ESSAYS ON TAXATION AND INEQUALITY

MUHAMMAD KHUDADAD CHATTHA



St Hugh's College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy in Public Policy

Michaelmas 2020

ABSTRACT

This article-based dissertation uses administrative data to examine questions about inequality, taxation, and election administration that are important topics of study in public economics and political science. The thesis is composed of three original papers that explore different questions, explained below. Chapters 1 and 5 supplement this work by providing a brief introduction and conclusion respectively.

In Chapter 2, I develop a new methodology to estimate income inequality for developing countries. The leading existing estimation technique explained in Alvaredo, Atkinson, Piketty, Saez and Zucman (2016) is ill-suited to conduct top income adjustment in low tax compliance countries. The new estimation technique addresses this limitation by matching corresponding income quantiles in survey and tax data to conduct top income adjustment. I implement both these methodologies using a combination of administrative data of all tax returns filed in Pakistan from 2008 to 2016, and nationally representative surveys. The results firstly show the existing technique breaks down in Pakistan's low tax compliance context while the method proposed by this paper produces more robust estimates. It secondly shows that income inequality in Pakistan remained high and stable from 2008 to 2016.

In Chapter 3, I analyze the effectiveness of a tax broadening tool used in a low tax compliance, developing country context. The Board of Revenue in Pakistan regularly gave out extensions to the tax filing deadline under the assumption that time is a binding constraint to filing behavior. I conduct the analysis by using administrative data of all tax returns filed in Pakistan from 2007 to 2017 and combine this with data extracted from official government circulars containing details about deadline extensions. The results firstly show that deadline extensions are associated with individuals and unincorporated businesses delaying the filing of their tax returns by 88% and 70% of the extension duration respectively. Secondly, extensions reduce compliance of individuals and unincorporated businesses with the final extended deadline by 4.5 and 1.1 percentage points respectively. Finally, extensions do not encourage tax filing behavior and hence are ineffective in expanding the tax base.

Chapter 4 explores the impact of gender-specific polling stations on turnout. Several countries, including Pakistan, have created separate polling stations for males and females with the aim of increasing female voter turnout. The paper uses an administrative dataset of the polling station-level results from Pakistan's recent general election in 2018 to estimate the impact of separate gender polling stations on voter turnout. The results show that separate gender polling stations reduce voter turnout for females by 2 percentage points on average, and for males by 1 percentage point on average.

Together, these three articles use administrative data to contribute towards the existing literature in public economics and political science. Chapter 2 contributes towards the literature on income inequality, Chapter 3 towards tax enforcement, and Chapter 4 towards political participation. The thesis also shows how administrative data can be used to answer large policy-relevant questions.

DECLARATION

Chapter 4 of this thesis was prepared jointly with Shruti Lakhtakia. I worked in collaboration with Shruti on every aspect of this paper. Hence, my contribution to this chapter was substantial. All other chapters of this dissertation are solely my own work.

This thesis consists of approximately 56,672 words. This is calculated using the total number of words on page 6 (322) multiplied by the total number of pages (176).

ACKNOWLEDGEMENTS

I was very fortunate to have Martin Williams and Julien Labonne as my supervisors who not only provided me the right guidance but were also especially kind to me throughout my time at Oxford. A D.Phil can be quite a challenging experience without having the right support structure. Having Martin and Julien as my supportive mentors made the experience so much more rewarding and enjoyable than it would have been otherwise.

The Blavatnik School of Government played a key role in enabling my work. From the generous research funding at the school to the constant logistical support provided by the indefatigable Ruosi Zhang, the department made sure that I had whatever I needed. This support went all the way up to the dean of the school, Ngaire Woods, which made me feel at home.

The academic community at Oxford was vital in helping me refine my work and in opening up new ways of thinking for me. I would particularly like to thank Stefan Dercon, Gabriel Ulyssea, Clare Leaver, Kate Orkin, Adnan Khan, Lant Pritchett, Anandi Mani, Pepper Culpepper, Simon Quinn, Pramila Krishnan, Kate Vyborny, and the DPhil community at Oxford for their detailed feedback during various vivas and seminars. Their fresh perspective helped me see things in ways that I wouldn't have been able to on my own.

Officials at Pakistan's Federal Board of Revenue, Bureau of Statistics and Election Commission constantly supported me throughout my work. This thesis would not have been possible without their valuable insights and help in securing data access.

Two of my close friends were an important part of this journey both at a professional and personal level. Bilal Qasim, who is my civil service batchmate, engaged deeply with my research and was invested in my success. Countless discussions with him helped me in developing good research proposals and understanding the mechanisms at play. Shruti Lakhtakia, who is part of my DPhil cohort and a long time friend, provided me with detailed feedback at every step of my work. Being a doctoral candidate herself, she understood the challenges of doing a DPhil and supported me throughout the process. My DPhil journey would have been much more challenging without their friendship.

Finally, I thank my parents and siblings for their unflinching support throughout my time at Oxford. In particular, I thank my brother, Qasim, for inspiring me to achieve my dreams every day, my sister, Namrah, for teaching me about mental strength by exemplifying it in her daily life, my father, Zubair, for showing me the path to perseverance in the face of adversity and my mother, Rashida, for making me the person I am today. There were plenty of moments when I doubted myself, but there was never a moment when my family doubted me.

CONTENTS

1 INTRODUCTION **1**

2 MEASURING INCOME INEQUALITY IN DEVELOPING COUNTRIES **13**

2.1 Introduction 14

2.2 Methodology 17

2.3 Data 27

2.4 Results 31

2.5 Robustness 40

2.6 Conclusion 43

Appendix **48**

2.A Income Definition and Inequality Measures 48

2.B Alternative Estimation Techniques Considered 51

2.C Primary Income and Pensions 54

2.D Comparison of Upgrade Factors 57

2.E Pakistan Social and Living Standards Measurement Surveys 59

2.F Changes in the Selected Sample 62

2.G Income Inequality and Gender 64

2.H Top Income Adjustment and Age 67

2.I Income Shares for Urban and Rural Pakistan 69

2.J Change in Income Distribution from Top Income Adjustment 71

2.K Upgrade Factors and Income 73

3 DO DEADLINE EXTENSIONS ENCOURAGE TAX FILING? EV-

IDENCE FROM PAKISTAN **75**

3.1 Introduction . . . 76

3.2 Institutional Context and Data . . . 79

3.3 Methodology and Results . . . 88

3.4 Mechanisms . . . 103

3.5 Robustness . . . 108

3.6 Conclusion . . . 118

Appendix **122**

3.A Structure of the FBR . . . 122

3.B Relevant Definitions from Income Tax Ordinance (2001) . . . 124

3.C Summary Statistics . . . 125

3.D Extension Circulars . . . 127

4 THE GAP WITHIN: IMPACT OF SEPARATE-GENDER POLLING STATIONS ON VOTER TURNOUT **129**

4.1 Introduction . . . 129

4.2 Gender Gap and Creation of Polling Stations . . . 131

4.3 Data . . . 136

4.4 Methodology and Results . . . 137

4.5 Threats to Identification . . . 139

4.6 Mechanisms . . . 142

4.7 Discussion and Conclusion . . . 145

Appendix **151**

4.A Gender Turnout Gap . . . 151

4.B Voting Process . . . 152

4.C Index Definitions and Descriptions 154

4.D Data 157

4.E Polling Station Descriptives 158

5 CONCLUSION **159**

LIST OF FIGURES

2.1 Alvaredo et al. Leads to Counter-Intuitive Upgrade Factors for 2016 . 25

2.2 Averaged Top Income Shares Across Different Estimation Methods (2008-2016) 32

2.3 Alvaredo et al. (2016) Significantly Underestimates Income Inequality 33

2.4 High Level of Income Inequality in Pakistan Compared to Other Countries 35

2.5 Top Income Adjustment Significantly Increases Gini Coefficient . . . 36

2.6 Top Income Adjustment Modifies Income Distribution Above Threshold 38

2.7 Upgrade Factors Increase with Income Above Threshold 39

2.8 Adjustment Increases as People Enter Working Age 39

2.9 Estimates Remain Robust to Changes in Adjustment Threshold . . . 41

2.10 Estimates Remain Robust with Uniform Adjustment Thresholds . . . 41

2.11 Estimates Remain Robust to Income Percentiles as Adjustment Thresholds 42

2.12 Estimates Remain Robust to Changes in Matching Method 42

2.C.1 Pension Included in Income Definition 55

2.C.2 Pension Included in Income Definition (Continued) 56

2.D.1 Upgrade Factors for This Paper vs Alvaredo et al. (2016) 57

2.D.2 Upgrade Factors for This Paper vs Alvaredo et al. (2016) (Continued) 58

2.F.1 Including Non-Earners (Age 20+) 62

2.F.2 Alvaredo et al. (2016) Estimates with a Sample of Age 20+ 63

2.G.1 Men Dominate Earnings Throughout Income Distribution (No Attribution) 64

2.G.2 Estimates Using All Income Heads Method 65

2.G.3 Men Dominate Earnings Across Time 65

2.G.4Men Dominate Earnings Across Time 66

2.H.1Adjustment Increases as People Enter Working Age 67

2.H.2Adjustment Increases as People Enter Working Age (Continued) . . . 68

2.I.1 Urban Areas are Slightly More Unequal Compared to Rural 69

2.I.2 Top Income Adjustment Increases Gini for Urban and Rural Areas . . 70

2.I.3 Similar Levels of Unadjusted Inequality in Urban and Rural Areas . . 70

2.J.1 Top Income Adjustment Modifies Income Distribution Above Threshold 71

2.J.2 Top Income Adjustment Modifies Income Distribution Above Threshold (Continued) 72

2.K.1Upgrade Factors Increase with Income Above Threshold 73

2.K.2Upgrade Factors Increase with Income Above Threshold (Continued) 74

3.1 Pakistan's Stagnant and Low Tax-to-GDP Ratio 80

3.2 Illustration of Deadline Extensions in 2010 80

3.3 Extensions Regularly Given to Individuals and AOPs 81

3.4 Daily Tax Filing and Extensions For Individuals and AOPs 86

3.5 Daily Tax Filing and Extensions For Companies 87

3.6 Decreasing Compliance with Original Deadline by Individuals and AOPs 90

3.7 Delayed Filing by Individuals and AOPs 92

3.8 Distributions of Delayed Filing by Tax Filer Type 95

3.9 Decreasing Compliance with Final Deadline by Individuals and AOPs 96

3.10 No Obvious Association Between Extensions and Filing Rate 99

3.D.1Sample Extension Circular for Tax Year 2009 128

4.1 Pakistan has a Large Difference between Female and Male Turnout . 132

4.2 Gender Gap in Turnout in Pakistan 133

4.3 Voting Process in the 2018 General Election 136

4.4 Illustration of Main Empirical Specifications 138

4.5 Combined Polling Stations Have Higher Turnout than Separate-Gender Ones 138

4.A.1Males are More Likely to Vote in National Elections than Females . . 151

LIST OF TABLES

2.1 Thresholds used for Top Income Adjustment 20

2.2 High Adjustment Percentiles Using Alvaredo et al. (2016) 24

2.3 Lower Adjustment Percentiles Compared to Alvaredo et al. (2016) . . 26

2.4 Comparison of Alvaredo et al. (2016) with This Paper 28

2.5 Pakistan Social and Living Standards Measurement Survey (PSLM) . 29

2.6 Administrative Data of all Tax Returns Filed by Individuals 31

2.7 Difference between Estimates (Percentage Points) 34

2.8 Difference between Gini Estimates 36

2.9 Adjustment Thresholds for Robustness Check (in PKR) 40

2.E.1 Pakistan Social and Living Standards Measurement Survey (PSLM) . 60

2.E.2 Income Variables in Pakistan Social and Living Standards Measurement Surveys (PSLM) 61

3.1 Pakistan's Administrative Tax Data From 2007 to 2017 84

3.2 Extensions are Associated with Increased Probability of Delayed Filing 93

3.3 Extensions are Associated with Delayed Filing 94

3.4 Extensions are Associated with Reduced Compliance with Final Extended Deadline 97

3.5 Deadline Extensions are not Associated with Increased Probability of Tax Filing 100

3.6 Excluding Time Periods Also Shows that Extensions are not Associated with Increased Probability of Tax Filing 101

3.7 Earlier Announcements Associated with More Delays in Tax Filing . 105

3.8 Lag of Extensions Associated with Delayed Filing 107

3.9 Heterogeneities in Response to Deadline Extensions 109

3.10 Results for Delayed Filing Remain Robust With Probit Model 110

3.11 Results for Delayed Filing Remain Robust With Logit Model 111

3.12 Results for Delayed Filing Remain Robust to Time Exclusion 113

3.13 Results For Delayed Filing Remain Robust to Large Company Exclusion 116

3.14 Individual-Company Comparison with Only Eligible Individuals . . . 117

3.A.1 FBR Field Offices 123

3.C.1 Summary Statistics from Tax Data 125

3.C.2 Summary Statistics by Tax Filer Type 126

4.1 Distribution of Polling Locations with Different Polling Station 133

4.2 Distribution of Polling Station Types 134

4.3 Lower Turnout in Separate-Gender Polling Stations 140

4.4 Summary of Indices Used for Analysis 144

4.5 Results are Consistent after Including Role of Resources, Norms, Voting Together 146

4.D.1 Data Sources Used for Analysis 157

4.E.1 Distribution of Polling Stations by Polling Location 158

4.E.2 Share of Different Types of Polling Stations by Province 158

Chapter 1

INTRODUCTION

Background and Motivation

Administrative data is often hard to secure in the developing world but is being increasingly used in empirical research (Chetty 2012). This is because it allows researchers to answer ambitious academic and policy-relevant questions thanks to larger sample sizes and 'virtually no attrition' (Chetty 2012). Some recent examples of the use of administrative data in public economics include Best et al. (2015) where they use tax data from Pakistan to analyze the impact of a minimum tax scheme and Piketty et al. (2018) where they use tax data from the United States to estimate income inequality.

In similar vein, this dissertation aims to answer such questions that connect the world of academia and policy by using detailed administrative data. It is composed of three original stand-alone articles (Chapters 2, 3 and 4) that fill key empirical gaps in the income inequality, tax enforcement and political participation literatures respectively. While this thesis is not about administrative data per se, such data is a major connecting thread between all three articles. Hence, this dissertation provides illustrative examples of how administrative data can play an invaluable role in answering large questions of interest to researchers and policymakers.

Chapter 2 builds on the existing work on income inequality estimation and develops a new estimation technique that overcomes limitations of the existing method. Chapter 3 contributes towards the tax enforcement literature by analyzing the effectiveness of a previously unexplored tax-broadening policy tool of deadline extensions. Chapter 4 contributes towards the political participation literature by

analyzing the effectiveness of a previously unexplored policy instrument designed to improve female political participation. All three articles provide examples of how administrative data can be used in developing countries to answer large academic and policy-relevant questions.

This chapter provides a broad overview of the main results and the key contributions of each article. The thesis concludes with Chapter 5 that contains a discussion about policy implications and additional research questions that stem from the findings of this dissertation.

Main Results

In Chapter 2, I develop a new methodology to estimate income inequality for countries with low levels of tax compliance. The existing method explained in Alvaredo, Atkinson, Piketty, Saez and Zucman (2016) is unable to conduct top-income adjustment in developing countries with low levels of tax compliance. This is because the existing method uses the total number of tax filers to estimate the minimum threshold for top-income adjustment.[1] This method underestimates income inequality in countries with low tax filing rates because it relies on the total number of tax filers to estimate the threshold for top income adjustment. A low number of tax filers inflates the adjustment threshold for top income adjustment and hence leads to virtually no top income adjustment.

To address this gap in the literature, I create a new method that matches corresponding income quantiles of survey and tax data to conduct top-income adjustment. This method does not rely on the total number of tax filers and is able to produce inequality estimates in low tax compliance, developing country contexts.

[1]Top-income adjustment makes upward adjustments to high incomes in survey data using another data source (such as administrative tax data). This is required since surveys alone are unable to capture the top end of the income distribution (Blanchet et al., 2018).

I then apply both the existing and new methodology to Pakistan - a country with low levels of tax compliance where less than 1% of the population files their income tax return. I use an administrative dataset of all tax returns filed by individuals between 2008-2016 along with nationally representative surveys to develop income inequality estimates.

The results show that the method set out in Alvaredo et al. (2016) is ill-suited to conduct top-income adjustment and hence underestimates income inequality in Pakistan's low tax compliance setting. On the contrary, the new method works to produce robust estimates. Specifically, the share of the Top 10% is 31.57 percentage points higher using the new method compared to the estimates using Alvaredo et al. (2016) per year on average.

The estimates from the new method show that income inequality in Pakistan has remained high and stable with individuals in the Top 10% earning an average of 54.22% of the total annual income from 2008-2016. This compares to a Top 10% share of 22.65% using the Alvaredo et al. (2016) method. The results also show stark gender disparity with females only earning 4.15% of the income in the Top 10%.

In Chapter 3, I analyze the effectiveness of a tax broadening tool used in a low tax compliance, developing country context with a narrow tax base. Pakistan has struggled with a low and stagnating tax to GDP ratio hovering around 10% (IMF, 2017). To broaden Pakistan's narrow tax base and resultantly increase tax collection, the Board of Revenue in Pakistan regularly gave extensions to the income tax filing deadline under the assumption that time is a binding constraint to tax filing behavior. However, it is unclear whether such extensions are successful at improving filing rates, or whether they instead undermine the credibility of future deadlines and hence lead to even more delay among taxpayers.

The tax enforcement literature for developing countries has analyzed the effective-

ness of various policy instruments to improve tax compliance, such as third-party information reporting (Carrillo et al., 2017; Jensen, 2019; Kleven et al., 2011; Naritomi, 2018; Pomeranz, 2015). The effectiveness of deadline extensions remains unexplored by this literature which is a gap that this chapter aims to fill.

To conduct the analysis, I use an administrative data of all tax returns filed from 2007-2017 and combine this with data extracted from official government circulars containing extension announcements. The empirical analysis utilizes the variation in deadline extensions given to different types of tax filers - these extensions were frequently given to individuals and unincorporated businesses, but not to companies. The results firstly show that extensions lead individuals and unincorporated businesses to delay the filing of their tax return by 88% and 70% of the extension duration respectively. Secondly, extensions reduce the compliance of individuals and unincorporated businesses with the final extended deadline by 4.5 and 1.1 percentage points respectively. Finally, extensions do not encourage tax filing behavior.

I supplement the main findings by exploring various underlying mechanisms. I find that information provision in the form of the date of the extension announcement helps explain the results: earlier announcements are associated with more delays in tax filing. Learning also plays an important role where extensions given in one year are associated with delayed filing in the future.

Together, these findings show that extensions are ineffective in broadening the tax base as they impose additional costs, such as lost interest revenue due to delayed filing and lower compliance with government set deadlines, while not providing any benefits.

In Chapter 4, which is a co-authored article, we analyze the impact of separate-gender polling stations on voter turnout. Such polling stations have been implemented in various countries around the world, including Pakistan, with the aim of

increasing female voter turnout.[2] It is important to note that Pakistan has one of the largest gender gaps of voter turnout in the world (Inglehart et al., 2014).

The literature on female political participation in developing countries has analyzed the effectiveness of various policy instruments aimed to increase female voting, such as voter awareness and canvassing campaigns (Cheema et al., 2019; Giné and Mansuri, 2018). A key gap in the literature is that it has not analyzed the effectiveness of separate-gender polling stations - a policy in wide use by election commissions across the world. Chapter 4 addresses this gap in the literature.

To conduct the analysis, we use an administrative data of polling station level election results from Pakistan's recent general election in 2018. We find that separate-gender polling stations reduce turnout for females and males by 2 and 1 percentage points respectively when compared to mixed-gender polling stations.

We explore underlying mechanisms by combining the administrative data with multiple surveys. From this analysis, we are able to rule out the possibility that the results are driven by females and males wanting to vote together. We also show that there is no hetrogeniety in response to the policy of separate-gender polling stations along a range of socio-demographic characteristics, such as education, awareness, norms, female bargaining power and partner's characteristics. These findings point towards the results being driven by a feature of these separate-gender polling stations that do not differ by geography. While our analysis did not reveal the precise mechanism driving these results, we can rule out a range of socio-demograhic factors, and men and women wanting to vote together in the same location.

[2]Governments in Afghanistan, Algeria, Argentina, Chile, Ecuador, Egypt, Jordan, Kuwait, Lebanon, Mozambique, Pakistan, Sri Lanka, Zanzibar, among others, have adopted the use of separate-gender polling locations, polling stations and priority queues (Electoral Knowledge Network, 2013).

Contributions

Chapter 2 has three main contributions to the literature. First, it adds to the larger methodological discussion on estimating income inequality in developing countries. It does this by providing researchers with an estimation technique that works in developing countries with low tax compliance. The existing method is unable to conduct top income adjustment in such contexts because it relies on the total number of tax filers to arrive at the estimates. Chapter 2 instead introduces a novel way to combine survey and tax data that matches corresponding income quantiles in both datasets to conduct top income adjustment. The new method works better than the existing technique as it does not rely on the tax compliance of the country.

Second, it provides a blueprint to researchers estimating inequality in cases where they have access to survey and tax microdata in developing countries. Alvaredo et al. (2016) utilize survey and tax data in the form of tabulations and interpolate the income distribution from this data using the Pareto Interpolation Technique described in Blanchet et al. (2017).[3] Chapter 2 adds to this work by providing a new tool that is able to use microdata instead of tabulations, which is becoming increasingly available in the developing world.

Third, it informs the larger academic and policy debate around income inequality by being one of the first papers to provide estimates of income inequality using top income adjustment for Pakistan. By analyzing the inequality trends for various sub-groups, Chapter 2 also contributes towards the literature on hetrogeneities within the income distribution. For instance, there is a literature that focuses on inequality by gender (Atkinson et al., 2018; Blau and Kahn, 2017; Guvenen et al., 2014) and on compensation inequality (Monaco and Pierce, 2015; Pierce, 2001).

Chapter 3 has four distinct contributions. First, it adds to the tax enforcement

[3]Tax tabulations typically provide different income brackets and the total number of individuals within these brackets.

literature by analyzing an unstudied tax broadening policy that has been in consistent use in Pakistan over the last decade. The existing literature has focused on third-party information reporting as an effective enforcement tool (Carrillo et al., 2017; Jensen, 2019; Kleven et al., 2011; Naritomi, 2018; Pomeranz, 2015). Chapter 3 adds to this literature by conducting the first analysis of deadline extensions as a tool to encourage tax filing and broadening the tax base.

Second, it adds to our understanding of tax filing behavior by presenting new insights. The results show that tax filers respond to information provision, learn to change their behavior over time, prefer to file taxes close to the deadline and could potentially change their perception about the credibility of government set deadlines. These insights supplement existing work in public finance and more broadly, in work related to individual behavior.

Third, it shows that ineffective tax broadening policies implemented over large periods of time can have significant costs. In this case, these costs are forgone interest revenue due to delayed tax filing and lower credibility of government set deadlines. Chapter 3 provides evidence for these costs and consequently adds to our understanding of unintended consequences of policy choices.

Fourth, it informs the larger discussion around tax broadening in developing countries that typically struggle with narrow tax bases. By providing empirical evidence for a common tax broadening tool, this chapter takes a step towards a better understanding of tax broadening tools commonly using by tax administrations.

Chapter 4 has three distinct contributions. First, it contributes towards the literature on political participation of women in the developing world. This literature has analyzed various policy tools aimed at increasing female turnout. For instance, recent work has analyzed the effectiveness of voter canvassing and awareness campaigns in increasing female turnout (Cheema et al., 2019; Giné and Mansuri, 2018). Chapter 4 supplements this work by providing the first analysis of separate-gender

polling stations - a policy that shares the objective of increasing female turnout. It does this in a context with the one of the largest gender gaps in political participation in the world.

Second, it provides evidence for how policies targeted at one group, can have unintended consequences for other groups. While the policy of separate-gender polling stations backfired for females, it also had consequences for males who were not the target of this policy.

Third, the findings tentatively point towards the importance of policy implementation, supplementing existing work that shows the same. We are able to rule out the possibility that the results are driven by females and males wanting to vote together, or by a range of socio-demographic characteristics, such as education, awareness, norms, female bargaining power and partner's characteristics. While the paper does not provide direct evidence of the importance of policy implementations, the results tentatively point towards the possibility of implementation related challenges that are important to analyze.

Together, these three chapters fill key empirical gaps in the existing literature with the help of detailed administrative data. By making a methodological innovation that addresses existing limitations (Chapter 2) and analyzing the effectiveness of unexplored policies (Chapters 3 and 4), this dissertation furthers our understanding within the fields of income inequality, tax enforcement and political participation.

In addition to filling important gaps in the existing literature, all three articles also have the common goal of answering important policy questions that are relevant to governments and policy professionals. Chapter 2 informs the broader policy discussion around income inequality by providing estimates for Pakistan. The contribution here is twofold. First, the methodological discussion on the estimation of income inequality in the developing world is itself inextricably linked to policy. This is because the choice of the method can lead to very different estimates and hence,

very different policy choices. Second, Chapter 2 directly informs the policy discussion around income inequality not just in Pakistan, but also globally, by providing estimates for another large developing country.

Chapter 3 informs the active policy discussion around domestic revenue mobilization in the developing world. Broadening the tax base remains a key priority for governments and international institutions. Broadening the tax base has been an important feature of the discussion on tax reform in the developing world. Chapter 3 contributes to this important discussion by analyzing a unique tax broadening tool. This makes the findings of Chapter 3 relevant for tax administrations in the developing world.

Finally, Chapter 4 contributes towards the policy discussion around ways to bridge the gender gap in political participation in the developing world. With many policies that aim to bridge this gender gap, Chapter 4 provides evidence for one such policy in a setting that has one of the largest gender turnout gaps in the world. This makes the findings from this chapter directly relevant for election commissions around the world.

To summarize, this thesis uses administrative data to answer a range of academic and policy relevant questions in the field of public economics and political science. It provides a new inequality estimation technique for low tax compliance developing countries, analyzes the impact of deadline extensions as a tax broadening tool and studies the impact of separate-gender polling stations on turnout. This work fills important gaps in the literature by creating a novel estimation technique and analyzing previously unexplored policies.

References

Alvaredo, Facundo, Anthony B Atkinson, Lucas Chancel, Thomas Piketty, Emmanuel Saez, and Gabriel Zucman, "Distributional National Accounts Guidelines: Methods and Concepts Used in WID.world," *WID Working Paper*, 2016.

Atkinson, Anthony, Alessandra Casarico, and Sarah Voitchovsky, "Top incomes and the gender divide," *The Journal of Economic Inequality*, 2018, *16* (2), 225–256.

Best, Michael Carlos, Anne Brockmeyer, Henrik Jacobsen Kleven, Johannes Spinnewijn, and Mazhar Waseem, "Production versus Revenue Efficiency with Limited Tax Capacity: Theory and Evidence from Pakistan," *Journal of Political Economy*, 2015, *123* (6), 1311–1355.

Blanchet, Thomas, Juliette Fournier, and Thomas Piketty, "Generalized Pareto Curves : Theory and Applications," Technical Report 201703, World Inequality Lab 2017.

Blau, Francine and Lawrence Kahn, "The Gender Wage Gap: Extent, Trends, and Explanations," *Journal of Economic Literature*, 2017, *55* (3), 789–865.

Carrillo, Paul, Dina Pomeranz, and Monica Singhal, "Dodging the Taxman: Firm Misreporting and Limits to Tax Enforcement," *American Economic Journal: Applied Economics*, 2017, *9* (2), 144–164.

Cheema, Ali, Sarah Khan, Shandana Khan-Mohmand, and Asad Liaqat, "Canvassing the Gatekeepers: A Field Experiment to Increase Women's Electoral Turnout in Pakistan," *Working Paper*, 2019.

Chetty, Raj, "Time Trends in the Use of Administrative Data for Empirical Research," 2012.

Electoral Knowledge Network, "Systems with Separate Polling Stations for Men and Women," 2013. Reporter: Ace Project.

Giné, Xavier and Ghazala Mansuri, "Together We Will: Experimental Evidence on Female Voting Behavior in Pakistan," *American Economic Journal: Applied Economics*, 2018, *10* (1), 207–235.

Guvenen, Fatih, Greg Kaplan, and Jae Song, "The Glass Ceiling and The Paper Floor: Gender Differences among Top Earners, 1981-2012," *NBER Working Paper Series*, 2014.

Inglehart, R., C. Haerpfer, A. Moreno, C. Welzel, K. Kizilova, J. Diez-Medrano, M. Lagos, P. Norris, E. Ponarin, and B. Puranen, "World Values Survey: Round Six - Country-Pooled Datafile," 2014.

International Monetary Fund, "World Revenue Longitudinal Data," 2017.

Jensen, Anders, "Employment Structure and the Rise of the Modern Tax System," *IDEAS Working Paper Series from RePEc*, 2019.

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

Monaco, Kristen and Brooks Pierce, "Compensation inequality: evidence from the National Compensation Survey," *Monthly Labor Review*, 2015, pp. B1–B11.

Naritomi, Joana, "Consumers as Tax Auditors," *American Economic Review*, 2019, *109* (9), 3031–3072.

Pierce, Brooks, "Compensation Inequality," *The Quarterly Journal of Economics*, 2001, *116* (4), 1493–1525.

Piketty, Thomas, Emmanuel Saez, and Gabriel Zucman, "Distributional National Accounts: Methods and Estimates for the United States," *The Quarterly Journal of Economics*, 2018, *133* (2), 553–609.

Pomeranz, Dina, "No Taxation without Information: Deterrence and Self-Enforcement in the Value Added Tax," *American Economic Review*, 2015, *105* (8), 2539–2569.

Chapter 2

MEASURING INCOME INEQUALITY IN DEVELOPING COUNTRIES

Abstract

I develop a new methodology to estimate income inequality for developing countries. The leading existing method explained by Alvaredo, Atkinson, Piketty, Saez and Zucman (2016) is unable to conduct top income adjustment in countries with low levels of tax compliance. I address this issue by matching observations in survey and tax data using corresponding income quantiles for both datasets. I then use nationally representative surveys and administrative data of all tax returns filed in Pakistan from 2008 to 2016 to compare estimates from the two methods. The results show that Alvaredo et al. (2016) significantly underestimates income inequality in Pakistan's low tax compliance context, while the methodology developed by this paper is able to conduct top income adjustment to produce robust inequality estimates. The share of the Top 10% is 31.57 percentage points higher using the methodology developed by this paper compared to the estimates using Alvaredo et al. (2016) per year on average. The new estimates show that inequality in Pakistan has remained high, but stable from 2008 to 2016, with individuals in the Top 10% earning an average of 54.22% of the overall income within the country per year. They also show that there is a significant gender imbalance with females earning only 4.15% of the income in the Top 10%.

2.1 Introduction

The methodology to estimate income inequality has evolved since Kuznets' pioneering effort to measure the top income shares for the United States in 1953. These methodological improvements have allowed researchers to produce robust estimates and consequently helped in the design of economic policies (Atkinson, 2005; Atkinson and Piketty, 2007; Chancel and Piketty, 2017; Garbinti et al., 2018; Kopczuk et al., 2010; Piketty, 2003; Piketty et al., 2018, 2019)

Alvaredo, Atkinson, Piketty, Saez and Zucman (2016) synthesize these methodological improvements and provide guidelines for researchers estimating income inequality.[4] Since household surveys are unable to capture the top end of the income distribution (Blanchet et al., 2018), the estimation technique for developing countries conducts top income adjustment using administrative tax data. However, this methodology underestimates income inequality in low tax compliance contexts. This is because it relies on the total number of tax filers to estimate the threshold for top income adjustment - a low number of tax filers inflates the adjustment threshold and hence leads to virtually no top income adjustment in low tax compliance contexts.

I overcome this limitation by developing a new estimation technique that matches corresponding income quantiles in survey and administrative tax data to conduct top income adjustment. This method works in low tax compliance settings because unlike Alvaredo et al. (2016), it does not rely on the total number of tax filers. Additionally, the method relies on weaker assumptions and provides a blueprint to future researchers for combining survey and tax microdata in developing countries.[5]

I then implement this new method in Pakistan's low tax compliance setting where

[4] These guidelines provide a different methodology for developed and developing countries. This paper focuses on the methodology for developing countries.

[5] Alvaredo et al. (2016) relies on a distributional assumption that allows researchers to interpolate an income distribution by using tabulations, this paper does not rely on any distributional assumption.

less than 1% of the population filed the income tax return every year during the period of the study, and compare the results with Alvaredo et al. (2016). For this purpose, I use nationally representative survey microdata and administrative tax data of all tax returns filed by individuals from 2008 to 2016.[6]

The results show that Alvaredo et al. (2016) systematically underestimates income inequality - its estimates are lower even compared to the unadjusted estimates that only use nationally representative surveys. Specifically, the unadjusted estimates and this paper's estimates for the share of the Top 10% are 17.13 and 31.57 percentage points higher compared to Alvaredo et al. (2016) respectively. Any methodology for top income adjustment should lead to estimates that are greater than the unadjusted estimates.[7] Since Pakistan has a low level of compliance with the income tax law, the methodology proposed by Alvaredo et al. (2016) leads to these counter-intuitive results.

The new estimation technique developed by this paper significantly changes the top income shares and the Gini coefficient compared to estimates that only use the nationally representative surveys. The share of the Top 10% increases on average by 14.55 percentage points per year.[8] Similarly, top income adjustment leads to an increase of 0.11 points in the Gini coefficient on average per year.

The results from implementing the estimation technique developed by this paper show that income inequality in Pakistan remains high but stable throughout the period of the study. Individuals in the Top 10% earn an average of 54.22% of the overall income within the country per year throughout the duration of the study.[9]

[6]The tax micro data is not publicly available and was obtained from the Government of Pakistan after receiving an approval from the federal cabinet. Access to tax data in Pakistan is restricted due to data privacy concerns. The policy relevance of this research played a key role in helping secure access to the tax data.

[7]This is because survey data underestimates top incomes and the purpose of top-income adjustment is to make upward adjustment to high incomes in survey data (Blanchet et al., 2018).

[8]The share of the Middle 40% drops by 10.70 percentage points on average, while that of the Bottom 50% drops by 3.75 percentage points on average.

[9]The Middle 40% earn an average of 33.88% and the Bottom 50% earn an average of 11.90% of the total income per year.

The average Gini coefficient from 2008 to 2016 is 0.64 and generally remains within the 0.62-0.66 range. Income inequality in Pakistan is high when compared to other countries - Pakistan's income share of the Top 10% is 6.3 percentage points higher than countries at the same level of GDP per capita for 2016.[10] The results also show significant imbalance between males and females across Pakistan where females earn only 4.15% of total income within the Top 10% bracket from 2008 to 2016.[11]

This paper contributes towards the broader methodological discussion in income inequality estimation for the developing world (Alvaredo et al., 2016; Anand and Segal, 2008; Blanchet et al., 2018; Chancel and Piketty, 2017; Piketty et al., 2017; Sala-I-Martin, 2006). By addressing key limitations of the leading existing method for developing countries, this paper furthers our understanding of income inequality estimation. A second contribution of this paper is that it is one of the first papers for Pakistan that estimates income inequality using a combination of survey and tax data. Hence, the paper adds to the existing literature by providing robust estimates for an additional country and would help inform the discussion on income inequality.

The remainder of the paper is organized as follows. Section 2.2 explains the new methodology developed by this paper and compares it to the existing methodology. Section 2.3 describes the dataset used to estimate income inequality for Pakistan. Section 2.4 contains an extensive discussion of the results. Section 2.5 conducts multiple robustness checks to ensure that the main estimates are consistent. Finally, Section 2.6 concludes.

[10]To compare the estimates for Pakistan with other countries, this paper used a simple OLS regression between the share of Top 10% and GDP per capita on a sample of 105 countries which have income inequality estimates available at the World Inequality Database website (`www.wid.world`). The OLS model predicts the share of the Top 10% to be 46.6% for 2016 while the actual estimate is 52.9%. This leads to a difference of 6.3 percentage points.

[11]Females in Pakistan earned an average of 17.59% of income in the Bottom 50% bracket and 4.02% in the Middle 40% bracket from 2008 to 2016.

2.2 Methodology

This section first explains the new estimation technique and then shows how it improves upon the existing method. A detailed discussion of the income definitions and inequality measures used is given in Appendix 2.A. Alternative methodologies that were considered while developing the estimation technique for this paper are given in Appendix 2.B.

2.2.1 New Estimation Technique

To address the limitations of the existing methodology, this paper proposes a new methodology to estimate income inequality for developing countries. A formal definition of this new technique is given below.

Let Y_t^s be the ordered vector of income obtained from the survey data with N_t observations for year t with each element of this vector being (y_{ti}^s) where $i = 1, ., N_t$. Let Y_t^τ be the ordered of income obtained from the tax data with X_t observations for year t with each element of this vector being (y_{ti}^τ) where $i = 1, ., X_t$. Let $\bar{k}_t$ the minimum tax threshold for year t and n_t, the number of observations in the survey data for $Y_t^s < \bar{k}_t$. Finally, $X_t > N_t$.

For all values of $Y_t^s \geq \bar{k}_t$, divide Y_t^s into $max_t = N_t - n_t + 1$ quantiles depicted by the ordered vector Q_t. Each element of this ordered vector (q_{ta}) represents the a^{th} quantile of income in the survey data for all values of $Y_t^s \geq \bar{k}_t$.

$$Q_t = (q_{t1}, q_{t2}, ..., q_{tmax}) \tag{2.1}$$

Similarly for all values of $Y_t^\tau \geq \bar{k}_t$, divide Y_t^τ into $max_t = N_t - n_t + 1$ quantiles depicted by ordered vector R_t. Each element of this ordered vector (r_{ta}) represents

the a^{th} quantile of income in the survey data for all values of $Y_t^{\tau} \geq \bar{k}_t$.

$$R_t = (r_{t1}, r_{t2}, ..., r_{tmax}) \tag{2.2}$$

Based on Equations 2.1 and 2.2, the upgrade factor, U_t is calculated as follows.

$$U_t = \frac{R_t}{Q_t} = (\frac{r_{t1}}{q_{t1}}, \frac{r_{t2}}{q_{t2}}, ..., \frac{r_{tmax}}{q_{tmax}}) = (u_{t1}, u_{t2}, ..., u_{tmax}) \tag{2.3}$$

Using the upgrade factors calculated from Equation 2.3, the adjusted distribution of income, depicted by Y_t would be as follows.

$$Y_t = \begin{cases} Y_t^s & \text{if } Y_t^s < \bar{k}_t \\ Y_t^s * U_t & \text{if } Y_t^s \geq \bar{k}_t \end{cases} \tag{2.4}$$

Let p_{tx} be the x^{th} percentile of the adjusted income distribution, Y_t for year t. From this, the share of the Top 10%, Middle 40% and Bottom 50%, depicted by $Share_{Tt}$, $Share_{Mt}$ and $Share_{Bt}$ respectively will be calculated as follows.

$$Share_{Tt} = \frac{\int_{p_{t90}}^{p_{t100}} Y_t}{\int_{p_{t0}}^{p_{t100}} Y_t} \tag{2.5}$$

$$Share_{Mt} = \frac{\int_{p_{t50}}^{p_{t90}} Y_t}{\int_{p_{t0}}^{p_{t100}} Y_t} \tag{2.6}$$

$$Share_{Bt} = \frac{\int_{p_{t0}}^{p_{t50}} Y_t}{\int_{p_{t0}}^{p_{t100}} Y_t} \tag{2.7}$$

Implementing this methodology for all values of t would provides estimated income

shares for multiple years and would yield a trend of income inequality over time.

The abovementioned methodology has seven distinct steps that researchers would have to follow to estimate income inequality. First, researchers need to identify the minimum income tax threshold above which individuals have to file their tax return and pay tax. For instance, in 2016, this threshold was PKR 400,000 (approximately USD 3,900) per annum for Pakistan. Table 2.1 shows the minimum income tax threshold for each year. This leads to the primary assumption that this estimation technique relies on: The tax data is a better predictor of income compared to survey data above the minimum tax threshold (*Assumption A*). It is important to highlight two points regarding this assumption. First, Assumption A does not require income declarations in the tax data to be free of bias. It simply requires that the tax data is a better predictor of income above the minimum tax threshold. Second, while this paper cannot test the assumption in Pakistan's context due to data limitations, other researchers could test the assumption conditional on data availability. In ideal circumstances, nationally representative surveys could use identification numbers of surveyed individuals to link survey data with tax data perfectly. This would help test the assumption in Pakistan and other country contexts. This assumption can be justified because the costs of misreporting income in the tax returns are greater than that of surveys because of the threat of audits and penalties. The fact that we observe upgrade factors above 1 in this paper empirically shows that incomes reported in the tax data are higher than those in survey data above the adjustment threshold. It is important to mention that the existing literature on top income adjustment for developing countries also makes a similar assumption that tax data is a better predictor of income compared to survey data above a certain threshold (Alvaredo et al., 2016; Chancel and Piketty, 2017; Piketty et al., 2019).[12] Surveys underestimate higher incomes because they do not capture the tail of the income distribution well, which bolsters the case to use tax data for top income adjustment (Blanchet et al., 2018).

[12]Further details about the assumption made in the existing literature are given in Section 2.2.2.

Table 2.1: Thresholds used for Top Income Adjustment

Year	Adjustment Threshold (PKR)
2008	100,000
2011	300,000
2012	350,000
2014	400,000
2016	400,000

Notes: The adjustment threshold is the minimum amount of annual income above which individuals are liable to pay income tax. These thresholds are taken from the Income Tax Ordinance (2001), which is the law that governs income taxes in Pakistan.

Second, take the survey microdata and drop all observations that have income below the minimum tax threshold. Then use the richness of the survey microdata to divide the distribution into the maximum number of quantiles of income possible. This is equal to the number of observations in the data plus one. For instance, if the dataset has 1,000 observations for income, we can divide the dataset into a maximum of 1,001 quantiles. We will hence get 1,000 values for income by this method.

Third, take the tax microdata and drop all observations that have income below the minimum tax threshold. This is important because there are individuals below the minimum tax threshold who file their tax return. Then use the richness of the tax microdata to create the same number of income quantiles as in step 2. In our previous example, the survey data was divided into 1,001 quantiles. Hence, the tax microdata should also be divided into 1,001 quantiles to obtain 1,000 values for income.

Fourth, using the output from steps 2 and 3, calculate the ratio of tax to survey income for each corresponding quantile. This will provide the upgrade factor. In our example, we will get 1,000 upgrade factors corresponding to different levels of income in the survey data.

Fifth, take the average of the upgrade factors from the previous step for a given level of survey income. For instance, the income value PKR 410,000 might appear several times in the survey dataset. This same value might be assigned slightly different

upgrade factors when it is matched with values from the tax dataset. Taking an average will provide an upgrade factor for each unique value of survey income.

Sixth, merge the upgrade factors for each unique value of survey income above the minimum tax threshold with the original survey microdata. The survey microdata will now contain information on income and their associated upgrade factors.

Seventh, multiply the income with the associated upgrade factors. Since there is no adjustment for incomes below the minimum tax threshold, the associated upgrade factor for incomes below the threshold is 1 by default. This step yields the distribution of adjusted income. Finally, repeat this for all years and calculate preferred measure of adjusted income inequality.

2.2.2 Existing Estimation Technique

The latest technique to estimate income inequality is explained in Alvaredo et al. (2016) where the authors present various methods depending on the extent of data access and coverage. For developed countries, the authors recommend using income declarations to estimate inequality because they have rich tax microdata that covers most of the population. Developing countries, however, suffer from two limitations due to which tax data alone cannot be used. First, the coverage of the tax data is limited. Hence, any findings that simply rely on tax data would only apply to high earners and not the whole population. Second, tax microdata is often not available in developing countries. There are, however, cases in which tabulations of tax filers according to different brackets of income are available. To address these issues, Alvaredo et al. (2016) develop an estimation technique when researchers have access to survey and tax tabulations. An explanation of this technique is given below.

First, using tax tabulations and population data, researchers should calculate the

percentage of individuals in the total population above 20 years of age who are in the tax tabulations. For instance, if 5 individuals are included in the tax tabulations out of a total population of 100, then this percentage would be 5%. Subtract this from 100% and let this be p_2. In our example, p_2 would equal 95. This step yields three foundational assumptions of the estimation technique, which are given below.

Assumption 1: The tax data is an accurate predictor of income above percentile p_2 in the survey data. In our example, the tax data would be an accurate predictor of income above the 95^{th} percentile.

Assumption 2: The survey data is an accurate predictor below an arbitrarily chosen percentile p_1. This is usually assumed to be the 80^{th} percentile of the survey data.

Assumption 3: Between p_1 and p_2, incomes are higher than the survey data, and lower than the tax data. Together, assumptions 2 and 3 help create a continuous final distribution of income. This will become apparent in the next steps.

Second, use survey income tabulations that are consistent with definitions of income in National Accounts to obtain the income distribution. This is done by applying the Generalized Pareto Interpolation technique developed by Blanchet et al. (2017) which uses tabulations to provide the distribution by assuming a Generalized Pareto Distribution. Third, based on Assumption 1, calculate the ratio of tax to survey income from the aforementioned income distributions above percentile p_2. This ratio is then used to upgrade survey incomes above percentile p_2. Fourth, between p_1 and p_2, increase the upgrade ratio from 1 at p_1, to the observed ratio at p_2. For instance, the ratio can be increased in a linear manner. This is done so that there is no discontinuity in the income distribution at p_2, resulting from the survey incomes above percentile p_2 being upgraded. Assumptions 2 and 3 are useful in resolving this potential discontinuity. Fifth, reconcile these figures with macro statistics from National Accounts. The paper will not go into further details of this step as it is beyond the scope of this paper and the methodological contribution of this paper

pertains to previous steps. Finally, repeat this for all years and calculate an adjusted and preferred measure of income inequality.

2.2.3 Improvement Over Existing Estimation Technique

The method proposed by this paper improves upon three key limitations of the method given in Alvaredo et al. (2016). These key limitations, along with the improvements are explained below.

First, the existing estimation technique relies heavily on Assumption 1 which is problematic and can be contested. The following example illustrates this: Assume that there are 5 individuals who file their tax return out of a total population of 100. It would be incorrect to assume that the incomes of these 5 individuals represent that of everyone in the top 5% of the overall population. These individuals could be drawn from anywhere in the income distribution, which depends on the nature of tax coverage and compliance in a country. In countries like Pakistan, where compliance with the income tax law is low, this poses a serious challenge to implementing the methodology recommended by Alvaredo et al. (2016). Since the number of filers above the minimum tax threshold are low, they represent a very small share of the overall population and lead to an inflated estimate for p_2. This can be seen from Table 2.2 where the number for p_2 is above the 99^{th} percentile. Using this adjustment threshold on the data for Pakistan, the upgrade factor turns out to be less than 1, implying that the survey incomes above the adjustment threshold are greater than the income declared in the tax data. This practically means no top income adjustment. Note that 'adjustment percentile' is the percentile of income in the survey data above which top income adjustment is conducted. The methodology proposed by this paper uses the minimum tax threshold as the adjustment percentile. On the other hand, Alvaredo et al. (2016) estimate the adjustment percentile (referred to as p_2 above) using the total number of tax filers as a proportion of total population

above the age of 20.

Table 2.2: High Adjustment Percentiles Using Alvaredo et al. (2016)

Year	Population (Age 20+)	Filers Above Threshold	Adjustment Percentile
2008	86,389,420	473,047	99.45
2011	94,917,930	211,730	99.78
2012	97,943,922	210,135	99.79
2014	104,288,388	521,225	99.50
2016	109,569,018	815,875	99.26

Source: UN Population Data and Administrative Tax Data.

Notes: The population data is from the United Nations population database while the number of filers above the minimum tax threshold is from the administrative dataset of all tax returns filed by individuals. These two numbers are used to calculate the adjustment percentile that Alvaredo et al. (2016) recommend using in order to estimate income inequality for developing countries.

The estimation technique recommended by this paper overcomes this limitation because it does not rely on estimating the adjustment percentile in the same way. Instead, it relies on the minimum tax threshold as the adjustment threshold and then divides the tax data into the maximum number of quantiles allowed by the survey data.[13] All that is required from here is that these quantiles from the tax data are a better representation of the income distribution compared to the same quantiles from the survey data. Whether compliance with the income tax law is high or low, any available tax microdata can be divided into a certain number of quantiles and used for top income adjustment. Hence, the methodology proposed by this paper does not rely on high levels of compliance and simply requires that even if compliance levels are low, the quantiles extracted from the tax data represent the income distribution in a better manner compared to the survey data.

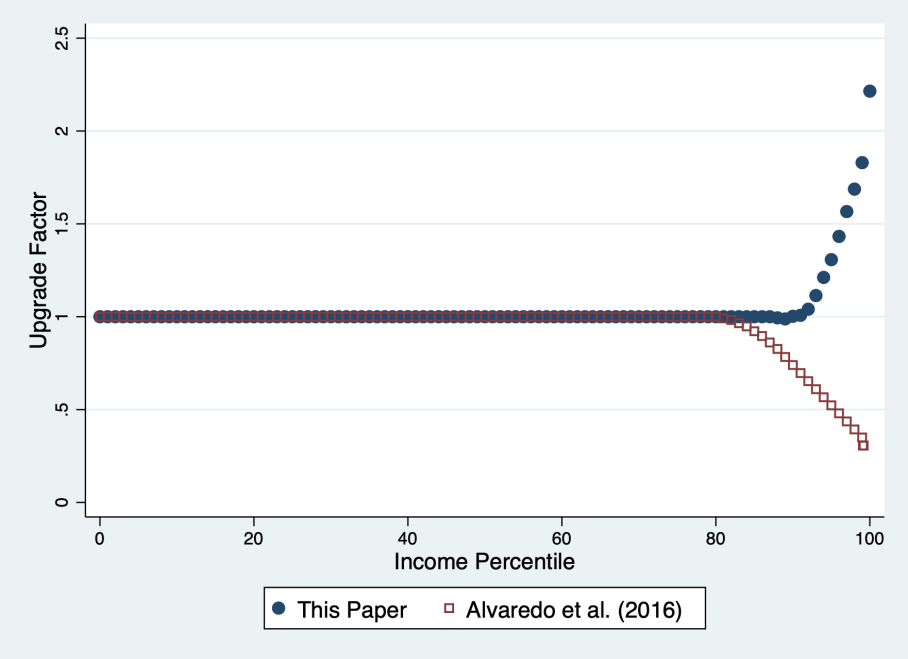
Figure 2.1 illustrates this point in greater detail by presenting a comparison of the upgrade factors using the methodology developed by this paper with the ones using Alvaredo et al. (2016) for the year 2016.[14] For any top income adjustment to take place, the upgrade factors need to be above 1. Implementing Alvaredo et al. (2016)

[13] Assumption A uses the tax threshold set by the income tax law rather than the total number of tax filers as used in Assumption 1 for the adjustment threshold to conduct top income adjustment.

[14] Graphs similar to Figures 2.1 for all other years can be found in Appendix 2.D.

in Pakistan's context results in upgrade factors below 1, which is counter-intuitive.[15] At the same time, the methodology presented by this paper leads to significant top income adjustment. Hence, for countries that have low levels income tax compliance, the methodology recommended by Alvaredo et al. (2016) cannot be implemented.

Figure 2.1: Alvaredo et al. Leads to Counter-Intuitive Upgrade Factors for 2016



The adjustment percentiles from the methodology proposed by this paper are lower in comparison to that of Alvaredo et al. (2016) leading to significant top income adjustment.[16] This can be seen from Table 2.3 while applying the methodology to Pakistan. In settings where compliance with the income tax law is low, the methodology proposed by this paper gets around the issue of low compliance.

Second, Assumptions 2 and 3 are arbitrary, and are in place to ensure that there is a continuous distribution of income when survey and tax data are combined. Since the upgrade factors calculated by the methodology jumps above 1 at p_2, this creates a discontinuity in the income distribution. Such a distribution would not depict the

[15] The upgrade factor is multiplied with the corresponding income. Hence, an upgrade factor of 1 leads to no top income adjustment, greater than 1 leads to adjustment of the income upwards, and less than 1 leads to adjustment of income downwards. The main of point of using administrative tax data is to adjust survey income upwards. This makes any upgrade factors less than 1 counter-intuitive.

[16] Alvaredo et al. (2016) refer to their adjustment percentile as p_2 which is explained in Section 2.2.2.

Table 2.3: Lower Adjustment Percentiles Compared to Alvaredo et al. (2016)

Year	**Alvaredo et al. (2016)**	**This Paper (Income>0)**	**This Paper (Age 20+)**
2008	99.45	74.90	86.06
2011	99.78	91.90	95.69
2012	99.79	92.22	95.72
2014	99.50	91.50	95.13
2016	99.26	87.80	93.13

Notes: The adjustment percentile using the method by Alvaredo et al. (2016) is estimated using the population above the age of 20 and the number of tax filers above the tax threshold. The details of the calculation can be found in Table 2.2. This paper calculates the adjustment percentile based on the sample of earning individuals (Income>0). For comparison, the corresponding adjustment percentile for this paper's methodology using a sample of individuals above the age of 20 is also given. Alvaredo et al. (2016) refer to their adjusmtent percentile as p_2.

real income distribution which is expected to be continuous. Hence, Assumptions 2 and 3 play a vital role in making the distribution realistic. However, this comes at the cost of the arbitrariness of these assumptions. Researchers make sure that these two assumptions do not matter by testing the sensitivity of the final estimates to them. However, the arbitrary nature of these assumptions means that they can be contested.

Assumption A from the new estimation technique addresses the concerns about Assumptions 1, 2 and 3 from the existing method. Assumption A simply requires the tax data to be a better predictor of income above the minimum tax threshold compared to survey data. This is a weaker assumption when compared with Assumptions 1, 2 and 3 from the previous method. Since, developing countries often struggle with tax evasion, under-reporting and non-filing, Assumption A would only lead to an underestimate of income inequality. The main limitation of the new estimation technique is that tax data may be underreported in the presence of tax evasion. This might lead to an underestimate of income inequality. It is important to mention that the existing estimation technique also has this limitation, in addition to others described above. Additionally, the upgrade factors in the new estimation technique gradually increase as survey income increases above the minimum tax threshold. As a result, Assumptions 2 and 3 are not required to ensure

that there is a continuous distribution of income after the adjustment from the tax data has been done.

Third, the existing method does not address cases in which researchers have access to rich tax and survey microdata instead of tabulations. A separate methodology that takes advantage of the richness of microdata would help create more robust estimates for developing countries. Such a method would become increasingly useful as developing countries improve their data quality and access.

This paper's estimation technique uses the extensive variation in income available in the survey and tax microdata to provide more robust estimates of income inequality. While the existing estimation technique relied on an explicit distributional assumption about income in order to interpolate the distribution from tabulations, the technique recommended by this paper does not require any such assumption. Instead, the new technique matches each observation in the survey data above the minimum tax threshold with tax data to calculate upgrade factors for every unique value of income in the survey data.

A brief comparison of the methodology proposed by this paper and that in Alvaredo et al. (2016) is given in Table 2.4. This table provides a short summary of the discussion in this section.

2.3 Data

2.3.1 Nationally Representative Household Surveys

The paper uses two sources of data to estimate top income shares. The first of these are nationally representative household surveys conducted by Pakistan Bureau of Statistics (called Pakistan Social and Living Standards Measurement (PSLM)). These surveys measure a variety of socio-economic indicators and have a dedicated

Table 2.4: Comparison of Alvaredo et al. (2016) with This Paper

	Alvaredo et al. (2016)	**This Paper**
Assumptions	1. The tax data is an accurate predictor of income above percentile p_2 in the survey data. 2. The survey data is an accurate predictor below an arbitrarily chosen percentile p_1. 3. Between p_1 and p_2, incomes are higher than the survey data, and lower than the tax data.	1. The tax data is a better predictor of income compared to survey data above the minimum tax threshold.
Data Required	1. Nationally representative survey tabulations. 2. Administrative tax data tabulations.	1. Nationally representative survey microdata. 2. Administrative tax microdata.
Top Income Adjustment	1. Uses the number of tax filers as a percentage of population to estimate adjustment threshold, p_2, which is sensitive to compliance with tax law. 2. Compares income in survey data and tax data above p_2 to estimate upgrade factor. 3. Linearly increases upgrade factor from p_1 to p_2.	1. Uses the minimum tax threshold as the adjustment threshold. 2. Matches survey and tax data by using the order of income in both datasets. 3. Uses matched observations to estimate upgrade factor.
Distributional Assumption	Yes	No

Notes: This table provides a brief comparison between the methodology for estimating income inequality for developing countries proposed in Alvaredo et al. (2016) with the methodology proposed by this paper. The row titled 'Top Income Adjustment' provides a shortened version of the two methodologies. For a detailed description of these estimation techniques, please refer to Section 2.2.

section on income data. The paper uses the income and biographical sections of these surveys in conducting the analysis.

The PSLM is structured such that it alternates annually between being conducted at a district-level and a province-level. This paper uses the province-level surveys to estimate income inequality. These surveys have an average of 17,978 households per year. For a detailed discussion on the reasons for using the province-level surveys instead of the district-level surveys, please refer to Appendix 2.E. The number of observations and households for each year is given in the Table 2.5.

Table 2.5: Pakistan Social and Living Standards Measurement Survey (PSLM)

Year	**Observations**	**Households**
2008	107,635	15,512
2011	108,937	16,342
2012	106,497	15,807
2014	119,019	17,989
2016	157,636	24,238

Notes: The PSLM are nationally representative household surveys composed of two types that alternate every year. These two types are called the district and province-level surveys. These surveys measure slightly different development outcomes with some overlapping variables. This table only shows the province-level surveys that were used to develop the estimates. For details about the district-level surveys and an explanation of the reasons for using the province-level surveys, please refer to Appendix 2.E.

The household surveys measure a wide range of variables for socioeconomic outcomes. Of these, there are four categories of variables that are relevant to estimate income inequality. The first category is that of income variables which disaggregate income according to different sources, such as primary income, pension, rent and dividends. The second category contains information on the biographical characteristics of individuals in the survey, such as age, gender and marital status. The third category has location data such as province, and the urban/rural classification of the place the individual lives in. The final category includes information on the household identification number, and the individual identification number within the household.

2.3.2 Tax returns from 2008 to 2016

The second source of data is an administrative dataset of all tax returns filed by individuals in Pakistan from 2008 to 2016. Access to tax data in Pakistan is extremely restricted due to data privacy concerns. This dataset was acquired after receiving an approval from the federal cabinet of the Government of Pakistan.[17]

The key variable of interest in the dataset is the total income declared by all tax filing individuals. The Income Tax Ordinance (2001), which is the law governing income taxes in Pakistan, defines total income in Section 10 as the sum of all income, both taxable and exempt. This declaration of total income is used in combination with the income data from the household surveys mentioned previously to develop estimates for income inequality. The dataset also contains a variety of other information for each tax return, including the tax filing year, tax paid, tax credits claimed, city of tax filer, the tax office jurisdiction relevant to the tax filer, taxpayer's principle activity and wealth declaration.[18]

Since the household surveys only have a consistent and disaggregated version of income in the province-level surveys, the tax data utilized for this paper is also of the same years (conducted in 2008, 2011, 2014 and 2016). One observation in the tax dataset represents the tax return filed by one individual. The total number of observations per year in the tax dataset are given in Table 2.6.

[17] To ensure that the dataset did not violate any data privacy laws, the tax identification numbers of tax filers were anonymized by officials at the Federal Board of Revenue, which is Pakistan's federal tax collection agency.

[18] From tax year 2014 onward, the Government of Pakistan legislated that anyone who files the tax return would also be required to file a wealth statement as per Section 116 of the Income Tax Ordinance (2001). Hence, the dataset also has information on the total asset declaration of a tax filer as per their wealth statement. However, this data is only available from 2014 onwards.

Table 2.6: Administrative Data of all Tax Returns Filed by Individuals

Tax Year	Total Observations
2008	715,784
2011	763,572
2012	768,825
2014	1,040,560
2016	1,356,057
Total	**4,644,798**

Notes: One observation represents a unique tax return filed by an individual for a particular tax year. An individual can only file the tax return once per tax year.

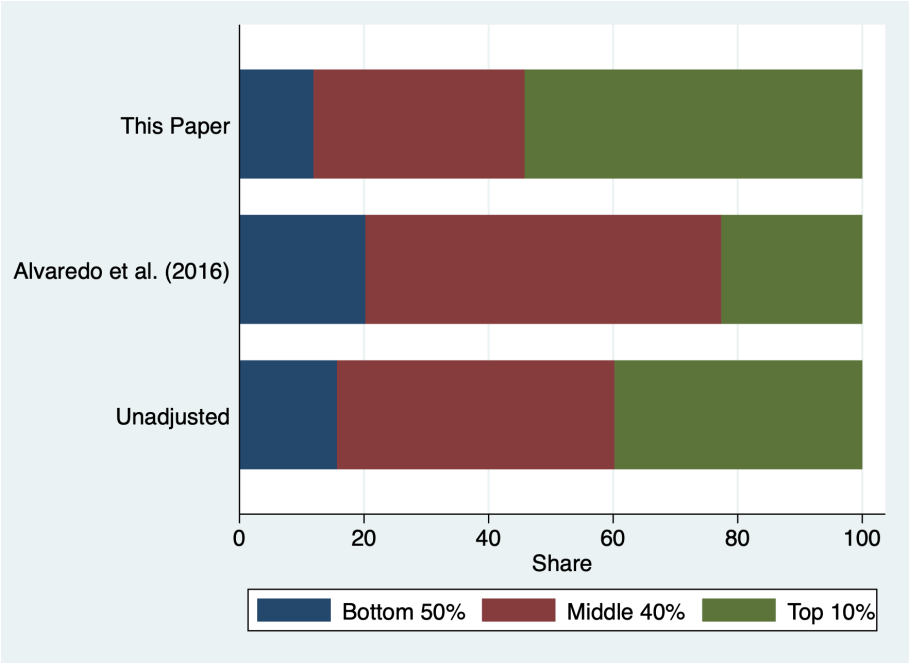
2.4 Results

2.4.1 Income Inequality in Pakistan from 2008 to 2016

This section presents the estimates for income inequality in Pakistan after implementing the new methodology explained in Section 2.2 and contrasts them with the estimates using the method laid out in Alvaredo et al. (2016). It starts by discussing the overall trend of top income shares and the Gini coefficient in Pakistan from 2008 to 2016, and then discusses multiple heterogeneities. Figure 2.2 shows how the estimates of the top income shares change when using the methodology developed by this paper, in contrast to the methodology given in Alvaredo et al. (2016). The unadjusted graph shows the trend of top income shares by only using the nationally representative household surveys, the Alvaredo et at. (2016) graph implements the existing methodology explained in that paper and 'This Paper' implements the new estimation technique explained in Section 2.2.

The figure shows that method laid out by Alvaredo et al. (2016) significantly underestimates income inequality even when compared to the unadjusted estimates. Specifically, the share of the Top 10% is 31.57 percentage points higher using the methodology developed by this paper compared to the estimates using Alvaredo et al. (2016) per year on average. At the same time, the unadjusted estimates that

Figure 2.2: Averaged Top Income Shares Across Different Estimation Methods (2008-2016)

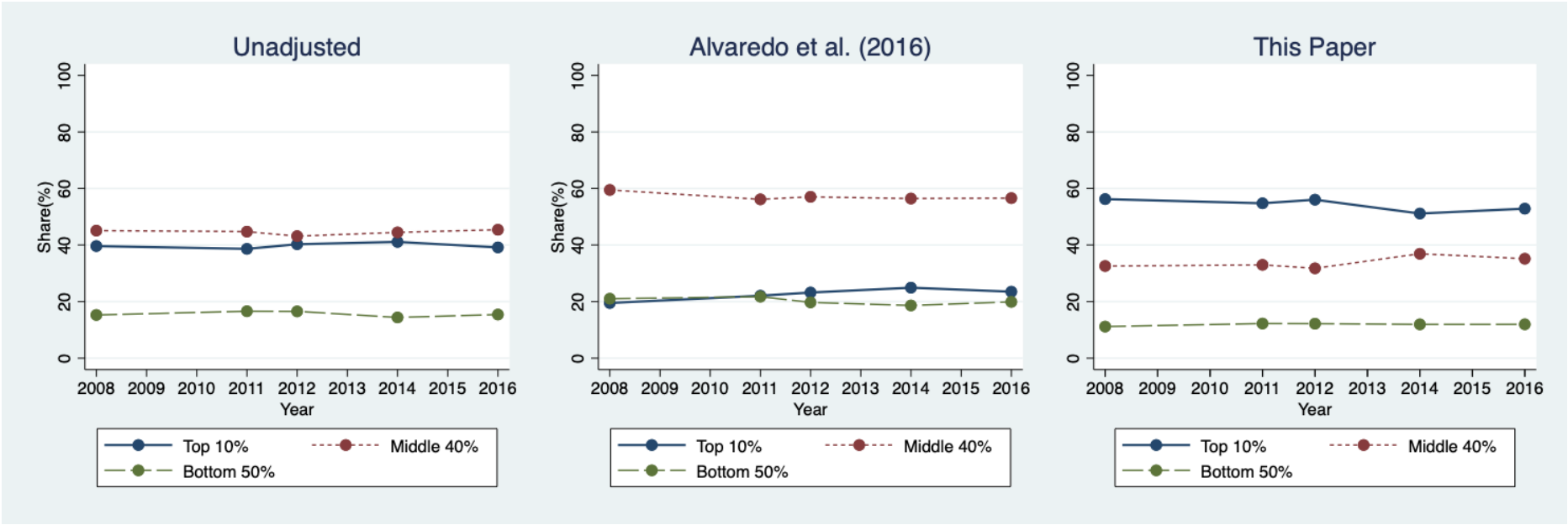


simply use nationally representative surveys have a share of the Top 10% that is 17.13 percentage points higher on average per year compared to Alvaredo et al. (2016). Any methodology for top income adjustment should lead to estimates that are greater than the unadjusted estimates. Since Pakistan has a low level of compliance with the income tax law, the methodology proposed by Alvaredo et al. (2016) breaks down and leads to these counter-intuitive results.[19] A detailed time-series of the top income share estimates using the three methods is given in Figure 2.3. The difference between this paper's income shares and the ones from Alvaredo et al. (2016) are given in Table 2.7.

When compared to the unadjusted estimates, the estimation technique developed by this paper increases the share of the Top 10% by 14.55 percentage points per year on average from 2008 to 2016. The share of the Middle 40% drops by 10.70 percentage points on average, while that of the Bottom 50% drops by 3.75 percentage points

[19]The estimates using Alvaredo et al. (2016) are lower compared to the unadjusted estimates because there is low level of compliance with the income tax law in Pakistan. Lower number of tax filers leads to an inflated adjustment percentile and upgrade factors that are less than 1. For an extensive discussion of these upgrade factors and adjustment percentiles, please refer to Section 2.2.

Figure 2.3: Alvaredo et al. (2016) Significantly Underestimates Income Inequality



Notes: For comparison, the Alvaredo et al. (2016) estimates for this graph use a sample of earning individuals ($income > 0$). Alvaredo et al. (2016) prefer to use a sample of individuals above the age of 20 which includes non-earners. Appendix 2.F shares the results of the methodology presented by this paper and that of Alvaredo et al. (2016) using a sample of individuals above the age of 20. In this case as well, Alvaredo et al. (2016) underestimates income inequality even in comparison to the simple estimates using survey data.

on average. Hence, the increase in the share of the Top 10% comes primarily at the expense of the Middle 40%, and also from the Bottom 50% in a minor way. The difference between this paper's income shares and the unadjusted ones is given in Table 2.7.

Table 2.7: Difference between Estimates (Percentage Points)

Panel A: This Paper vs Alvaredo et al. (2016)

Top Income Share	**2008**	**2011**	**2012**	**2014**	**2016**
Top 10%	36.75	32.68	32.85	26.23	29.36
Middle 40%	-26.91	-23.19	-25.30	-19.54	-21.42
Bottom 50%	-9.84	-9.49	-7.55	-6.69	-7.94

Panel B: This Paper vs Unadjusted

Top Income Share	**2008**	**2011**	**2012**	**2014**	**2016**
Top 10%	16.62	16.11	15.76	10.02	13.71
Middle 40%	-12.53	-11.75	-11.39	-7.57	-10.24
Bottom 50%	-4.09	-4.36	-4.37	-2.45	-3.47

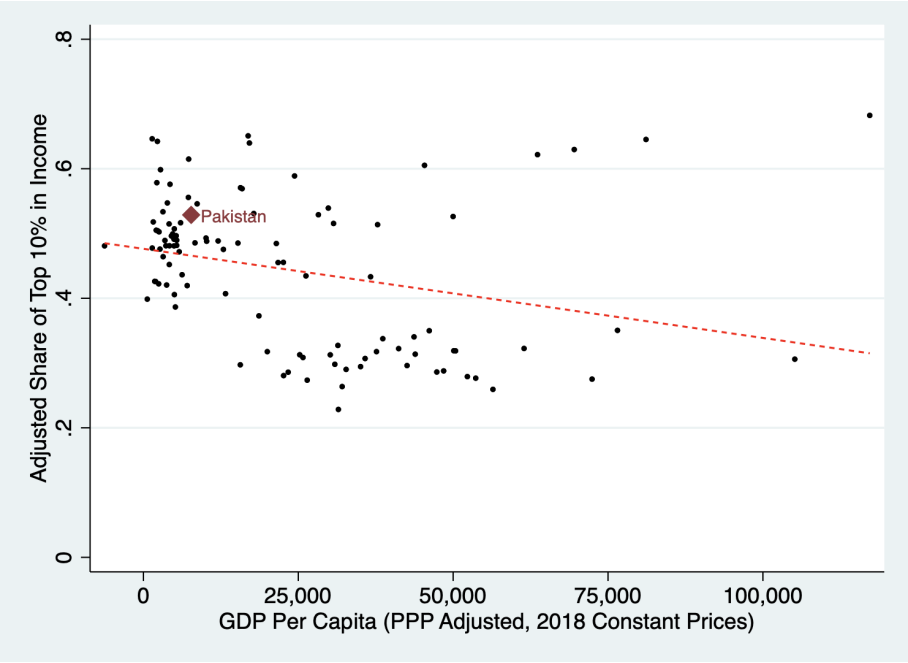
Notes: The difference in the income shares is calculated by subtracting this paper's income shares with the estimates using Alvaredo et al. (2016) in Panel A, and subtracting this paper's income shares with the unadjusted ones in Panel B. 'This Paper' implements the methodology given in Section 2.2 to conduct top income adjustment, the Alvaredo et al. (2016) estimates implements the existing methodology and the unadjusted shares simply use nationally representative surveys.

This paper's estimates given in Figure 2.3 show that income inequality in Pakistan remained high, but stable from 2008 to 2016. Individuals in the Top 10% earn an average of 54.22% of the overall income within the country per year throughout the duration of the study. On the other hand, the Middle 40% earn an average of 33.88% and the Bottom 50% earn an average of 11.90% of the total income per year. Pakistan's income inequality has remained high when compared to other countries for which estimates have been developed. Figure 2.4 shows Pakistan's income share of Top 10% in comparison to other countries. This figure uses data from the World Inequality Database and adds the estimates for Pakistan's income inequality from this paper.[20] The figure shows firstly, that Pakistan's share of the top 10% is high

[20] All estimates shown in Figure 2.4 use a consistent pre-tax definition of income and hence can be compared.

compared to other countries, and secondly, that this income share is high even when compared to countries at a similar GDP per capita.

Figure 2.4: High Level of Income Inequality in Pakistan Compared to Other Countries

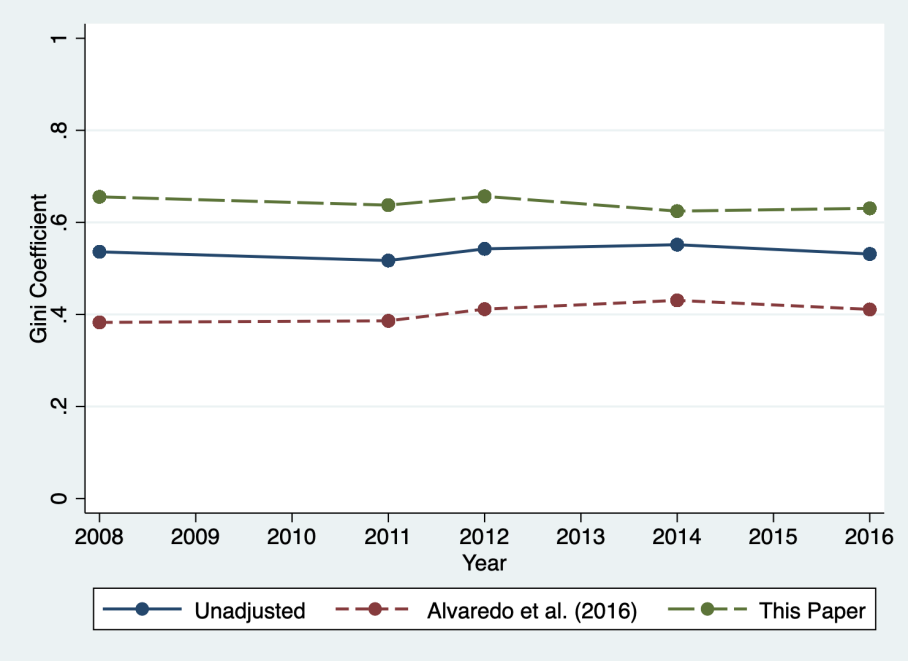


Source: World Inequality Database Combined With This Paper's Estimates

The Gini coefficient is another way to quantify income inequality. Figure 2.5 shows the unadjusted, Alvaredo et al. (2016) and this paper's estimates for the Gini coefficient in Pakistan from 2008 to 2016. Similar to the results for the top income shares, Alvaredo et al. (2016) leads to estimates of the Gini coefficient that are lower than both the unadjusted and this paper's estimates. The Gini coefficient is 0.24 points higher on average using the methodology presented by this paper compared to Alvaredo et al. (2016). Additionally, the methodology proposed by this paper leads to a significant increase of 0.11 points in the Gini coefficient on average per year when compared to the unadjusted estimates. Table 2.8 shows the difference between this paper's and Alvaredo et al. (2016)/unadjusted Gini coefficient for different years.

The trend of the Gini coefficient corroborates the findings from the trend of top income shares showing that income inequality in Pakistan remained high, but stable.

Figure 2.5: Top Income Adjustment Significantly Increases Gini Coefficient



Notes: For comparison, the Alvaredo et al. (2016) estimates for this graph use a sample of earning individuals ($income > 0$).

Table 2.8: Difference between Gini Estimates

	2008	**2011**	**2012**	**2014**	**2016**	**Average**
This Paper vs Alvaredo et al. (2016)	0.27	0.25	0.25	0.19	0.22	0.24
This Paper vs Unadjusted	0.12	0.12	0.11	0.07	0.10	0.11

Notes: The difference in the Gini coefficient is calculated by subtracting the Gini coefficient from this paper with the Alvaredo et al. (2016)/unadjusted estimates for each year. 'This Paper' implements the methodology given in Section 2.2 to conduct top income adjustment, the Alvaredo et al. (2016) estimates implements the existing methodology and the unadjusted shares simply use nationally representative surveys.

The average Gini coefficient from 2008 to 2016 is 0.64 and generally remains within the 0.62-0.66 range. The magnitude of the Gini coefficient is high when compared to other countries with similar characteristics.

The paper also conducts an analysis of various heterogeneities with the overall estimates of income inequality. For a detailed discussion of income inequality by gender, please refer to Appendix 2.G. This appendix shows stark gender disparities where females in Pakistan only earn 4.15% of the overall income in the Top 10%. Appendix 2.I presents the estimates of income inequality separately for urban and rural Pakistan.

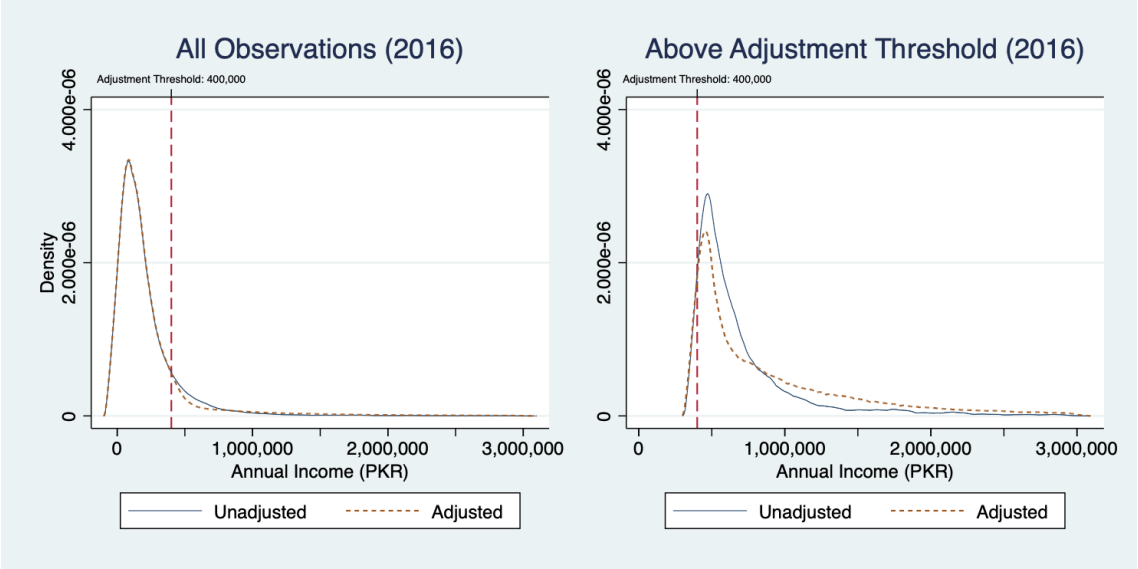
2.4.2 Change in the Income Distribution

The estimation technique described in Section 2.2 uses the income distribution from survey data and conducts top income adjustment above the minimum income tax threshold. This methodology modifies the income distribution from the survey data such that the top incomes are higher. Figure 2.6 shows the unadjusted and adjusted distribution of income for 2016 with a full sample and a restricted sample of observations above the adjustment threshold.[21] This figure shows that the distribution changes above the tax threshold such that the density of higher incomes rise while that of lower incomes closer to the threshold decreases. This can be seen in greater detail in the restricted sample version, which shows the 2016 unadjusted and adjusted income distribution for a restricted sample of observations that are above the tax threshold. These two graphs in combination show how the distribution below the tax threshold remains similar with and without adjustment, while the distribution above the adjustment threshold changes such that the density of higher incomes increases while that of lower incomes close to the tax threshold decreases. Graphs similar to Figures 2.6 for all other years can be found in Appendix 2.J which show

[21] The graph with the restricted sample starts below the adjustment threshold because the kdens function in Stata plots a distribution graph that can start below the minimum value in the dataset.

similar changes in the income distribution.

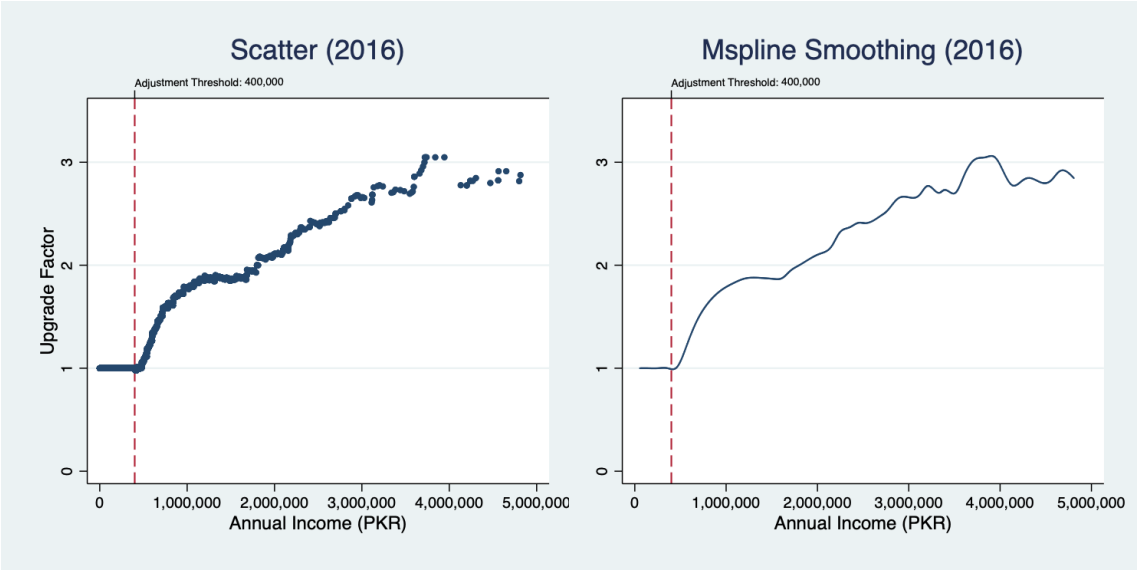
Figure 2.6: Top Income Adjustment Modifies Income Distribution Above Threshold



The top income adjustment is done such that each observation in the survey data above the minimum income tax threshold is assigned an adjustment factor that is calculated from a combination of the nationally representative surveys and the administrative tax data. Figure 2.7 shows the change in adjustment factors with income for 2016. The figure shows that the adjustment factor is 1 for all incomes below the adjustment threshold which implies that there is no adjustment to these incomes. The adjustment factor gradually increases with income leading to a higher adjustment for higher incomes. This is consistent with the existing literature where the adjustment factor generally rises with income. Graphs similar to Figure 2.7 for all other years can be found in Appendix 2.K which show a similar trend of adjustment factors.

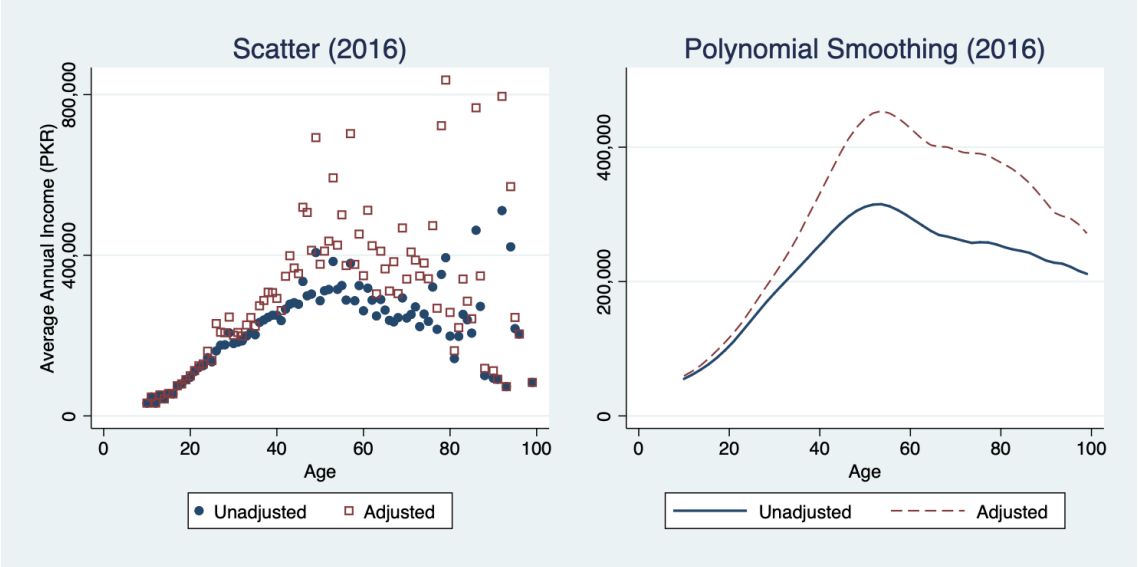
The top income adjustment upgrades the income of individuals differentially across age. This can be seen from Figure 2.8 which depicts a scatterplot of the average adjusted and unadjusted income with age for 2016. The figure also shows the same graph with polynomial smoothing instead of a scatterplot for ease of analysis. This figure shows that the difference between adjusted and unadjusted income first rises

Figure 2.7: Upgrade Factors Increase with Income Above Threshold



with age and then later decreases. This is expected because as people enter into working age they earn more, and subsequently experience a decline in their overall earnings later in life. Appendix 2.H contains similar graphs showing a similar trend across different years.

Figure 2.8: Adjustment Increases as People Enter Working Age



2.5 Robustness

To ensure that the estimates for income inequality are robust, the paper conducts several checks. Since the estimation technique proposed by this paper relies on one assumption, this section first tests the sensitivity of the results to this assumption in multiple ways. The section then tests the robustness of the results to modifications in the process of matching the survey and tax data.

To recap, the estimation technique relies on the assumption that the tax data is a better predictor of income above the minimum tax threshold as compared to survey data.[22] As a first check, the paper changes the minimum adjustment threshold in four different ways: increase by 10%, increase by 20%, decrease by 10% and decrease by 20% for the threshold of each respective year. Table 2.9 shows the original adjustment threshold along with the four versions of the adjustment thresholds used for this robustness tests. Figure 2.9 shows how the trend of the Top 10%, Middle 40% and Bottom 50% changes when the adjustment threshold is changed. The graphs also show the original estimates from the main results in Section 2.4 for comparison. The figures show that the results remain robust to changes in the adjustment threshold for each year.

Table 2.9: Adjustment Thresholds for Robustness Check (in PKR)

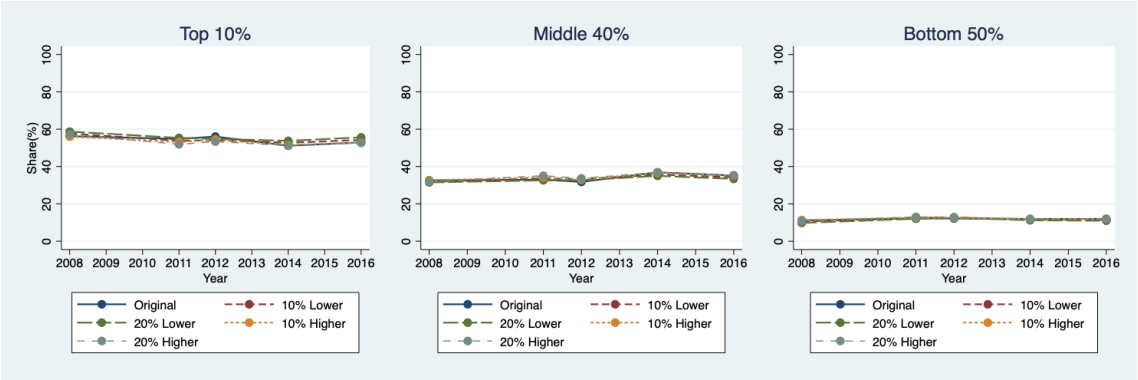
Year	Original Threshold	10% Less	20% Less	10% More	20% More
2016	400,000	360,000	320,000	440,000	480,000
2014	400,000	360,000	320,000	440,000	480,000
2012	350,000	315,000	280,000	385,000	420,000
2011	300,000	270,000	240,000	330,000	360,000
2008	100,000	90,000	80,000	110,000	120,000

Notes: The original adjustment threshold is the minimum amount of annual income above which individuals are liable to pay income tax. These thresholds are taken from the Income Tax Ordinance (2001), which is the law that governs income taxes in Pakistan. The threshold is increased and decreased by 10% and 20% for this robustness check.

A possible source of bias in the estimates is that the minimum adjustment threshold

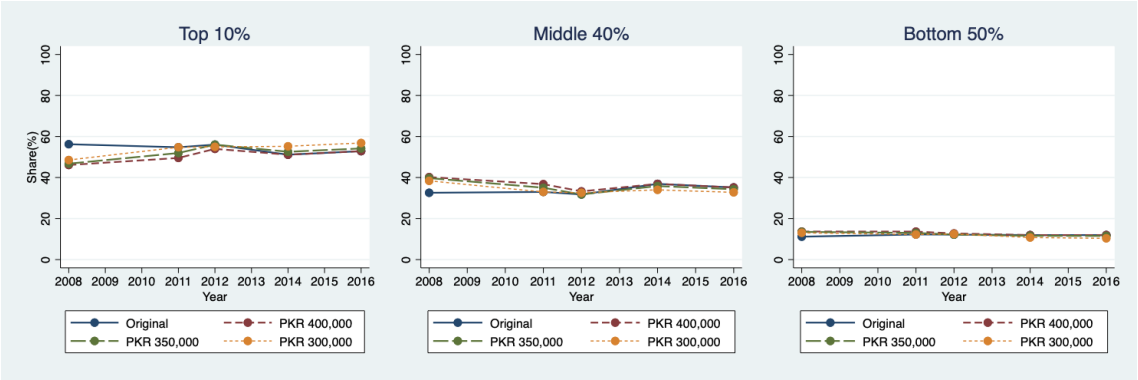
[22] This assumption is named Assumption A and is explained in Section 2.2.

Figure 2.9: Estimates Remain Robust to Changes in Adjustment Threshold



is not the same for all years. To address this concern, this section replicates the estimation technique using the same adjustment threshold for each year. The paper uses three different thresholds, PKR 400,000, PKR 350,000 and PKR 300,00 as the uniform adjustment thresholds. Figure 2.10 shows the trend of the Top 10%, Middle 40% and Bottom 50% respectively with these uniform adjustment thresholds along with the original estimates for comparison. The graphs show that the estimates remain similar to the original results, except for 2008 where the share of the Top 10% and Middle 40% change. This is because the adjustment threshold of 2008 is PKR 100,000 and the robustness test increases the adjustment threshold by a maximum multiple of 4. This leads to lower top income adjustment. Apart from 2008, the results remain similar to the original results.

Figure 2.10: Estimates Remain Robust with Uniform Adjustment Thresholds



To supplement the robustness tests related to the adjustment threshold shown in

Figures 2.9 and 2.10, this paper also tests whether the results remain consistent if the adjustment threshold is defined as a specific income percentile in a year. Instead of taking the adjustment threshold as the minimum tax threshold, this test takes the adjustment threshold as a specific income percentile. The test considers three scenarios: top income adjustment above the 85^{th}, 90^{th} and 95^{th} income percentiles of survey income respectively for each year. The results can be seen in Figure 2.11 and remain robust to this alternative definition of the adjustment threshold.

Figure 2.11: Estimates Remain Robust to Income Percentiles as Adjustment Thresholds

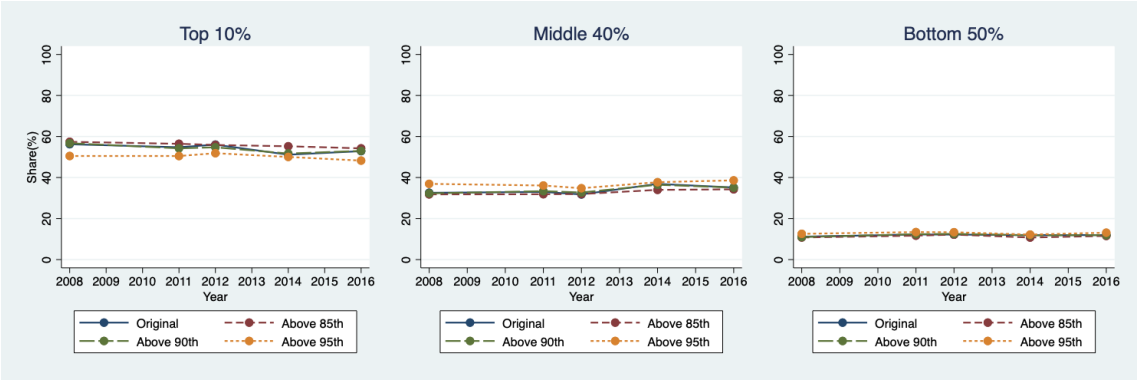
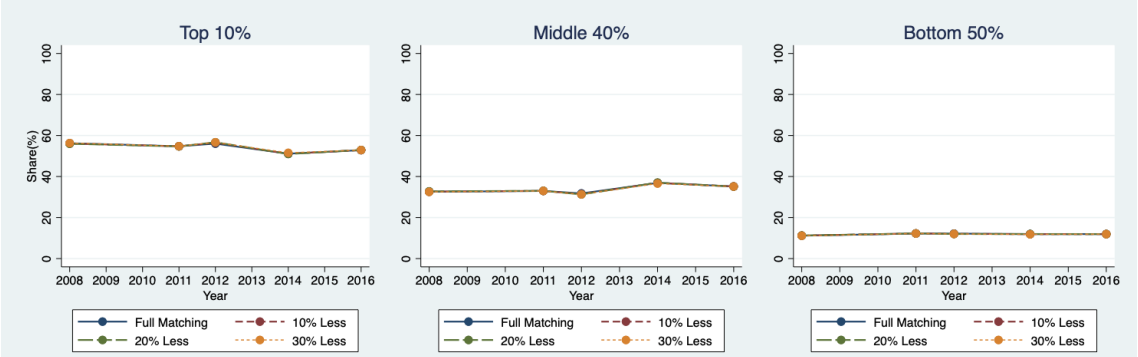


Figure 2.12: Estimates Remain Robust to Changes in Matching Method



A key step in the estimation technique is that each observation in the survey data is matched with an observation of declared income in the tax data. This is done by dividing the survey data into the maximum number of quantiles, then dividing the tax data into the same number of quantiles and finally matching them both. Please refer to the estimation technique explained in Section 2.2 for further details. To

test if this matching process has any impact on the estimates, this paper changes the matching process in three ways. Instead of dividing the survey and tax data into the maximum number of quantiles possible in the survey data, it successively divides the data into 10% less, 20% less and 30% less quantiles than the original estimates. For instance, if the original estimates divide both the survey and tax data into 1,000 quantiles to calculate the upgrade factors, the first test divides them into 900 quantiles, the second test into 800 and the third into 700. The observations that are not matched in these robustness tests are assigned a conservative adjustment factor, that of the nearest lower level of income that was matched.

The results from this robustness test are given in Figure 2.12 which show how the trend of the Top 10%, Middle 40% and Bottom 50% changes respectively with different types of matching. The figures show that the matching method makes little difference in the estimates.

The baseline estimates of the paper use a sample of individuals who earn income (Income>0). An alternative definition could be to take the sample of individuals above the age of 20. The results of this robustness test along with a discussion are shared in Appendix 2.F.

2.6 Conclusion

This paper develops a new methodology to estimate income inequality for developing countries with low levels of tax compliance. The new method addresses key limitations of the existing method by not relying on the number of tax filers to estimate the threshold for top income adjustment. It further improves upon the existing methodology by imposing no distributional assumptions on the income distribution and by using the rich variation in the tax and survey microdata.

I then implement this methodology for Pakistan using a combination of nationally

representative household surveys and administrative data of all tax returns filed in Pakistan from 2008 to 2016. The results show that the existing methodology given in Alvaredo et al. (2016) underestimates income inequality in Pakistan's case while the methodology presented by this paper does not.

This work opens up five areas of future research that would help refine the methodology for inequality estimation further. First, national accounts data for a lot of developing countries are not detailed enough to allow researchers to reconcile the micro income figures with the national accounts. This is the case for Pakistan as well. As a result, I was unable to undertake the final step of reconciling the macro and micro figures. As developing countries add more detail to their national accounts, future researchers can undertake this last step.

Second, this paper shows the pre-tax trend of income inequality across Pakistan. Since most developing countries do not have detailed tax data as part of household surveys, it becomes difficult to calculate the post-tax trend of income inequality. From a distributional standpoint, the difference between the pre- and post-tax trend explains how redistributive the system is. As developing countries start collecting data on tax payments, future researchers can estimate the post-tax trend of income inequality.

Third, the results show very stark gender disparities in Pakistan which can be explored further. While this paper is not able to match the survey observations with the tax data using gender because the administrative tax data does not have details about gender, future work can conduct a closer match. This can provide us with a more detailed understanding of gender disparities.

Fourth, future researchers who have access to tax and survey microdata for developing countries could use the estimation technique explained in this paper to estimate the trend of income inequality for other countries. The generalizability of the methodology proposed by this paper depends on the availability of income tax

microdata in developing countries. Even if researchers might not have direct access to such proprietary data, tax administrations in the developing world do have such data for recent years. Hence, they could, in theory, implement the methodology proposed by this paper with ease in order to produce income inequality estimates. Such work will add to the literature on, and the global discussion of income inequality.

Fifth and finally, future work could empirically test the generalizability of Alvaredo et al. (2016) in low tax compliance settings. From the analysis conducted by this paper, we would expect Alvaredo et al. (2016) to increasingly underestimate income inequality as tax compliance reduces. Given that there would be some level of tax evasion that exists in every country, Alvaredo et al. (2016) would always produce underestimates of income inequality, with greater underestimates for lower tax compliance countries. Future cross-country work could empirically test this relationship and consequently improve our understanding of income inequality estimation.

References

Alvaredo, Facundo, Anthony Atkinson, Emmanuel Saez, Gabriel Zucman, and Thomas Piketty, "World Inequality Database (WID)," 2018.

_ , _ , Lucas Chancel, Thomas Piketty, Emmanuel Saez, and Gabriel Zucman, "Distributional National Accounts Guidelines: Methods and Concepts Used in WID.world," *WID Working Paper*, 2016.

Anand, Sudhir and Paul Segal, "What Do We Know about Global Income Inequality?," *Journal of Economic Literature*, 2008, *46* (1), 57–94.

Atkinson, Anthony, "Top incomes in the UK over the 20th century," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 2005, *168* (2), 325–343.

_ and Thomas Piketty, *Top incomes over the twentieth century: a contrast between continental European and English-speaking countries*, Oxford University Press, 2007.

Blanchet, Thomas, Ignacio Flores, and Marc Morgan, "The Weight of the Rich: Improving Surveys Using Tax Data," *WID Working Paper*, 2018.

Chancel, Lucas and Thomas Piketty, "Indian Income Inequality, 1922-2014: From British Raj to Billionaire Raj?," *WID Working Paper*, 2017.

Garbinti, Bertrand, Jonathan Goupille-Lebret, and Thomas Piketty, "Income inequality in France, 1900–2014: Evidence from Distributional National Accounts (DINA)," *Journal of Public Economics*, 2018, *162* (C), 63–77.

Government of Pakistan, "Income Tax Ordinance, 2001," 2001.

Kopczuk, Wojciech, Emmanuel Saez, and Jae Song, "Earnings Inequality and Mobility in the United States: Evidence from Social Security Data since 1937," *The Quarterly Journal of Economics*, 2010, *125* (1), 91–128.

Kuznets, Simon, *Shares of upper income groups in income and savings* Publications of the National Bureau of Economic Research; 55, National Bureau of Economic Research, 1953.

Pakistan Bureau of Statistics, "Pakistan Social And Living Standards Measurement Survey (PSLM) 2008-16: Provincial Dataset," Technical Report, Islamabad, Pakistan 2016.

Piketty, Thomas, "Income Inequality in France, 1901–1998," *Journal of Political Economy*, 2003, *111* (5), 1004–1042.

_ , Emmanuel Saez, and Gabriel Zucman, "Distributional National Accounts: Methods and Estimates for the United States," *The Quarterly Journal of Economics*, 2018, *133* (2), 553–609.

_ , Li Yang, and Gabriel Zucman, "Capital Accumulation, Private Property, and Rising Inequality in China, 1978–2015," *American Economic Review*, 2019, *109* (7), 2469–2496.

Sala-I-Martin, Xavier, "The World Distribution of Income: Falling Poverty and... Convergence, Period," *The Quarterly Journal of Economics*, 2006, *121* (2), 351–397.

United Nations, European Commission, International Monetary Fund, Organisation for Economic Co-operation and Development, and World Bank, eds, *System of National Accounts 2008*, New York: United Nations, 2009.

Appendix

2.A Income Definition and Inequality Measures

A detailed explanation of the different income definitions used in the recent literature on income inequality is given in Alvaredo et al. (2016). In addition to explaining the methodology to estimate income inequality, the document provides four different definitions of income that map onto different heads in the national accounts. These definitions are pre-tax national income, pre-tax factor income, post-tax disposable income and post-tax national income. This paper uses the pre-tax national income to estimate income inequality for four reasons. First, the survey data in Pakistan does not have information about taxes paid by individuals. As a result, the post-tax version of income inequality cannot be estimated. This leaves a choice between pre-tax national income and pre-tax factor income. Second, the variables for income that are available in the survey data map on to the pre-tax national income definition and hence are used for the analysis. Third, Alvaredo et al. (2016) mention that they prefer pre-tax national income because it includes pension income. This is explained in more detail below. Fourth, the pre-tax national income definition has been used widely in the existing literature and most income inequality estimates with top income share use this definition. Using this definition allows for the comparison of income inequality with different countries.

In this definition of income, there are two things that require a special mention. First, as mentioned previously, pensions are added to the total income while social security contributions are deducted. This is because if pension is not included in the income definition, we run the risk of overestimating income inequality which can be explained by people going into retirement. Alvaredo et al. (2016) refer to adding pensions and deducting the social security contributions as the 'narrow' definition

of pensions.[23] Second, the market rent of owner-occupied housing is also included in the definition of income as per the System of National Accounts (SNA), 2008 (United Nations et al., 2009).[24] The survey data includes a variable that shows the market rent of owner-occupied housing which makes it possible to do this. The main reason why this rent is included is that the income of people living in their own houses would be underestimated when compared to people earning money from rented houses.

While the income definitions used in this paper are in line with the ones used in the National Accounts, this paper does not reconcile the income figures from the estimates with the national accounts. This is because Pakistan's national accounts are not detailed enough to allow for this step. As and when the national accounts have more detail, this final step can be completed.

There are several measures of income inequality that have been used in the literature. Some examples include top income shares, the Gini coefficient, the 20-20 ratio, the Palma ratio, ratio of percentiles, the Theil index and the Hoover index. This paper uses top income shares and the Gini coefficient for three reasons. First, recent estimates developed as part of of the World Inequality Database through a series of papers have used top income shares as a preferred measure of income inequality (Atkinson, 2005; Atkinson and Piketty, 2007; Chancel and Piketty, 2017; Garbinti et al., 2018; Kopczuk et al., 2010; Piketty, 2003; Piketty et al., 2018, 2019). In similar vein, this paper also uses top income shares that allow for the comparison of inequality in Pakistan with other countries. Second, the top income shares are simple and relatively easy to understand. The recent literature on income inequality also cites this as a reason behind the choice of using top income shares. Third, the Gini coefficient has been consistently used in the literature on income inequality over

[23] For further details about pensions, please refer to Appendix 2.C.

[24] Owner-occupied housing is included in household consumption expenditures as per the SNA 2008 (United Nations et al., 2009). Alvaredo et al. (2016) synchronize their income definitions with the SNA 2008 so that the income definitions used for inequality estimates are the same as the ones used in the National Accounts. As a result, they include owner-occupied rent.

the years. Hence, the estimates of the Gini coefficient act as a good supplement to the overall findings of this paper.

2.B Alternative Estimation Techniques Considered

To develop the new methodology explained in Section 2.2, the paper experimented with two more methods that had their own set of limitations. This iterative process helped develop the final methodology proposed in the paper. The alternative estimation techniques considered, along with their limitations are described below.

Alternative 1

First, divide the survey data into equal brackets of income. The size of the brackets in this case is arbitrary. For instance, the whole survey data can be divided into equal brackets of PKR 100,000. Second, calculate the total survey income for individuals in the income brackets from step 1. This simply involves adding up all incomes in a particular bracket. Third, repeat the previous steps using tax data with the same income brackets used for survey data. This will yield the total income for every income bracket using tax data. Fourth, compare the total income of each income bracket from the survey and tax data to see where the tax data exceeds the survey data. Let p_2 be the percentile of income where the income bracket from the tax data exceeds the survey data. This percentile would be for the lower bound of the income bracket.

Assumption 1: The tax data is an accurate predictor of income above percentile p_2 in the survey data.

Assumption 2: The survey data is an accurate predictor below an arbitrarily chosen percentile p_1.

Assumption 3: Between p_1 and p_2, incomes are higher than the survey data, and lower than the tax data.

Fifth, based on assumption 1, calculate the ratio of tax and survey income from the

aforementioned income distributions above percentile p_2. This ratio is then used to upgrade survey incomes above percentile p_2. Sixth, between p_1 and p_2, the upgrade ratio is increased from 1 at p_1, to the observed ratio at p_2. Finally, repeat the steps above for all years and calculate preferred measure of income inequality.

Please note that this method is similar to the existing estimation technique because Assumptions 1, 2 and 3 are the same. The big difference is the way to determine p_2. If the total income for a bracket in the tax data exceeds that of survey data, it would make a case for top income adjustment.

This methodology has three major limitations. First, the value of p_2 from the estimation technique is very sensitive to the definition of income brackets. Since survey data has population weights for each observation, a smaller income bracket would lead to a significantly lower value of p_2. This is even more significant because the choice of income brackets is completely arbitrary. Second, due to the non-filing of tax returns, the sum of all incomes within a bracket using tax data is unlikely to ever exceed the sum for the same bracket in the nationally representative survey. This inflates the value of p_2 and leads to practically no top income adjustment using the tax data. This methodology somewhat addresses the concern about Assumption 1 explained in the existing estimation technique. However, it ends up being too conservative leading to no top income adjustment. Third, the concerns about Assumptions 2 and 3 explained in Section 2.2 still hold.

Alternative 2

First, divide the survey data into equal brackets of income. Second, calculate the total number of individuals in the income brackets from step 1 using survey data. Third, repeat the previous steps using tax data with the same income brackets used for survey data. This will yield the number of individuals in every income bracket using tax data. Fourth, compare the total number of individuals in each income bracket from the survey and tax data to see where the tax data exceeds the survey

data. Let p_2 be the percentile of income where the total number of individuals from the tax data exceeds the survey data. This percentile would be for the lower bound of the income bracket. Finally, use the same estimation technique explained in the Alternative 1 that uses p_1 and p_2.

This methodology also has the same limitations as Alternative 1. Briefly, the value of p_2 from the estimation technique is very sensitive to the definition of income brackets, non-filing of tax returns leads to lack of top income adjustment, and Assumptions 2 and 3 explained in the previous section can be contested.

2.C Primary Income and Pensions

This appendix shows how the primary income and pensions in the survey data vary with age for all years. The existing literature on income inequality recommends using a definition of income that combines the primary income being earned by an individual along with the pensions net of any social security contributions. This is because if pension is not included in the income definition, we run the risk of overestimating income inequality which can be explained by older people going into retirement. This can be seen from the figures below where pensions rise at higher ages as primary income falls. Hence, the paper uses a definition of income that includes pensions and deducts social security contribution. The province-level surveys have the right level of disaggregation to implement this definition of income. As discussed in the paper, the district-level surveys do not have disaggregated and consistent variables for income. Please refer to Table 2.E.2 for details about the exact variables in the survey data.

Figure 2.C.1: Pension Included in Income Definition

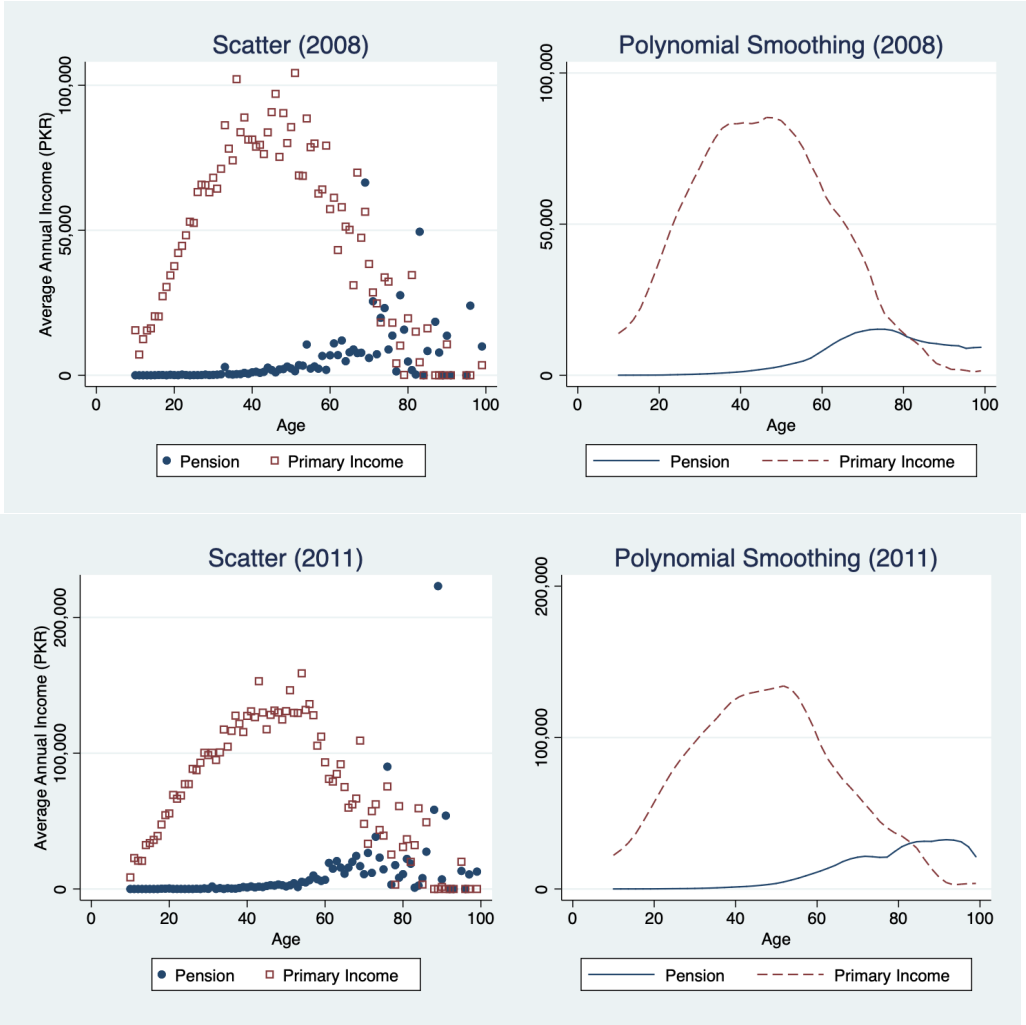
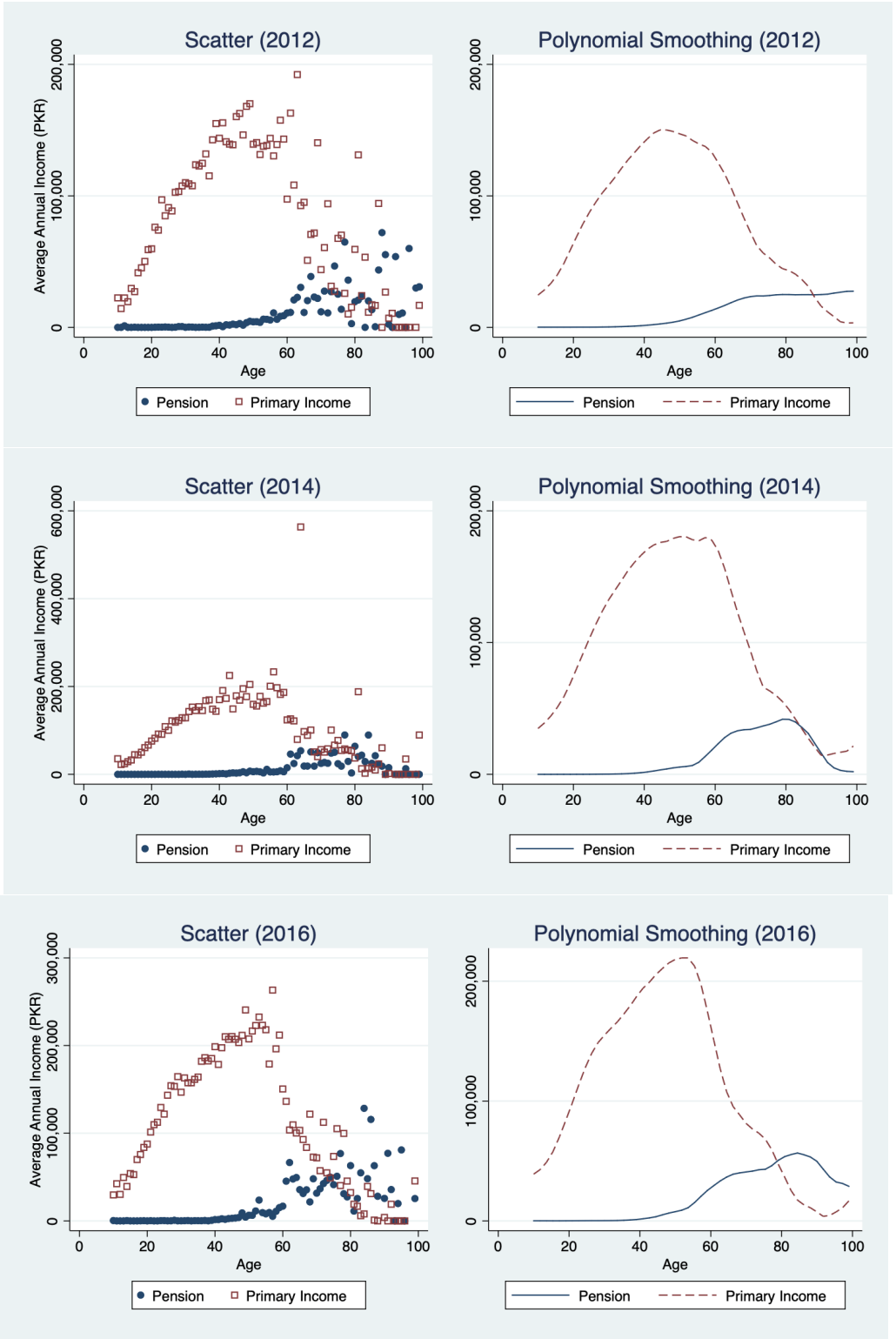


Figure 2.C.2: Pension Included in Income Definition (Continued)



2.D Comparison of Upgrade Factors

This appendix supplements Section 2.2 of the main paper and shares the comparison between the upgrade factors using the methodology developed by this paper, and the one in Alvaredo et al. (2016). The main paper presented the results for 2016. This appendix presents the graphs for other years, given in Figures 2.D.1 and 2.D.2. The results for 2016 are re-included in this appendix for convenience.

Figure 2.D.1: Upgrade Factors for This Paper vs Alvaredo et al. (2016)

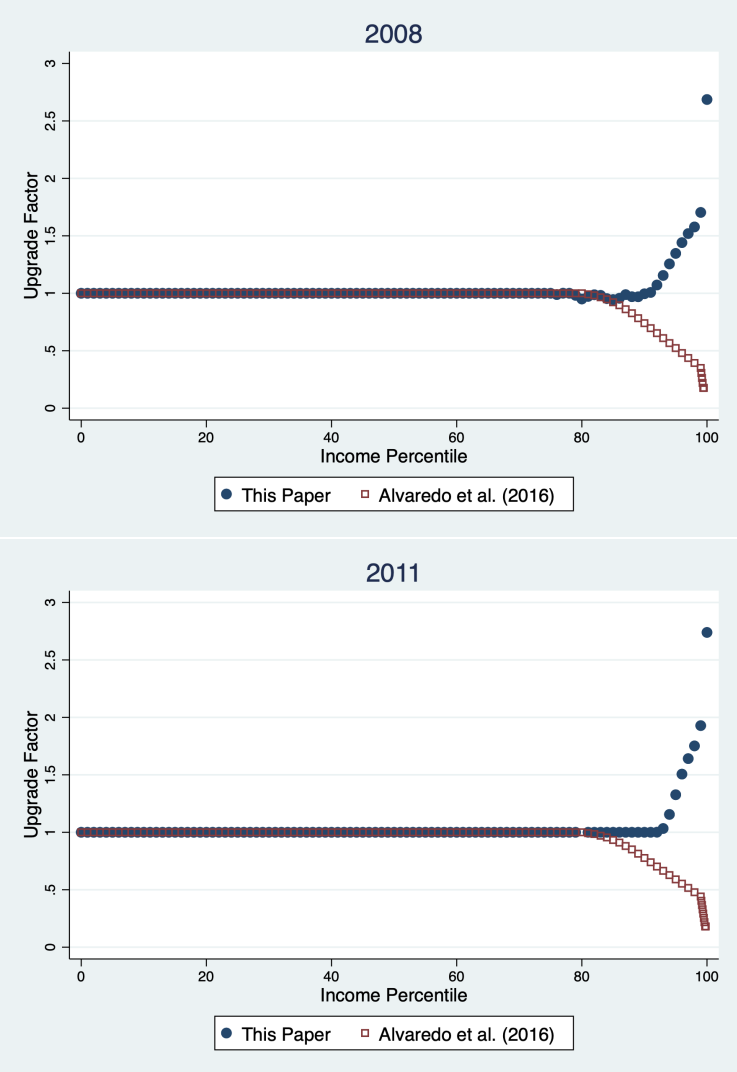
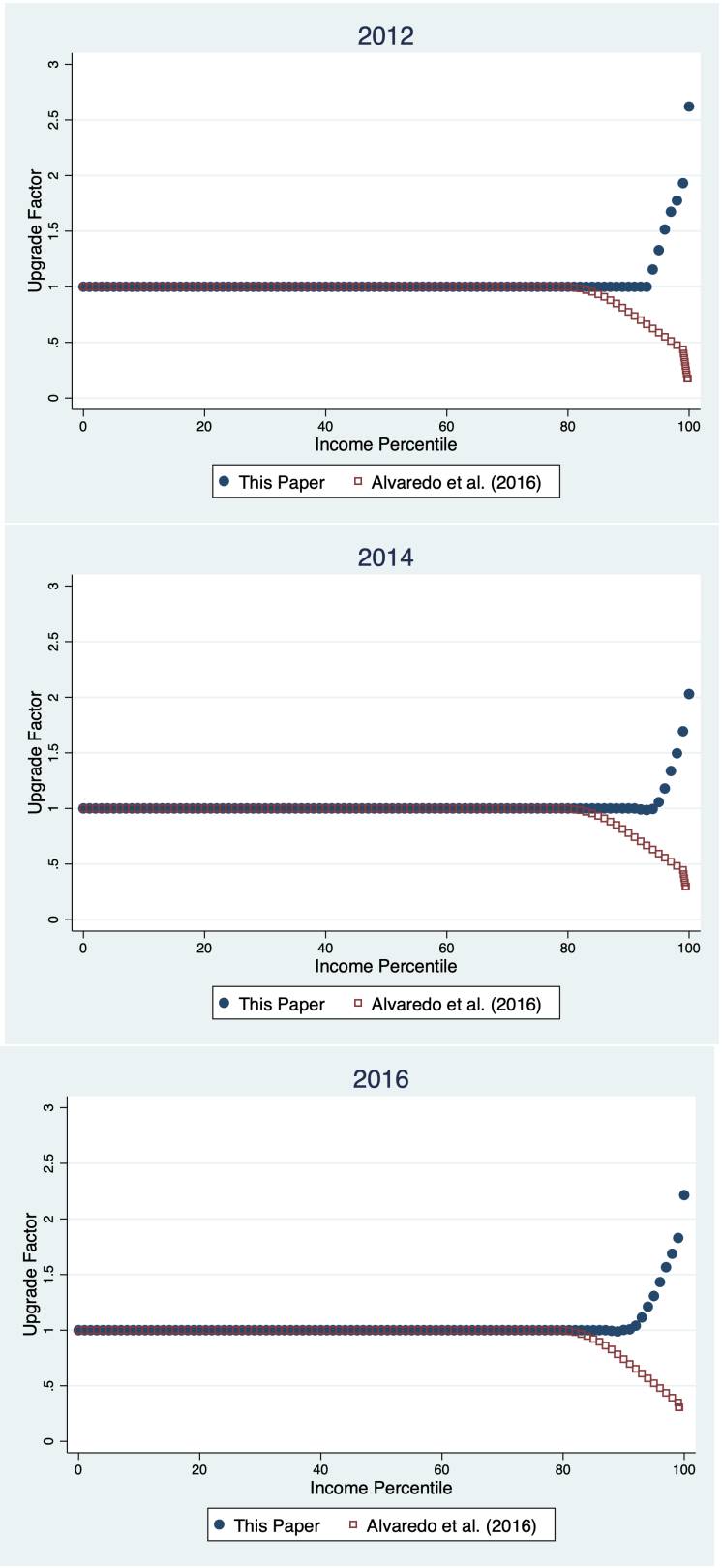


Figure 2.D.2: Upgrade Factors for This Paper vs Alvaredo et al. (2016) (Continued)



2.E Pakistan Social and Living Standards Measurement Surveys

The PSLM is structured such that it alternates annually between being conducted at a district-level and a province-level. From 2007 to 2016, the PSLM has an average of 75,969 households in the district-level survey and 17,978 households in the province-level surveys. The number of observations and households for these surveys is given in Table 2.E.1. Please note that the data for 2010 is not available because the PSLM was not conducted during this year. The province-level survey that was supposed to take place in 2010 was implemented in 2011 instead. As a result, there are two surveys for 2011. The micro-files for all these surveys are publicly available on the Pakistan Bureau of Statistics website (`www.pbs.gov.pk`).

The province- and district-level surveys contain different variables that record income. To estimate income inequality, this paper uses province-level household surveys rather than district-level ones due to three reasons. First, province-level surveys have the most detailed classifications of income that is not present in the district-level survey. This can be seen from Table 2.E.2. The province-level surveys have 17 variables depicting different types of income while district-level surveys only have 1 variable for 2007 and 2009, 5 variables for 2011, and 8 variables for 2013 and 2015. Second, the province-level surveys disaggregate income in a way that allows for the calculation of pre-tax post-replacement national income, which is in line with the existing literature in Alvaredo et al. (2016). As per this definition, pensions are included in income, while social security contributions are deducted from income. District-level surveys do not have the appropriate disaggregation to be able to measure this definition of income. Third, the income variables for the province-level survey have remained the same throughout the duration of the study. On the other hand, the district-level surveys for 2007 and 2009 only measure employment income, and the 2011 version did not measure rental income. This lack of consistency in the

district-level surveys can be seen in Table 2.E.2, and posed a challenge for using data from these surveys.

While the micro files for the province-level surveys consistently had the same income variables throughout the duration of the study, the survey files for each year were not in a standard format. The PSLM also measures other development outcomes that were not relevant to this paper. To conduct the analysis, relevant data from each year was first extracted, then converted into a standard format and finally combined into one file.

Table 2.E.1: Pakistan Social and Living Standards Measurement Survey (PSLM)

Year	**Survey Type**	**Observations**	**Households**
2007	District Level	496,060	73,953
2008	Province Level	107,635	15,512
2009	District Level	499,798	75,188
2011-A	Province Level	108,937	16,342
2011-B	District Level	499,220	76,548
2012	Province Level	106,497	15,807
2013	District Level	492,657	75,521
2014	Province Level	119,019	17,989
2015	District Level	513,099	78,635
2016	Province Level	157,636	24,238

Notes: The PSLM are nationally representative household surveys composed of two types that alternate every year. These two types are called the district- and province-level surveys. These surveys measure slightly different development outcomes with some overlapping variables.

Table 2.E.2: Income Variables in Pakistan Social and Living Standards Measurement Surveys (PSLM)

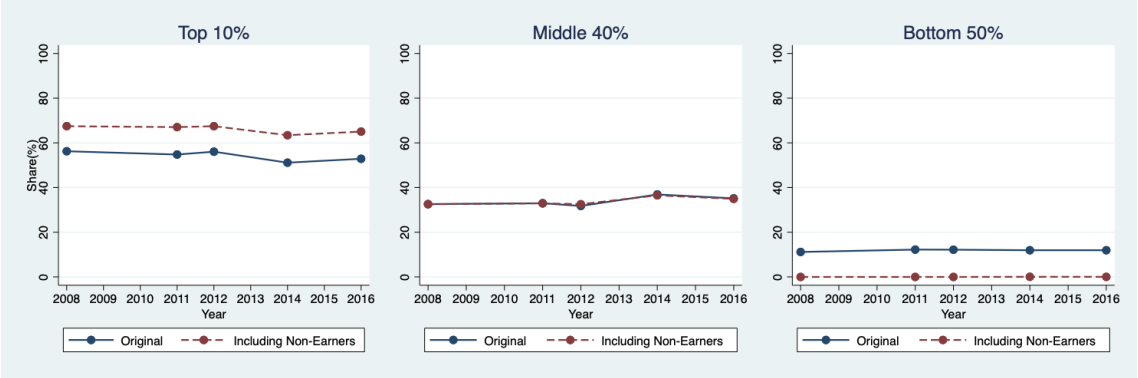
		Variables																	
Year	**Survey Type**	**Primary Income**	**Secondary Income**	**Other Income**	**Income in Kind**	**Pension**	**Net International Remittances**	**Rental Income**	**Social Security Contribution**	**Interest on Savings**	**Dividends**	**Provident Fund Income**	**Profit on Loan**	**Group Insurance**	**Agricultural Land Rent**	**Agricultural Equipment Rent**	**Non-Agricultural Establishment Rent**	**Owner Occupied Rent**	**Any Other Income**
2007	District	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
2008	Province	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗
2009	District	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
2011-A	Province	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗
2011-B	District	✓	✓	✓	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
2012	Province	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗
2013	District	✓	✓	✓	✓	✓	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓
2014	Province	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗
2015	District	✓	✓	✓	✓	✓	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓
2016	Province	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗

Notes: The PSLM are nationally representative household surveys composed of two types that alternate every year. These two types are called the district and province-level surveys. These two types of surveys measure slightly different development outcomes with some overlapping variables. Province-level surveys are used in developing the estimates and are highlighted light-grey in the table. District-level surveys are not used in developing the estimates and are not highlighted.

2.F Changes in the Selected Sample

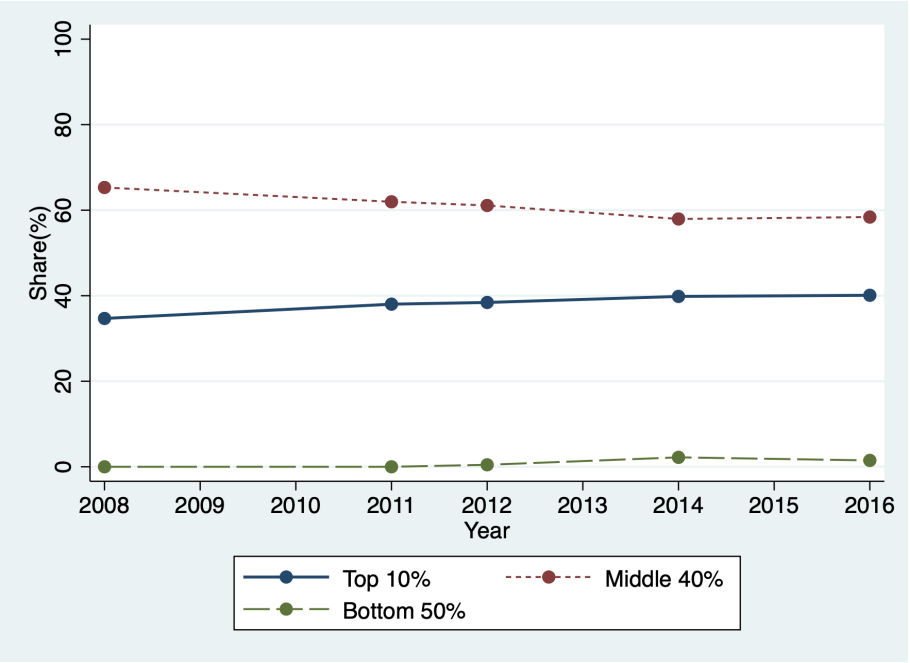
This appendix supplements Sections 2.4 and 2.5 of the main paper, and shares how the estimates of income inequality change when non-earners are included in the sample. The baseline estimates looked at a sample of earning individuals (Income>0). An alternative definition of the sample could be to look at all adults above the age of 20 regardless of whether they are in the labor force. The results of using this sub-sample are given in Figure 2.F.1. Including non-earners in the sample increases the share of the Top 10% while decreasing the share of the Bottom 50%. This happens because there is a significant number of individuals in the sample who do not earn any income which reduces the share of the Bottom 50%. The share of the Middle 40% remains stable.

Figure 2.F.1: Including Non-Earners (Age 20+)



The main paper also shares the estimates that implement the methodology laid out by Alvaredo et al. (2016). For comparison, the main paper used a sample of earning individuals while implementing their methodology. Figure 2.F.2 shows the estimates with a sample of individuals above the age of 20. Here as well, the share of the Top 10% is significantly lower compared to the methodology presented by this paper and the unadjusted estimates. As result, the results presented in the main paper remain robust to this alternative definition of income.

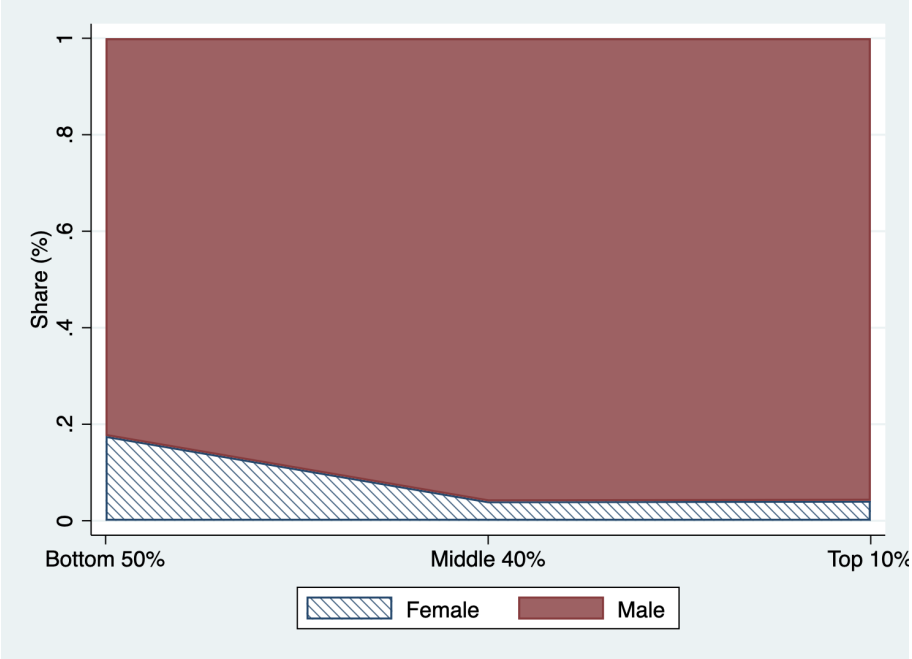
Figure 2.F.2: Alvaredo et al. (2016) Estimates with a Sample of Age 20+



2.G Income Inequality and Gender

From a gender standpoint, the results from this paper show a stark difference in the income earned by males and females. Generally, males in Pakistan earn the majority of income across the income distribution. This can be seen from Figure 2.G.1 which shows the average share of males and females in different brackets of the income distribution.[25] Females in Pakistan earn an average of 17.59% of income within the Bottom 50% bracket, 4.02% within the Middle 40% bracket and 4.15% within the Top 10% bracket from 2008 to 2016. Hence, females in Pakistan not only earn a very low share of income within different brackets, but this share also decreases at higher levels of income.

Figure 2.G.1: Men Dominate Earnings Throughout Income Distribution (No Attribution)

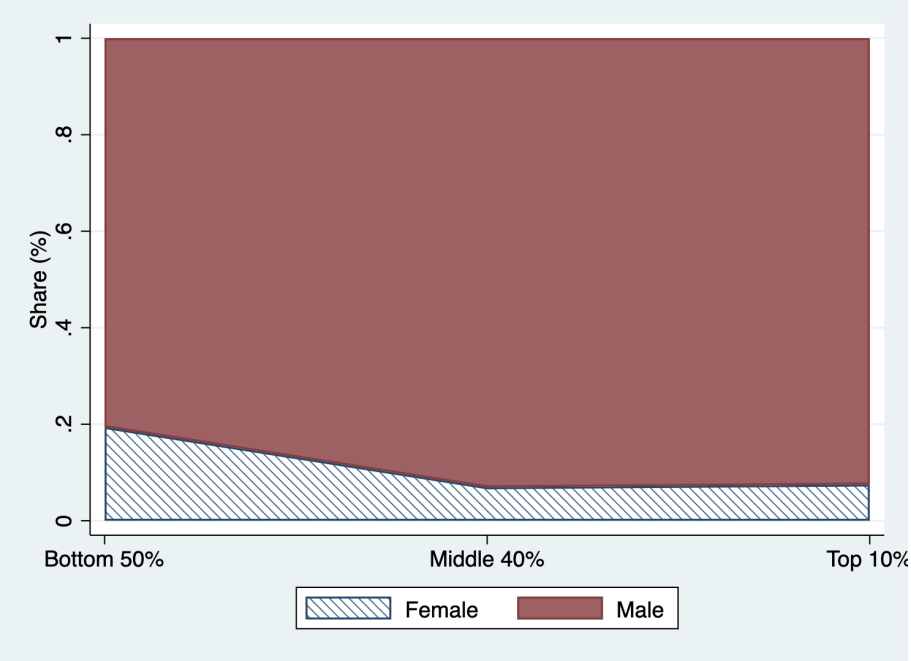


This paper estimates the income shares by gender using two different methods. The first method, called 'No Attribution', only uses the income variables in the survey that can be assigned to an individual and are not at a household level. The second

[25]This graph only uses income sources that can be assigned to a particular individual. For instance, rental income that accrues to the entire household is excluded from the analysis. The results remain similar if income sources that are at a household level are included and assigned.

method, called 'All Income Heads', assigns the income variables at the household level to the household head. Figure 2.G.1 shows the averaged estimates using the 'No Attribution' method while Figure 2.G.2 uses the 'All Income Heads' method.

Figure 2.G.2: Estimates Using All Income Heads Method



Figures 2.G.3 and 2.G.4 show the detailed time trends of the estimates from these two methods respectively. From these two figures, we can make four observations. First, males dominate the earnings across the income distribution, second, the gender disparity increases with income, third, income inequality by gender is stable across time, and last, that the method of attribution makes little difference to the estimates.

Figure 2.G.3: Men Dominate Earnings Across Time

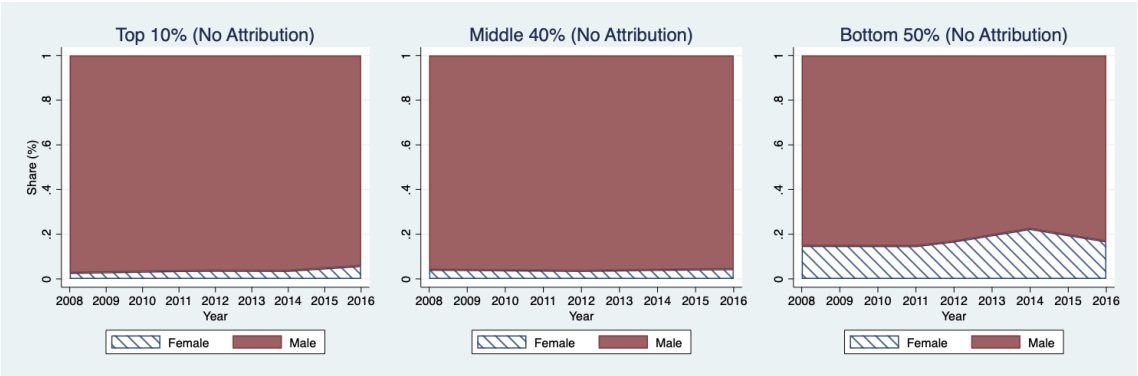
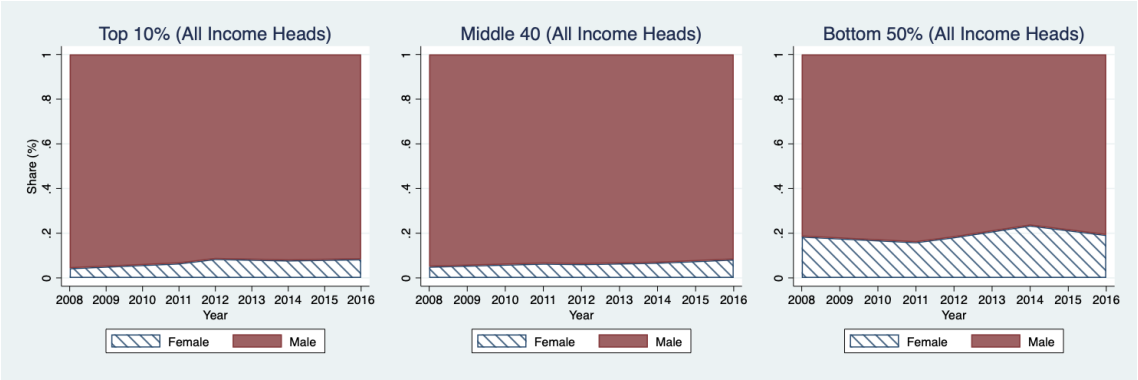


Figure 2.G.4: Men Dominate Earnings Across Time



It is important to mention here that while the existing literature on income inequality splits the earnings between males and females within the household, with a preferred split of 50-50, this paper does not split the earnings. This is because any form of income splitting within a household relies on strong assumptions of intra-household bargaining power. The paper instead does not make any assumptions about income splitting within the household. This is especially important in the South Asian context where female labor force participation is low with a myriad of challenges around female empowerment, and splitting assumptions such as the 50-50 split preferred in the existing literature might lead to bias and an underestimate of gender disparities.

2.H Top Income Adjustment and Age

This appendix supplements Section 2.4 of the main paper that showed how the top income adjustment changes income with age. The main paper presented the graphs for 2016. This appendix presents the graphs for other years which are given below. The graphs for 2016 are re-included in this appendix for convenience.

Figure 2.H.1: Adjustment Increases as People Enter Working Age

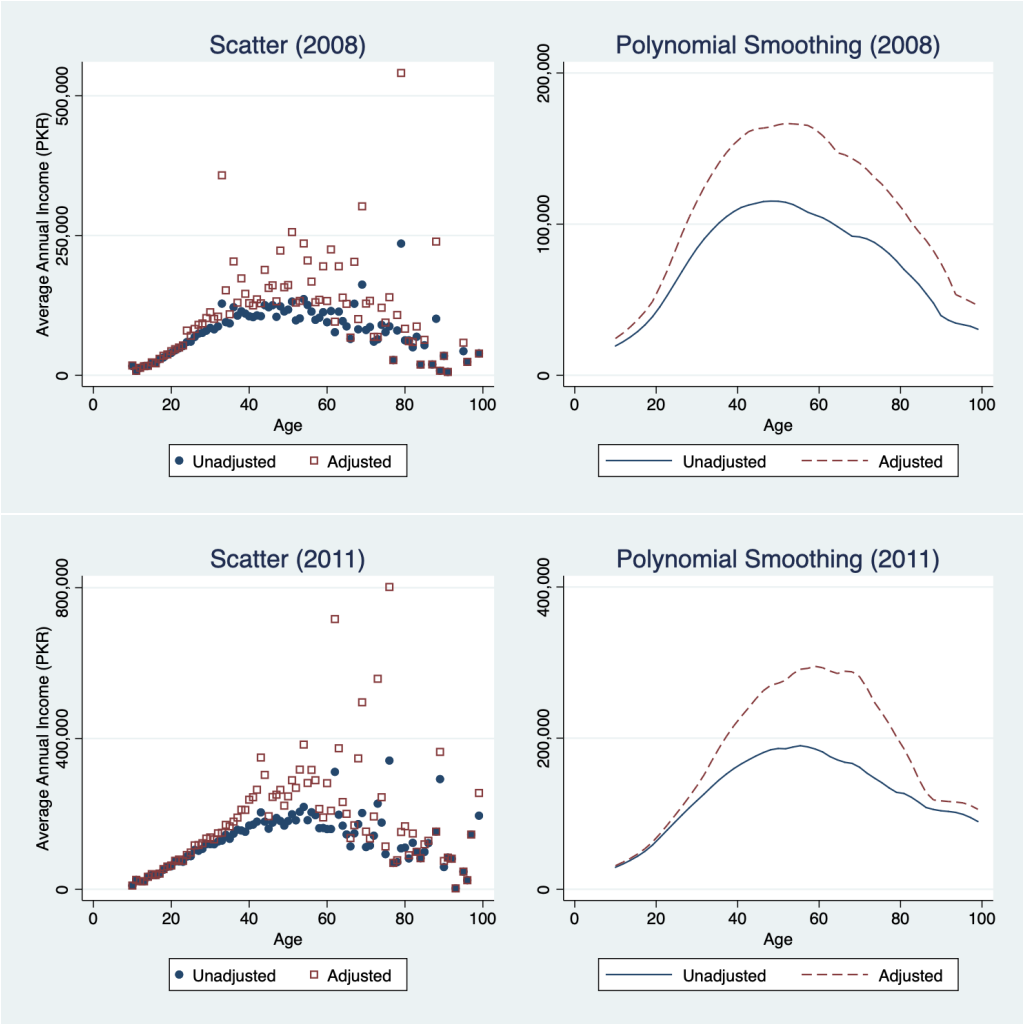
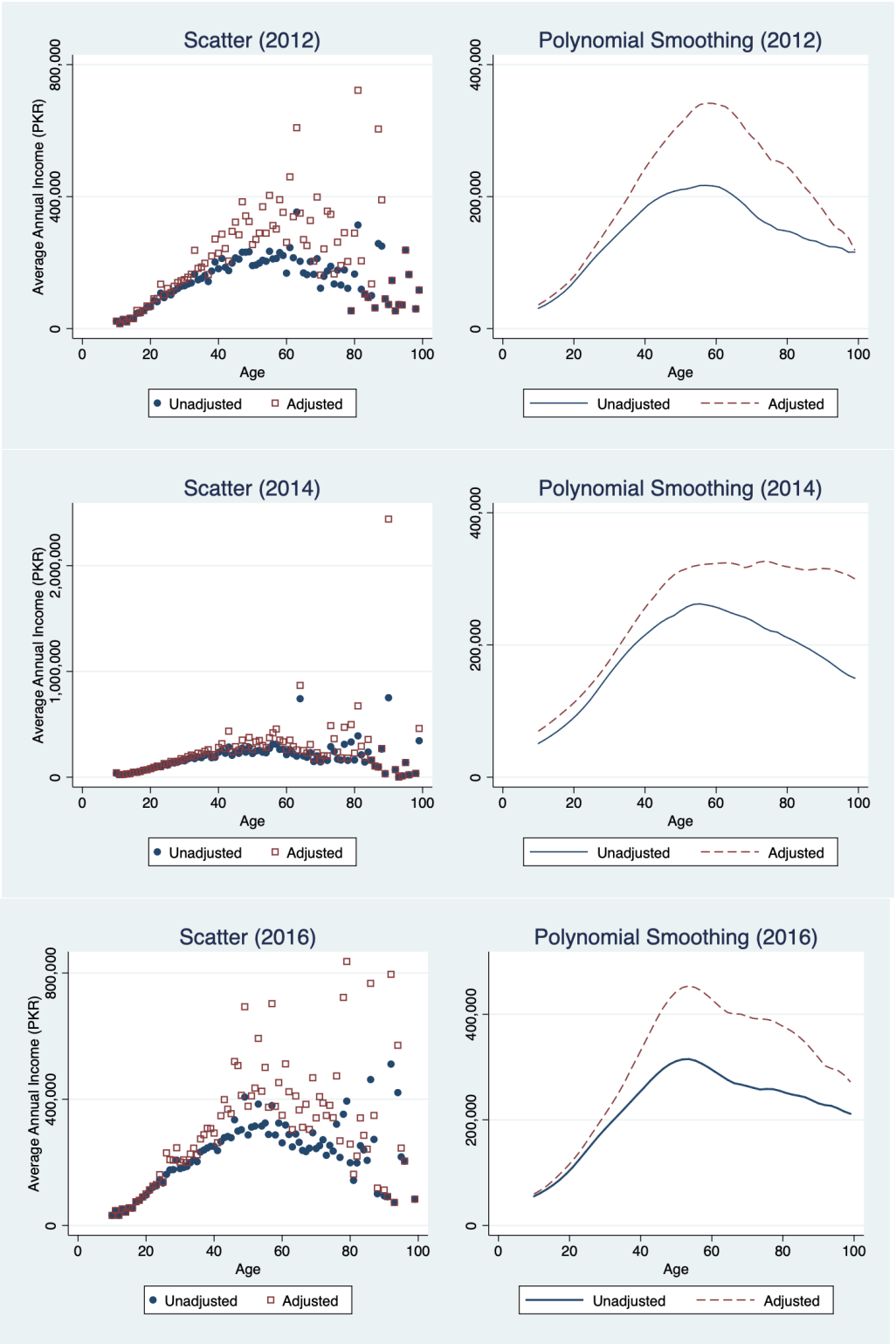


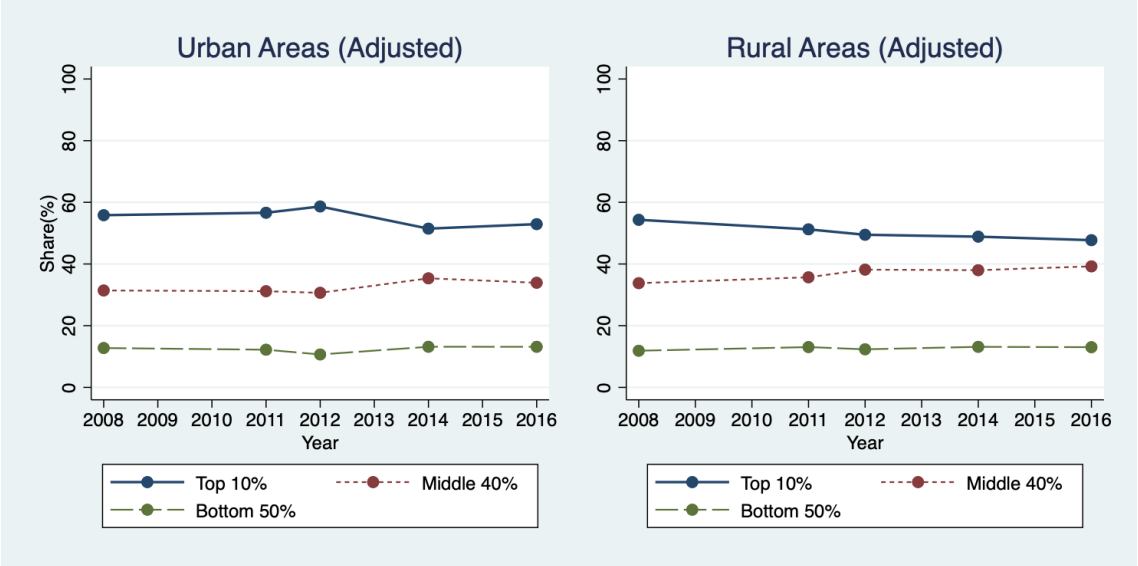
Figure 2.H.2: Adjustment Increases as People Enter Working Age (Continued)



2.I Income Shares for Urban and Rural Pakistan

Income inequality in Pakistan's urban and rural areas is similar to the national trend where the top income shares have remained high, but stable. Figure 2.I.1 shows the adjusted top income shares for urban and rural areas. From these graphs, we can see that top income adjustment significantly increases the share of the Top 10% while reducing the share of the Middle 40% and the Bottom 50%. We can also see that while income inequality is high in both urban and rural areas, urban areas are slightly more unequal when compared to rural areas. Figure 2.I.2 shows the adjusted and unadjusted Gini coefficient for urban and rural areas which corroborates the evidence from top income shares.

Figure 2.I.1: Urban Areas are Slightly More Unequal Compared to Rural



The unadjusted shares for urban and rural areas in Pakistan are given in Figure 2.I.3. These figures, combined with the adjusted shares presented in Figure 2.I.1, show that top income adjustment makes a significant difference to the income shares by increasing the share of the Top 10% and decreasing the share of the Middle 40% and Bottom 50%. This is in line with the estimates for the whole country developed by this paper.

Figure 2.I.2: Top Income Adjustment Increases Gini for Urban and Rural Areas

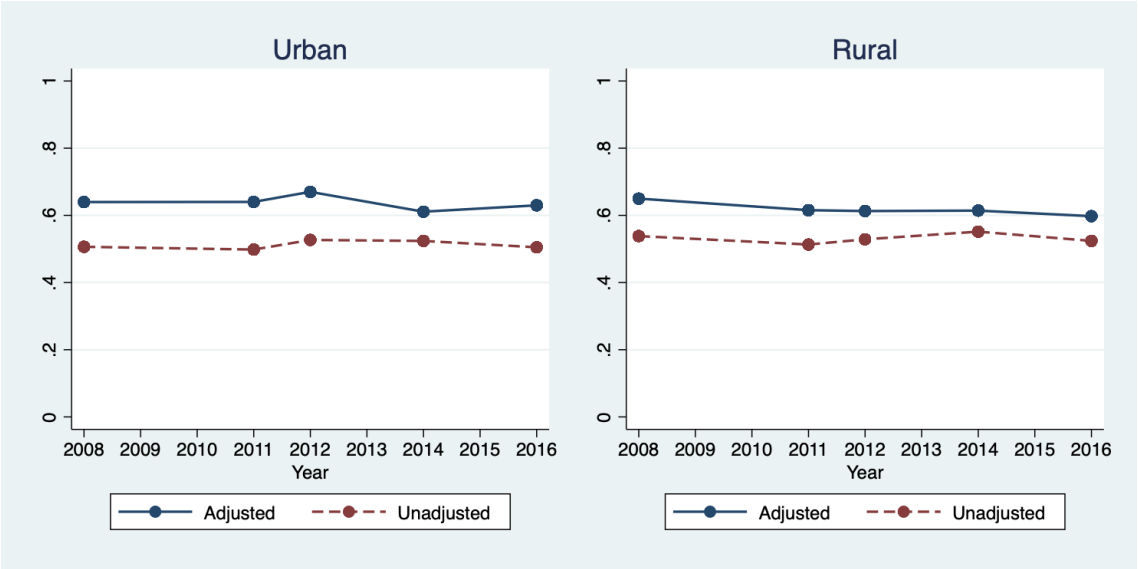
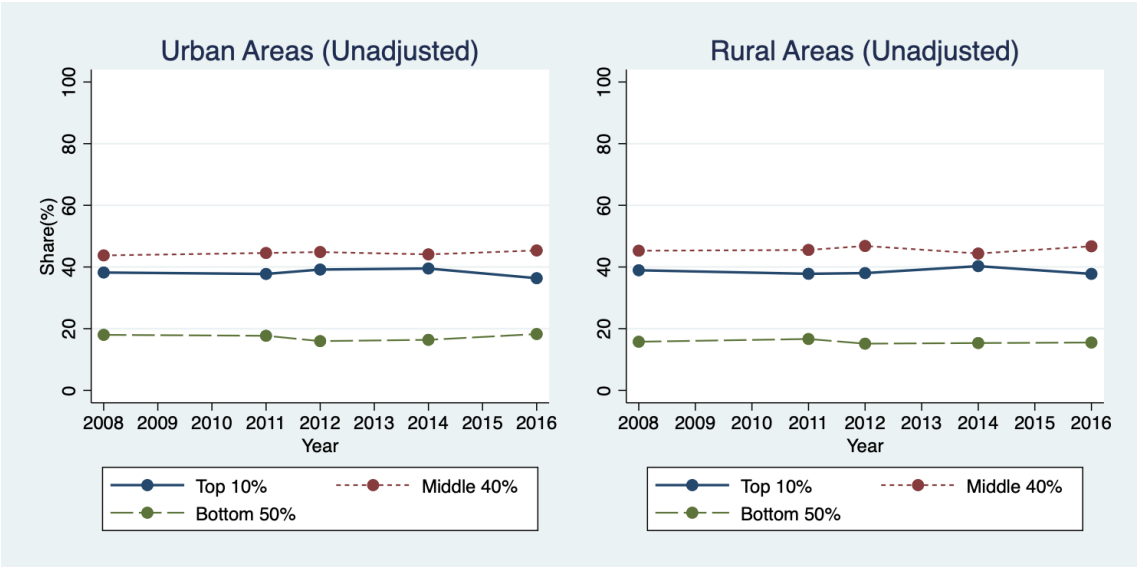


Figure 2.I.3: Similar Levels of Unadjusted Inequality in Urban and Rural Areas



2.J Change in Income Distribution from Top Income Adjustment

This appendix supplements Section 2.4 of the main paper that showed how the income distribution changes due to the top income adjustment. The main paper presented the graphs for 2016. This appendix presents the graphs for other years which are given below. The graphs for 2016 are also re-included in this appendix for convenience.

Figure 2.J.1: Top Income Adjustment Modifies Income Distribution Above Threshold

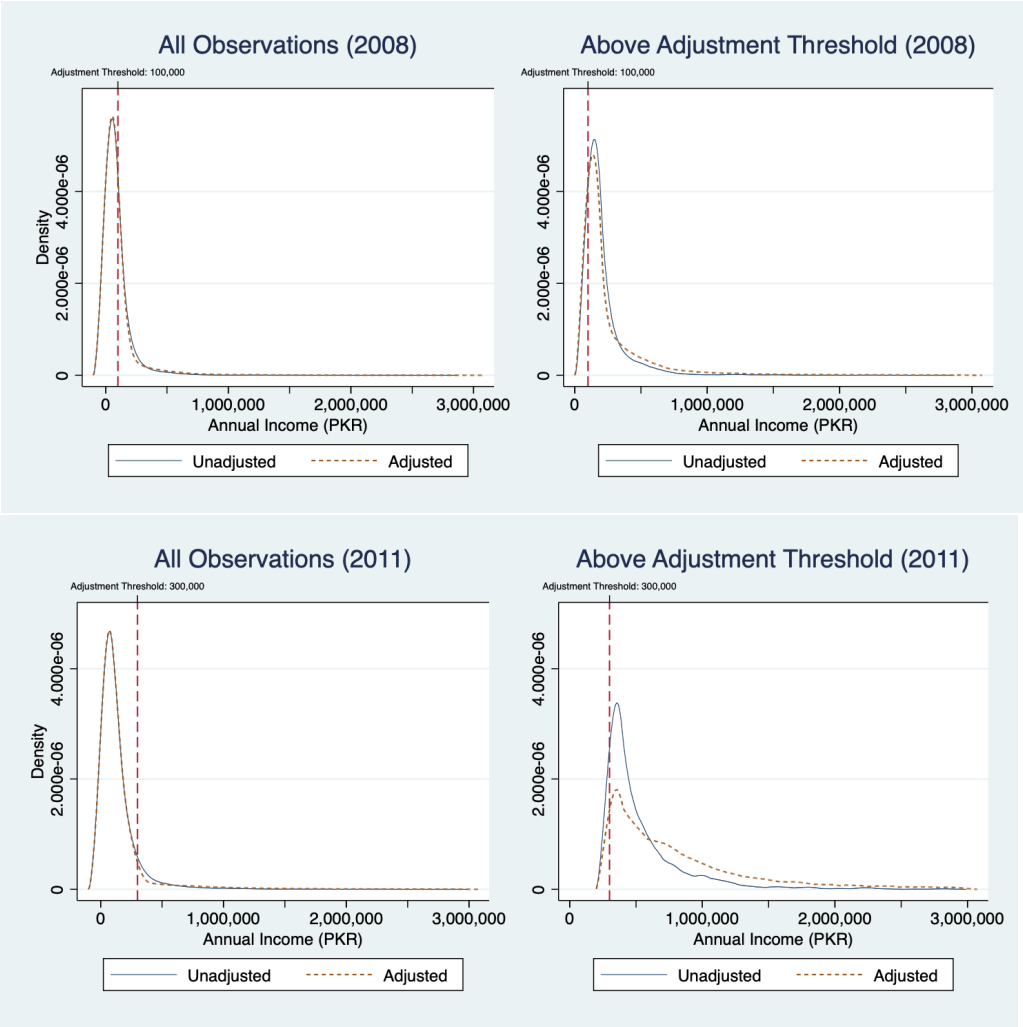
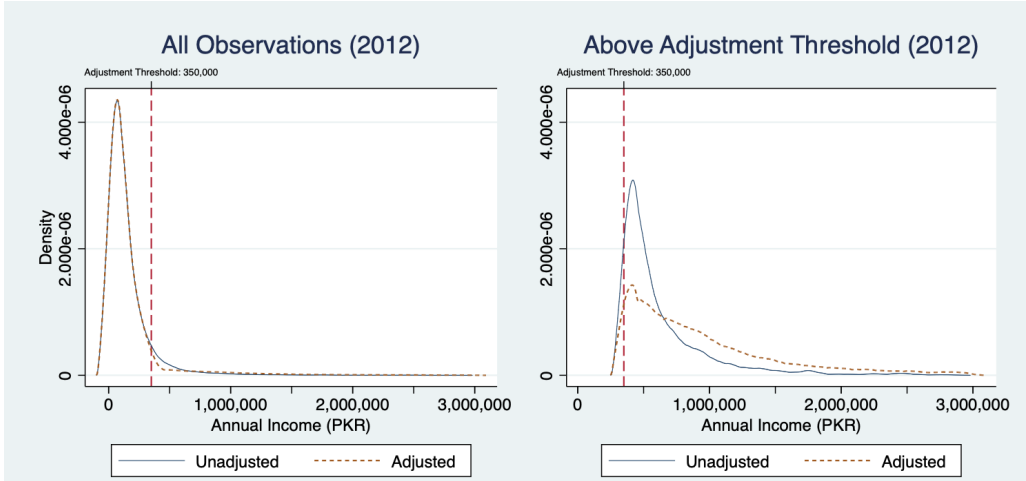
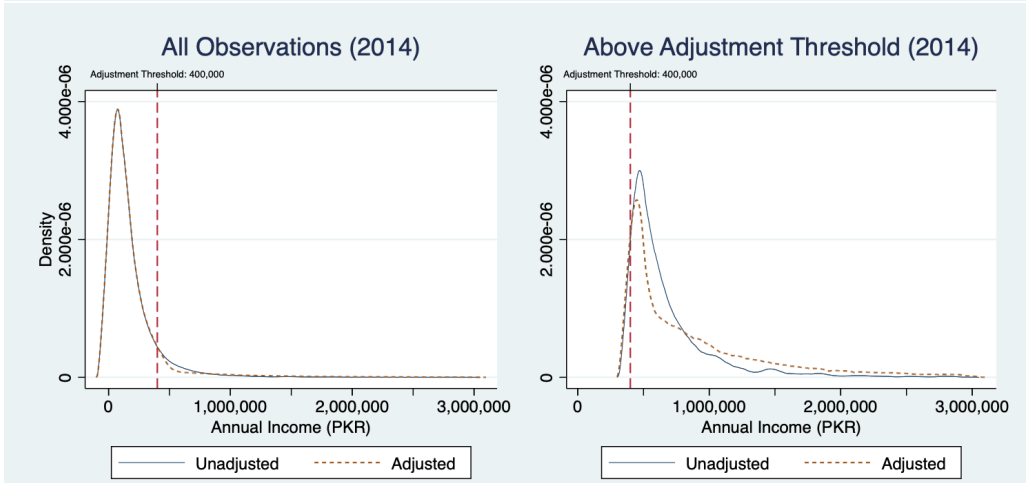
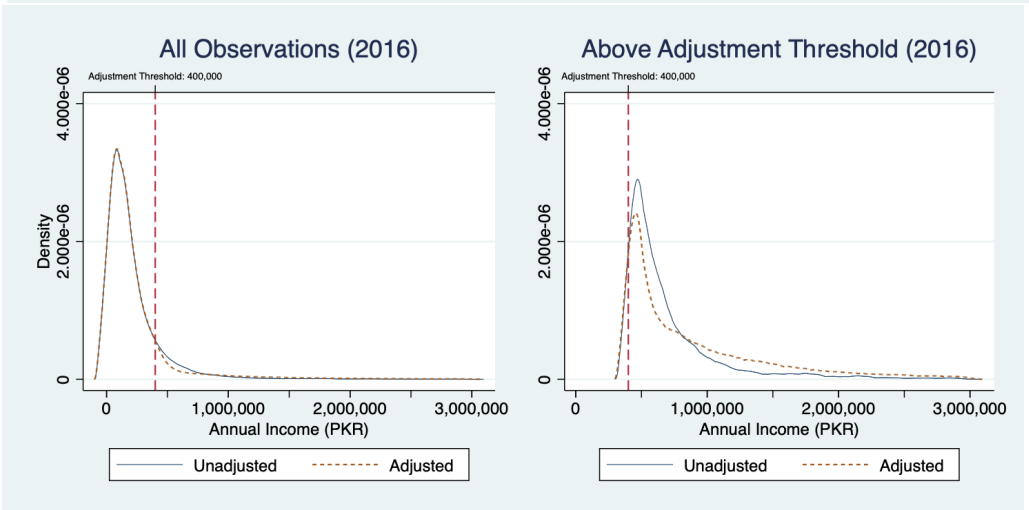


Figure 2.J.2: Top Income Adjustment Modifies Income Distribution Above Threshold (Continued)







2.K Upgrade Factors and Income

This appendix supplements Section 2.4 of the main paper that showed how the upgrade factors from the estimation technique change with income. The main paper presented the graphs for 2016. This appendix presents the graphs for other years which are given below. The graphs for 2016 are also re-included in this appendix for convenience.

Figure 2.K.1: Upgrade Factors Increase with Income Above Threshold

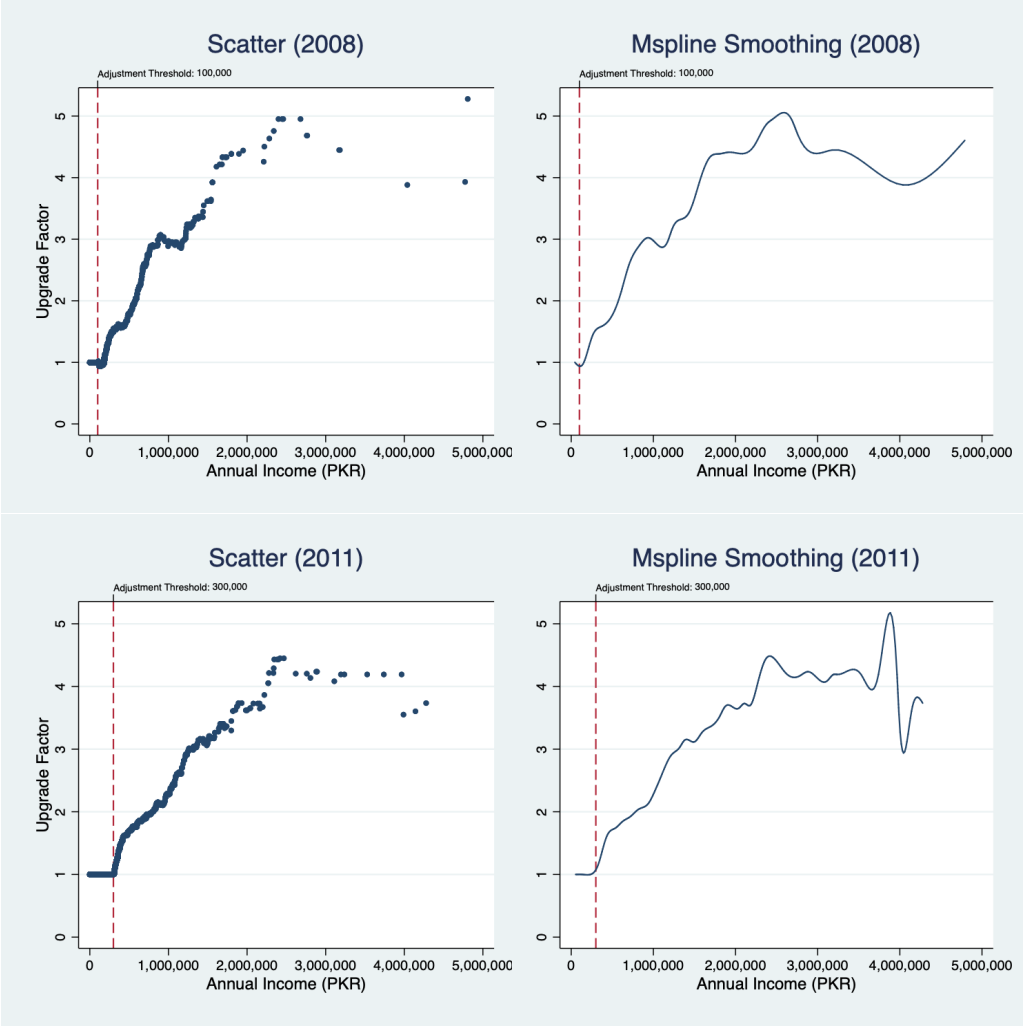
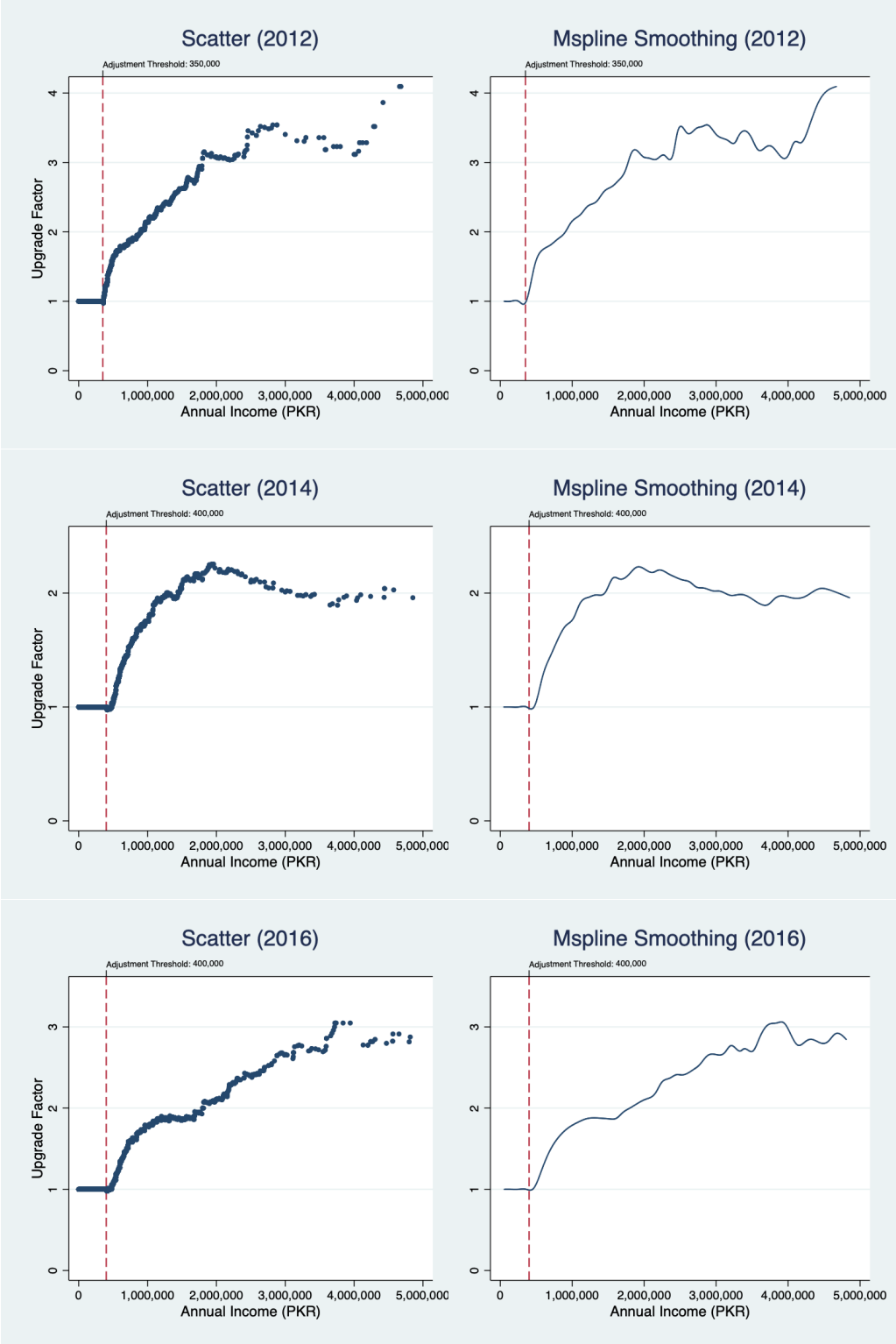


Figure 2.K.2: Upgrade Factors Increase with Income Above Threshold (Continued)



Chapter 3

DO DEADLINE EXTENSIONS ENCOURAGE TAX FILING? EVIDENCE FROM PAKISTAN

Abstract

This paper analyzes a tax broadening policy in a low compliance, developing country context with a narrow tax base. The Board of Revenue in Pakistan regularly gave out extensions to the tax filing deadline under the assumption that time is a binding constraint to filing behavior. To study the relationship between extensions and tax compliance behavior, I use a rich panel dataset of all tax returns filed in Pakistan from 2007 to 2017 and combine this with data extracted from official government circulars containing details about deadline extensions. I find that deadline extensions are associated with individuals and unincorporated businesses delaying the filing of their tax returns by 88% and 70% of the extension duration respectively. Extensions also reduce the compliance of individuals and unincorporated businesses with the final extended deadline by 4.5 and 1.1 percentage points respectively. I then provide evidence for learning and information provision as the underlying mechanisms at play. Finally, I provide evidence that extensions do not encourage individuals and unincorporated businesses to file their tax returns, suggesting that time is not a binding constraint. Together, these findings reveal that deadline extensions are ineffective in broadening the tax net as they impose additional costs, such as lost interest revenue due to delayed filing and lower compliance with government set deadlines, without providing any benefits.

3.1 Introduction

Developing countries typically tax a lower share of GDP, rely less on income taxation and have narrower tax bases when compared to developed countries (Besley and Persson, 2014).[26] Less than 1% of Pakistan's population files the income tax return, illustrating a narrow tax base that is common in the developing world. Governments in the developing world know the value of expanding the tax base as a potent lever to increase revenue, and hence regularly experiment with policies. However, little is known in the academic literature about the effectiveness of such policies targeted towards widening the tax net.

In this paper, I contribute towards our understanding of tax enforcement by analyzing a tax broadening policy used regularly by Pakistan's tax administration over the last decade. As part of this policy, the government frequently gave out extensions to the tax filing deadline between 2007 and 2017 with a clear objective of encouraging tax filing. These extensions differed across time and type of tax filers, with varying durations. We would expect deadline extensions to help widen the tax net if time is a binding constraint to tax filing.[27]

These extensions were consistently given out to individuals and unincorporated businesses - an average of 66 extra days to file their tax return from 2007-2017. However, companies were given extensions only in exceptional circumstances.[28] The empirical analysis utilizes this variation in extensions across multiple types of tax filers over time to estimate the relationship between extensions and tax filing.[29] To conduct this analysis, I combine an administrative panel dataset of all tax returns filed in

[26] Developing countries typically collect only 10% to 20% of their GDP in taxes while developed countries often collect above 40% (Besley and Persson, 2014).

[27] Official communication from the government usually cites that time could be a binding constraint as a justification for these extensions. For instance, documents from 2009 mention the "pre-occupation of taxpayers with the Holy month of Ramzan" as justification for a 20 day extension (circular issued by the tax collection agency on 30 September 2009).

[28] Companies were only given an average of 5.5 extra days to file their tax return from 2007-2017.

[29] The panel nature of the tax dataset allows for robust estimation that controls for covariates, time fixed effects and individual fixed effects.

Pakistan from 2007 to 2017 with data extracted from official government documents containing details about deadline extensions.[30]

The results reveal that both individuals and unincorporated businesses respond to deadline extensions by delaying the filing of their tax return to a later date. Individuals delay their filing by 88% of the extension duration while unincorporated businesses delay by 70%.[31] Deadline extensions also reduce compliance of individuals and unincorporated businesses with the final extended deadline by 4.5 and 1.1 percentage points respectively.

To explain the underlying mechanisms at play, I provide evidence for learning - previous experience of receiving an extension leads tax filers to delay their filing in future years. I also provide evidence for responding to information - earlier announcements of deadline extensions provide more time for tax filers to change their behavior and are associated with more delays in filing. One additional day of earlier announcement is associated with a delayed filing of 0.44 days for individuals and 0.42 days for unincorporated businesses.

Finally, I answer the question about the effectiveness of this policy to broaden the tax base by showing that deadline extensions do not increase the probability of tax filing for both individuals and unincorporated businesses.[32] Back of the envelope calculations reveal that deadline extensions in Pakistan were associated with a total of USD 5 million lost in interest revenue from 2007-2017. To put this into perspective, this amount is equivalent to the entire annual federal budget for Environment Protection in 2019-20 (Government of Pakistan, 2020).

[30]The panel dataset of tax returns was obtained from the Government of Pakistan after receiving an approval from the federal cabinet. Unfortunately, such data prior to 2007 is not available with the government, which could have helped establish the parallel trends assumption and hence causality. This remains an important caveat to the findings of this paper.

[31]The average duration of the extension is 23.09 days. Individuals delay the filing of their tax return by 20.24 days while unincorporated businesses delay by 16.21 days.

[32]I conduct this analysis using a sample of individuals, unincorporated businesses and companies that filed their tax return at least once between 2007-2017. This practically meant converting the unbalanced panel of filed tax returns into a balanced one.

Together, the findings of this paper show that deadline extensions simply result in tax filers putting off their tax filing to a later date while being ineffective in encouraging tax filing. As a result, the policy incurs additional costs such as loss of interest revenue due to delayed filing and loss of enforcement credibility of the tax administration in the form of lower compliance with the final deadline, without providing any benefits. This makes it a sub-optimal policy, implying that the government should instead look towards other policy alternatives to broaden the tax base.

This paper contributes to an understanding of tax enforcement in developing countries. Following the seminal work by Allingham and Sandmo (1972) where they laid out a model of tax evasion based on the probability of audit and penalties, the recent tax enforcement literature has explored factors behind weak enforcement in developing countries. These factors include informality (Emran and Stiglitz, 2005; Gordon and Li, 2009), tax administration incentives (Khan et al., 2016), information asymmetry (Best et al., 2015) and cultural norms (Luttmer and Singhal, 2014). A strand of the tax enforcement literature has analyzed the effectiveness of different policies aimed at encouraging tax compliance. For instance, a series of papers explore the effectiveness of third-party information reporting in improving tax compliance (Carrillo et al., 2017; Kleven et al., 2011; Naritomi, 2018; Pomeranz, 2015). However, this literature has so far not analyzed the effectiveness of deadline extensions in tax base broadening. This paper fills the gap in the tax enforcement literature by quantifying the relationship between deadline extensions and the probability of tax filing.

The findings are important because they take a step towards a better understanding of tax broadening by analyzing a unique policy that has been in regular use in a low tax compliance, developing country context. The paper provides evidence that time is not a binding constraint to tax filing in this context, making deadline extensions ineffective in broadening the tax net.

The remainder of the paper is organized as follows. Section 3.2 describes the institutional context and the dataset used for the econometric analysis. Section 3.3 explains the empirical methodology along with an extensive discussion of the results. Section 3.4 tests various underlying mechanisms that help explain the results. Section 3.5 discusses some of the confounding factors and undertakes multiple robustness checks to ensure that the main findings are consistent. Finally, Section 3.6 concludes.

3.2 Institutional Context and Data

3.2.1 Tax Compliance and Filing Process in Pakistan

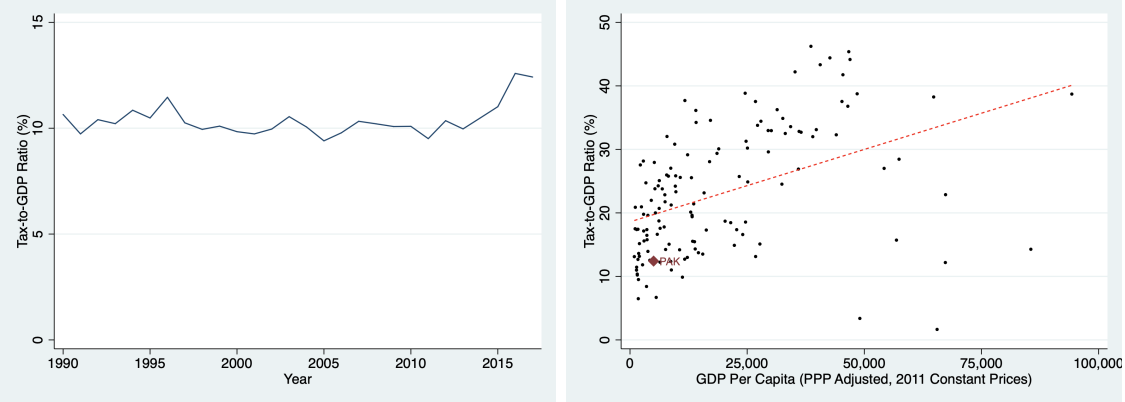
Pakistan's tax administration is broadly organized into two levels - at the federal level and at the provincial levels. Constitutionally, the federal government is responsible for the collection of Income Tax, Sales Tax on Goods and Federal Excise Duty via the Board of Revenue (FBR). To achieve this objective, the FBR has a number of field offices staffed by civil servants, located in major cities across the country.[33]

Pakistan has struggled to improve its tax collection as a percentage of GDP over the last decade, which has stagnated around 10% since 1990 (IMF, 2017). Even when compared to countries with similar GDP per capita, Pakistan collects a lower proportion of taxes as a share of GDP (Figure 3.1). For instance, India collected 17.76% and Nepal collected 20.57% of GDP in tax revenues in 2017 (IMF, 2017).

Digging deeper reveals that Pakistan collects only 3-4% of GDP in income taxes (IMF, 2017). Furthermore, Pakistan has a low income tax base with less than 1% of the Pakistani population filing their income tax return in 2017. This is admittedly a crude measure of tax compliance but provides valuable insights into the structure and administration of income taxes in the country.

[33]For a detailed discussion of the FBR's organizational structure, please refer to Appendix 3.A.

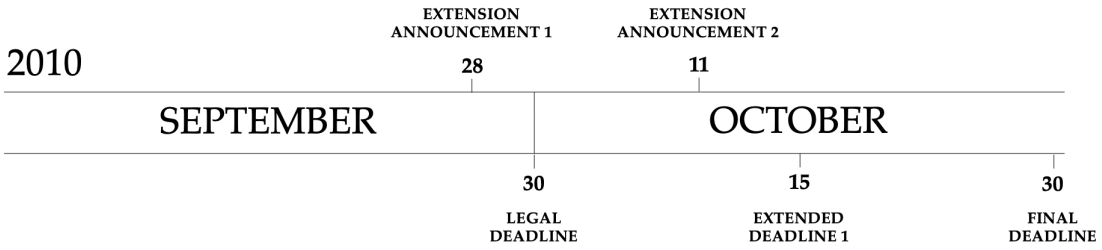
Figure 3.1: Pakistan's Stagnant and Low Tax-to-GDP Ratio



Source: International Monetary Fund and University of Oxford

To address this problem, the FBR has experimented with various policies to broaden the tax base.[34] One policy has stood out in the last decade for its consistent and increasingly frequent usage over time: giving extensions to the tax filing deadline with the underlying assumption that time is a binding constraint to tax filing.[35]

Figure 3.2: Illustration of Deadline Extensions in 2010



Explanation: In the tax year 2010, the original deadline for individuals and AOPs was 30 September 2010. This deadline was extended on 28 September 2010 to 15 October 2010.[36] On the 11 October, the FBR extended the deadline yet again until 30 October, which ended up being the final deadline.

These extensions were given out consistently to two out of three types of tax filers: individuals and Association of Persons (AOPs).[37] However, very few extensions were given to companies. This can be seen from the fact that individuals and AOPs were

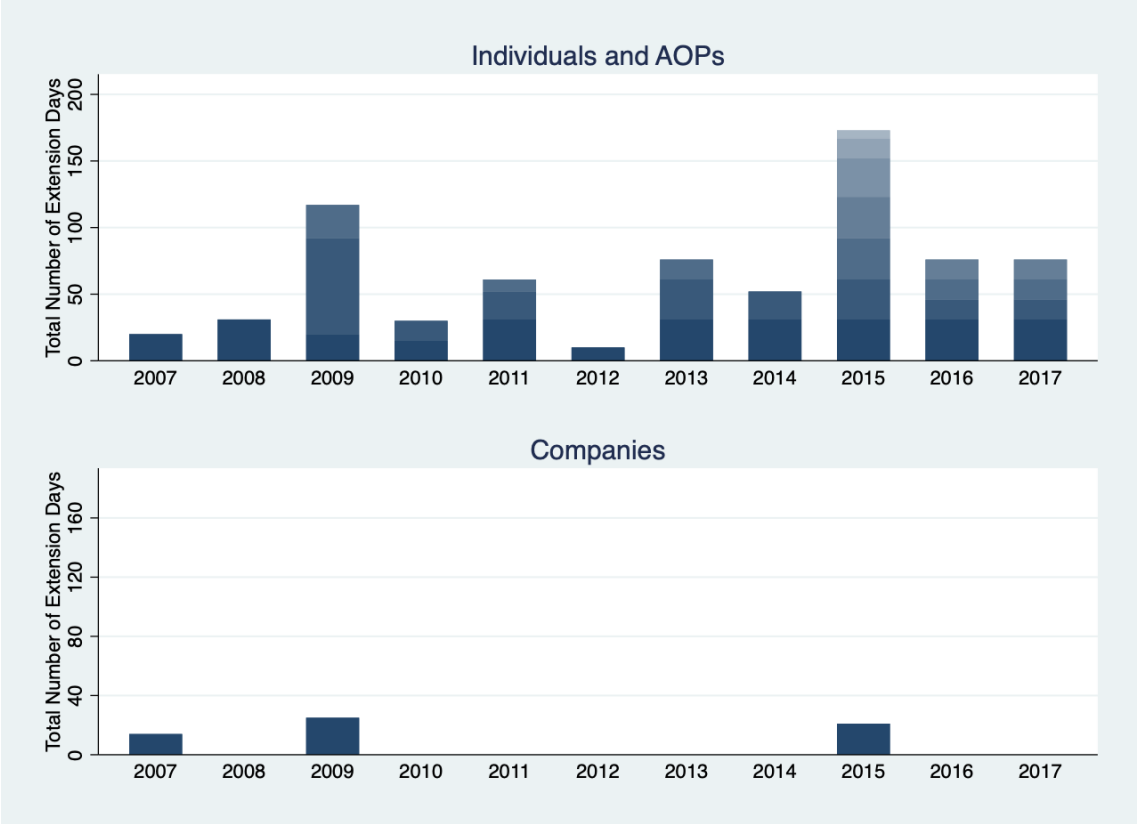
[34]The FBR has been repeatedly criticized for relying on a narrow tax base, especially during fiscal crises in the recent past. For instance, the fiscal crises of 2008, 2013 and most recently of 2018 put the spotlight on the FBR and its inability to create more fiscal space in the country.

[35]The underlying assumption by officials at the FBR was explicitly stated as the justification for the extensions in the official circulars that announced the deadline extensions. The circulars are discussed in detail in Section 3.2.2

[37]AOPs are unincorporated businesses.

given an average of 2.82 extensions per year between 2007 and 2017, while companies were only given 0.27. This amounted to an average of 66 extra days for filing the tax return for individuals and AOPs per year, and only 5.5 days for companies. A summary of the total days of extension and number of extensions for different types of tax filers is given in Figure 3.3.

Figure 3.3: Extensions Regularly Given to Individuals and AOPs



Notes: The figure shows the total number of extension days given to different types of tax filers across time. It also shows the duration of each extension, shaded by different colors in the figure.

The Income Tax Ordinance (2001) is the current legislative instrument that governs income tax in Pakistan. Section 114 of the ordinance explains the type of persons that are liable to file the tax return.[38] This section also delegates the design and implementation of tax returns to the FBR.

Based on this, the FBR has explained the exact format of the income tax return as part of the Income Tax Rules (2002). Rule 34 defines the format of the tax return

[38]For example, all companies, all individuals with a declared income above the minimum tax threshold and all non-profit organizations are required to file the tax return.

while Rule 73 explains the ways in which it can be submitted. According to Rule 73, the tax return can be submitted by courier, delivered by hand in the tax office or submitted electronically. The tax returns that are submitted by courier or by hand are manual returns while the ones submitted from FBR's online portal are electronic.[39]

Prior to the tax filing, potential filers are required to register with the FBR by submitting verification documents, such as a copy of the national identification card and a copy of a recent pay slip. The FBR verifies these documents and then assigns a unique tax identification number to the registrant.[40] Completing this one-time process makes potential tax filers eligible to file taxes for all future years.

Once the registration process is complete, individuals can file their tax returns. Manual returns are received by the tax office, their date of submission verified by tax officials and then they are keyed into the IT system. Electronic returns can simply be filed on the FBR's online portal and submitted. If the tax filer owes the government income tax after adjusting for all tax credits, then the return cannot be submitted until the money is deposited, and proof of the payment is attached. Some additional technical definitions from the Income Tax Ordinance (2001) that are used in this paper are given in Appendix 3.B.

3.2.2 Administrative Tax Data and Extension Circulars

This paper utilizes two sources of data to conduct the empirical analysis. The first of these is an administrative dataset of all tax returns filed from 2007 to 2017. This dataset was acquired after receiving an approval from the federal cabinet of the

[39]Both manual and electronic tax returns are of the same format.

[40]Recently, this tax identification number was made the same as the national identification number in order to simplify the process. The registration process can be done either manually or electronically.

Government of Pakistan.[41]

The key variable used in the analysis is that of the tax filing date of all returns. This is a system generated date in cases where the tax return is filed electronically, and manually entered date on the tax return that is verified by the tax authorities when the tax return is filed manually. In both cases, the variable represents the date when the FBR receives the tax return. Those who submit their returns later than the official deadline can be penalized based on this date record.

The dataset also contains a variety of other information for each tax return, including the tax filing year, taxable income declared, total income declared, tax paid, tax credits claimed, city of tax filer, the tax office jurisdiction relevant to the tax filer, taxpayer's principal activity and tax filer type (individual, company or AOP).[42] From tax year 2014 onwards, the Government of Pakistan legislated that anyone who files the tax return would also be required to file a wealth statement as per Section 116 of the Income Tax Ordinance (2001). Hence, the dataset also has information on the total asset declaration of a tax filer as per their wealth statement from 2014.

Of the variables mentioned above, the taxpayer's principal activity, that contains information on the main economic activity of the tax filer, is an optional field. This variable is present in 58% of observations in the entire dataset and was not declared in the other 42%. This variable has a detailed disaggregation of the tax filer's economic activity. Where present, the variable is able to identify the broad sector of work in the economy (e.g. agriculture) and follow that up with details about a specific activity (e.g. growing rice).

[41] To ensure that the dataset did not violate any data privacy laws, the tax identification numbers of tax filers were anonymized. Officials at the FBR replaced all tax identification numbers with dummy numbers and removed all names and residential addresses from the dataset before sharing it. However, the dummy identification number for each taxpayer in the dataset is unique. Hence, this is a panel dataset which allows for the incorporation of individual fixed effects in the empirical results.

[42] This information needs to be filled up in the tax return as per Rule 34 of the Income Tax Rules (2002).

Each observation in the dataset represents a unique tax return filed by an individual, company or AOP. Hence, by definition, there is only one observation of a tax filer per tax year. In total, the entire dataset consists of 11,362,104 observations and is disaggregated by year in Table 3.1.[43]

Table 3.1: Pakistan's Administrative Tax Data From 2007 to 2017

Tax Year	Total Observations
2007	914,872
2008	768,113
2009	855,275
2010	826,408
2011	829,100
2012	835,874
2013	936,441
2014	1,115,238
2015	1,264,649
2016	1,431,349
2017	1,584,785
Total	11,362,104

Notes: One observation represents a unique tax return filed by an individual, AOP or company for a particular tax year. An individual, AOP or company can only file the tax return once per tax year.

The second source of data is put together from official documents of the FBR which contain announcements of deadline extensions to the tax filing date. These announcements are made through official circulars that are administrative instruments used to communicate changes in policies to the general public.[44] Such circulars are widely used by all government departments in Pakistan to make announcements and are often forwarded to media outlets so that the news can be disseminated.[45] In the past, the FBR has used circulars for multiple purposes such as clarifying the meaning of legal amendments, announcing changes in tax treaties with other countries and explaining administrative procedures.

This dataset was created by extracting information about all deadline extensions in

[43]For detailed summary statistics of the administrative tax data, please refer to Appendix 3.C.

[44]The FBR has made all circulars publicly available on their website (www.fbr.gov.pk).

[45]The news about the deadline extensions is extensively covered in media outlets. Some examples of the media coverage are Arshad (2016), Daily Times (2016), Dawn News (2008) and Geo News (2016).

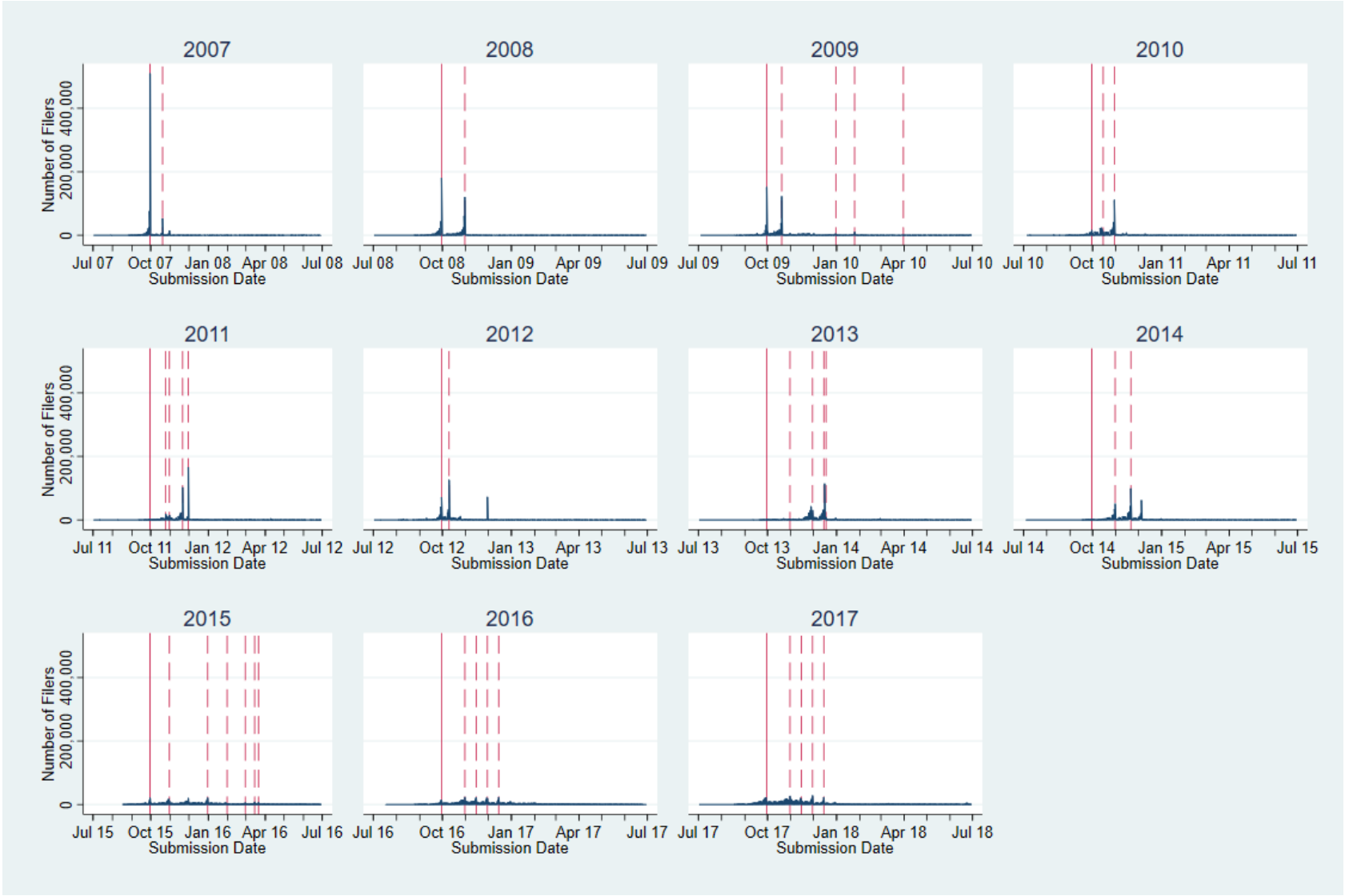
the abovementioned circulars from 2007 to 2017. It contains information such as the date of the announcement of the extension, the new extended deadline, the tax year applicable, class of taxpayers the extension applied to, number of extensions given in a tax year, number of days the deadline is extended for and the circular reference number.

To make sure that extensions were primarily given only during this period, all circulars from 2001 to 2006 were also analyzed. This exercise showed that there was only one extension given prior to 2007, in the year 2005, where the extension was given under exceptional circumstances as the country suffered from a major earthquake. In total, the FBR gave 31 extensions to individuals and AOPs, and 3 extensions to companies from 2007 to 2017.[46]

Figures 3.4 and 3.5 provide a description of daily filing and extensions using a combination of the administrative tax data and extension circulars. These two graphs firstly shows that individuals, AOPs and companies prefer to file their tax returns on or closer to the deadline or extended deadline. This is because most spikes in daily filing are on the date of the deadline or extended deadline. Secondly, companies keep their behavior of filing most tax returns on the date of the legal deadline consistent through time, except when they receive a deadline extension where they file their tax returns on the extended deadline. Thirdly, individuals and AOPs initially also prefer to file most tax returns on the legal deadline. However, as time passes and extensions are given out regularly, daily filing on the legal and extended deadlines drops.

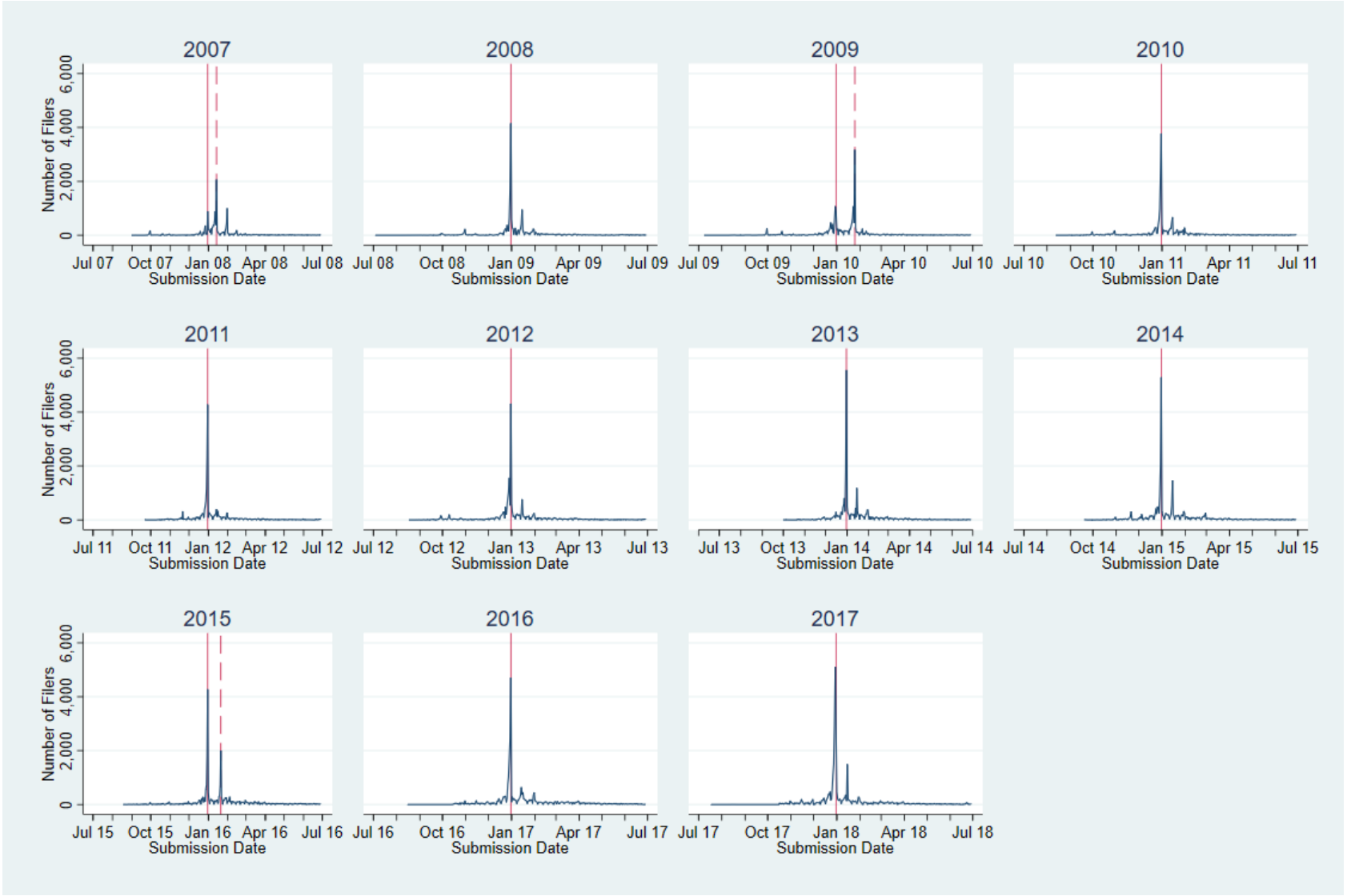
[46]For further details about the extension circulars, please refer to Appendix 3.D.

Figure 3.4: Daily Tax Filing and Extensions For Individuals and AOPs



Notes: The solid line represents the filing deadline in the tax law and the dotted lines represent extensions

Figure 3.5: Daily Tax Filing and Extensions For Companies



Notes: The solid line represents the filing deadline in the tax law and the dotted lines represent extensions

3.3 Methodology and Results

3.3.1 Extensions are Associated with Delayed Filing of Tax Returns

To measure the association between the deadline extensions and tax filing behavior, I utilize the variation in the number of extensions given out to individuals and companies.[47] As a first step, I quantify the relationship between extensions and the probability of filing after the submission deadline defined in the tax law. This would inform us about whether tax filers are likely to delay the filing of their tax returns when they are given a deadline extension. For this purpose, I use a linear probability model with the empirical specification given in Equation 3.1.[48]

$$Delay_{it} = \phi_0 + \phi_1 Extension_{it} + \phi_2 X_{it} + \delta_t + \gamma_i + \epsilon_{it} \tag{3.1}$$

Here, $Delay_{it}$ is a dummy variable that is 1 if tax filer i files the return after the legal deadline defined in the tax law for tax year t, and 0 otherwise.[49] $Delay_{it}$ captures the likelihood of a delay in the filing of the tax return.

$Extension_{it}$ is a dummy variable that is 1 if tax filer i is given one or more extensions to file the return in tax year t, and 0 otherwise. For instance, individuals were given 2 extensions in tax year 2010 while companies were given 0. Hence, the $Extension_{it}$ would be 1 for individuals and 0 for companies in tax year 2010. The covariates are represented by X_{it}, the time fixed effect by δ_t, individual fixed effects by γ_i.[50]

Please note that I run all specifications in this paper first with a sample of all

[47] Deadline extensions were regularly given out to individuals and AOPs but not to companies.

[48] I also runs probit and logit specifications in addition to the linear probability model discussed here. The results from the probit and logit specifications are similar to results from the linear probability model and are discussed in Section 3.5.

[49] The legal deadline for individuals and AOPs is 30 September, while that for companies is 31 December. For further details, please refer to Appendix 3.B

[50] Individual fixed effects control for each unique tax filer in the dataset.

individuals and companies, and then with a sample of all AOPs and companies. I present the results for the individual-company and AOP-company comparison separately throughout the paper. Since AOPs and companies are both businesses and are expected to be similar in their behavior, the AOP-company results act as an important check on the individual-company results. Presenting the two analyses separately also has an added advantage of helping us understand the differential response of individuals and unincorporated businesses to extensions. In both of these analyses, companies represent the counterfactual where they only receive extensions in exceptional circumstances.

To measure the magnitude of delay by tax filers when they are given deadline extensions, I change the outcome variable from a binary variable as in the LPM to the number of days after the original deadline as per the tax law that the tax filer submits the return. I also change the independent variable from a binary variable as in the LPM to the number of extensions. The specification is given in Equation 3.2.

$$DelayDays_{it} = \beta_0 + \beta_1 NumExtension_{it} + \beta_2 X_{it} + \delta_t + \gamma_i + \epsilon_{it} \tag{3.2}$$

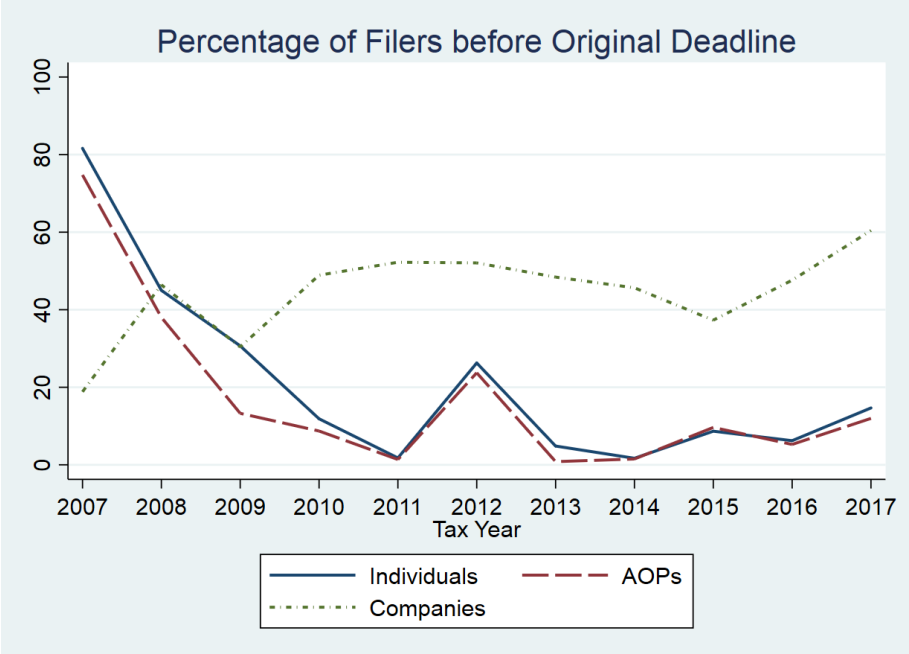
Here, $DelayDays_{it}$ represents the number of days that a person i files the tax return after the original deadline set in the Income Tax Ordinance for tax year t. For instance, if an individual files the tax return for tax year 2010 on 15$^{\text{th}}$ October 2010, while the original deadline set by the law was 30$^{\text{th}}$ September 2010, $DelayDays_{it}$ would be 15. Similarly, if an individual files the tax return for tax year 2010 on 15$^{\text{th}}$ September 2010, $DelayDays_{it}$ would be -15.

$NumExtensions_{it}$ represents the number of extensions that the government gives to tax filer i in tax year t. For instance, the government gave two extensions to individuals to file their return while it gave none to companies in tax year 2010.

Hence, *NumExtensions* would be 2 for individuals and 0 for companies in tax year 2010.

A visual depiction of how filing behavior changes with extensions over time is given in Figures 3.6 and 3.7. Figure 3.6 shows that the percentage of individuals who file their return before the original deadline legislated in the tax law decreases dramatically with time. Around 80% of individuals and AOPs filed their tax return before the original deadline in 2007 while this number decreased to less than 20% over time. On the other hand, the percentage of companies that file before the original deadline is stable across time. Companies were only given three extensions in 2007, 2009 and 2015. The graph shows that the only case where the share of companies that file before the original deadline dips are for these three years in which extensions were given out to companies. AOPs were given the same number of extensions as individuals and depict very similar behavior over time.

Figure 3.6: Decreasing Compliance with Original Deadline by Individuals and AOPs



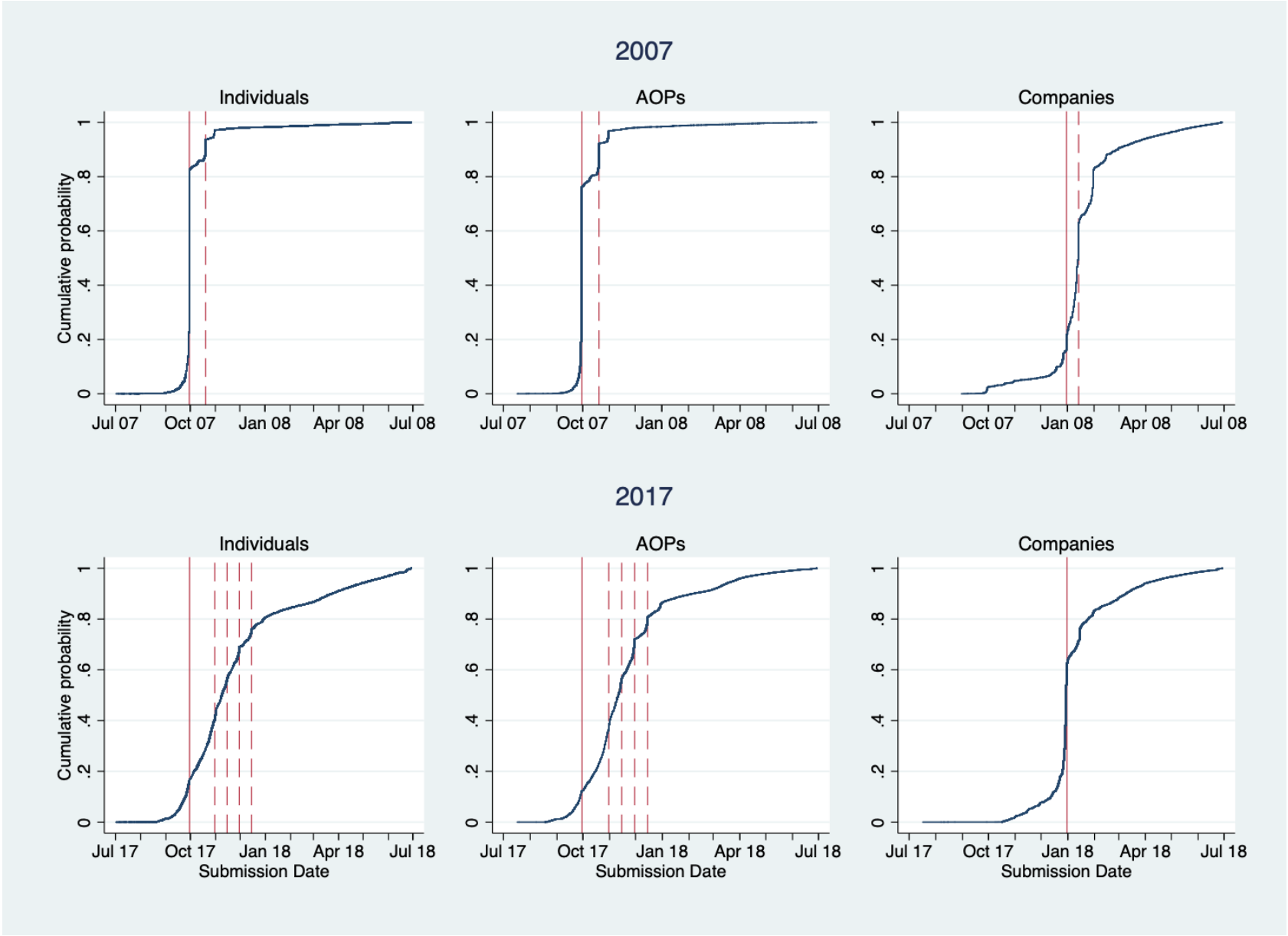
Notes: Companies were given deadline extensions only in 2007, 2009 and 2015. The compliance of companies remained stable apart from these three years when companies also delayed the filing of their tax returns after the deadline set in the tax law.

Figure 3.7 shows the cumulative distributions of tax filing dates for individuals, AOPs and companies for 2007 and 2017 along with the original deadlines and ex-

tensions. This figure shows that the cumulative distribution of individuals and AOPs has shifted to the right over time, implying that individuals have increasingly delayed their return filing when extensions were given.[51] In stark contrast, the cumulative distribution of companies has remained stable.

[51] Graphs for each successive year shows a pattern of gradual shift with time. Figure 3.7 shows the results for 2007 and 2017 that summarize the same pattern with the distribution eventually shifting to the right.

Figure 3.7: Delayed Filing by Individuals and AOPs



Notes: The solid line represents the filing deadline in the tax law and the dotted lines represent extensions

The results from the linear probability model are presented in Table 3.2. Specification (4) of both panels show that one deadline extension is associated with an increase in the probability of individuals and AOPs to file after the legal deadline by 38.3 and 33.5 percentage points respectively. I also run probit and logit models in addition to the linear probability model discussed here and find consistent results.[52]

Table 3.2: Extensions are Associated with Increased Probability of Delayed Filing

Panel A: Individual-Company Comparison

Dep Variable: Delay	(1)	(2)	(3)	(4)
Extension Dummy	0.307***	0.313***	0.374***	0.383***
	(0.002)	(0.002)	(0.002)	(0.002)
Observations	10690722	10220544	10220544	10220544
Adjusted R^2	0.012	0.018	0.307	0.292
Controlling for Covariates	No	Yes	Yes	Yes
Time Fixed Effects	No	No	Yes	Yes
Individual Fixed Effects	No	No	No	Yes

Panel B: AOP-Company Comparison

Dep Variable: Delay	(1)	(2)	(3)	(4)
Extension Dummy	0.337***	0.292***	0.385***	0.335***
	(0.002)	(0.002)	(0.002)	(0.002)
Observations	732669	718823	718823	718823
Adjusted R^2	0.121	0.133	0.201	0.093
Controlling for Covariates	No	Yes	Yes	Yes
Time Fixed Effects	No	No	Yes	Yes
Individual Fixed Effects	No	No	No	Yes

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table uses a linear probability model to measure the association between extensions and the probability of filing after the deadline set in the tax law. Panel A compares the behavior of all individuals and companies while Panel B compares AOPs and companies that filed tax returns from 2007 to 2017. Column (1) conducts a simple OLS regression. Column (2) controls for covariates such as taxable income, city and tax office. Columns (3) and (4) additionally control for time fixed effects and individual fixed effects. Standard errors are clustered at the level of the tax filer.

The results that quantify the magnitude of delay in tax filing are presented in Table 3.3. Specification (4) shows that giving one extension to tax filers is associated with

[52]The results from the probit and logit models are presented in Section 3.5

individuals delaying tax filing by 20.24 days and AOPs by 16.21 days, on average. The average duration of an extension is 23.09 days which means that individuals delay their tax filing by 88% of the extension duration while AOPs delay by 70% of the extension duration. This amounts to a delay of 0.88 and 0.70 day for each extension day for individuals and AOPs respectively. Hence, when individuals and AOPs are given extra time to file their returns, they simply put off the filing to a later date.

Table 3.3: Extensions are Associated with Delayed Filing

Panel A: Individual-Company Comparison

Dep Variable: Delay (Days)	(1)	(2)	(3)	(4)
Number of Extensions	17.262***	16.403***	15.062***	20.243***
	(0.043)	(0.043)	(0.187)	(0.338)
Observations	10690722	10220544	10220544	10220544
Adjusted R^2	0.020	0.030	0.065	0.073
Controlling for Covariates	No	Yes	Yes	Yes
Time Fixed Effects	No	No	Yes	Yes
Individual Fixed Effects	No	No	No	Yes

Panel B: AOP-Company Comparison

Dep Variable: Delay (Days)	(1)	(2)	(3)	(4)
Number of Extensions	6.319***	5.577***	7.654***	16.214***
	(0.174)	(0.190)	(0.280)	(0.408)
Observations	732669	718823	718823	718823
Adjusted R^2	0.003	0.020	0.033	0.036
Controlling for Covariates	No	Yes	Yes	Yes
Time Fixed Effects	No	No	Yes	Yes
Individual Fixed Effects	No	No	No	Yes

Standard errors in parentheses

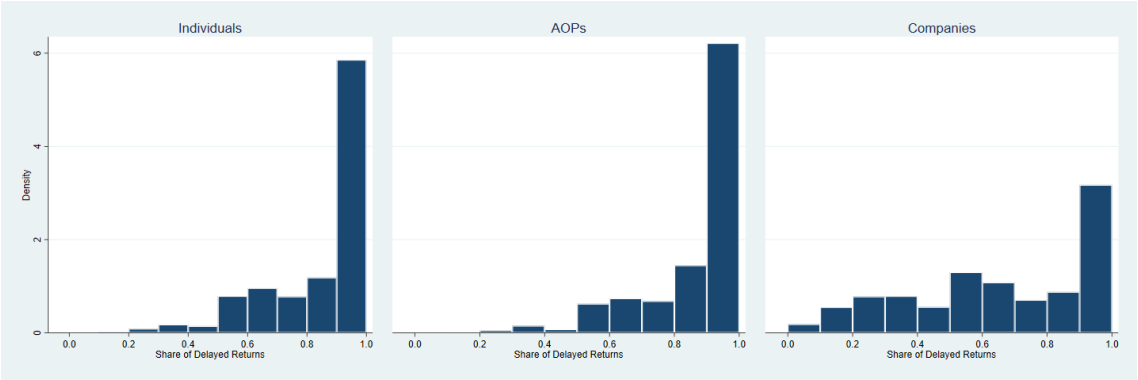
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table measures the association between the number of extensions and the number of days the tax return is filed after the deadline set in the tax law. Panel A compares the behavior of all individuals and companies while Panel B compares AOPs and companies that filed tax returns from 2007 to 2017. Column (1) conducts a simple OLS regression. Column (2) controls for covariates such as taxable income, city and tax office. Columns (3) and (4) additionally control for time fixed effects and individual fixed effects. Standard errors are clustered at the level of the tax filer.

Figure 3.8 shows that distribution of delayed filing for individuals, AOPs and com-

panies. Here, the share of delayed returns for a tax filer is defined as the total number of returns filed by the tax filer after the legal deadline divided by the total number of returns filed by the tax filer from 2007-2017. The figure shows that most individuals and AOPs delay the filing of almost all of their tax returns. This shows that the results are not being driven by a few outliers.

Figure 3.8: Distributions of Delayed Filing by Tax Filer Type



Notes: The share of delayed returns for a tax filer is defined as the total number of returns filed by the tax filer after the legal deadline divided by the total number of returns filed by the tax filer from 2007-2017.

3.3.2 Extensions are Associated with Reduction in Compliance with Final Deadline

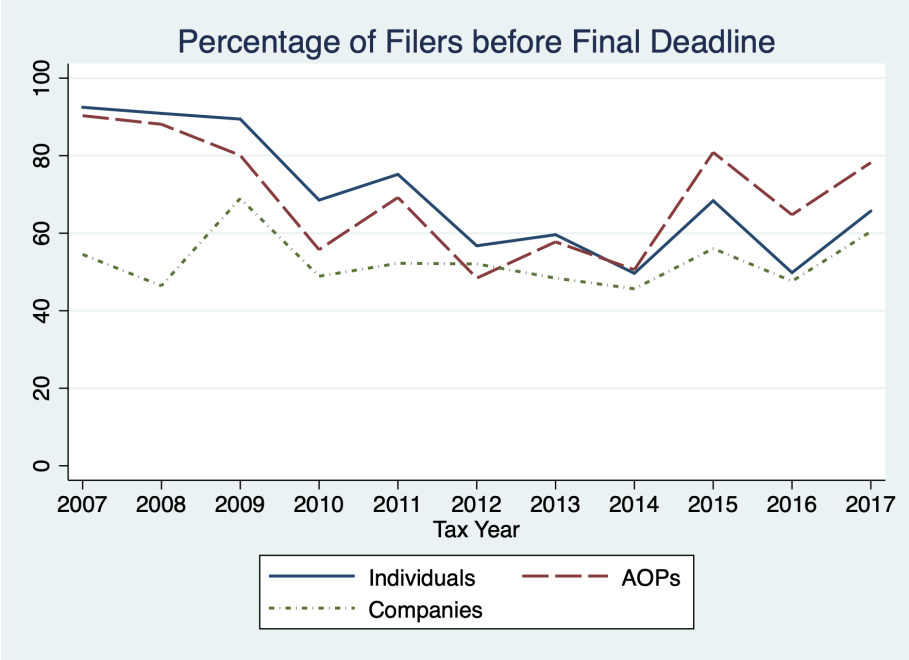
Repeated extensions might reduce the credibility of the deadline and resultantly reduce compliance with the final extended deadline. To conduct this analysis, I run the linear probability model in Equation 3.3.

$$DelayFinal_{it} = \beta_0 + \beta_1 NumExtension_{it} + \beta_2 X_{it} + \delta_t + \gamma_i + \epsilon_{it} \tag{3.3}$$

Here, $DelayFinal_{it}$ is a dummy variable that is 1 if tax filer i files the return after the final extended deadline and 0 otherwise. $NumExtensions_{it}$ represents the number of extensions given out to taxpayer i in tax year t. The covariates are represented by X_{it}, the time fixed effects by δ_t, individual fixed effects by γ_i.

A visual depiction is presented in Figure 3.9 which shows that compliance of individuals and AOPs with the final extended deadline declines with time, while the trend for companies remains stable. Specifically, the percentage of filers before the final extended deadline drops by 26.77 and 12.09 percentage points respectively for individuals and AOPs from 2007 to 2017. The compliance for companies instead increases by 5.90 percentage points for companies.

Figure 3.9: Decreasing Compliance with Final Deadline by Individuals and AOPs



The results are presented in Table 3.4. Specification (4) of both panels shows that one extension is associated with a decrease in the compliance with the final deadline by 4.5 and 1.1 percentage points for individuals and AOPs respectively.

3.3.3 Extensions are not Associated with Increased Tax Filing Probability

The FBR explicitly gave deadline extensions throughout the period of the study under the assumption that it encourages filing behavior and that time is a binding constraint to tax filing. To test whether this assumption is true, I convert the panel dataset used in the previous section into a balanced panel where tax filers in a given

Table 3.4: Extensions are Associated with Reduced Compliance with Final Extended Deadline

Panel A: Individual-Company Comparison

Dep Var: Delay Final	(1)	(2)	(3)	(4)
Number of Extensions	0.018***	0.017***	-0.032***	0.045***
	(0.000)	(0.000)	(0.000)	(0.001)
Observations	10690722	10220544	10220544	10220544
Adjusted R^2	0.005	0.021	0.114	0.096
Controlling for Covariates	No	Yes	Yes	Yes
Time Fixed Effects	No	No	Yes	Yes
Individual Fixed Effects	No	No	No	Yes

Panel B: AOP-Company Comparison

Dep Var: Delay Final	(1)	(2)	(3)	(4)
Number of Extensions	-0.047***	-0.039***	-0.033***	0.011***
	(0.000)	(0.000)	(0.001)	(0.001)
Observations	732669	718823	718823	718823
Adjusted R^2	0.033	0.051	0.076	0.063
Controlling for Covariates	No	Yes	Yes	Yes
Time Fixed Effects	No	No	Yes	Yes
Individual Fixed Effects	No	No	No	Yes

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table uses a linear probability model to measure the association between extensions and the probability of filing after the final extended deadline. Panel A compares the behavior of all individuals and companies while Panel B compares AOPs and companies that filed tax returns from 2007 to 2017. Column (1) conducts a simple OLS regression. Column (2) controls for covariates such as taxable income, city and tax office. Columns (3) and (4) additionally control for time fixed effects and individual fixed effects. Standard errors are clustered at the level of the tax filer.

year were assigned the dummy variable of value 1 when they filed and 0 when they did not. Then, I use the variation in the extensions given out to individuals/AOPs and companies to draw conclusions about the probability of filing. The empirical specification is given in Equation 3.4.

$$FileProbability_{it} = \theta_0 + \theta_1 NumExtensions_{it} + \theta_2 X_{it} + \delta_t + \gamma_i + \epsilon_{it} \tag{3.4}$$

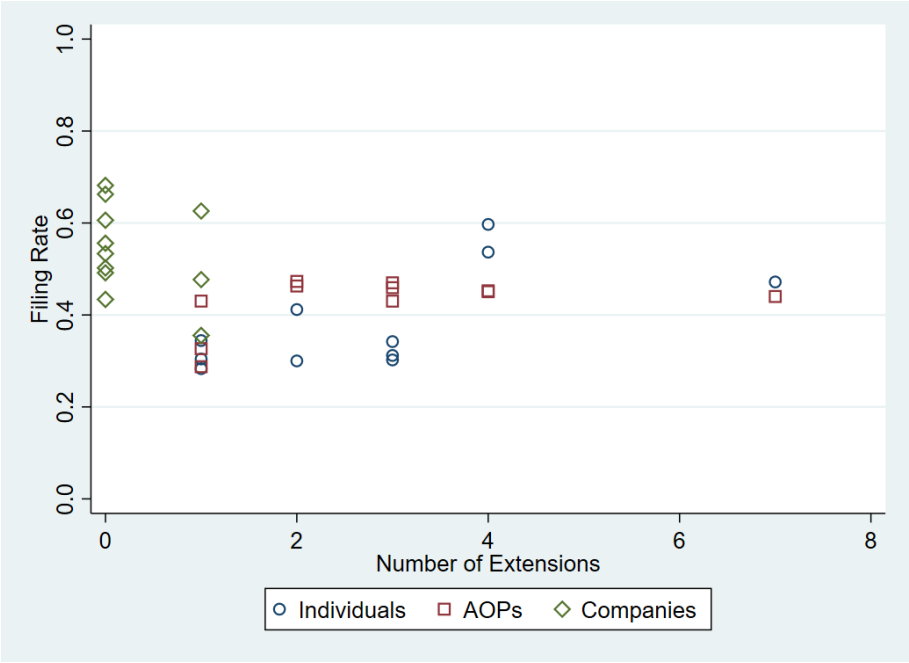
Here $FileProbability_{it}$ is a binary variable that is 1 if taxpayer i filed the tax return in tax year t and 0 otherwise. $NumExtensions_{it}$ represents the number of extensions given out taxpayer i in tax year t.

Figure 3.10 shows how the filing rates for individuals, AOPs and companies vary with the number of extensions given over time.[53] From this graph, we do not observe an association between extensions and filing rates.

The results are presented in Table 3.5. Specification (4) shows that the association of extensions with the probability of filing is close to zero for individuals and negative for AOPs. The coefficient size for the individual-company comparison is 0.002, which means that even in the best case scenario, extensions improve the filing probability by 0.2 percentage points. This amounts to only 4,750 additional tax filers per deadline extension where on average 966,031 individuals file their tax returns in the sample. This represents a growth of only 0.4% in the tax base per deadline extension. The results for AOPs also depict that deadline extensions do not encourage tax filing. Crucially, the result that deadline extensions do not encourage tax filing behavior for both individuals and AOPs becomes even stronger when we exclude certain time periods from the analysis (Table 3.6). The FBR introduced tax filing incentives for individuals in 2015 where filers had to pay lower withholding taxes on transactions. Excluding 2015-2017 from the analysis makes the coefficient for extensions negative

[53]The filing rate for each individuals is the number of individuals who file in a particular year divided by the total number of unique individuals who filed from 2007-2017. It is similarly calculated for AOPs and companies.

Figure 3.10: No Obvious Association Between Extensions and Filing Rate



Notes: The filing rate is calculated separately for each type. For instance, the filing rate for each individuals is the number of individuals who file in a particular year divided by the total number of unique individuals who filed from 2007-2017.

for individuals and close to zero for AOPs.

3.3.4 Threats to Identification

The empirical strategy of this paper relies on a difference-in-difference design with an important caveat that data for years in which no deadline extensions were given (i.e. prior to 2007) is not available. In ideal circumstances, access to such data would have helped test the parallel trends assumption to establish causality. Such data could be used to run placebo difference-in-difference tests to check whether individuals, AOPs and companies had parallel trends of tax filing behavior before extensions were given. This would increase our confidence in whether parallel trends were likely to have persisted after 2007 and consequently helped make a causal claim. However, the Board of Revenue in Pakistan was not able to provide such data due to which the paper could not test this identifying assumption.

In this scenario, there could be three main threats to identification. First, individuals

Table 3.5: Deadline Extensions are not Associated with Increased Probability of Tax Filing

Panel A: Individual-Company Comparison

Dep Variable: File Probability	(1)	(2)	(3)	(4)
Number of Extensions	0.034***	0.039***	-0.029***	0.002***
	(0.000)	(0.000)	(0.000)	(0.001)
Observations	28329692	26122950	26122950	26122950
Adjusted R^2	0.014	0.032	0.067	0.086
Controlling for Covariates	No	Yes	Yes	Yes
Time Fixed Effects	No	No	Yes	Yes
Individual Fixed Effects	No	No	No	Yes

Panel B: AOP-Company Comparison

Dep Variable: File Probability	(1)	(2)	(3)	(4)
Number of Extensions	-0.012***	-0.000	-0.032***	-0.026***
	(0.000)	(0.000)	(0.001)	(0.001)
Observations	1585074	1554836	1554836	1554836
Adjusted R^2	0.002	0.019	0.042	0.031
Controlling for Covariates	No	Yes	Yes	Yes
Time Fixed Effects	No	No	Yes	Yes
Individual Fixed Effects	No	No	No	Yes

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table measures the association between the number of extensions and the probability of filing the tax return. Panel A compares the behavior of all individuals and companies while Panel B compares AOPs and companies that filed tax returns from 2007 to 2017. Column (1) conducts a simple OLS regression. Column (2) controls for covariates such as taxable income, city and tax office. Columns (3) and (4) additionally control for time fixed effects and individual fixed effects. Standard errors are clustered at the level of the tax filer.

Table 3.6: Excluding Time Periods Also Shows that Extensions are not Associated with Increased Probability of Tax Filing

Panel A: Individual-Company Comparison

Dep Variable: File Probability	(1) No Incentives	(2) Early Years
Number of Extensions	-0.044***	-0.059***
	(0.001)	(0.001)
Observations	18998160	11874954
Adjusted R^2	0.020	0.005
Controlling for Covariates	Yes	Yes
Time Fixed Effects	Yes	Yes
Individual Fixed Effects	Yes	Yes

Panel B: AOP-Company Comparison

Dep Variable: File Probability	(1) No Incentives	(2) Early Years
Number of Extensions	0.004***	0.010***
	(0.001)	(0.001)
Observations	1130726	706700
Adjusted R^2	0.036	0.044
Controlling for Covariates	Yes	Yes
Time Fixed Effects	Yes	Yes
Individual Fixed Effects	Yes	Yes

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table measures the association between the number of extensions and the probability of filing the tax return. Panel A compares the behavior of all individuals and companies while Panel B compares AOPs and companies that filed tax returns. Columns (1) excludes the tax years 2015-2017 from the analysis because these years introduced incentives for filing tax returns. Column (2) run the analysis for 2007-2011 when the notifications for deadline extensions had an explicitly stated reason of encouraging tax filing. Standard errors are clustered at the level of the tax filer.

and AOPs might not have parallel trends of tax filing when compared to companies. This would introduce bias in the estimates. In theory, companies might be a better counterfactual for AOPs because both of these categories are businesses. However, this cannot be shown in the absence of data prior to 2007 when no extensions were given.

Second, there might be concerns around reverse causality i.e. extensions were given because tax filing rates in a particular year were low. This could be a factor in the last years of the policy when the filing rates before the original deadline were low.

Third, in some years extensions might not have been given solely to broaden the tax base. There might be a situation where the government gets a political benefit from giving extensions which is hard to quantify.

The paper deals with such threats to identification in three ways. First, instead of making a causal claim of the impact of deadline extensions, this paper treats the results between deadline extensions and tax filing as an association, and interprets the observed association accordingly.

Second, the paper conducts several robustness tests given in Section 3.5 to deal with specific threats to identification. For instance, to address the possibility of reverse causality, the paper conducts regressions with a restricted time sample of years in which extensions were explicitly given with the purpose of broadening the tax net. Similarly, to address concerns about the possibility of large companies being on a different trend compared to individuals and AOPs, the paper uses six different methods of excluding large companies from the sample and shows that the results remain consistent.

Finally, the paper shows two sets of results for all analyses conducted: one with a sample of individuals and companies, and the other with a sample of AOPs and companies. Both AOPs and companies are businesses with the only difference between

the two being that AOPs are unincorporated while companies are incorporated. This could make AOPs a better counterfactual compared to individuals. Since both sets of results are in a similar direction, this increases our confidence in the findings, and seeks to address the concern that the behavior of individuals and companies might be different.

3.4 Mechanisms

3.4.1 Earlier Extension Announcements Associated with More Delays

The announcement date of the extension relative to the deadline could be an important mechanism. This is because earlier announcement would provide potential tax filers with more time to change their filing behavior. To test this mechanism, I treat every extension period differently and create an outcome variable that is the number of days after the previous deadline that tax filer files the return. The empirical specification is given in Equation 3.5.

$$DelayDays_{it} = \alpha_0 + \alpha_1 EarlyAnnouncement_{it} + \alpha_2 X_{it} + \delta_t + \gamma_i + \epsilon_{it} \tag{3.5}$$

Here, $DelayDays_{it}$ represents the number of days that tax filer i files the return after the previous extended deadline for tax year t. For instance, if the original deadline for an individual for tax year 2010 was 30 September 2010 which was later extended to 15 October 2010 and then again to 30 October 2010, and if the individual filed the tax return on 17 October 2010, the value for the variable be 2 days.

$EarlyAnnouncement_{it}$ represents the number of days prior to the upcoming deadline that the extension announcement is made for person i in tax year t. For instance, using the previous example of an individual filing on 17 October 2010, if the relevant extension availed by the tax filer was announced on the 11 October 2010 with an

upcoming deadline of 15 October 2019, the value of the variable for the individual will be 4. The covariates are represented by X_{it}, the time fixed effect by δ_t, individual fixed effects by γ_i.

The results are presented in Table 3.7. From specification (4) for both panels, we can see that making the extension announcement a day earlier prior to the upcoming deadline, is associated with an additional delay of 0.44 days for individuals and 0.42 days for AOPs. Hence, when extension announcements are made earlier, tax filers delay the filing of their returns more compared to announcements that are made close to the upcoming deadline.

3.4.2 Past Extensions are Associated with Future Changes in Behavior

Extensions were consistently given out to individuals and AOPs, which means they could learn over time and change their behavior. To test this mechanism, I run the empirical specification given in Equation 3.6.

$$DelayDays_{it} = \beta_0 + \beta_1 NumExtension_{it} + \beta_2 NumExtensionLag_{it} + \beta_3 NumExtensionLag_{it} * FileLag_{it} + \beta_4 FileLag_{it} + \beta_5 X_{it} + \delta_t + \gamma_i + \epsilon_{it} \quad (3.6)$$

Here $NumExtensionsLag_{it}$ represents the lag of extensions for person i. For instance, the government gave 3 extensions to individuals in 2009. Hence, the lag of extensions for individuals in tax year 2010 will be 3. This variable measures whether people learn that the FBR consistently gives extensions and modify their behavior accordingly.

Here, $FileLag_{it}$ is a dummy variable that is 0 if person i did not file the return in tax year $t-1$ and 1 if she did. This variable proxies for familiarity with the tax

Table 3.7: Earlier Announcements Associated with More Delays in Tax Filing

Panel A: Individual-Company Comparison

Dep Variable: Delay (Days)	(1)	(2)	(3)	(4)
Early Announcement (Days)	0.284***	0.286***	0.421***	0.442***
	(0.001)	(0.001)	(0.001)	(0.001)
Observations	4649383	4478578	4478578	4478578
Adjusted R^2	0.039	0.042	0.287	0.321
Controlling for Covariates	No	Yes	Yes	Yes
Time Fixed Effects	No	No	Yes	Yes
Individual Fixed Effects	No	No	No	Yes

Panel B: AOP-Company Comparison

Dep Variable: Delay (Days)	(1)	(2)	(3)	(4)
Early Announcement (Days)	0.209***	0.215***	0.387***	0.419***
	(0.002)	(0.002)	(0.003)	(0.003)
Observations	217378	212741	212741	212741
Adjusted R^2	0.028	0.034	0.307	0.337
Controlling for Covariates	No	Yes	Yes	Yes
Time Fixed Effects	No	No	Yes	Yes
Individual Fixed Effects	No	No	No	Yes

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table measures the relationship between the announcement date of the extension on the number of days the tax return is filed after the deadline set in the tax law. Panel A compares the behavior of all individuals and companies while Panel B compares AOPs and companies that filed tax returns from 2007 to 2017. Column (1) conducts a simple OLS regression. Column (2) controls for covariates such as taxable income, city and tax office. Columns (3) and (4) additionally control for time fixed effects and individual fixed effects. Standard errors are clustered at the level of the tax filer.

filing system. $NumExtensionLag_{it} * FileLag_{it}$ is simply an interaction of the two terms defined above. This interaction term seeks to understand how the learning of people who filed their return in the previous year differs from people who did not. I also run the specification with $DelayFinal_{it}$ instead of $DelayDays_{it}$ as the outcome variable.

The results are presented in Table 3.8. The positive and significant coefficient of 'Extensions Lag' provides evidence that individuals and AOPs learn to change their behavior over time. Specifically, an extension in one year is associated with delayed filing and lower compliance with the final extended deadline in the next year.

Columns (2) and (4) depict variation in learning behavior. Column (2) shows that individuals who file their tax returns and receive an extension in the same year delay their filing less than non-filers while AOPs respond by delaying more than non-filers. Column (4) instead shows that both individuals and AOPs who file their tax returns and receive an extension in the same year reduce their compliance with the final extended deadline less than non-filers in the next year.

3.4.3 Association with Extensions Similar Across Income and Tax Payments

We might expect tax filers that have to pay taxes along with the submission of the return to delay their filing more because they can make interest income on the tax payment by delaying. Additionally, we might also expect the association to vary by income. To test this mechanism, I run the empirical specification in Equation 3.7.

$$DelayDays_{it} = \beta_0 + \beta_1 NumExtension_{it} + \beta_2 NumExtension_{it} * TaxPaid_{it} + \beta_3 TaxPaid_{it} + \beta_4 X_{it} + \delta_t + \gamma_i + \epsilon_{it} \quad (3.7)$$

Table 3.8: Lag of Extensions Associated with Delayed Filing

Panel A: Individual-Company Comparison

Dep Var:	(1) Delay (Days)	(2) Delay (Days)	(3) Delay Final	(4) Delay Final
Number of Extensions	16.650***	16.793***	0.024***	0.024***
	(0.281)	(0.280)	(0.001)	(0.001)
Extensions Lag	10.652***	25.898***	0.062***	0.082***
	(0.267)	(0.279)	(0.001)	(0.001)
Extensions Lag*Filing Lag		-16.045***		-0.021***
		(0.100)		(0.000)
Observations	10220544	10220544	10220544	10220544
Adjusted R^2	0.073	0.142	0.098	0.127
Controlling for Covariates	Yes	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes	Yes
Individual Fixed Effects	Yes	Yes	Yes	Yes

Panel B: AOP-Company Comparison

Dep Var:	(1) Delay (Days)	(2) Delay (Days)	(3) Delay Final	(4) Delay Final
Number of Extensions	13.642***	13.868***	-0.005***	-0.004***
	(0.351)	(0.345)	(0.001)	(0.001)
Extensions Lag	8.483***	8.225***	0.053***	0.068***
	(0.325)	(0.531)	(0.001)	(0.001)
Extensions Lag*Filing Lag		1.866***		-0.014***
		(0.475)		(0.001)
Observations	718823	718823	718823	718823
Adjusted R^2	0.037	0.057	0.073	0.080
Controlling for Covariates	Yes	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes	Yes
Individual Fixed Effects	Yes	Yes	Yes	Yes

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table shows evidence for learning in behavior. Panel A compares the behavior of all individuals and companies while Panel B compares AOPs and companies that filed tax returns from 2007 to 2017. Standard errors are clustered at the level of the tax filer.

Here, $TaxPaid_{it}$ is the amount of tax person i had to pay along with the tax return after adjusting for tax credits in year t. $NumExtension_{it} * TaxPaid_{it}$ is an interaction term that helps understand whether tax filers who pay respond differently to deadline extensions compared to filers who do not have to pay. On average, 10.12% of all tax filers in the dataset pay taxes along with the submission of the tax return. In cases where withholding taxes claimed by the tax filer exceed the final tax liability, they do not have to pay any taxes along with the submission of the return. For instance, most salaried individuals have their tax withheld at source and hence only have to file the tax return. Instead of $TaxPaid_{it}$, I also use $Income_{it}$ which is the taxable income of tax filer i in yeaar t.

The results are presented in Table 3.9 and shows that filers with different levels of tax payments and income respond in the same way to deadline extensions for both individuals and AOPs.[54]

3.5 Robustness

3.5.1 Results Remain Robust With Probit and Logit Models

The main results showing that extensions increase the probability of individuals and AOPs to file after the legal deadline use a linear probability model. To test whether the results are sensitive to the type of probability model, I also run probit and logit models. The results are presented in Tables 3.10 and 3.11.

These tables show that the results remain robust to the type of model used. Specifically, column (3) in both tables show a similar association between extensions and delaying behavior as in the main results using the linear probability model.

[54]The coefficient for the interactions terms in columns (1) and (2) of Table 3.9 are insignificant.

Table 3.9: Heterogeneities in Response to Deadline Extensions

Panel A: Individual-Company Comparison

Dep Var: Delay (Days)	(1)	(2)
Number of Extensions	20.247***	20.241***
	(0.338)	(0.338)
Extensions*Tax Paid	0.003	
	(0.003)	
Extensions*Income		-0.000
		(0.000)
Observations	10220544	10220544
Adjusted R^2	0.073	0.073
Controlling for Covariates	Yes	Yes
Time Fixed Effects	Yes	Yes
Individual Fixed Effects	Yes	Yes

Panel B: AOP-Company Comparison

Dep Var: Delay (Days)	(1)	(2)
Number of Extensions	16.214***	16.214***
	(0.408)	(0.408)
Extensions*Tax Paid	-0.002	
	(0.002)	
Extensions*Income		0.000
		(0.000)
Observations	718823	718823
Adjusted R^2	0.036	0.036
Controlling for Covariates	Yes	Yes
Time Fixed Effects	Yes	Yes
Individual Fixed Effects	Yes	Yes

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table shows heterogeneities in behavior. Panel A compares the behavior of all individuals and companies while Panel B compares AOPs and companies that filed tax returns from 2007 to 2017. Standard errors are clustered at the level of the tax filer.

Table 3.10: Results for Delayed Filing Remain Robust With Probit Model

Panel A: Individual-Company Comparison

Dep Variable: Delay	(1)	(2)	(3)
Extension Dummy	0.242***	0.245***	0.266***
	(0.001)	(0.001)	(0.001)
Observations	10690722	10220336	10220336
Pseudo R^2	0.009	0.016	0.279
Controlling for Covariates	No	Yes	Yes
Time Fixed Effects	No	No	Yes

Panel B: AOP-Company Comparison

Dep Variable: Delay	(1)	(2)	(3)
Extension Dummy	0.287***	0.244***	0.346***
	(0.001)	(0.001)	(0.002)
Observations	732669	718525	718525
Pseudo R^2	0.098	0.111	0.182
Controlling for Covariates	No	Yes	Yes
Time Fixed Effects	No	No	Yes

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table uses a probit model to measure the association between extensions and the probability of filing after the deadline set in the tax law. Panel A compares the behavior of all individuals and companies while Panel B compares AOPs and companies that filed tax returns from 2007 to 2017. Results show average marginal effect using margins command in Stata. Standard errors are clustered at the level of the tax filer.

Table 3.11: Results for Delayed Filing Remain Robust With Logit Model

Panel A: Individual-Company Comparison

Dep Variable: Delay	(1)	(2)	(3)
Extension Dummy	0.228***	0.231***	0.263***
	(0.001)	(0.001)	(0.001)
Observations	10690722	10220336	10220336
Pseudo R^2	0.009	0.016	0.278
Controlling for Covariates	No	Yes	Yes
Time Fixed Effects	No	No	Yes

Panel B: AOP-Company Comparison

Dep Variable: Delay	(1)	(2)	(3)
Extension Dummy	0.279***	0.238***	0.347***
	(0.001)	(0.001)	(0.002)
Observations	732669	718525	718525
Pseudo R^2	0.098	0.110	0.179
Controlling for Covariates	No	Yes	Yes
Time Fixed Effects	No	No	Yes

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table uses a logit model to measure the association between extensions and the probability of filing after the deadline set in the tax law. Panel A compares the behavior of all individuals and companies while Panel B compares AOPs and companies that filed tax returns from 2007 to 2017. Results show average marginal effect using margins command in Stata. Standard errors are clustered at the level of the tax filer.

3.5.2 Results Remain Robust to With Restricted Time Sample

The FBR introduced another policy aimed at encouraging filing behavior from 2015 onward. As part of the Finance Act 2014, the government legislated that all tax filers would have a lower amount of income tax withheld from transactions where withholding tax was applicable compared to non-filers. For instance, tax filers who bought a new car would have to pay a lower withholding tax as part of the final price compared to non-filers. This withholding tax can then be adjusted against the final liability in the tax return. Since this policy could potentially impact individuals, AOPs and companies in a different manner, it could be a confounding factor. To address this, I run the preferred baseline specification with a restricted time sample of returns filed prior to 2015. The preferred baseline specification controls for various covariates, time fixed effects and individual fixed effects.

Another confounding factor is that in the later years of the policy, there were concerns that the extensions were given out due to low number of filers close to the original deadline and due to pressure from tax bar associations. This might create a problem of reverse causality and omitted variables. To address this concern, I run the preferred baseline specification with a restricted time sample of returns filed prior to 2011. This is because the extension circulars from 2007 to 2010 explicitly mention tax broadening and facilitation of tax filers as the main justification of giving the extensions. With this clear and explicit purpose of the extensions, this robustness test addresses the concerns of reverse causality and omitted variables.

Since these two robustness checks are both time exclusion tests, they are clubbed together in Table 3.12 for convenience. The results show that all coefficients are consistent with the main findings.

Table 3.12: Results for Delayed Filing Remain Robust to Time Exclusion

Panel A: Individual-Company Comparison

Dep Var: Delay (Days)	(1) No Incentives	(2) No Incentives	(3) Early Years	(4) Early Years
Number of Extensions	34.196***	27.196***	35.811***	32.791***
	(0.644)	(0.572)	(0.937)	(1.206)
Extensions Lag		41.093***		-0.160
		(0.697)		(1.429)
Extensions Lag*Filing Lag		-22.613***		0.666
		(0.373)		(0.464)
Observations	6168285	6168285	3598456	3598456
Adjusted R^2	0.017	0.047	0.014	0.028
Controlling for Covariates	Yes	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes	Yes
Individual Fixed Effects	Yes	Yes	Yes	Yes

Panel B: AOP-Company Comparison

Dep Var: Delay (Days)	(1) No Incentives	(2) No Incentives	(3) Early Years	(4) Early Years
Number of Extensions	32.170***	29.241***	39.867***	32.892***
	(0.730)	(0.669)	(1.098)	(1.417)
Extensions Lag		17.555***		4.681*
		(1.401)		(2.247)
Extensions Lag*Filing Lag		1.249		5.753**
		(1.255)		(1.806)
Observations	498533	498533	289732	289732
Adjusted R^2	0.017	0.031	0.017	0.022
Controlling for Covariates	Yes	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes	Yes
Individual Fixed Effects	Yes	Yes	Yes	Yes

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table measures the relationship between the number of extensions and the number of days the tax return is filed after the deadline set in the tax law with time-restricted samples. Panel A compares the behavior of all individuals and companies while Panel B compares AOPs and companies that filed tax returns from 2007 to 2017. Columns (1) and (2) exclude the tax years 2015-2017 from the analysis because these years introduced incentives for filing tax returns. Column (3) and (4) run the analysis for 2007-2011 when the notifications for deadline extensions had an explicitly stated reason of encouraging tax filing. All columns control for covariates,such as taxable income, time fixed effects and individual fixed effects. Standard errors are clustered at the level of the tax filer.

3.5.3 Results Remain Robust to Large Company Exclusion

The behavior of large corporations might be different compared to individuals and AOPs.[55] Large corporations also receive more attention from the FBR because they are the largest contributors to income tax revenue. The jurisdiction of these corporations is assigned to Large Taxpayer Units (LTUs) as opposed to the regular Regional Tax Offices (RTOs). LTUs have a higher number of civil servants per tax filer which is related to the extra focus on these large companies. To address this concern, I exclude large companies from the sample by 6 different methods.

The first method excludes companies with a declared income higher than the maximum or lower than the minimum income declared by individual or AOPs for a particular year. For instance, if the maximum income declared by an AOP for the tax year 2007 is PKR 1,000,000 and the minimum is PKR 0, then all companies in the year 2007 that have a declared income above PKR 1,000,000 or lower than PKR 0 are dropped from the sample.[56] The paper refers to this method as 'Min-Max'.

The second method excludes companies with a declared income higher than the 90^{th} percentile or lower than the 10^{th} percentile of income declared by individuals or AOPs for a particular year. I refer to this method as '90-10'. The third method excludes companies with a declared income higher than the 80^{th} percentile or lower than the 20^{th} percentile of income declared by individuals or AOPs for a particular year. The paper refers to this method as '80-20'.

While the first three methods drop companies both above a maximum threshold and below a minimum one, the fourth, fifth and sixth methods simply drop companies above a maximum threshold. The fourth method excludes companies with a declared income higher than the 75^{th} percentile of income declared by individuals or AOPs for a particular year. The paper refers to this method as 'Above 75^{th}'. The fifth

[55] AOPs are usually medium-sized enterprises.

[56] Please note that losses can also be declared in tax returns.

and sixth methods exclude companies with a declared income above the 90^{th} and 99^{th} percentiles of incomes respectively. I refer to these methods as 'Above 90^{th}' and 'Above 99^{th}'. The preferred specifications for all six methods is shown in Table 3.13. Together, these different methods of excluding large companies from the analysis address the potential problem posed by large corporations.

Table 3.13 shows the results from all six methods of excluding large companies in both panels. From this table, we can see that not only are the coefficients across all six methods extremely consistent and stable, but also that the value of the coefficients is very similar to the preferred specification of the main results in Table 3.3 presented in Section 3.3. We can conclude that excluding large companies makes no major difference to the findings of the paper and that it does not confound the results.

Table 3.13: Results For Delayed Filing Remain Robust to Large Company Exclusion

Panel A: Individual-Company Comparison

Dep Var: Delay (Days)	(1) Min-Max	(2) 90-10	(3) 80-20	(4) Above 75^{th}	(5) Above 90^{th}	(6) Above 99^{th}
Number of Extensions	16.712***	16.682***	18.167***	17.277***	17.023***	16.824***
	(0.281)	(0.340)	(0.472)	(0.333)	(0.321)	(0.301)
Extensions Lag	25.779***	26.243***	30.134***	26.658***	26.394***	26.161***
	(0.281)	(0.333)	(0.542)	(0.329)	(0.318)	(0.300)
Extensions Lag*Filing Lag	-16.043***	-15.956***	-15.955***	-15.977***	-15.987***	-16.016***
	(0.100)	(0.100)	(0.101)	(0.100)	(0.100)	(0.100)
Observations	10218248	10156459	10103690	10171393	10180509	10197538
Adjusted R^2	0.142	0.143	0.143	0.142	0.142	0.142
Controlling for Covariates	Yes	Yes	Yes	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Individual Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes

Panel B: AOP-Company Comparison

Dep Var: Delay (Days)	(1) Min-Max	(2) 90-10	(3) 80-20	(4) Above 75^{th}	(5) Above 90^{th}	(6) Above 99^{th}
Number of Extensions	13.900***	14.480***	14.801***	14.890***	14.430***	13.982***
	(0.346)	(0.401)	(0.413)	(0.397)	(0.382)	(0.358)
Extensions Lag	8.219***	8.243***	8.510***	8.741***	8.424***	8.276***
	(0.531)	(0.566)	(0.573)	(0.564)	(0.555)	(0.539)
Extensions Lag*Filing Lag	1.860***	2.693***	2.717***	2.389***	2.348***	2.014***
	(0.475)	(0.483)	(0.484)	(0.482)	(0.480)	(0.476)
Observations	717227	650950	643059	663669	675000	699801
Adjusted R^2	0.057	0.060	0.061	0.059	0.058	0.057
Controlling for Covariates	Yes	Yes	Yes	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Individual Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table measures the relationship between the number of extensions and the number of days the tax return is filed after the deadline set in the tax law. Panel A compares the behavior of all individuals and companies while Panel B compares AOPs and companies that filed tax returns from 2007 to 2017 by excluding large companies from the sample using six different methods. All columns control for covariates, such as taxable income, time fixed effects and individual fixed effects. Standard errors are clustered at the level of the tax filer.

3.5.4 Results Remain Robust with Eligible Individuals Only

The main results in the individual-company comparison shared in Panel A of Table 3.3 use a sample of all individuals who filed their tax return from 2007-2017. This includes individuals below the minimum tax threshold who might not be required by the law to file their tax return but nonetheless do file. To ensure that the results are not being driven by individuals below the tax threshold, and purely by people who are required by law to file, I conduct an additional test that excludes all individuals below the minimum tax threshold for each tax year. The results are presented in Table 3.14, and show the they remain robust with a restricted sample of eligible individuals.

Table 3.14: Individual-Company Comparison with Only Eligible Individuals

Dep Var: Delay (Days)	(1)	(2)
Number of Extensions	19.199***	15.403***
	(0.341)	(0.283)
Extensions Lag		26.146***
		(0.292)
Extensions Lag*Filing Lag		-13.927***
		(0.140)
Observations	5406185	5406185
Adjusted R^2	0.069	0.129
Controlling for Covariates	Yes	Yes
Time Fixed Effects	Yes	Yes
Individual Fixed Effects	Yes	Yes

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table measures the relationship between the number of extensions and the number of days the tax return is filed after the deadline set in the tax law. It compares the behavior of all individuals above the minimum tax threshold and companies that filed tax returns from 2007 to 2017. All columns control for covariates, such as taxable income, time fixed effects and individual fixed effects. Standard errors are clustered at the level of the tax filer.

3.6 Conclusion

This paper contributes to the tax enforcement and broadening literature by analyzing the effectiveness of deadline extensions in a low compliance, developing country context. The analysis uses a combination of administrative data of all tax returns filed in Pakistan from 2007-2017 and data extracted from official government documents containing details about deadline extensions. The findings reveal that extensions are associated with individuals and unincorporated businesses delaying the filing of their tax returns, without encouraging tax filing. This shows that time is not a binding constraint to tax filing, making extensions ineffective in broadening the tax base. Back of the envelope estimates using the average deposit interest rate reveal that this policy is associated with forgone interest revenue of USD 5 million from 2007-2017, which is equivalent to the entire federal budget for Environment Protection in 2019-2020 (Government of Pakistan, 2020).[57]

While revealing the relationship between deadline extensions and taxpayer behavior in detail, the findings lead to two further questions that would add to our understanding. First, a key question is whether the effectiveness of deadline extensions interacts with the enforcement capability of the tax administration in any way. If the enforcement capability of the tax administration is better, regular tax filers might not respond to extensions by delaying their tax filing and individuals who do not usually file might consider the extension as an amnesty scheme.

Second, it is important to know whether deadline extensions could work if they are used in combination with disincentives in the future. If one-off extensions are combined with a future credible commitment to large scale implementation of fines

[57]Individuals approximately pay a total of USD 37 million income tax per year after adjusting for tax credits. We also know that individuals extend their tax filing by an average of 57 days per year due to extensions. Using a daily compound interest rate of 0.0181% from 2007-2017 (International Monetary Fund, 2020), we can estimate the loss in interest revenue from delayed filing. Interest Lost For Individuals= $(37 * (1 + 0.0002)^{57} - 37) * 11 \approx$ USD 4 Million. Similar calculations for AOPs yield a total loss of approximately USD 5 million from 2007-2017.

for non-filing, we could observe extensions broaden the tax net. Such a policy of combining carrots and sticks might be effective. Answering these questions would help us understand deadline extensions in more detail and help contribute to the tax enforcement literature.

References

Allingham, Michael G. and Agnar Sandmo, "Income tax evasion: a theoretical analysis," *Journal of Public Economics*, 1972, *1* (3), 323–338.

Arshad, Muhammad, "FBR extended deadline for filing of tax returns to facilitate taxpayers," *Customs Today Newspaper*, May 2016.

Besley, Timothy and Torsten Persson, "Why Do Developing Countries Tax So Little?," *Journal of Economic Perspectives*, 2014, *28* (4), 99–120.

Best, Michael Carlos, Anne Brockmeyer, Henrik Jacobsen Kleven, Johannes Spinnewijn, and Mazhar Waseem, "Production versus Revenue Efficiency with Limited Tax Capacity: Theory and Evidence from Pakistan," *Journal of Political Economy*, 2015, *123* (6), 1311–1355.

Carrillo, Paul, Dina Pomeranz, and Monica Singhal, "Dodging the Taxman: Firm Misreporting and Limits to Tax Enforcement," *American Economic Journal: Applied Economics*, 2017, *9* (2), 144–164.

Daily Times, "FBR extends income tax returns deadline," September 2016.

Dawn News, "FBR extends tax filing deadline up to October 31," September 2008.

Emran, M. Shahe and Joseph E. Stiglitz, "On selective indirect tax reform in developing countries," *Journal of Public Economics*, 2005, *89* (4), 599–623.

Federal Board of Revenue, "Income Tax Rules, 2002," 2002.

Geo News, "FBR extends date for filing of income tax returns till Nov 30," 2016.

Gordon, Roger and Wei Li, "Tax structures in developing countries: Many puzzles and a possible explanation," *Journal of Public Economics*, 2009, *93* (7), 855–866.

Government of Pakistan, "Income Tax Ordinance, 2001," 2001.

_ , "Federal Budget 2020-21," 2020.

International Monetary Fund, "World Revenue Longitudinal Data," 2017.

_ , "Interest Rates - IMF Data," 2020.

Khan, Adnan, Asim Khwaja, and Benjamin Olken, "Tax Farming Redux: Experimental Evidence on Performance Pay for Tax Collectors," *The Quarterly Journal of Economics*, 2016, *131* (1), 219–271.

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

Luttmer, Erzo F. P. and Monica Singhal, "Tax Morale," *Journal of Economic Perspectives*, 2014, *28* (4), 149–168.

Naritomi, Joana, "Consumers as Tax Auditors," *American Economic Review*, 2019, *109* (9), 3031–3072.

Pomeranz, Dina, "No Taxation without Information: Deterrence and Self-Enforcement in the Value Added Tax," *American Economic Review*, 2015, *105* (8), 2539–2569.

University of Oxford, "Tax Revenue and GDP per Capita," 2017. Our World in Data.

Appendix

3.A Structure of the FBR

This appendix describes the structure of the Federal Board of Revenue in more detail and acts as a supplement to the brief discussion about the FBR in the paper. The FBR operates through two types of tax offices namely, Large Taxpayer Units (LTUs) and Regional Tax Offices (RTOs). LTUs are assigned the jurisdiction of the largest corporations in the country. They typically have more civil servants per case compared to RTOs. There are four LTUs in Pakistan, with two in Karachi, one in Lahore and one in Islamabad. Big corporations have head offices based in the biggest cities in the country. Hence, the number of LTUs across different cities reflects this trend. Since Karachi is the largest economic hub of Pakistan, it has two LTUs. Both Lahore and Islamabad also have head offices of large corporations due to which both cities have one LTU each. This can be seen from Table 3.A.1.

RTOs have the jurisdiction of all individuals, AOPs and companies not assigned to LTUs. These offices are spread across big cities in Pakistan. These tax offices also have smaller tax offices in relatively smaller cities referred to as *Mufassil* areas within the FBR. The spread of RTOs across multiple cities across Pakistan is shown in Table 3.A.1.

Within these tax offices, a senior civil servant with the title of Chief Commissioner is responsible for the whole office. Below the Chief Commissioner, there is a whole hierarchy of tax officials that are assigned to different departments, such as withholding taxes, legal and administration. The jurisdiction of tax officials can be divided along functional, geographic and taxpayer type lines.

The FBR headquarter, located in Islamabad, monitors all of these field offices. The headquarter has its own hierarchy headed by the Chairperson, who is responsible

Table 3.A.1: FBR Field Offices

Type of Tax Office	**City**	**Number**
Large Taxpayer Units	Karachi	2
	Lahore	1
	Islamabad	1
Regional Tax Offices	Karachi	3
	Lahore	2
	Islamabad	1
	Bahawalpur	1
	Faisalabad	1
	Gujranwala	1
	Hyderabad	1
	Multan	1
	Peshawar	1
	Quetta	1
	Rawalpindi	1
	Sahiwal	1
	Sargodha	1
	Sialkot	1
	Sukkur	1

for the whole organization. The FBR itself forms one part of the Finance Ministry within the federal government.

3.B Relevant Definitions from Income Tax Ordinance (2001)

Tax Year: This is usually the period from 1 July to 30 June. For instance, Tax Year 2010 means 1 July 2009 to 30 June 2010 (Section 74).

Association of Persons (AOPs): These are unincorporated businesses and partnerships (Section 80).

Filing deadlines: Individuals and AOPs must file their income tax return by 30 September. For instance, the filing deadline for an individual for the Tax Year 2010 would be 30 September 2010. Companies on the other hand must file their returns by 31 December (Section 118).

Tax Payable: This is the tax that must be paid by the filer along with the income tax return. In other words, it is tax chargeable on the taxable income minus any tax credits such as withholding tax already paid (Section 4).

3.C Summary Statistics

This appendix presents the summary statistics from the administrative tax data of all tax returns filed in Pakistan from 2007 to 2017. It supplements Section 3.2 of the main paper. Table 3.C.1 contains the mean, standard deviation, minimum and maximum values of multiple variables in the dataset. Table 3.C.2 disaggregates this by tax filer type.

Table 3.C.1: Summary Statistics from Tax Data

	Mean	Std. Dev.	Min.	Max.
Taxable Income (PKR)	751,234.80	186,183,120.41	-217,969,772,917.00	101,711,880,998.00
Tax Chargeable (PKR)	383,212.93	38,280,668.88	-95,268,035.00	36,963,745,646.00
Tax Credit (PKR)	328,705.03	33,014,118.04	-11,241,774.00	42,751,050,345.00
Tax Paid (PKR)	14,450.10	2,254,468.04	0.00	3,000,000,000.00
Electronic Filing (%)	0.53	0.50	0.00	1.00
Individual Dummy (1/0)	0.94	0.25	0.00	1.00
AOP Dummy (1/0)	0.04	0.19	0.00	1.00
Company Dummy (1/0)	0.03	0.16	0.00	1.00
Observations	11,362,104			

Notes: This table provides the summary statistics from the dataset of all tax returns filed in Pakistan by any category of tax filer from 2007 to 2017.

Table 3.C.2: Summary Statistics by Tax Filer Type

	Individuals		AOPs		Companies	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Taxable Income (PKR)	573,835.48	15,581,659.16	343,198.23	172,975,062.45	7,910,684.10	1,143,933,113.33
Tax Chargeable (PKR)	100,956.02	1,752,505.35	683,200.52	43,855,861.55	10,305,261.32	233,379,203.45
Tax Credit (PKR)	76,655.88	1,085,156.19	486,286.74	44,011,709.05	9,359,601.83	199,395,423.86
Tax Paid (PKR)	3,820.02	148,985.13	18,134.50	490,233.63	399,908.96	14,093,878.77
Electronic Filing (%)	0.50	0.50	0.79	0.41	1.00	0.05

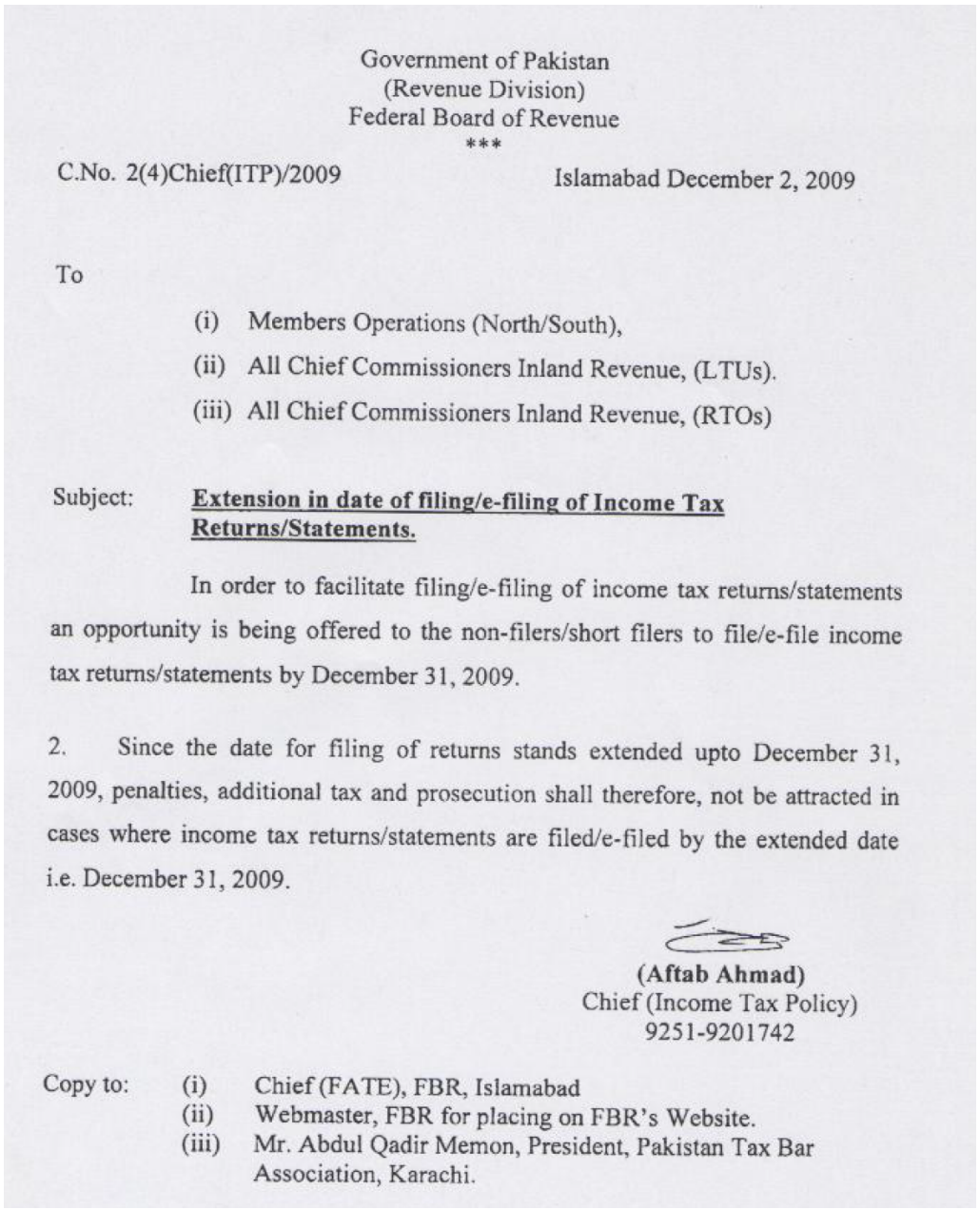
Notes: This table provides the summary statistics from the dataset of all tax returns filed in Pakistan by of tax filer from 2007 to 2017 disaggregated by tax filer type. The income tax law in Pakistan lists three types of tax filers, namely, individuals, Association of Persons (AOPs) and companies.

3.D Extension Circulars

This appendix provides one sample of an extension circular discussed in Section 3.2 of the paper. To recap, the FBR announced extensions to the tax filing deadline via administrative instruments called official circulars. For the empirical analysis, the paper extracted data from these circulars that are publicly available on the FBR website. This data was later merged with the tax data to conduct the analysis.

Figure 3.D.1 provides a sample of the aforementioned circulars from the year 2009. The circular has multiple pieces of information that were extracted and keyed in manually. The extraction process could not be automated because all circulars were scanned copies. Some examples of the information extracted are: circular reference number at the top left (2(4)Chief(ITP)2009), date of the circular at the top right (2nd December 2009), reason for extension explained in the main body of the circular (Facilitate Filing) and extended deadline (31st December 2009).

Figure 3.D.1: Sample Extension Circular for Tax Year 2009



Government of Pakistan
(Revenue Division)
Federal Board of Revenue
***

C.No. 2(4)Chief(ITP)/2009 Islamabad December 2, 2009

To

(i) Members Operations (North/South),
(ii) All Chief Commissioners Inland Revenue, (LTUs).
(iii) All Chief Commissioners Inland Revenue, (RTOs)

Subject: **Extension in date of filing/e-filing of Income Tax Returns/Statements.**

In order to facilitate filing/e-filing of income tax returns/statements an opportunity is being offered to the non-filers/short filers to file/e-file income tax returns/statements by December 31, 2009.

2. Since the date for filing of returns stands extended upto December 31, 2009, penalties, additional tax and prosecution shall therefore, not be attracted in cases where income tax returns/statements are filed/e-filed by the extended date i.e. December 31, 2009.

(Aftab Ahmad)
Chief (Income Tax Policy)
9251-9201742

Copy to:
(i) Chief (FATE), FBR, Islamabad
(ii) Webmaster, FBR for placing on FBR's Website.
(iii) Mr. Abdul Qadir Memon, President, Pakistan Tax Bar Association, Karachi.

Chapter 4

THE GAP WITHIN: IMPACT OF SEPARATE-GENDER POLLING STATIONS ON VOTER TURNOUT

Abstract

Governments in several countries create separate-gender polling stations to encourage female voting. Yet, little is known about the effectiveness of this policy. We estimate the impact of separate-gender polling stations on voter turnout using administrative data from the 2018 general election in Pakistan. We find that being assigned to separate-gender polling stations is associated with a reduction in turnout for both women and men by 2 and 1 percentage points respectively, when compared to the turnout of women and men in mixed-gender polling stations. Our results remain consistent after controlling for the role of resources, norms and mobility-related costs. Together, these results indicate that separate-gender polling stations do not alleviate gender-based voting constraints faced by women.

4.1 Introduction

There are large and persistent gender gaps in political participation in the developing world (Prillaman, 2017). Women are less likely to vote, run for office, campaign, canvass or make demands on those in government (Chhibber, 2002; Parthasarathy et al., 2017; Khan 2017). Female political participation could be hindered by gender-specific factors such as access to fewer resources, a 'gendered psyche', or the limited mobilization of women (Burns et al., 2001; Fox and Lawless, 2010; Preece, 2016;

Giné and Mansuri, 2018; Cheema et al., 2019; Iyer and Mani, 2019; Goyal, 2020). It could also be affected by non-gendered factors that affect both women and men. For example, additional costs imposed by administrative voting processes could have large effects on turnout (Downs, 1957; Brady and McNulty, 2011; de Kadt, 2017).

Governments in certain developing countries have adopted the use of separate-gender polling stations to encourage female voting (UN Women and United Nations Development Programme, 2015).[58] This policy aims to overcome constraints around safety and cultural norms faced by women by providing them with a separate space to vote. In cultures where women tend not to interact with unrelated men, having separate spaces to vote, with women staff and other facilities, is expected to increase access for women voters (UN Women and United Nations Development Programme, 2015). There is little evidence that the policy is effective in increasing female turnout.

In this paper, we study the impact of separate-gender polling stations on voter turnout using administrative data from the 2018 general elections in Pakistan. Comparing female turnout in separate-gender polling stations to female turnout in mixed-gender polling stations within a polling location, we find that separate-gender polling stations reduce female turnout by 2 percentage points.[59] A similar analysis for males shows that male turnout is reduced by 1 percentage point.

Our results also show that there is no heterogeneity in the response to separate-gender polling stations with respect to factors such as resources and norms at an aggregated level, including the role of education, awareness, mobility norms, bargaining power and partner's characteristics.

[58] Governments in Afghanistan, Algeria, Argentina, Chile, Ecuador, Egypt, Jordan, Kuwait, Lebanon, Mozambique, Pakistan, Sri Lanka, Zanzibar, among others, have adopted the use of separate-gender polling locations, polling stations and priority queues (Electoral Knowledge Network, 2013).

[59] A polling location is usually a government building, such as a school. In practice, multiple polling stations are often situated in the same polling location.

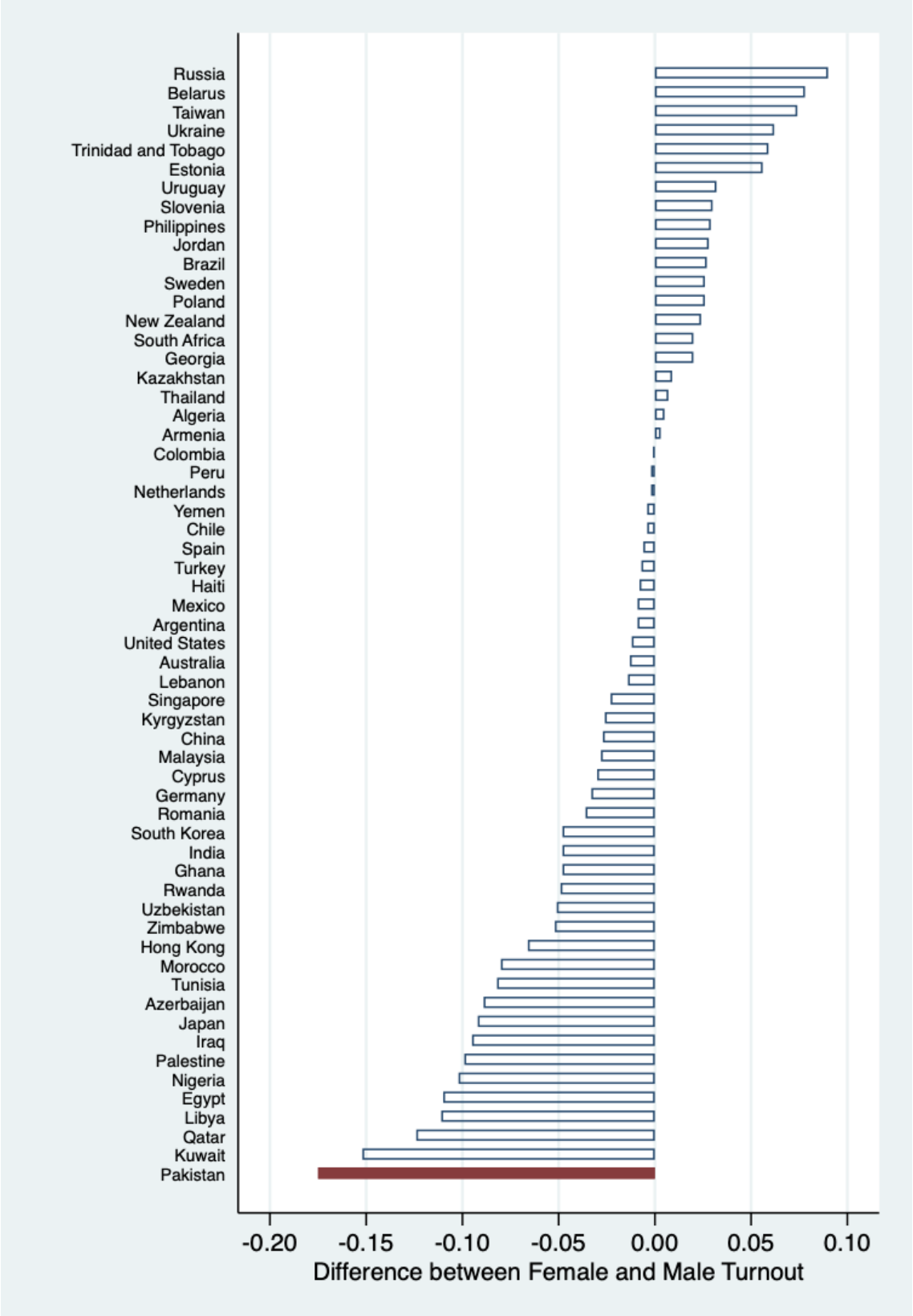
Our findings are important in understanding the constraints faced by women in voting - they suggest that the creation of separate-gender polling stations does not resolve the key constraint to low female turnout, and in fact, interacts with administrative, non-gendered constraints in a way that aggravates the turnout difference in separate-gender polling stations. Incorporating the effect of the interaction between gendered and non-gendered costs on female turnout could help better understand the constraints that prevent women and men from voting in separate-gender polling stations, such as a negative perception of separate-gender polling stations, or other norms that prevent women from voting alone. Understanding these barriers plays an important role in improving female turnout in developing countries.

4.2 Gender Gap and Creation of Polling Stations

The gap between male and female turnout in Pakistan is one of the highest in the world, and parts of the country face enormous resistance from men to women's voting (Figure 4.1). The Election Commission of Pakistan (ECP) has implemented the policy of separate-gender polling stations since the 2008 elections in order to encourage women's voting. Figure 4.2 shows that the gender turnout gap between males and females is large and persistent.

The ECP creates three kinds of polling stations – Female-Only, Male-Only, and Combined. The Female-Only and Male-Only polling stations are gender-specific, while Combined polling stations are mixed-gender. A polling station can be situated in any Government building within the constituency, or when this is unavailable, in an improvised polling station on public property, or in a building owned by a private, registered educational institution. We define 'polling location' as a building within which one or multiple polling stations might be present. In practice, multiple polling stations are situated within a given polling location. For example, 'City District Government Junior Model School' is a government school in the city of Lahore

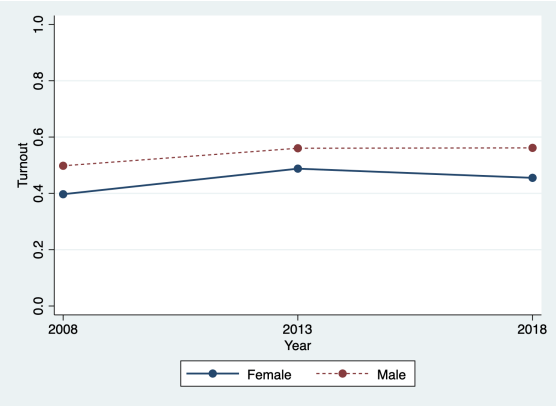
Figure 4.1: Pakistan has a Large Difference between Female and Male Turnout



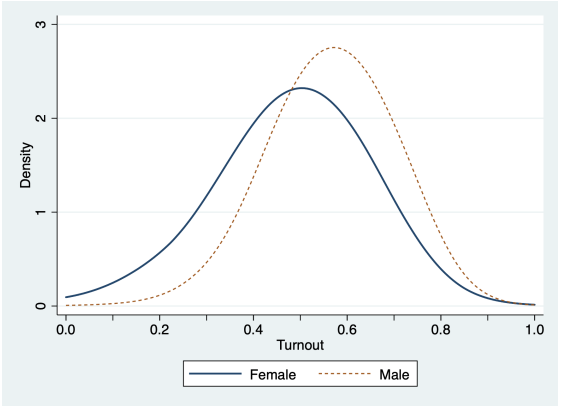
Notes: The graph uses self-reported likelihood of voting. It depicts the difference between the share of female and males in the country who report that they always vote. *Source:* World Values Survey 2014.

Figure 4.2: Gender Gap in Turnout in Pakistan

A. Persistent Gender Turnout Gap



B. Higher Male Turnout in 2018 Elections



Notes: Panel A shows female and male turnout in Pakistan's general elections in 2008, 2013 and 2018. Polling station level turnout, disaggregated by gender, was made available for the first time in the 2018 election. Turnout rates for all elections are calculated using only separate-gender polling stations for consistency. Panel B shows the distributions of female and male turnout at the polling station level in Pakistan's 2018 general elections. *Panel A Source:* Free and Fair Election Network 2009, 2013.

that had 5 polling stations in the 2018 election (2 Female-Only, 2 Male-Only and 1 Combined). Here, 'City District Government Junior Model School' is the polling location. The 2018 election dataset has 79,091 unique polling stations and 61,548 unique polling locations. Table 4.1 provides a distribution of polling locations with different combinations of separate- and mixed-gender polling stations. Table 4.2 provides the distribution of polling stations for the 2018 election by polling station type. This table shows that both separate- and mixed-gender polling stations are approximately half of the polling stations across the country.

Table 4.1: Distribution of Polling Locations with Different Polling Station

		Number of Separate-Gender Polling Stations			
		0	1	2	3+
Number of Mixed-Gender Polling Stations	0		33.49%	10.26%	1.52%
	1	47.57%	0.72%	0.63%	0.11%
	2	3.94%	0.06%	0.09%	0.03%
	3+	0.67%	0.01%	0.03%	0.01%

Notes: This table presents the distribution of polling locations with different polling station combinations. There are a total of 61,548 polling locations. 536 polling locations do not have information on polling station type (0.87% of total polling locations). Hence the total of the percentages of this table is 99.13%.

Table 4.2: Distribution of Polling Station Types

Polling Station Type	**Frequency**	**Share**
Combined	38,009	48%
Female-Only	19,424	25%
Male-Only	21,081	27%
Unknown	577	1%
Total	**79,091**	**100%**

Notes: The table presents the share of Male-Only, Female-Only and Combined polling stations in the 2018 election dataset. There are a total of 79,091 polling stations in the dataset with less than 1% observations that do not have information about the polling station type.

We refer to internal ECP documents that dictate the policy on creating polling stations. In preparing the polling scheme, the ECP allocates polling stations by census blocks. Here, four considerations are essential. First, where separate-gender polling stations are created, corresponding Male-Only and Female-Only polling stations are to be within the same polling location, as far as possible (Election Commission of Pakistan, 2018). Where this is not possible, Male-Only polling stations might be placed in a different polling location than Female-Only polling stations. Second, where Combined polling stations are created, corresponding male and female voters of a census block are to be assigned to be same Combined polling station (Election Commission of Pakistan 2018). Third, as far as possible, the ECP ensures that the distance between the polling station and the voters assigned to it is less than 1 kilometre (Government of Pakistan, 2017). This radius was adhered to in practice - a post-election assessment of polling stations found that 94 percent of surveyed individuals said that the distance between the polling stations and their place of residence was less than 1 km (Free and Fair Election Network, 2018b). Finally, no more than 1,200 voters can be assigned to any polling station (Government of Pakistan, 2017).[60] The last census for Pakistan prior to the 2018 election was conducted in 1998, which means that little updated socio-economic information was available at the time of preparing the polling scheme.[61]

[60]The final implementation of the policy in Pakistan is dictated by the availability of resources, such as female staff, etc.

[61]The allocation of voters to their polling stations is known in advance, and cannot be changed.

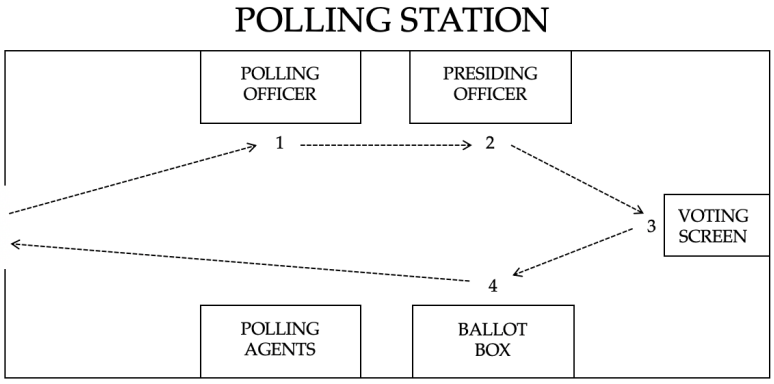
In practice, the ECP attempts to create polling stations that satisfy these four considerations within resource constraints. There are two important resource constraints that play an important part in designing the polling scheme. First, the availability of government buildings and resources (such as infrastructure and election budget) determines the number of polling stations in one location, the number of voters assigned to a particular polling stations in a given area and the distance of residents from polling stations. For instance, if a particular locality with a large population just has one government building or other appropriate public property or private property, the ECP will either have to create multiple polling stations in the government building or increase the number of voters assigned to each polling station beyond 1,200 if making additional polling stations is costly. This will also have consequences for the distance between residents and polling stations.

Second, the availability of trained female staff determines whether separate-gender polling stations can be made. If a particular locality does not have trained female polling staff, then the ECP can opt for a Combined polling station. The resource constraint plays an important role in determining the number of separate- and mixed-gender polling stations made.

Once the number of polling stations, their size and location are determined, the last step is to assign census blocks to polling stations. Here, a census block is either assigned to a Combined polling station or split up and assigned to Female-Only and Male-Only polling stations. Census blocks and polling stations have an average of 457 and 1,268 registered voters respectively. Hence, one polling station on average has voters from several census block codes. To the best of our knowledge, there is no criteria used to assign census blocks to separate- or mixed-gender polling stations.

The voting process once a voter enters a polling station is shown in Figure 4.3. Further details are given in the Appendix 4.B.

Figure 4.3: Voting Process in the 2018 General Election



Explanation: 1. Individual enters polling station with original national identity card. Polling officer verifies identity and informs polling agents. 2. Presiding officer hands ballot paper to individual. 3. Individual puts stamp on ballot paper for the candidate she/he wishes to vote for behind the voting screen. 4. Individual places ballot paper with stamp inside ballot box. 5. Individual exits the polling station.

4.3 Data

In this paper, we use an administrative dataset with polling station-level electoral information for Pakistan's most recent general election conducted on 25 July 2018 (Sonnet, 2019).[62] There are a total of 79,091 unique polling stations in the data, divided into the three types of polling stations - Female-Only, Male-Only and Combined. 48% of the total 79,091 polling stations are mixed-gender, 25% are Female-Only and 27% are Male-Only. Gender-disaggregated turnout data is available for all three polling station types.

The number of observations in the dataset is nearly twice the number of polling stations. This is because the election in 2018 was conducted at both the national and provincial levels. Individuals voted for both a National Assembly and a Provincial Assembly seat in their polling stations. As a result, polling stations are duplicated for National and Provincial Assembly seats.

We combine this data with the 2015 Pakistan Social and Living Standards Mea-

[62]Election results data at the polling station level, disaggregated by gender, is not available for previous national elections. Hence, we could not use a panel dataset for our analysis.

surement (PSLM) and the 2017 Demographic and Health Survey (DHS) to analyze mechanisms. The 2015 PSLM and the 2017 DHS are representative at the district and national levels, respectively, and contain information on income, employment, education, amongst other development outcomes.[63]

4.4 Methodology and Results

To estimate the impact of polling station type on voter turnout, we compare the turnout of a particular gender in a separate-gender polling station to the turnout of the same gender in a mixed-gender polling station (Figure 4.4). Since there can be multiple polling stations within the same location or building, for example, a school, we control for the polling location, hence, comparing only separate-gender and mixed-gender polling stations within a polling location.[64]

The empirical specification used to estimate this impact is given in Equation 4.1, which is our preferred specification. Here, $FemaleTurnout_{ijc}$ is the turnout of females in polling station i, polling location j and constituency c.[65] $Female_{ijc}$ is a dummy variable that takes the value 1 if polling station i is a Female-Only polling station, and 0 if it is a Combined polling station. μ_c and γ_{ij} represent the constituency and location fixed effects, respectively.[66] In addition to this preferred specification, we also show results for a simple OLS that does not control for constituency and/or location. We conduct a similar analysis for males.

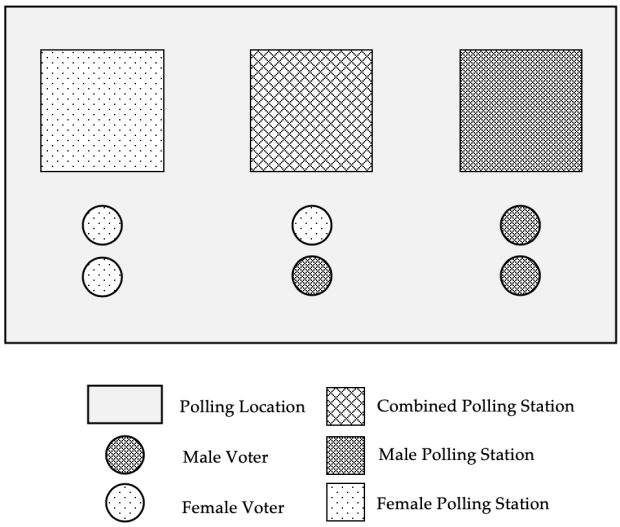
[63]The 2015 PSLM was conducted by the Pakistan Bureau of Statistics (PBS) which has made the micro-files for the survey publicly available on their website (`www.pbs.gov.pk`). The DHS is conducted by USAID in countries across the world.

[64]To identify the polling station location, we clean the names of all polling stations and then group polling stations with the same names together. Hence, we are able to control for location fixed effects. 3.77% of all polling stations are in locations that have at least one mixed-gender and at least one separate-gender polling station (2,982 polling stations out of a total of 79,091).

[65]Turnout is defined as the total number of votes cast on election day as a proportion of the total number of registered voters.

[66]In this context, provincial and national elections are held at the same time. Individuals may cast votes for either provincial or national, or both elections at their polling station. Our constituency fixed effects control for national and provincial seats separately.

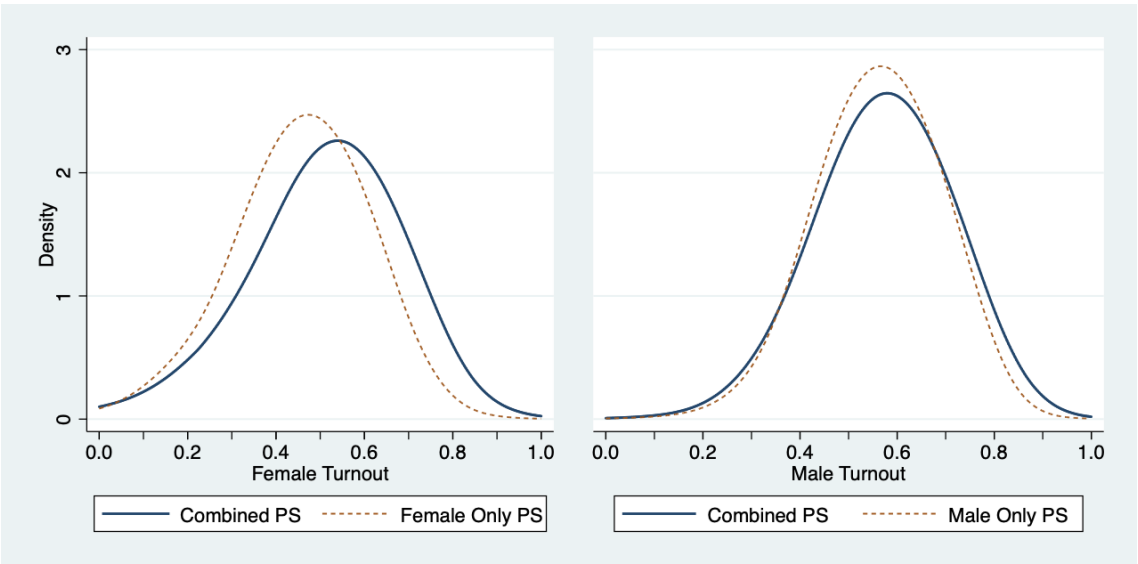
Figure 4.4: Illustration of Main Empirical Specifications



Explanation: This figure illustrates our preferred empirical specification which controls for polling location. In Equation 4.1, we compare the female voter turnout in a Female-Only polling station with the female voter turnout in a Combined polling station within the *same polling location*. Similarly, we compare male voter turnout in a Male-Only polling station with male voter turnout in a Combined polling station within the *same polling location*.

$$FemaleTurnout_{ijc} = \beta_0 + \beta_1 Female_{ijc} + \mu_c + \gamma_{ij} + \epsilon_{ijc} \tag{4.1}$$

Figure 4.5: Combined Polling Stations Have Higher Turnout than Separate-Gender Ones



Notes: The graph shows the polling station-level distributions of turnout in mixed and separate-gender polling stations for females and males in Pakistan's 2018 general elections.

It is important to highlight that the incorporation of location fixed effects significantly reduces the variation in the sample that the results rely on. This is because the results from the preferred specification with location fixed effects are from polling stations that have at least one mixed-gender and at least one separate-gender polling station in the same polling location, which represents only 3.77% of total polling stations and 1.69% of total polling locations in the dataset. In other words, the results with location fixed effects are driven by a total of 2,982 polling stations and 1,039 polling locations in the dataset. We acknowledge that this an important caveat to the results from our preferred specification. To mitigate this risk, we provide the results with and without the incorporation of location fixed effects. As can be seen from Table 4.3, the results with and without the incorporation of location fixed effects are in a similar direction which increases our confidence in the findings.

Column (3) in Table 4.3, which presents results from our preferred specification, shows that Female-Only polling stations are associated with a reduction in female turnout by 2 percentage points, on average, when compared to Combined polling stations. Male-Only polling stations are associated with a reduction in male turnout by 0.9 percentage points on average when compared to Combined polling stations. The results remain consistent across all three specifications presented.

4.5 Threats to Identification

The ECP creates polling stations across the country based on several factors such availability of infrastructure, number of registered voters, availability of female staff etc. Hence, a simple comparison between the turnout in separate- and mixed-gender polling stations has several threats to identification. This is because of possible omitted variable bias where the type of polling station might be correlated with some other factor that could explain the difference between the turnout in separate- and mixed-gender polling stations. For instance, it might be the case that there

Table 4.3: Lower Turnout in Separate-Gender Polling Stations

Panel A: Female Turnout in Female-Only vs Combined Polling Stations

Dep Var: Female Turnout	(1)	(2)	(3)
Female Only PS	-0.056***	-0.054***	-0.020***
	(0.001)	(0.001)	(0.006)
Observations	69303	69303	69303
Adjusted R^2	0.036	0.542	0.932
Constituency Fixed Effects	No	Yes	Yes
Location Fixed Effects	No	No	Yes

Panel B: Male Turnout in Male-Only vs Combined Polling Stations

Dep Var: Male Turnout	(1)	(2)	(3)
Male Only PS	-0.011***	-0.020***	-0.009**
	(0.001)	(0.001)	(0.004)
Observations	72370	72370	72370
Adjusted R^2	0.003	0.465	0.878
Constituency Fixed Effects	No	Yes	Yes
Location Fixed Effects	No	No	Yes

Notes: This table measures the impact of having separate-gender polling stations on turnout. Panel A compares the female turnout in Female-Only polling stations with that in Combined polling stations. Panel B compares the male turnout in Male-Only polling stations with that in Combined polling stations. Column (1) conducts a simple OLS regression. Column (2) controls for constituency fixed effects. Column (3) controls for location fixed effects. Each location has one national assembly and one provincial assembly constituency. Standard errors are clustered at the location level and are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

are more mixed-gender polling stations in urban areas, which have different socio-demographic characteristics when compared to rural areas. In this case, the results might be driven by the socio-demographic characteristics and would not capture the impact of the polling station type.

The empirical strategy of this paper relies on location fixed effects to mitigate the threat of omitted variables. By introducing location fixed effects, the paper compares female turnout in Female-Only polling stations with female turnout in Combined polling stations within a polling location, which is typically a government building. It is also important to mention that the ECP ensures that voters reside within 1km of the polling station - 94 percent of surveyed individuals said that the distance between the polling stations and their place of residence was less than 1 km for the 2018 election (Free and Fair Election Network, 2018b). This means that the empirical strategy of this paper compares females assigned to separate- and mixed-gender polling stations within a 1km radius. This helps mitigate the threat of omitted variables.

One threat to identification may still remain after incorporating location fixed effects - it might be the case that even within a 1km radius, the characteristics of voters assigned to different polling stations are different. We acknowledge this potential threat to inference. Our identifying assumption is that the allocation of voters to polling stations is conditionally random within these narrowly-defined locations. While we cannot test this assumption directly at present, because the data for Pakistan's census conducted in 2018 has not been released and the census before 2018 was conducted in 1998, we believe that this is a reasonable assumption since the characteristics of voters within these 1km radius locations are likely to be similar across polling stations. Furthermore, since there was no recent census data available to the government when designing the polling scheme, it would have been difficult for them to systematically assign voters to polling stations differentially. When the 2018 census data is publicly released, we will be able to test this assumption directly.

4.6 Mechanisms

To understand the mechanisms that could drive these results, we look at the role of various gendered and non-gendered factors that determine turnout and are identified in the literature. First, we test the role of factors that discourage women's voting, such as lack of access to resources or information (Burns et al., 2001; Giné and Mansuri, 2018). We use measures of education and awareness to test the role of this factor. Second, we test the role of social norms that reinforce a 'gendered psyche' (Fox and Lawless, 2010; Preece, 2016; Gottlieb, 2016; Cheema et. al, 2019). We use district-level measures of female mobility-related norms, and female bargaining power to test this channel. Finally, we look at the role of voting costs, such as the ability for males and females of a household to be able to vote in the same polling location (Downs, 1957; Brady and McNulty, 2011).

We use 30 variables from the 2015 Pakistan Social and Living Standards Measurement survey (PSLM) and from the 2017 Demographic and Health Survey (DHS) to test the role of resources and norms. In the interest of conciseness and power, these variables are aggregated at the district level, and combined into seven indices, following Anderson (2008). The seven indices of interest are male and female education, male and female awareness, mobility, bargaining power, and husband's characteristics.[67] The variables used to construct the indices are shown in Table 4.4.

Higher levels of education or awareness might mean that individuals are more politically involved or active. Politically inclined individuals, in turn, might be more likely to vote and, hence, less likely to be affected by the policy of separate polling stations. Lower female bargaining power, or less equitable mobility and gender norms could be important if the policy of separate-gender polling stations encourages females in conservative areas to vote, especially if the demand for separate-gender polling stations comes from the women in conservative areas themselves. Alternatively, it

[67]The results for the individual variables are consistent with the indices.

might have the opposite impact if the demand for this policy does not represent the constraints faced by the women. Finally, the characteristics of husbands might have an impact on women's constraints. For instance, more educated husbands might be associated with higher mobility of their wives. Detailed information on the definitions of the variables can be found in Appendix 4.C.

We test the role of various indices in explaining the results by running the empirical specification in Equation 4.2. Here, $FemaleTurnout_{ijdc}$ is the female turnout in polling station i, polling location j, constituency c and district d. $Index_d$ represents the district level value of the index variable explained above. We run a similar analysis for males. With the inclusion of location fixed effects, the interaction term $Female_{ijcd} * Index_d$ compares polling locations across districts with higher and lower values of the respective index. This helps us check whether the type of polling station is associated with differential turnout as the index changes across districts.

$$FemaleTurnout_{ijcd} = \beta_0 + \beta_1 Female_{ijcd} + \beta_2 Female_{ijcd} * Index_d + \mu_c + \gamma_{ij} + \epsilon_{ijcd} \quad (4.2)$$

We also examine the costs of voting or mobility-related costs, by testing whether males and females being able to vote together drives the results. When mixed-gender polling stations are created, an entire census block with all its females and males is assigned to that polling station. However, when the ECP creates separate-gender polling stations, a census block is divided into two polling stations, one for females and one for males. While the corresponding female and male polling stations may be located in the same polling location, this is not always the case. Combining our data with information on the assignment of census blocks to polling stations shows that where blocks are allocated into gender-specific polling stations, only 32 percent of the corresponding Male-Only and Female-Only polling stations are in the same polling location.

Table 4.4: Summary of Indices Used for Analysis

#	Index	Types of Variables Used
1	Education	Female and male ability to read, school attendance, highest level of education attained
2	Awareness	Female and male consumption of newspapers, TV, radio, mobile and internet
3	Mobility	Norms against free mobility of women, including the need for permission to go to a doctor and acceptability of being hit for not taking permission
4	Bargaining Power	Decision-making power of women in different household decisions, including the use of contraception and household spending
5	Husband	Husband's education and labor force participation

Notes: This table provides a brief summary of the different indices created to various mechanisms. The Education Index uses data from the Pakistan Social and Living Standards Measurement survey conducted in 2015. All other indices use data from the Demographic and Health Survey conducted in 2017. For further details about the indices, please refer to the Appendix 4.C.

We test the role of mobility-related characteristics by computing the variable $Corresp\ Male_{ij}$, which tells us the share of all block codes in the polling station that are assigned to the same polling location for males. The coefficient on the interaction term $Female_{ij} * CorrespMale_{ij}$ tells us whether the difference in female turnout in Combined and Female-Only polling stations is different when corresponding Male-Only polling stations are in the same polling location. We run a similar analysis for males.

$$FemaleTurnout_{ijc} = \beta_0 + \beta_1 Female_{ijc} + \beta_2 Female_{ijc} * CorrespMale_{ij} + \mu_c + \gamma_{ij} + \epsilon_{ijc} \tag{4.3}$$

The results of the abovementioned analysis (Table 4.5) shows that the interaction terms are statistically insignificant at the 10 percent level. This implies that districts with different values of various indices have a similar association between

the type of polling station and turnout. The statistically insignificant results for 'Female*Corresponding Male' and 'Male*Corresponding Female' show that the results are not being driven by females and males wanting to vote together in the same polling location. By ruling out several potential mechanisms, the analysis tentatively point towards the possibility of the results being driven by a feature of separate- or mixed-gender polling stations that is common throughout the country and does not vary by norms, resources or voting together.

4.7 Discussion and Conclusion

Together, these results suggest that separate-gender polling stations do not address the gender-based constraints to voting. To be effective, the policy of separate-gender polling stations would overcome the resource-, norms- or mobilization-based constraints faced by women that are highlighted in the literature. Having separate spaces for women to vote, could in theory, address the 'gendered psyche' by reducing men's resistance to female voting, or making women feel more comfortable in the public sphere. However, our analysis shows that separate-gender polling stations do not alleviate such constraints.

Instead, our results suggest that separate-gender polling stations increase non-gendered costs of voting for both females and males, and consequently, reduce turnout of both genders. Our analysis shows that the results are consistent across different levels of education, awareness, female bargaining power, and mobility norms in the country. This suggests that the results are driven by a feature of these polling stations in themselves. The administrative implementation of the policy of separate-gender polling stations could interact with the gendered component of the policy, and consequently, have large effects on turnout. While we can control for the effects of placement of separate-gender polling stations by looking at within-polling location differences, there might be other systematic differences in the manner this policy is

Table 4.5: Results are Consistent after Including Role of Resources, Norms, Voting Together

Panel A: Female Turnout in Female-Only vs Combined Polling Stations

Dep Var: Female Turnout	(1)	(2)	(3)	(4)	(5)	(6)
Female Only PS	-0.024***	-0.020***	-0.021***	-0.020***	-0.021***	-0.021***
	(0.009)	(0.007)	(0.008)	(0.007)	(0.007)	(0.008)
Female*Corresponding Male	0.005					
	(0.008)					
Female*Education Index		0.000				
		(0.009)				
Female*Awareness Index			0.005			
			(0.018)			
Female*Mobility Index				-0.001		
				(0.012)		
Female*Bargaining Power Index					-0.018	
					(0.021)	
Female*Husband Index						0.005
						(0.012)
Observations	69303	68293	44692	44692	44692	44692
Adjusted R^2	0.932	0.931	0.937	0.937	0.937	0.937
Constituency Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Location Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes

Panel B: Male Turnout in Male-Only vs Combined Polling Stations

Dep Var: Male Turnout	(1)	(2)	(3)
Male Only PS	-0.011**	-0.010**	-0.008
	(0.005)	(0.005)	(0.005)
Male*Corresponding Female	0.003		
	(0.006)		
Male*Education Index		0.002	
		(0.006)	
Male*Awareness Index			0.004
			(0.011)
Observations	72370	71273	47026
Adjusted R^2	0.878	0.874	0.871
Constituency Fixed Effects	Yes	Yes	Yes
Location Fixed Effects	Yes	Yes	Yes

Notes: This table tests for potential mechanisms. All specifications control for constituency and location fixed effects. Each location has one national assembly and one provincial assembly constituency. Standard errors are clustered at the location level and are reported in parentheses. For a detailed description of the indices, please refer to the Appendix 4.C. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

implemented, in terms of staff or other resources made available to separate-gender polling stations, or in terms of the general or historical perceptions of such polling stations.

Our research opens up three avenues for future work. First, it is important to understand whose preferences the policy of separate-gender polling stations represents. It might be the case that this policy is being driven by demands from men or policymakers, rather than addressing the actual constraints faced by women.

Second, further research is required to understand the exact administrative features of separate-gender polling stations that decrease female and male turnout. Systematic differences in implementation could have a large impact on electoral outcomes (Neggers, 2018). Future work could help uncover the mechanisms at play, which would help our understanding of the administrative determinants of voter turnout in developing countries.

Finally, future research could test the effectiveness of this policy in other contexts where it is implemented. It might be the case that the impact of this policy in conflict-affected areas such as Afghanistan and Yemen is different from its impact in places where the implementation is driven by concerns regarding women's safety or harassment by men, for example, or where it is driven by cultural and religious norms.

Understanding the impact of the policy of creating separate-gender polling stations on women's turnout behaviour is important for governments and electoral bodies around the world. It can help design policies that encourage the participation of women in the political arena and lead to better outcomes for all.

References

Anderson, Michael L., "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects," *Journal of the American Statistical Association*, 2008, *103* (484), 1481–1495.

Brady, Henry E. and John E. McNulty, "Turning Out to Vote: The Costs of Finding and Getting to the Polling Place," *American Political Science Review*, 2011, *105* (1), 115–134.

Burns, Nancy, Kay Lehman Schlozman, and Sidney Verba, *The Private Roots of Public Action: Gender, Equality, and Political Participation*, Cambridge, MA: Harvard University Press, 2001.

Cheema, Ali, Sarah Khan, Shandana Khan-Mohmand, and Asad Liaqat, "Canvassing the Gatekeepers: A Field Experiment to Increase Women's Electoral Turnout in Pakistan," *Working Paper*, 2019.

Chhibber, Pradeep, "Why are Some Women Politically Active? The Household, Public Space, and Political Participation in India," *International Journal of Comparative Sociology*, 2002, *43* (3-5), 409–429.

de Kadt, Daniel, "Bringing the Polls to the People: How Electoral Access Encourages Turnout But Exacerbates Political Inequality," Technical Report SSRN 2922125, Social Science Research Network, Rochester, NY 2017.

Downs, Anthony, "An Economic Theory of Political Action in a Democracy," *Journal of Political Economy*, 1957, *65* (2), 135–150. Publisher: University of Chicago Press.

Election Commission of Pakistan, "Preparation of Polling Scheme for General Elections - 2018 (Order No. F. 7(1)/2018 - Cord)," 2018.

Electoral Knowledge Network, "Systems with Separate Polling Stations for Men and Women," 2013.

Fox, Richard L. and Jennifer L. Lawless, "If Only They'd Ask: Gender, Recruitment, and Political Ambition," *The Journal of Politics*, 2010, *72* (2), 310–326.

Free and Fair Election Network, "Pakistan General Elections 2008: Election Observation Summary and Recommendation for Electoral Reforms," Technical Report, Islamabad, Pakistan 2009.

_ , "General Election 2013 - FAFEN Observation," Technical Report, Islamabad, Pakistan 2013.

_ , "Analysis of Voter Turnout in GE-2018," Technical Report, Islamabad, Pakistan 2018.

_ , "General Election 2018 Polling Stations Assessment Report," Technical Report, Islamabad, Pakistan 2018.

Giné, Xavier and Ghazala Mansuri, "Together We Will: Experimental Evidence on Female Voting Behavior in Pakistan," *American Economic Journal: Applied Economics*, 2018, *10* (1), 207–235.

Gottlieb, Jessica, "Why Might Information Exacerbate the Gender Gap in Civic Participation? Evidence from Mali," *World Development*, 2016, *86*, 95–110.

Government of Pakistan, "Elections Act," 2017.

Goyal, Tanushree, "How Women Mobilize Women Into Politics: A Natural Experiment in India," *SSRN Electronic Journal*, 2020.

ICF, "Demographic and Health Surveys Standard Recode Manual for DHS7," Technical Report, Rockville, Maryland, USA 2018.

Inglehart, R., C. Haerpfer, A. Moreno, C. Welzel, K. Kizilova, J. Diez-Medrano, M. Lagos, P. Norris, E. Ponarin, and B. Puranen, "World Values Survey: Round Six - Country-Pooled Datafile," 2014.

Iyer, Lakshmi and Anandi Mani, "The Road not Taken: Gender Gaps along Paths to Political Power," *World Development*, 2019, *119*, 68–80. Publisher: Elsevier Ltd.

Khan, Sarah, "What Women Want: Gender Gaps in Political Preferences," in Matt Golder and Sona Golder, eds., *Symposium: Women/Gender and Comparative Politics. CP: Newsletter of the Comparative Politics Organized Section of the American Political Science Association.*, Vol. 27(1) 2017, pp. 42–49.

Neggers, Yusuf, "Enfranchising Your Own? Experimental Evidence on Bureaucrat Diversity and Election Bias in India," *American Economic Review*, 2018, *108* (6), 1288–1321.

Pakistan Bureau of Statistics, "Pakistan Social And Living Standards Measurement Survey (PSLM) 2014-15: Provincial/District Dataset," Technical Report, Islamabad, Pakistan 2016.

Parthasarathy, Ramya, Vijayendra Rao, and Nethra Palaniswamy, "Deliberative Inequality: A Text-As-Data Study Of Tamil Nadu's Village Assemblies," Policy Research Working Papers 8119, The World Bank 2017.

Preece, Jessica Robinson, "Mind the Gender Gap: An Experiment on the Influence of Self-Efficacy on Political Interest," 2016, *12* (1), 198–217.

Prillaman, Soledad Artiz, "Strength in Numbers: How Women's Groups Close India's Political Gender Gap," *Working Paper*, 2017.

Sonnet, Luke, "2018 Pakistani General Election Polling Station Data," 2019.

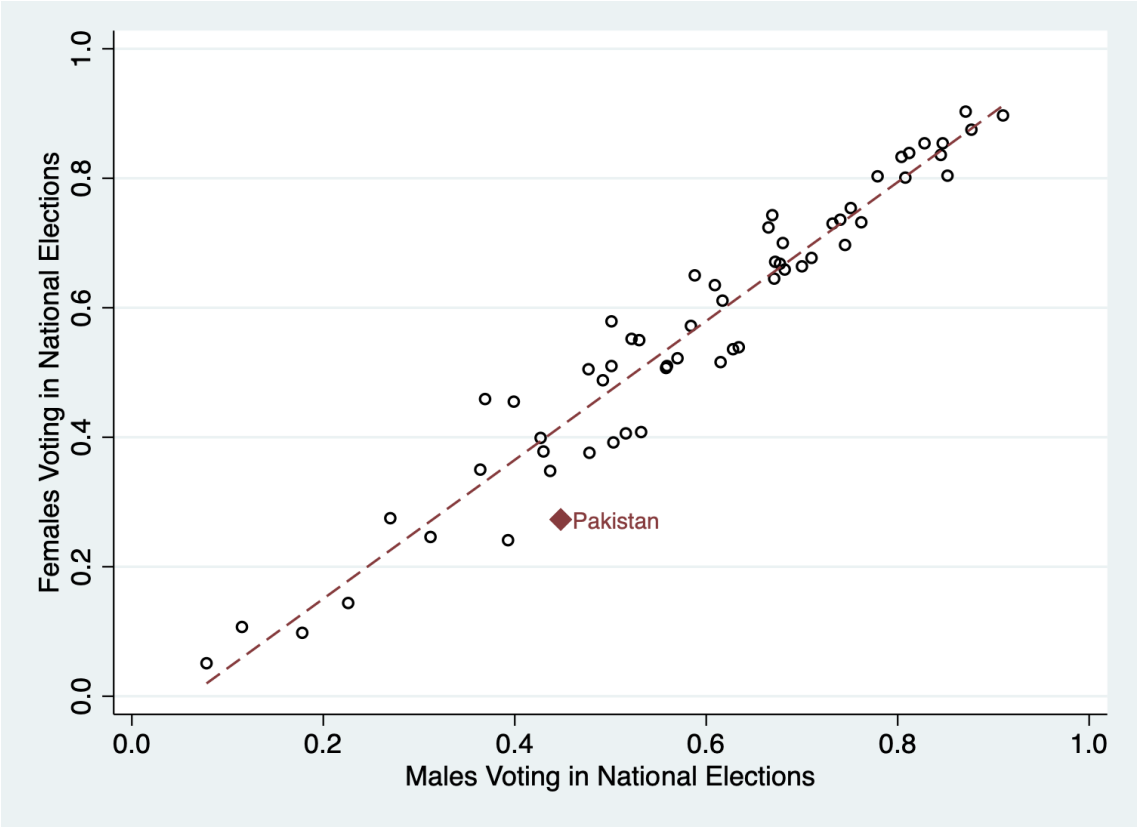
UN Women and United Nations Development Programme, "Inclusive Electoral Processes: A Guide for Electoral Management Bodies on Promoting Gender Equality and Women's Participation," Technical Report 2015.

Appendix

4.A Gender Turnout Gap

Figure 4.A.1 presents a scatter plot of the average share of females and males in the country who say they always vote. While males are always more likely to vote compared to females, Pakistan remains an outlier with one of the largest gender turnout gaps in the world.[68]

Figure 4.A.1: Males are More Likely to Vote in National Elections than Females



Notes: The graph uses self-reported answers on the likelihood of voting. It shows the percentage of males and females who report that they always vote. The dashed line is the line of best fit. *Source:* World Values Survey, 2014

[68] While survey data such as the one used here runs the risk of over-reporting turnout, it allows us to make cross-country comparisons.

4.B Voting Process

The voting process for the 2018 general election in Pakistan is given in the Elections Act 2017. On election day, voters are required to bring their original national identification card (CNIC) for identity verification.[69] No other form of identification (e.g. passport, drivers license) is accepted. Individuals can vote with their original CNIC even if it has expired. Women and men usually have separate waiting queues in Combined polling stations.

Upon entering the polling station, individuals are required to go to the desk of the polling officer and present their CNIC. The polling officer will verify the details of the individual given on the CNIC with a list of registered voters. The individual is required to show her/his face so that the polling officer can verify their identity. If the name of the individual is present in the voter list, then the polling officer will call out the name and the serial number of the voter in the voting list out loud so that the polling agents seated close by can hear this and keep a track of the voters on their list. Each candidate can assign two polling agents per polling station. The polling agents represent the candidates and act as a check on the administrative staff. Then the polling officer will mark the right thumb of the individual with indelible ink which serves as a visual identifier of having voted. After this, the polling officer will ask the individual to place a thumb print on the voter list which is proof that the individual voted.

Once the polling officer has crossed off the individuals name from the voter list, the individual will then move towards the adjacent desk where the assistant presiding officer will hand over the ballot paper after taking a thumb print as receipt, and stamping the back of the ballot paper.[70] The assistant presiding officer will give the individual a stamp and keep her/his CNIC. Then the individual goes behind the voting screen to mark a stamp on the candidate that the individual wants to vote

[69] Copies of the CNIC are cannot be used for identity verification.

[70] The are two assistant presiding officers, one for the national assembly seat and one for the provincial assembly seat. Both hand over separate ballot papers for both seats to the individual.

for. After this, the individual will return the stamp to the presiding officer, collect the CNIC, and put the voting ballot paper in the ballot box.[71]

[71] There are two ballot boxes, one for the national assembly and one for the provincial assembly.

4.C Index Definitions and Descriptions

Female Indices

Education Index: Education could be an important mechanism that could help explain the main results. To test this channel, we created an Education Index using three education-related variables from the 2015 Pakistan Social and Living Standards Measurement survey (PSLM), namely, $Read_d$, $Attend_d$ and $Level_d$. $Read_d$ is the percentage of women in district d who know how to read, $Attend_d$ is the percentage of women in district d who have attended school and $Level_d$ is the average highest grade/level of education for women in district d. These three variables are used to create the Education Index for district d and combined with the administrative data of the 2018 election results.

Awareness Index: We create an Awareness Index using six variables from the 2017 Demographic and Health Survey, namely, $News_d$, $Radio_d$, TV_d, $Mobile_d$, $Internet_d$ and $InternetUsage_d$. $News_d$ is the percentage of women who read the news at least once a week in district d. $Radio_d$ is the percentage of women who listen to the radio at least once a week, TV_d is the percentage of women who watch TV at least once a week, $Mobile_d$ is the percentage of women who own a mobile phone, $Internet_d$ is the percentage of women who have used the internet at least once in a year and $InternetUsage_d$ is the percentage of women who have used the internet at least once in the last week. These six variables are used to create the Awareness Index for district d and combined with the administrative data of the 2018 election results.

Mobility Index: To test how female mobility and the norms around female mobility might explain the main results, we create a Mobility Index. For this purpose, we extract three mobility-related variables from the 2017 Demographic and Health Survey for Pakistan, namely, $Permission_d$, $Beating_d$ and $Alone_d$. $Permission_d$ is

the percentage of married women in district d who report that getting permission from their partners to go to the doctor will be a problem. $Beating_d$ is the percentage of women in district d who believe that it is okay for their partner/husbands to hit them if they go outside without getting permission from their partner/husband. $Alone_d$ is the percentage of women in district d who report that wanting to go alone to the doctor would be a big issue. These three variables are used to create the Mobility Index for district d and combined with the administrative data of the 2018 election results.

Bargaining Power Index: We create the Bargaining Power Index by using seven variables from the 2017 Demographic and Health survey. $HusDecidesCont_d$ is the percentage of women whose husbands exclusively decide the use of contraceptives in district d, as a proxy for decision-making within the household. $HusDecidesNoCont_d$ is the percentage of women who report that their husbands exclusively decide not to use contraceptives, $OthersDecideHealth_d$ is the percentage of women who do not have a role in the decision about their own health, $OthersDecidePurchase_d$ is the percentage of women who do not have a role in making the decision about household purchases, $OthersDecideVisit_d$ is the percentage of women who do not have a role in making the decision about visiting relatives and $HusbandMoney_d$ is the percentage of women who have no role in making the decision about the way their husband's money is spent. $Worked_d$ is the percentage of women who did not work in the last 12 months. These seven variables are used to create the Bargaining Power Index for district d and combined with the administrative data of the 2018 election results.

Husband Index: We test the channel of husband's characteristics to see if it might help explain the main results for females. We test this channel by creating a Husband Index using two variables from the 2017 Demographic and Health survey, namely, $HusbandEdu_d$ and $HusbandWork_d$. $HusbandEdu_d$ is the average highest level of husband's education in district d. $HusbandWork_d$ is average percentage of

husbands who worked in the last 12 months. These two variables are used to create the Husband Index for district d and combined with the administrative data of the 2018 election results.

Male Indices

Education Index: Similar to females, we also create an Education Index for males to test for underlying mechanisms. The Education Index for males uses three education-related variables from the 2015 Pakistan Social and Living Standards Measurement survey (PSLM), namely, $Read_d$, $Attend_d$ and $Level_d$. $Read_d$ is the percentage of men in district d who know how to read, $Attend_d$ is the percentage of men in district d who have attended school and $Level_d$ is the average highest grade/level of education for men in district d. These three variables are used to create the Education Index for district d and combined with the administrative data of the 2018 election results.

Awareness Index: We create an Awareness Index for males to test the awareness channel. The index uses six variable from the 2017 Demographic and Health Survey, namely, $News_d$, $Radio_d$, TV_d, $Mobile_d$, $Internet_d$ and $InternetUsage_d$. $News_d$ is the percentage of men who read the news at least once a week in district d. $Radio_d$ is the percentage of men who listen to the radio at least once a week, TV_d is the percentage of men who watch TV at least once a week, $Mobile_d$ is the percentage of men who own a mobile phone, $Internet_d$ is the percentage of men who have used the internet at least once in a year and $InternetUsage_d$ is the percentage of men who have used the internet at least once in the last week. These six variables are used to create the Awareness Index for district d and combined with the administrative data of the 2018 election results.

4.D Data

A summary of the different sources of data used in this paper and their use is given in Table 4.D.1.

Table 4.D.1: Data Sources Used for Analysis

#	Source	Description	Use
1	Administrative Election Results Data (2018)	Dataset with the 2018 election results at the polling station level	Produce the main results of the paper showing the impact of separate gender polling stations on turnout, explore mechanisms
3	Pakistan Social and Living Standards Measurement (2015)	Survey representative at the national and district level with various development outcomes, conducted by the Pakistan Bureau of Statistics (PBS)	Analyze various mechanisms
4	Demographic and Health Survey (2017)	Nationally representative survey with various development outcomes, conducted by USAID	Analyze various mechanisms

Notes: The table presents a summary of different datasets that are used for the analysis conducted in this paper. It provides a brief description of the datasets and their specific use in the analysis.

4.E Polling Station Descriptives

Table 4.E.1: Distribution of Polling Stations by Polling Location

	Number of Polling Stations	**Share**
1 Polling Stations per Location	50,399	67%
2 Polling Stations per Location	11,149	15%
3+ Polling Stations per Location	14,140	19%
Total	**75,688**	**100%**

Notes: The table presents the distribution of polling stations with shared locations for the 2018 General Election conducted in Pakistan. It uses the administrative election results data at the polling station level that is used for the main analysis in this paper.

Table 4.E.2: Share of Different Types of Polling Stations by Province

Province	**Male-Only**	**Female-Only**	**Combined**	**Observations**
Balochistan	30%	25%	45%	8,338
FATA	24%	19%	57%	1,160
Federal Capital	46%	45%	9%	570
KP	34%	29%	37%	22,906
Punjab	27%	25%	48%	87,932
Sindh	22%	20%	58%	34,550
Total				**155,456**

Notes: The table presents the share of Male-Only, Female-Only and Combined polling stations in the 2018 election dataset by province. It also shows the total observations per province. The total number of observations with details about the polling station type across all provinces are 155,456. There are 1,150 observations that do not have details about the polling station type, and the total number of observations in the dataset is 156,606.

Chapter 5

CONCLUSION

This dissertation answers academic and policy-relevant questions with the help of administrative data. Chapter 2 develops a new methodology to estimate income inequality for developing countries that addresses key limitations of the leading existing method. Chapter 3 analyzes the effectiveness of deadline extensions as a tax broadening tool in the developing world. Chapter 4 examines the impact of separate gender polling stations on voter turnout. These three articles have policy implications and open up further research questions, discussed below.

Policy Implications

Chapter 2 has three policy distinct implications. First, developing countries that simply rely on survey data to estimate income inequality are likely to underestimate inequality. This can have adverse consequences for the design of economic policies. Hence, the developing world needs to be wary of using only survey data to capture top incomes. Second, when developing countries do rely on a combination of survey and tax data to estimate income inequality, the choice of the method needs to be carefully thought out. For countries with low levels of tax compliance, the existing method of top income adjustment leads to counter-intuitive estimates. Chapter 2 provides a tool for such low tax compliance countries and uses administrative data that should in theory, be easily available with the government. This tool has enough flexibility that it can be modified and refined according to the data availability of any developing country. Third, the chapter shows descriptively that income inequality in Pakistan has remained high and stable, which has significant policy implications for the country.

Chapter 3 has two different policy implications. First, governments in the developing world need to be wary of using ineffective policies such as deadline extensions as a tool to broaden the tax base. While there might be conditions under which extensions could help bring additional people into the tax net, the evidence in Chapter 3 advises caution. The results are likely to be similar in other contexts where extensions are consistently given out and are not combined with disincentives in future years. A stand-alone policy of extensions with low tax enforcement will likely lead to delayed filing without an expansion in the tax base. Second, the findings point towards the importance of producing timely evidence to see if the policy is working and iterate accordingly. With timely evidence, policy-makers could have seen that extensions do not expand the tax base and tried out a new policy.

Chapter 4 has two different policy implications. First, well intentioned policies that might seem technically sound at the outset can backfire. Providing women a separate space to vote might make logical sense, but could entail other costs to voting. On the other hand, such policies might not address actual constraints faced by women but the perceived constraints by policymakers. It is hence important to ground policy tools in rigorous evidence of the actual constraints. It is equally important to make sure that the implementation of such policies is carefully thought out. Second, policies aimed to improve outcomes for one group can have unintended consequences for another group. Chapter 4 shows that separate-gender polling stations not only reduce voter turnout for females but also for males. Males were not a target of the policy and yet responded to the policy.

Future Research

Chapter 2 paves the way to refine the inequality estimation technique for developing countries further. The article matches corresponding income quantiles from survey and tax data. In the presence of more detailed data, this matching can be done

more closely to reduce potential bias in the estimates. For instance, a government could in theory match incomes by particular sub-groups to get more granular details about the income distribution. In its current form, the methodology explained in Chapter 2 works well at an aggregated national level, but would be inadequate for sub-group analysis. Future researchers could push the methodological discussion forward by fine-tuning the matching method according to the analysis.

In theory, the method proposed by Chapter 2 could work better than the existing method even in developing countries where tax compliance is high. This is because the new method has weaker assumptions and a simple way of matching that can be implemented in most developing countries. However, the article is not able to show this because it focuses on one country. We know from Chapter 2 that the new method works better than the alternative method in low tax compliance settings, but researchers could also test whether it works better for developing countries with better tax compliance. Such work will go a long way in helping researchers and practitioners pick the best method for their context.

The findings from Chapter 3 raise further questions about deadline extensions. While the article shows that extensions are ineffective in broadening the tax net, there might be conditions under which such extensions work. For instance, the policy might be effective if these extensions are given on a one-off basis and are coupled with future disincentives to file late. In some ways, extensions are like amnesty schemes on the penalty for non-filing or late-filing. An amnesty is unlikely to work if it is given consistently and is not attached to a future credible threat. If there are other contexts where a different version of extensions is given, researchers could analyze them to help supplement our understanding.

Chapter 3 also points towards the importance of analyzing tax broadening policies that are regularly used by governments in the developing world. Such policies are often experimented with and are usually not analyzed in a rigorous manner. The

current literature in tax enforcement and broadening leaves significant room to analyze these policies. Chapter 3 only takes a small step in this direction. Future research could take further steps to take the discussion forward.

Chapter 4 leaves room for further research that could pin down the exact feature of separate-gender polling stations that reduces turnout. The article rules out a plethora of possible underlying mechanisms and points towards a particular feature of separate-gender polling stations that reduces turnout. Future research could pinpoint these exact features and help explain the results further.

Another question for future research is whether the policy of separate-gender polling stations represent actual constraints faced by females or those perceived by policymakers/male members of the household. If separate-gender polling stations target actual constraints, then the results could be explained by implementation flaws, while if these are constraints perceived by others, the policy itself is not designed well. Such research could go a long way to help us understand decision making in policy settings and about the actual constraints to voting.

Administrative data provides researchers with exciting opportunities to empirically test large academic and policy relevant questions. One of the objectives of this dissertation was to provide three illustrative examples of answering such large questions that could foster win-win collaborations between academics and policymakers. Increasing collaboration with proprietary data in the developing world will open up new possibilities for researchers and policymakers. One of the key advantages of administrative data is that the same dataset can be used to answer multiple questions. While this dissertation uses the proprietary tax data provided by the Government of Pakistan to analyze questions of income inequality and tax enforcement, many other questions can also be answered. I discuss four types of questions that researchers could analyze below.

First, the tax data has the unique property of containing the actual date of submis-

sion of all tax returns filed from 2007-2018. This rich data could be used to conduct fine-grained temporal analysis that relies on dates. This could allow researchers to analyze tax filing patterns in detail. Some examples of questions that could be answered are: testing if individuals file more taxes after a major patriotic event, analyzing whether conflict (e.g. terrorism) has any impact on tax filing behavior and simple descriptive analysis of when people prefer to file taxes. These questions could help us understand tax filing behavior in more detail.

Second, similar to the empirical design for Chapter 3, the tax data could be used to analyze the impact of policy shocks between 2007 and 2017 that affected certain types of taxpayers, but not others. For instance, researchers could analyze the impact of changes in tax rates for a certain class of taxpayers on tax revenue and income declarations. The tax rates have been changed on numerous occasions from 2007 to 2017 which would provide good variation for empirical analysis.

Third, researchers could test whether income shocks lead to changes in tax declarations. In a world with high levels of tax compliance and low evasion, we should see income shocks translating into tax declarations after adjusting for exemptions. However, in countries will low levels of tax compliance, this might not be the case. Such an exercise would be able to give us with a better sense of tax evasion. The hetrogeneities of the response of different sectors/type of taxpayers will help us understand the factors that are important in controlling evasion. For instance, if sector X's income declarations respond to income shocks and sector Y's do not, then the next question would be to conduct a comparative analysis to see what factors contribute towards this variation in response.

Finally, an exciting question could be to analyze the impact of mass information provision of tax avoidance on tax compliance. The best example of this would be the Panama Papers leak that contained details of shell companies and tax avoidance by individuals across the world. Being a significant exogenous information shock,

researchers could look into how such a shock modified peoples tax compliance. The key challenge in this research would be to create an information shock that varies across time and taxpayers. The dataset could also be combined with publicly available information on politician's tax declarations to see if the shock had any impact on them that was different from other people.

To conclude, this dissertation opens up an exciting research agenda for the future. I hope to continue down this exciting path and contribute towards answering such questions.