

Table 3.5: Effect of “No Elder Brother” on Learning Outcomes

Dependent Variable (Math Level)	(1) Village FE	(2) Village FE	(3) Mother FE
(Robust SE in parentheses)	Whole Sample	Completed Fertility	# of children>1
Girl	-0.0351***	-0.0357***	-0.0313***
	(0.00348)	(0.00400)	(0.00406)
Second Child	-0.0330***	-0.0257***	-0.0490***
	(0.00242)	(0.00304)	(0.00422)
Later Child	-0.0808***	-0.0736***	-0.112***
	(0.00369)	(0.00503)	(0.00738)
No Elder Brother	0.0127***	0.0174***	0.0208***
	(0.00323)	(0.00367)	(0.00422)
Girl x No Elder Brother	-0.0144***	-0.0154***	-0.0119**
	(0.00395)	(0.00458)	(0.00573)
Constant	-0.0697***	0.0219	0.200***
	(0.00990)	(0.0164)	(0.0157)
(p-value in parentheses)			
NEB + Girl x NEB	-0.002	0.002	0.009
	(0.630)	(0.642)	(0.202)
Hh Controls	YES	YES	NO
Observations	259,109	198,963	202,793

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Robust standard errors are clustered at the village level. NEB stands for No Elder Brother dummy. Average math level in the data is 0.38. The coefficients should be interpreted as change in the probability of attaining the minimum level of learning. Column (2) restricts sample to the mothers who might have completed their fertility and thus, allows to control for total number of children. All the regressions also include birth year Fixed effects. Child level controls include age and the class in which she is studying. Household level controls include the parents' education, parents' age and index for household infrastructure.

[Source] All the data for this table is derived from ASER-2014.

Table 3.6: Heterogeneity in the Effect of "No Elder Brother" on Learning Outcomes according to District-level Sex Ratio at Last Birth (SRLB)

Dependent Variable (Math Level)	(1) Village FE	(2) Village FE	(3) Mother FE
(Robust SE in parentheses)	Whole Sample	Completed Fertility	# of children>1
Girl	-0.0355***	-0.0363***	-0.0321***
	(0.00348)	(0.00401)	(0.00406)
Second Child	-0.0331***	-0.0261***	-0.0486***
	(0.00242)	(0.00305)	(0.00421)
Later Child	-0.0810***	-0.0743***	-0.111***
	(0.00369)	(0.00503)	(0.00738)
No Elder Brother	-0.0305***	-0.0319***	-0.0371***
	(0.0108)	(0.0122)	(0.0130)
Girl x No Elder Brother	0.0247**	0.0264**	-0.0250
	(0.00991)	(0.0115)	(0.0157)
No Elder Brother x Dist. SRLB	0.0299***	0.0340***	0.0400***
	(0.00713)	(0.00803)	(0.00841)
Girl x No Elder Brother x Dist. SRLB	-0.0271***	-0.0290***	0.00938
	(0.00635)	(0.00745)	(0.00999)
Constant	-0.0689***	0.0221	0.199***
	(0.00990)	(0.0164)	(0.0157)
(p-value in parentheses)			
NEB + NEB x SRLB	-0.001	0.002	0.003
	(0.890)	(0.676)	(0.609)
NEB + NEB x SRLB x 1.36	0.010	0.014	0.017
	(0.002)	(0.000)	(0.000)
NEB + NEB x SRLB x 1.88	0.026	0.032	0.038
	(0.000)	(0.000)	(0.000)
NEB + NEB x SRLB + Girl x NEB + Girl x NEB x SRLB	-0.003	0.000	-0.013
	(0.521)	(0.930)	(0.162)
NEB + NEB x SRLB x 1.36 + Girl x NEB + Girl x NEB x SRLB x 1.36	-0.002	0.001	0.005
	(0.559)	(0.762)	(0.469)
NEB + NEB x SRLB x 1.88 + Girl x NEB + Girl x NEB x SRLB x 1.88	-0.001	0.004	0.031
	(0.896)	(0.529)	(0.000)
Hh Controls	YES	YES	NO
Observations	259,109	198,963	202,793

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Robust standard errors are clustered at the village level. NEB and SRLB stand for No Elder Brother dummy and Sex Ratio at Last Birth which is defined as number of boys per girl in a district, conditional on being the last born. This ratio is calculated using families where mothers might have completed their fertility. 1.36 is median value and 1.88 is top decile value of SRLB across districts. Average math level in the data is 0.38. The coefficients should be interpreted as change in the probability of attaining the minimum level of learning. Column (2) restrict sample to the mothers who might have completed their fertility and thus, allows to control for total number of children. All the regressions also include birth year Fixed effects. Child level controls include age and the class in which she is studying. Household level controls include the parents' education, parents' age and index for household infrastructure.

[Source] All the data for this table is derived from ASER-2014.

Table 3.7: Potential Mechanisms

Dependent Variable	(1) Pvt Tuition	(2) Pvt Tuition	(3) Pvt Tuition	(4) Pvt School	(5) Pvt School	(6) Pvt School
	(Village FE)	(Village FE)	(Mother FE)	(Village FE)	(Village FE)	(Mother FE)
(Robust SE in parentheses)	Whole Sample	Completed Fertility	# of children>1	Whole Sample	Completed Fertility	# of children>1
Girl	-0.0291***	-0.0279***	-0.0250***	-0.0472***	-0.0433***	-0.0479***
	(0.00282)	(0.00319)	(0.00251)	(0.00311)	(0.00349)	(0.00281)
Second Child	-0.0230***	-0.0162***	-0.00200	-0.0395***	-0.0230***	-0.0321***
	(0.00187)	(0.00233)	(0.00250)	(0.00203)	(0.00254)	(0.00299)
Later Child	-0.0499***	-0.0316***	0.000266	-0.0826***	-0.0494***	-0.0589***
	(0.00317)	(0.00398)	(0.00455)	(0.00359)	(0.00447)	(0.00526)
No Elder Brother	0.00609**	0.00691**	0.0142***	0.0102***	0.0147***	0.00426
	(0.00257)	(0.00289)	(0.00254)	(0.00279)	(0.00316)	(0.00285)
Girl x No Elder Brother	-0.0146***	-0.0166***	-0.00987***	-0.0159***	-0.0159***	-0.0113***
	(0.00324)	(0.00372)	(0.00348)	(0.00354)	(0.00399)	(0.00399)
Constant	0.141***	0.186***	0.135***	0.170***	0.232***	0.358***
	(0.00866)	(0.0145)	(0.0103)	(0.0103)	(0.0164)	(0.0120)
(p-value in parentheses)						
NEB + Girl x NEB	-0.008	-0.010	0.004	-0.006	-0.001	-0.007
	(0.009)	(0.011)	(0.304)	(0.104)	(0.765)	(0.151)
Hh Controls	YES	YES	NO	YES	YES	NO
Observations	245,103	188,700	191,436	258,505	198,557	202,327

*** p<0.01, ** p<0.05, * p<0.1

Robust standard errors are clustered at the village level. NEB stands for No Elder Brother dummy. Mean of the dummy for attending private tuition and private school is 0.22 and 0.33 respectively. The coefficients should be interpreted as change in the probability of attending private tuition or private school. Columns 2 and 5 restrict sample to the mothers who might have completed their fertility and thus, allows to control for total number of children. All the regressions also include birth year Fixed effects. Child level controls include age and the class in which she is studying. Household level controls include the parents' education, parents' age and index for household infrastructure.

[Source] All the data for this table is derived from ASER-2014.

Table 3.8: Test of Fertility Stopping Rule at Family Level

Dependent variable (Total No. of Children)	(1) OLS	(2) Village FE	(3) Village FE
(Robust SE in parentheses)	No. of children ≥ 2	No. of children ≥ 2	No. of children ≥ 2 & Completed Fertility
Dummy for Both Girls	-0.155***	-0.156***	-0.236***
	(0.0240)	(0.0252)	(0.0337)
Dummy for Both Girls x SRLB	0.312***	0.317***	0.384***
	(0.0168)	(0.0176)	(0.0238)
Mother Age	0.0111***	0.00511***	0.00325***
	(0.000702)	(0.000761)	(0.000958)
Father Age	-0.000979	0.00355***	-0.000634
	(0.000616)	(0.000659)	(0.000838)
Dummy Mother School	-0.157***	-0.0929***	-0.0654***
	(0.00583)	(0.00626)	(0.00800)
Dummy Father School	-0.0121*	-0.0167**	-0.00887
	(0.00689)	(0.00722)	(0.00911)
HH Infra	-0.0741***	-0.0376***	-0.0449***
	(0.00218)	(0.00274)	(0.00363)
Constant	2.168***	2.157***	2.348***
	(0.0150)	(0.0167)	(0.0242)
(p-value in parentheses)			
DBG + DBG x SRLB	0.157	0.161	0.148
	(0.000)	(0.000)	(0.000)
DBG + DBG x SRLB x 1.36	0.269	0.275	0.286
	(0.000)	(0.000)	(0.000)
DBG + DBG x SRLB x 1.88	0.433	0.442	0.487
	(0.000)	(0.000)	(0.000)
Hh Controls	YES	YES	YES
Village FE	NO	YES	YES
Observations	103,337	103,337	66,344

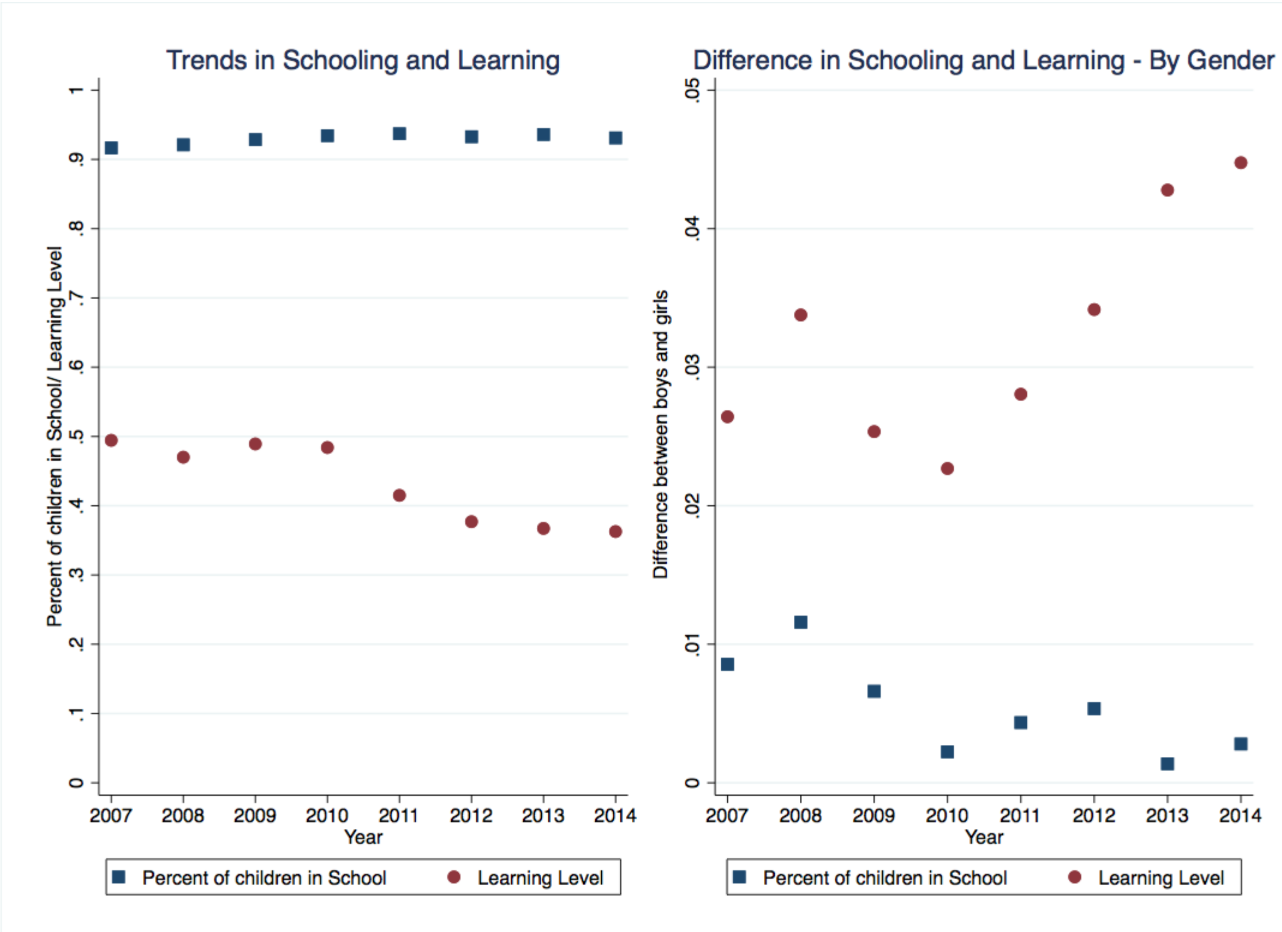
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Robust standard errors are clustered at the village level. DBG stands for Dummy for Both Girls . It takes a value of 1 if the two oldest children in the family are girls and 0 otherwise. SRLB stands for Sex Ratio at Last Birth which is defined as the number of boys per girl in a district, conditional on being the last born. 1.36 is the median value and 1.88 is the top decile value of SRLB across districts. Each observation is one family. Sample is restricted to households who have at least 2 children. Column 3 further restricts the sample to women who might have completed fertility i.e. above 35 years of age or not had children for more than 4 years.

[Source] All the data for this table is derived from ASER-2014.

3.10 Figures

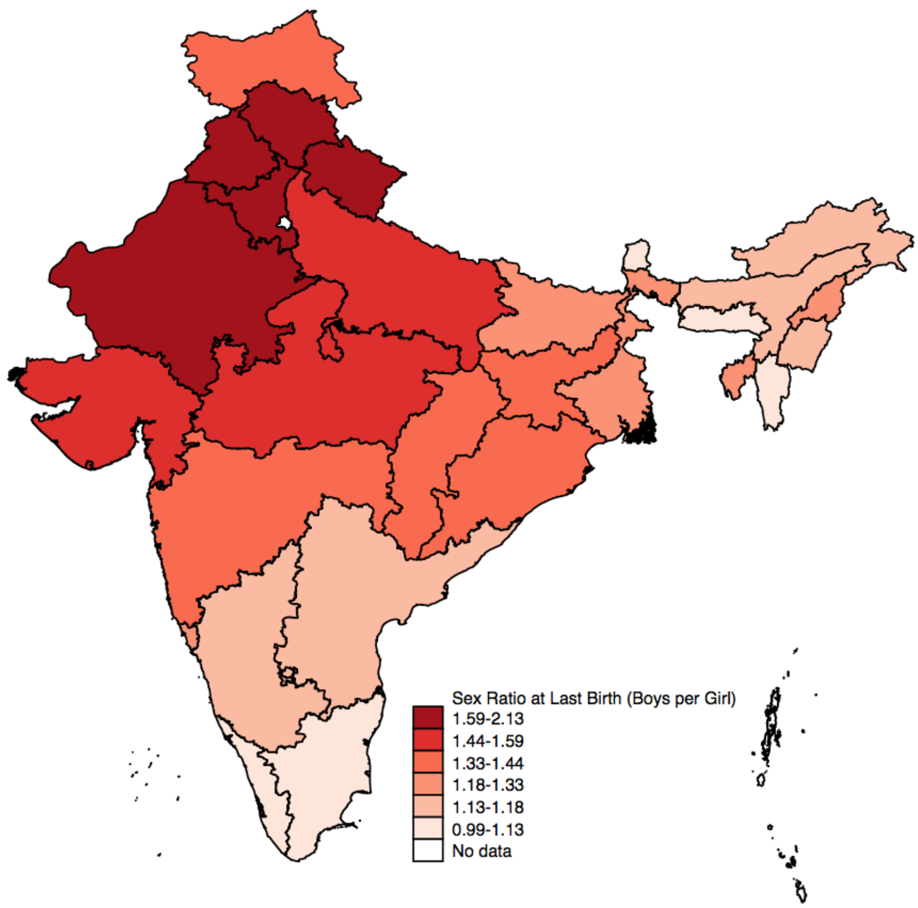
Figure 3.1: Trends in School Enrollment and Learning



The graph on the left panel depicts the percent of children going to school and their standardized learning outcomes. The learning outcome variable is a dummy which takes a value of 1 if the child has attained the learning according to the grade in which she is studying. Right panel depicts the difference between boys and girls in terms of enrollment and learning outcomes over the years.

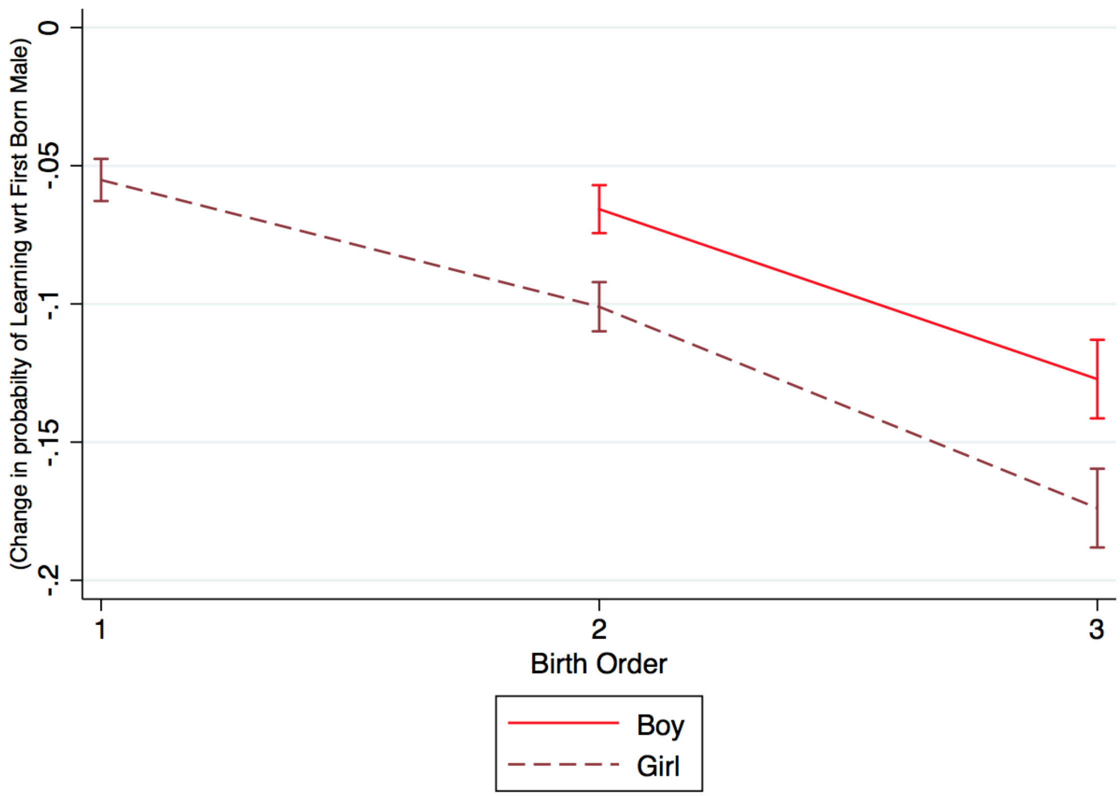
[Source] ASER reports 2007 to 2014. Based on author's calculations.

Figure 3.2: Average district-level SRLB across Indian states.



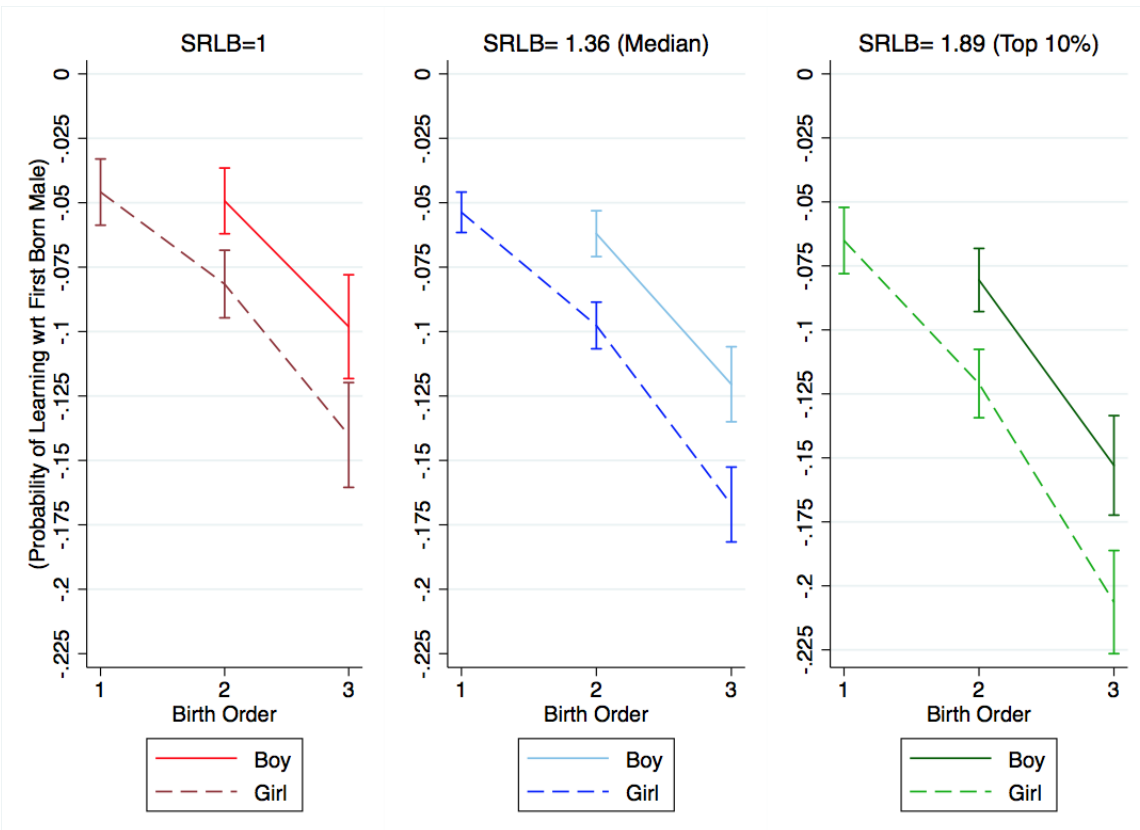
Sex Ratio at Last Birth (SRLB) is defined as the number of last born boys per girl at the district level. The map shows the average of district-level SRLB for each state. SRLB is calculated using ASER data.
[Source] ASER 2014.

Figure 3.3: Gender-wise birth order gradient of learning outcomes



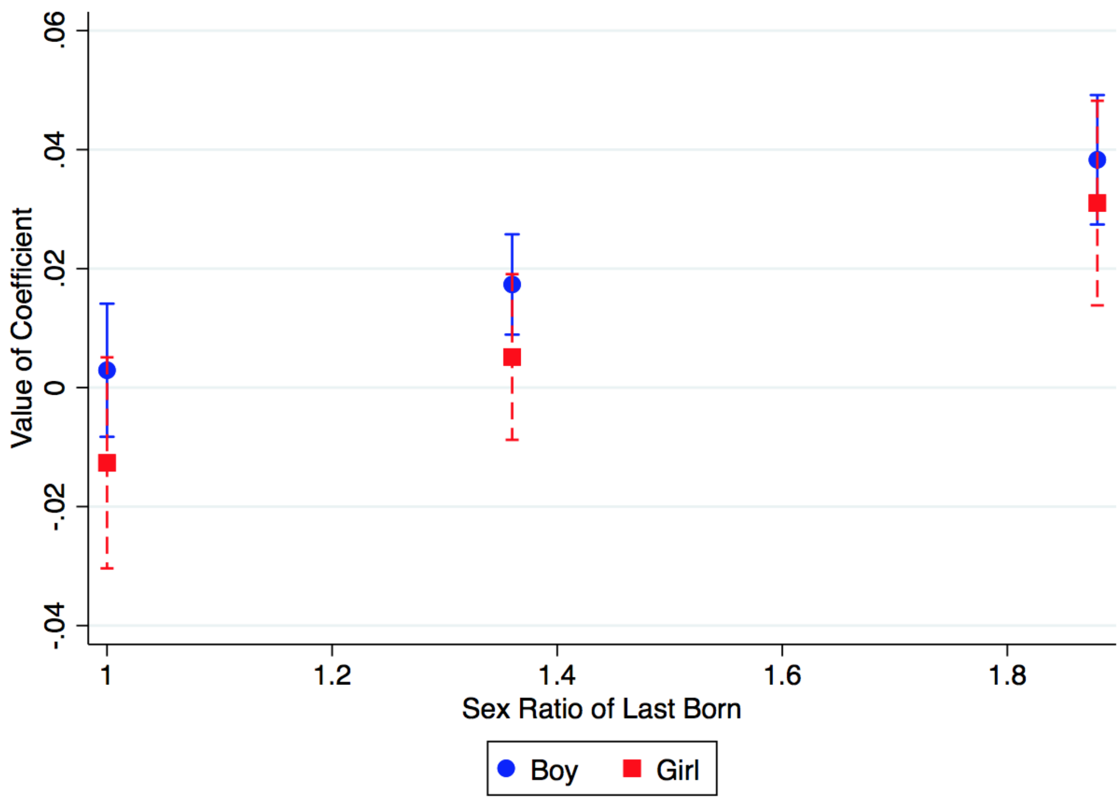
The graph plots the coefficients and their 95 percent confidence interval from Table 3.3, column 3. The omitted category is the first born boy in the family. The first born is given birth order value of 1, second-born a value of 2, and all the other children a value of 3.
[Source] Data for the analysis comes from ASER 2014.

Figure 3.4: Birth Order effect and son-preferences



The Sex Ratio at Last Birth(SRLB) is defined as the number of last-born boys per girl at the district level. Higher SRLB implies stronger elder son preference. The graph uses coefficients from column 3 of Table 3.4 and Appendix Table C.1. The left panel depicts the birth order gradient at the gender-parity level of SRLB. The right-most panel shows the birth order gradient for the districts with SRLB equal to highest decile value. [Source] Data for the analysis comes from ASER 2014.

Figure 3.5: No-Elder-Brother effect and son-preferences



The Sex Ratio at Last Birth(SRLB) is defined as the number of last-born boys per girl at the district level. Higher SRLB implies stronger elder son preference. The graph plots coefficients and confidence intervals from Table 3.6, column 3.

[Source] Data for the analysis comes from ASER 2014.

APPENDIX A

Chapter I Supporting Material

A.1 Appendix

A.1.1 Differences between private and public company.

In this study, we label private firms as restricted share transfer (RST) firms and public firms as non-RST firms.The Company Act (2013) defines a public company as i)Not a private company and ii) has a minimum paid-up share capital of Rs.500K. Private company differs from a public company as they have restrictions on raising capital by selling shares. Section 2(68) requires private companies to restrict the sale of shares under Articles of Association. On the other hand, Section 58(2) provides that the securities or other interest of any member in a public company shall be 'freely transferable', unless there is a sufficient cause. Second, a private company must have at least 2 shareholders, while a public company must have at least 7 shareholders. Additionally, a private company must have at least 2 directors whereas a public company must have at least 3 directors. There are no restrictions on the managerial remunerations of a private company whereas they are capped at 11 percent of the net profit for the public company. These differences imply that managers of public company face a potential moral hazard where they have to split the gains of tax evasion with more shareholders but incur the same penalty as that of private company's manager if they get caught.

If a public company wants to raise capital by selling its shares, then it has to issue a prospectus. Among other things, it has to provide information on "reports by the auditors of the company with respect to its profits and losses and assets and liabilities.." and "any litigation or legal action pending or taken by a Government Department or a statutory body during the last five years immediately preceding the year of the issue of prospectus against the promoter of the company".

In our context, a public company has incentive to report honestly to the tax-authorities even when it doesn't undergo third-party tax audit because it might have to issue a prospectus at some point in the future. The Chartered Accountants also have an incentive to perform statutory-audit more rigorously as their reports will become public once the company decides to raise capital in a stock exchange.

A.1.2 Model of evasion with dynamic considerations.

In this appendix, we present a stylized model which develops the intuition that firms misreport their income in a staggered way if they believe that their chances of getting caught increases when they report zero-growth to the tax-authorities.

Static case - Consider a two-period model where a firm has to choose reported income, $\bar{y}_1$ and $\bar{y}_2$ at period $t = 1$ and $t = 2$. For analytical simplicity, assume that firm's true income y doesn't change in both the periods. Thus, the under-reporting in time periods 1 and 2 is given by $u_1 = y - \bar{y}_1$ and $u_2 = y - \bar{y}_2$, respectively.

Let τ be the tax rate on reported profit. There is a strictly increasing, continuous and convex resource cost of under-reporting given by $k(u)$. The probability of the firm getting caught, in any period, is given by $\delta = \phi h(u)$, where ϕ is the effective audit intensity faced by the firm. $h(.)$ is increasing and convex in u. If the firm get caught, it faces a penalty rate of θ on evaded taxes.

In the static model, the firm maximizes identical objective function in both the periods : $E[\pi] = (1-\tau)[y-u] - k(u) + u - \phi h(u)[\tau u + \theta\tau u]$. The FOC characterizing the optimal level of under-reporting is given by:

$$k_u(u) + \tau u(1+\theta)\phi h_u(u) - \tau[1 - \phi h(u)(1+\theta)] = 0 \tag{1}$$

Let u^* and $\bar{y}^*$be the optimal level of under-reporting and reported income in the static-model.

Dynamic Case: To model the strategic concern, we assume that the probability of getting caught in time period 2 also depends on the growth of reported income. Let $\delta_1 = \phi h(u_1)$ and $\delta_2 = \phi h(u_2) + f(g)$, where $g = (\bar{y}_2 - \bar{y}_1)/\bar{y}_1$. Lower reported growth results in higher probability of getting caught which implies that $f'(g) < 0$. The firm now jointly chooses u_1 and u_2 to maximize:

$$\begin{aligned} E[\pi] =& (1-\tau)[y-u_1] - k(u_1) + u_1 - \delta_1[\tau u_1 + \theta\tau u_1] \\ &+ (1-\tau)[y-u_2] - k(u_2) + u_2 - \delta_2[\tau u_2 + \tau\theta u_2] \end{aligned}$$

The FOCs with respect to u_1 and u_2 is given by:

$$k_u(u_1) + \tau u_1(1+\theta)\phi h_u(u_1) - \tau[1 - \phi h(u_1)(1+\theta)] + [\tau u_2 \bar{y}_2(1+\theta) f'(g)]/\bar{y}_1^2 = 0 \quad (2)$$

$$k_u(u_2) + \tau u_2(1+\theta)\phi h_u(u_2) - \tau[1 - \phi h(u_2)(1+\theta)] + \tau(1+\theta)[f(g) - u_2 f'(g)/\bar{y}_1] = 0 \quad (3)$$

Let u_1^* and u_2^* solve the above equations. Now, assume that $A(u)$ represents the LHS of equation 1, then equation 2 can be written as:

$$A(u_1) + B(u_1, u_2) = 0,$$

where $B(u_1, u_2) = [\tau u_2 \bar{y}_2(1+\theta) f'(g)]/\bar{y}_1^2 < 0$. Simple comparative stats reveal that $\frac{\partial A}{\partial u} > 0$. Combining the fact that $A(u^*) = 0$, $B(u_1, u_2) < 0$ and $\frac{\partial A}{\partial u} > 0$ gives us the result that,

$$u_1^* > u^* \implies \bar{y}_1^* < \bar{y}^*$$

Similarly, we can argue that

$$u_2^* < u^* \implies \bar{y}_2^* > \bar{y}^*$$

Thus, if the firms believe that reporting zero growth will increase the probability of getting caught, then they don't report the same level of income across different time periods and stagger their growth. For our main analysis, this implies that the excess mass of bunchers can be diffused rather than concentrated at the notch.

A.1.3 Appendix Tables

Table A.1: Test for change in probability in the treatment bins versus control bins

Dependent Variable: (Proportion of firms in bunching region)	(1) RST Firms	(2) RST Firm	(3) RST Firms	(4) Non-RST Firms
Treatment Bin x After	0.0218***	0.00755	0.00857	0.0161
	(0.00336)	(0.00551)	(0.00517)	(0.02244)
After	0.000861	0.00643**	0.00470	-.00038
	(0.00144)	(0.00274)	(0.00406)	(0.00575)
Constant	0.00926***	0.04734***	0.0391***	0.0105***
	(0.000640)	(.0013735)	(0.00197)	(0.00257)
Treatment Bins $[k_1, k_2]$	[8,15]	[17.5,21]	[21.5,26.5]	[8,15]
Control Bins	[1,8) ∪ (15,20]	[15,17.5) ∪ (21,25]	[20,21.5) ∪ (26.5,29]	[1,8) ∪ (15,20]
Bunching Region	[9,10)	[18,19)	[24,25)	[9,10)
Observations	105,499	26,153	19,209	3,473
R-squared	0.980	0.851	0.779	0.702

Robust standard errors (clustered at firm level) in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$
Note–The bin value represents revenue of firms in million of Rupees. In this table we do a bin-level difference-in-differences analysis as described by equation 1.5. We test if firms that report revenue in the treatment bins are more likely to be in the bunching region after the policy change in 2012 as compared to firms that are in the control bins. In all the specifications, we use bin fixed effects. In column 1, we restrict the sample to RST firms. If the firm has revenue between Rs.8-15 million in 2011, then it is in the treatment bin (see Figure 1.4). We test if there is higher probability that it reports revenue in the bunching region defined as area below the notch Rs.9-10 million. The control group is firms having revenue in the control bins in 2011. The baseline is probability of reporting revenue in the bunching region in 2011 based on revenue in 2009. In columns 2 & 3 we conduct a placebo test by incorrectly specifying the bunching region. The treatment bins for these columns are inferred from Appendix Figures A.10 and A.11. Finally, in column 4, we repeat the analysis for non-RST firms. We use the same treatment and control bins as column 1 to show that non-RST firms are not as responsive to notch as RST firms. All the data for this table is derived from Corporate Income Tax returns from 2009-13.

Table A.2: Sensitivity of estimates to the restrictions on the sample

VARIABLES	(1) PBITD	(2) Taxable Income	(3) Tax Paid	(4) Audit Fee
Panel A: Variables winsorized at 95th percentile				
Treat x Post2012	-522,468***	-58,884*	-23,960**	-2,197***
	(114,964)	(35,565)	(11,268)	(624.2)
Panel B: Variables winsorized at 99th percentile				
Treat x Post2012	-1.100e+06***	-195,057**	-87,277***	-2,868***
	(232,121)	(75,811)	(25,411)	(1,022)
Panel C: Unbalanced sample of firms				
Treat x Post2012	-638,932***	-79,117*	-34,297***	-2,393***
	(132,366)	(41,867)	(13,297)	(693.1)
Panel D: Inactive firms are included in the sample				
Treat x Post2012	-843,848***	-77,235*	-33,644**	-2,518***
	(150,410)	(43,137)	(13,680)	(666.7)
Panel E: Including firms that switch between RST and non-RST status				
Treat x Post2012	-197,062***	-72,561***	-23,108***	-730.4**
	(47,535)	(19,266)	(6,072)	(328.1)
Panel F: No restrictions on the sample				
Treat x Post2012	-223,985***	-50,204***	-14,794***	-708.7**
	(47,118)	(16,080)	(5,130)	(283.2)

Robust standard errors (clustered at firm level) in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Note–In this table, we test the sensitivity of estimates to the sample selection. Each panel relaxes a restriction on the sample used in the main analysis. For specification details, please refer to notes below Tables 1.2 and 1.3. All the data for this table is derived from Corporate Income Tax returns from 2009-16.

Table A.3: Estimates using a sub-sample of firms that have no potential selection bias.

VARIABLES	(1) PBITD	(2) Taxable Income	(3) Tax Paid	(4) Audit Fee
Treat x Post2012	-831,327***	-183,172**	-48,566**	-4,121***
	(235,198)	(74,799)	(22,320)	(1,195)
Observations	89,379	89,379	89,379	89,379
R-squared	0.581	0.609	0.611	0.733

Robust standard errors (clustered at firm level) in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Note- In this table, the treated group consists of RST firms with revenue between Rs.10-15 million in 2011, a year before the notch was moved from Rs.6 million to Rs.10 million. The comparison group consists of non-RST within the same revenue bandwidth in 2011. We want to test if excluding the firms between Rs.6 million to Rs.10 million changes the coefficients qualitatively. There is a potential concern that firms between Rs.6 million to Rs.10 million had an opportunity to bunch in 2011, when the notch was at Rs.6 million. Excluding these firms from the sample will remove any potential bias from the estimates. For specification details, please refer to notes below Tables 1.2 and 1.3. All the data for this table is derived from Corporate Income Tax returns from 2009-16.

Table A.4: Placebo test by mis-specifying the treatment neighborhood.

VARIABLES	(1) PBITD	(2) Taxable Income	(3) Tax Paid	(4) Audit Fee
Treat x Post2012	-368,030	-100,730	-44,932	-2,316*
	(246,508)	(97,147)	(31,412)	(1,206)
Observations	114,138	114,138	114,138	114,138
R-squared	0.611	0.615	0.616	0.737

Robust standard errors (clustered at firm level) in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Note– In this table, we show estimates from equation 1.3 on a group of firms that are unaffected by the policy change. The treatment group consists of restricted share transfer (RST) firms with revenue between Rs.16-28 million in 2011, a year before the audit threshold was moved from Rs.6 million to Rs.10 million. The comparison group consists of non-RST firms within the same revenue bandwidth in 2011. The treatment is removal of third-party audit requirement because of the change in the threshold. All the regressions in the table include firm fixed effects, year fixed effects and sector-specific time trends. The data comes from Corporate Income Tax returns from 2009-16.

Table A.5: Differences in the effect of removal of third-party audit between MAT and CIT firms.

VARIABLES	(1) Tax Paid	(2) Taxable Income	(3) Tax Paid	(4) Taxable Income
Treat x Post2012	-44,559	-95,218	-34,019**	-121,723**
	(31,946)	(78,613)	(16,425)	(52,457)
Sample	MAT firms	MAT firms	CIT firms	CIT firms
Observations	22,115	22,115	148,012	148,012
R-squared	0.695	0.595	0.638	0.640

Robust standard errors (clustered at firm level) in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Note– In this table we divide the sample between firms that pay taxes under regular corporate income tax (CIT) and the ones that have to pay the minimum alternate tax (MAT).The treatment group consists of restricted share transfer (RST) firms with revenue between Rs.6-15 million in 2011, a year before the audit threshold was moved from Rs.6 million to Rs.10 million. The comparison group consists of non-RST firms within the same revenue bandwidth in 2011. The treatment is removal of third-party audit requirement because of the change in the threshold. All the regressions in the table include firm fixed effects, year fixed effects and sector-specific time trends. The data comes from Corporate Income Tax returns from 2009-16.

Table A.6: Static bunching analysis using different bin-sizes

Bin Size (million Rs.)	(1) 0.1 mil	(2) 0.3mil	(3) 0.5mil
Upper Bound	12	13	11.5
(Standard Error)	(8.204)	(4.548)	(8.795)

Note- This table provides estimates of the upper-bound of the treatment neighborhood using different bin-sizes for the static analysis. The standard errors are estimated using a bootstrap procedure using 50 iterations. The bin-size used in the main analysis is Rs.0.5 million. All the data for this table is derived from Corporate Income Tax returns from 2012-16.

Table A.7: Estimates using the upper bound of treatment neighborhood from static bunching analysis.

VARIABLES	(1) PBITD	(2) Taxable Income	(3) Tax Paid	(4) Audit Fee
Treat x Post2012	-649,838***	-81,007	-40,462**	-1,641*
	(173,333)	(55,502)	(18,002)	(891.2)
Observations	113,336	113,336	113,336	113,336
R-squared	0.579	0.595	0.594	0.722

Robust standard errors (clustered at firm level) in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Note – In this table, we use the upper bound of the treatment neighborhood obtained by static bunching analysis instead of difference in probabilities method. The treatment group consists of restricted share transfer (RST) firms with revenue between Rs.6-11.5 million in 2011, a year before the audit threshold was moved from Rs.6 million to Rs.10 million. The comparison group consists of non-RST firms within the same revenue bandwidth in 2011. The treatment is removal of third-party audit requirement because of the change in the threshold. All the regressions in the table include firm fixed effects, year fixed effects and sector-specific time trends. The data comes from Corporate Income Tax returns from 2009-16.

Table A.8: Table of upstream ratios of industries.

Description	ITR Code	Upstream Ratio
Automobile and Auto parts	102	0.33
Cement	103	0.99
Drugs and Pharmaceuticals	105	0.74
Electronics including Computer Hardware	106	0.38
Engineering goods	107	0.15
Fertilizers, Chemicals, Paints	108	0.84
Flour & Rice Mills	109	0.85
Petroleum and Petrochemicals	113	0.87
Power and energy	114	0.79
Printing & Publishing	115	0.45
Rubber	116	0.89
Steel	117	0.88
Sugar	118	0.17
Tea, Coffee	119	0.69
Textiles, handloom, Power looms	120	0.45
Tobacco	121	0.54
Vanaspati & Edible Oils	123	0.53
Chain Stores	201	0
Retailers	202	0
Wholesalers	203	1
Builders	401	0.06
Estate Agents	402	0.12
Property Developers	403	0.56
Others	404	0.12
Civil Contractors	501	0.12
Legal professionals	603	0.50
Medical professionals	604	0
Nursing Homes	605	0
Specialty hospitals	606	0
Beauty Parlours	702	0

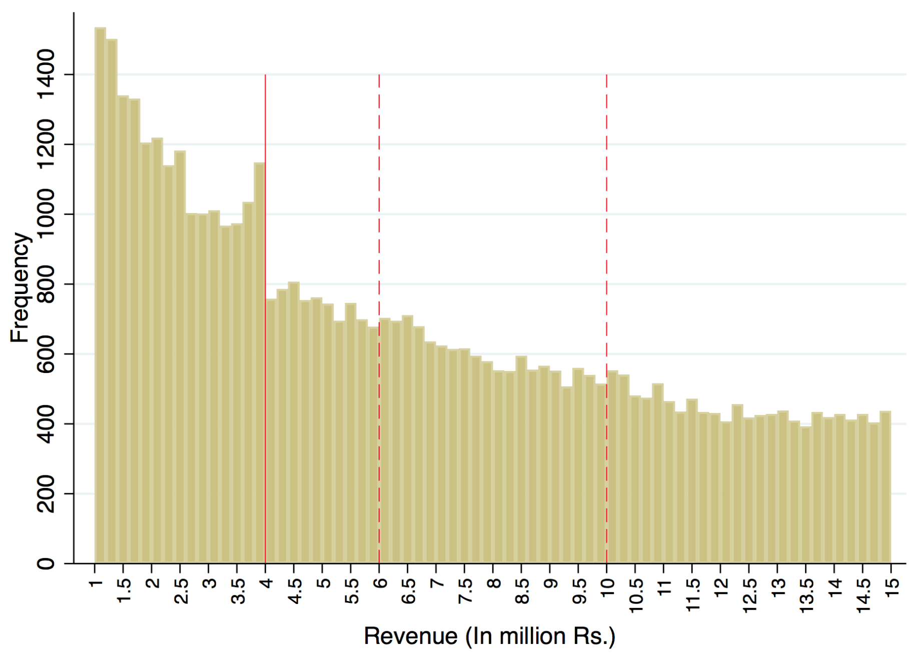
Continued on next page

Description	ITR Code	Upstream Ratio
Consultancy services	703	0
Courier Agencies	704	0
Computer training/educational and coaching institutes	705	0.02
Forex Dealers	706	0.00
Hospitality services	707	0.43
Hotels	708	0.43
I.T. enabled services, BPO service providers	709	0.03
Security agencies	710	0
Software development agencies	711	0.03
Transporters	712	0.27
Banking Companies	801	0.67
Chit Funds	802	0.67
Financial Institutions	803	0.67
Financial service providers	804	0.67
Leasing Companies	805	0.67
Non-Banking Finance Companies	807	0.67
Cable T.V. productions	901	0.06
Film distribution	902	0.06
Film laboratories	903	0.06
Motion Picture Producers	904	0.06
Television Channels	905	0.06

Note – This table was created by matching the description of industry codes in Income Tax forms to the Supply-Use Tables compiled by the Ministry of Statistics and Program Implementation in 2011. The upstream ratio is the proportion of sales to the intermediate consumer to the total consumption.

A.1.4 Appendix Figures

Figure A.1: Distribution of RST firms in 2009 when audit threshold was Rs.4 million.



]

Figure A.2: Distribution of RST firms in 2010 when audit threshold was Rs.6 million.

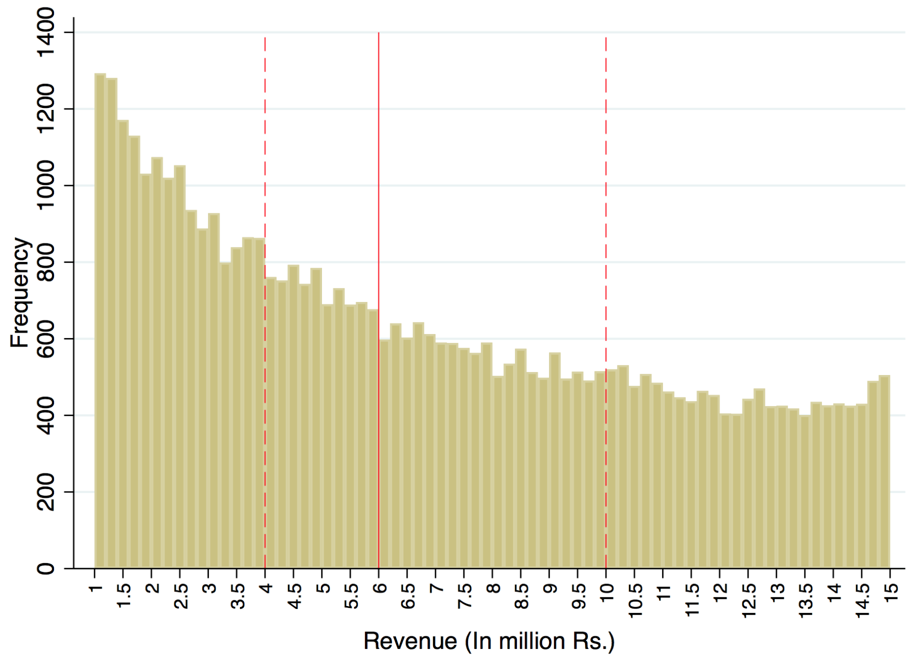


Figure A.3: Distribution of RST firms in 2011 when audit threshold was Rs.6 million.

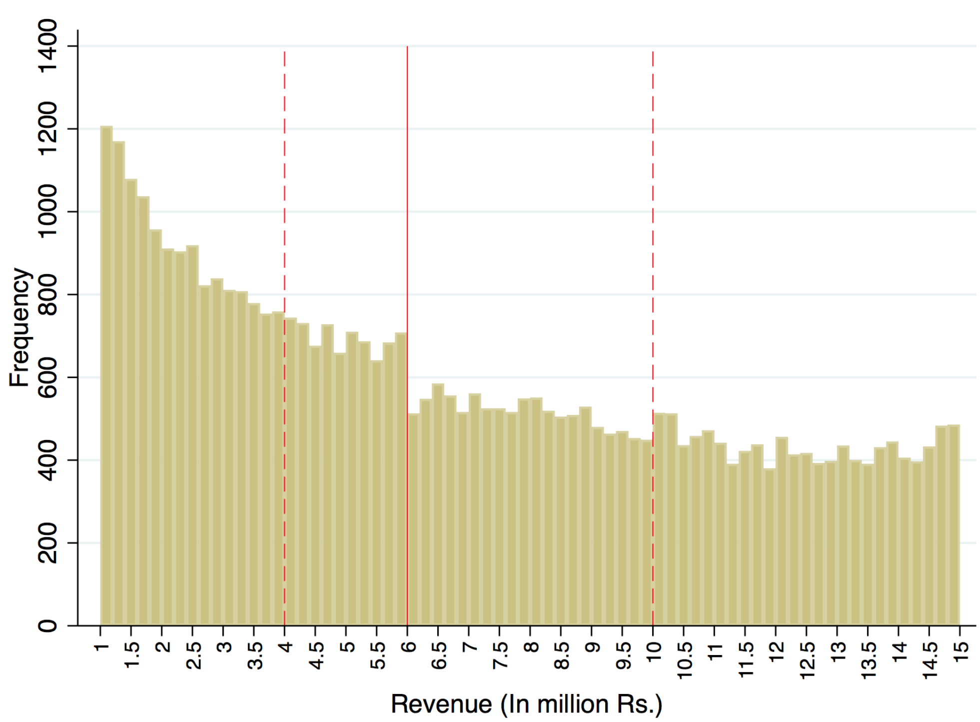


Figure A.4: Distribution of RST firms in 2012 when audit threshold was Rs.10 million.

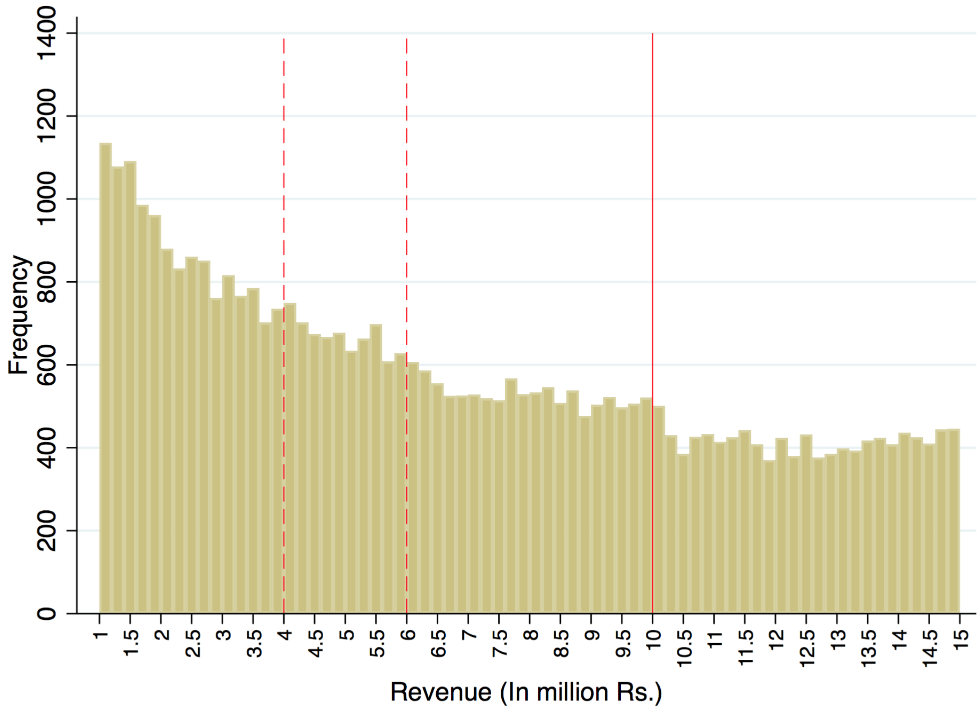


Figure A.5: Distribution of RST firms in 2013 when audit threshold was Rs.10 million.

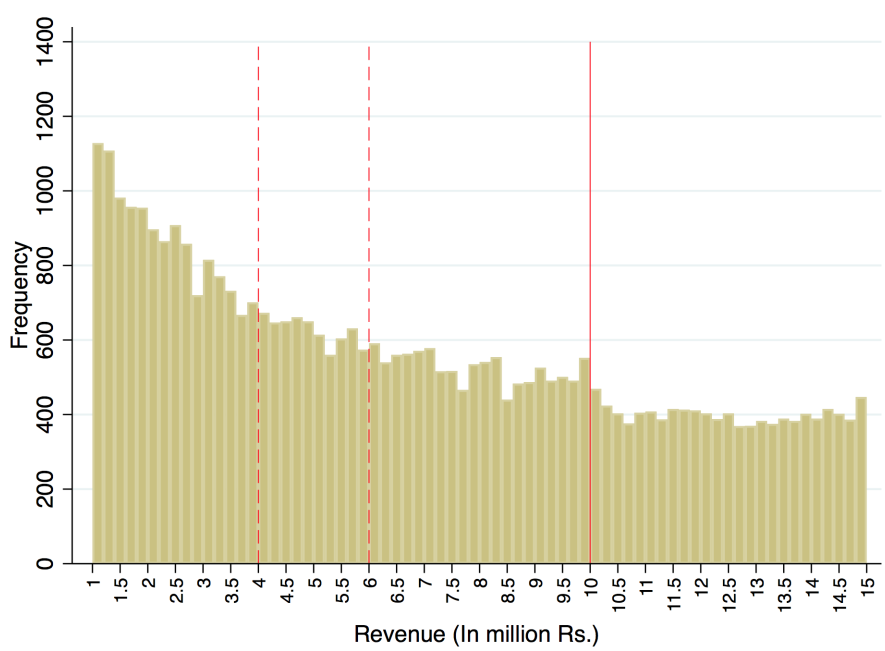


Figure A.6: Distribution of RST firms in 2014 when audit threshold was Rs.10 million.

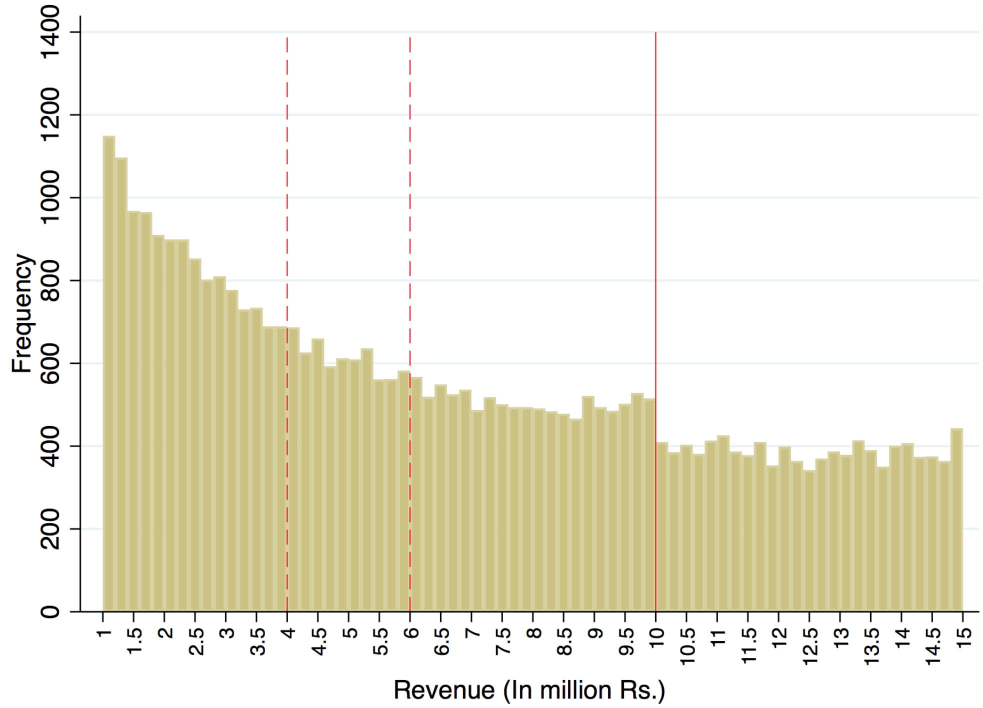


Figure A.7: Distribution of RST firms in 2015 when audit threshold was Rs.10 million.

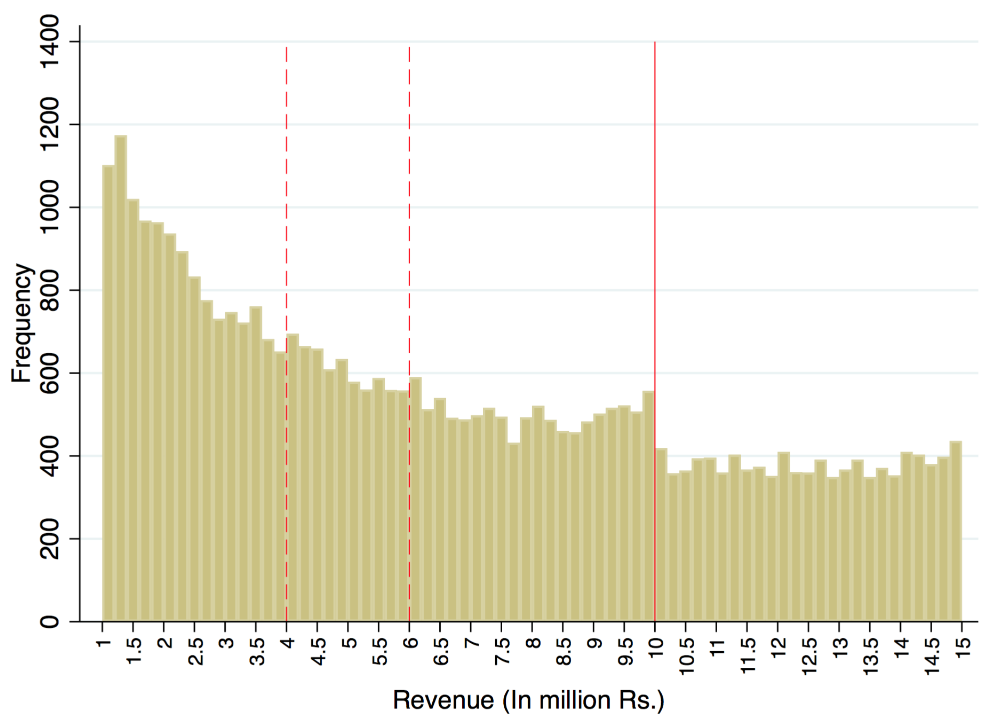


Figure A.8: Distribution of RST firms in 2016 when audit threshold was Rs.10 million.

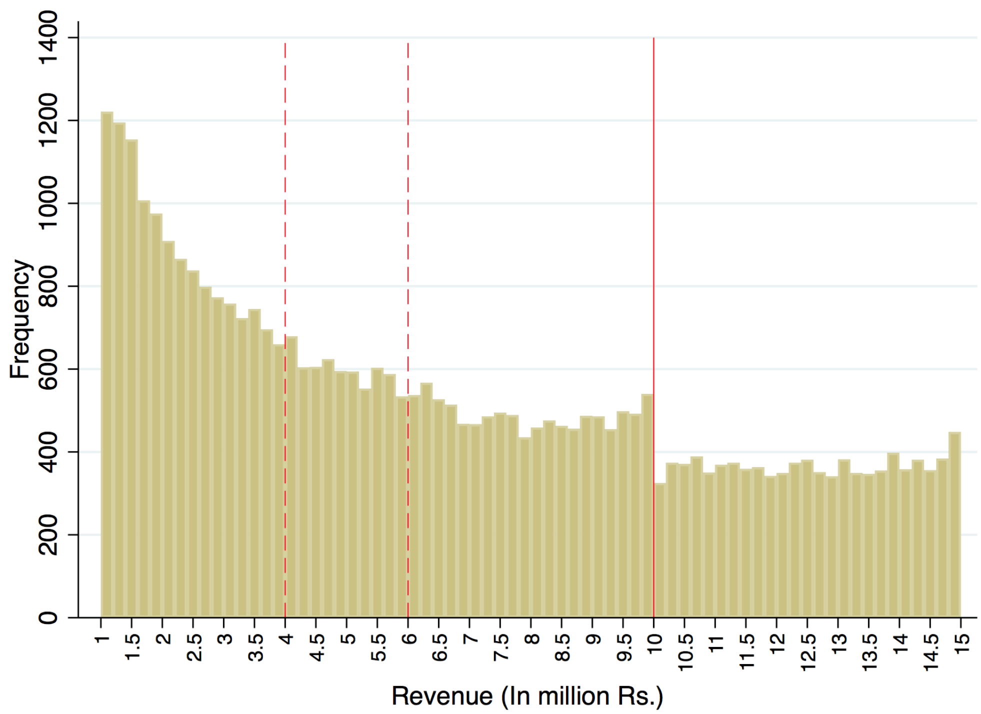
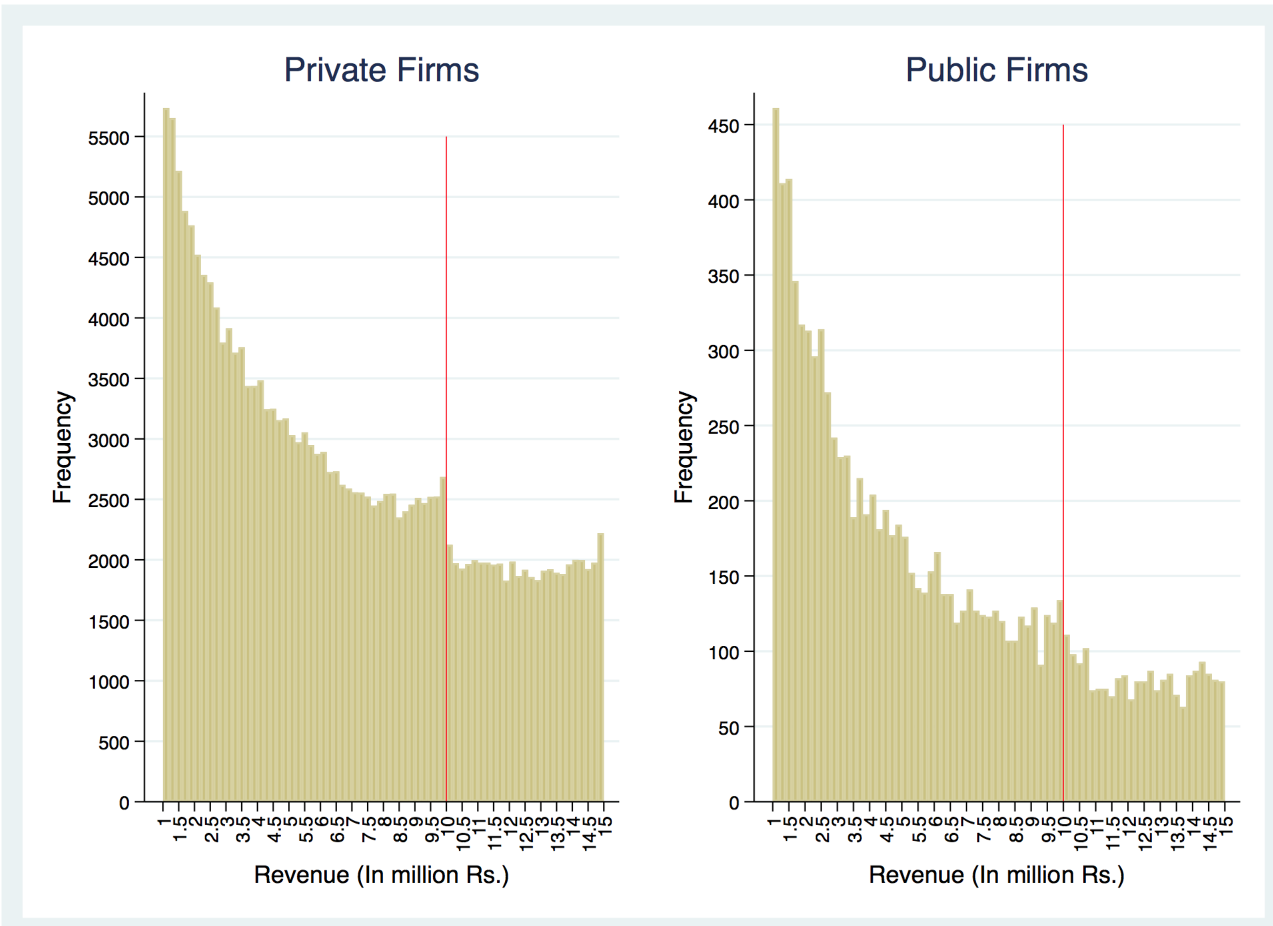
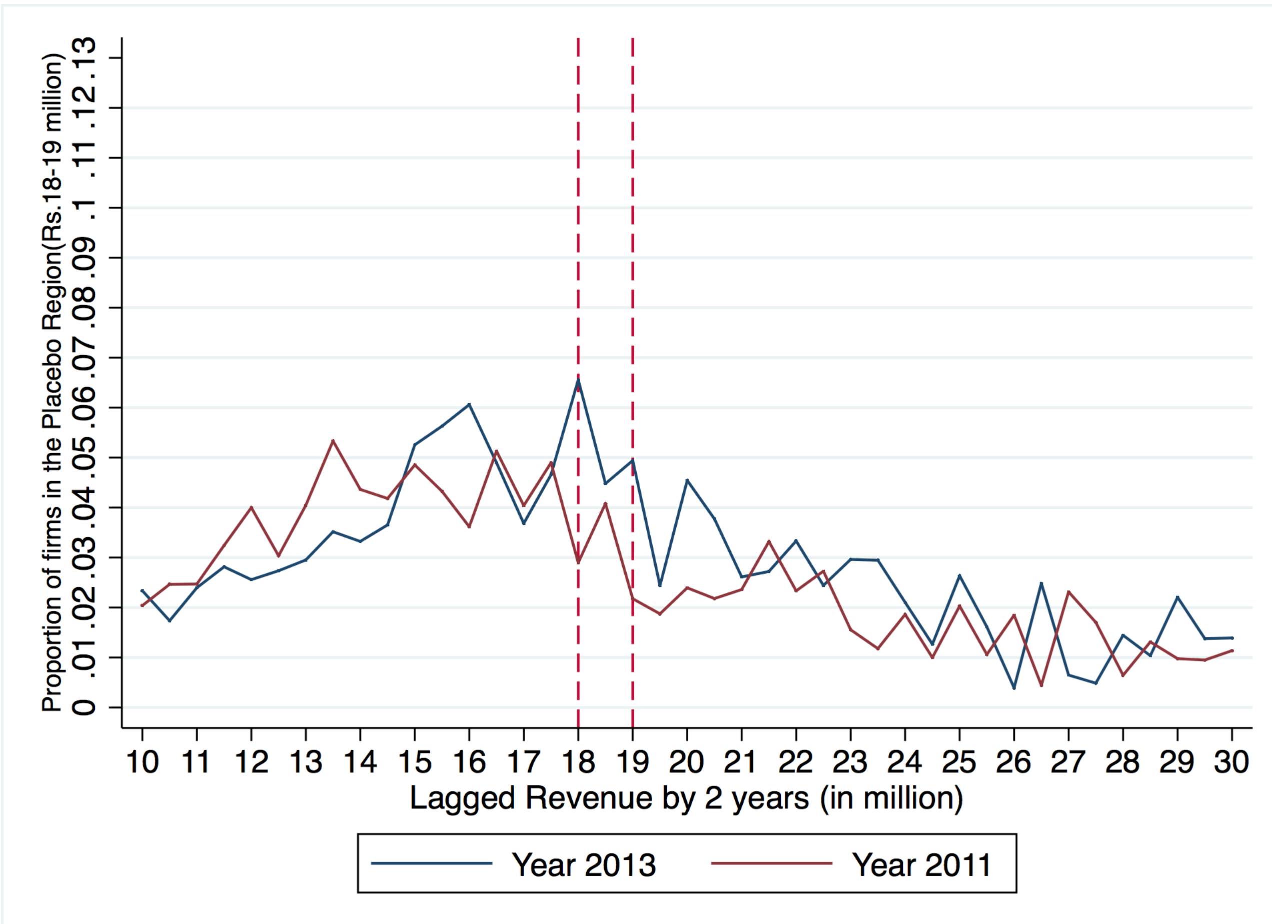


Figure A.9: Histograms of RST and non-RST firms from 2012-16.



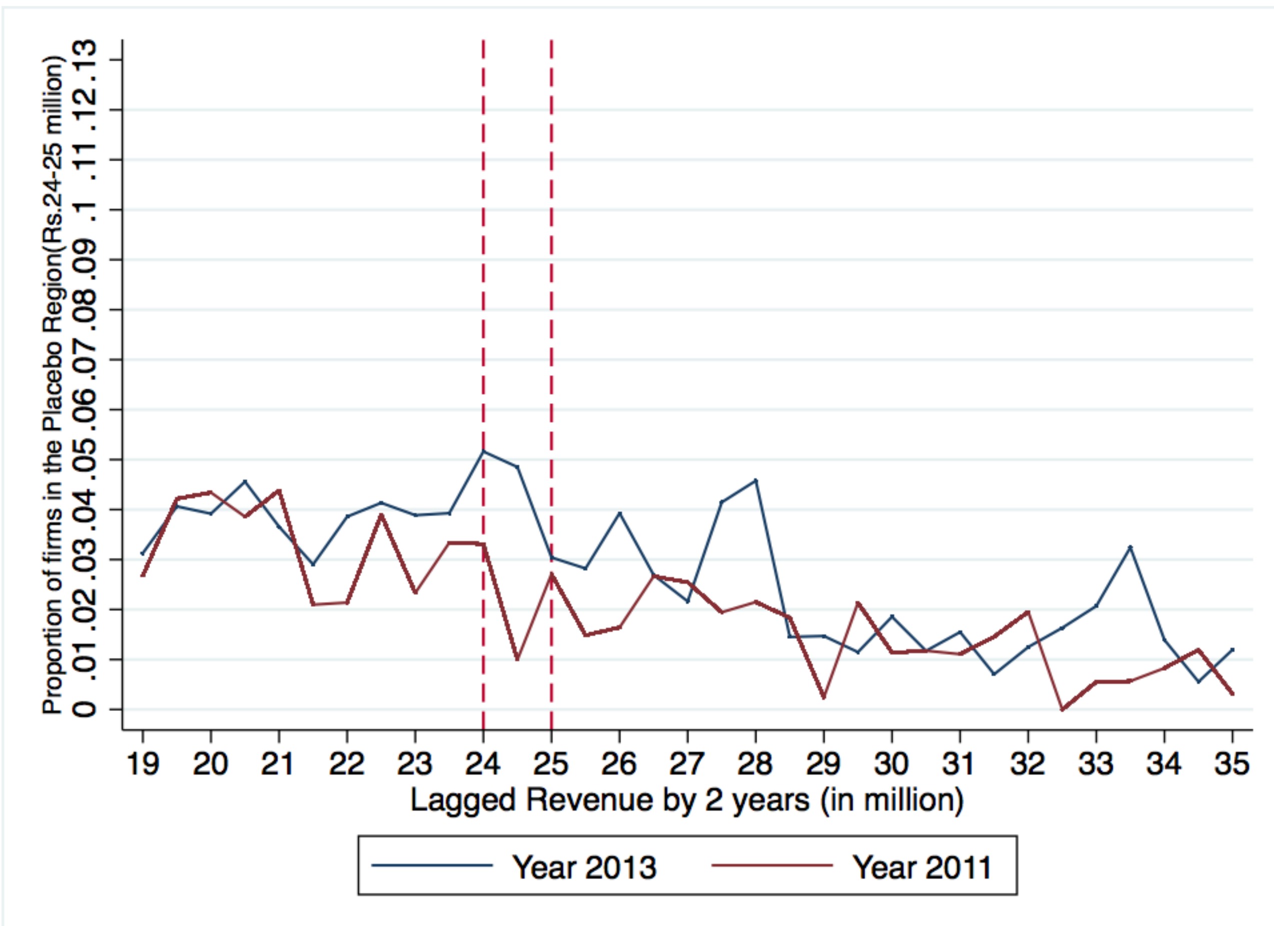
Note – In this graph, we show the difference in bunching behavior between RST and non-RST firms using data from 2012-16. The third-party audit threshold was Rs.10 million during this period. All the data for this graph is derived from Corporate Income Tax returns from 2012-16.

Figure A.10: Placebo test for difference in probabilities method



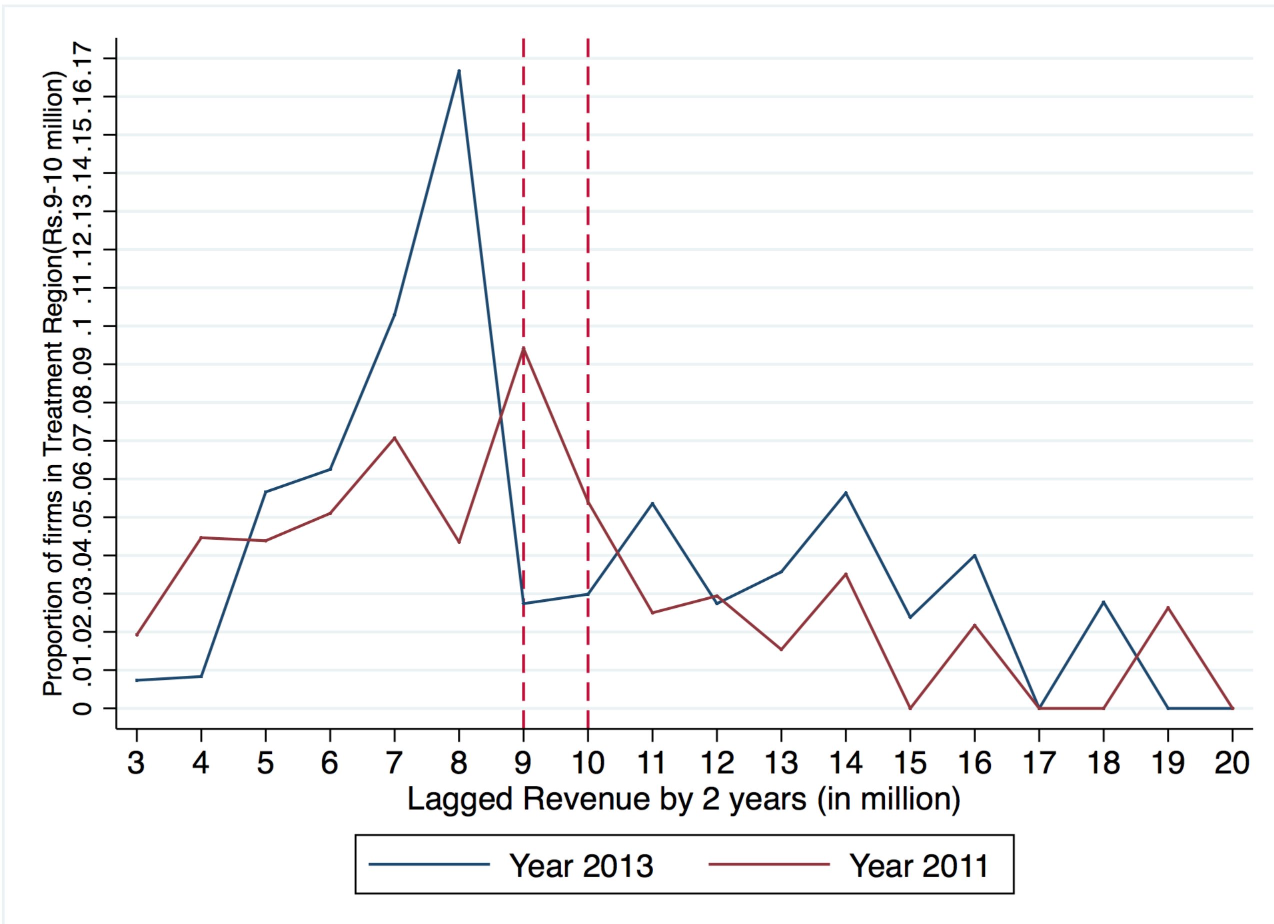
Note–In this graph, we plot the probability of being in a placebo region based on two-year lagged revenue. The notch was introduced at Rs.10 million in 2012. The blue line represents the probability of reporting revenue between Rs.18-19 million (region unrelated to the audit threshold) in 2013, conditional on revenue in 2011. Similarly, the red line represents the probability of reporting revenue in the same range in 2011, conditional on revenue in 2009 – both the years are before the change in policy. The bin size used in this graph is Rs.0.5 million. All the data for this graph is derived from Corporate Income Tax returns from 2009-13.

Figure A.11: Placebo test for difference in probabilities method



Note–In this graph, we plot the probability of being in another placebo region, Rs.24-25 million, based on two-year lagged revenue.

Figure A.12: Estimating upper bound using the difference in probability method for non-RST firms



Note–In this graph, we plot the probability of being in the bunching region based on two-year lagged revenue. The notch was introduced at Rs.10 million in 2012. The blue line represents the probability of reporting revenue between Rs.9-10 million (bunching region) in 2013, conditional on revenue in 2011. Similarly, the red line represents the probability of reporting revenue in the same range in 2011, conditional on revenue in 2009 – both the years are before the change in policy. The difference between the two probabilities shows the effect of the notch on firm's bunching response. The upper bound estimated from this graph is Rs.15 million, which is larger than the upper bound estimated by static bunching analysis. The bin size used in this graph is Rs.1 million. All the data for this graph is derived from Corporate Income Tax returns from 2009-13.

APPENDIX B

Chapter II Supporting Material

B.1 Appendix

B.1.1 Various Definitions of Formality

The third round of Nepal Living Standards Measurement Survey (LSMS), conducted in 2010-11, asks every respondent if she is engaged in self or wage employment and further if the employment is in agriculture or non-agriculture sector. We construct four different definitions of formality using this information. While we use the first definition of formality in the main analysis, we also test if our results are robust to alternative definitions in the sensitivity analysis.

- Definition 1 - A wage-earning person is assigned to the formal sector if *any* of the jobs she works in displays *any* of the following features - tax is deducted by the employer, employee contributes to the employee provident fund, employee will receive pension on retirement, or subsidized medical care. Further, if the person is self-employed and the firm is registered with the government, then also she gets assigned to the formal sector. Conversely, if the person is engaged in agriculture then she is assigned to informal sector, as the agricultural income is exempt from personal income-tax in Nepal. Unemployed, household workers and students are also assigned to the informal sector.

- Definition 2 - A person who is wage-employed is assigned to formal sector only if *any* of the jobs she works in displays *all* the job features listed above. All else is similar to Definition 1.

- Definition 3 - For this definition, we consider one more job criterion mentioned in the survey - if the employer has more or less than 10 workers employed. Several labor law regulations, particularly those pertaining to hiring and firing of workers are only applicable to firms that employ more than 10 workers. All else is similar to Definition 1.

- Definition 4 - In this case, all the self-employed people are considered to be in the informal sector. All else is similar to Definition 1.

B.1.2 Methodology of Imputing Income Tax Payments

We impute and assign the income tax payments to each household by constructing a tax table and merging it with the LSMS data. We only do this for households heads that are employed in the formal sector as defined in the previous appendix. Households heads who are employed in the informal sector, including agriculture, are assumed to pay zero income tax.

First, we create an income tax table that shows tax payments for each possible level of income by using the income tax schedule of Nepal. Then, household's consumption, c, equals $f(s) - T(s)$, where $f(s)$ is the taxable income and $T(s)$ is the income tax payment. For example, if the taxable income of a single male is NPR 2,00,000, then the income tax—according to the tax schedule—is NPR 7,600 and the after-tax consumption is NPR 192,400.

Next, we merge the income tax table with the LSMS data by matching the imputed consumption from the tax table to the reported consumption in the LSMS data. The income tax schedule of Nepal, like most countries, has a non-decreasing average tax rate over income. This ensures a monotonic relation between consumption and taxable income. Thus, we can assign the taxable income and tax payment to each household in the LSMS data using the tax table. Using the previous example, if some household(s) reports consumption of NPR 192,400 in the LSMS survey, then the income tax payments of that household is imputed as NPR 7,600. In this exercise, we are making three assumptions. First, it is the household head who is earning all the income and therefore, a single tax return is filed per household. Second, a proportion of the household consumption may be deductible from income for tax purposes. For instance, a self-employed person might claim business expenses as itemized deductions. In the absence of access to administrative data, we do not know whether such claims are made. Third, we assume that there is full tax compliance by people employed in the formal sector.

Finally, in the above analysis we create separate income tax tables for married and single household heads as they face different tax schedules according to the income tax law. Furthermore, single women face the same tax schedule as single men, however, they get a 10% deduction from their total tax liability. Thus, we create separate tax tables for each category of taxpayers and match them to the LSMS data using the process described above. The income tax schedules are described below:

Table B.1: Marginal income tax schedule for a single male filer

Taxable Income (In NPRs)	Marginal tax rate(In percent)
Less than 160,000	1
Between 160,000 & 260,000	15
Between 260,000 & 2,500,000	25
Above 2,500,000	35

Marginal income tax schedule for joint filing by a married couple

Taxable Income (In NPRs)	Marginal tax rate(In percent)
Less than 200,000	1
Between 200,000 & 300,000	15
Between 300,000 & 2,500,000	25
Above 2,500,000	35

B.1.3 Construction of variables used in the analysis.

1. **Identity and demographics of the household head** – We use the "household roster" of the LSMS survey to determine the identity of the household head. This section also has information on the age and gender of the family members. Additionally, we create a unique id for each respondent using the PSU id, household id and serial number of the respondent. This id is used to merge other relevant sections of the LSMS to the household roster.

2. **Education** – We use the "education" module of the LSMS questionnaire to construct this variable. For the grades 1 to 12, the years of education completed is equal to the education grade. Household members with bachelors and masters degree are coded as having completed 15 and 17 years of education, because the duration of bachelors and masters degree in Nepal is of three and two years respectively. We assume that professional degree holders have completed 15 years of education.

3. **Public Schools** – All the schools that are coded as Community/Government; community (public) campus; constituent campus and others, in the LSMS data are considered as public schools. The rest of the schools are considered as private schools.

4. **Jobs and Formality** – The "Wage jobs" section of LSMS documents job characteristics of household members who report themselves as wage earners, as opposed to self-employed. Using the job characteristics, we construct several measures of formality (details are given in the section B.1.1). All the self-employed members are also considered as part of the informal sector. If any person is working in a non-agricultural enterprise, we use the "Non-agriculture enterprises/activities" section to ascertain whether the enterprise is registered with the government or not. Persons working in the registered enterprises are also considered to be in the formal sector. Finally, we use the "Jobs and time use" section of the LSMS to identify household members who are engaged in household work or not employed at all. Such members are considered to be in the informal sector.

5. **Household Consumption** – We construct measures of annual food and total consumption by multiplying the annual per capita consumption of each household, recorded in the "Poverty" file of the LSMS, with the total number of household members.

6. **Benefits** –The benefits received by each member of the household are recorded in the "Transfers, social assistance and Other Income" section of the LSMS. First, we annualize the benefits received by each member of the household under all the programs mentioned in the section. Then, we create a measure of total annual benefits received by the household by adding the benefits received by all the members of the household. This allows us to deduct the annual benefits from the annual tax payments to calculate the net tax payments of a household. We use this measure in the sensitivity analysis as the main dependent variable. We also create a dummy variable for benefits received under programs where transfers are given in-kind rather than in cash. This variable is used to measure the gradient of take-up of welfare programs with respect to the education level.

7. **Value Added Tax** – The Value Added Tax, in our analysis, is equal to 13 percent of the total household consumption. An alternate measure of VAT excludes food consumption from the total consumption, because essential food items are zero-rated in Nepal.

8. **Mean remittances** -– We use the "Absentees information" section of the LSMS data to calculate the average remittances sent by migrants across different education levels.

9. **Education subsidies** -– We use National Education Accounts complied by the Nepalese government, UNESCO Institute for Statistics (UIS), International Institute for Educa-

tional Planning (IIEP) and Global Partnership for Education (GPE), to measure the subsidies incurred on education. Education subsidies are equal to the non-household expenditure which is defined as the expenditure by federal, state and local governments; NGOs, and school. The expenditure by government includes money spent by the ministry of education, district and village development committees, grants on budget, technical assistance off budget and administrative offices. Expenditure by NGOs includes both local and international NGOs. Finally, expenditure by schools is recorded under the header of "Internally generated funds" in the data. We merge this data with the LSMS data using the education level of the household head and the type of school attended by him (public or private).

B.1.4 Appendix Tables

Table B.2: Summary Statistics of variables not included in Table 2.1

	Age (in years)	Sum of education grades, in years, of other household members (excluding the head)	Highest grade of education, in years, within the household (excluding the head)	Household Size	Male Dummy for Household head	Benefit received (In NPRs)	Dummy for Benefits received	Dummy for In-Kind Transfers
Panel A: Primary Education (Grades 0-5)								
Count	3,787	3,787	3,635	3,787	3,787	3,787	3,787	3,787
Median	48	10	7	5	1	0	0	0
25th percentile	38	3	3	3	0	0	0	0
75th percentile	59	20	10	6	1	0	0	0
Mean	48.60	13	6.60	4.86	0.67	1,043	0.16	0.09
Standard deviation	13.74	12.83	4.36	2.37	0	2,940	0.37	0.28
Panel B: Secondary Education (Grades 6-10)								
Count	1,081	1,081	1,067	1,081	1,081	1,081	1,081	1,081
Median	40	14	9	5	1	0	0	0
25th percentile	32	8	5	3	1	0	0	0
75th percentile	50	25	12	6	1	0	0	0
Mean	41.37	18	8.29	4.94	0.82	784	0.14	0.06
Standard deviation	11.62	14.60	4.26	2.18	0	2,223	0.35	0.24
Panel C: Higher Education (Grade 11 to Bachelor's Degree)								
Count	893	893	862	893	893	893	893	893
Median	40	19	11	4	1	0	0	0
25th percentile	32	11	9	3	1	0	0	0
75th percentile	50	30	15	5	1	0	0	0
Mean	41.31	22	10.82	4.41	0.86	711	0.12	0.05
Standard deviation	12.20	16.92	4.23	2.19	0	2,293	0.33	0.22

Note: The primary data source is Nepal Living Standards Survey - 2010. The subsidy data comes from National Education Accounts reports compiled by International Institute for Educational Planning (IIEP), UNESCO Institute for Statistics (UIS) and Global Partnership for Education [IIEP Reports 2016a 2016b]. This data can be accessed at http://uis.unesco.org/en/news/national-education-accounts

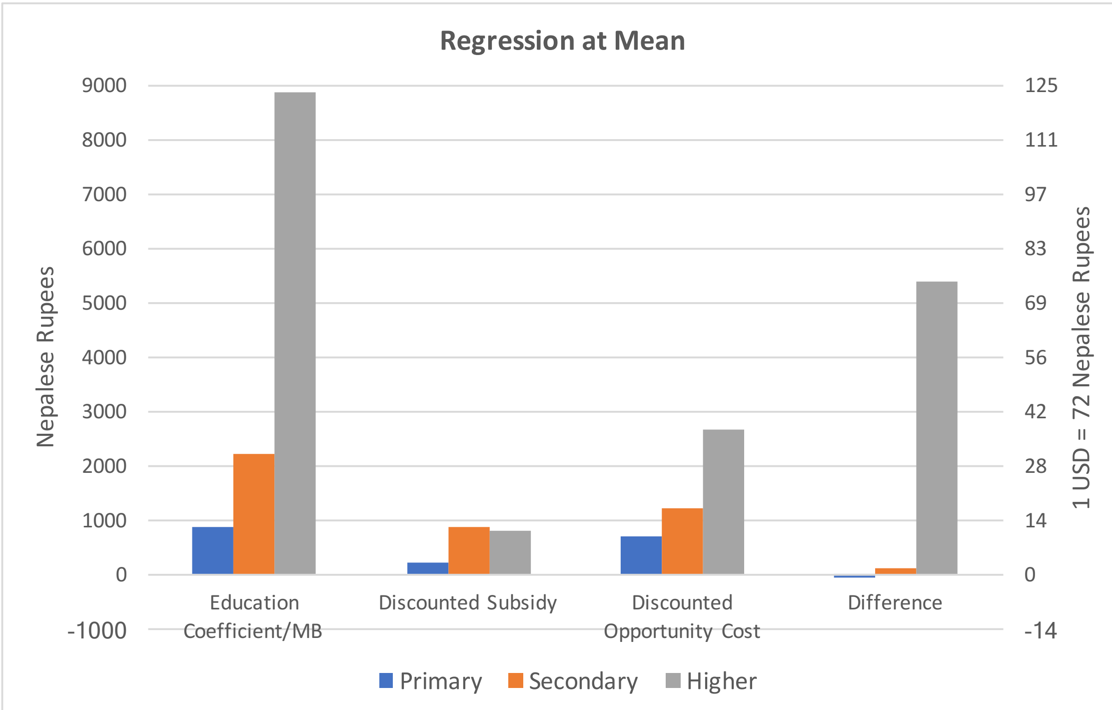
Table B.3: Migration, remittances, and taxes

	1 Education level of Household head	0	1	2	3	4	5	6	7	8	9	10	11	12	15
2	Probability of migration	0.04	0.05	0.13	0.12	0.16	0.22	0.21	0.21	0.24	0.27	0.23	0.24	0.18	0.13
3	Average income tax paid by the household (in NPR)	717	570	1864	1804	2235	1818	6684	3044	8008	6457	12862	13361	25840	40388
4	Average VAT paid by the household (in NPR)	17854	19365	19995	20390	20456	22883	25312	24490	27889	30717	33566	35058	39628	53032
5	Average net remittances sent by migrants who have years of schooling equal to the respective column (in NPR)	37752	64993	35546	42977	48176	83664	66230	69472	83358	88860	120688	104615	63537	62204
6	Average migration-adjusted taxes=[0.13*(5)*(2)] + [[(3)+(4)] * [1-(2)]]	18000	19352	19605	20274	20005	21677	27198	23700	29834	30335	39401	40178	55387	82754

Note: In this table, we adjust the tax paid by the household for the possibility of migration. The probability of migration is equal to the proportion of people who migrated conditional on school grade. If the household head migrates, then the household pays VAT on the remittances. These remittances, we assume, are equal to the average remittance sent by the migrants with similar years of schooling. If the household head doesn't migrate, then there is no change in the tax payments. Thus, migration-adjusted taxes = [0.13 X Average net remittances X Prob. of migration] + [(Direct taxes + VAT) X (1-Prob. of Migration)]. The VAT rate is 13 percent. The data for this table comes from the Nepal Living Standards Survey - 2010.

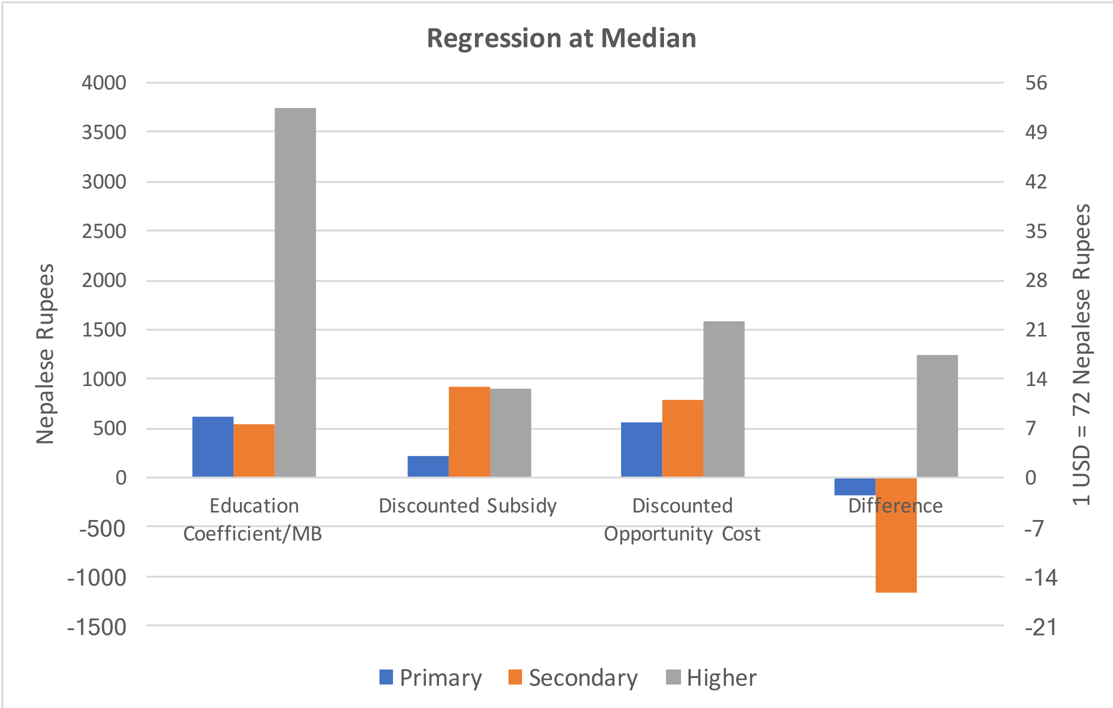
B.1.5 Appendix Figures

Figure B.1: Marginal Benefit versus Marginal Cost



Note: This figure uses the coefficients from Table 2.2 regression at the mean. It plots the different components of marginal cost and marginal benefit. The discount rate used in the analysis is 3 percent. Primary Education is from 0 to 5 grade, secondary from 6 to 10 and higher from 11 to Bachelors.

Figure B.2: Marginal Benefit versus Marginal Cost



Note: This figure uses the coefficients from Table 2.2 regression at the median.

APPENDIX C

Chapter III Supporting Material

C.1 Appendix

C.1.1 Appendix Tables

Table C.1: Value of coefficients of Table 4 at different levels of SRLB

Specifications same as Table 4	(1) Village FE	(2) Village FE	(3) Mother FE
	Whole Sample	Completed Fertility	# of children>1
(p-value in parentheses)			
Girl + Girl x SRLB	-0.037	-0.039	-0.046
	(0.000)	(0.000)	(0.000)
Girl + Girl x SRLB x 1.36	-0.045	-0.047	-0.054
	(0.000)	(0.000)	(0.000)
Girl + Girl x SRLB x 1.88	-0.057	-0.060	-0.065
	(0.000)	(0.000)	(0.000)
Second Child + Second child x SRLB	-0.033	-0.028	-0.049
	(0.000)	(0.000)	(0.000)
Second Child + Second child x SRLB x 1.36	-0.036	-0.032	-0.062
	(0.000)	(0.000)	(0.000)
Second Child + Second child x SRLB x 1.88	-0.042	-0.039	-0.081
	(0.000)	(0.000)	(0.000)
Later Child + Later child x SRLB	-0.069	-0.065	-0.098
	(0.000)	(0.000)	(0.000)
Later Child + Later child x SRLB x 1.36	-0.080	-0.077	-0.120
	(0.000)	(0.000)	(0.000)
Later Child + Later child x SRLB x 1.88	-0.097	-0.094	-0.153
	(0.000)	(0.000)	(0.000)
Girl x Second Child + Girl x Second child x SRLB	0.005	0.005	0.014
	(0.411)	(0.520)	(0.125)
Girl x Second Child + Girl x Second child x SRLB x 1.36	0.003	0.005	0.018
	(0.454)	(0.299)	(0.001)
Girl x Second Child + Girl x Second child x SRLB x 1.88	0.000	0.005	0.025
	(0.953)	(0.535)	(0.006)
Girl x Later Child + Girl x Later child x SRLB	0.004	0.005	0.004
	(0.663)	(0.658)	(0.737)
Girl x Later Child + Girl x Later child x SRLB x 1.36	-0.003	-0.004	0.007
	(0.541)	(0.551)	(0.286)
Girl x Later Child + Girl x Later child x SRLB x 1.88	-0.014	-0.016	0.012
	(0.147)	(0.151)	(0.300)

This table calculates the magnitude of the coefficients from the regressions in Table 3.4 at different levels of Sex Ratio at Last Birth (SRLB) which is defined as number of boys per girl in a district, conditional on being the last born. This ratio is calculated using families where mothers might have completed their fertility. 1.36 boys per girl is the median value and 1.88 is the top decile value of SRLB across districts. The coefficients should be interpreted as change in the probability of attaining the minimum level of learning.
[Source] All the data for this table is derived from ASER-2014.

Table C.2: Robustness Check – Using Raw Math Scores and Flexibly Controlling for School Grade

Dependent Variable (Raw Math Score)	(1) Village FE	(2) Mother FE	(3) Village FE	(4) Mother FE
(Robust SE in parentheses)	Completed fertility	# of children>1	Completed fertility	# of children>1
Girl	-0.117***	-0.145***	-0.0906***	-0.0791***
	(0.00552)	(0.00733)	(0.00842)	(0.00802)
Second Child	-0.0852***	-0.200***	-0.0651***	-0.155***
	(0.00651)	(0.00836)	(0.00600)	(0.00807)
Later Child	-0.206***	-0.370***	-0.195***	-0.332***
	(0.0108)	(0.0141)	(0.0103)	(0.0143)
Girl x Second Child	0.0106	0.0567***		
	(0.00881)	(0.00970)		
Girl x Later Child	-0.0254**	0.0251**		
	(0.0126)	(0.0123)		
No Elder Brother			0.0466***	0.0530***
			(0.00744)	(0.00814)
Girl x No Elder Brother			-0.0331***	-0.0344***
			(0.00945)	(0.0110)
Constant	1.274***	1.745***	1.235***	1.681***
	(0.0361)	(0.0310)	(0.0367)	(0.0319)
(p-value in parentheses)				
NEB + Girl x NEB			0.0135	0.0185
			(0.137)	(0.169)
Hh Controls	YES	NO	YES	NO
Observations	198,963	202,793	198,963	202,793

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Robust standard errors are clustered at the village level. The learning outcome is the raw score which ranges discretely from 1 to 5. The average score is 3.25. The coefficients should be interpreted as change in the raw score. Column (1) and (3) restricts the sample to the mothers who might have completed their fertility and thus, allows to control for the total number of children. Column (2) and (4) use Mother Fixed effects. All the regressions also include birth year Fixed effects. They also include school grade Fixed effects to flexibly control for class in which child is studying. Child level control includes the child age. Household level controls include the parents' education, parents' age and index for household infrastructure.

[Source] All the data for this table is derived from ASER-2014.

Table C.3: Robustness Check – Using Standardized Reading Score as Learning Outcome

Dependent Variable (Standard Reading Level)	(1) Village FE	(2) Mother FE	(3) Village FE	(4) Mother FE
(Robust SE in parentheses)	Completed fertility	# of children>1	Completed fertility	# of children>1
Girl	0.00539**	-0.00255	0.0171***	0.0125***
	(0.00244)	(0.00347)	(0.00367)	(0.00371)
Second Child	-0.0346***	-0.0688***	-0.0257***	-0.0619***
	(0.00313)	(0.00408)	(0.00284)	(0.00388)
Later Child	-0.0566***	-0.106***	-0.0550***	-0.107***
	(0.00494)	(0.00671)	(0.00476)	(0.00685)
Girl x Second Child	0.00752*	0.0123**		
	(0.00407)	(0.00480)		
Girl x Later Child	-0.0142***	-0.00499		
	(0.00539)	(0.00587)		
No Elder Brother			0.0177***	0.00945**
			(0.00343)	(0.00392)
Girl x No Elder Brother			-0.0139***	-0.0150***
			(0.00422)	(0.00531)
Constant	-0.0720***	0.0951***	-0.0870***	0.0861***
	(0.0134)	(0.0137)	(0.0136)	(0.0143)
(p-value in parentheses)				
NEB + Girl x NEB			0.0038	-0.006
			(0.331)	(0.395)
Hh Controls	YES	NO	YES	NO
Observations	198,739	202,572	198,739	202,572

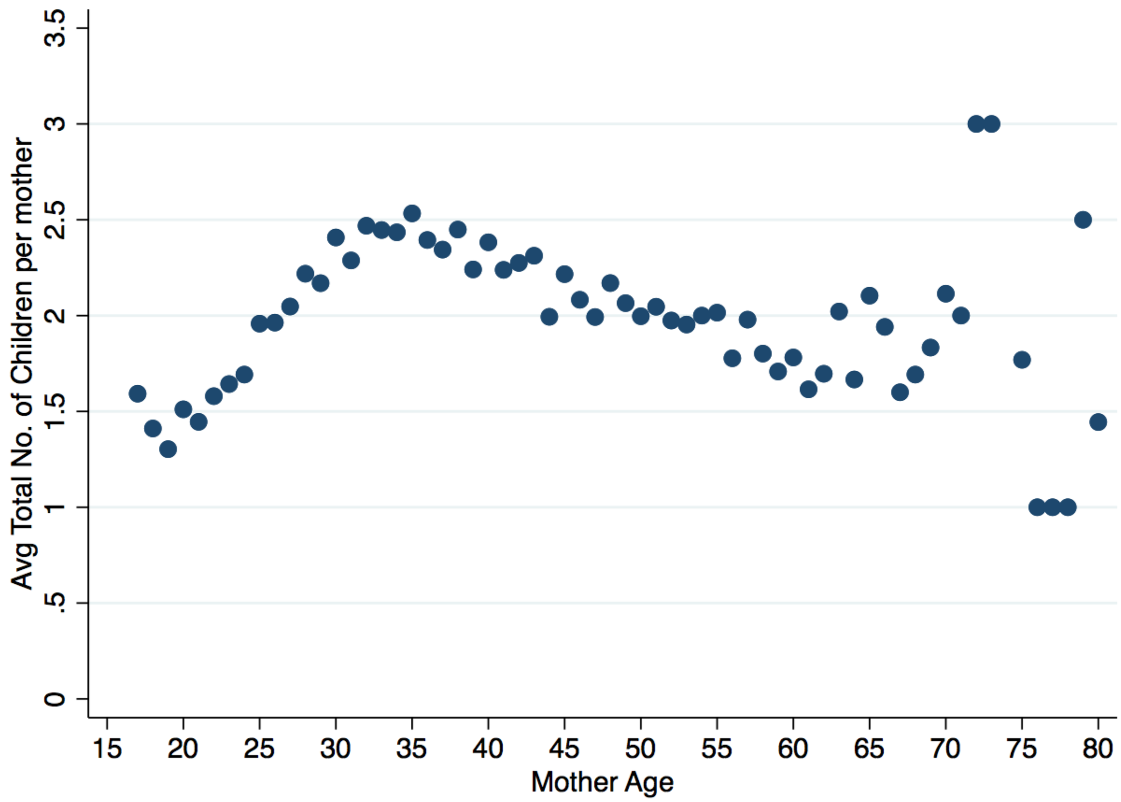
*** p<0.01, ** p<0.05, * p<0.1

Robust standard errors are clustered at the village level. The learning outcome is the reading score standardized according to the class in which the child is studying. The average score is 0.48. The coefficients should be interpreted as the change in the probability of attaining the minimum level of learning. Columns (1) and (3) restrict the sample to mothers who might have completed their fertility and thus, allows to control for the total number of children. Columns (2) and (4) use Mother Fixed effects. All the regressions also include birth year Fixed effects. Child level controls include the child age and school class. Household level controls include the parents' education, parents' age and index for household infrastructure.

[Source] All the data for this table is derived from ASER-2014.

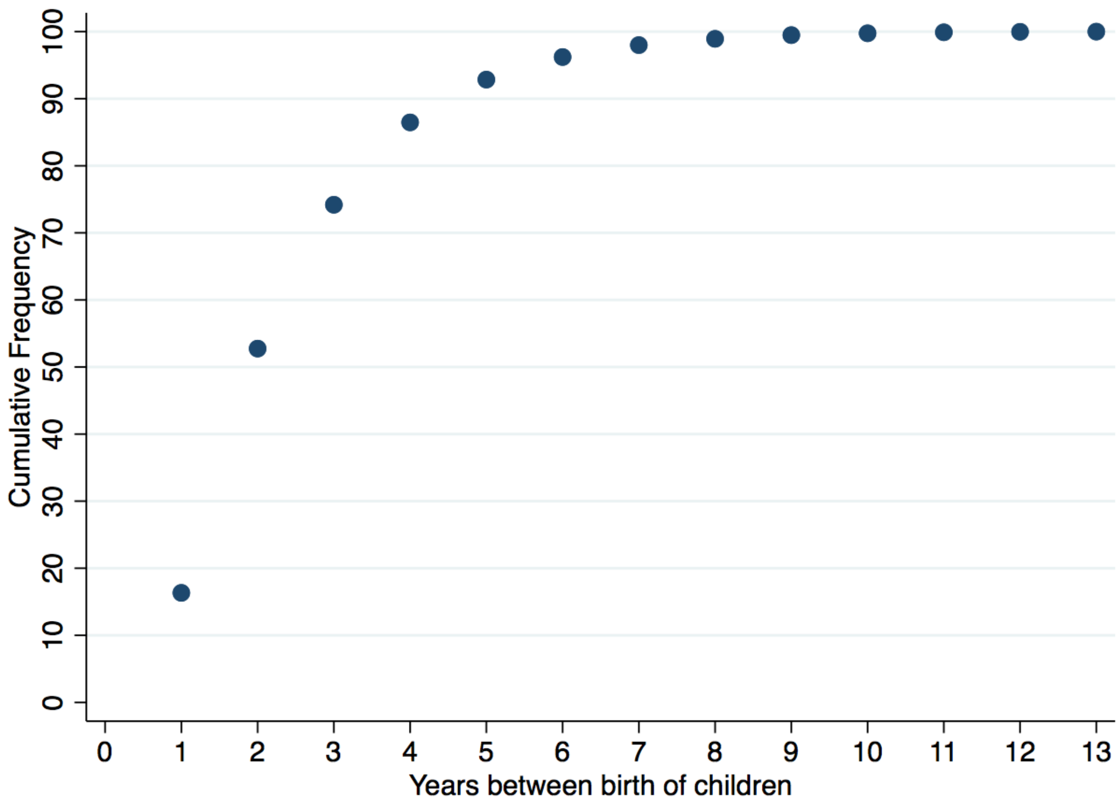
C.1.2 Appendix Figures

Figure C.1: Distribution of total children in a family according to the age of the mother



The average number of children per mother increases till the mother's age is 35, and then declines. This is used to determine the age by which women might have completed their fertility.
[Source] Data for the analysis comes from ASER 2014.

Figure C.2: Frequency distribution of years between birth of closest siblings



This graph shows cumulative frequency of the gap between closest siblings in terms of years. More that 86 percent of the siblings are born less than 5 years apart. This is used to identify women who might have completed their fertility.
[Source] Data for the analysis comes from ASER 2014.

BIBLIOGRAPHY

Almond, Douglas, and Janet Currie. 2011. "Killing me softly: The fetal origins hypothesis." *Journal of economic perspectives*, 25(3): 153–72.

Almond, Douglas, Lena Edlund, and Kevin Milligan. 2013. "Son preference and the persistence of culture: evidence from South and East Asian immigrants to Canada." *Population and Development Review*, 39(1): 75–95.

Almunia, Miguel, and David Lopez-Rodriguez. 2018. "Under the radar: The effects of monitoring firms on tax compliance." *American Economic Journal: Economic Policy*, 10(1): 1–38.

Andrews, Matt, Michael Woolcock, and Lant Pritchett. 2017. *Building state capability: Evidence, analysis, action.* Oxford University Press.

Anukriti, S. 2018. "Financial incentives and the fertility-sex ratio trade-off." *American Economic Journal: Applied Economics*, 10(2): 27–57.

Azam, Mehtabul, and Chan Hang Saing. 2018. "Is there really a trade-off? Family size and investment in child quality in India." *The BE Journal of Economic Analysis & Policy*, 18(1).

Azam, Mehtabul, and Geeta Gandhi Kingdon. 2013. "Are girls the fairer sex in India? Revisiting intra-household allocation of education expenditure." *World Development*, 42: 143–164.

Azuara, Oliver, and Ioana Marinescu. 2013. "Informality and the expansion of social protection programs: Evidence from Mexico." *Journal of health economics*, 32(5): 938–950.

Bachas, Pierre, and Mauricio Soto. 2018. *Not (ch) your average tax system: corporate taxation under weak enforcement.* The World Bank.

Basri, M Chatib, Mayara Felix, Rema Hanna, and Benjamin A Olken. 2019. "Tax Administration vs. Tax Rates: Evidence from Corporate Taxation in Indonesia." National Bureau of Economic Research.

Becker, Gary S, and H Gregg Lewis. 1973. "On the Interaction between the Quantity and Quality of Children." *Journal of political Economy*, 81(2, Part 2): S279–S288.

Behrman, Jere R. 1988. "Intrahousehold allocation of nutrients in rural India: Are boys favored? Do parents exhibit inequality aversion?" *Oxford Economic Papers*, 40(1): 32–54.

Bertanha, Marinho, Andrew H McCallum, and Nathan Seegert. 2019. "Better bunching, nicer notching." *Nicer Notching (December 23, 2019)*.

Best, Michael Carlos, and Henrik Jacobsen Kleven. 2018. "Housing market responses to transaction taxes: Evidence from notches and stimulus in the UK." *The Review of Economic Studies*, 85(1): 157–193.

Black, Maureen M, Susan P Walker, Lia CH Fernald, Christopher T Andersen, Ann M DiGirolamo, Chunling Lu, Dana C McCoy, Günther Fink, Yusra R Shawar, Jeremy Shiffman, et al. 2017. "Early childhood development coming of age: science through the life course." *The Lancet*, 389(10064): 77–90.

Black, Sandra E, Paul J Devereux, and Kjell G Salvanes. 2005. "The more the merrier? The effect of family size and birth order on children's education." *The Quarterly Journal of Economics*, 120(2): 669–700.

Blake, Judith. 1989. *Family size and achievement.* Vol. 3, Univ of California Press.

Bleakley, Hoyt. 2018. "Longevity, Education, and Income: How Large is the Triangle?" National Bureau of Economic Research.

Blomquist, Soren, Whitney K Newey, Anil Kumar, and Che-Yuan Liang. 2018. "Identifying the Effect of Taxes on Taxable Income." *Unpublished manuscript*.

Bolton, Patrick, Xavier Freixas, and Joel Shapiro. 2012. "The credit ratings game." *The Journal of Finance*, 67(1): 85–111.

Butcher, Kristin F, and Anne Case. 1994. "The effect of sibling sex composition on women's education and earnings." *The Quarterly Journal of Economics*, 109(3): 531–563.

Camacho, Adriana, Emily Conover, and Alejandro Hoyos. 2014. "Effects of Colombia's social protection system on workers' choice between formal and informal employment." *The World Bank Economic Review*, 28(3): 446–466.

CBS. 2011. "Nepal Living Standards Survey Statistical Report Volume 1." *Available at https://microdata.worldbank.org/index.php/catalog/1000/related-materials*.

CBS. 2012. *National Population and Housing Census 2011.* Central Bureau of Statistics, Kathmandu, Nepal. Available at https://unstats.un.org/unsd/demographic-social/census/documents/Nepal/Nepal-Census-2011-Vol1.pdf.

Chetty, Raj. 2009. "Is the taxable income elasticity sufficient to calculate deadweight loss? The implications of evasion and avoidance." *American Economic Journal: Economic Policy*, 1(2): 31–52.

Coffey, Diane, and Dean Spears. 2019. "Neonatal death in India: birth order in a context of maternal undernutrition."

Conley, Dalton, and Rebecca Glauber. 2006. "Parental educational investment and children's academic risk estimates of the impact of sibship size and birth order from exogenous variation in fertility." *Journal of human resources*, 41(4): 722–737.

Crocker, Keith J, and Joel Slemrod. 2005. "Corporate tax evasion with agency costs." *Journal of Public Economics*, 89(9-10): 1593–1610.

Dahl, Gordon B, and Enrico Moretti. 2008. "The demand for sons." *The Review of Economic Studies*, 75(4): 1085–1120.

Das Gupta, Monica, Jiang Zhenghua, Li Bohua, Xie Zhenming, Woojin Chung, and Bae Hwa-Ok. 2003. "Why is son preference so persistent in East and South Asia? A cross-country study of China, India and the Republic of Korea." *The Journal of Development Studies*, 40(2): 153–187.

Dayioğlu, Meltem, Murat G Kirdar, and Aysit Tansel. 2009. "Impact of sibship size, birth order and sex composition on school enrolment in urban Turkey." *Oxford Bulletin of Economics and Statistics*, 71(3): 399–426.

Duflo, Esther. 2001. "Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment." *American economic review*, 91(4): 795–813.

Duflo, Esther, Michael Greenstone, Rohini Pande, and Nicholas Ryan. 2013*a*. "Truth-telling by third-party auditors and the response of polluting firms: Experimental evidence from India." *The Quarterly Journal of Economics*, 128(4): 1499–1545.

Duflo, Esther, Michael Greenstone, Rohini Pande, and Nicholas Ryan. 2013*b*. "What does reputation buy? differentiation in a market for third-party auditors." *American Economic Review*, 103(3): 314–19.

Duflo, Esther, Rema Hanna, and Stephen P Ryan. 2012. "Incentives work: Getting teachers to come to school." *American Economic Review*, 102(4): 1241–78.

Dyson, Tim, and Mick Moore. 1983. "On kinship structure, female autonomy, and demographic behavior in India." *Population and development review*, 35–60.

Ejrnæs, Mette, and Claus C Pörtner. 2004. "Birth order and the intrahousehold allocation of time and education." *Review of Economics and Statistics*, 86(4): 1008–1019.

Fracassi, Cesare, Stefan Petry, and Geoffrey Tate. 2016. "Does rating analyst subjectivity affect corporate debt pricing?" *Journal of Financial Economics*, 120(3): 514–538.

Griffin, John M, and Dragon Yongjun Tang. 2011. "Did credit rating agencies make unbiased assumptions on CDOs?" *American Economic Review*, 101(3): 125–30.

Hamilton, Steven. 2018. "Optimal deductibility: Theory, and evidence from a bunching decomposition." *The Tax and Transfer Policy Institute-Working paper*, 14.

Härmä, Joanna. 2009. "Can choice promote Education for All? Evidence from growth in private primary schooling in India." *Compare*, 39(2): 151–165.

Hotz, V Joseph, and Juan Pantano. 2015. "Strategic parenting, birth order, and school performance." *Journal of population economics*, 28(4): 911–936.

IIEP, UIS and IIEP Pôle de Dakar. 2016*a*. "Methodology of National Education Accounts." *Available at http://unesdoc.unesco.org/images/0024/002457/245781e.pdf.*

IIEP, UIS and IIEP Pôle de Dakar. 2016*b*. "Who Pays For What In Education?" *Available at http://unesdoc.unesco.org/images/0024/002462/246277e.pdf.*

Inland Revenue Department Report. 2015. *Reform Plan 2015/16 - 2017/18.* Ministry of Finance, Government of Nepal, Kathmandu, Nepal. Available at https://ird.gov.np/Content/ContentAttachment/4/ReformPlanIRD8122016114507AM.pdf.

Institute of Chartered Accountants of India Year Book. 2018. *Year Book.* Public Relations Group.

IRS. 2013-18. *Internal Revenue Service Data Book.* Washington DC.

Ito, Koichiro, and James M Sallee. 2018. "The economics of attribute-based regulation: Theory and evidence from fuel economy standards." *Review of Economics and Statistics*, 100(2): 319–336.

Jayachandran, Seema. 2015. "The roots of gender inequality in developing countries." *economics*, 7(1): 63–88.

Jayachandran, Seema. 2017. "Fertility decline and missing women." *American Economic Journal: Applied Economics*, 9(1): 118–39.

Jayachandran, Seema, and Ilyana Kuziemko. 2011. "Why do mothers breastfeed girls less than boys? Evidence and implications for child health in India." *The Quarterly journal of economics*, 126(3): 1485–1538.

Jayachandran, Seema, and Rohini Pande. 2017. "Why are Indian children so short? The role of birth order and son preference." *American Economic Review*, 107(9): 2600–2629.

Jiang, John Xuefeng, Mary Harris Stanford, and Yuan Xie. 2012. "Does it matter who pays for bond ratings? Historical evidence." *Journal of Financial Economics*, 105(3): 607–621.

Johnson, William R. 2006. "Are public subsidies to higher education regressive?" *Education Finance and Policy*, 1(3): 288–315.

Kang, Changhui. 2011. "Family size and educational investments in children: Evidence from private tutoring expenditures in South Korea." *Oxford Bulletin of Economics and Statistics*, 73(1): 59–78.

Kantarevic, Jasmin, and Stéphane Mechoulan. 2006. "Birth order, educational attainment, and earnings an investigation using the PSID." *Journal of human resources*, 41(4): 755–777.

Keen, Michael, and Joel Slemrod. 2017. "Optimal tax administration." *Journal of Public Economics*, 152: 133–142.

Khan, Adnan Q, Asim I Khwaja, and Benjamin A Olken. 2016. "Tax farming redux: Experimental evidence on performance pay for tax collectors." *The Quarterly Journal of Economics*, 131(1): 219–271.

Kleven, Henrik Jacobsen. 2016. "Bunching." *Annual Review of Economics*, 8: 435–464.

Kleven, Henrik Jacobsen, Claus Thustrup Kreiner, and Emmanuel Saez. 2016. "Why can modern governments tax so much? An agency model of firms as fiscal intermediaries." *Economica*, 83(330): 219–246.

Kleven, Henrik Jacobsen, Martin B Knudsen, Claus Thustrup Kreiner, Søren Pedersen, and Emmanuel Saez. 2011. "Unwilling or unable to cheat? Evidence from a tax audit experiment in Denmark." *Econometrica*, 79(3): 651–692.

Kleven, Henrik J, and Mazhar Waseem. 2013. "Using notches to uncover optimization frictions and structural elasticities: Theory and evidence from Pakistan." *The Quarterly Journal of Economics*, 128(2): 669–723.

Kugler, Adriana D, and Santosh Kumar. 2017. "Preference for boys, family size, and educational attainment in India." *Demography*, 54(3): 835–859.

Lee, Jungmin. 2008. "Sibling size and investment in children's education: An Asian instrument." *Journal of Population Economics*, 21(4): 855–875.

Marx, Benjamin M. 2018. "Dynamic Bunching Estimation with Panel Data."

Mincer, Jacob. 1958. "Investment in human capital and personal income distribution." *Journal of political economy*, 66(4): 281–302.

Naritomi, Joana. 2019. "Consumers as tax auditors." *American Economic Review*, 109(9): 3031–72.

OECD. 2019. *Tax Administration 2019.*

Parish, William L, and Robert J Willis. 1993. "Daughters, education, and family budgets Taiwan experiences." *Journal of Human Resources*, 863–898.

Pavan, Ronni. 2016. "On the production of skills and the birth-order effect." *Journal of Human Resources*, 51(3): 699–726.

Pomeranz, Dina. 2015. “No taxation without information: Deterrence and self-enforcement in the value added tax.” *American Economic Review*, 105(8): 2539–69.

Price, Joseph. 2008. “Parent-child quality time does birth order matter?” *Journal of human resources*, 43(1): 240–265.

Psacharopoulos, George, and Harry Anthony Patrinos. 2018. *Returns to investment in education: a decennial review of the global literature.* The World Bank.

Qureshi, Javaeria A. 2018. “Additional returns to investing in girls’ education: Impact on younger sibling human capital.” *The Economic Journal*, 128(616): 3285–3319.

Rivkin, Steven G, Eric A Hanushek, and John F Kain. 2005. “Teachers, schools, and academic achievement.” *Econometrica*, 73(2): 417–458.

Sahoo, Soham. 2017. “Intra-household gender disparity in school choice: Evidence from private schooling in India.” *The Journal of Development Studies*, 53(10): 1714–1730.

Sample Registration System Report. 2013. *SRS Statistical Report.* Office of the Registrar General & Census Commissioner, India.

Savage, Tim, José GB Derraik, Harriet L Miles, Fran Mouat, Wayne S Cutfield, and Paul L Hofman. 2013. “Birth order progressively affects childhood height.” *Clinical Endocrinology*, 79(3): 379–385.

Sen, Amartya, and Sunil Sengupta. 1983. “Malnutrition of rural children and the sex bias.” *Economic and political weekly*, 855–864.

Shah, Manisha, and Bryce Millett Steinberg. 2017. “Drought of opportunities: Contemporaneous and long-term impacts of rainfall shocks on human capital.” *Journal of Political Economy*, 125(2): 527–561.

Short, Jodi L, Michael W Toffel, and Andrea R Hugill. 2016. “Monitoring global supply chains.” *Strategic Management Journal*, 37(9): 1878–1897.

Slemrod, Joel. 1994. “Fixing the leak in Okun’s bucket optimal tax progressivity when avoidance can be controlled.” *Journal of Public Economics*, 55(1): 41–51.

Slemrod, Joel. 2007. “Cheating ourselves: The economics of tax evasion.” *Journal of Economic perspectives*, 21(1): 25–48.

Slemrod, Joel, and Shlomo Yitzhaki. 1985. “The optimal size of a tax collection agency.” National Bureau of Economic Research.

Stolper, Anno. 2009. “Regulation of credit rating agencies.” *Journal of Banking & Finance*, 33(7): 1266–1273.

Tantri, Prasanna L. 2017. “Impact of Third Party Audits on Compliance and Corruption: Evidence Using a Policy Experiment.” *Indian School of Business WP*, 2919091.

Velayudhan, Tejaswi. 2018. “Misallocation or Misreporting? Evidence from a Value Added Tax Notch in India.” *Unpublished Paper*.

World Bank Development Report. 2018. “Learning to Realize Education’s Promise. World Development Report 2018.” *A World Bank Group Flagship Report*.

Yoo, Sam Hyun, Sarah R Hayford, and Victor Agadjanian. 2017. “Old habits die hard? Lingering son preference in an era of normalizing sex ratios at birth in South Korea.” *Population Research and Policy Review*, 36(1): 25–54.